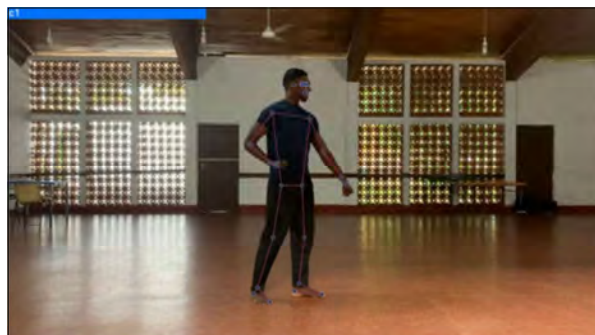# Journal

of the

National Science Foundation of Sri Lanka

JOURNAL OF THE
NATIONAL SCIENCE FOUNDATION
OF SRI LANKA

C O N T E N T S

Guidelines for Contributors

Journal
of the
National Science Foundation
of Sri Lanka

Cover: Cover caption – Computer vision based evaluation of Taekwondo Poomsae movements (top) and badminton smash shots (bottom) using skeleton points
See *J.Natn.Sci.Foundation Sri Lanka* 2024 **52**(1): 125 – 156

EDITORIAL

The culture of laziness

The development of science and technology has enabled humans to carry out more or less effortlessly many activities which previously would have entailed much physical effort and expenditure of time. While this development enables people to enjoy more time for leisure, it has also encouraged a culture of physical laziness.

The same trend appears in the intellectual field as well. Modern computers can in a few moments, carry out tasks which would have entailed an enormous amount time and mental effort, earlier. Again, this saving of time enables one to focus on the more creative and insightful activities that are specific to human beings.

However, the culture of laziness is widespread. It would appear that there is at the present time, a tendency to even attempt to transfer creative intellectual efforts to machines, through the use of generative AI. The writing up of a good piece of research for publication is one of the creative and enjoyable activities afforded to a scientist. While the use of generative AI to obtain help in deciding the most effective way of "saying what one has to say" can be considered acceptable in situations where the writer's language skills are not well developed, to attempt to use generative AI to decide on the content of the paper, "what to say", is to attribute to generative AI an ability it does not have.

Recently, scientists and sociologists have written about the deficiencies and dangers of this approach to creating knowledge and understanding. To be meaningful, any generative AI output needs to be moderated and evaluated by human intelligence.

Ajit Abeysekera

RESEARCH ARTICLE

Germplasm Characterization

Variation in plant morphology and leaf essential oil composition of a representative *Cinnamomum verum* collection from Sri Lanka[†]

MR Prathibhani[1], R Azad,[1] S Ranaweera[2], L Jayasekara[3], RAAK Ranawaka[4], G Senanayake[5], S Abeynayake[6] and S Geekiyanage[5*]

[1] *Board of Study in Agriculture, Faculty of Graduate Studies, University of Ruhuna, Matara, Sri Lanka.*
[2] *Department of Chemistry, Faculty of Science, University of Ruhuna, Sri Lanka.*
[3] *Department of Mathematics, Faculty of Science, University of Ruhuna, Sri Lanka.*
[4] *Department of Cinnamon Development, Gunarathna Weerakoon Mawatha, Borakanda, Karanadeniya, Sri Lanka.*
[5] *Department of Agricultural Biology, Faculty of Agriculture, University of Ruhuna, Mapalana, Kamburupitiya, Sri Lanka.*
[6] *Department of Animal, Plant and Soil Sciences, La Trobe University, Melbourne, Australia.*

Abstract: Sri Lankan cinnamon germplasm is an underexploited genetic resource of high breeding potential. A representative collection of a cultivated cinnamon germplasm with 48 accessions, is established at the University of Ruhuna, Sri Lanka. From that collection, 40 three-year-old vegetative-propagated accessions, were used in this study to study the variations in plant morphology and leaf oil composition. Flush colour was determined as four main categories of pink, red, brown, and green in all 48 accessions. Leaf length, leaf width, leaf length to width ratio and petiole length varied from 7.9 to 14.5 cm, 3.8 to 7.3 cm, 1.8 to 3.07 cm, and 1 to 2.2 cm, respectively, in the selected 40 accessions. Both the good fragrant aroma category of bark fragrance and the sweet pungent category of bark taste were recorded in 21 of the 40 accessions. Twenty (20) accessions from above collection were selected for leaf oil analysis. Gas chromatography-mass spectrometry (GC-MS) of leaf oil revealed the presence of 146 chemical compounds. Eugenol was the major compound of 17 accessions (52.2% to 79.5%). Two accessions with zero eugenol contained 86.8% and 91.9% of benzyl benzoate (BB), in contrast to that of 0%-0.65% from other accessions. One accession produced 16.6% eugenol and 22.3% BB. A green flush colour associated with a higher BB percentage suggested a potential morphological marker for BB. The PCA for 20 accessions explained 88.11% of total variance. The cinnamon accessions with high BB, KA11 and GB17, were clustered together, and HB12 was isolated on a different cluster. Positive correlations were detected between linalool and β-caryophyllene (0.649), linalool and BB (0.770) ($p < 0.01$), and negative correlations between eugenol and linalool (-0.630) and eugenol and BB (-0.886) ($p < 0.01$).

Keywords: *Cinnamomum verum* collection, GC-MS, leaf essential oil composition, leaf morphology, Sri Lanka.

INTRODUCTION

The genus *Cinnamomum* in plant Family Lauraceae, to which Sri Lankan cinnamon (*Cinnamomum verum* J. Presl) belongs, consists of 250 species and sub-species. Most of these species are distributed in Asia, some parts of South and Central America, and Australia (Mabberley, 2008). *Cinnamomum verum* exhibits three major phenological phases of leaf flushing, flowering and fruiting. According to Hansika *et al.* (2022) flushing is controlled by seasonality and rainfall after a dry period. Azad *et al.* (2018 and 2019a) reported the wide variation of leaf and flower morphology in *Cinnamomum verum* germplasm explored from major growing areas of Sri Lanka.

* Corresponding author (sudarshanee@agbio.ruh.ac.lk; https://orcid.org/0000-0002-3771-2680)

There are seven wild cinnamon species in Sri Lanka, which comprise the secondary gene pool of cultivated cinnamon (Kumarathilake *et al.*, 2010). Prathibhani *et al.* (2021) have reported the protogynous dichogamy, leaf morphology, and leaf essential oil composition of selected wild cinnamon species from an *ex situ* conservation site. In 2018, the export volume of 17860 metric tons of cinnamon earned 37,315 million rupees (Central Bank of Sri Lanka, 2020). The leaf and bark essential oils of cinnamon are used primarily for spice, cosmetics, and pharmaceuticals due to their important chemical profiles (Joy *et al.*, 1998). Cinnamon bark, leaf, root, and fruit essential oils bear unique chemical profiles with cinnamaldehyde, eugenol, camphor, and cadinene, respectively, as major components (Senanayake *et al.*, 1990; Paranagama *et al.*, 2001). Cinnamon possesses several anti-diabetic, anti-inflammatory, anti-microbial, insecticidal, and antioxidant properties (Ranasinghe & Galappaththy, 2016). Many traditional Asian cultures use *Cinnamomum verum* as a medicine mainly for bloating, nausea, flatulence, colic and gastro-intestinal tract spastic conditions (Toriizuka, 1998).

Azad *et al.* (2019a) reported an eco-geographical survey in major cinnamon growing areas of the Matara, Galle, Ratnapura, Kalutara, Kurunegala, and Hambantota districts of Sri Lanka to develop a core collection of Sri Lankan cinnamon germplasm. Intensity of leaf spot and rough bark diseases in major cinnamon growing areas was reported by Azad *et al.* (2019b). A vegetative propagated cinnamon collection was established at the Faculty of Agriculture, University of Ruhuna for *ex situ* conservation (Azad *et al.*, 2019a). Azad *et al.* (unpublished data) have reported that there is a variation in essential oil composition of cinnamon bark, collected from plantations more than 30 years old, in major cinnamon growing areas of Sri Lanka. Sri Lankan cinnamon germplasm exhibits a wide variation in morphology that may be associated with distinct chemical profiles (Azad *et al.*, 2016; Azad *et al.*, 2019a).

Determination of morphological markers for the identification of the variation recorded in the chemical composition is important in breeding new varieties. Wijesinghe and Gunarathna (2001) reported a relationship between leaf shape and size with the yield in seven genotypes of true cinnamon, as trees with large round inwardly curved leaves had produced a higher cinnamaldehyde content in bark oil, while high quality leaf oil was obtained from the small round leaves. Flush colour of cinnamon was recorded to be a highly variable qualitative leaf character (Azad *et al.*, 2016; Azad *et al.*,

2019a). The anthocyanins cyanidin glucoside, cyanidin xyloside, and cyanidin galactoside were reported to be the pigments responsible for cinnamon flush colour (Joy *et al.*, 1998). The present study is an attempt to determine the variations of leaf essential oil chemical profiles and plant morphology in Sri Lankan cinnamon germplasm at one location, irrespective of differential environmental effects. Further, the information on characterization of chemical constituents may provide an insight to the biochemical pathways of these chemical constituents. The reports on cinnamon chemical composition by Wijesekera *et al.* (1974), Senanayake and Wijesekera (1990), and Paranagama *et al.* (2001) are based on a limited number of cinnamon genotypes. Morphological characterizations of whole cinnamon plant by Azad *et al.* (2016; 2019a), and flower by Azad *et al.* (2018) were based on *in situ* data from different locations. Azad *et al.* (2015) carried out the *ex situ* evaluation of age and the environmentally independent morphological characters of leaf shape, leaf apex, and leaf base of two *Cinnamomum verum* progenies and mother plants at the same location. Both progenies were different from their mother plants and produced new phenotypes for leaf shape and leaf base. Cross pollination due to protogynous dichogamy in genus *Cinnamomum* including the seven wild cinnamon species in Sri Lanka (Kumarathilake *et al.*, 2010), had led to the wide variation among accessions. Prathibhani *et al* (2021) have reported the protogynous dichogamy in selected wild cinnamon species from an *ex situ* conservation site and suggested its effect on variation of leaf morphology and leaf essential oil composition. There are no reports on chemical and morphological characterization of a representative cinnamon germplasm collection established at one location in Sri Lanka, which may be independent of environmental effects on morphology and chemical composition. Therefore, this study was conducted to determine the variations of leaf oil composition and plant morphology in a collection of three-year-old vegetative-propagated accessions, established at University of Ruhuna, which were originally collected from major cinnamon growing areas (Azad *et al*, 2019a).

MATERIALS AND METHODS

Location

The cinnamon collection, established at the Faculty of Agriculture, University of Ruhuna (GPS 6°03'27.7"N 80°34'02.9"E) was used as a representative collection for Sri Lankan cinnamon germplasm (Azad *et al.,* 2019a).

The Faculty of Agriculture, University of Ruhuna is located at Mapalana, Kamburupitiya, in Sri Lanka in the agro-ecological region WL2 (low country wet zone with >1900 mm of annual rainfall). Mean monthly temperatures of the location during the research period of August to October, 2018 were 28.8 °C, 30.4 °C, and 28.4 °C, respectively. Mean monthly rainfalls were 2.14 mm, 4.91 mm, and 9.79 mm, respectively.

Morphological characterization

Material: The representative cinnamon collection (Azad *et al*., 2019a) was comprised of 48 vegetative-propagated, three-year old accessions, which were used for flush colour determination. For data collection of other characters, only 40 accessions were selected based on leaf availability and plant growth. All morphological characters except the flush colour of the accessions were based on the Descriptors for Cinnamon (Team of TURIS 2013 Project, 2016).

Leaf characters

The above mentioned 40 cinnamon accessions in the collection were characterized based on both quantitative and qualitative leaf morphological characters, *viz.*, leaf length (LL), leaf width (LW), Petiole length (PL), leaf arrangement (LA), leaf shape (LS), leaf apex (LAP), leaf base (LB), leaf texture (LT), upper surface leaf texture (UST), lower surface leaf texture (LST), leaf venation (LV), and leaf margin (LM), during August and September, 2018. For flush colour (FC) determination, a young active bud within two days of emerging was selected from each accession. Observations were taken on five consecutive days. At the fourth to fifth day of the fully emerged flush, colour was recorded.

Bark characters

Bark characters *viz.*, bark fragrance (BF), bark thickness (BTH), bark taste (BT), bark colour (BC), peel quality (PQ) and bark surface (BS) were recorded .

Tree characters

Tree characters, *viz.*, tree height (TH), tree vigor (TV), tree shape (TS), trunk circumference (TC), trunk surface (TSU), crotch angle of main branches (CA), inter-nodal length of twigs (INLT), distribution of branches (DIBR), twig diameter (TWD), and branching pattern (BP) were recorded. The plants were observed for the presence or absence of flower buds, flowers and fruits at the time of data recording. The occurrence of leaf spot disease was recorded in order to determine the level of resistance to leaf spot disease among cinnamon accessions. The variations of selected qualitative characters are graphically presented, while quantitative character variations are given in the text.

Chemical characterization of leaf oil

Leaf oil extraction

Material: Twenty (20) accessions were selected from the collection, considering the diversity of leaf morphology and high leaf availability. For oil extraction, healthy, mature, fresh leaves were collected from selected accessions in October 2018.

Methods: Collected leaves were shade dried at room temperature for two weeks until the colour changed from green to metallic brown. Dried cinnamon leaves were cut into small pieces of ~1-1.5 cm before hydro-distillation in a light oil clevenger-type apparatus to extract cinnamon leaf essential oil at the National Cinnamon Research and Training Center (NCRTC), Pallolpitiya, Thihagoda, Sri Lanka. The round bottom flask with 50 g of dried leaf sample and 250 mL of tap water was settled on the heater. The settled Clevenger arms were filled with 2 mL of hexane ether solvent (2 : 1 ratio of hexane to ether) and 10 mL of distilled water. Hydro-distillation was performed for 4 h for one sample. Cinnamon leaf essential oil was separated from the mixture of hexane ether, water and leaf essential oil after an overnight settling period.

GC-MS analysis

Material: Twenty (20) cinnamon leaf essential oil samples were prepared for the GC-MS analysis.

Method: Each essential oil sample of 50 μL was dissolved in 950 μL of absolute methanol. All samples were filtered using Agilent 0.22 μm PTFE syringe filters. An Agilent 7890A gas chromatograph system (Agilent, USA) coupled with a 5975C inert XL EI/CI triple axis mass selective detector was used to perform GC-MS analysis at the Instrumental Center, University of Sri Jayawardenapura, Sri Lanka. A 19091s-433HP-5MS 5% phenyl methylpolysiloxane capillary column (of 30 m length and 250 μm inner diameter, 0.25 μm film thickness) was used for separation. One microliter of the sample was injected using a split injector with a split ratio of 1:100. The initial oven temperature was set at 40 °C and increased up to 230 °C at a rate of 5 °C/min.

Total run time was 38 min. The injector temperature was maintained at 250 ℃. The ion source temperature was 230 ℃. The carrier gas was Helium (He) with a flow rate of 1 mL/min. Leaf essential oil constituents were identified based on their mass spectra with authentic standard samples or with those recorded in the National Institute of Standards and Technology database. The relative abundance of identified chemical compounds of each accession were recorded. Morphological and chemical variables were subjected to PCA followed by cluster analysis through FactoMine R (Le *et al*., 2008).

RESULTS AND DISCUSSION

Morphological characterization

Leaf characters: Leaf is one of the most useful morphological characters in genus *Cinnamomum* as it varies greatly among species (Ravindran *et al*., 2004). Mean values of LL, LW, length to width ratio and PL were 11.40 cm (±0.24), 5.2 cm (±0.14), 2.2 (±0.04) and 1.5 cm (±0.04) respectively. In Table 1, the variation of quantitative leaf characters such as LL, LW, leaf length to width ratio, and PL of the accessions in the collection established at the Faculty of Agriculture, University of Ruhuna, Sri Lanka was compared and contrasted with the reported ranges of above characters at the original locations by Azad *et al.* (2019). In the same table, the means of above characters of selected wild cinnamon species (Prathibhani *et al*, 2021), are compared. Leaf arrangement patterns of opposite, sub-opposite, opposite or sub-opposite in different branch but in same plant, and opposite to sub-opposite in the same branch in the same plant, as indicated, are illustrated in Figure 1(a), following the Descriptors for Cinnamon by Team of TURIS 2013 Project (2016). The above leaf arrangement patterns were recorded in 7/40, 3/40, 4/40, and 26/40 accessions, respectively (Figure 1A). There were eight types of leaf shapes as described in the Descriptors from elliptic to oblong-lanceolate (Figure 1(b), in different frequencies. Elliptic, broadly elliptic, narrowly elliptic, ovate, broadly ovate, oval, lanceolate and ovate-lanceolate categories were recorded in 15/40, 1/40, 12/40, 2/40, 2/40, 3/40, 4/40 ,and 1/40 accessions, respectively (Figure 1B). Six types of leaf apexes, *viz.*, acute, acuminate, long acuminate, narrowly acuminate, acuminate with broad acumen and obtuse (Figure 1(c) were recorded in 20/40, 9/40, 7/40, 2/40, 1/40, and 1/40 accessions, respectively (Figure 1(C). Five categories of leaf base, acute, sub-acute, rounded, obtuse, and obtuse

contracted into petiole then shortly cuneate (Figure 1(d) were recorded in 6/40, 8/40, 3/40, 10/40, and 13/40 accessions, respectively from the collection (Figure 1(D). Three types of leaf venation patterns, *viz.*, three-veined, five-veined and three- and five-veined on the same plant were observed in 26/40, 1/40, and 13/40 accessions, respectively in the *Cinnamomum verum* collection. Two types of leaf margins, entire and undulate, were recorded at frequencies of 6/40 and 34/40 accessions, respectively. In *Cinnamomum verum*, observations were made on the flushes emerging directly from the main stem, from branches, from twigs, or from the cut base of the stem. The colour patterns varied in flushes of different origin on the same plant. A gradual colour change was observed in flush with time from emergence to maturation. The variation of flush colour of a plant in terms of spatial and temporal variations may be due to genetic response to the environmental cues. The frequency of pink, red, brown, and green colour categories of the same aged flushes is depicted in Figure 2. Photographs of different flushes under each colour category are presented: The pink and red are in Figure 3, while brown and green are in Figure 4. In Figure 3, the GAM2 (A), KE5 (B) and GK 17 (C) accessions represent the light pink leaf blade with a light green base. A pink leaf blade with a light green base is represented by accessions GB13-2 (D), KA11 (E) and KA14 (F). Accessions KB 14 (G), GB13-1(H) and GE7 (I) represent the red leaf blade with a green leaf base. In Figure 4, the MaDS5 (J), 2MDS5-2 (K) and Palol-1(L) accessions represent the light brown throughout the leaf. A light brown leaf blade with a green leaf base was represented by accession MKG14 (M). Brown colour throughout the leaf is represented by 2 MDS4 (N), while a brown leaf blade with a light green leaf base was represented by 2-MDS5-1(O) and MKG 13 (P). A light green colour throughout the leaf was represented in HB12 (Q) while a whitish green colour throughout the leaf blade with light pink leaf apex was represented by GB17 (R).

Bark characters: The frequencies of different bark fragrance types, bark tastes, bark peeling qualities, and bark surface types of the cinnamon accessions in the collection were reported (Figure 5). The good fragrant aroma category of bark fragrance and the sweet pungent category of bark taste were recorded at the frequency of 21/40 accessions. Weak bark peeling qualities were recorded at the frequency of 16/40, while the good peeling quality was at 10/40 accessions. Slightly rough bark surface was at the frequency of 22/40 accessions.
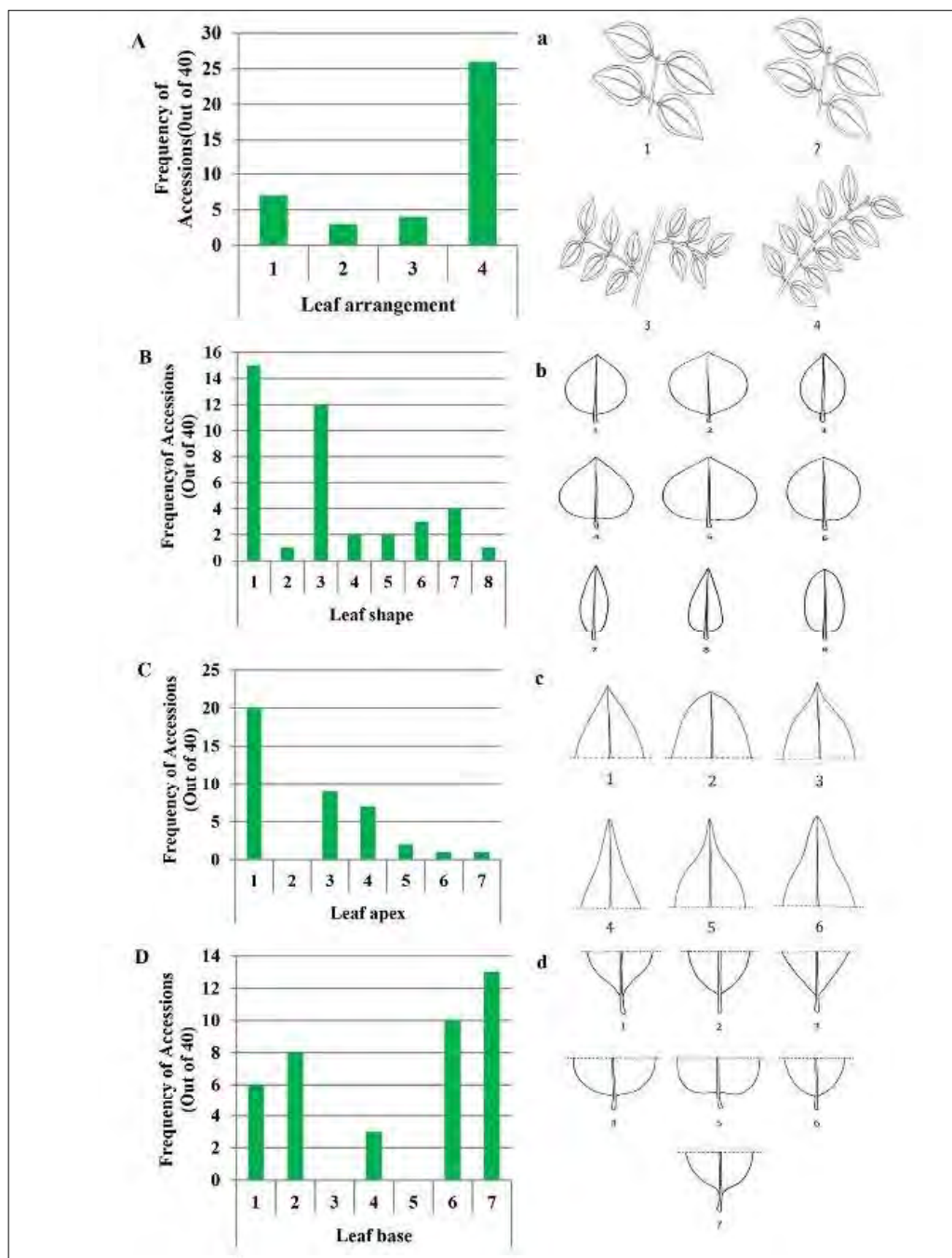
Figure 1: Variation of leaf morphology in the cinnamon collection at University of Ruhuna based on the Descriptors for Cinnamon (*Cinnamomum verum*) (Team of TURIS 2013 Project, 2016); (A) Frequency of leaf arrangement types (a) Descriptors for leaf arrangement patterns of *Cinnamomum verum*; 1. opposite 2. sub-opposite 3. opposite to sub-opposite in a different branch in same plant 4. opposite to sub–opposite in the same branch in the same plant (B) Frequency of different leaf shapes (b) Descriptors for leaf shapes of *Cinnamomum verum*; 1. elliptic 2. broadly elliptic 3. narrowly elliptic 4. ovate 5. broadly ovate 6. oval 7. lanceolate 8. ovate-lanceolate 9. oblong-lanceolate (C) Frequency of leaf apex types (c) Descriptors for leaf apex types of *Cinnamomum verum*; 1. acute 2. obtuse 3. acuminate 4. long acuminate 5. narrowly acuminate 6. acuminate with broad acumen 7. other (D) Frequency of leaf base types (d) Descriptors for leaf base types of *Cinnamomum verum*; 1. acute 2. subacute 3. cuneate 4. rounded 5. subcordate 6. obtuse 7. obtuse, contracted into petiole, then shortly cuneate.
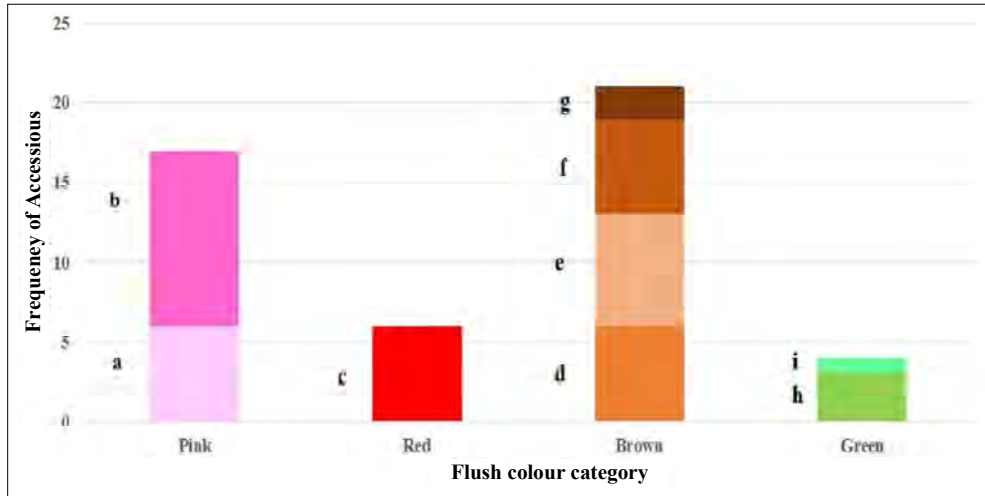
Figure 2: Variation of flush colour of the cinnamon collection at University of Ruhuna Pink; (a) light pink colour leaf blade with light green colour base (6/48) (b) pink colour leaf blade with light green leaf base (11/48) Red; (c) red colour leaf blade with green leaf base (6/48) Brown; (d) light brown colour throughout the leaf (6/48) (e) light brown colour leaf blade with green leaf base (7/48) (f) brown colour throughout the leaf (6/48) (g) brown colour leaf blade with light green leaf base (2/48) Green; (h) light green colour throughout the leaf (3/48) (i) whitish green colour throughout the leaf blade with light pink leaf apex (1/48)

Table 1: Variation of leaf quantitative characters of cultivated and wild cinnamon species

Leaf character	Range						
	1	2	3	4	5	6	7
Leaf length (cm)	7.9 to 14.5	7.43 to 20.26	14.02 ± 2.13	11.88 ± 1.22	13.98 ± 1.44	12.52 ± 1.53	7.44 ± 0.44
Leaf width (cm)	3.8 to 7.3	3.4 to 10.08	6.68 ± 1.13	5.92 ± 0.64	4.20 ± 0.67	3.78 ± 0.05	2.66 ± 0.23
Leaf length to width ratio	1.08 to 3.07	1.58 to 3.1	2.10	2.01	3.33	3.31	2.81
Petiole length (cm)	1.0 to 2.2	0.9 to 2.7	1.60 ± 0.26	1.16 ± 0.21	1.00 ± 0.24	0.66 ± 0.21	0.88 ± 0.21

1 Accessions of the collection at Faculty of Agriculture, University of Ruhuna, during the current study, 2 Accessions at the original locations (Azad *et al*., 2019), 3 Cultivated cinnamon variety *Sri Gemunu* at Mid-Country Research Station, 4 Cultivated cinnamon variety *Sri Wijaya* at Mid-Country Research Station 5 *Cinnamomum capparu-coronde* Blume at Mid-Country Research Station, 6 *Cinnamomum litsaeifolium* Thwaites at Mid-Country Research Station (Prathibhani *et al*., 2021)

Tree characters: TH of accessions ranged from 0.53 m to 2.51 m. TV of accessions was reported as weak (8/40), intermediate (20/40) and strong (12/40). Pyramidal (7/40), circular (3/40), and irregular (30/40) shaped TS were reported. The frequencies of TSU of accessions were reported as smooth (10/40), rough (25/40) and very rough (5/40). CA of accessions was reported as 90° (14/40) and other (26/40). The average INLT of accessions was reported from 1.2 to 8 cm. The frequencies of DIBR was reported as ascendant (26/40), irregular (3/40), axial (1/40), and horizontal (10/40). The presence of leaf spot disease was reported in all accessions. The average TWD of accessions ranged from 0.9 to 2.3 cm. Nine accessions were reported with the extensive BP and 31 were reported with the intensive BP. Six accessions were at flowering during the period of data collection.

Figure 3: Representative accessions for different flush colour categories of pink and red in the cinnamon core collection: Light pink colour leaf blade with light green colour base (A) GAM2 (B) KE5 (C) GK17, Pink colour leaf blade with light green leaf base (D) GB13-2 (E) KA11 (F) KA14 and Red colour leaf blade with green leaf base (G) KB14 (H) GB13-1 (I) GE7.

Figure 4: Representative accessions for different flush colour categories of brown and green in the cinnamon core collection: Light brown colour throughout the leaf (J) MaDS5 (K) 2MDS5-2 (L) Palol-1, Light brown colour leaf blade with green leaf base (M) MKG14, Brown colour throughout the leaf (N) 2MDS4, Brown colour leaf blade with light green leaf base (O) 2MDS5-1 (P) MKG13, Light green colour throughout the leaf (Q) HB12 and Whitish green colour throughout the leaf blade with light pink leaf apex (R) GB17.

Variation in quantitative leaf characters LL, LW, length to width ratio, PL, and in qualitative leaf characters LS, LAP, LB, LA, and LM of cinnamon accessions at the collection at the Faculty of Agriculture, in contrast to their original locations, reflects the intra-specific genetic diversity of *Cinnamomum verum*. Plants acquire morphological plasticity with confounding environmental conditions. In most plant species, leaf morphological plasticity is incompletely understood. In *Quercus acutissima,* stable leaf morphological characters at species level are the main characters that adapt to environmental changes (Zhang *et al*., 2018). During data collection, the accessions KA 12, KA14, KD5-1, GAK9, RL16-1 and RL16-2 were at the first flowering. Environment dependent leaf morphological characters could be identified based on the variability of the characters, which was valid for wild cinnamon as well (Table 1).

GC-MS analysis: One hundred and forty-six chemical compounds were reported during the GC-MS analysis of 20 leaf oil samples. Eugenol was the major compound of

seventeen accessions with the content ranging from 52.2-79.5%, while BB was the major compound of other three accessions ranging from 22.3-91.1%. Zero percentage of eugenol with 86.8 and 91.9% of BB were recorded in two accessions. The accession which produced 22.3% of BB, recorded 16.5% of eugenol (Table 3). The variation of eugenol, linalool, *β*-caryophyllene and BB in twenty cinnamon accessions collected from different cinnamon growing districts and established at the Faculty of Agriculture is illustrated in Figure 6. The range of selected fatty acids from 20 cinnamon leaf oil samples detected through GC-MS is listed (Table 3).

Principal component analysis (PCA): Fifty-five morphological and chemical variables (including 31 morphological variables and 24 chemical variables) were subjected to PCA. Eight PCs explained 83.11% total variance. In Table 4, percentage of variance of

each PC is presented. Methyl palmitate (MetP), methyl stearate (MS), methyl myristate (MeM), methyl laurate (MeL), toluene (To), prococene-2 (HB), ethyl palmitate (EthP), methyl linolelaidate (MeLIn), methyl elaidate (Octa), methyl iso-eugenol (MeIE), 2-oxazolidone

(Oxa), ethyl oleate (EO), α-caryophyllene (AlCa), benzyl benzoate (BB), (1~{Z},4~{Z},7~{Z})-1,5,9,9-tetramethylcycloundeca-1,4,7-triene (Cyc) and linalool (Li) were among major components of total variance ($p < 0.05$).
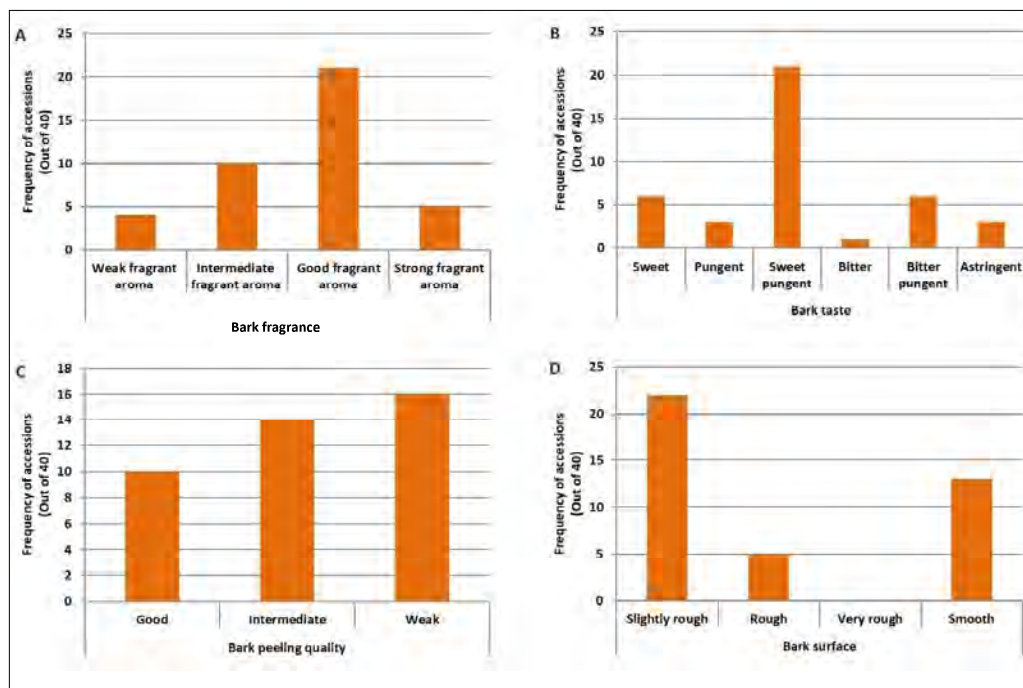
Figure 5: (A) Frequency of bark fragrance types in the cinnamon collection at University of Ruhuna based on Descriptors for Cinnamon (Team of TURIS2013 Project, 2016): (B) Frequency of bark tastes in the cinnamon core collection: (C) Frequency of bark peeling qualities in the cinnamon core collection: (D) Frequency of bark surface types in the cinnamon core collection

Correlation matrixes of major chemical compounds: Pearson correlation coefficients were generated among eight major chemical variables, *viz.*, toluene, eucalyptol, linalool, eugenol, α-caryophyllene, β-caryophyllene, caryophyllene oxide, and BB, through IBM SPSS version 24.0 statistical software (Table 5). Strong positive correlations were found between linalool and β-caryophyllene (+0.649), and linalool and BB (+0.770), at the 0.01 level of significance. Strong negative correlations were found between linalool and eugenol (-0.630), and eugenol and BB (-0.886), at the 0.01 level of significance. At the 0.05 level of significance, positive correlations between eucalyptol and α-caryophyllene

(0.472), and β-caryophyllene and BB (0.498), and negative correlations between toluene and linalool (0.483), eucalyptol and caryophyllene oxide (0.457), and eugenol and β-caryophyllene (0.497) were significant.

Cluster analysis: Cluster analysis, followed by PCA using the average linkage method, classified the selected twenty cinnamon accessions into five clusters at an average distance of 2 (Figure 7). Cluster analysis assigned the two cinnamon accessions with highest BB percentages, KA11 and GB17 (86.8% and 91.9% respectively) into one cluster and HB12 of (22.3% of BB) into a different cluster.

Table 2: A comparison of eugenol, *β*-caryophyllene, linalool, benzyl benzoate, methyl eugenol, eucalyptol, and alpha-terpineol in leaf oil of cultivated and selected wild *Cinnamomum* species from Sri Lanka

Source	Relative abundance (%)						
	Eugenol	β-Caryophyllene	Linalool	Benzyl benzoate	Methyl eugenol	Eucalyptol	Alpha-terpineol
Wijesekara *et al.*, 1974	87	1.85	1.5	2.67	-	-	0.15
Paranagama *et al.*, 2001	76.74	3.47	2.77	4.01	-	-	0.28
Senanayake *et al.*, 1978	68.5	3.33	2.4	4.06	0.01	-	0.4
Schmidt *et al.*, 2006	74.9	4.1	2.5	3.0	-	-	0.3
Liyanage *et al.*, 2017	85.66	1.08	0.97	-	-	-	-
Accessions of collection at the Faculty of Agriculture, Univ. of Ruhuna during the current study	0-79.47	0.46-3.08	0-5.59	0-91.92	-	0.06-0.43	0.11-0.42
Cinnamomum Capparu-coronde Blume (Prathibhani *et al.*, 2021)	33.11	1.81	11.76	-	-	-	0.37
Cinnamomum dubium Nees (Prathibhani *et al.*, 2021)	-	-	3.41	-	-	51.19	13.18
Cinnamomum litsaeifolium Thwaites (1) (Prathibhani *et al.*, 2021)	24.99	-	7.16	-	59.27	-	0.63
Cinnamomum litsaeifolium Thwaites (2) (Prathibhani *et al.*, 2021)	-	3.04	30.93	-	-	0.12	10.01

According to previous records on leaf chemical composition of *Cinnamomum verum* from Sri Lanka, eugenol accounts for the highest percentage in leaf essential oil mainly with low percentages of linalool, *β*-caryophyllene, cinnamaldehyde, and BB (Wijesekera *et al.*, 1974; Liyanage *et al.*, 2017; Ravindran *et al.*, 2004; Schmidt *et al.*, 2006; Paranagama *et al.*, 2001). In agreement with those reports, seventeen accessions in the core collection were detected with eugenol ranging from 52.2-79.5%. However, *Cinnamomum dubium* is reported to possess less or zero eugenol in leaves (Kumarathilake, 2009; Prathibhani *et al.*, 2021). Deviating from other accessions, three accessions from Kalutara (KA11), Hambantota (HB12), and Galle (GB17), produced chemical profiles, mainly with BB (Figure 7). Both

HB12 and GB17 were associated with green flush colour, which needs to be further investigated as a potential morphological marker. Determination of flavonoid composition and variation of Sri Lankan cinnamon germplasm would be important in the production of therapeutic products. This is a recent record of the presence of BB as the major constituent in cinnamon leaf essential oil and the first record from Sri Lanka. GC and GC/MS analysis of leaf and bark essential oils from *Cinnamomum verum* grown in the Brahmaputra valley, India, recorded BB as the main constituent at 65.42% and 84.69%, respectively (Nath *et al.*, 1996). GC-MS analysis reported cinnamaldehyde, cinnamyl acetate, eugenol, BB, *β*-caryophyllene, and linalool as the major chemical compounds in bark essential oil profiles from the

accessions at their original locations. There were positive correlations between eugenol and BB (0.49), and linalool and *β*-caryophyllene (0.52), and negative correlations between cinnamaldehyde and cinnamyl acetate (0.81), cinnamaldehyde and BB (0.37), and cinnamaldehyde and

β-caryophyllene (0.37) (Azad *et al*., unpublished data). There was a positive correlation between eugenol and BB in bark essential oil profiles, which had cinnamaldehyde as the major constituent. In contrast, there is a negative correlation between eugenol and BB in leaf essential oil
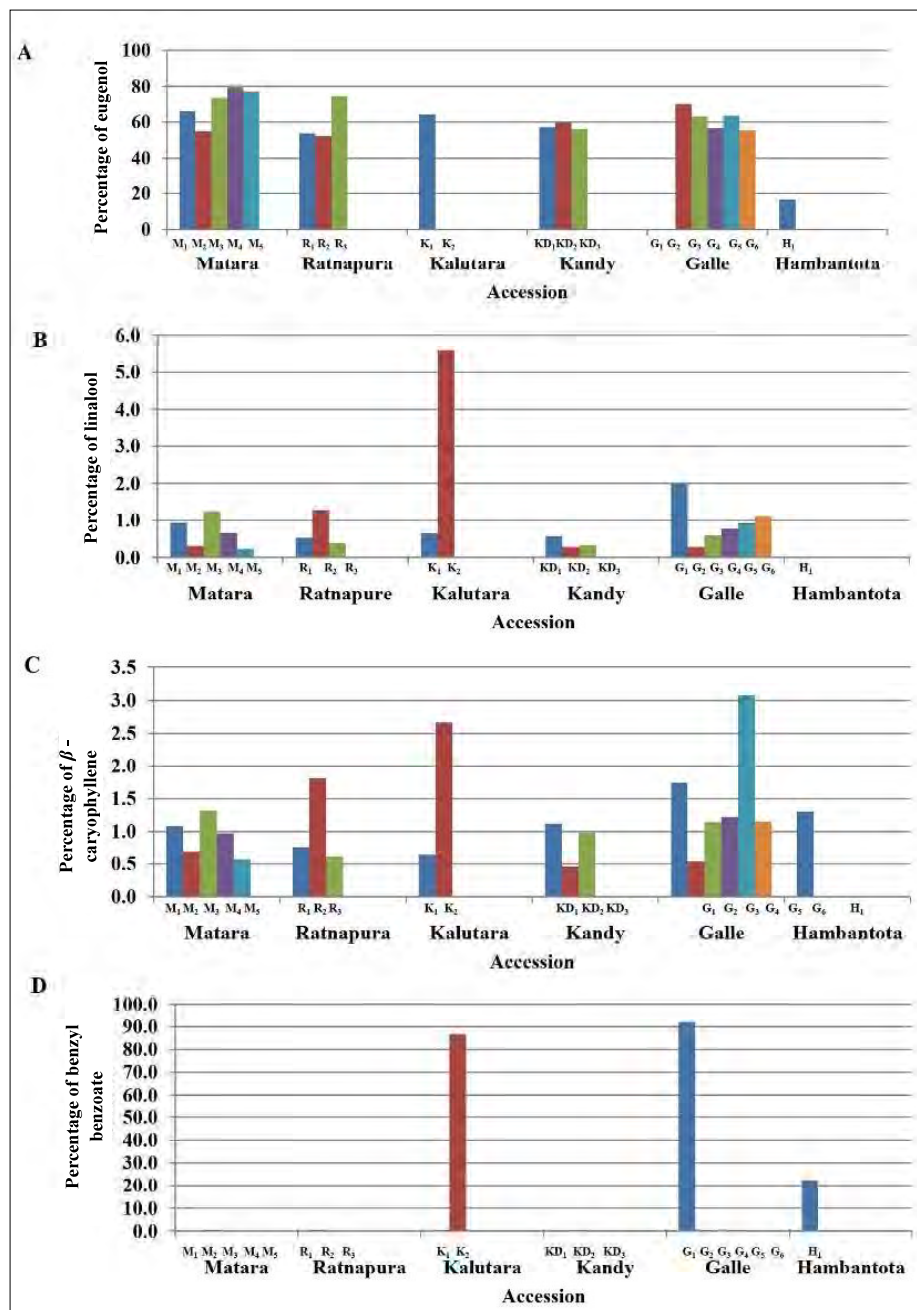
Figure 6:   Variation of percentages of eugenol (A), β-Caryophyllene (B), Linalool (C) and Benzyl benzoate (D) in selected accessions of core collection originally from different districts and grown at Faculty of Agriculture, University of Ruhuna. M1, M2, M3, M4, M5, R1, R2, R3, K1, K2, KD1, KD2, KD3, G1, G2, G3, G4, G5, G6 and H1 represent accessions MKG14, MKG13, MaDS5, 2MDS5-1, 2MDS4-1, RL16-2, RL16-1, RL15, KA14, KA11, KD5-2, KD5-1, KD2-1, GB17,GK17, GE11, GAM2, GAK9, GAK3-1 and HB12 respectively.

profiles in the current study. The fatty acid methyl esters methyl laurate, and methyl myristate are used as flavoring agents and may be responsible for the characteristic

odour and taste of cinnamon. Most of these fatty acids recorded from cinnamon leaf essential oil are identified as plant metabolites. Fatty acid analysis of *Cinnamomum verum* bark oil using gas liquid chromatography has shown the presence of undecanoic acid (2.43%), lauric acid (0.5%), palmitic acid (15.58%), linoleic acid (26.08%), stearic acid (47.68%), and oleic acid (7.73%) (Jamil *et al.,* 2012). Schmidt *et al.* (2006) reported that the presence of considerably higher concentrations of the phenolic compound, eugenol, was responsible for the antioxidant capacity of cinnamon leaf essential oil. Cinnamon essential oil could be an alternative to replace the synthetic carcinogenic antioxidants of butylated hydroxytoluene (BHT) and butylated hydroxyanisole (BHA), with a higher antioxidant activity and the lowest IC_{50} value among them. Cinnamon oil was effective over BHT in inhibition of secondary product formation

Table 3: Range of fatty acids in cinnamon leaf essential oil from 20 cinnamon accessions

Fatty acid	Range of relative abundance%
Methyl laurate	0-3.74
Methyl myristate	0-1.89
Methyl palmitate	0-8.96
Ethyl palmitate	0-0.20
Methyl stearate	0-1.37
Ethyl oleate	0-0.21
Methyl linolelaidate	0-2.95

Table 4: Variance (%) of each principal component (PC) of principal component analysis (PCA)

PC	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
Variance	9.250	5.057	3.346	3.209	2.107	1.840	1.350	1.155	0.986	0.737
% Variance	29.838	16.314	10.794	10.351	6.798	5.936	4.356	3.726	3.182	2.378
Cumulative % of variance	29.838	46.153	56.947	67.297	74.095	80.031	84.387	88.113	91.295	93.673

Table 5: The correlation matrix of eight identified chemical compounds in cinnamon leaf essential oil collected from 20 accessions of the core collection.

Variable	To	Euc	Li	Eu	Alca	Ca	CaOx	BB
To	1.000	0.191	-0.483*	0.145	0.437	0.004	0.205	-0.417
		0.420	0.031	0.543	0.054	0.987	0.386	0.067
Euc		1.000	-0.185	0.298	0.472*	0.285	-0.457*	-0.345
			0.434	0.203	0.036	0.223	0.043	0.136
Li			1.000	-0.630**	-0.238	0.649**	0.007	0.770**
				0.003	0.311	0.002	0.977	0.000
Eu				1.000	0.035	-0.497*	0.171	-0.886**
					0.884	0.026	0.472	0.000
Alca					1.000	0.369	-0.056	-0.205
						0.110	0.813	0.386
Ca						1.000	0.017	0.498*
							0.942	0.026
CaOx							1.000	0.011
								0.962
BB								1.000

To: tolune; Eu: eucalyptol; Li: linalool; Eu: eugenol; Alca: α-caryophyllene; Ca: β-caryophyllene: CaOx: caryophyllene oxide and BB: benzyl benzoate

by lipid peroxidation. However, all twenty accessions have distinct chemical profiles differentiating from each other even in the same environmental conditions. The different chemical cultivars in the cinnamon plantations had been recognized as sweet, honey, camphoraceous, mucilaginous, wild, and bloom by planters based on sensory evaluation. Only the sweet and honey varieties had been selected for cultivation (Ravindran *et al.*, 2004). Accessions with more than 70% of eugenol and zero percentage of BB were recorded as having a sweet pungent taste (RL15, MaDS5, 2MDS5-1 and 2MDS4-1). KA11 and GB17 with more than 85% of BB and zero percentage of eugenol recorded sweet pungent and sweet tastes, respectively. HB12 with 22.3% of BB and 16.6% of eugenol produced a bitter pungent taste. According to the above results, there is a possibility that bulk oil yields of cinnamon bark and leaf from commercial plantations contain variable chemical compositions suggesting a great potential in utilization for diverse industries of food, pharmaceuticals, and cosmetics. The results of the study showed that there is an immediate requirement in standard grading systems and quality estimating for chemical profiles in export bulks of cinnamon products

from Sri Lanka. Continuous cross pollination over generations may have created a wide chemical diversity and leaf diversity including a variety of flush colours among accessions of Sri Lankan cinnamon germplasm. Accessions GB 17 and HB 12 recorded zero or low eugenol and exceptionally higher benzyl benzoate, in contrast to other tested accessions contained flushes of green colour variations. Our results suggest the utilization of germplasm to optimize economic potential: The identified accessions with distinct chemical profiles could be released to the farmers after evaluation of their agronomic feasibility by the Department of Cinnamon Development. Simultaneously, the results provide an insight into the importance of the cinnamon germplasm: It would be a breeding resource for future cultivars of novel chemical compositions leading to desired aroma and flavour. On the other hand, this germplasm will provide information on elucidating the relationship between the pigment composition of flush colour and the essential oil biosynthesis pathway, to reveal the potential morphological markers for distinct chemical cultivars.

Figure 7: Dendrogram classified the twenty accessions of the core collection at the Faculty of Agriculture, University of Ruhuna, based on both morphological and chemical variables.

CONCLUSION

There is a variation in leaf morphological characters among accessions in the collection. There are four main categories of flush colour – pink, red, brown, and green. Leaf length, leaf width, leaf length to width ratio, and petiole length vary from 7.9 to 14.5 cm, 3.8 to 7.3 cm, 1.8 to 3.07 cm, and 1 to 2.2 cm, respectively. The collection comprised different bark fragrance types. According to the analysis of gas chromatography-mass spectrometry (GC-MS) of leaf oil, we report that there are eugenol (52.2% to 79.5%) rich accessions and, for the first time in Sri Lanka, zero to 16.6% eugenol containing, benzyl benzoate (BB) rich (23.3%, 86.8% and 91.9%) accessions. The green flush colour is associated with a higher BB percentage, which needs to be further investigated as a visual marker.

The dendrogram followed by the PCA, which is derived through variations of plant morphological characters and leaf oil constituents, explained 88.11% of total variance. The cinnamon accessions with exceptionally high BB, were clustered as distinct.

Acknowledgements

(2018) to SG. Authors wish to acknowledge the Staff of the Division of Post-Harvest Technology, National Cinnamon Research and Training Center, Pallolpitiya, Sri Lanka and the Instrument Center, Faculty of Applied Sciences, University of Sri Jayawardhanapura, Sri Lanka for technical support of this study.

REFERENCES

Azad R., Senanayake G., Kumara K.L.W., Ranawaka R.A.A.K., Pushpakumara D.K.N.G., Wijesinghe, K.G.G. & Geekiyanage S. (2015). Morphological variation within progeny and deviations from mother plant reveal the allele richness in *Cinnamomum verum* germplasm: a case study from Deiyandara, Matara collection at the early vegetative stage. *Tropical Agricultural Research & Extension* **18**(4): 163–167.
DOI: https://doi.org/10.4038/tare.v18i4.5380

Azad R., Jayasekara L., Ranawaka R.A.A.K., Senanayaka G., Kumara K.L.W., Pushpakumara, D.K.N.G. & Geekiyanage S. (2019a). Development of a core collection for Sri Lankan cinnamon germplasm based on morphological characterization using an eco-geographical survey. *Australian Journal of Crop Science.* **13**(09): 1473–1485.
DOI: https://doi.org/10.21475/ajcs.19.13.09.p1561

Azad R., Kumara K.W., Senanayake G., Ranawaka R.A.A.K., Pushpakumara D.K.N.G. & Geekiyanage S. (2019b). Intensity of leaf spot and rough bark diseases in cinnamon accessions collected from major cinnamon growing areas of Sri Lanka. *Journal of the National Science Foundation of Sri Lanka* **47**(3): 297–305.
DOI: https://doi.org/10.4038/jnsfsr.v47i3.9403

Azad R., Ranawaka R.A.A.K., Senanayake G., Kumara K.L.W., Wijesinghe K.G.G., Pushpakumara D.K.N.G. & Geekiyanage S. (2016). Morphological variation of cinnamon (*Cinnamomum verum* Presl) germplasm in Matara district of Sri Lanka. *International Journal of Minor Fruits, Medicinal and Aromatic plants* **2**(1): 6–14.

Central Bank of Sri Lanka. (2020). *Economic and Social Statistics of Sri Lanka*. Statistics Department, Central Bank of Sri Lanka, Colombo, Sri Lanka.

Hansika K.A.S., Kodikara K.A.S., Wijeweera A.A., Senanayake G. & Jayatissa L.P.(2022). Vegetative and reproductive phenology of *Cinnamomum verum* J. Presl grown in intermediate climatic zone of Sri Lanka. *Journal of National Science Foundation Sri Lanka.* 50 (1): 87–99.
DOI: https://doi.org/10.4038/jnsfsr.v50i1.10256

Jamil M., Salahuddin S., Akter H. & Shajahan M. (2012). Fatty acid analysis of *Cinnamomum zeylanicum* (Dalchini) fatty oil. *University of Information Technology and Sciences Journal* **1**(2): 70–74.

Joy P.P., Thomas J. & Samuel M. (1998). Cinnamon (*Cinnamomum verum* Presl) for flavour and fragrance. *Pafai Journal* **20**(2):37–42.

Kumarathilake D.M.H.C. (2009). Study on the extinction risks, conservation and domestication of endemic wild *Cinnamomum* species in Sri Lanka. *M.Phil. thesis*, Department of Agricultural Biology, Faculty of Agriculture, University of Ruhuna, Kamburupitiya.
DOI: https://doi.org/10.4038/tar.v21i3.3298

Kumarathilake D.M.H.C., Senanayake S.G.J.N., Wijesekara G.A.W., Wijesundara D.S.A. & Ranawaka, R.A.A.K. (2010). Extinction risk assessment at the species level: National red List status of endemic wild cinnamon species in Sri Lanka. *Tropical Agricultural Research.* **21**(3): 247–257.

Liyanage T., Madhujith T. & Wijesinghe K.G.G. (2017). Comparative study on major chemical constituents in volatile oil of true cinnamon (*Cinnamomum verum* Presl. Syn. *C. zeylanicum* Blum.) and five wild cinnamon species grown in Sri Lanka. *Tropical Agricultural Research* **28**(3): 270–280.
DOI: https://doi.org/10.4038/tar.v28i3.8231

Le S., Josse J., & Husson F. (2008). FactoMineR: An R Package for multivariate analysis. *Journal of Statistical Software* **25**(1):1–18.
DOI: https://doi.org/10.18637/jss.v025.i01

Mabberley D.J., 2008. *Mabberley's Plant-Book: A Portable Dictionary of Plants, Their Classifications and Uses,* 3rd edition, Cambridge University Press, UK.

Nath S.C., Pathak M.G. & Baruah A. (1996). Benzyl benzoate, the major component of the leaf and stem bark oil of *Cinnamomum zeylanicum* Blume. *Journal of Essential Oil Research* **8**(3):327–328.
DOI: https://doi.org/10.1080/10412905.1996.9700626

Paranagama P.A., Wimalasena S., Jayatilake G.S., Jayawardena A.L., Senanayake U.M. & Mubarak A.M. (2001). A comparison of essential oil constituents of bark, leaf, root and fruit of cinnamon (*Cinnamomum zeylanicum* Blume) grown in Sri Lanka. *Journal of the National Science Foundation of Sri Lanka.* **29**(3-4): 147–153.
DOI: https://doi.org/10.4038/jnsfsr.v29i3-4.2613

Prathibhani M.R., Ranawaka R.A.A.K., Samantha A.R. & Geekiyanage S. (2021). Protogynous dichogamy, leaf morphology and leaf essential oil composition of selected *Cinnamomum* species in Sri Lanka. *Tropical Agricultural Research and Extension* **24**(3): 185–197.
DOI: https://doi.org/10.4038/tare.v24i3.5515

Ranasinghe P. & Galappaththy P. (2016). Health benefits of Ceylon cinnamon (*Cinnamomum zeylanicum*): a summary of the current evidence. *Ceylon Medical Journal* **61**(1):1–5.
DOI: https://doi.org/10.4038/cmj.v61i1.8251

Ravindran P.N., Nirmal-Babu K. & Shylaja M. (2004). *Cinnamon and Cassia: the Genus Cinnamomum.* CRC press, New York, USA.
DOI: https://doi.org/10.1201/9780203590874

Schmidt E., Jirovetz L., Buchbauer G., Eller G.A., Stoilova I., Krastanov. A., Stoyanova A. & Geissler M. (2006). Composition and antioxidant activities of the essential oil of cinnamon (*Cinnamomum zeylanicum* Blume) leaves from Sri Lanka. *Journal of Essential Oil Bearing Plants* **9**(2):170–182.
DOI: https://doi.org/10.1080/0972060X.2006.10643490

Senanayake U.M., Lee T.H. & Wills R.B. (1978). Volatile constituents of cinnamon (*Cinnamomum zeylanicum*) oils. *Journal of Agricultural and Food Chemistry* **26**(4): 822–824. DOI: https://doi.org/10.1021/jf60218a031

Senanayake U.M. & Wijesekera R.O.B. (1990). The volatiles of the *Cinnamomum* species. *Proceedings of the 11th International Congress of essential oils, fragrances and flavours,* 12-16 November. New Delhi, India, pp. 103–120.

Team of TURIS 2013 Project (2016). *Descriptors for Cinnamon (Cinnamomum verum)*. University of Ruhuna, Matara, Sri Lanka.

Toriizuka K. (1998). Basic lecture of Kampo medicine: pharmacological effect of cinnamon. *Kampo Medicine* **11**:431–436.

Wijesekera R.O.B., Jayewardene A.L. & Rajapakse L.S. (1974). Volatile constituents of leaf, stem and root oils of cinnamon (*Cinnamomum zeylanicum*). *Journal of the Science of Food and Agriculture* **25**(10): 1211–1220. DOI: https://doi.org/10.1002/jsfa.2740251004

Wijesinghe K.G.G. & Gunarathna W.D.L. (2001). Characterization of true cinnamon (*Cinnamomum verum* Presl) based on leaf morphology and their relationship with yield and quality. *Proceedings of 57th Annual Session of Sri Lanka Association for the Advancement of Science*. SLAAS, Colombo, Sri Lanka, pp. 42.

Zhang H., Yang X., Yu M. & Wu T. (2018). Effect of genetics and environment on leaf morphology and SLA for Quercus 1 acutissima 2 Available at: *http://www. Researchgate.net/ publication/325118040,* Accessed 15 January 2019.

RESEARCH ARTICLE

Statistical Quality Control

The modified control charts for monitoring the error detection of process control under different estimators

Azam Zaka[1], Riffat Jabeen[2], Amirah Saeed Alharthi[3] and Hassan M. Aljohani[3]
[1] *Department of Statistics, Government Graduate College of Science, Wahdat Road, Lahore, Pakistan.*
[2] *Department of Statistics, COMSATS University of Islamabad, Lahore Campus, Lahore, Pakistan.*
[3] *Department of Mathematics and Statistics, College of Science, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia.*

Abstract: The reflected power function distribution (RPFD) has increasing importance in practical life due to its application in diversified fields of life. Organisations often face difficulty monitoring operations to identify and remove errors during production. That is why there is a need to introduce control charts that effectively monitor the processes, mainly when the number of errors follows RPFD and the manufacturing process is in control. The current study suggested memory-based control charts as a solution to the problem. The control charts are based on the estimation methods and play a remarkable role in enhancing the machine process reliability. The parameters of RPFD are evaluated through the percentile estimator (PE) and modified maximum likelihood estimator methods (MMLM). Further, we create memory-based control charts, *i.e.*, hybrid exponentially weighted moving average (HEWMA) and extended exponentially weighted moving average (EEWMA), using the PE and MMLM. The findings reflect that HEWMA control charts based on PE provide a better result in estimating the defects. The implications of the study will be helpful for practitioners and policy makers from reliability engineering, management sciences, and statisticians.

Keywords: Control charts, machine errors, manufacturing process, modified maximum likelihood estimator, percentile estimator, reflected power function distribution.

INTRODUCTION

The reflected power function was introduced by Zaka *et al.* (2020) as a modification to power function distribution. The reason for introducing the reflected

power function distribution is to enhance the application of the power function distribution in reliability, medical, and engineering sciences. The probability density function (PDF) of the reflected power function distribution is given as

$$f(x) = \frac{\gamma(\theta-x)^{\gamma-1}}{\beta^{\gamma}}, \quad \theta - \beta < x < \theta, \text{and } \beta, \theta, \gamma > 0,$$

where θ is the reflecting parameter that will reflect the distribution towards positive skewed to negative skewed or negative skewed to positive skewed. Also γ and β are the shape and scale parameters, respectively.

The RPFD is more flexible to fit the data sets obtained from the medical and engineering sciences,

The RPFD can be negative-skewed or positive-skewed, whereas the HRF can be J-shape, monotonically increasing and decreasing shapes, which shows its more flexible nature to fit any data set. The graphs for the probability density function and hazard rate function (HRF) are given in Figure 1 and Figure 2, respectively.

We see that the shape parameter of the RPFD plays a vital role in defining the shape of any process in reliability, medical, and engineering sciences. A little shift in the shape parameter affects the shape of the distribution of any process. It is essential to control these small shifts in the process. We use statistical process control (SPC)

* Corresponding author (azamzka@gmail.com; https://orcid.org/0000-0002-0497-2304)

to control the small shifts in the process that follows reflected power function distribution.
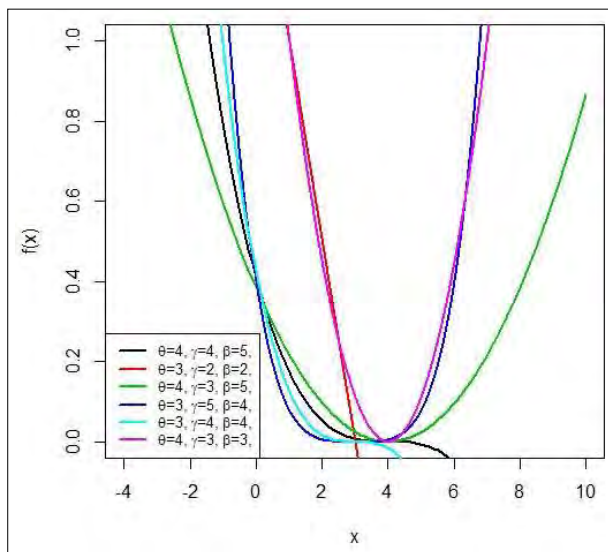
Figure 1:   PDF plots for RPFD.
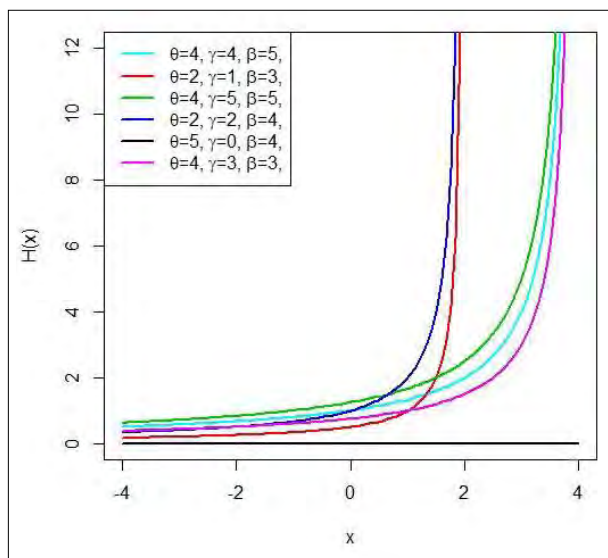
Figure 2:   HRF plots for RPFD

Quality control issues have gained the attention of scholars and practitioners in the recent era. It is an important topic of discussion as the credibility of firms depends on it, and it is also considered a source of customer retention. But in real life, sometimes firms fail to accurately assess the error trends during production, affecting product quality and, ultimately, reputation. That is why the managers and reliability engineers are keen to determine the failure trends for the manufacturing units to confirm the validity of pre-production testing results in real-life. One of the examples is from the pharmaceutical industry, where life-saving drugs are manufactured. The monitoring of the production process is necessary at every stage. The companies adopt strict checking procedures to avoid any mishaps. So, the quality control check in these types of industries is essential not only for the firms' credibility but also for the well-being of humans.

Similarly, another example is from the software industry, which is expected to be error-free when delivered in the market. That is why it has become essential for the product to pass through the careful quality assessment processes during testing. Whenever defects in the machine appear, they will be handled skillfully so that the final product will be perfect. Therefore, the companies try to control the errors, but the underlying distribution of errors may not be normal and often follows a reflected power function distribution (RPFD).

The current study tries to introduce the control charts that can assist the firms in identifying and eliminating the variations in the production process through well-established monitoring of the product reliability when the number of errors follows RPFD.

Prior studies related to the control charts mostly followed the normality assumption, such as the hybrid exponential weighted moving average (HEWMA) control charts due to Shamma and Shamma (1992) and Haq (2013). When the distribution of the process is normal, the extended exponentially weighted moving averages (EEWMA) control chart was introduced by Naveed *et al* (2018). But what can organizations do if the normality assumption fails in practical life. There are very few studies available in the literature that highlighted this issue, including Qiu and Li (2011), Noorossana *et al*. (2016), Zhang *et al*. (2017), Lin *et al*. (2017), Erto *et al*. (2018), Li *et al*. (2018), and Liang *et al*. (2019). As many production processes do not necessarily follow the normality assumption, the study aims to introduce the memory-based control charts, *i.e.*, HEWMA and EEWMA control charts assume that the underlying process distribution follows an RPFD. Further, the proposed control charts are based on the different estimators of the shape parameter of RPFD. We use the shape parameter of the RPFD instead of the usual mean as the process parameter. The percentile estimator (PE) and modified maximum likelihood method (MMLM)

estimator are discussed by Zaka *et al.* (2020) for the shape parameter of RPFD, which is used to construct the plotting statistics. Both are equally efficient, and the comparison is made to see which performs better in statistical process control. Monte Carlo simulation is used to estimate the shape parameter of the RPFD as well as the plotting statistics. Average run lengths are used to evaluate the performances of each of the control charts. It is observed that the proposed EEWMA control chart of the PE detects smaller shifts quicker than the HEWMA control charts.The simulation study and real-life application are discussed in the end.

MATERIALS AND METHODS

In this section, the percentile estimator (PE) and modified maximum likelihood method (MMLM) to estimate the shape parameter of RPFD are discussed using Zaka *et al.* (2020). Then, some traditional HEWMA and EEWMA are modified to monitor the shape parameter of RPFD using PE and MMLM.

Proposed estimators of process monitoring for an RPFD

The following section uses PE and MMLM to construct memory-based control charts to monitor the shape parameter of a process that follows an RPFD. From Zaka *et al.* (2020), it is assumed that the process random variables $x_1, x_2, x_3, ..., x_t$ are independently and identically distributed following RPFD, with the probability density function (PDF) and cumulative density function (CDF) given respectively by

$$f(x) = \frac{\gamma(\theta - x)^{\gamma-1}}{\beta^{\gamma}}, \theta - \beta < x < \theta, \text{ and } \beta, \theta, \gamma > 0,$$

and

$$F(x) = 1 - \frac{(\theta - x)^{\gamma}}{\beta^{\gamma}},$$

where θ is the reflecting parameter that will reflect the distribution towards positive skewed to negative skewed or negative skewed to positive skewed. Also $\gamma \text{ and } \beta$ are the shape and scale parameters.

     MMLM and PE estimators for the shape parameter of RPFD defined by Zaka *et al.* (2020) are given below.

$$\hat{\gamma}_{MMLM} = \left(\frac{n(1 + ln(0.5))}{(n \, ln(\theta - \tilde{x}) - \sum_{i=1}^{n} ln(\theta - x_i))} \right). \qquad ...(1)$$

$$\hat{\gamma}_{PE} = \frac{ln\left(\frac{1-H}{1-L}\right)}{ln\left(\frac{\theta - P_H}{\theta - P_L}\right)}. \qquad ...(2)$$

Traditional hybrid exponentially weighted moving average (HEWMA) control chart

Let the distribution of the underlying process having the sequence $\{X_i\}$ be normal. Also, let $0 \le \lambda_i \le 1$ for i = 1, 2 be a known constant. Now, consider a new sequence HW_t as

$$HW_t = \lambda_1 W_t + (1 - \lambda_1)HW_{t-1}, \qquad ...(3)$$

where

$$W_t = \lambda_2 \hat{\mu}_t + (1 - \lambda_2)W_{t-1}, \qquad ...(4)$$

where

$HW_0 = W_0 = \mu$ (population mean) and HW_t is a plotting statistic. By placing (4) in (3), we get the following

$$HW_t = \lambda_1 \lambda_2 \sum_{i=0}^{t-1} (1 - \lambda_1)^i \sum_{j=0}^{t-i-1} \left(\frac{1-\lambda_1}{1-\lambda_2}\right)^j Y_i +$$

$$\lambda_1 \sum_{i=0}^{t-1} (1 - \lambda_1)^i (1 - \lambda_2)^{t-i} \mu + (1 - \lambda_1)^t \mu.$$

The mean and the variance of HW_t are given below as

$$E(HW_t) = \mu$$

where μ is the population mean and

$$V(HW_t) = \left(\frac{\lambda_1 \lambda_2}{(\lambda_1 - \lambda_2)}\right)^2 \left[\sum_{i=1}^{2} (1 - \lambda_1)^2 (1 - (1 - \lambda_i)^{2t}/1 - (1 - \lambda_i)^2) \right.$$

$$\left. - \frac{2(1 - \lambda_1)(1 - \lambda_2)\{1 - (1 - \lambda_1)^t (1 - \lambda_2)^t\}}{1 - (1 - \lambda_1)(1 - \lambda_2)} \right] \sigma^2,$$

Where σ^2 is the population variance.

The control limits for the HEWMA control chart are explained as

$$UCL_{HW_t} = \mu + L\sqrt{V(HW_t)}$$

and

$$UCL_{HW_t} = \mu + L\sqrt{V(HW_t)}.$$

Proposed HEWMA control chart based on PE of the shape parameter of RPFD

The HEWMA statistic using PE of the shape parameter of the RPFD is given by

$$HEWPE_t = \lambda_1 EWPE_t + (1 - \lambda_1)HEWPE_{t-1}, \qquad ...(5)$$

where $EWPE_t$ is a usual EWMA statistic based on PE, given as

$$EWPE_t = \lambda_2 \hat{\gamma}_{PE(t)} + (1 - \lambda_2)EWPE_{t-1}, \qquad ...(6)$$

and $HEWPE_{t-1}$ is the statistic on previous time. Also λ_1 and λ_2 are smoothing constants here. The control limits are given for the HEWMA control chart using PE for the shape parameter of the RPFD; we get,

$$HEWPE_t = \lambda_1 \left(\lambda_2 \hat{\gamma}_{PE(t)} + (1 - \lambda_2) \sum_{j=0}^{1} \left(\frac{(1-\lambda_1)}{(1-\lambda_2)} \right)^j \hat{\gamma}_{PE(t-1)} \right.$$

$$+ (1 - \lambda_2)^2 \sum_{j=0}^{2} \left(\frac{(1-\lambda_1)}{(1-\lambda_2)} \right)^j \hat{\gamma}_{PE(t-2)} +$$

$$(1 - \lambda_2)^3 \sum_{j=0}^{3} \left(\frac{(1-\lambda_1)}{(1-\lambda_2)} \right)^j \hat{\gamma}_{PE(t-3)} + ... +$$

$$(1 - \lambda_2)^{t-1} \sum_{j=0}^{t-1} \left(\frac{(1-\lambda_1)}{(1-\lambda_2)} \right)^j \hat{\gamma}_{PE(1)} \right) +$$

$$\lambda_1(1 - \lambda_2)^t \sum_{j=0}^{t} \left(\frac{(1-\lambda_1)}{(1-\lambda_2)} \right)^j \hat{\gamma}_{PE(0)} +$$

$$(1 - \lambda_2)^t HEWPE_0.$$

$$LCL_{HEWPE_t} = \gamma - L$$
$$* \frac{\lambda_1 \lambda_2}{(\lambda_1 - \lambda_2)} \sqrt{ \left(\sum_{i=1}^{2} \frac{(1-\lambda_i)^2(1-(1-\lambda_i)^{2t})}{1-(1-\lambda_i)^2} - \frac{2(1-\lambda_1)(1-\lambda_2)\{1-(1-\lambda_1)^t(1-\lambda_2)^t\}}{1-(1-\lambda_1)(1-\lambda_2)} \right) var(\hat{\gamma}_{PE}) }$$

$$CL_{HEWPE_t} = \gamma$$

$$UCL_{HEWPE_t} = \gamma + L$$
$$* \frac{\lambda_1 \lambda_2}{(\lambda_1 - \lambda_2)} \sqrt{ \left(\sum_{i=1}^{2} \frac{(1-\lambda_i)^2(1-(1-\lambda_i)^{2t})}{1-(1-\lambda_i)^2} - \frac{2(1-\lambda_1)(1-\lambda_2)\{1-(1-\lambda_1)^t(1-\lambda_2)^t\}}{1-(1-\lambda_1)(1-\lambda_2)} \right) var(\hat{\gamma}_{PE}) }.$$

Proposed HEWMA control chart of the shape parameter of RPFD using MMLM

Let $x_1, x_2, ..., x_n$ be a sequence of independent and identical random variables generated from a process which follows an RPFD with shape parameter "γ". It is important to note that we use the estimator of the shape parameter of the process instead of an average of observations or single observations by assuming that $E(\hat{\gamma}_{MMLM}) = \gamma$.

The HEWMA statistic by using MMLM of RPFD may be written as

$$HEWMMLM_t = \lambda_1 EW_t + (1 - \lambda_1)HEWMMLM_{t-1} \qquad ...(7)$$

where PEW_t is a usual EWMA statistic given as

$$EWMMLM_t = \lambda \hat{\gamma}_{MMLM(t)} + (1 - \lambda)EWMMLM_{t-1},$$

where $\hat{\gamma}_{MMLM}$ is the modified maximum likelihood estimator for RPFD and $EWMMLM_{t-1}$ is the statistic on previous time. Also "λ" is a smoothing constant. We may get the control limits for HEWMA as

$$LCL_{HEWMMLM_t} = \gamma - L \left(\frac{\lambda_1 \lambda_2}{(\lambda_1 - \lambda_2)} \sqrt{ \left(\sum_{i=1}^{2} \frac{(1-\lambda_i)^2(1-(1-\lambda_i)^{2t})}{1-(1-\lambda_i)^2} - \frac{2(1-\lambda_1)(1-\lambda_2)\{1-(1-\lambda_1)^t(1-\lambda_2)^t\}}{1-(1-\lambda_1)(1-\lambda_2)} \right) Var(\hat{\gamma}_{MMLM}) } \right), \qquad ...(8)$$

$$\text{CL}_{HEWMMLM_t} = \gamma$$

$$\text{JCL}_{HEWMMLM_t} = \gamma + L\left(\frac{\lambda_1\lambda_2}{(\lambda_1-\lambda_2)}\sqrt{\left(\sum_{i=1}^{2}\frac{(1-\lambda_i)^2(1-(1-\lambda_i)^{2t})}{1-(1-\lambda_i)^2} - \frac{2(1-\lambda_1)(1-\lambda_2)\{1-(1-\lambda_1)^t(1-\lambda_2)^t\}}{1-(1-\lambda_1)(1-\lambda_2)}\right)\text{Var}(\hat{\gamma}_{MMLM})}\right). \qquad \dots(9)$$

Algorithm for HEWMA control charts under PE and MMLM

1. Generate random samples of size $n = 150$ on "X_t" from the RPFD where with parameters $(\beta, \gamma, \theta) = (1, 2, 1)$.
2. Compute $\hat{\gamma}_{*t}$ where $* = \text{PE}$, and MMLM.
3. Repeat steps 1 and 2 5000 times and compute $\hat{\gamma}_{*t}$, $E(\hat{\gamma}_{*t})$ and $V(\hat{\gamma}_{*t})$.
4. Repeat step 3 5000 times and compute $\hat{\gamma}_{*t}$.
5. Compute control limits to construct HEWMA control charts based on $\hat{\gamma}_{*t}$.
6. Compute the ARL value for each HEWMA control chart that is based on $\hat{\gamma}_{*t}$, given that process is an in-control state.
7. Now fix $\text{ARL}_0 = 500$ for the in-control state of the process, and search the suitable value of L so that ARL_0 for the in-control state of a process is achieved.
8. Now assume that the process parameter γ is shifted by its true value and compute ARL_s. This step is repeated for different shift values 0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.25 0.30, and 0.35. The shifts are selected based on the change observed in the shape parameter during the simulation process. Also, compute ARL_1 in each case of shift values.
9. Plot ARL_s values against the values of shift that are used in step 7 & 8.
10. It is to be noted that the procedure of the HEWMA control chart is based on $\hat{\gamma}_{*t}$, observe whether the process following the RPFD is in-control or out of control. If the process is in-control, go to Step 1. Otherwise, record the Run Length, *i.e.*, the process remaining in control before being declared out-of-control.
11. Repeat this process 5000 times to obtain the ARLs, SDRLs, and fractiles.

The traditional extended exponentially weighted moving averages (EEWMA) control chart

When the distribution of process is normal, the EEWMA control chart was introduced by Naveed *et al*. (2018). The EEWMA control chart by Naveed *et al*. (2018) is given as

$$Z_t = \lambda_1 T_t - \lambda_2 T_{t-1} + (1 - \lambda_1 + \lambda_2)Z_{t-1}$$

where $0 \le \lambda_1 \le 1$ and $0 \le \lambda_2 \le \lambda_1$. T_{t-1} represents the previous value of the variable and Z_{t-1} denotes the previous value of statistic.

The mean and the variance are given as

$$E(Z_t) = \mu$$

and

$$var(Z_t) = \sigma^2\left[\left(\lambda_1^2 + \lambda_2^2\right)\left\{\frac{1-(1-\lambda_1+\lambda_2)^{2t}}{2(\lambda_1-\lambda_2)-(\lambda_1-\lambda_2)^2}\right\}\right.$$
$$\left. - 2a\lambda_1\lambda_2\left\{\frac{1-(1-\lambda_1+\lambda_2)^{2t-2}}{2(\lambda_1-\lambda_2)-(\lambda_1-\lambda_2)^2}\right\}\right],$$

respectively.

Proposed EEWMA control chart based on PE of the shape parameter of RPFD

The EEWMA statistic using PE of the shape parameter of RPFD using Zaka *et al*. (2020) and Naveed *et al*. (2018) is given by

$$EEWPE_t = \lambda_1\hat{\gamma}_{PE(t)} - \lambda_2\hat{\gamma}_{PE(t-1)} + (1 - \lambda_1 + \lambda_2)EEWPE_{t-1}.$$

For t = 1

$$EEWPE_1 = \lambda_1\hat{\gamma}_{PE(1)} - \lambda_2\hat{\gamma}_{PE(0)}(1 - \lambda_1 + \lambda_2)EEWPE_0.$$

For t = 2

$$EEWPE_2 = \lambda_1\hat{\gamma}_{PE(2)} - \lambda_2\hat{\gamma}_{PE(1)}(1 - \lambda_1 + \lambda_2)EEWPE_1.$$

$$EEWPE_2 = \lambda_1\hat{\gamma}_{PE(2)} - \lambda_2\hat{\gamma}_{PE(1)}$$
$$+ (1 - \lambda_1 + \lambda_2)\big(\lambda_1\hat{\gamma}_{PE(1)} - \lambda_2\hat{\gamma}_{PE(0)}$$
$$+ (1 - \lambda_1 + \lambda_2)EEWPE_0\big).$$

Let $a = (1 - \lambda_1 + \lambda_2)$

$$EEWPE_2 = \lambda_1\hat{\gamma}_{PE(2)} + b\hat{\gamma}_{PE(1)} - a\lambda_2\hat{\gamma}_{PE(0)} + a^2EEWPE_0.$$

$$EEWPE_2 = \lambda_1\hat{\gamma}_{PE(2)} + (a\lambda_1 - \lambda_2)\hat{\gamma}_{PE(1)} - a\lambda_2\hat{\gamma}_{PE(0)}$$
$$+ a^2EEWPE_0.$$

Let $b = (a\lambda_1 - \lambda_2)$.

$$EEWPE_2 = \lambda_1\hat{\gamma}_{PE(2)} + b\hat{\gamma}_{PE(1)} - a\lambda_2\hat{\gamma}_{PE(0)} + a^2EEWPE_0.$$

$$EEWPE_3 = \lambda_1 \hat{\gamma}_{PE(3)} + b\hat{\gamma}_{PE(2)} + ab\hat{\gamma}_{PE(1)} - a^2 \lambda_2 \hat{\gamma}_{PE(0)} + a^3 EEWPE_0.$$

$$EEWPE_t = \lambda_1 \hat{\gamma}_{PE(t)} + b\hat{\gamma}_{PE(t-1)} + ab\hat{\gamma}_{PE(t-2)} + a^2 b\lambda_2 \hat{\gamma}_{PE(t-3)}$$
$$+ a^3 \lambda_2 \hat{\gamma}_{PE(0)} + \dots + a^{t-2} b\hat{\gamma}_{PE(1)}$$
$$- a^{t-1} \lambda_2 \hat{\gamma}_{PE(0)} + a^t EEWPE_0.$$

By taking expectation, we get

$$E(EEWPE_t) = \lambda_1 \gamma + b\gamma + ab\gamma + a^2 b\lambda_2 \gamma + a^3 \lambda_2 \gamma + \dots + a^{t-2} b\gamma - a^{t-1} \lambda_2 \gamma + a^t \gamma.$$

Replacing $b = (a\lambda_1 - \lambda_2)$,

$$E(EEWPE_t) = \gamma\big(\lambda_1 + (a\lambda_1 - \lambda_2) + a(a\lambda_1 - \lambda_2) + a^2(a\lambda_1 - \lambda_2) + \dots + a^{i-2}a^2(a\lambda_1 - \lambda_2) - a^{t-1}\lambda_2 + a^t\big).$$

$$E(EEWPE_t) = \gamma\{\lambda_1(1 + a + a^2 + a^3 + \dots + a^{t-1}) - \lambda_2(1 + a + a^2 + a^3 + \dots + a^{t-1}) + a^t\}.$$

$$E(EEWPE_t) = \gamma\{(\lambda_1 - \lambda_2)(1 + a + a^2 + a^3 + \dots + a^{t-1}) + a^t\}.$$

By using geometric series, we get

$$E(EEWPE_t) = \gamma\left\{(\lambda_1 - \lambda_2)\left(\frac{1-a^t}{1-a}\right) + a^t\right\}.$$

So, $E(EEWPE_t) = \gamma\{1 - a^t + a^t\}$

$$E(EEWPE_t) = \gamma \qquad \dots(10)$$

Since we know $Var(\hat{\gamma}_{PE(t)}) = V_{PE} = E(\hat{\gamma}_{PE} - \gamma)^2$, we get

$$Var(EEWPE_t) = \lambda_1^2 V_{PE} + b^2 V_{PE} + a^2 b^2 V_{PE} + a^4 b^2 \lambda_2^2 V_{PE} + \dots + a^{2(t-2)} b^2 V_{PE} - a^{2(t-1)} \lambda_2^2 V_{PE}.$$

Let $b = (a\lambda_1 - \lambda_2)$. Then,

$$Var(EEWPE_t) = V_{PE}\{\lambda_1^2 + (a\lambda_1 - \lambda_2)^2 + a^2(a\lambda_1 - \lambda_2)^2 + a^4(a\lambda_1 - \lambda_2)^2 + \dots + a^{2(t-2)}(a\lambda_1 - \lambda_2)^2 - a^{2(t-1)}\lambda_2^2\}.$$

$$Var(EEWPE_t) = V_{PE}\{\lambda_1^2 + a^2\lambda_1^2 - 2a\lambda_1 + \lambda_2^2 + a^4\lambda_1^2 - 2a^3\lambda_1 + a^2\lambda_2^2 + (a^6\lambda_1^2 - 2a^5\lambda_1^2 + a^4\lambda_2^2) + \dots + a^{2(t-2)}\lambda_1^2 - 2a^{2t-3}\lambda_1 + a^{2(t-2)}\lambda_2^2 + a^{2(t-1)}\lambda_2^2\}.$$

$$Var(EEWPE_t) = V_{PE}\{\lambda_1^2 + a^2\lambda_1^2 - 2a\lambda_1 + \lambda_2^2 + a^4\lambda_1^2 - 2a^3\lambda_1 + a^2\lambda_2^2 + (a^6\lambda_1^4 - 2a^5\lambda_1^3 + a^4\lambda_2^4) + \dots + a^{2(t-2)}\lambda_1^2 - 2a^{2t-3}\lambda_1 + a^{2(t-2)}\lambda_2^2 + a^{2(t-1)}\lambda_2^2\}.$$

$$Var(EEWPE_t) = V_{PE}\left\{(\lambda_1^2 + \lambda_2^2)\left(\frac{1-a^{2t}}{1-a^2}\right) - 2a\lambda_1\lambda_2\left(\frac{1-a^{2t-2}}{1-a^2}\right)\right\}.$$

$$Var(EEWPE_t) = V_{PE}\left\{(\lambda_1^2 + \lambda_2^2)\left(\frac{1-(1-\lambda_1+\lambda_2)^{2t}}{1-(1-\lambda_1+\lambda_2)^2}\right) - 2a\lambda_1\lambda_2\left(\frac{1-(1-\lambda_1+\lambda_2)^{2t-2}}{1-(1-\lambda_1+\lambda_2)^2}\right)\right\} \qquad \dots(11)$$

The control limits are given as

$$UCL_{EEWPE_t}$$
$$= \gamma + L\sqrt{V_{PE}\left\{(\lambda_1^2 + \lambda_2^2)\left(\frac{1-(1-\lambda_1+\lambda_2)^{2t}}{1-(1-\lambda_1+\lambda_2)^2}\right) - 2a\lambda_1\lambda_2\left(\frac{1-(1-\lambda_1+\lambda_2)^{2t-2}}{1-(1-\lambda_1+\lambda_2)^2}\right)\right\}}$$

$$CL_{EEWPE_t} = \gamma$$

$$LCL_{EEWPE_t}$$
$$= \gamma - L\sqrt{V_{PE}\left\{(\lambda_1^2 + \lambda_2^2)\left(\frac{1-(1-\lambda_1+\lambda_2)^{2t}}{1-(1-\lambda_1+\lambda_2)^2}\right) - 2a\lambda_1\lambda_2\left(\frac{1-(1-\lambda_1+\lambda_2)^{2t-2}}{1-(1-\lambda_1+\lambda_2)^2}\right)\right\}}$$

Proposed EEWMA control chart for shape parameter of RPFD using MMLM

Let $x_1, x_2, \dots, x_n$ be a sequence of independent and identical random variables generated from a process which follows a RPFD with shape parameter γ. It is important to note that we use MMLM of the shape parameter of the process instead of the average of observations assuming that $E(\hat{\gamma}_{MMLE}) = \gamma$.

The EEWMA statistic using MMLM of the shape parameter of RPFD using Zaka *et al.* (2020) and Naveed *et al.* (2018) is given by

By using the mean and variance for the EEWMA statistic given in (10) and (11), the control limits are given as

UCL_{EEWMMLM_t}

$$= \gamma + L \sqrt{V_{\text{MMLM}}\left\{(\lambda_1{}^2 + \lambda_2{}^2)\left(\frac{1 - (1 - \lambda_1 + \lambda_2)^{2t}}{1 - (1 - \lambda_1 + \lambda_2)^2}\right) - 2a\lambda_1\lambda_2\left(\frac{1 - (1 - \lambda_1 + \lambda_2)^{2t-2}}{1 - (1 - \lambda_1 + \lambda_2)^2}\right)\right\}}$$

$$CL_{\text{EEWMMLM}_t} = \gamma$$

UCL_{EEWMMLM_t}

$$= \gamma + L \sqrt{V_{\text{MMLM}}\left\{(\lambda_1{}^2 + \lambda_2{}^2)\left(\frac{1 - (1 - \lambda_1 + \lambda_2)^{2t}}{1 - (1 - \lambda_1 + \lambda_2)^2}\right) - 2a\lambda_1\lambda_2\left(\frac{1 - (1 - \lambda_1 + \lambda_2)^{2t-2}}{1 - (1 - \lambda_1 + \lambda_2)^2}\right)\right\}}$$

$$CL_{\text{EEWMMLM}_t} = \gamma$$

Algorithm for EEWMA control charts under PE and MMLM

1. Generate a random sample of size $n_1 = 150$ on X_t from the RPFD, *i.e.*, $x = \theta - \beta(1 - R)^{\overline{\gamma}}$ with parameters $(\beta, \gamma, \theta) = (1, 2, 1)$.
2. Compute $\hat{\gamma}_*$ where $* = $ PE and MMLM.
3. Repeat steps 1 and 2 5000 times and compute $E(\hat{\gamma}_*)$ and $V(\hat{\gamma}_*)$.
4. Repeat step 3 5000 times and compute the average of $E(\hat{\gamma}_*)$ and $V(\hat{\gamma}_*)$.
5. Compute control limits for EEWMA control chart based on $\hat{\gamma}_*$.
6. Compute ARL value for each EEWMA control chart that based on $\hat{\gamma}_*$ given that process is in the in-control state.
7. Now fix $ARL_0 = 500$ for in-control state of the process and search the suitable value of L so that ARL_0 for in-control state of process is achieved.
8. Now assume that the process parameter γ is shifted from its true value and then compute ARL_1. This step is repeated for different shift values 0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, and 0.35. The shifts are selected based on the change observed in shape parameter during the simulation process. Also compute ARL_1 for each of the shift values.
9. Plot ARL_s values against the values of shift which are used in step 7 and 8.
10. It is to be noted that the procedure of EEWMA control chart is based on $\hat{\gamma}_*$. observe whether the process following the RPFD is in-control or out of control. If the process is in-control, go to Step 1. Otherwise, record the run length (RL), *i.e.*, the process remained in-control before it is declared to be out-of-control.
11. Repeat this process 5000 times to obtain the ARLs, SDRLs, and fractiles.

RESULTS AND DISCUSSION

Interpretation

From Figure 3 to Figure 8, we observe that EEWMA control charts based on PE and MMLM for the shape parameter of RPFD perform more efficiently as compared to the PE and MMLM based HEWMA control charts. In order to compare the performance of EEWMA and HEWMA control charts based on PE and MMLM, different choices of λ_1 (0.3, 0.5 and 0.9) and λ_2 (0.2, 0.6, and 0.75) are used. The ARL values for the EEWMA control chart based on MMLM and PE are given in Table 3 and Table 4. From Table 4, it is observed that for larger of values of λ_1, ARL values are getting larger. For instance, taking $\lambda_1 = 0.30$, $\lambda_2 = 0.20$ and shift = 0.05, ARL for EEWMA based on PE is 5.291, and taking $\lambda_1 = 0.90$, $\lambda_2 = 0.20$, and shift = 0.05, ARL is 6.312. Now from this picture it is clearly noted that the ARL for the EEWMA control chart increases if the value of λ_1 is increased, and this is similar to the kind of performance of the MMLM based EEWMA control chart for larger values of λ_1. However this conclusion remains the same for the PE based HEWMA control chart if, in Table 2, we compare the ARLs taking $\lambda_1 = 0.30$, $\lambda_2 = 0.20$, and shift = 0.05 with the ARLs taking $\lambda_1 = 0.50$, $\lambda_2 = 0.20$, and shift = 0.4. One point that is interesting to note is that the PE based EEWMA control chart is consistently getting smaller ARLs than those of the MMLM based EEWMA and HEWMA control charts for each value of λ_1 as well as for each shift value. It is clear from Figure 4 to Figure 6 that the line of ARLs for the PE based EEWMA control chart remains below the line of ARLs for the MMLM based EEWMA control chart. So we can say that the PE based EEWMA control chart performs more efficiently than the MMLM based EEWMA control charts, to detect an out of control state of a process having RPFD.

The ARLs of EEWMA based on MMLM and PE presented respectively in Tables 3 and 4 are compared with the ARLs of HEWMA based on MMLM and PE. It is clearly observed that the ARLs of PE based EEWMA are smaller than the ARLs of PE based HEWMA, taking shift to be 0.05, $\lambda_1 = 0.3$, and $\lambda_2 = 0.20$. For instance, see that the ARL for PE based EEWMA is 5.291 at shift = 0.4, $\lambda_1 = 0.3$, $\lambda_2 = 0.2$, while the ARL for PE based HEWMA is 18.191 at shift = 0.05 and $\lambda = 0.20$. So it is clear from the ARLs presented in Tables 1 to 4 that ARLs for the PE based EEWMA control charts remains smaller than the PE based HEWMA control chart as well as the MMLM based EEWMA and HEWMA control charts.

Furthermore, it is also observed that use of PE makes each type of control chart more efficient as compared to the control charts that are based on MMLM for monitoring a process which is following RPFD.

Table 1: ARL$_S$ using MMLM estimators for the parameters of RPFD using HEWMA control chart.

Estimation methods		0	0.05	0.10	Shift 0.15	0.20	0.25	0.30	0.35
HEWMA	ARL	500.848	23.08	13.36	12.57	1.10	1.005	1.001	1
$\lambda_1 = 0.30$,	SDRL	483.706	7.5235	1.408	0.5807	0.3130	0.0772	0.0316	0
$\lambda_2 = 0.20$	P10	53.90	6	2	1	1	1	1	1
L= 3.10	P25	157.50	8	2	1	1	1	1	1
	P50	358.50	13	3	2	1	1	1	1
	P75	712.25	18	4	2	1	1	1	1
	P90	1117.50	24	5	2	2	1	1	1
HEWMA	ARL	500.865	30.533	13.19	11.38	1.045	1.002	1	1
$\lambda_1 = 0.30$,	SDRL	511.57	20.6742	1.605	0.530	0.2074	0.0446	0	0
$\lambda_2 = 0.60$	P10	46.00	5	1	1	1	1	1	1
L= 3.08	P25	149.75	9	2	1	1	1	1	1
	P50	334.50	17	3	1	1	1	1	1
	P75	712.25	33	4	2	1	1	1	1
	P90	1159.00	52	5	2	1	1	1	1
HEWMA	ARL	500.618	34.77	13.38	11.41	1.051	1.002	1.001	1
$\lambda_1 = 0.30$,	SDRL	512.501	20.548	1.768	0.551	0.2201	0.0446	0.031	0
$\lambda_2 = 0.75$	P10	46.90	4	1	1	1	1	1	1
L= 3.08	P25	144.25	11	2	1	1	1	1	1
	P50	329.5	18	3	1	1	1	1	1
	P75	713.25	38	4	1	1	1	1	1
	P90	1164.6	59	6	1	1	1	1	1
HEWMA	ARL	500.836	24.50	13.21	11.46	1.067	1.003	1.001	1
$\lambda_1 = 0.5$,	SDRL	492.8587	9.544	1.4527	0.5558	0.2501	0.0547	0.0316	0
$\lambda_2 = 0.20$	P10	56.80	4	1	1	1	1	1	1
L= 8.70	P25	153.75	7	2	1	1	1	1	1
	P50	358.50	11	3	1	1	1	1	1
	P75	707.00	17	4	2	1	1	1	1
	P90	1124.10	28	5	2	1	1	1	1
HEWMA	ARL	500.748	44.614	13.661	11.453	1.059	1.002	1.001	1
$\lambda_1 = 0.5$,	SDRL	496.0317	30.778	2.1005	0.5693	0.2357	0.04469	0.0316	0
$\lambda_2 = 0.60$	P10	53.90	6.0	1	1	1	1	1	1
	P25	146.75	11.0	2	1	1	1	1	1
	P50	337.50	25.0	3	1	1	1	1	1
	P75	708.75	52.0	5	2	1	1	1	1
	P90	1156.20	82.1	6	2	1	1	1	1

- continued from previous page

Estimation methods		Shift							
		0	0.05	0.10	0.15	0.20	0.25	0.30	0.35
HEWMA	ARL	500.222	57.909	14.239	11.526	1.077	1.005	1.001	1
$\lambda_1 = 0.5$,	SDRL	495.0974	45.73822	2.846	0.6047	0.26672	0.07056	1.001	0
$\lambda_2 = 0.75$	P10	60.0	6	1	1	1	1	1	1
L = 11.65	P25	147.0	14	2	1	1	1	1	1
	P50	337.5	33	4	1	1	1	1	1
	P75	711.0	64	5	2	1	1	1	1
	P90	1217.7	107	8	2	1	1	1	1
HEWMA	ARL	500.382	31.324	13.361	11.424	1.053	1.002	1.001	1
$\lambda_1 = 0.90$,	SDRL	515.024	16.76437	1.6695	0.5572	0.2241	0.0446	0.0316	0
$\lambda_2 = 0.20$	P10	46.00	5	1	1	1	1	1	1
L = 9.725	P25	142.00	9	2	1	1	0	1	1
	P50	332.00	16	3	1	1	1	1	1
	P75	712.25	29	4	2	1	1	1	1
	P90	1160.10	46	6	2	1	1	1	1
HEWMA	ARL	500.865	82.193	15.892	11.666	1.12	1.006	1.001	1
$\lambda_1 = 0.90$,	SDRL	486.214	66.673	4.607	0.7368	0.3281	0.0772	0.03162	0
$\lambda_2 = 0.60$	P10	64.9	9.0	2	1	1	1	1	1
L = 4.14	P25	154.0	22.0	3	1	1	1	1	1
	P50	343.5	53.0	5	2	1	1	1	1
	P75	688.0	101.0	8	2	1	1	1	1
	P90	1174.4	160.1	12	3	2	1	1	1
HEWMA	ARL	500.2	98.447	18.41	11.81	1.146	1.007	1.001	1
$\lambda_1 = 0.90$,	SDRL	490.6586	86.438	7.653	0.9214	0.3561	0.08341	0.0316	0
$\lambda_2 = 0.75$	P10	63.9	11.00	2.0	1	1	1	1	1
L = 12.00	P25	155.0	29.00	3.0	1	1	1	1	1
	P50	346.0	66.00	6.0	2	1	1	1	1
	P75	693.0	132.25	11.0	2	1	1	1	1
	P90	1178.7	206.00	18.10	3	2	1	1	1

Table 2: ARL_S using PE estimators for the parameters of RPFD using HEWMA control chart.

Estimation method		Shift							
		0	0.05	0.10	0.15	0.20	0.25	0.30	0.35
HEWMA	ARL	500.291	18.191	15.41	1	1	1	1	1
$\lambda_1 = 0.30$,	SDRL	493.5543	2.246	0.5120	0	0	0	0	0
$\lambda_2 = 0.20$	P10	61.00	2	1	1	1	1	1	1
L = 5.80	P25	139.75	3	1	1	1	1	1	1
	P50	341.00	5	1	1	1	1	1	1
	P75	666.50	6	2	1	1	1	1	1
	P90	1190.50	8	2	1	1	1	1	1
HEWMA	ARL	500.291	27.575	15.161	1	1	1	1	1
$\lambda_1 = 0.30$,	SDRL	515.416	2.604	0.3757	0	0	0	0	0
$\lambda_2 = 0.60$	P10	51.00	2	1	1	1	1	1	1
L = 7.144	P25	141.75	3	1	1	1	1	1	1
	P50	349.50	4	1	1	1	1	1	1
	P75	686.75	6	1	1	1	1	1	1
	P90	1158.00	8	2	1	1	1	1	1

- continued from previous page

Estimation method		0	0.05	0.10	Shift 0.15	0.20	0.25	0.30	0.35
HEWMA $\lambda_1 = 0.30$, $\lambda_2 = 0.75$ L = 7.27	ARL	500.948	21. 83	15.164	1	1	1	1	1
	SDRL	505.03	2.938	0.375	0	0	0	0	0
	P10	51.90	2	1	1	1	1	1	1
	P25	141.75	3	1	0	1	1	1	1
	P50	349.0	4	1	1	1	1	1	1
	P75	681.0	6	1	1	1	1	1	1
	P90	1170.4	9	2	1	1	1	1	1
HEWMA $\lambda_1 = 0.5$, $\lambda_2 = 0.20$ L = 7.00	ARL	500.48	26.814	14.28	1	1	1	1	1
	SDRL	493.1145	2.4391	0.4602	0	0	0	0	0
	P10	65.00	2	1	1	1	1	1	1
	P25	145.75	3	1	1	1	1	1	1
	P50	363.50	5	1	1	1	1	1	1
	P75	673.50	6	2	1	1	1	1	1
	P90	1232.20	8	2	1	1	1	1	1
HEWMA $\lambda_1 = 0.5$, $\lambda_2 = 0.60$ L = 7.35	ARL	500.326	26.104	14.294	1	1	1	1	1
	SDRL	496.0043	3.5320	0.375	0	0	0	0	0
	P10	60	2	1	1	1	1	1	1
	P25	161	3	1	1	1	1	1	1
	P50	349	4	1	1	1	1	1	1
	P75	648	7	1	1	1	1	1	1
	P90	1195	10	2	1	1	1	1	1
HEWMA $\lambda_1 = 0.5$, $\lambda_2 = 0.75$ L = 7.44	ARL	500.405	25.456	11.166	1	1	1	1	1
	SDRL	480.0373	4.0849	0.37760	0	0	0	0	0
	P10	52.00	2	1	1	1	1	1	1
	P25	150	3	1	1	1	1	1	1
	P50	361.00	4	1	1	1	1	1	1
	P75	703.25	7	1	1	1	1	1	1
	P90	1195.20	10	2	1	1	1	1	1
HEWMA $\lambda_1 = 0.90$, $\lambda_2 = 0.20$ L = 7.25	ARL	500.174	24.34	11.228	1	1	1	1	1
	SDRL	490.6912	2.519	0.429	0	0	0	0	0
	P10	56.00	2	1	1	1	1	1	1
	P25	142.75	3	1	1	1	1	1	1
	P50	363.00	4	1	1	1	1	1	1
	P75	708.25	6	1	1	1	1	1	1
	P90	1185.40	8	2	1	1	1	1	1
HEWMA $\lambda_1 = 0.90$, $\lambda_2 = 0.60$ L = 7.59	ARL	500.452	26.589	11.18	1	1	1	1	1
	SDRL	471.5114	5.013	0.3996	0	0	0	0	0
	P10	50.90	2	1	1	1	1	1	1
	P25	150.75	3	1	1	1	1	1	1
	P50	366.50	5	1	1	1	1	1	1
	P75	708.50	9	1	1	1	1	1	1
	P90	1172.60	13	2	1	1	1	1	1
HEWMA $\lambda_1 = 0.90$, $\lambda_2 = 0.75$ L = 7.66	ARL	500.005	28.112	11.191	1	1	1	1	1
	SDRL	478.2393	6.746	0.4274	0	0	0	0	0
	P10	52.0	2	1	1	1	1	1	1
	P25	155.0	3	1	1	1	1	1	1
	P50	353.0	6	1	1	1	1	1	1
	P75	712.50	11	1	1	1	1	1	1
	P90	1192.0	17	2	1	1	1	1	1
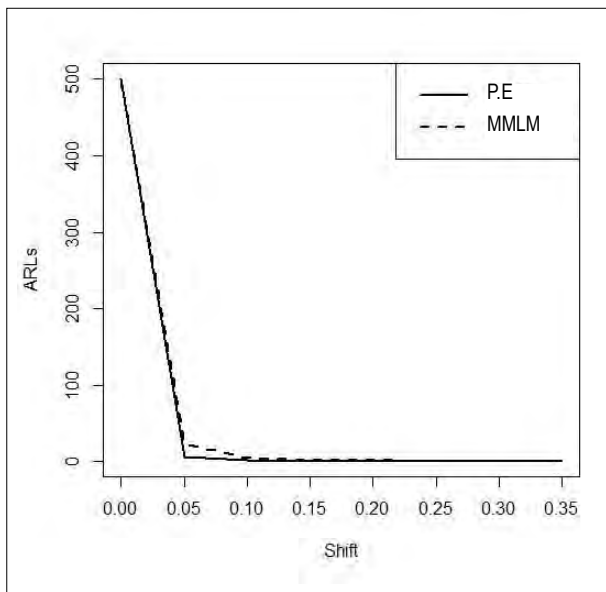
Figure 3: ARLs for the shape parameter of RPFD under HEWMA control chart and $\lambda_1 = 0.90$, $\lambda_2 = 0.20$
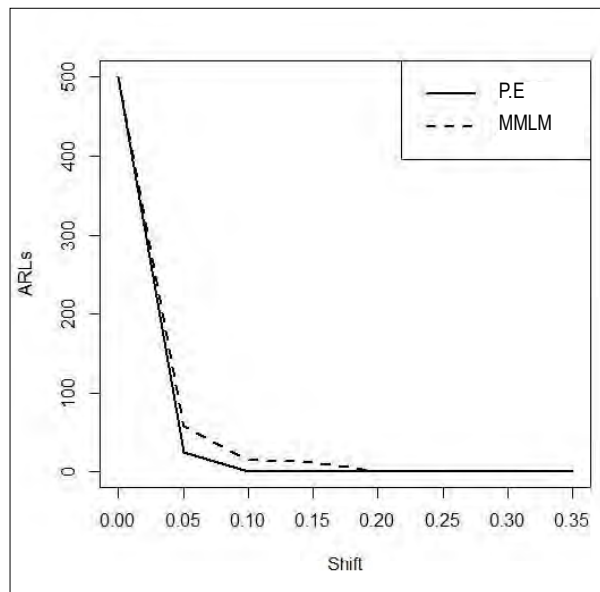
Figure 4: ARLs for the shape parameter of RPFD under HEWMA control chart and $\lambda_1 = 0.50$, $\lambda_2 = 0.75$.
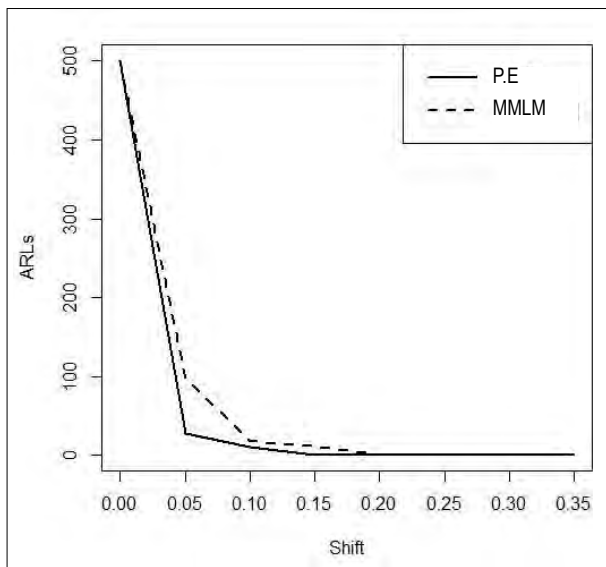
Figure 5: ARLs for the shape parameter of RPFD under HEWMA control chart and $\lambda_1 = 0.90$, $\lambda_2 = 0.75$.

Figure 6: ARLS for MMLM and PE Based EEWMA control charts taking $\lambda_1 = 0.90$, $\lambda_2 = 0.20$

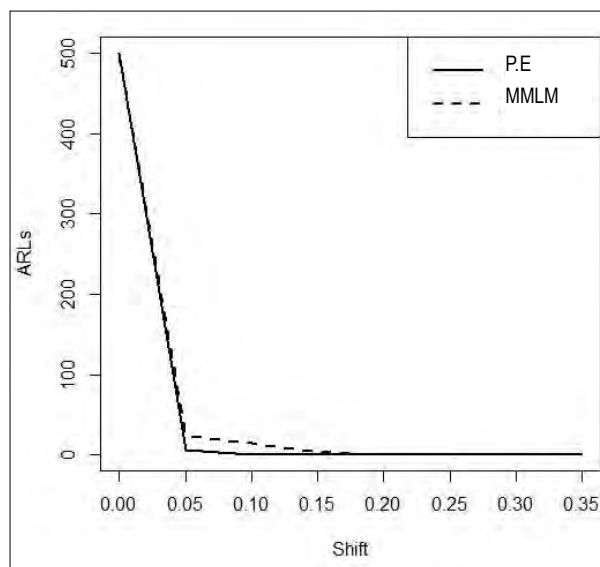**Table 3:** ARLS for MMLM Based EEWMA control charts

Estimation methods						Shift			
		0	0.05	0.10	0.15	0.20	0.25	0.30	0.35
MMLM	ARL	500.4	5.291	2.41	1.571	1.11	1.006	1.001	1
$\lambda_1 = 0.30$,	SDRL	483.7063	2.246	0.5120	0.5807	0.3130	0.0772	0.03162	0
$\lambda_2 = 0.20$	P10	53.90	2	1	1	1	1	1	1
L = 4.30	P25	157.50	3	1	1	1	1	1	1
	P50	358.50	5	1	2	1	1	1	1
	P75	712.25	6	2	2	1	1	1	1
	P90	1117.50	8	2	2	2	1	1	1
MMLM	ARL	500.6	5.575	2.161	1.388	1.045	1.002	1	1
$\lambda_1 = 0.30$,	SDRL	511.57	2.604	0.3757	0.5307	0.2074	0.0446	0	0
$\lambda_2 = 0.60$	P10	46.00	2	1	1	1	1	1	1
L = 3.98	P25	149.75	3	1	1	1	1	1	1
	P50	334.50	4	1	1	1	1	1	1
	P75	712.25	6	1	2	1	1	1	1
	P90	1159.00	8	2	2	1	1	1	1
MMLM	ARL	500.1	5.783	2.164	1.419	1.051	1.002	1.001	1
$\lambda_1 = 0.30$,	SDRL	512.5012	2.938	0.375	0.55112	0.2201	0.0446	0.031	0
$\lambda_2 = 0.75$	P10	46.90	2	1	1	1	1	1	1
L = 3.98	P25	144.25	3	1	1	1	1	1	1
	P50	329.5	4	1	1	1	1	1	1
	P75	713.25	6	1	1	1	1	1	1
	P90	1164.6	9	2	1	1	1	1	1
MMLM	ARL	500.3	5.814	2.28	1.464	1.067	1.003	1.001	1
$\lambda_1 = 0.5$,	SDRL	492.8587	2.4391	0.4602	0.5558	0.2501	0.0547	0.0316	0
$\lambda_2 = 0.20$	P10	56.80	2	1	1	1	1	1	1
L = 8.80	P25	153.75	3	1	1	1	1	1	1
	P50	358.50	5	1	1	1	1	1	1
	P75	707.00	6	2	2	1	1	1	1
	P90	1124.10	8	2	2	1	1	1	1
MMLM	ARL	500.8	6.104	2.294	1.453	1.059	1.002	1.001	1
$\lambda_1 = 0.5$,	SDRL	496.0317	3.5320	0.375	0.5693	0.2357	0.04469	0.0316	0
$\lambda_2 = 0.60$									
	P10	53.90	2	1	1	1	1	1	1
	P25	146.75	3	1	1	1	1	1	1
	P50	337.50	4	1	1	1	1	1	1
	P75	708.75	7	1	2	1	1	1	1
	P90	1156.20	10	2	2	1	1	1	1
MMLM	ARL	500	7.456	2.166	1.526	1.077	1.005	1.001	1
$\lambda_1 = 0.5$,	SDRL	495.0974	4.0849	0.3776	0.6047	0.26672	0.07056	1.001	0
$\lambda_2 = 0.75$,	P10	60.0	2	1	1	1	1	1	1
L = 10.65	P25	147.0	3	1	1	1	1	1	1
	P50	337.5	4	1	1	1	1	1	1
	P75	711.0	7	1	2	1	1	1	1
	P90	1217.7	10	2	2	1	1	1	1

- continued from previous page

Estimation Methods		0	0.05	0.10	0.15	Shift 0.20	0.25	0.30	0.35
MMLM	ARL	500.2	6	3.228	1.2	1.053	1.002	1.001	1
$\lambda_1 = 0.90,$	SDRL	515.02	2.519	0.429	2.519	0.2241	0.0446	0.0316	0
$\lambda_2 = 0.20,$	P10	46.00	2	1	2	1	1	1	1
L = 9.725	P25	142.00	3	1	3	1	0	1	1
	P50	332.00	4	1	4	1	1	1	1
	P75	712.25	6	1	6	1	1	1	1
	P90	1160.10	8	2	8	1	1	1	1
MMLM	ARL	500.6	8.589	4.18	1.666	1.12	1.006	1.001	1
$\lambda_1 = 0.90,$	SDRL	486.214	5.013	0.3996	0.7368	0.3281	0.0772	0.03162	0
$\lambda_2 = 0.60$	P10	64.9	2	1	1	1	1	1	1
L= 4.14	P25	154.0	3	1	1	1	1	1	1
	P50	343.5	5	1	2	1	1	1	1
	P75	688.0	9	1	2	1	1	1	1
	P90	1174.4	13	2	3	2	1	1	1
MMLM	ARL	500	12.112	4.191	1.816	1.146	1.007	1.001	1
$\lambda_1 = 0.90,$	SDRL	490.6586	8.746	0.4274	0.9214	0.3561	0.08341	0.0316	0
$\lambda_2 = 0.75$	P10	63.9	2	1	1	1	1	1	1
L =12.00	P25	155.0	3	1	1	1	1	1	1
	P50	346.0	8	1	2	1	1	1	1
	P75	693.0	11	1	2	1	1	1	1
	P90	1178.7	17	2	3	2	1	1	1

Table 4: ARLS for PE Based EEWMA control charts.

Estimation method		0	0.05	0.10	0.15	Shift 0.20	0.25	0.30	0.35
PE	ARL	500.01	5.01	1.05	1	1	1	1	1
$\lambda_1 = 0.35,$	SDRL	490.525	3.050	0.896	0	0	0	0	0
$\lambda_2 = 0.25$	P10	61.00	3	1	1	1	1	1	1
L= 4.2	P25	138.5	4	1	1	1	1	1	1
	P50	352.00	5	1	1	1	1	1	1
	P75	600.50	6	1	1	1	1	1	1
	P90	1050.50	8	1	1	1	1	1	1
PE	ARL	500.01	3.575	2.161	1	1	1	1	1
$\lambda_1 = 0.35,$	SDRL	505.416	1.604	0.567	0	0	0	0	0
$\lambda_2 = 0.60$	P10	52.20	2	1	1	1	1	1	1
L= 4.3	P25	140.5	3	1	1	1	1	1	1
	P50	350.50	4	1	1	1	1	1	1
	P75	686.75	6	1	1	1	1	1	1
	P90	1158.00	8	2	1	1	1	1	1

- continued from previous page

Estimation method		0	0.05	0.10	0.15	Shift 0.20	0.25	0.30	0.35
PE	ARL	500.01	4.783	1.164	1	1	1	1	1
$\lambda_1 = 0.35$,	SDRL	504.03	2.938	0.375	0	0	0	0	0
$\lambda_2 = 0.75$	P10	53.90	2	1	1	1	1	1	1
$L = 4.22$	P25	138.75	3	1	1	1	1	1	1
	P50	355.0	4	1	1	1	1	1	1
	P75	675.0	6	1	1	1	1	1	1
	P90	1150.4	9	2	1	1	1	1	1
PE	ARL	500.01	5.814	1.12	1	1	1	1	1
$\lambda_1 = 0.5$,	SDRL	490.05	2.4391	0.4602	0	0	0	0	0
$\lambda_2 = 0.25$	P10	55.00	2	1	1	1	1	1	1
$L = 5.00$	P25	185.75	3	1	1	1	1	1	1
	P50	353.50	5	1	1	1	1	1	1
	P75	670.50	6	2	1	1	1	1	1
	P90	1232.20	8	2	1	1	1	1	1
PE	ARL	500.01	4.105	1.34	1	1	1	1	1
$\lambda_1 = 0.5$,	SDRL	496.0043	2.520	0.375	0	0	0	0	0
$\lambda_2 = 0.60$	P10	60	2	1	1	1	1	1	1
$L = 5.05$	P25	161	3	1	1	1	1	1	1
	P50	349	4	1	1	1	1	1	1
	P75	648	7	1	1	1	1	1	1
	P90	1195	10	2	1	1	1	1	1
PE	ARL	500.01	5.456	1.166	1	1	1	1	1
$\lambda_1 = 0.5$,	SDRL	485.0373	4.0849	0.37760	0	0	0	0	0
$\lambda_2 = 0.75$	P10	51.50	2	1	1	1	1	1	1
$L = 5.2$	P25	150	3	1	1	1	1	1	1
	P50	360.00	4	1	1	1	1	1	1
	P75	702.25	7	1	1	1	1	1	1
	P90	1185.20	10	2	1	1	1	1	1
PE	ARL	500.01	6.312	1.208	1	1	1	1	1
$\lambda_1 = 0.90$,	SDRL	493.912	3.419	0.429	0	0	0	0	0
$\lambda_2 = 0.20$	P10	52.00	3	1	1	1	1	1	1
$L = 5.25$	P25	132.75	4	1	1	1	1	1	1
	P50	361.50	5	1	1	1	1	1	1
	P75	705.15	6	1	1	1	1	1	1
	P90	1175.40	8	2	1	1	1	1	1
PE	ARL	500.01	6.589	1.15	1	1	1	1	1
$\lambda_1 = 0.90$,	SDRL	471.5114	5.013	0.4996	0	0	0	0	0
$\lambda_2 = 0.60$	P10	50.90	2	1	1	1	1	1	1
$L = 4.59$	P25	150.75	3	1	1	1	1	1	1
	P50	366.50	5	1	1	1	1	1	1
	P75	708.50	9	1	1	1	1	1	1
	P90	1172.60	13	2	1	1	1	1	1
PE	ARL	500	8.112	1.191	1	1	1	1	1
$\lambda_1 = 0.90$,	SDRL	478.2393	6.746	0.4274	0	0	0	0	0
$\lambda_2 = 0.75$,	P10	52.0	2	1	1	1	1	1	1
$L = 4.62$	P25	155.0	3	1	1	1	1	1	1
	P50	353.0	6	1	1	1	1	1	1
	P75	712.50	11	1	1	1	1	1	1
	P90	1192.0	17	2	1	1	1	1	1

Table 5: Simulated Data

| EEWMA for MMLM			EEWMA for PE		
λ_1=0.90		λ_2=0.75	λ_1=0.90		λ_2=0.75
L=12			L=7.66		
EEWMMLM$_t$	LCL	UCL	EEWPE$_t$	LCL	UCL
2.792763	0.2380000	3.802000	2.174939	1.389540	2.630460
2.757910	0.1320048	3.907995	2.055011	1.352634	2.667366
2.890396	0.1240272	3.915973	1.884097	1.349857	2.670143
2.031693	0.1234880	3.916512	1.851996	1.349669	2.670331
1.607408	0.1234533	3.916547	1.845604	1.349657	2.670343
2.457326	0.1234511	3.916549	2.209770	1.349656	2.670344
2.706825	0.1234509	3.916549	2.468172	1.349656	2.670344
2.578589	0.1234509	3.916549	2.109744	1.349656	2.670344
2.277098	0.1234509	3.916549	2.124111	1.349656	2.670344
2.033351	0.1234509	3.916549	2.082389	1.349656	2.670344
2.089141	0.1234509	3.916549	2.135626	1.349656	2.670344
2.007338	0.1234509	3.916549	1.958650	1.349656	2.670344
2.080522	0.1234509	3.916549	1.994993	1.349656	2.670344
1.995012	0.1234509	3.916549	2.039405	1.349656	2.670344
1.985600	0.1234509	3.916549	1.999159	1.349656	2.670344
2.404916	0.1234509	3.916549	1.954418	1.349656	2.670344
1.881724	0.1234509	3.916549	2.070869	1.349656	2.670344
2.660258	0.1234509	3.916549	2.120229	1.349656	2.670344
2.194834	0.1234509	3.916549	2.278065	1.349656	2.670344
2.066108	0.1234509	3.916549	2.277322	1.349656	2.670344
2.170824	0.1234509	3.916549	2.277124	1.349656	2.670344
1.925193	0.1234509	3.916549	2.230325	1.349656	2.670344
1.711615	0.1234509	3.916549	2.109841	1.349656	2.670344
2.086559	0.1234509	3.916549	2.112682	1.349656	2.670344
2.466933	0.1234509	3.916549	2.196142	1.349656	2.670344
3.217173	0.1234509	3.916549	2.478410	1.349656	2.670344
3.268923	0.1234509	3.916549	2.428640	1.349656	2.670344
3.453694	0.1234509	3.916549	2.249904	1.349656	2.670344
2.431717	0.1234509	3.916549	2.218243	1.349656	2.670344
1.925132	0.1234509	3.916549	2.212262	1.349656	2.670344
2.943783	0.1234509	3.916549	2.649364	1.349656	2.670344
3.242825	0.1234509	3.916549	2.959301	1.349656	2.670344
3.089214	0.1234509	3.916549	2.529532	1.349656	2.670344
2.728023	0.1234509	3.916549	2.546798	1.349656	2.670344
2.436000	0.1234509	3.916549	2.496781	1.349656	2.670344
2.502835	0.1234509	3.916549	2.560622	1.349656	2.670344
2.404841	0.1234509	3.916549	2.348429	1.349656	2.670344
2.492511	0.1234509	3.916549	2.391968	1.349656	2.670344
2.390080	0.1234509	3.916549	2.445245	1.349656	2.670344
2.378799	0.1234509	3.916549	2.396851	1.349656	2.670344
2.881183	0.1234509	3.916549	2.343292	1.349656	2.670344
2.254360	0.1234509	3.916549	2.482961	1.349656	2.670344
3.187066	0.1234509	3.916549	2.542158	1.349656	2.670344
2.629467	0.1234509	3.916549	2.731320	1.349656	2.670344
2.475245	0.1234509	3.916549	2.730486	1.349656	2.670344
2.600721	0.1234509	3.916549	2.730273	1.349656	2.670344
2.306430	0.1234509	3.916549	2.674169	1.349656	2.670344
2.050553	0.1234509	3.916549	2.529708	1.349656	2.670344
2.499746	0.1234509	3.916549	2.533056	1.349656	2.670344
2.955474	0.1234509	3.916549	2.633164	1.349656	2.670344
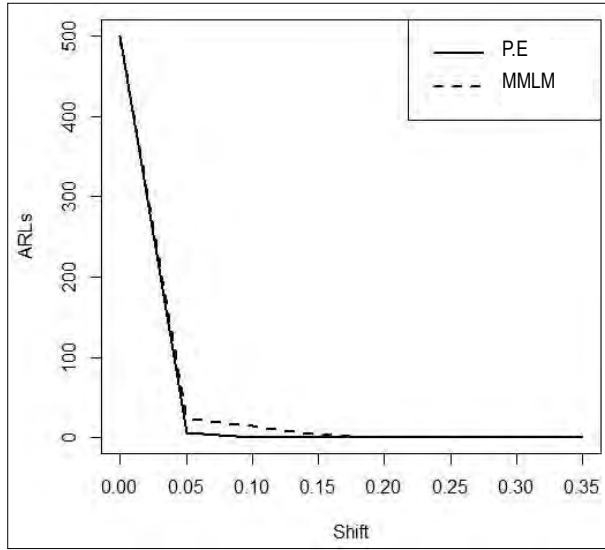
Figure 7: ARLS for MMLM and PE Based EEWMA control charts taking $\lambda_1 = 0.50$, $\lambda_2 = 0.75$.
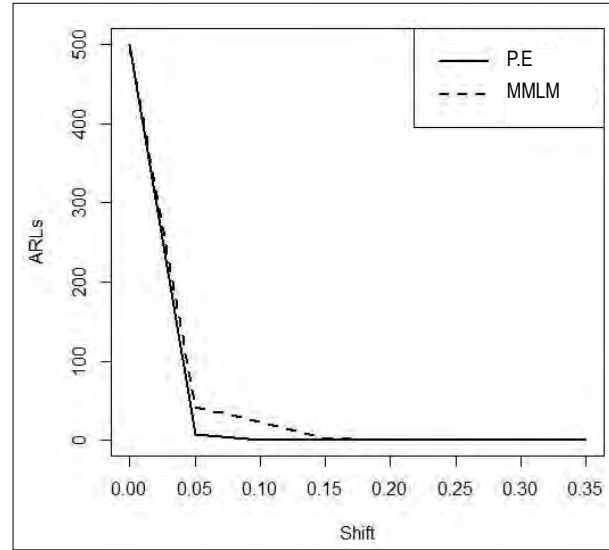
Figure 8: ARLS for MMLM and PE Based EEWMA control charts taking $\lambda_1 = 0.90$, $\lambda_2 = 0.75$.
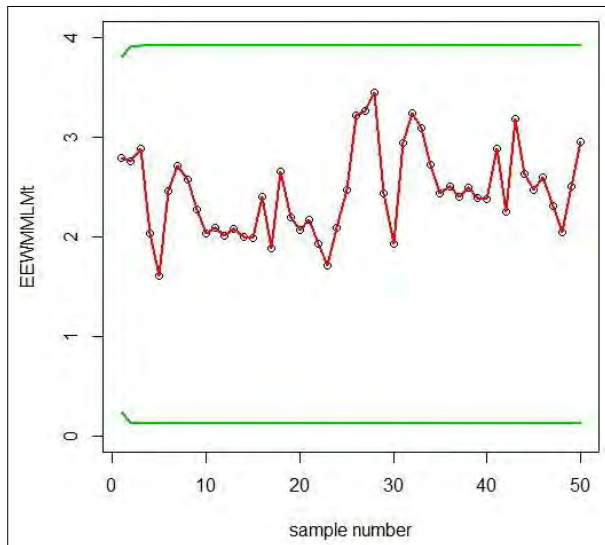
Figure 9: Graph of simulated data of the proposed EEWMA control chart under MMLM when $\lambda_1 = 0.90$ and $\lambda_2 = 0.75$
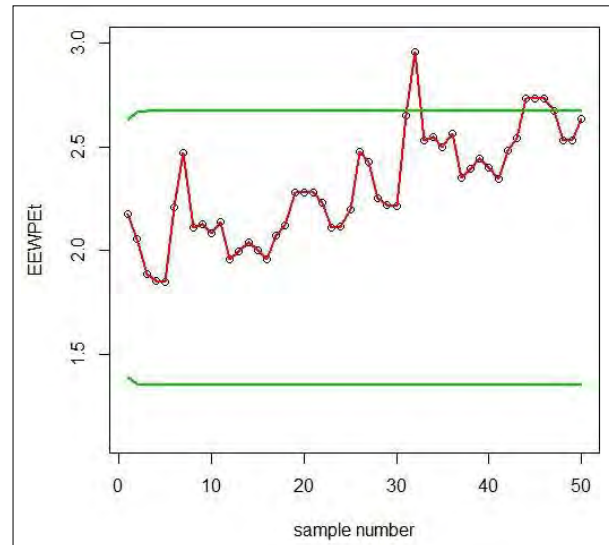

Figure 10: Graph of simulated data of the proposed EEWMA control chart under PE when $\lambda_1 = 0.90$ and $\lambda_2 = 0.75$

Simulation

In order to see the working procedure of the proposed control charts, a simulation study was carried out. For this purpose, we generated 25 observations from an RPFD for the in-control process, and the next 25 observations were generated from the shifted process. The estimated values of the proposed EEWMA statistic (MMLM and PE) were

computed for the selected levels of the proposed control charts parameters with $\lambda_1 = 0.90$ and $\lambda_2 = 0.75$. The data and values of the proposed and existing statistics are listed in Table 5, and plotted values of these statistics are shown in Figures 9 and 10. In Figure 10, we noted that the proposed EEWMA control chart under PE detected a shift at the 32nd sample, while in Figure 9, the EEWMA control chart under MMLM could not detect the shift.
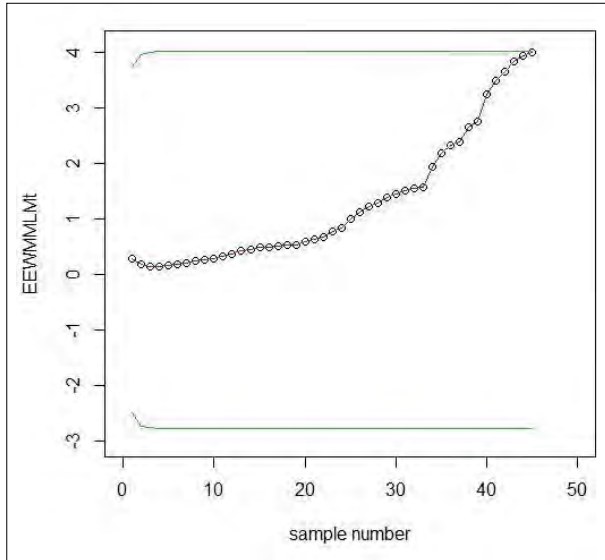
Figure 11: Graph of real data of the EEWMA control chart under MMLM when L=12, λ_1= 0.90 and λ_2= 0.75
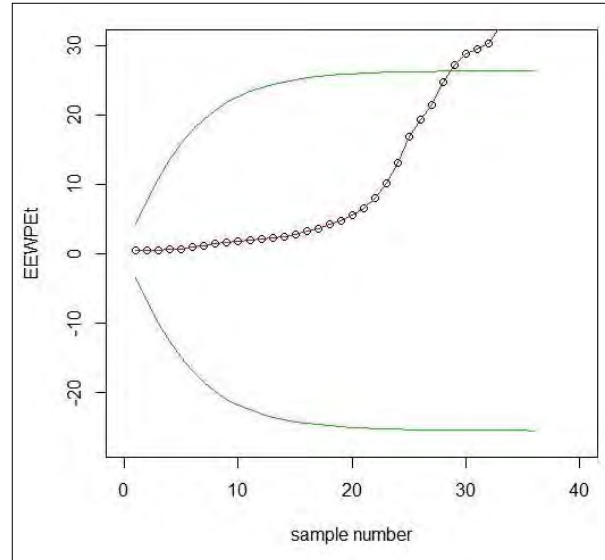
Figure 12: Graph of real data of the EEWMA control chart under PE when L=7.66, λ_1= 0.90 and λ_2= 0.75

Table 6: Earnings per share (EPS) of the National Refinery Ltd.

Year	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
EPS	0.8	1.8	1.8	1.8	1.8	6.96	3.09	4.3	2.9	4.5	4.9	3	2.2
Year	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
EPS	7.4	7.3	10.3	10.8	11.2	12.4	16.9	26.7	30.2	52.8	61.4	73.96	22.37
Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019			
EPS	37.81	88.16	34.68	33.96	12.03	46.38	96.14	100.61	22.14	108.7			

Hence, this shows that the proposed EEWMA control chart under PE has a more remarkable ability to see more minor changes earlier than the EEWMA control chart under MMLM.

Real life application

Real-life data for earnings per share (EPS) of the National Refinery Ltd. were taken from the State Bank of Pakistan (SBP) report for non-financial companies from the year 1984-2019. The data follows the RPFD and are plotted for EEWMA control charts under MMLM and PE, as shown in Figure 11 and Figure 12.

We have constructed EEWMA control charts on real life data under MMLE and PE. In Figure 11 and Figure 12, we see that EEWMA under PE detects the process shift early as compare to EEWMA under MMLE, which shows that EEWMA under PE is better to be used in real life when the distribution of underlying process is RPFD.

CONCLUSION

The current study explained the application of RPFD functions in management sciences and reliability engineering fields. It introduces control charts based on PE and MMLM estimators under the assumption that the proposed distribution follows RPFD. We constructed a memory-based control chart, *i.e.*, HEWMA and EEWMA control charts under MMLM and PE estimators. The findings indicate that the performance of PE remains

consistently good in all control charts, while EEWMA based PE is proved to be the best among all the proposed control charts. Finally, it is expected that these findings will be helpful for the scholars and practitioners in different field of applied sciences. We can use different estimators for reflected power function distribution to see their performance in statistical process control.

Acknowledgement

REFERENCES

Abbas N., Riaz M. & Does R.J.M.M. (2013). Mixed exponentially weighted moving average-Cumulative sum charts for process monitoring. *Quality and Reliability Engineering International* **29**: 345–356.
DOI: https://doi.org/10.1002/qre.1385

Ajadi J.O. & Riaz M. (2017). Mixed multivariate EWMA-CUSUM control charts for an improved process monitoring. *Communications in Statistics-Theory and Methods* **46**(14): 6980–6993.
DOI: https://doi.org/10.1080/03610926.2016.1139132

Erto P., Pallotta G., Palumbo B. & Mastrangelo C.M. (2018). The performance of semi empirical Bayesian control charts for monitoring Weibull data. *Quality Technology and Quantitative Management* **15**(1): 69–86.
DOI: https://doi.org/10.1080/16843703.2017.1304036

Haq A. (2013). A new hybrid exponentially weighted moving average control chart for monitoring process mean. *Quality and Reliability Engineering International* **29**: 1015–1025.
DOI: https://doi.org/10.1002/qre.1453

Haq A. & Munir W. (2018). Improved CUSUM charts for monitoring process mean. *Journal of Statistical Computation and Simulation* **88**(9): 1684–1701.
DOI: https://doi.org/10.1080/00949655.2018.1444040

Hossain M.P., Sanusi R.A., Omar M.H. & Riaz M. (2019). On designing Maxwell CUSUM control chart: An efficient way to monitor failure rates in boring processes. *The International Journal of Advanced Manufacturing Technology* **100**(8): 1923–1930.
DOI: https://doi.org/10.1007/s00170-018-2679-1

Li Z., Xie M. & Zhou M. (2018). Rank-based EWMA procedure for sequentially detecting changes of process location and variability. *Quality Technology and Quantitative Management* **15**(3): 354–373.
DOI: https://doi.org/10.1080/16843703.2016.1208941

Li Z., Zou C., Gong Z. & Wang Z. (2014). The computation of average run length and average time to signal: An overview, *Journal of Statistical Computation and Simulation* **84**(8): 1779–1802.
DOI: https://doi.org/10.1080/00949655.2013.766737

Liang W., Xiang D., Pu X., Li Y. & Jin L. (2019). A robust multivariate sign control chart for detecting shifts in covariance matrix under the elliptical directions distributions. *Quality Technology and Quantitative Management* **16**(1): 113–127.
DOI: https://doi.org/10.1080/16843703.2017.1372852

Lin Y.C., Chou C.Y. & Chen C.H. (2017). Robustness of the EWMA median control chart to non- normality. *International Journal of Industrial and Systems Engineering* **25**(1): 35–58.
DOI: https://doi.org/10.1504/IJISE.2017.080687

Naveed M., Azam M., Khan N. & Aslam M. (2018). Design of a control chart using extended EWMA Statistic. *Technologies* **6**(108): 2–15.
DOI: https://doi.org/10.3390/technologies6040108

Nguyen H.D., Tran K.P. & Heuchenne C. (2019). Monitoring the ratio of two normal variables using variable sampling interval exponentially weighted moving average control charts. *Quality and Reliability Engineering* **35**(1): 439–460.
DOI: https://doi.org/10.1002/qre.2412

Noorossana R., Fathizadan S. & Nayebpour M.R. (2016). EWMA control chart performance with estimated parameters under non-normality. *Quality Technology and Quantitative Management* **32**(5): 1638–1654.
DOI: https://doi.org/10.1002/qre.1896

Page E.S. (1954). Continuous Inspection Scheme. *Biometrika* **41**(1-2): 100–115.
DOI: https://doi.org/10.1093/biomet/41.1-2.100

Qiu P. & Li Z. (2011). Distribution-free monitoring of univariate processes. *Statistics and Probability Letters* **81**(12): 1833–1840.
DOI: https://doi.org/10.1016/j.spl.2011.07.004

Qiu P. & Li Z. (2011). On nonparametric statistical process control of univariate processes. *Technometrics* **53**(4): 390–405.
DOI: https://doi.org/10.1198/TECH.2011.10005

Roberts S.W. (1959). Control chart tests based on geometric moving averages. *Technometrics* **1**: 239–250.
DOI: https://doi.org/10.1080/00401706.1959.10489860

Sanusi R.A., Riaz M. & Abbas N. (2017). Combined Shewhart CUSUM charts using auxiliary variable. *Computers and Industrial Engineering* **105**: 329–337.
DOI: https://doi.org/10.1016/j.cie.2017.01.018

Shamma S.E. & Shamma A.K. (1992). Development and evaluation of control charts using double exponentially weighted moving averages. *International Journal of Quality and Reliability Management* **9**(6): 18–25.
DOI: https://doi.org/10.1108/02656719210018570

Zaka A., Akhter A.S. & Jabeen R. (2020). The new reflected power function distribution: theory, simulation and application. *AIMS Mathematics* **5**(5): 5031–5054.
DOI: https://doi.org/10.3934/math.2020323

Zhang J., Li E. & Li Z. (2017). A Cramér-von Mises test-based distribution free control chart for joint monitoring of location and scale. *Computers and Industrial Engineering* **110**: 484–497.
DOI: https://doi.org/10.1016/j.cie.2017.06.027

RESEARCH ARTICLE

Opthalmic Genetics

Prevalence and *in silico* analysis of p.D658G variant of *WDR36* gene in patients affected with primary open angle glaucoma from Punjab Pakistan

Khazeema Yousaf [1], Zernish Shabbir[1], Rasheeda Bashir[1*], Syed Mohsin Raza[2], Saiqa Ilyas[1] and Rukhama Haq[1]

[1] *Department of Biotechnology, Lahore College for Women University, Lahore, Pakistan.*

[2] *Institute of Biomedical and Health Sciences, University of Health Sciences, Lahore, Pakistan.*

Abstract: The aim of the present study was to check the frequency of genetic variants in exons 8, 11, 13, and 17 of the *WDR36* gene among primary open angle glaucoma (POAG) patients from Punjab, Pakistan, and to perform the *in silico* analysis of identified variants on protein function. Ninety-two individuals affected with primary open angle glaucoma were enrolled for this study. The clinical investigation involved the examination of the optic nerve head, visual field loss and elevated intraocular pressure (IOP). Selected exons (8, 11, 13, and 17) of the *WDR36* gene was screened by Sanger sequencing. Sequencing results revealed a previously reported missense mutation p.D658G in exon 17 in two out of ninety-two POAG patients, while no mutation has been identified in the exons 8, 11, and 13. To predict the structural and functional effect of the p.D658G variant, SIFT, Polyphen-2, PROVEAN, mutation taster, I-mutant 3.0, and MuPRO were used. The MODELLER-CABS based hybrid approach was used for protein structure modelling. *In silico* analysis predicted the p.D658G variant to be deleterious, and it may affect the stability of protein and protein-protein interaction. The findings of this study suggested that the genetic variant p.D658G of the *WDR36* gene is a rare genetic cause of POAG in Pakistani patients. The *in silico* tools predicted the variant p.D658G to be deleterious; however the modelled normal and mutant structure showed no effect on protein structure and function. To further confirm the pathogenic effect of this SNP, *in vivo* experiments, X-ray Crystallography of the WDR36 protein and population-based studies are needed.

INTRODUCTION

Primary open angle glaucoma (POAG) is a heterogeneous group of optic neuropathies that lead to optic nerve damage and permanent vision loss (Mbacham *et al*., 2020). POAG is the most prevalent form of glaucoma particularly in Asia. A study from Pakistan reported that the prevalence of POAG is 40.3% among the 3021 glaucoma patients enrolled while the overall glaucoma prevalence was 8.5% (Akhtar *et al*., 2010). Clinical symptoms of POAG includes ocular hypertension, elevated intraocular pressure (> 22 mmHg) called high tension glaucoma (HTG) (Bui *et al*., 2005) and low-tension glaucoma (LTG) or normal tension glaucoma (NTG) with IOP < 22 mmHg (Miyazawa *et al*., 2007). Individuals having age more than 40 years showed greater susceptibility to glaucoma; also, men are more prone to POAG than females (Tham *et al*., 2014).

The precise molecular basis of POAG is still not well known. It is a genetically heterogeneous disorder caused

* Corresponding author (rashidasbs@yahoo.com; https://orcid.org/0000-0003-1270-3742)

by the involvement of multiple genes and environmental factors. Till now, at least 14 loci from Glaucoma (GLC)1A to GLC1N have been linked to POAG (Monemi *et al.*, 2005) and the following genes, *viz.*, Myocilin (*MYOC*), Optineurin (*OPTN*) (Rezaie *et al.*, 2002), WD repeat domain 36 (*WD40-Repeat36*) (Monemi *et al.*, 2005), neurotrophin *4 (NTF4)* (Rezaie *et al.*, 2002), optic atrophy *1 (OPA1)* (Monemi *et al.*, 2005)*,* cytochrome P450 family 1, subfamily B *(CYP1B1)* (Kaur *et al.*, 2011), and latent transforming beta binding protein 2 *(LTBP2)* are associated with POAG (Su *et al.*, 2017). The gene *WDR36* is known as a gene, out of many other genes, which can cause POAG and is recognized as a cause of POAG pathogenesis (Youngblood *et al.*, 2019). WD repeat domain 36 is a class of a nucleolar protein encoded by WDR36 and is involved in the maturation of 18s rRNA. *WDR36* is a member of the *WD40* repeat protein family, and helps in T- cell activation (Rezaie *et al.*, 2002). The T-cell mediated response participates in optic nerve degeneration (Padgett & Glaser 2003).

POAG caused by mutations at in *WDR36* gene alone contribute 5% of all reported cases (Fuse 2010). The worldwide prevalence of *WDR36* gene mutations in patients affected with POAG is 1.6-1.7% (Mbacham *et al.*, 2020). The WD repeat domain 36 is a class of nucleolar protein encoded by WDR36 and is involved in the maturation of 18s rRNA. The exact role of *WDR36* gene mutations in causing POAG is controversial in previously published data (Weisschuh *et al.*, 2007). In a study from German population, it is reported that *WDR36* gene variants are rare causes of NTG (Weisschuh *et al.*, 2007). It is also suggested in another study by Pasutto *et al.* (2008) that WDR36 plays a minor contributing role in causing POAG in same population, whereas another investigation suggested that WDR36 sequence variants can lead to diseased phenotype in polygenic forms of glaucoma (Footz *et al.*, 2009). Four alterations (N355S, D658G, R529Q, and A449T) were identified between 17 distinct POAG subjects (Monemi *et al.*, 2005). Different mutation patterns in p.D658G of *WDR36* of about 1-2% were reported in a German cohort, 1.7% in St. Petersburg, Russia (Motushchuk *et al.*, 2009), and 1% in Italy (Frezzotti *et al.*, 2011).

The genetics of POAG due to *WDR36* gene has not been explored in Pakistan, so the aim of this study was to check the pathogenic variants in selected exon 8, 11, 13, and 17 in the *WDR36* gene among POAG patients of Punjab, Pakistan and to analyze the damaging effect of identified variants on the function of the WDR36 protein by using various *in silico* tools. The *WDR36* Protein structure was modeled to analyze the deleterious effect of variant p.D658G on protein 3-D structure.

MATERIALS AND METHODS

The study was carried out at the Department of Biotechnology, Lahore College for Women University. Approval was granted from the Institutional Review Board (IRB), Lahore. Ninety-two POAG affected patients were enrolled after diagnosis by ophthalmologists from the Mughal Eye Hospital, Lahore. One hundred (100) controls belonging to the same ethnic background were also enrolled. Written consent was obtained from all individuals. All clinical parameters were recorded, *i.e.*, age of onset of disease, family history, visual acuity, cup to disc ratio (C/D) and increased intraocular pressure. An intravenous blood sample (5 cc) was taken from each individual in EDTA coated tubes (BD, USA) and stored at $4\,^{\circ}$C.

Molecular genetic analysis

DNA was extracted using sucrose lysis and the method of salt precipitation. PCR was performed for the amplification of exons 8, 11, 13, and 17 (details provided in supplementary Table 1). The primers were designed using Primer 3 software (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) (supplementary Table 1). The specific set of primers was prepared with a final volume of 25 µL, with 0.15 - 0.25 units of *Thermus aquaticus* DNA (TaqDNA) polymerase (MBI Fermentas, Vilnius, Lithuania) having a volume of 2.5 µL of Taq buffer, 2 mM $MgCl_2$, 0.2 mM DNTPs, 0.24 µL forward primer and 0.24 µL reverse primer (with 0.24 µM concentration of each primer), and 50 ng genomic DNA. Initial denaturation of DNA was performed at 94 ºC for 4 min followed by 36 cycles of denaturation at 95 ºC for 30 s, and annealing at 55 ºC for exon 11 and 17, and 56 ºC and 57 ºC for exon 8 and 13, respectively, for 30 s. The initial extension was at 75 ºC for 1 min and final extension was performed at 75 ºC for 5 min. PCR products were purified by a gene cleaning column (Gel Extraction Kit Thermo Scientific GeneJet) and were sequenced by direct sequencing with Big Dye Terminator® (V 3.1) Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). The samples were resolved on an ABI 3730 capillary machine.

In silico analysis

Once the sequencing files were acquired, they were exported to SEQMAN, a software tool in the Lasergene suite from DNASTAR. Normal sequences of *WDR36* gene exons 8, 11, 13, and 17 were obtained from the UCSC genome browser website (http://genome.ucsc.edu/) and were used for analyzing the trace files. ClustalOmega was used for multiple alignment of WDR protein sequences

of different species to check the conservation domain on D658 residue (Figure 1E). To check the damaging effect on protein structure Polyphen, Mutation taster and PROVEAN were applied. The PolyPhen server provides a damaging scale ranging from 0.00-1, wherein a score near to 1 has been regarded as potentially damaging. The PROVEAN tool predicts whether an amino acid variation has a damaging impact on the biological functionality of a protein. To check the damaging effects of mutation on the stability of the protein, I-Mutant2.0 and MuPRO tools were used. The I-Mutant server predicts the effect of mutations on the stability of a protein.

Modelling the complete WDR36 protein

Currently, there is no crystal structure available for the WDR36 protein. Thus, an NCBI BLASTp (PSI-BLAST) (Altschul *et al*., 1997) search was performed, using PDB as the search source, to find closely related protein structures. Among the results, the top 3 structures (Table 3) were chosen for homology modelling. For the modelling of the WDR36 protein, the MODELLER-CABS hybrid approach was adopted (Van Koolwijk *et al*., 2007; Jamroz & Kolinski, 2010). The part of the protein which was covered by templates was modelled by homology modelling using MODELLER v9.24 (Eswar *et al*., 2006; Webb & Sali., 2016). MODELLER utilizes python scripts, to perform various steps of homology modelling. The basic steps involved were alignment of the structures of the three templates, their alignment with query sequence, model building based on multiple templates-query sequence alignment, followed by model evaluation by DOPE (discrete optimized protein energy) score. The top 10 resulting models with the lowest DOPE score were selected for the modelling of the N- and C-terminal regions for which templates provided no coverage. The chosen top 10 models were submitted to the online CABS-fold server (Blaszczyk *et al*., 2013) (http://biocomp.chem.uw.edu.pl/CABSfold/index.php) as templates along with the WDR36 protein sequence, to model the N- and C-terminal part of protein. The structures provided by CABS-Fold server were in the form of clusters (models) ranked according to the cluster densities. The cluster with highest density was selected as the predicted model. The loops of this model were further refined again by MODELLER v9.24, and the resulting model with the lowest DOPE score was saved as the WDR36 native model.

Mapping of p.D658G variant on the structure of the protein

The best structure chosen after the loop refinement was used as native protein. The mutated model was generated by using the Swiss-PDBViewer ver. 4.10, which allows browsing through a rotamer library to change amino acids (Guex *et al*., 1997). In order to substitute native amino acid with mutated one, the mutation tool was used. The mutation tool allows the placement of the best rotamer for the new amino acid. The mutated structure was saved in the .pdb format. The energy minimization of both the native and mutant structure was also performed by the GROMOS96 implementation of Swiss-PDBViewer by the conjugate gradient method.

Calculation of RMSD value of modelled protein

For the calculation of root-mean-square deviation of the atomic positions in the mutated modeled protein, USCF Chimera ver 1.14 was used (Pettersen *et al*., 2004). The native and the mutated proteins were superposed, and the matchmaker command was applied to calculate the RMSD value between the native and mutated protein structure. The extent of structural deviation calculated based on RMSD value shows that the higher the RMSD value, the higher will be the deviation of structure and related function of the protein (Coutsias *et al*., 2004).

Validation of native and mutant models

The Ramachandran plot calculates the dihedral angles of the amino acid residues. These phi and psi dihedral angles help in calculating the energetically allowed resides, helping in understanding the structural and functional properties of protein structure. Both the native and mutant protein structures were evaluated by the online tool PROCHECK (https://servicesn.mbi.ucla.edu/PROCHECK/) (Laskowski *et al*., 1993).

RESULTS AND DISCUSSION

The study was designed to check the contribution of the selected exons (8, 11, 13, and 17) of the *WDR36* gene among POAG patients from Punjab. The age range of patients involved in our study was 41-65 years (mean, 52.7 ± 9.8). The average IOP was 23.5 ± 8.10. Table 1 represents the baseline characteristics of the patients. Clinical features of POAG patients showed watery eyes,

enlarged globe and hazy cornea. In patients (POAG II & III) the IOP ranged from 40 to 50 mmHg. The P-value showed a significant difference between control and diseased individuals' IOP and C/D ratio (Table 1). It was also observed in POAG patients that IOP of the right eye was greater than that of the left eye, and the percentage of men affected with POAG was greater than that of women (Figures 1 A and 1 B).

Table 1: Clinical parameters of POAG and control groups

Parameters	Status	POAG	Control	P Value	
Sex	Male	56	50		
	Female	36	50		
Age	Mean ± SD	52.78 ± 9.8	65.9 ± 11.6		
	Range	41 - 65	41 - 93		
Max. IOP	Mean ± SD	23.6 ± 8.1	15.0 ± 2.7	Left eye	Right eye
		8 – 60	8–21	0.0001*	0.01*
CD ratio	Male &	0.4 – 0.9	0.4	Left eye	Right eye
	Female			0.01*	0.01*

* indicates the significant values
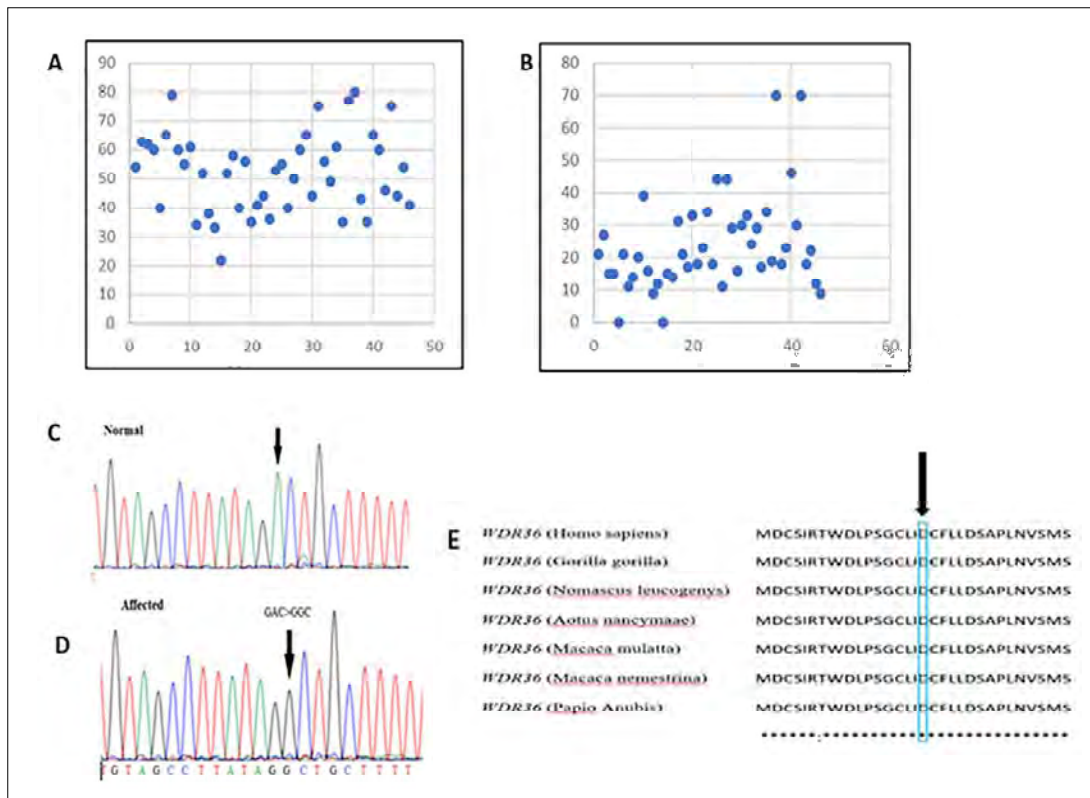
Figure 1: (A) IOP of right eye with respect to age among POAG patients; (B) IOP of left eye with respect to age among POAG patients; (C) Sequencing Chromatogram of Exon 17 of *WDR36* gene in normal and (D); POAG affected patient, the homozygous mutation c.1973A>G (p.D658G) is indicated by an arrow in the trace from normal individual; (E) ClustalOmega output, showing conservation of p.D658 residue among WDR protein from different species.

Sanger sequencing results of exon 8, 11, and 13 showed no mutation while mutation p.D658G has been identified in exon 17. PCR amplification of mutant variant p.D658G using a specific set of primers showed the amplicon size of 443 bp. The sequencing results revealed p.D658G mutation in 2 out of ninety-two POAG patients at exon 17. The mutational analysis of sequencing results was done by Edit Seq and Chromas revealed a previously reported homozygous missense mutation at nucleotide position 1973 where adenine was changed into guanine (c.1973A > G) in the affected individual (Figures 1C and 1 D).

Only 2 out of 92 patients showed mutation p.D658G thus giving the evidence that prevalence of this variant is low in Pakistani patients.

In silico analysis

The mutational effect of p.D658G using the Polyphen server was predicted to be 1 which showed the damaging effect. PROVEAN algorithm results showed that variant p.D658G is deleterious (Table 2). To check the damaging effects of the mutation on the stability of the protein, I-Mutant2.0 and MuPRO tools were used (Table 2). The I-Mutant server predicts the effect of mutations on the stability of protein. A score below 0 shows decreased stability of protein and a score higher than 0 indicates increased stability of protein. The results showed the DDG score of -3.57, indicating that the p.D658G substitution decreases the stability of the protein. Mu-PRo tool showed that the substitution reduces the stability of protein structure with a DDG score of -1.448.

Table 2: Summary of *in silico* tools used to predict the pathogenicity of p.D658G variant

Prediction tools	Prediction result	Severity score
Mutation taster	Disease causing	Deleterious
PolyPhen-2	Probably damaging	1
PROVEAN	Deleterious,	-6.409
	Decrease stability	-1.4485842
I-Mutant	Decrease stability	-3.57
SIFT	Intolerable	0.42

Modelling of the *WDR36* protein & model validations

Due to the absence of the PDB structure of the WDR36 protein, structure modelling was performed. The PSI-BLAST revealed that there were very few structures which shared similarity with our protein sequence,

the highest being 33.99% similar (Table 3). The query coverage was also less between 67-89%. Keeping in view the low similarity templates and less coverage, a hybrid approach of modelling was adopted. MODELLER v9.24 was commanded to generate 30 more similar models using the 3 aligned templates 5JPQ, 4NSX, and 5TZS, and "WDR36-mult.ali" file. The tope 10 models were selected according to the widely practiced criteria for selecting top models, *i.e.*, DOPE score must be lowest and GA341 score highest. These models were presented to the CABS-Fold server to model the non-template region of the protein. The server provides various predicted structures, each representing a cluster of structures, while the structure with highest cluster density is chosen as the best structure. The CABS-Fold server provided 8 clusters and model 1 had the highest cluster density of 213, with an average cluster RMSD of 1.8. The loops for the model 1 were refined by using MODELLER to get the final structure WDR36.BL00030001.pdb, which was later renamed as WDR36 Native Protein (Figure 2A). The modelled structure was validated by PROCHECK. The analysis of WDR36 Native Protein by PROCHECK showed that 77.2% of the residues lie in the most favored regions, 18.7% were in additional allowed regions, 2.2% in possible generously allowed regions and 2.0% were in the disallowed regions (Figure 2B).

Table 3: Results of closely related protein search (NCBI PSI-BLAST)

Sr. No	PDB ID	Query cover	Per. identity	Chain
1	5JPQ	89%	31.88%	I
2	4NSX	68%	33.83%	A
3	5TZS	67%	33.99%	T

Mapping of p.D658G variant on the protein structure

The p.D658G variant was mapped onto the WDR36_ Native Protein by using the mutation tool in Swiss-PDBViewer. The resulting Structure was denoted as WDR36_Mutant Protein (Figure 3B). To mimic the *in vivo* folding settings and parameters, energy minimization of both the native and mutant structures was performed by Swiss-PDBViewer using conjugate gradients. The total post-energy minimization and energy of the WDR36_Native Protein was -42975.512 kJ/mol, while the WDR36_Mutant Protein showed a total energy of -42878.793 kJ/mol, *i.e.*, a 96.719 kJ/mol increase in the energy. This slight increase in the protein energy may slightly affect the protein stability and function.
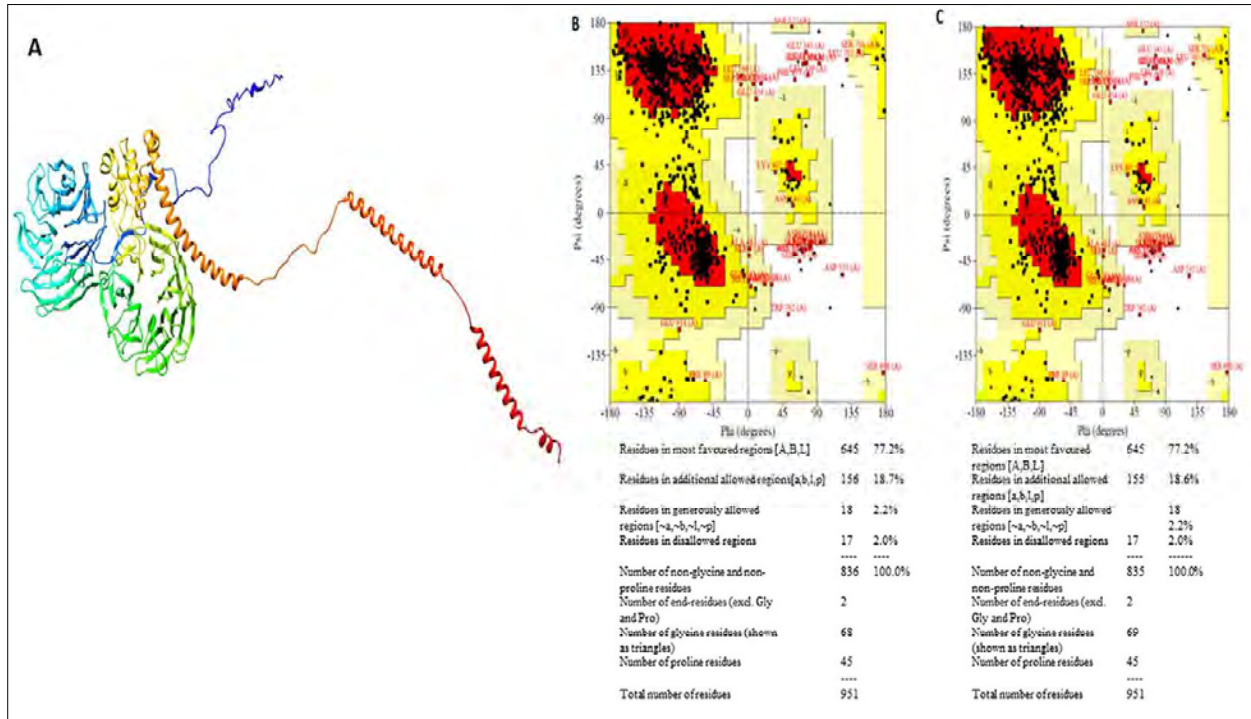
Figure 2: A) Representative model of WDR36 complete normal protein modelled by MODELLER-CABS hybrid approach B) Ramachandran plot of normal protein. C) Ramachandran plot of mutated protein.
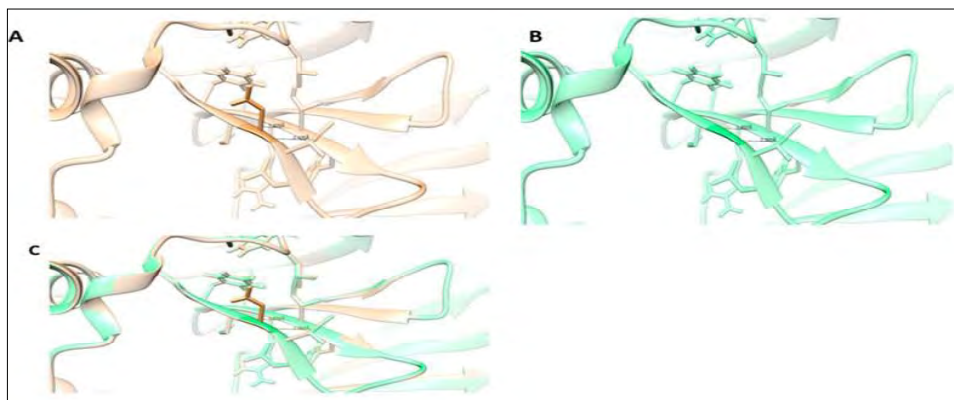
Figure 3: A) Representative model of WDR36 normal protein modelled by MODELLER-CABS hybrid approach; B) representative model of mutated protein generated by Swiss-PDBViewer.; C) superimposed structures of WDR36 normal protein and WDR36 mutant protein (RMSD = 0.004 Å). The images were modeled using USCF Chimera ver 1.14.

Calculating the RMSD value of the mutated protein

One of the important parameters for measuring the rate of deviation of normal and mutated protein structures is the method of root-mean-square deviation (RMSD). The mutant structure showed an RMSD value of 0.004 Å which shows that structure is not much deviated from the native structure. The hydrogen bond analysis shows that the number of hydrogen bonds are the same in both proteins on this position, with only some minor differences in the length of hydrogen bonds (Figures 3B and 3C).

Despite studies that have examined the involvement of the p.D658G variant of the *WDR36* gene in POAG patients in different population groups, no study has yet been reported from Pakistan to show the contributing role of the p.D658G mutation in Punjab POAG patients. Therefore, the present study was conducted to check the contribution of p.D658G of the *WDR36* gene in Punjab patients affected with POAG. The result of the study showed low prevalence of this mutant variant (c.1973A>G, p.D658G) in POAG patients in Punjab.

An analysis of published data revealed that this variant is associated with HTG cases (Miyazawa *et al*., 2007). In the current study, 2 patients carried the p.D658G variant. The POAG II patient was a 53 year old male, the IOP of his left eye was 45 mmHg and that of his right eye was 50 mmHg, with a 0.7/1 CD ratio. He had blurred vision with photophobia. The patient III was 55 years old; he had symptoms of blurred vision and photophobia. Clinical examination showed IOPs of 40 mmHg and 35 mmHg in the right and left eye, respectively. The CD ratio was 0.8/0.7 in the right and left eye, respectively.

From previous reports the p.D658G mutation has been attributed to about 1.94% of unrelated POAG patients in the US (Monemi *et al*., 2005). This mutation was previously found to be the more recurrent disease-causing allele in familial and non-familial cases (Monemi *et al*., 2005). The *WDR36* gene consists of four domains, which help in the T cell activation pathway, and any change in this gene is hypothesized to participate in optic nerve degeneration and a different form of glaucoma (Monemi *et al*., 2005). The p.D658G missense mutation is an exonic alteration and an important part of the *WDR36* gene, which results in the replacement of aspartic acid with lysine (acidic to basic amino acid) (Frezzotti *et al*., 2011).

The mutation p.D658G was also reported in the US as 1% (Monemi *et al*., 2005). However, further data on the involvement of this mutation in causing POAG was controversial. In the present study, the p.D658G mutation has been identified in two individuals out of 97. The mutation frequency of p.D658G in PAOG varied in different ethnic groups, with no mutation in Chinese and Japanese (Miyazawa *et al*., 2007; Huang *et al*., 2014), Australian and Russian populations (Motushchuk *et al*.,2009), but ranging from 1% to 3% in Spanish (Kaur *et al*., 2011), German, American (Monemi *et al*., 2005), and Italian (Frezzotti *et al*., 2011) populations. This data showed that, although it is a pathogenic mutation,

its prevalence is low in different population, just as reported in the Pakistani population. These variations in mutational frequency may be due to different geographical regions, and environmental factors such as age, sex, race, and family history also contribute to causing POAG. Investigations of POAG patients from Australia and Germany (Monemi *et al*., 2005) suggest that variant p.D658G is a neutral variant, which signifies that variants of the *WDR36* gene may act as a causative factor in certain populations, or may act as a modifier gene for POAG in some other populations.

Residue D658 is a beta-sheet of the protein. D658 residue is located in separate G-beta WD40 repeats. The G proteins belong to a family of membrane-associated proteins and act as mediators in signal transduction induced by transmembrane receptors. The G-beta subunit is essential for anchoring to the membrane and recognition by the receptor. Structurally, the G-beta subunit comprises eight tandem repeats having 40 residues. Each residue contains a central Trp-Asp dipeptide bonding, thus named as WD40 repeats. Hence, mutations disturbing the structure of WD40 repeats may cause hindrance in the WDR36 interaction with other proteins (Tham *et al*., 2014). The extent of structural deviation due to the p.D658G variant between the modelled normal and mutated structures of protein was found to have a minimal effect on the protein structure and stability. However, the mutations in the structure of WD40 repeats may hinder the interacting ability of WDR36 with other proteins or may disrupt protein-protein interactions. *In vitro* studies and X-ray crystallography of WDR36 protein are needed to validate the functional effect of the p.D658G variant.

CONCLUSION

It is suggested that the p.D658G mutation is a rare cause of POAG in Punjab region of Pakistan. Overall, the prevalence of mutation p.D658G in the *WDR36* gene is low in our enrolled patients and no variants have been identified in other selected exons. However, mutational frequency may change when a large number of individuals from different regions of Pakistan are screened for this variant. Although the effect of p.D658G on the WDR36 protein structure was explained to be deleterious using *in silico* tools, the exact mechanism and the pathogenicity of this SNP should further be validated by X-Ray crystallography of the WDR36 protein, *in vitro* and *in vivo* experiments and by large population-based studies.

Acknowledgments

We are grateful to all the patients for their participation in this research study. We also appreciate the doctors and staff at the Mughal Eye Hospital, Lahore for their help in recruitment and clinical assessment of the patients.

REFERENCES

Akhtar S., Micheal S., Khan M.I., Yousaf S., Bilal M., Ahmed A. & Qamar R. (2010). Does gender have an effect in the prevalence of types of glaucoma in Pakistani population? *Al-Shifa Journal of Ophthalmology* **6**(1):30–36.

Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W. & Lipman D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**(17):3389–3402.
DOI: https://doi.org/10.1093/nar/25.17.3389

Bui C.M., Chen H., Shyr Y. & Joos K.M. (2005). Discontinuing nasal steroids might lower intraocular pressure in glaucoma. *Journal of Allergy and Clinical Immunology* **116**(5):1042–1047.
DOI: https://doi.org/10.1016/j.jaci.2005.07.031

Blaszczyk M., Jamroz M., Kmiecik S., & Kolinski A. (2013). CABS-fold: server for the de novo and consensus-based prediction of protein structure. *Nucleic Acids Research* **41**(W1): 406–411.
DOI: https://doi.org/10.1093/nar/gkt462

Coutsias E.A., Seok C. & Dill K.A. (2004). Using quaternions to calculate RMSD. *Journal of Computational Chemistry* **25**(15):1849–1857.
DOI: https://doi.org/10.1002/jcc.20110

Eswar N., Webb B., Marti-Renom M.A., Madhusudhan M.S., Eramian D., Shen M.Y., Pieper U. & Sali A. (2006). Comparative protein structure modeling using Modeller. *Current Protocols in Bioinformatics* **15**(1): 5–6.
DOI: https://doi.org/10.1002/0471250953.bi0506s15

Frezzotti *et al.* (14 authors) (2011). Association between primary open-angle glaucoma (POAG) and WDR36 sequence variance in Italian families affected by POAG. *British Journal of Ophthalmology* **95**(5): 624–626.
DOI: https://doi.org/10.1136/bjo.2009.167494

Fuse N. (2010). Genetic bases for glaucoma. *Tohoku Journal of Experimental Medicine* **221**: 1–10.
DOI: https://doi.org/10.1620/tjem.221.1

Footz T.K., Johnson J.L., Dubois S, Boivin N., Raymond V. & Walter M.A. (2009). Glaucoma-associated WDR36 variants encode functional defects in a yeast model system. *Human Molecular Genetics* **18**(7): 1276–1287.
DOI: https://doi.org/10.1093/hmg/ddp027

Gemenetzi M., Yang Y. & Lotery A.J. (2012). Current concepts on primary open-angle glaucoma genetics: A contribution to disease pathophysiology and future treatment. *Eye* **26**(3):355–369.
DOI: https://doi.org/10.1038/eye.2011.309

Guex N. & Peitsch M.C. (1997). SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *Electrophoresis* **18**(15): 2714–2723.
DOI: https://doi.org/10.1002/elps.1150181505

Huang X., Li M., Guo X., Li S., Xiao X., Jia X., Liu X. & Zhang Q. (2014). Mutation analysis of seven known glaucoma-associated genes in Chinese patients with glaucoma. *Investigative Ophthalmology and Visual Science* **55**(6): 3594–3602.
DOI: https://doi.org/10.1167/iovs.14-13927

Jamroz M. & Kolinski A. (2010). Modeling of loops in proteins: a multi-method approach. *BMC Journal of Structural Biology* **10**(1):1–9.
DOI: https://doi.org/10.1186/1472-6807-10-5

Kaur K., Mandal A.K. & Chakrabarti S. (2011). Primary congenital glaucoma and the involvement of CYP1B1. *Middle East African Journal of Ophthalmology* **18**(1): 7–16.
DOI: https://doi.org/10.4103/0974-9233.75878

Laskowski R.A., MacArthur M.W., Moss D.S. & Thornton J.M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**(2): 283–291.
DOI: https://doi.org/10.1107/S0021889892009944

Mbacham F., Bilong Y., Chedjou, J.P., Nomo A. & Bella A.L. (2020). Variants of WDR36 in Cameroonian glaucoma patients. *Biomedical Science and Engineering* **6**(1): 043–047.
DOI: https://doi.org/10.17352/abse.000021

Motushchuk A.E., Komarova T.Y., Grudinina N.A., Rakhmanov V.V., Mandelshtam M.Y., Astakhov Y.S. & Vasilyev V.B. (2009). Genetic variants of CYP1B1 and WDR36 in the patients with primary congenital glaucoma and primary open angle glaucoma from Saint-Petersburg. *Russian Journal of Genetics* **45**(12):1467–1474.
DOI: https://doi.org/10.1134/S1022795409120102

Miyazawa A., Fuse N., Mengkegale M., Ryu M., Seimiya M., Wada Y. & Nishida K. (2007). Association between primary open-angle glaucoma and WDR36 DNA sequence variants. *Molecular Vision* **13**: 1912–1919.

Monemi *et al.* (11 authors) (2005). Identification of a novel adult-onset primary open-angle glaucoma (POAG) gene on 5q22. 1. *Human Molecular Genetics* **14**(6): 725–733
DOI: https://doi.org/10.1093/hmg/ddi068

Padgett D.A. & Glaser R. (2003). How stress influences the immune response. *Trends in Immunology* **24**(8): 444–448.
DOI: https://doi.org/10.1016/S1471-4906(03)00173-X

Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C. & Ferri T.E. (2004). UCSF Chimera: a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**(13): 1605–1612.
DOI: https://doi.org/10.1002/jcc.20084

Pasutto F., Mardin C.Y., Michels-Rautenstrauss K., Weber B.H.F., Sticht H., Chavarria-Soley G., Rautenstrauss B., Kruse F. & Resis A. (2008). Profiling of WDR36 missense variants in German patients with glaucoma. *Investigative Ophthalmology and Visual Science* **49**:270–274
DOI: https://doi.org/10.1167/iovs.07-0500

Rezaie T., *et al.* (12 authors) (2002). Adult-onset primary open-

angle glaucoma caused by mutations in optineurin. *Science* **295**(5557): 1077–1079.
DOI: https://doi.org/10.1126/science.1066901

Su H.A., Li S.Y., Yang J.J. & Yen Y.C. (2017). An application of NGS for WDR36 gene in Taiwanese patients with juvenile-onset open-angle glaucoma. *International Journal of Medical Sciences* **14**(12): 1251–1256.
DOI: https://doi.org/10.7150/ijms.20729

Tham Y.C., Li X., Wong T.Y., Quigley H.A., Aung T. & Cheng C.Y. (2014). Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology* **121**(11): 2081–2090
DOI: https://doi.org/10.1016/j.ophtha.2014.05.013

Van Koolwijk L.M, Despriet D.D., van Duijn C.M., Cortes L.M., Vingerling J.R., Aulchenko Y.S., Oostra B.A., Klaver C.C.W. & Lemij H.G. (2007). Genetic contributions to glaucoma: heritability of intraocular pressure, retinal nerve fiber layer thickness, and optic disc morphology. *Investigative Ophthalmology and Visual Science* **48**(8): 3669–3676.
DOI: https://doi.org/10.1167/iovs.06-1519

Webb B. & Sali A. (2016). Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics* **54**(1): 5–6.
DOI: https://doi.org/10.1002/cpbi.3

Weisschuh N., Wolf C., Wissinger B. & Gramer E. (2007). Variations in the WDR36 gene in German patients with normal tension glaucoma. *Molecular Vision* **13**: 724–729.

Youngblood H., Hauser M.A. & Liu Y. (2019). Update on the genetics of primary open-angle glaucoma. *Experimental Eye Research* **188**: 107795.
DOI: https://doi.org/10.1016/j.exer.2019.107795

Supplementary Table 1: Primer sequences for Exon 8, 11, 13 and 17 of WDR36 gene

Gene	Exon	TM(C')	Forward Primer	Reverse Primer	Product Size
	8	56°C	ACTGGAGTAAAAGGCAAAGAAA	TCATCTTCTAGGTTGAAAGCTGAT	454bp
	11	55°C	CAGTGGTAATAACATCTTTGTTTTG	TGCTGGTCTTGTGTGAAG ACT	400bp
	13	57°C	GGAATGTTTTAACTTTGGCTTTG	TCTTCTTACACCCCTAAAATC CA	807bp
WDR36	17	55°C	GTGGTGACTTTCTGATCAATC	AGCTGTCTACATTATCAAGCAG	443bp

RESEARCH ARTICLE

Geotechnical Engineering

The microstructure and the behaviour of low organic clayey soils in Sri Lanka

Y Wang[1], D Vidurapriya[2*], X Qin[1] and S Thilakasiri[2]
[1] *China MCC20 Group Corp. Ltd, Shanghai, China.*
[2] *Sri Lanka Institute of Information Technology, Malabe, Sri Lanka.*

Abstract: Numerous studies focus on investigating organic soils with high organic content. However, only a limited amount of literature discusses the properties and behaviour of low organic content soils. Sri Lankan organic soil is an excellent example of a low organic content soil, demonstrating different behaviour from highly organic fibrous soil. Scanning electron microscope (SEM) images of Sri Lankan organic soil were studied to determine the reasons for these observed variations. Sri Lankan organic soils were observed to consist of highly decomposed organic matter in an amorphous state with higher proportions of mineral matter. The microstructure characteristics observed through SEM images mostly explained the variation of properties and behaviour of the Sri Lankan organic soil in contrast to highly organic fibrous soil. Correlations were also derived between properties of Sri Lankan organic soil and compared with similar relationships derived from other geological conditions. While some correlations derived in this study closely follow the relationships for other organic soils with similar microstructure characteristics, the acidity of the deposit resulted in different behaviour of a few correlations. This study improves the current knowledge of the effect of microstructure characteristics on the properties and behaviour of Sri Lankan low-organic clayey soils, which otherwise primarily depend on theories for highly organic peaty soils. Furthermore, the correlations derived and validated for low organic soils can be utilized in any organic clay deposit with similar microstructure characteristics.

INTRODUCTION

Organic soil is formed by the accumulation of fully or partially decomposed organic matter under wet and anaerobic conditions. In engineering practice, these soils are often considered problematic soils due to high compressibility, low shear strength, localised sinking, slip failure, variability in material properties, difficulty in sampling, shrinkage, and change in material (chemical and biological) (Edil & Den Haan, 1994; Den Haan & Kruse, 2006; Haut, 2006). The physical and mechanical properties of organic soils are governed mainly by their microstructure (O'Kelly & Pichan, 2013). The degree of decomposition and the nature and type of plant components affect the microstructure of the organic soils. Because these factors vary greatly from one geological condition to another, the properties and behaviour of organic soil show significant variations across different deposits (Hobbs, 1986). According to Vasander (2014), there is a considerable variation between organic soil/ peat deposits in temperate regions and tropical regions. Vasander (2014) further mentioned that differences

* Corresponding authors (dinitha.v@sliit.lk; https://orcid.org/0000-0003-4457-0251)
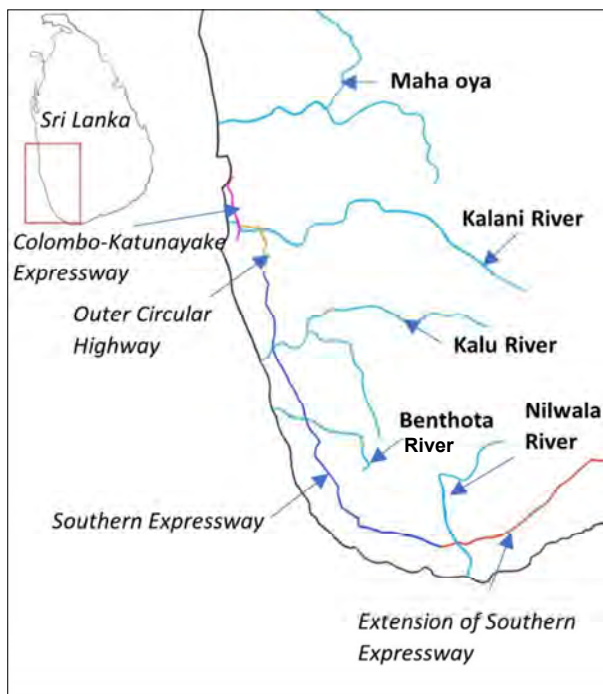
could be noted in high-altitude deposits and low-altitude deposits, even within tropical regions. Therefore, the study of the geotechnical behaviour of Sri Lankan organic soils formed under tropical climatic conditions is very important.

Sri Lankan organic soil

Sri Lankan organic soil deposits can mainly be found in the Western and Southern coastal regions, where low-lying grounds with poor water drainage are a common topographical feature (Karunawardene, 2007; Ariyarathna *et al*., 2010). As shown in Figure 1, major rivers supply these low-lying grounds with flood water intermittently throughout the year, and these conditions provide an ideal environment for the accumulation and formation of organic deposits. In the analysis of organic soils in Sri Lanka, it was found that these deposits are mainly characterised by low organic content (less than 75%). In most classification systems including the ASTM system, organic soils with more than 75% organic content are referred to as peat and the soils with organic content

of less than 75% are just called organic soils (Leong & Chin, 2000). Since the ASTM classification system is followed in the current study, the term organic soil will be used for Sri Lankan deposits in the text instead of peat. However, many studies do not discuss the organic content limit for defining inorganic soil. In the present study, soils with an organic content of less than 10% will be considered inorganic soils following the classification system defined by Kearns & Davidson (1983).

As mentioned above, the microstructure of organic soil mainly affects its physical and mechanical properties. There are only a limited number of studies conducted on the microstructure of Sri Lankan organic soil. Zimar *et al.* (2020) recently studied the microstructure of the organic soil samples recovered from the surface of the deposits. These surficial deposits consist of partially decomposed fibrous material and are mostly classified as Fibric to Hemic and contain a higher percentage of macropores (Zimar *et al.*, 2020). However, these organic soil deposits are about 15 m deep. To properly understand the behaviour of the entire organic soil deposit, the properties and behaviour of these deep layers should be studied. Therefore, in this research, the microstructure of the Sri Lankan organic soil recovered from more than 5 m below the surface was studied using Scanning Electron Microscope (SEM) images. These microstructure characteristics were then used to explain the variations observed in Sri Lankan organic soil in comparison to organic soils formed under different geological conditions.

MATERIALS AND METHODS

Data description

For the present research, three primary data sources were utilised. An extensive data set was obtained from the pre-construction and post-construction ground investigation data of the Colombo-Katunayake Expressway (CKE) project (obtained from the Road Development Authority of Sri Lanka). The other two sources were data from the literature and the results of laboratory tests conducted on undisturbed soil samples extracted in the present study. Data from the above sources were combined to create a database for the present study, including properties such as natural water content, initial void ratio, consistency limits, specific gravity, soil densities, compression indices, and settlement monitoring data.

Figure 1: Rivers and expressway network in the Western and Southern coastal areas of Sri Lanka.

Laboratory testing program

Several sections of the expressway network that connects the Western, Southern and Central provinces of Sri Lanka were built over the previously discussed organic soil deposits (Figure 1). For the present investigation, undisturbed organic soil samples from various depths up to 12 m were recovered in 8 locations along this expressway network. According to relevant ASTM standards, one-dimensional consolidation tests, moisture content tests, organic content tests, liquid limit tests, and specific gravity tests were conducted on these undisturbed soil samples.

Organic soil specimens with different organic contents obtained from the undisturbed soil samples were then used to determine the microstructure characteristics. SEM images of 1,000x, 2,000x, 5,000x and 10,000x magnifications and Energy-dispersive X-ray spectroscopy (EDAX) data were obtained for the samples mentioned above.

Comparison of Sri Lankan organic soil with other soil types

The index soil parameters of Sri Lankan organic soil obtained from the above data set were compared with similar properties of highly organic peaty and inorganic soils. The field settlement monitoring data of the CKE embankment was then used to determine the actual consolidation behaviour of Sri Lankan organic soil deposits. Finally, the compressibility properties obtained from the laboratory results were compared to highly organic peaty and inorganic soils. Numerous variations observed in the comparison of properties and behaviour of Sri Lankan organic soils were then explained using the microstructure characteristics determined by the SEM images.

As mentioned above, the laboratory test results obtained from this study were combined with similar properties available from the literature to form a wider database than previously existed. Empirical correlations between different properties in this database were derived, and a cross-validation analysis was carried out using Python programming language to validate these relationships statistically. These correlations were then compared with similar correlations from other geological conditions.

RESULTS AND DISCUSSION

Table 1 contains the properties of Sri Lankan organic soil compared with the properties of peat from the USA, Canada, and Malaysia (Adams, 1965; Mesri *et al*., 1997; Mesri & Ajlouni, 2007; Adnan & Wijeyesekara, 2008; Kazemian *et al*., 2011; Lee *et al*., 2015; Raghunandan & Anirudh, 2017), and mineral soil from the USA, Bangladesh, and India (Force, 1998; Santagata, 1999; Sridharan & Nagaraj, 2004; Vikas *et al*., 2015; Wasif *et al*., 2016). Soil properties given in Karunawardena (2007) are combined with the current data set to represent the properties of Sri Lankan organic soil. Here, the variations observed in the index soil parameters of Sri Lankan organic soil are noticeable.

Apart from the index soil parameters, noticeable variations in the consolidation behaviour of Sri Lankan organic soil were also observed. According to the literature, the primary consolidation of highly organic soils is completed almost immediately after loading (Mesri & Ajlouni, 2007). Hence, in peaty soils, the time taken for primary consolidation settlement to occur is considered insignificant. Figure 2 shows a typical embankment height *vs* settlement graph for an embankment on a Sri Lankan organic soil deposit, and upon the analysis of similar field settlement monitoring data of several embankments on Sri Lankan organic soil

Table 1: Comparison of index soil properties of Sri Lankan organic soils with similar properties of peat and inorganic soils from other locations

Property	Peat	Sri Lankan Organic Soil	Inorganic Clay
Natural water content (%)	623–1,340	200–800	25.2–43.4
Initial void ratio	10.1–23.5	2.0–8.0	0.64–1.24
Organic content (%)	90–96	20–50	4.4
Specific gravity	1.53–1.65	1.5–2.2	2.77–2.81
Liquid limit		100–400 (Present study)	30–47
Plastic limit		50–250 (Present study)	11.7–25.0

deposits, it was discovered that completion of primary consolidation takes about 6 to 8 months on average. This implies that the time taken for the completion of primary consolidation is also important in these deposits, unlike other highly organic soils. Additionally, a considerable amount of secondary consolidation settlement seems to continue after the primary consolidation settlement is over.
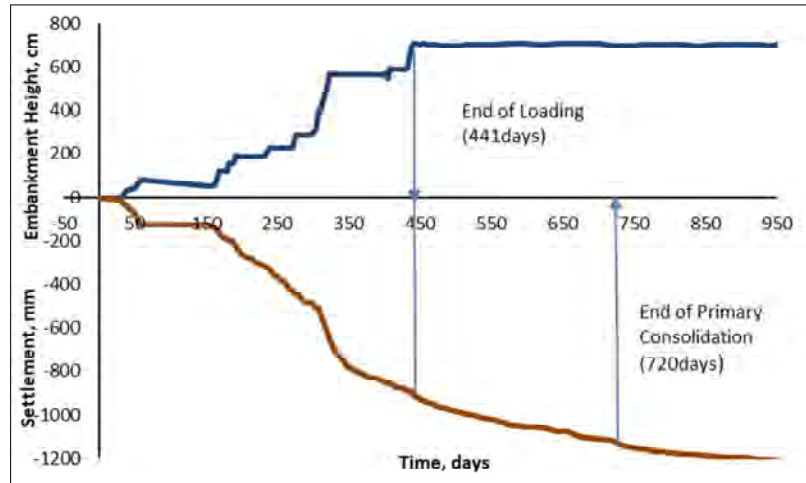

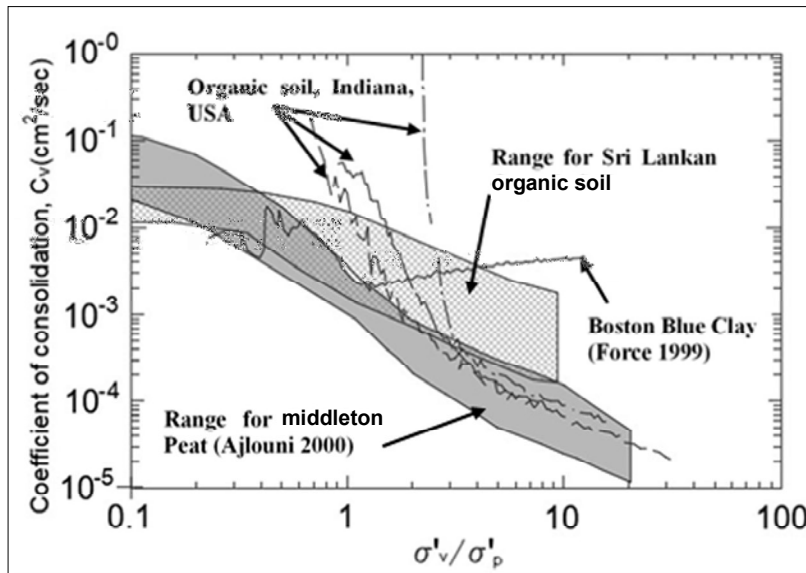**Figure 2:** End of loading vs end of primary consolidation of Sri Lankan organic soil deposits
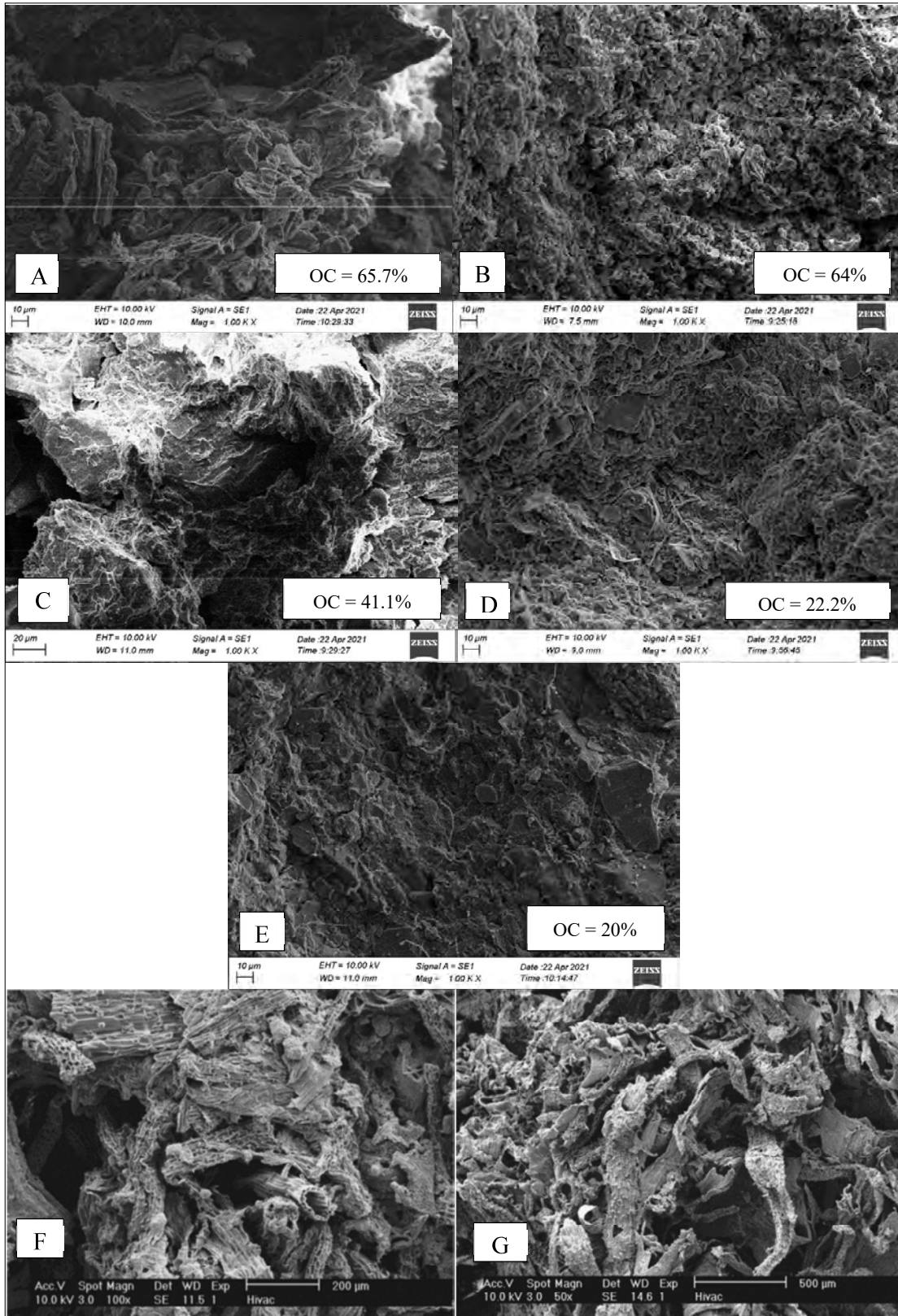

Figure 3: Variation of C_v with normalized effective vertical stress [Source: Santagata *et al*. (2008)]

Figure 4: A to E: SEM images of Sri Lankan organic soil specimens; F and G: SEM images of James Bay Peat (Mesri & Ajlouni, 2007)

The variation of coefficient of consolidation (C_v) with normalised effective overburden pressure (σ'_v/σ'_p) for highly organic Middleton peat and inorganic Boston Blue clay in Figure 3 was developed by Santagata *et al*. (2008). A similar variation between C_v vs σ'_v/σ'_p for Sri Lankan organic soil obtained in the present study is also included in Figure 3 for comparison. The effective overburden pressure (σ'_v) was normalised with respect to the preconsolidation pressure (σ'_p) for reasonable comparison among different soil types and samples. It was observed that the C_v of Sri Lankan organic soil remains unchanged in the overconsolidated region ($\sigma'_v/\sigma'_p < 1$) and starts to slowly decrease around $\sigma'_v/\sigma'_p = 1$. In comparison, the C_v of Middleton peat decreases considerably in all aforementioned ranges of σ'_v/σ'_p.

To clarify the above-observed variations in the properties and behaviour of Sri Lankan organic soil, its microstructure characteristics were studied using SEM images.

Microstructure characteristics of Sri Lankan organic soil

Figures 4A to 4E carry SEM images of organic soil samples of Sri Lankan deposits extracted in the present study. When compared to the microstructure of uncompressed surficial soil, the compressed soil from deep layers of the deposit contains a lesser amount of macropores. Figures 4A and 4B illustrate microstructures of non-surficial specimens with comparatively high organic contents. Though some partially decomposed fibrous material can be observed in these two specimens, the structure is mainly comprised of highly decomposed amorphous materials, and a visible cellular structure of organic matter is absent. Despite the lack of macropores, a higher number of micropores can be identified. However, this observed amount of macro and micropores is remarkably low when compared with the microstructure of highly organic fibrous peat such as James Bay peat (Mesri & Ajlouni, 2007) with highly perforated hollow fibrous material (Figures 4F and 4G). In the medium to low organic content specimens of Sri Lankan organic soil (Figures 4C to 4E), the microstructure comprises a small proportion of highly decomposed organic material and mineral soil grains, as clearly visible in the SEM images. These specimens include only a limited number of micropores except in Figures 4C, where voids are observed between the large mineral grains. The following sections discuss the effect of these microstructure characteristics on the index soil parameters, consolidation properties and behaviour.

Effect of microstructure characteristics on index soil parameters

The SEM images demonstrate that most Sri Lankan organic soil has a lesser amount of both macro and micropores when compared to James Bay peat (Figures 4F and 4G) (Mesri & Ajlouni, 2007). It is also noted that, unlike the James Bay peat, no proper cellular structure can be identified in the microstructure of Sri Lankan organic soil. This lack of pore spaces and cellular structure explains the lower levels of natural water content (w_0) and initial void ratio (e_0) of Sri Lankan organic soil when compared to highly organic fibrous peat (Table 1). The specimens with lower organic contents (OC) (Figures 4 C-F) have visible mineral particles and a faint amount of decomposed organic matter. Such specimens are associated with the high specific gravity (G_s) values recorded in Table 1. In contrast, occasional specimens with high organic content and partially decomposed organic matter contribute to the lower specific gravity values in the Sri Lankan organic soil. On the other hand, the microstructure of highly organic peat (Figures 4F and 4G) consists of slightly decomposed organic matter. The specific gravity of this organic matter is usually in the range of 1.4 to 1.5 (O'Kelly & Pichan, 2013). With the lack of mineral particles in their microstructure (Mesri & Ajlouni, 2007), these highly organic peats usually have very low specific gravity values, as mentioned in Table 1.

Generally, the liquid limit (LL) and plastic limit (PL) values are not considered important in highly organic fibrous peat above H_3 in the von Post humification scale, as they cannot be accurately determined due to the lack of required minimum mineral soil proportion and difficulty of testing (Hobbs, 1986). However, when comparing the Sri Lankan organic soils with such fibrous soils, the mineral portion is significantly high. The EDAX analysis showed that the Sri Lankan organic clayey samples contained Aluminium (12% - 18% by weight), Silicon (16% - 22% by weight), and Iron (5% - 7% by weight), among other elements such as magnesium and potassium in lower percentages. These elements account for more than 50% of the samples by weight. Hence, when discussing the index soil properties of Sri Lankan organic soil, consistency limits cannot be disregarded.

Effect of microstructure characteristics on compressibility properties and behaviour of Sri Lankan organic soil

In the previous section, the effect of microstructure characteristics on the index soil properties was discussed.

It is also important to understand the effect of these microstructure characteristics on the compressibility properties and behaviour of the organic soil. Figure 3 contains the variation of the coefficient of consolidation (C_v) with the normalised effective vertical stress for different soil types. Santagata *et al.* (2008) explained the significant decrease observed in C_v for peat and organic soils in the normally consolidated (NC) region as a result of the subsequent reduction in hydraulic conductivity (k_v) with the increasing stress level offsetting the increase in constrained modulus ($D = 1/m_v$) and vice versa for Boston Blue Clay in which the C_v gradually increases in the NC region. The C_v variation of the Sri Lankan organic soil falls in between the above-mentioned highly organic soil and inorganic clay. This demonstrates that, compared to the highly organic soils, the reduction in k_v does not significantly offset the increase in constrained modules in Sri Lankan organic soil. The presence of high mineral content can be identified as the main reason for the above observations. Furthermore, in the overconsolidated region, the proportion of macropores in fibrous peat is greater than in amorphous organic soil, as illustrated in the above SEM images (Figure 4). In this region, the higher proportion of macropores in fibrous peat facilitates faster water drainage and hence, higher C_v, whereas in amorphous Sri Lankan organic soil, the limited number of macropores blocks easy dissipation of excess pore water, resulting in lower C_v values than for fibrous peat. Hence, constant C_v values for Sri Lankan organic soil can be expected up to $\sigma'_v/\sigma'_p = 1$. In the normally consolidated region, the reduction of C_v for Sri Lankan organic soil is less than that for fibrous peat. The inability of Sri Lankan amorphous peat towards further compression due to a lack of macropores and fibrous organic matter has resulted in the C_v variation observed in Figure 3.

According to Mesri and Ajlouni (2007), the secondary compression component of fibrous peat is more significant than the primary consolidation as it completes almost immediately after loading. As evident from Figures 4F and 4G, the highly organic fibrous peat, has a larger proportion of macropores in its microstructure. These macropores, combined with the undecomposed porous fibrous organic material, facilitate easy and rapid dissipation of pore water upon loading. Hence, the primary consolidation of these soils happens almost immediately after loading. In contrast, upon analysis of settlement monitoring data in the Colombo-Katunayake Expressway project to determine the end of primary consolidation using the Asaoka method (Asaoka, 1978), it was found that it takes about 6 to 8 months for completion of primary consolidation (Figure 2) in Sri Lankan

organic soil deposits. Zimar *et al.* (2020) claimed that the surficial deposits of the Sri Lankan organic soil have a considerable amount of macropores and micropores, and the oedometer tests reported achieving the end of primary consolidation in a short period. However, even these proportions of pores are considerably lower than that reported for highly organic fibrous peat. Furthermore, when the deep layers of organic soils are considered, the proportion of both macropores and micropores is reduced (Figure 4). This absence of enough macropores reduces the rate of dissipation of pore water and consequently increases the time taken to achieve the end of primary consolidation. Therefore, unlike highly organic fibrous deposits, Sri Lankan organic deposits take a considerably longer time to achieve primary consolidation under an embankment load.

Validity of commonly used empirical correlations for organic soils with low organic content.

Empirical correlations can be used to determine the engineering parameters of organic soils with considerable accuracy without the need to conduct expensive ground investigations. Numerous such correlations reported in the literature have been derived for different geological conditions across the world. It was observed in the previous sections that the Sri Lankan organic soils are significantly different from the highly organic soil deposits. Thus, it is required to compare the already existing correlations for organic soils to check whether they are valid for low organic soil deposits in Sri Lanka and whether it is required to have a separate set of correlations derived for such soils. The present study used data from the literature, data from pre-construction and post-construction ground investigation reports of the CKE project (obtained from RDA), and data from laboratory tests conducted on undisturbed soil samples (present study).

Rather than using simple curve fitting, the cross-validation analysis technique was applied using the Python programming language to check the most suitable degree of the polynomial of the correlations. Each correlation was then compared with similar correlations reported in the literature to validate the derived correlations and to check the effects of microstructure characteristics on the correlations. Figures 5 -12 contain the comparison of correlations derived in the present study with similar correlations available in the literature. Figures 5, 6, 8, 9, 10, 11, and 12 were reproduced from Hobbs (1986), whereas Figure 7 was replicated from Huat *et al.* (2007). The correlations obtained in the current study were then plotted on these reproduced graphs for comparison

Correlations that depend on the microstructure characteristics and type of deposit

The microstructure of the Sri Lankan organic soil is mostly amorphous with no traces of a proper cellular structure and most voids are filled with mineral particles. Therefore, the w_0, e_0, LL, and C_C values were low compared to highly organic soils. This is reflected in Figures 5 -12 as most of the data points representing Sri Lankan organic soils fall in lower w_0, e_0, LL, and C_C ranges. Figures 5 and 6 represent the comparison of the correlation between w_0 vs e_0 and C_C vs w_0 respectively. The correlations derived in the present study between the above parameters are given in equations 1 and 2.

$$e_0 = 0.0156 \, w_0 + 0.942 \qquad \text{...(1)}$$

$$c_c = 0.0078 \, w_0 \qquad \text{...(2)}$$

It can be observed in Figures 5 and 6 that the correlation derived in the present study falls among the correlations for organic soils with similar conditions, such as amorphous microstructure and high mineral content. Hobbs (1986) stated that the UK Fen peat has an amorphous microstructure, and Cook (1956) mentioned that most of the peat samples used in his study were naturally mixed with contaminants (mineral soil). Lea & Brawner (1963), Miyakawa (1960), and Mickleborough (1961) did not specify the type of peat that mostly contributed to the derivation of the correlation. Furthermore, in the above two comparisons, the correlation derived for the present study shows clear variation from the correlation for highly organic fibrous soils such as Quebec peat (Brochu & Pare, 1964) and Welsh bog peat (Hobbs, 1986).

There is an excess scatter between the correlations for different organic soil deposits in Figure 6. This is mainly due to the differences in the amounts of water retained (water-holding ability) in the following states: (i) intracellular water, (ii) interparticle water held by capillary force, and (iii) adsorbed water (Ramanov, 1968; Hayward & Clymo, 1982; Hobbs, 1986; Farrell, 2012; O'Kelly, 2014). Only the water in states (i) and (ii) will be dissipated during consolidation. Therefore, the consolidation of soil by expelling water does not entirely depend on the total amount of moisture but depends mostly on the amount of water held in each state (Hobbs, 1986). When oven drying at 60 °C to determine w_0, a small quantity of water in conditions (i) and (ii) above will not be entirely removed (O'Kelly, 2004). On the other hand, oven drying at 105 °C to determine w_0 will cause the fibrous material to char, eventually overpredicting the w_0 (O'Kelly, 2004). Therefore, the w_0 determined at 60 °C and 105 °C can have different values and slightly misinterpret the definite amount of water held in each state. This can be considered a major contributor to the excess data scatter and significant variation between correlations observed in Figure 6.

As evident by the above observations, the correlations derived in the present study (equations 1 and 2) may be successfully used in a low organic content soil with amorphous microstructure and high mineral content.
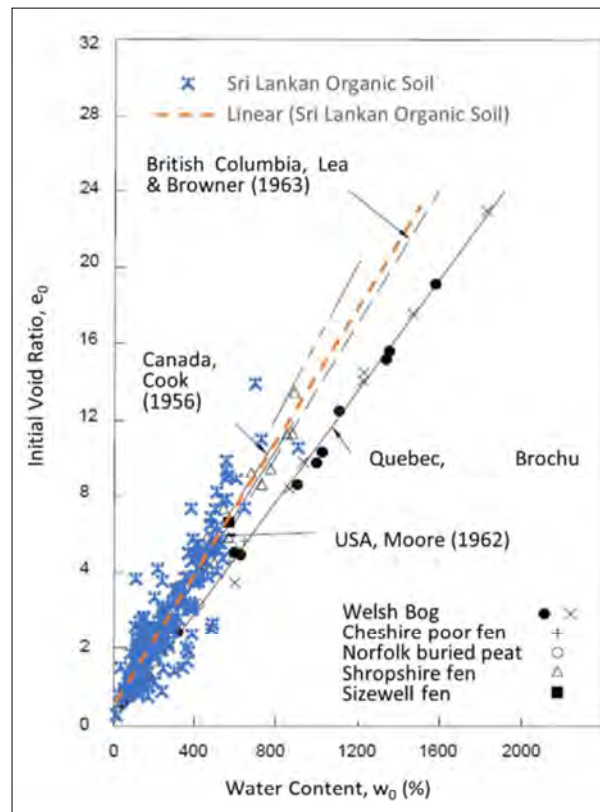

Figure 5: Comparison of correlation between w_0 and e_0

The correlation between ρ_d and w_0 for Sri Lankan organic soil is given in equation (3) and illustrated in Figure 7, along with some correlations from different geological conditions. The relationships obtained by Den Haan (1997) and Huat *et al*. (2007) are given in equations (4) and (5) respectively for comparison.

$$\rho_d = 33.462 \, w_0^{-0.842} \qquad \text{...(3)}$$

$$\rho_d = 35.075 \, w_0^{-0.856} \qquad \text{...(4)}$$

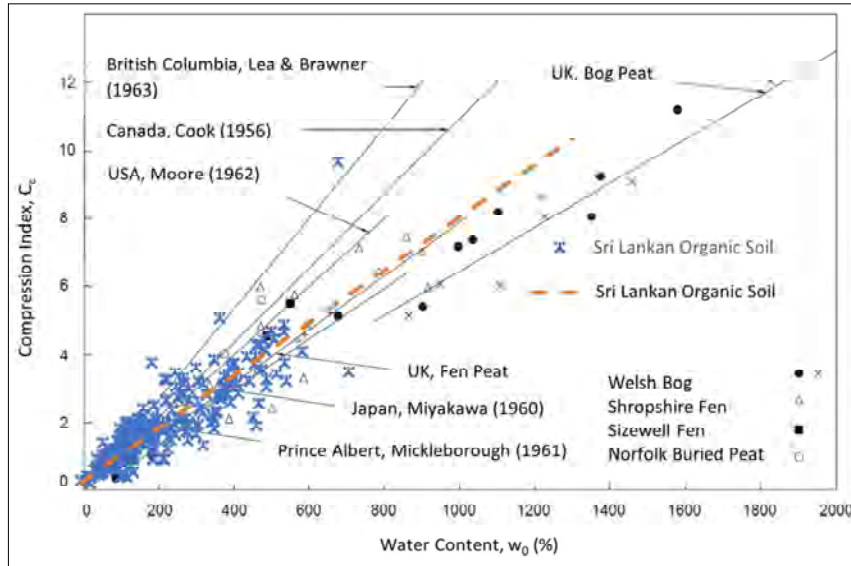$$\rho_d = 24.422 \, w_0^{-0.804} \qquad \text{...(5)}$$

Figure 6: Comparison of the correlation between C_c and w_0
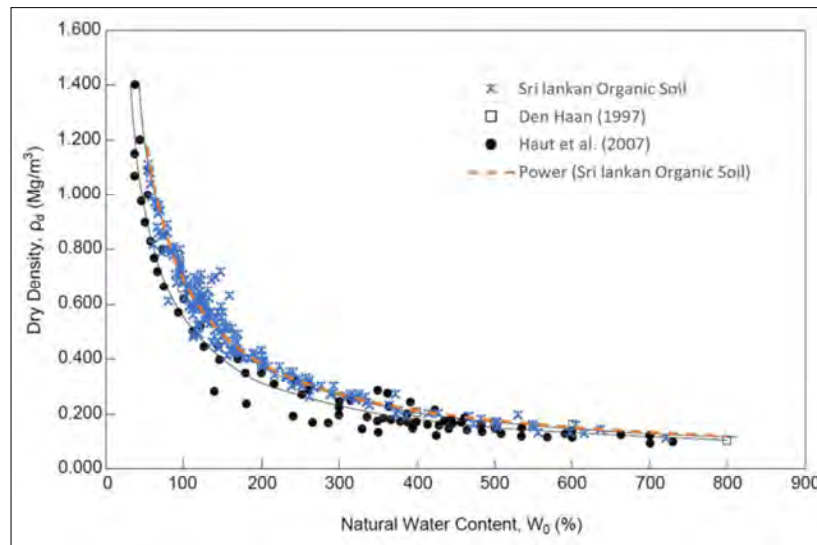

Figure 7: Comparison of correlation between ρ_d and w_0

Den Haan (1997) observed that the ρ_d *vs* w_0 correlation does not vary much for different organic contents and specific gravity values. According to Figure 7, data from the present study follow a trend similar to the relationship proposed by Den Haan (1997) which was derived based on organic soil with low organic content mixed with impurities. The relationship given by Huat *et al.* (2007) falls below the regression for Den Haan (1997) and the present study, which depicts that the dry densities of the local soil deposits are slightly high. The organic content of the samples that were considered by Huat *et al.* (2007) is in the range of 70% - 88%, with an average specific gravity of 1.5. As evident by the SEM images in Figure 2, Sri Lankan organic soil has decomposed organic matter with fewer macro and micropores. This lack of pores, combined with the presence of a higher percentage of mineral particles has resulted in higher dry density values when compared to soils with high organic content. The correlation derived in this study (equation 3) might be used for organic soils with low organic content and similar microstructure characteristics.
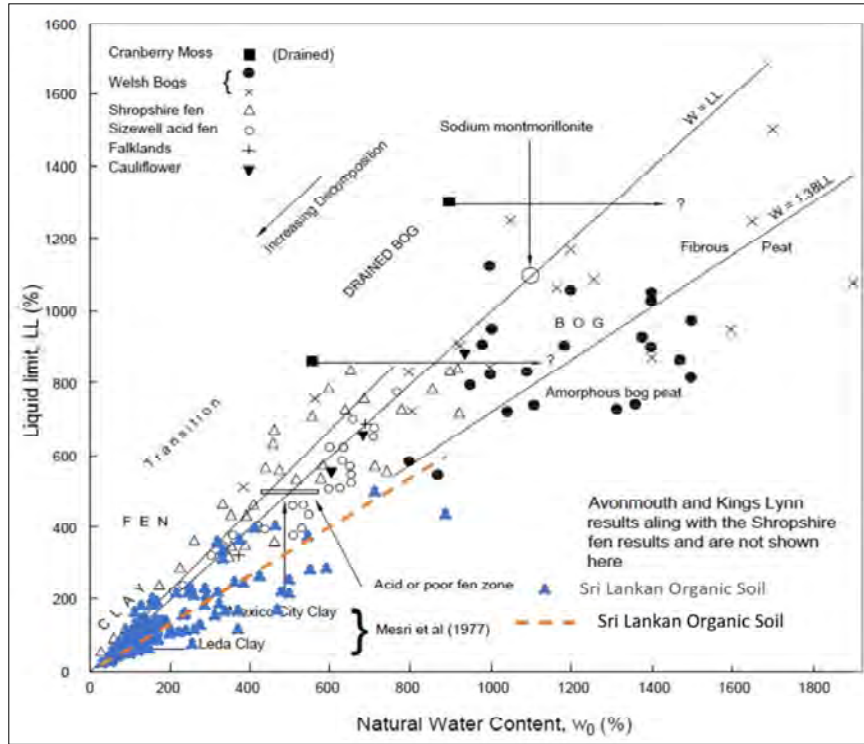
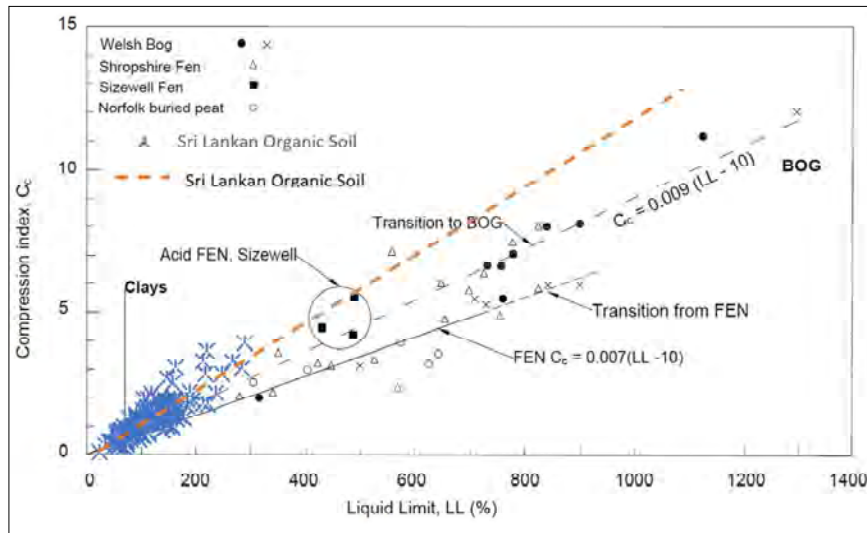**Figure 8:** Comparison of correlation between w_0 and L_L

Figure 9: Comparison of correlation between C_c and L_L

Special observations

In the above section, it was observed that the correlations for Sri Lankan organic soils follow trends similar to other amorphous organic soils. However, in Figures 8 and 9, Sri Lankan organic soil follows different trends.

Organic soils with high mineral content and amorphous structure have LL closer to their w_0 as the interparticle voids are filled with liquid clay (Hobbs, 1986). This is shown by the correlation of $w_0 = LL$ obtained for amorphous UK Fen peat and NC clay in Figure 8. In highly organic soils such as UK Bog peat, the interparticle voids

are filled with water. Therefore, their LL is less than the w_0. Even though having an amorphous structure, organic soil deposits such as acidic amorphous peat in Sizewell, acidic amorphous UK bog peat, and acidic organic soil deposits in Sri Lanka (pH = 5.0) (Dissanayake *et al.*, 1982; Dissanayake, 1987; Karunawardena, 2007)) has LL lower than its w_0 (Figure 8). Hobbs (1986) proposed that the acidity in organic soil reduces its cation exchangeability. According to Yilmaz (2004), cation exchangeability is directly proportional to LL. Thus, these acidic amorphous organic soils have lower LL values for a given w_0 than their non-acidic counterparts. This behaviour is further observed in Figure 9, where local organic soil and Sizewell peat have a lower LL value than the other amorphous organic soils for a given C_C value. Therefore, the correlations derived in the present study between w_0 vs LL and C_C vs LL, given in equations 6 and 7, respectively, may be used for acidic organic soils with similar microstructure characteristics.

$$w_0 = 1.5124 \, LL \qquad \qquad ...(6)$$

$$C_c = 0.0118 \, (LL - 7.31) \qquad ...(7)$$

Correlations that are independent of the microstructure characteristics and type of deposit

Correlations discussed in the previous sections showed behavioural differences depending on their microstructure characteristics. However, when comparing correlations given in Figures 10, 11, and 12, such variations were not observed.

Generally, the recompression index of an organic soil sample is 10% of its compression index value, which may vary within a range of 5% to 20%. The correlation developed for Sri Lankan organic soil given in equation 8 was compared with the relationship in Hobbs (1986), as shown in Figure 10. It was observed that the data points relevant to different types of organic soils strictly satisfy one relationship. Hobbs (1986) then concluded that the correlation between C_c and C_r is not affected by morphological differences between the two deposits. However, the type of testing may affect the relationship. The variation between the local correlation and the relationship suggested by Hobbs (1986) may be due to differences in the methods of testing.

$$C_r = 0.0987 \, C_c \qquad \qquad ...(8)$$

Generally, organic soil is undersaturated due to the presence of gas emitted from humification. Its bulk density (ρ) is less than mineral soils and is related to its organic content (which influences the specific gravity (G_S)), natural water content (w_0) and degree of saturation (S_r) as evidenced by equation (9) (Hobbs, 1986).

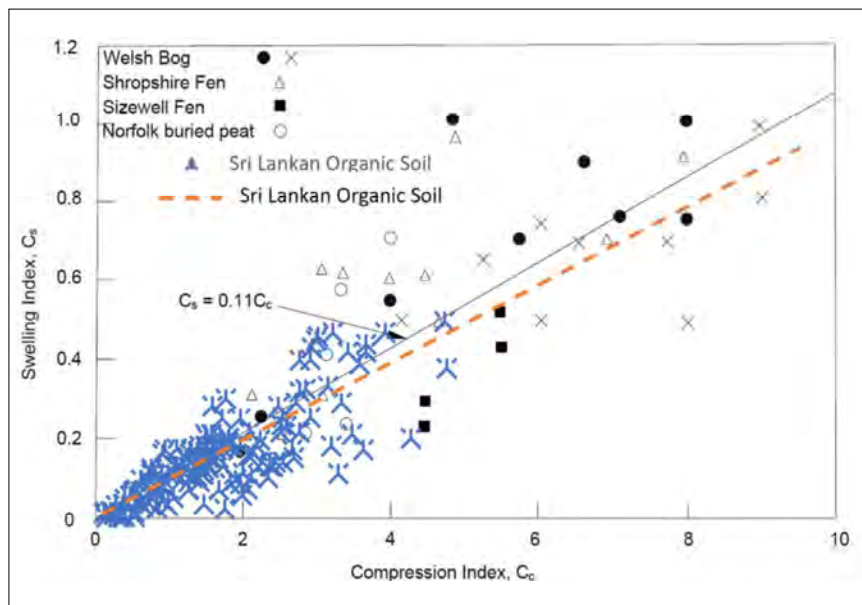$$\rho = \left(\frac{1+w_0}{s_r + G_S \, w_0} \right) G_s \, S_r \, \gamma_w \qquad ...(9)$$

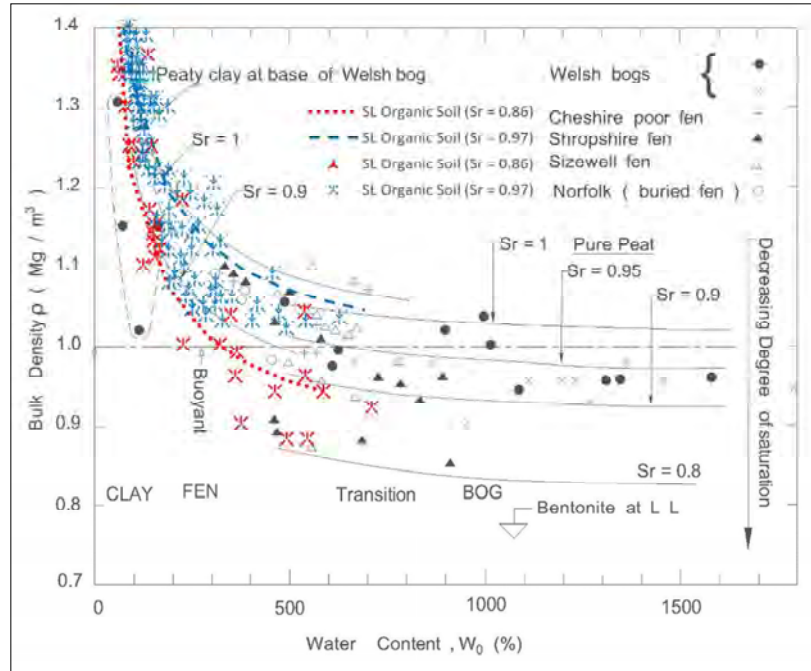**Figure 10:** Comparison of the correlation between C_c and C_r

Figure 11: Comparison of correlation between ρ and w₀

The data available for Sri Lankan organic soil was divided into two ranges depending on the S_r values. These two sets of data were then used to graphically represent equation 9 in Figure 11. It is observed that the curves representing the Sri Lankan organic soil deposits follow the curves for other organic soils drawn using equation 9 but with different G_s and S_r values. Such evidence suggests that equation 9 can be used regardless of the difference in the microstructure characteristics. The inconsistency in degree saturations of different samples at higher water contents causes the excessive scatter of data points observed in Figugre 11.

The correlation between e_0 and C_c for Sri Lankan organic soil is given in equation 10 and represented in Figure 12. Here, Hobbs (1986) only defines one correlation for both amorphous and fibrous organic soil, indicating that the correlation is valid for deposits with different microstructure characteristics. The correlation derived in the present study follows a similar trend as the correlation proposed by Hobbs (1986). Therefore, the correlation proposed in the present study may be used for a deposit with similar microstructure characteristics.

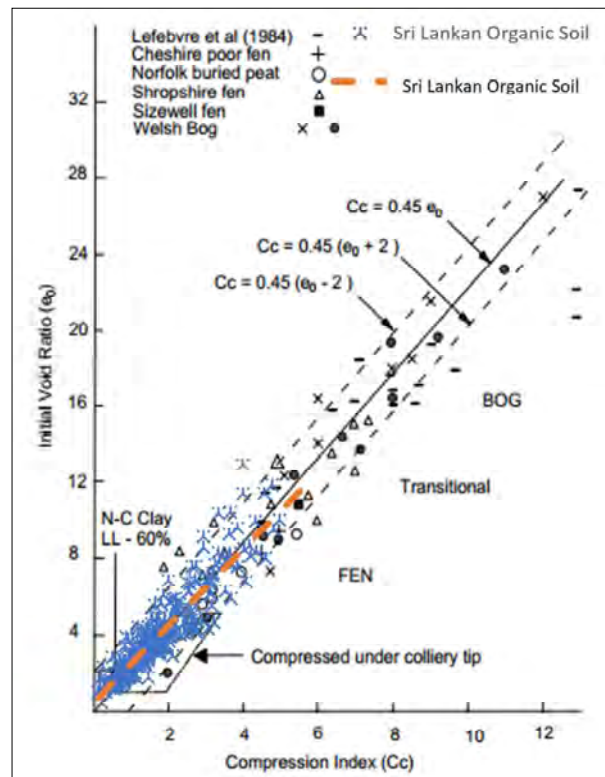$$C_c = 0.4272\, e_0 - 0.0536 \qquad \qquad ...(10)$$

Figure 12: Comparison of correlation between e_0 and C_c

CONCLUSIONS

It was observed that the soil properties and compressibility behaviour of Sri Lankan low-organic clayey soils differ from the highly organic soils in other parts of the world. Thus, in this study, the microstructure of Sri Lankan organic soil was evaluated to understand the observed variations. The microstructure analysis through SEM images showed that the Sri Lankan organic soil has an amorphous microstructure that contains slightly to highly decomposed organic matter and a considerable proportion of mineral particles. As a result, the w_0, e_0, LL, C_c and G_s values considerably vary from those reported for highly organic soils. Due to the considerable amounts of mineral matter in the Sri Lankan organic soil, consistency limits can be considered a critical soil parameter, contrary to the highly organic peaty soils, where consistency limits are often disregarded. The lack of macropores observed in the microstructure combined with low amounts of highly compressible organic matter resulted in reduced rates of pore water dissipation and, hence, slow rates of primary consolidation. Due to these observed differences in properties and behaviour of Sri Lankan organic soil deposits, it was required to derive and update correlations between its properties. It was observed that the e_0 vs w_0, w_0 vs C_c, and ρ_d vs w_0 correlations for Sri Lankan organic soil agree well with the relationships for other amorphous organic soil types worldwide. In contrast, correlations such as C_c vs C_r, ρ vs w_0 and e_0 vs C_c do not vary depending on the type of deposit or microstructure. Due to the acidity of the Sri Lankan organic soil, the correlations related to liquid limits demonstrated slightly different behaviour from the non-acidic amorphous soils. The correlations suggested in this study can be used to predict soil properties in any low organic clayey deposits with similar microstructures. In conclusion, this study highlights the importance of understanding the effect of microstructure characteristics on the engineering behaviour of low organic clayey soil.

Acknowledgements

REFERENCES

Adams J.I. (1965). The engineering behavior of a Canadian muskeg. *Proceeding of the 6th International Conference on Soil Mechanics and Foundation Engineering,* University of Toronto Press, Toronto.

Adnan Z. & Wijeyesekara D.C. (2008). Geotechnical Characteristics of Peat. *Proceedings of the 3rd Annual Conference on Advances in Computing and Technology Conference,* University of East London, Crowne Plaza Hotel, Docklands, London.

Ariyarathna P.R.C., Thilakasiri H.S. & Karunawardane W.A. (2010). Vacuum consolidation of Sri Lankan peaty soils. *Proceedings of the Annual Sessions of Institution of Engineers,* Institution of Engineers, Colombo, Sri Lanka.

Asaoka A. (1978). Observational procedure of settlement prediction. *Soils and Foundations* **18**(4): 87–101.
DOI: https://doi.org/10.3208/sandf1972.18.4_87

Brochu P.A. & Pare J.J. (1964). Construction des routes sur tourbieres dans la Province de Quebec. *Proceedings of the 9th Muskeg Research Conference,* 21 May. National Research Council, Ottawa, Ontario,Canada. pp. 267.

Cook P.M. (1956). Consolidation characteristics of organic soils. *Proceedings of the 9th Canadian Soil Mechanics Conference,* 15 December. National Research Council, Ottawa, Ontario, Canada. pp 105.

Den Haan E.J. (1997). An overview of the mechanical behavior of peats and organic soils and some appropriate construction techniques. *Proceedings of the Conference of Recent Advances in Soft Soil Engineering,* 5-7 March. Kuching, Sarawak, Malaysia.

Den Haan E.J. & Kruse G.A.M. (2006). Characterisation and engineering properties of Dutch peats. *Proceedings of the 2nd International Workshop on Characterisation and Engineering Properties of Natural Soils,* 29 November-1 December. Singapore.

Dissanayake C.B. (1987). Metals in a lateritic peat deposit - a case study from Sri Lanka. *Proceedings of the International Seminar on Laterite,* 14 -17 October. Tokyo, Japan.

Dissanayake C.B., Senaratne A. & Gunatilaka A.A.L. (1982). Organic geochemical studies of the Muthurajawela peat deposit of Sri Lanka. *Organic Geochemistry* **4**(1):19–26.
DOI: https://doi.org/10.1016/0146-6380(82)90004-3

Edil T.N. & Den Haan E.J. (1994). Settlement of peats and organic soils. *Proceedings of the Settlement '94,* 16-18 June. Department of Civil Engineering, Texas A&M University College, Texas, USA.

Farrell E.R. (2012). Chapter 35: Organics/Peat soils. In: *ICE Manual of Geotechnical Engineering*, pp. 63–479. Institution of Civil Engineers, UK.

Force E.A. (1998). Factors controlling pore-pressure generation during K_0-consolidation of laboratory tests. *Dissertation (MSc.)*, Department of Civil and Environmental Engineering, MIT, Massachusetts, USA.

Hayward P.M. & Clymo R.S. (1982). Profiles of water content and pore size in sphagnum and peat, and their relation to peat Bog ecology. *Proceedings of The Royal Society B* **215**: 299–325.
DOI: https://doi.org/10.1098/rspb.1982.0044

Hobbs N.B. (1986). Mire morphology and the properties and behaviour of some British and foreign peats. *Quarterly Journal of Engineering Geology* **19**:7–80.
DOI: https://doi.org/10.1144/GSL.QJEG.1986.019.01.02

Hsi J., Gunasekara C. & Nguyen V. (2015). Characteristics of soft peat, organic soils and clay, Colombo-Katunayake Expressway, Sri Lanka. In: *Ground improvement Case Histories,* (eds. B. Indrarathne & J. Chu), pp. 681–722. Butterworth-Heinemann, UK.
DOI: https://doi.org/10.1016/B978-0-08-100192-9.00025-9

Huat B.B.K. (2006). Deformation and shear strength characteristics of some tropical peat and organic soil. *Pertanika Journal of Science and Technology* **14**(1-2): 61–74.

Huat B.B.K., Duraisamy Y. & Aziz A.A. (2007). Engineering properties and compressibility behavior of tropical peat soil. *American Journal of Applied Science* **4**(10): 768–773.
DOI: https://doi.org/10.3844/ajassp.2007.768.773

Karunawardane A.W. (2007). Consolidation analysis of Sri Lankan peaty clay using elasto-viscoplastic theory. *Dissertation (PhD)*, Kyoto University, Japan.

Kazemian S., Huat B.B.K., Prasad A. & Barghchi M. (2011). A state of art review of peat: Geotechnical engineering perspective. *International Journal of the Physical Sciences* **6**(8): 1974–1981.
DOI: https://doi.org/10.5897/IJPS11.396

Kearns F.L. & Davison A.T. (1983). Field classification of organic rich soils. *Proceedings of the Conference Mineral Matter in Peat: Its Occurrence, Form and Distribution.* Los Alamos National Laboratory, Los Alamos, NM, USA., pp. 147–157.

Lea N.D. & Brawner C.O. (1963). Highway design and construction over peat deposits in Lower British Columbia. In: *Highway Research Record No. 7,* pp 1-31. Highway Research Board, Washington DC, USA.

Lee J.S., Seo S.Y. & Lee C. (2015). Geotechnical and geophysical characteristics of muskeg samples from Alberta, Canada. *Engineering Geology* **95**:135–141.
DOI: https://doi.org/10.1016/j.enggeo.2015.04.030

Leong E.C. & Chin C.Y. (2000). Geotechnical characteristics of peaty soils in southeast Asia. *Proceedings of the International Conference on Geotechnical and Geological Engineering (Geoengineering 2000),* 19 - 24 November, Melbourne, Australia.

Mesri G., Stark T., Ajlouni M. & Chen C. (1997). Secondary compression of peat with or without surcharging. *Journal of Geotechnical and Geoenvironmental Engineering* **123**: 411–421.
DOI:https://doi.org/10.1061/(ASCE)1090-0241(1997)123: 5(411).

Mesri G. & Ajlouni M. (2007). Engineering properties of fibrous peat. *ASCE Journal of Geotechnical and Geoenvironmental Engineering* **133**(7): 850–866.
DOI: https://doi.org/10.1061/(ASCE)1090-0241(2007)133: 7(850)

Mickleborough W.B. (1961). Embankment construction in Muskeg at Prince Albert. *Proceedings of 7th Muskeg Research Conference,* 18 April. National Research Council of Canada, Ottawa, Ontario, Canada.

Miyakawa I. (1960). *Some aspects of road construction in peaty or marshy areas in Hokkaido.* Civil Engineering Research Institute, Hokkaido Development Bureau, Sapporo, Japan.

O'Kelly B.C. (2004). Accurate determination of moisture content of organic soils using the oven drying method. *Drying Technology* **22**(7): 1767–1776,
DOI: https://doi.org/10.1081/DRT-200025642.

O'Kelly B.C. (2014). Characterisation and undrained strength of amorphous clay; Proceedings of the Institute of Civil Engineers. *Geotechnical Engineering* **167**(3): 311–320.
DOI: https://doi.org/10.1680/geng.11.00025

O'Kelly B.C. & Pichan S.P. (2013). Effects of decomposition on the compressibility of fibrous peat - A review. *Geomechanics and Geoengineering* **9**(4): 286–296.
DOI: https://doi.org/10.1080/17486025.2013.804210

Raghunandan M.E. & Anirudh S.S. (2017). An overview of the basic engineering properties of Malaysian peats. *Geoderma Regional* **11**:1–7.
DOI: https://doi.org/10.1016/j.geodrs.2017.08.003

Romanov V.V. (1968). *Hydrophysics of Bogs*. pp. 237-248. Israel Program for Scientific Translations, Jerusalem.

Santagata M.C. (1999). Factors affecting the initial stiffness and stiffness degradation of cohesive soils. *Dissertation (PhD)*, Department of Civil and Environmental Engineering, MIT, Cambridge, Massachusetts, USA.

Santagata M.C., Bobet A., Johnston C.T. & Hwang J. (2008). One-dimensional compression behavior of a soil with high organic matter content. *Journal of Geotechnical and Geoenvironment Engineering* **134**(1):1–13.
DOI:https://doi.org/10.1061/(ASCE)1090-0241(2008)134: 1(1)

Sridharan A. & Nagaraj H.B. (2012). Coefficient of consolidation and its correlation with index properties of remolded soils. *Geotechnical Testing Journal* **27**(5):1–6,
DOI: https://doi.org/10.1520/GTJ10784

Thavasuthan T. & Thilakasiri H.S. (2011). Empirical correlations for Sri Lankan peaty soils. *Geotechnical Journal of Sri Lankan Geotechnical Society* **5**:27–33.

Vasander H. (2014). Overview of types of peatlands. In: *Towards Climate-Responsible Peatlands Management. Mitigation Of Climate Change* on Agriculture Series, no. 9 (eds. R. Biancalani , A. Avagyan), pp 15–18. FAO, Rome, Italy.

Vikas K.J., Mahabir D. & Chitra R. (2015). Correlation of plasticity index and compression index of soil. *International Journal of Innovations in Engineering and Technology* **5**(3): 263–270.
DOI: https://doi.org/10.1515/9781400868537-041

Wasif Z., Rezwan H., Hossain S. & Akin A.A. (2016). A study on correlation between consolidation properties of soil with liquid limit, in situ water content, void ratio and plasticity index. *Proceedings of the 3rd International Conference on Geotechnics of Sustainable Infrastructure Development,* 24-25 November. Hanoi, Vietnam.

Yilmaz I. (2004). Relationships between liquid limit, cation exchange capacity, and swelling potentials of clayey soils. *Eurasian Soil Science* **37**: 506–512.

Zimar A.M.Z., Nasvi M.C.M. & Jayakody S. (2020). Geotechnical characterization of peats in Muthurajawela region in the western coast of Sri Lanka. *Geotechnical and Geological Engineering* **38**: 6679–6693,
DOI: https://doi.org/10.1007/s10706-020-01462-8

RESEARCH ARTICLE

Statistical Modelling

Comparison of methods for handling outliers in Cox regression model

N Alkan[1*], MC Pardo[2] and BB Alkan[3]
[1] *Department of Business Administration, Akdeniz University, Antalya, Turkey.*
[2] *Department of Statistics and Operational Research, Complutense University of Madrid, Spain.*
[3] *Department of Educational Sciences, Akdeniz University, Antalya, Turkey.*

Abstract: The Cox regression analysis is used to determine the relationship between a dependent variable and covariates in survival analysis involving censored data. The proportional hazards assumption is one of the most important assumptions of Cox regression. Outliers may have a strong influence on the Cox regression model's parameter estimates and lead to violation of the proportional hazard assumption. Therefore, having outliers in the data set is a problem for researchers. In this case, robust estimations are commonly used to infer the parameters in a more robust way. However, we explore a new approach consisting of considering an outlier as missing data and replacing it by the multiple imputation method. The aim of this study is to compare these two methods through simulation. Furthermore, an analysis of a lung cancer data set is considered for illustration. According to the results of the study carried out based on simulated data sets and a real data set, the multiple imputation method, which is a missing data analysis method, solves the problem of outliers better than the robust estimation method, as the outcome is closer to the results obtained through original data.

Keywords: Cox regression, multiple imputation, outliers, robust Cox regression.

INTRODUCTION

One of the most important assumptions of Cox regression is the proportional hazard assumption. Outliers in data could lead to violation of this assumption and it leads to the emergence of inaccurate estimates. Because outliers

may have a strong influence on the model's parameter estimates, the outliers may be a problem for researchers. In many studies, data have been remodelled by deleting the observation with outliers. But in this case the data set to be analysed becomes smaller, compromising the statistical power of the study and eventually the reliability of its results. For this reason, it is not right to drop an observation just because it is an outlier, *e.g.*, individuals that lived too long or died too early when compared with others with the same clinical conditions.

Robust statistics have good performance of data sets with outliers and other small departures from model assumptions. Robust estimators are a modified class of regression parameters estimators (Bernarski, 1989). In Cox regression analysis, the partial likelihood function is used to estimate the parameters and the modification process is executed via this function. Recently, Farcomeni and Viviani (2011) proposed a modified Cox model that is fitted by trimming the smallest contributions to the partial likelihood.

Also another suggestion for the outlier problem is to use the multiple imputation method which is one of the missing data analysis methods (Alkan & Alkan, 2018). In this method, instead of deleting all the values of the observation with an outlier, only the outlier is deleted and this missing value is imputed by using multiple imputation method.

* Corresponding author (nesrin.alkan@gmail.com; https://orcid.org/0000-0003-1452-4780)

So, in this study, the solution of the problem that results from a violation of assumptions is discussed using two different methods. Firstly, the problem caused by outliers is transformed into a missing value problem and it is solved by the multiple imputation method, and then Cox regression analysis was applied to the completed data set. Secondly, robust estimates which give accurate results in case of deviations from the assumptions are obtained for Cox regression analysis. Therefore, our aim is to compare the results of Cox regression analysis after multiple imputation methods and the results of robust Cox regression analysis.

For this purpose, we carried out a simulation study. The simulations were made to determine how much outliers influence the parameter estimates. The simulation scenarios were created according to different censor ratios (20%, 40%, 60%), different outliers ratios (10%, 20% 30%), and sample size N = 50, 100. So the multiple imputation and robust Cox regression have been compared in different situations. Furthermore, we revisit the popular NCCTG lung cancer data to illustrate both methods.

MATERIALS AND METHODS

Cox regression

Cox regression analysis is used extensively in biological and medical studies in survival analysis involving censored data. In survival analysis, the Cox regression analysis is used to determine the relationship between dependent variable and covariates (Cox, 1972). The Cox regression model can be written as:

$$\lambda_i(t) = \lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{X}_i) \qquad \ldots(1)$$

where $\lambda_i(t)$ represents the hazard function for the i-th individual, $\boldsymbol{\beta}'$ is the unknown parameter vector, $\mathbf{X_i}$ is the p-dimensional covariate vector for the i-th individual, and $\lambda_0(t)$ is called the baseline hazard function that represents the hazard when all the independent variables are equal to zero. The baseline hazard function is a non-parametric function that shows the change of hazard over time without considering the effects of specific covariates or independent variables (Hosmer & Lemeshow, 1999). This method estimates the parameters by maximizing the partial likelihood function (Kalbfleisch & Prentice, 1980). The partial likelihood is provided by the following equation:

$$\prod_{i=1}^{n}\left[\frac{\exp(\boldsymbol{\beta}'\mathbf{X}_i)}{\sum_{t_j \geq t_i}\exp(\boldsymbol{\beta}'\mathbf{X}_j)}\right]^{\delta_i} \qquad \ldots(2)$$

where t_i is the minimum of survival and censored time, $\delta_i = 0$ if censored and $\delta_i = 1$ if the event occurred. The proportional hazard assumption is the principal assumption of the Cox regression analysis. In this assumption, the hazard ratio (HR) of any two individuals is constant over the time axis in the model. Therefore, the Cox regression model is also known as a proportional hazard model. Furthermore, baseline hazard function is independent of the covariates. Reliable statistical inferences and estimates are obtained by providing this assumption (Kleinbaum & Klein, 1996).

In statistics, an outlier is an observation point that is different from the rest of the data. Outlier values in the dataset may have a great influence on parameter estimation. For this reason, model adequacy should be checked after the survival data set is modelled by Cox regression analysis (Xue & Schifano, 2017; Alonso & Pardo, 2020). Model diagnosis is one of the most important parts of the modelling process. Many diagnostic methods are based on the analysis of model residuals. One of the most well-known is the Schoenfeld residual analysis. The Schoenfeld residuals (1982) are the difference between the true value of the covariate and the average of weighted risk scores.

$$\hat{r}_k(\beta) = \mathbf{X}_{(k)} - \bar{x}(\beta, t_k), \qquad \ldots(3)$$

$$\bar{x}(\beta, t_k) = \frac{\sum_{i=1}^{n} Y_i(t_k)\exp(\boldsymbol{\beta}'\mathbf{X}_i(t_k))\mathbf{X}_i(t_k)}{\sum_{i=1}^{n} Y_i(t_k)\exp(\boldsymbol{\beta}'\mathbf{X}_i(t_k))} \qquad \ldots(4)$$

where $\bar{x}(\beta, t_k)$ is a weighted average of covariates over observations which are still at risk at time t_k. $\mathbf{X}_{(k)}$ is k-th covariate vector of a subject with event time t_k. $Y_i(t_k)$ indicates whether the i-th subject is still at risk at time t_k. The term $\exp(\beta'X_i(t_k))$ is the risk score for the ith observation.

The approach to test the proportional hazard assumption in Schoenfeld (1982) is generalized by Grambsch and Therneau (1994). They defined the following function to test the proportional hazard assumption for each covariate j,

$$T_j(g) = \frac{\sum(g_j\hat{r}_{jk})^2}{D_{jj}} \qquad \ldots(5)$$

where g_j is an element of G_k which is a diagonal matrix and shows how the survival times should be transformed, $\hat{r}_{jk}$ is the j-th element of Schoenfeld residual, D_{jj} is an element of $D = \sum G_k\hat{V}_kG_k^r - \left(\sum G_k\hat{V}_k\right)\left(\sum \hat{V}_k\right)^{-1}\left(\sum G_k\hat{V}_k\right)^T$ where $\hat{V}_k$ is the observed variance of $\hat{\beta}$ at the k –th time. This test statistic is used for testing the proportional

hazard assumption of each covariate. The test statistic is distributed as $\chi^2_{(1)}$ if the proportional hazard assumption is provided (Xue & Schifano, 2017). Graphs of the Schoenfeld residuals against transformed time are used for checking violations of the proportional hazard assumption. If the residuals are around a horizontal line, the proportional hazard assumption is satisfied (Schoenfeld, 1982). Both statistical tests and graphical diagnostics which are based on the scaled Schoenfeld residuals are used to check proportional hazard assumption.

Multiple imputation method

Missing data may be encountered even in a well-planned and controlled study. If there is a missing value in the data, statistical power is reduced, biased estimates are produced and invalid results are obtained. Therefore, missing data is an important problem for researchers. Also statistical methods and software suppose that all variables in a model are complete. To solve the missing data problem, either the observation which has a missing value is deleted, or the value is imputed. The multiple imputation (MI) method, which has a better performance than other imputation methods, is a missing data analysis method (Alkan *et al*, 2013).

The multiple imputation method develops the Bayesian approaches to solve the problem of missing value in the data (Enders, 2010). In the multiple imputation method, for generating *m* complete data sets, each of the missing values is filled in *m* times. Standard statistical methods analyze the imputed data sets and combine the results from these analyses for the inference. Rubin (1987) outlined the formulas for combined parameter estimates, which are based on the arithmetic mean of the *m* complete data estimates. Our first proposal for handling the outlier problem is to consider each outlier as missing data as proposed Alkan and Alkan (2018).

Robust cox regression

The partial likelihood estimator β used for parameter estimation in the Cox regression is very sensitive to deviations from the model assumptions. Outlier values cause a violation of the most important assumption of the Cox regression. In such a case, unreliable, mis-established models can occur. For this reason Farcomeni and Viviani (2011) proposed the Robust Cox Regression for data sets with outliers. Bretagnolle and Huber-Carrol (1988) have shown that exclusion of the relevant covariate with outlier gives biased results. Reid and Crepeau (1985) and Bednarski (1989) have shown that even slight departures

from the proportional hazard assumption lead to bias in the estimation of β. Bednarski (1993) showed how the proportional hazard estimator's prediction equation was modified to get robust estimates. Bednarski (1993) started by using equation (2) for this modification process and equation (6) was obtained

$$\sum_{i=1}^{n} \left[\mathbf{X_i} - \frac{\sum_{t_j \geq t_i} X_j \exp(\boldsymbol{\beta'X_j})}{\sum_{t_j \geq t_i} \exp(\boldsymbol{\beta'X_j})} \right] \delta_i = 0 \qquad ...(6)$$

The estimator solving this equation is much affected by the large values of $\exp(\boldsymbol{\beta'}\mathbf{X_j})$. One way of reducing the effect of large values is to regulate equation (6) Equation (6) is modified using the funtion A(t, X), which is zero. Take a smooth non-negative function to obtain equation (7).

$$\sum_{i=1}^{n} A(t_i, X_i) \left[\mathbf{X_i} - \frac{\sum_{t_j \geq t_i} A(t_i, X_j) x_j \exp(\boldsymbol{\beta'X_j})}{\sum_{t_j \geq t_i} A(t_i, X_j) \exp(\boldsymbol{\beta'X_j})} \right] \delta_i = 0$$
$$...(7)$$

The function A (t,X) is processed at two points. The outer sum is used to reduce the weight of uncensored observations which have large values of $\exp(\boldsymbol{\beta'}\mathbf{X_j})$. The A(t,X) in square brackets are computed for down-weighting all observations with relatively large values of $\boldsymbol{\beta'}\mathbf{X_j}$. Such double correction leads to the consistency of the estimator. The robust estimator of $\boldsymbol{\beta}$ is obtained by solving equation (7). Our second proposal for handling the outlier problem is to use robust estimation which traditionally addresses the consequences of having outliers in the data.

RESULTS AND DISCUSSION

In this study, the results of Cox regression analysis after multiple imputation methods and robust Cox regression analysis were compared for both simulation datasets and real datasets. In the simulation study, a Cox regression model with three covariates is considered. Standard normal distribution is assumed for the covariates. A Weibull distribution with scale parameter equal to 0.002 and shape parameter equal to 1 is used for baseline hazard function. A shape of 1 means a constant baseline hazard function. Censoring times are generated from a Weibull with scale parameter equal to 0.008 and shape parameter equal to 1 for a censoring proportion of 20%. And scale parameters of 0.004 and 0.002 are chosen to produce specific censoring proportions of 40% and 60%, respectively. The minimum of event time or censoring time was recorded as survival time. The true regression coefficients are fixed as $\beta_1 = 1$, $\beta_2 = -3$, $\beta_3 = 2$. Sample sizes n = 50 and 100 are selected.

In order to generate the data sets with outliers, some of the extreme values are given to the largest and smallest values of X_1 and X_2 in the simulated data sets. These values are evaluated as outliers. Changes are not made in the values of X_3. At this stage of the simulation study, different outlier proportions of 10%, 20%, and 30% are considered. In order to illustrate the use of multiple imputation method in the presence of outliers, the outlier values in the data are deleted and missing data sets are created. Cox regression analysis for clean simulated survival data sets, robust cox regression analysis for data sets with outliers, and multiple imputation method for missing data sets are used and estimation are repeated 500 times for each simulation. To determine how much outliers influence the parameter estimates and to compare the methods, Bias and RMSE are calculated as follows

$$\text{Bias} = \frac{\sum_{i=1}^{N}(\beta_i - \hat{\beta}_i)}{N}; \qquad \qquad ...(8)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(\beta_i - \hat{\beta}_i)^2}{N-1}} \qquad \qquad ...(9)$$

where N = 500 and β_i is the parameter estimation of the Cox regression model which is calculated from the data set with no outliers. $\hat{\beta}_i$ is the parameter estimation of robust Cox regression and Cox regression after Multiple İmputation. So $\text{Bias}_{\text{Robust}}$ and $\text{Bias}_{\text{MI+Cox}}$ are calculated. Similar calculations are made for RMSE and the results are given in Table 1 and Table 2.

Table 1: Bias and RMSE for different outlier and censor ratios and N = 50 for each of the methods.

N	Censor ratio	Outlier ratio	Parameter	Bias MI+Cox	Bias robust Cox	RMSE MI+Cox	RMSE robust Cox
			β_1	0.262	0.415	0.342	0.738
		10%	β_2	0.174	0.399	0.265	0.685
			β_3	0.194	0.309	0.371	0.563
			β_1	0.046	0.153	0.059	0.179
	20%	20%	β_2	0.108	0.293	0.159	0.341
			β_3	0.031	0.063	0.037	0.089
			β_1	0.726	1.176	0.948	1.497
		30%	β_2	1.743	2.432	2.359	3.042
			β_3	0.861	0.964	1.417	1.458
			β_1	0.174	0.416	0.249	0.545
		10%	β_2	0.569	1.291	0.671	1.901
			β_3	0.187	0.487	0.234	0.731
			β_1	0.229	0.765	0.293	0.912
50	40%	20%	β_2	0.441	0.932	0.551	1.545
			β_3	0.397	0.534	0.495	0.869
			β_1	0.320	0.883	0.459	0.948
		30%	β_2	0.766	1.098	0.912	1.409
			β_3	0.685	0.446	0.801	0.509
			β_1	0.351	0.610	0.499	0.738
		10%	β_2	0.499	0.509	1.067	0.702
			β_3	0.570	0.189	0.888	0.268
			β_1	0.550	0.706	0.721	0.920
	60%	20%	β_2	0.605	1.269	0.725	1.717
			β_3	0.415	0.339	0.532	0.533
			β_1	0.374	0.519	0.435	0.634
		30%	β_2	0.815	1.086	1.123	1.452
			β_3	0.639	0.506	0.904	0.749

Table 2: Bias and RMSE for different outlier and censor ratios and N = 100 for each of the methods.

N	Censor ratio	Outlier ratio	Parameter	Bias MI+Cox	Bias robust Cox	RMSE MI+Cox	RMSE robust Cox
			β_1	0.243	0.476	0.608	1.189
		10%	β_2	0.422	0.512	1.056	1.281
			β_3	0.233	0.247	0.583	0.617
			β_1	0.439	0.862	1.098	2.155
		20%	β_2	0.932	1.359	2.33	3.397
			β_3	0.612	0.679	1.532	1.699
	20%		β_1	0.659	0.832	1.649	2.081
		30%	β_2	1.351	2.295	3.379	5.739
			β_3	0.901	0.929	2.251	2.322
			β_1	0.281	0.456	0.702	1.140
		10%	β_2	0.395	0.523	0.988	1.308
			β_3	0.262	0.262	0.654	0.656
			β_1	0.502	0.717	1.256	1.792
		20%	β_2	0.839	1.798	2.097	4.494
			β_3	0.612	0.724	1.530	1.811
	40%		β_1	0.636	0.687	1.589	1.717
		30%	β_2	1.112	1.786	2.780	4.466
			β_3	0.696	0.677	1.742	1.692
100			β_1	0.447	0.578	1.118	1.446
		10%	β_2	0.648	1.244	1.620	3.109
			β_3	0.494	0.709	1.235	1.773
			β_1	0.447	0.725	1.118	1.812
		20%	β_2	0.654	1.531	1.635	3.828
			β_3	0.434	0.686	1.085	1.715
	60%		β_1	0.281	0.651	0.704	1.629
		30%	β_2	0.638	1.244	1.595	3.109
			β_3	0.584	0.462	1.460	1.150

Bias and RMSE are calculated to determine how close to the true value the estimates are calculated for each method. When Table 1 and Table 2 are examined, both methods obtained estimates near the parameter value for all censors and outlier ratios. Also, it was observed that as the outlier ratio is increased from 10% to 30%, the RMSE and BIAS values calculated in both methods generally tend to increase. Although the increase in the outliers ratio affected the methods, both methods made good estimates. However, the BIAS and RMSE values of the estimates obtained with the MI+ Cox method were consistently smaller than those obtained from the robust method for the coefficients of X_1 and X_2 which are covariates with outliers in both N = 50 and N = 100.

So, according to simulations, the multiple imputation method is a good alternative to robust methods in the presence of outliers.

In order to compare the proposed methods with a real data set, the NCCTG lung cancer data set from R library is used. There are 228 patients and 8 covariates in the data set; 61 observations were deleted due to missingness and as the result, N = 167 and number of events is 120 with a censored ratio of 28%. The prognostic factors in the data set are listed as follows: institution code (inst), age in years (age), sex, ECOG performance score (ph.ecog), Karnofsky performance score rated by physician (ph.karno), Karnofsky performance score as rated by patient (pat.

karno), calories consumed at meals (meal.cal), and weight loss in last six months (wt.loss) (Loprinzi *et al*, 1994).

Proportional hazard assumption is not provided for meal.cal which is one of the covariates in the data set. The aim of the study is to determine the method that gives the nearest estimation to the parameter estimation in the case where the assumption is achieved. Therefore, all variables in the data set that we will use as references

will have to provide the assumption. For this reason, the meal.cal covariate is not included in the model.

As first, a residual analysis for original data is carried out and no significant outliers were found but some of the observations have potential outliers in wt.loss variables. To obtain an outlier data set, extreme values were given to the values of these observations and a data set containing 10% outliers was created.
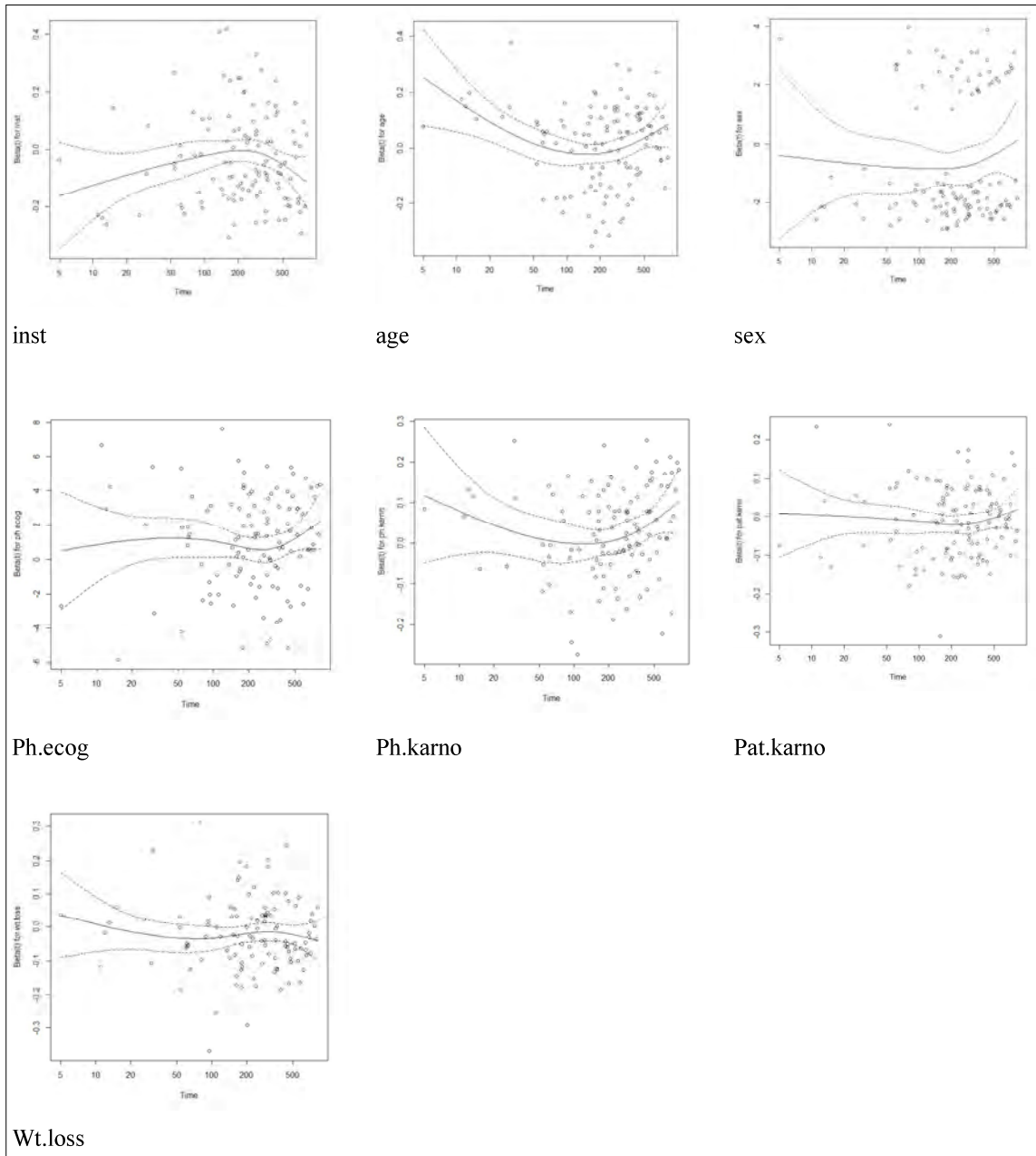

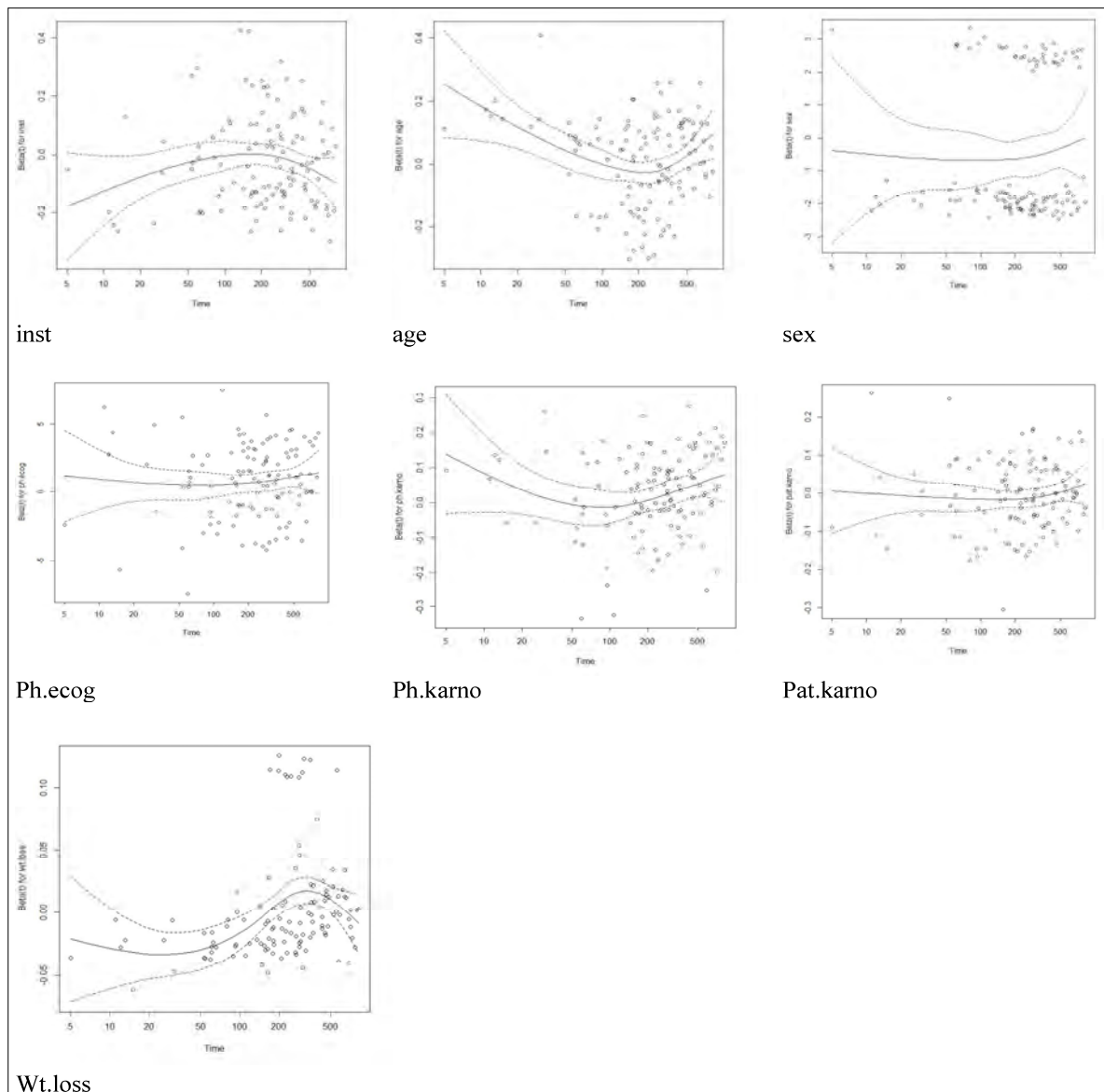**Figure 1:** Graphs of Schoenfeld residuals for the original data set

Testing of proportional hazard assumption for lung cancer data sets

The assumption test of the Cox regression analysis is performed using Schoenfeld residuals for the original data sets. The graphs of Schoenfeld residuals for each covariate of the original data set are given in Figure 1 and the statistical results of Schoenfeld residual analysis for the original data set are given Table 3.

Figure 1 shows that the residuals are randomly around a horizontal line for all of the covariates. In other words all variables provide proportional hazard assumption.

Table 3: Statistical results of Schoenfeld residual analysis for the original data set

	rho	chisq	p value
İnst	0.0298	0.125	0.723
Age	-0.0715	0.699	0.403
Sex	0.0613	0.416	0.519
Ph.ecog	0.0354	0.167	0.683
Ph.karno	0.0818	0.572	0.450
Pat.karno	0.0098	0.013	0.909
Wt.loss	-0.0283	0.114	0.736

inst

age

sex

Ph.ecog

Ph.karno

Pat.karno

Wt.loss

Figure 2: Graphs of Schoenfeld residuals for the data set with outlier

Table 3 contain the correlation between Schoenfeld residuals and transformed survival time, the test statistic which is given in equation (5) and the two sided p-value. From Table 3, we conclude that the same, proportional hazard assumption is provided for all variables (p > 0.05). Also the correlations are close to 0 so the assumption is provided. This shows that the original data set can be analyzed by Cox regression and the results obtained will be reliable.

The proportional hazard assumption has been tested after adding the outliers to the data set. Schoenfeld residual analysis graphs for each covariate are given in Figure 2. The statistical results of the residual analysis are also given in Table 4.

Table 4:   Statistical results of Schoenfeld residual analysis for data set with outliers

	rho	chisq	p value
İnst	0.0151	0.0334	0.855
Age	-0.134	2.5459	0.111
Sex	0.0524	0.3151	0.574
Ph.ecog	0.0588	0.463	0.496
Ph.karno	0.0868	0.732	0.392
Pat.karno	0.050	0.350	0.554
Wt.loss	0.2844	11.821	0.00058

Figure 2 shows that in case of outliers in the data set, the assumption of proportional hazard for the wt.loss covariate is violated due to some of the residuals being scattered in a different way. Also, according to the statistical results in Table 4, the wt.loss covariate has been seen as not providing the assumption (p < 0,05).

The existence of outliers is a problem for researchers because they affect the proportional hazard assumption. In order to solve this problem, we propose to consider the outlier values as missing data and a value is assigned for each missing value by the multiple imputation method so that 5 completed data sets are obtained. The assumption test was performed with Schoenfeld residual analysis for each of the data sets which is completed with the multiple imputation method (MI), and the result was obtained for all data sets. The statistical result of Schoenfeld analysis

for the first imputed data set is given in Table 5. Similar results were obtained for the other imputed data sets. That is, it was seen that the assumption provided for all imputed data sets.

According to Table 5, the proportional hazard assumption is satisfied for all variables (p > 0,05). That is, these results indicate that the problem of assumption violation caused by outliers can be solved by imputation.

Table 5:   Statistical results of Schoenfeld residual analysis for imputed data set by multiple imputation

	rho	chisq	p value
İnst	0.0236	0.0824	0.774
Age	-0.0946	1.2585	0.262
Sex	0.0773	0.6798	0.410
Ph.ecog	0.0258	0.0917	0.762
Ph.karno	0.0860	0.6854	0.408
Pat.karno	0.0434	0.2663	0.606
Wt.loss	0.0695	0.7826	0.376

Statistical inference of each method for lung cancer data

As there are no outliers in the data set during the first phase of the study then the parameters are estimated by applying Cox regression analysis. These results are compared as original results and used as reference for other results.

In the second phase of the application, the assumption test of the Cox regression analysis is performed using the data set containing the outlier values. As a result, the outliers violate the proportional hazard assumption, and in this case the results are unreliable. Robust estimation can used to overcome this problem. Therefore robust Cox regression analysis is applied to lung cancer data with outliers.

Also, the multiple imputation method overcomes the problem caused by outliers. In the third phase of the application, firstly the outlier values in the data set are deleted and the missing data set is obtained. A value is assigned for each missing value by the multiple imputation method, so that 5 completed data sets are obtained. Cox regression is separately applied to the 5 completed data sets and the combined results are calculated. All results are given in Table 6.

Table 6:   Parameter estimates

Parameters		Original			Results for data set with outliers.				
					Robust regression			Cox Regression with MI	
	β	SE	p value	β	SE	p value	β	SE	p value
inst.	-0.0304	0.013	0.0204	-0.0161	0.015	0.0285	-0.0302	0.013	0.0247
Age	0.0128	0.0117	0.2769	0.0065	0.0134	0.628	0.0134	0.012	0.2641
sex	-0.5669	0.2001	0.0046	-0.7041	0.2449	0.0041	-0.5699	0.201	0.0055
Ph.ecog	0.9073	0.2385	0.00014	0.7231	0.4397	0.0100	0.8571	0.239	0.001
Ph.karno	0.0266	0.0116	0.02214	0.0118	0.0202	0.5576	0.0234	0.0121	0.0480
Pat.karno	-0.0109	0.0080	0.173	-0.01272	0.0107	0.232	-0.0107	0.0079	0.1877
Wt.loss	-0.0167	0.0079	0.0346	-0.00306	0.0065	0.571	-0.0171	0.0083	0.0445

In order to compare the performances of robust Cox regression and Cox tegression with MI, Cox regression analysis results of the original data were used as a reference. Robust Cox regression analysis, which is recommended to be used safely in case of violation of the proportional hazard assumption, was applied to the data set containing 10% outliers. As a result, the analysis had similar results in terms of sign and magnitude to parameter estimations of the original data. Then, Cox regression with multiple imputation is applied to the outlier data set; the parameter estimates are very similar to the original estimates, as in the robust method. When these two methods were compared in terms of parameter estimates, the estimates of the Cox regression with MI method were found to be closer to the original estimates. Also, Cox regression with MI had a smaller standard error than robust Cox regression.

CONCLUSION

Outliers may lead to violation of the proportional hazard assumption, which is one of the most important assumptions of Cox regression. Therefore, outliers can have a great influence on parameter estimation. In such a case, the existence of outliers leads to incorrect results. For this reason, the outliers are a problem and there are different methods in the literature for solving this problem. Robust Cox regression is robust to outliers and can reduce the impact of outliers. Therefore, it is a recommended method as it gives safe results in the presence of outliers. The proposed method in this work is Cox regression with multiple imputation. In this method, the outliers are considered as missing data, and a value is assigned for each missing value by the multiple imputation method. Then the completed data

set is analyzed by Cox regression. In this study, robust Cox regression and the Cox regression with multiple imputation methods are evaluated and their results are compared with original results.

According to the results of the study carried out on simulated data sets and a real data set, the multiple imputation method, which is a missing data analysis method, solves the problem of outliers better than the robust estimation method, due to results closer to the original results being obtained.

Acknowledgement

REFERENCES

Alkan N. & Alkan B.B. (2018). A new approach for Cox regression analysis in the presence of outliers, *Süleyman Demirel University Journal of Natural and Applied Sciences* **22** (2): 637–643.
DOI: https://doi.10.19113/sdufbed.37782

Alkan N., Terzi Y., Cengiz M.A. & Alkan B.B. (2013). Comparison of missing data analysis methods in Cox proportional hazard models. *Turkiye Klinikleri Journal of Biostatistic* **5**(2): 49–54.

Alonso R. & Pardo M.C. (2020). Assessing influence on the estimated coefficients efficiency in a Cox regression. *Journal of Statistical Computation and Simulation* **90**(7): 1216–1229.
DOI: https://doi.org/10.1080/00949655.2020.1720986

Bednarski T. (1989). On sensitivity of Cox's estimators. *Statistics and Decisions* **7**: 215–228.

DOI: https://doi.org/10.1524/strm.1989.7.3.215

Bednarski T. (1993). Robust estimation in the Cox regression model. *Scandinavian Journal of Statistics* **20**:13–225.

Bretagnolle J. & Huber C. (1988). Effects of omitting covariates in Cox's model for survival data. *Scandinavian Journal of Statistics* **15**: 125–138.

Cox D.R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society* **34**: 187–220.
DOI: https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

Enders C.K. (2010). *Applied Missing Data Analysis,* pp.165-286. Guilford Pres, New York, USA.

Farcomeni A. & Viviani S. (2011), Robust estimation for the Cox regression model based on trimming. *Biometrical Journal* **53**(6): 956–73.
DOI: https://doi.org/10.1002/bimj.201100008

Grambsch P. & Therneau T. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrik*a **81**: 515–526.
DOI: https://doi.org/10.1093/biomet/81.3.515

Hosmer D.W. & Lemeshow S. (1999) *Applied Survival Analysis: Regression Modeling of Time to Event Data.* John Wiley & Sons, Inc., Canada.

Kalbfleisch J.D. & Prentice R. L. (1980). *The Statistical Analysis of Failure Time Data.* John Wiley and Sons, New York, USA..

Kleinbaum D.G. & Klein M. (1996). *Survival Analysis, A Self Learning Text*. Springer, USA.

Loprinzi C.L., Laurie J.A., Wieand H.S., Krook E., Novotny P.J., Kugler J.W., Bartel J., Law M., Bateman M. & Klatt N.E. (1994). Prospective evaluation of prognostic variables from patient-completed questionnaires. North Central Cancer Treatment Group. *Journal of Clinical Oncology* **12**(3): 601–607.
DOI: https://doi.org/10.1200/JCO.1994.12.3.601

Reid N. & Crepeau H. (1985). Influence function for proportional hazards regression. *Biometrika* **72**: 1–9.
DOI: https://doi.org/10.2307/2336329

Rubin D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley & Sons, New York, USA.

Schoenfeld D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika* **69**: 239–241.
DOI: https://doi.org/10.2307/2335876

Xue X. & Schifano E.D. (2017). Diagnostics for the Cox model. *Communications for Statistical Applications and Methods* **24**(6): pp. 583–604.
DOI: https://doi.org/10.29220/CSAM.2017.24.6.583

RESEARCH ARTICLE

Modeling and Simulation

Optimal fissile distribution in multiplying systems: Illustrative examples with Monte Carlo simulation and Pontryagin's maximum principle

H Khan[1]*, U Aziz[2], ZU Koreshi[2] and SR Sheikh[2]
[1] *National University of Computer and Emerging Sciences (NUCES-FAST), Islamabad, Pakistan.*
[2] *Air University, E-9, Islamabad, Pakistan.*

Abstract: In multiplying systems, such as nuclear reactors and criticality experiments, it is desirable to place the fissile material in the optimal or 'best' way to reduce the critical mass of the system as well as to achieve uniform fuel burnup. This paper considers two methods, namely Pontryagin's maximum principle (PMP) and Monte Carlo (MC) perturbation for estimating a minimum critical mass configuration. These methods are applied to an elementary multizone model of a pressurized water reactor (PWR) and a criticality experiment to estimate the minimum critical mass. It is found that while two-group diffusion theory with PMP predicts a minimum critical mass, more detailed MC simulations with MCNP5 show a consistent reduction in critical mass when fissile fuel is placed in inner zones. Such a distribution reduces the fissile material requirement but is undesirable due to the higher power peaking. MC simulations show that for a three-zone model of the KORI 1 PWR, a uniform fissile distribution gives criticality for 1.09 atomic percent (at.%) enrichment, whereas non-uniform fissile distribution (0.6, 1.6, 0.6 at.%) reduces the critical mass by 14%. The changes found from MC simulations were subsequently predicted from first- and second-order derivative sampling. It was found that substantial computational savings can be achieved for large-scale optimization problems. In the case of a criticality experiment, MC derivative sampling was also used to estimate optimal fissile distribution for minimizing the critical mass.

Keywords: Derivative sampling, Goertzel's distribution, MCNP, minimum critical mass, Pontryagin maximum principle.

INTRODUCTION

The concept of minimum critical mass (MCM) (Goertzel, 1956) has been a challenging and interesting research topic over the last few decades, beginning with Goertzel's two-group diffusion theory model, and including work by Wilkins (Wilkins Jr. & Srivastava, 1982), Williams (Williams, 2003; 2017) and others based on the transport model.

Critical mass refers to the minimum amount of fissile material required for a self-sustaining nuclear chain reaction. It is a fundamental concept in nuclear physics that ensures a stable and controlled reaction. The concept was systematically explored for the first time during the Manhattan Project (Gosling, 1999). Enrico Fermi and his team explored the relation between the amount of fissile material and the rate of fission reactions (Segrè, 1970). The construction of the first controlled experimental nuclear reactor Chicago Pile-1, in 1942, demonstrated the significance of both the placement and quantity of fissile material for achieving a self-sustainable nuclear chain reaction (Rowinski *et al.*, 2015).

This concept gained further validation through subsequent nuclear tests, most notably the Trinity test in 1945 (Szasz, 1984). The understanding of critical mass is not only crucial for the development but also for the

safe operation of nuclear reactors. Advancements in nuclear physics have refined the concept, incorporating considerations such as reflectors and control rods. This enhanced understanding has led to the concept of MCM, which reduces the required amount of fissile material even further for a self-sustaining nuclear chain reaction.

Goertzel concluded that a necessary condition for MCM in a reflected reactor is a flat thermal flux while literature shows that the necessary condition is that the integral of the product of the thermal neutron flux and a function related to the adjoint flux is a constant. The thermal flux flattening has also been shown for diffusion theory (Lewins, 2004) for heterogeneous systems of thin fuel 'foils' in a moderator (Dam, 2013; 2015). The basic purpose of non-uniformly distributing the fuel is to achieve criticality with a minimum amount of fissile material. For a nuclear power reactor, the placement of fuel rods in an optimal pattern is both an engineering and economic objective for achieving efficient fuel utilization and scheduled fuel reloading. A power reactor such as the AP1000, for example, has 157 fuel assemblies with a 17 × 17 rod array containing 41,448 fuel rods of three enrichments. To place these optimally is a large-scale optimization challenge that requires efficient methods and computational techniques. Thus, theoretical works which are mostly based on models with assumptions and artifacts are of great value; hence the motivation for this work. An idealized model of a PWR is considered for an optimization study using Monte Carlo (MC) simulation with the MCNP5 (Briesmeister, 2003) code. Apart from its use as a production code with detailed physics modeling, 'continuous' cross-section data libraries, and combinatorial 3D geometry modeling, the perturbation capability is particularly useful for sensitivity analyses and design optimization.

This work uses derivative sampling for illustrative cases to predict optimality from a 'reference' design; computational efficiency depends on how 'good' a reference is. To this end, a two-group diffusion model with the Pontryagin maximum principle (PMP) (Lee, 1973; Koreshi *et al.*, 2019) variational formulation can be used.

This paper considers two applications, namely a multi-zone bare cylindrical reactor, and a criticality experiment to obtain 'optimal' configurations with MC derivative sampling leading to minimization of critical mass. These outcomes are discussed in connection with Goertzel's result that more fissile material should be added to the central regions and progressively less material towards the end of the core region.

For a three-zone bare reactor, a two-group PMP formulation shows (Lee, 1973) an optimal configuration associated with minimum critical mass in a min-max-min (enrichment) fissile distribution. Such a configuration can also be predicted using sensitivity coefficients from MC derivative sampling based on the importance (Otsuka, 2017) of a zone to decide whether fissile material should be added or removed. This strategy is used in this paper to estimate the optimal fissile distribution for a criticality experiment by Clark (Clark, 1966) and also to demonstrate MCM in a U^{235}-H_2O system.

METHODOLOGY

The formulations considered for estimating MCM are two-group diffusion theory with PMP, and MC simulation with derivative sampling.

Notations used in the paper are defined in the Annex 1

Two-group diffusion model

In the two-group diffusion equations for a bare multiplying system (Lamarsh & Baratta, 2001)

$$D_1 \nabla^2 \phi_1 - \Sigma_1 \phi_1 + \frac{k_\infty}{p} \Sigma_2 \phi_2 = 0 \qquad ...(01)$$

and

$$D_2 \nabla^2 \phi_2 - \Sigma_2 \phi_2 + p \Sigma_1 \phi_1 = 0, \qquad ...(02)$$

for a uniformly distributed fuel, the coefficients in equations. 1 and 2 are constant. These equations can be solved to obtain exact solutions by applying the usual boundary conditions at the center, interfaces, and outer boundary for a multizone cylindrical reactor.

The optimal fissile nuclide distribution can be estimated from a variational formulation (Lee, 1973; Koreshi *et al.*, 2019; Saracco *et al.*, 2019), based on the above two-group diffusion equations. For a bare multizone reactor equations (1) and (2) are written with $D_1, D_2, \Sigma_1, \Sigma_2$ constant in each zone. The thermal absorption cross-section is expressed as a 'control' parameter u where $\Sigma_{2c} \equiv u = \gamma u' + \delta$, where u' is the atomic enrichment $u' = \frac{N^{235}}{N^U}$. The total number of uranium atoms $N^U = N^{235} + N^{238}$, is kept constant, while the enrichment is varied. Here $\gamma = N^U \sigma_{2F}^5$, and $\delta = \Sigma_{2,others}$ where 'others' represent the moderator, structural material, zirconium, etc.

The optimization problem is then formulated as:

$$\text{minimize } \tilde{J} = \int u(r)\, dV \qquad \qquad \ldots(03)$$

$$\text{subject to } N^U = constant$$

In the variational formulation (Lee, 1973), equations (1) and (2) are written as four first-order state-space equations for variables y_i, where $i = 1,2,3,4$, and with a performance index $\tilde{J}$, given as equation (3), defined as a function with 'control' u representing fissile mass in volume V, to be minimized.

The 'augmented' Hamiltonian can then be written as

$$H = \sum_{i=0}^{5} \varphi_i f_i \left(\underline{y}, u, x \right) \qquad \qquad \ldots(04)$$

where φ_i are adjoint functions and φ_0, f_0 represent $\tilde{J}$. From the stationarity conditions on H, coupled ODEs are obtained for λ_i and f_i and solved with the transversality conditions. Pontryagin's Maximum Principle, with admissible controls $u = u_m, u_M$ (minimum and maximum values) is subsequently expressed in the form

$$H = \psi_1(u, x) + \psi_2(x). \qquad \qquad \ldots(05)$$

Thus, in a bare multizone reactor, u is selected such that H is maximized and the shape of the 'switching function' $\psi_1(u, x)$, obtained from λ_i and f_i, determines the controls. Such an analysis leads to critical configurations in a multizone analysis. The MCM, for this 'restrictive criticality configuration', is subsequently found for (u_M, u_m) fuel enrichments in the inner and outer zones of a two-zone core while a (u_m, u_M, u_m) configuration is

found in a three-zone core.

In a single zone, there is no optimal arrangement, and the criticality condition is given as

$$\left| J_0(\mu R)\, I_0(\lambda R)\, S_1 J_0(\mu R)\, S_2 I_0(\lambda R) \right| = 0 \qquad \ldots(06)$$

where J_0 is zero-order Bessel function and I_0 is modified zero-order Bessel functions, μ^2 and λ^2 are constants given as equations (7) and (8) and $\tau_1 = \frac{D_1}{\Sigma_1}$, $\tau_2 = \frac{p\Sigma_1}{D_2}$, $\alpha = \frac{k_\infty}{pD_1}$, $\beta = 1/D_2$.

$$\mu^2 = \frac{1}{2}\left[-\left(\beta u + \frac{1}{\tau_1} \right) + \sqrt{ \left(\beta u + \frac{1}{\tau_1} \right)^2 + 4u\left(\frac{\alpha}{\tau_2} - \frac{\beta}{\tau_1} \right) } \right]$$

$$\ldots(07)$$

$$\lambda^2 = \frac{1}{2}\left[+\left(\beta u + \frac{1}{\tau_1} \right) + \sqrt{ \left(\beta u + \frac{1}{\tau_1} \right)^2 + 4u\left(\frac{\alpha}{\tau_2} - \frac{\beta}{\tau_1} \right) } \right]$$

$$\ldots(08)$$

$S_{1,2}$ are the coupling coefficients $S_1 = \tau_2(\mu^2 + \beta u)$, $S_2 = -\tau_2(\lambda^2 - \beta u)$.

For a two-zone bare cylindrical reactor, the MCM arrangement requires (u_M, u_m) configuration. In a three-zone reactor, with regions I, II, and III for $0 \le r \le r_1$, $r_1 \le r \le r_2$ and $r_2 \le r \le R$, a criticality configuration is obtained by finding values of minimum enrichment u_m and maximum enrichment u_M in the configuration u_m, u_M, u_m from the determinant of the 10×10 matrix given in equation 9. The three-zone matrix is given as

$$
\begin{bmatrix}
S_{11}J_0(\mu_1 r_1) & -S_{21}J_0(\mu_2 r_1) & 0 & -S_{22}Y_0(\mu_2 r_1) & 0 & S_{13}I_0(\lambda_1 r_1) & -S_{23}I_0(\lambda_1 r_1) & 0 & -S_{24}K_0(\lambda_2 r_1) & 0 \\
-S_{11}\mu_1 J_1(\mu_1 r_1) & S_{21}\mu_2 J_0(\mu_2 r_1) & 0 & S_{22}\mu_2 Y_1(\mu_2 r_1) & 0 & S_{13}\lambda_1 I_1(\lambda_1 r_1) & 0 & 0 & -S_{23}\lambda_1 I_1(\lambda_1 r_1) & 0 \\
0 & S_{21}J_0(\mu_2 r_2) & -S_{31}J_0(\mu_1 r_2) & -S_{22}Y_0(\mu_2 r_2) & -S_{32}Y_0(\mu_1 r_2) & 0 & S_{23}I_0(\lambda_2 r_2) & -S_{33}I_0(\lambda_1 r_2) & S_{24}K_0(\lambda_2 r_2) & -S_{34}K_0(\lambda_1 r_2) \\
0 & -S_{21}\mu_2 J_1(\mu_2 r_2) & S_{31}\mu_1 J_1(\mu_1 r_2) & -S_{22}\mu_2 Y_1(\mu_2 r_2) & S_{32}\mu_1 Y_1(\mu_1 r_2) & 0 & S_{23}\lambda_2 I_1(\lambda_2 r_2) & -S_{33}\lambda_1 I_1(\lambda_1 r_2) & -S_{24}\lambda_2 K_1(\lambda_2 r_2) & S_{34}\lambda_1 K_1(\lambda_1 r_2) \\
J_0(\mu_1 r_1) & -J_0(\mu_2 r_1) & 0 & -J_0(\mu_2 r_1) & 0 & I_0(\lambda_1 r_1) & -I_0(\lambda_2 r_1) & 0 & -K_0(\lambda_2 r_1) & 0 \\
-\mu_1 J_1(\mu_1 r_1) & \mu_2 J_1(\mu_2 r_1) & 0 & \mu_2 Y_1(\mu_2 r_1) & 0 & \lambda_1 I_1(\lambda_1 r_1) & -\lambda_2 I_1(\lambda_2 r_1) & 0 & S_{24}\lambda_2 K_1(\lambda_2 r_1) & 0 \\
0 & J_0(\mu_2 r_2) & -J_0(\mu_1 r_2) & -Y_0(\mu_2 r_2) & -Y_0(\mu_1 r_2) & 0 & I_0(\lambda_2 r_2) & -I_0(\lambda_1 r_2) & K_0(\lambda_2 r_2) & -K_0(\lambda_1 r_2) \\
0 & -\mu_2 J_1(\mu_1 r_2) & \mu_1 J_1(\mu_1 r_2) & -\mu_2 Y_1(\mu_2 r_2) & \mu_1 Y_1(\mu_1 r_2) & 0 & \lambda_2 I_1(\lambda_2 r_2) & -\lambda_1 I_1(\lambda_1 r_2) & -\lambda_2 K_1(\lambda_2 r_2) & \lambda_1 K_1(\lambda_1 r_2) \\
0 & 0 & S_{31}J_0(\mu_1 R) & 0 & S_{32}Y_0(\mu_1 R) & 0 & 0 & S_{33}I_0(\lambda_1 R) & 0 & S_{34}K_0(\lambda_1 R) \\
0 & 0 & J_0(\mu_1 R) & 0 & Y_0(\mu_1 R) & 0 & 0 & I_0(\lambda_1 R) & 0 & K_0(\lambda_1 R)
\end{bmatrix}
$$

$$\ldots(9)$$

where Y_0 is zero-order Bessel function and K_0 is a modified zero-order Bessel function. S_{ij} are the coupling coefficients for zone i and index j with $S_{11} = S_{21} = S_{22} = S_1, S_{13} = S_{23} = S_{24} = S_2$.

The condition for the determinant is Cramer's rule for the solution of a system of linear homogeneous equations. The equations arise out of four continuity conditions for flux and current at the interfaces r_1, r_2 and two zero-flux continuity conditions at the outer boundary R.

The homogeneous system of linear equations is thus solved for pairs of minimum and maximum enrichment values that correspond to criticality; one of these pairs $\underline{u}^* = (u_m, u_M, u_m)^*$. The condition that inner and outer zones have the same enrichment is a restriction imposed here which makes this criticality analysis different from that of Goertzel for which a continuous distribution is obtained. The optimal distribution is thus piece-wise continuous with fissile atomic densities contained in the 'optimal' control $\underline{u}^*$.

MC Simulation with derivative sampling

The optimal distribution $\underline{u}^*$ can be found using MC simulation by sampling derivatives (Rief, 1984; Koreshi & Lewins, 1990) in a 'reference configuration'; computational speed-up can be achieved by estimating the reference from a diffusion-based variational described in the preceding section.

Derivatives in an MC simulation are 'scored' during a 'history' in analog simulation from 'birth' to 'death'; the scores are based on the integral form of the transport equation. The collision density $\chi(P)$ in phase space $P(r, E, \hat{\Omega}, t)$ is given as

$$\chi(P) = S(P) + \int K\left(P' \to P\right) \chi\left(P'\right) dP' \qquad ...(10)$$

where $S(P)$ is the 'direct source' in P, and $K(P' \to P)$ is the transition kernel describing a neutron transition from $P' \to P$. This kernel $K(P' \to P)$ is expressed as the product of a 'collision' part, and a 'transport' part

$$K = \hat{C} \cdot \hat{T} \qquad ...(11)$$

so that the change in quantity such as k_{eff} can be estimated due to a perturbation in a parameter p by a Taylor series expressed as

$$k_{eff}(p) = k_{eff}(p_o) + \frac{dk_{eff}}{dp}\big|_{p_o}\delta p + \frac{1}{2!}\frac{d^2 k_{eff}}{dp^2}\big|_{p_o}(\delta p)^2 + \cdots \qquad ...(12)$$

Optimization using MC derivatives

The MC derivatives sampled for a zone can be interpreted as 'sensitivity' coefficients for that zone. Thus, the material is preferably placed in a zone of high importance and removed from a zone of low importance. This can lead to a net reduction in total fissile material in a critical system. The optimal distribution $\underline{u}^*$ can then be found by an optimization search to yield the MCM.

The optimization problem is thus expressed as

$$\text{minimize } \tilde{J} = m_{ref} + \sum_{i=1}^{N} \Delta m_i \qquad ...(13)$$

$$\text{subject to } \sum_{i=1}^{N} \Delta m_i \frac{\partial k_{eff}}{\partial \Delta m_i} = 0 \qquad ...(14)$$

where $m_{ref}, and \, \Delta m_i$ are the fuel mass in the reference design, and the mass added or removed from zone i respectively.

RESULTS AND DISCUSSION

Bare cylindrical reactor

For the decommissioned Korean (Westinghouse) KORI-1 Reactor (Clark, 1966), a bare cylindrical U-H$_2$O homogeneous model is considered of radius 122.6 cm, height of 365.8 cm with a fixed number of U atoms $N^U = N^{235} + N^{238} = 6.994 \times 10^{21}$ atoms cm^{-3}, the number of oxygen atoms in uranium $N_{Oxygen}^U = 1.399 \times 10^{22}$ atoms cm^{-3}, $and \, N_w = 1.947 \times 10^{22}$ molecules cm^{-3}. The two-group data are $D_1 = 1.367 \, cm$, $D_2 = 0.2294 \, cm$, $\Sigma_{1s} = 0.03922 \, cm^{-1}$, and $\Sigma_{2c}^{others} = 0.028058 \, cm^{-1}$.

The mass of U^{235} (in grams for volume V in cm^3) in terms of the enrichment u is

$$M_5 = 235.04 \frac{u N_U V}{N_{av}} \sim uV \qquad ...(15)$$

which is proportional to u for equal volume zones.

The one-zone two-group criticality determinant (equation 6) gives criticality at 1.10% enrichment compared with MCNP5 simulations, for 1000 neutrons simulated for 1000 cycles with 10 skip cycles, shown in Figure 1 giving $k_{eff} = 1.00029 \, (0.00052)$ at 1.09% (513.96 kg U^{235}) and 1.00422 (0.00053) at 1.1% (518.68 kg U^{235}) and associated values of fissile mass. The cross-sections were taken from the ENDF66a and ENDF66c libraries of MCNP5.

Table 1 lists the criticality pairs and associated relative critical masses $m_{r.}$ defined as $(u_M + 2u_m)/(u_M + 2u_m)_o$ where the subscript o denotes the critical pair at 1.09 at.%, obtained from MC simulations, for a three-zone bare critical reactor (radii 70.7831 cm, 100.1025 cm, 122.6 cm, height 365.8 cm).

Table 1: Criticality pairs and critical mass

u_M (%)	1.09	1.15	1.20	1.30	1.40	1.50	1.60
$\underline{u_m}$ (%)	1.090	1.0665	1.050	0.981	0.877	0.75	0.600
k_{eff}	1.00029 (0.00052)	1.00006 (0.00053)	1.00092 (0.00055)	1.00002 (0.00053)	1.00090 (0.00055)	1.00050 (0.00055)	1.00039 (0.00058)
m_r	1.00	1.00	1.01	1.00	0.96	0.92	0.86
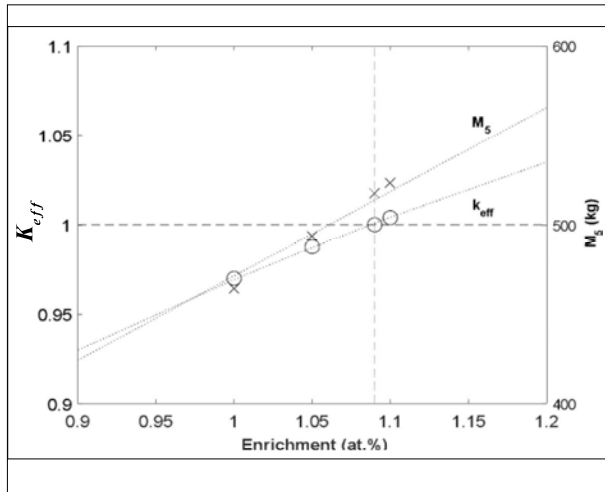
Figure 1: Effective multiplication k_{eff} and fissile mass in the KORI-1 cylindrical reactor
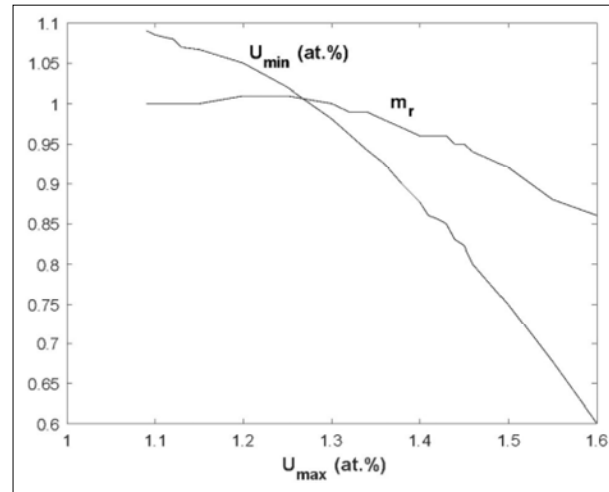
Figure 2: Critical pairs and relative critical mass

The results show that the relative critical mass remains steady till u_M is increased from 1.09 to 1.25 ($u_m = 1.02$) then falls to 0.86 for the min-max pair (0.60, 1.60) as shown in Figure 2. As u_M is increased to 1.8%, u_m decreases to 0.33% with m_r falling to 0.75. The critical mass can thus be reduced by 25% in this configuration. Beyond this, MC simulations give a critical configuration of $u_M = 1.85\%$ in the central zone surrounded by water in the inner and outer zones.

In the above simulations, the enrichments in the first and third zones were kept equal to compare with the 10 ×10 critical determinant obtained by Lee in the PMP formulation. This was shown by Lee to result in 'switching functions' requiring a u_m, u_M, u_m configuration for minimizing critical mass in a three-zone reactor.

As stated before, LWRs typically have fuels of three levels of enrichment, with the lowest enrichment fuel in the innermost zone followed by the next higher enrichment fuel and the highest enrichment in the outer zone. The Westinghouse KORI reactor, modeled by Lee in the preliminary PMP formulation, also had fuels of three levels of enrichment (2.10, 2.83, and 3.20 weight %) so that the outer zone does not have the same enrichment as the innermost zone. The Westinghouse AP1000 has three fuels of enrichment 2.35, 3.40, and 4.45 weight % with an initial core loading consisting of interspersed lower enrichment fuels and the outer periphery containing the highest enriched fuel. Such a distribution gives a favorable radial power distribution.

The MCNP simulations for the critical pairs reported by Lee using two-group diffusion theory give the following results: for the pairs (0.9, 0.89), (1.1, 0.71), (1.2, 0.65), (1.3, 0.61), (1.40, 0.58), and (1.50, 0.56) the estimates are $k_{eff} = 0.92734$ (0.00049), 0.92043 (0.00052), 0.93317 (0.00052), 0.94748 (0.00054), 0.96205 (0.00055), and 0.978978 (0.00056), respectively. By including structural materials, these estimates are likely to decrease. Thus,
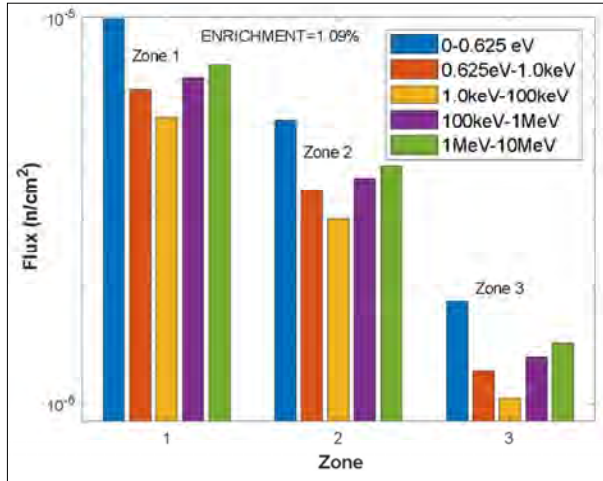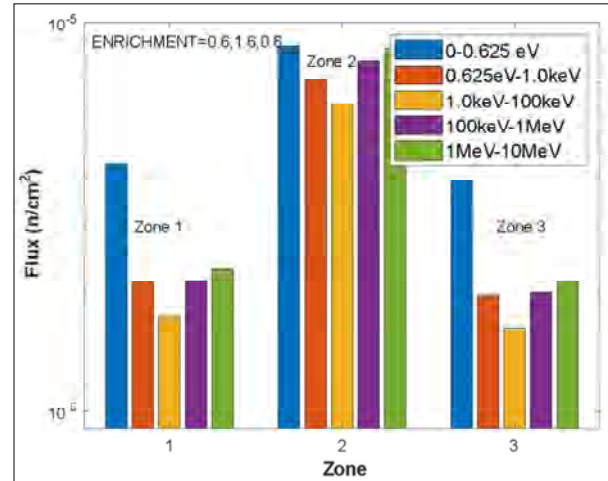
Figure 3: Group fluxes in for uniform fissile (1.09 atomic %) distribution

Figure 4: Group fluxes in for 0.6, 1.6, 0.6 atomic % distribution

the present MC simulations differ significantly from the two-group diffusion results of Lee. The relative critical masses for these pairs (1.00, 0.94, 0.93, 0.94, 0.96, and 0.98, respectively) led Lee to report a minimum critical mass.

We next consider the two pairs in Table 1 which give a 14% decrease in critical mass to see what effect they have on the relative fluxes in the zones. These results are shown in Figure 3 and Figure 4.

As seen from Figure 3 and Figure 4, in reducing the critical mass by increasing the fissile fuel in the central zone, the fluxes show a decrease in the inner zone and an increase in the central and last zones; the outer zone is now almost as 'good' as the inner zone. The problem now, clearly, is the increased peaking in the central zone. The ratios of thermal flux in the first zone to the second and third zones in Figure 3, 1.83 and 5.38, respectively, have now changed to 0.50 and 1.11. An intuitive remedy is the reduction of the central zone which will bring the flux down favorably but would need to be compensated by additional fuel in the outer zone. The decision of how much to remove from one zone and add in the other could be made based on the zone 'importances'.

From the MCNP5 neutron activity table for the results of Figure 4, we see that about 60% of the collisions take place in the central zone and 20% each in the inner and outer zones; the activity is thus three times enhanced in the central zone. Both the number-weighted energy and the flux-weighted energy are higher in the central zone

as well. The aim should thus be to reduce the central zone activity by a third and compensate for criticality by additions in the inner and outer zones.

By now it is well established that the fuel should be distributed such that the product of the thermal flux and the adjoint thermal flux should be a constant. Of course, that applies to a core-reflector configuration for which the constant in the core should not exceed that constant in the reflector. In our case, we have a bare homogenous reactor instead of a reflected reactor.

We now take the help of MC derivative sampling to estimate the sensitivity of the central zone. The MC derivatives were sampled with the following commands:

PERT1:N CELL=2 MAT=1090 to estimate the derivatives and changes in reducing the central zone enrichment from 1.6% to 1.09%, and

PERT2:N CELL=1,3 MAT=1090 to estimate the derivatives and changes in increasing the first and third zone enrichments from 0.6 to 1.09%. The identifier MAT=1090 denotes material with an enrichment of 1.09%.

The estimates for PERT1 and PERT2, for the predicted change in k_{eff} (track length flux estimator) are $-0.08181\,(0.00005)$ and $0.06901\,(0.00003)$ respectively for the MC simulation with 5000 neutrons sampled in 5000 cycles with 10 skip cycles. These add up to give a predicted change of about -0.01 whereas the

total is zero since both are critical configurations. The estimates are nonetheless useful as they are two 'stand-alone' estimates while the change is simultaneous. The simulation also gives predicted estimates in the track length scalar fluxes in each energy bin shown in Figure 3 and Figure 4. Thus, the full range of critical configurations can be predicted with reasonable accuracy.

Optimization

Derivatives estimated in a 'reference' MC simulation are used to estimate the amount of fissile material to be added or removed from each zone. For perturbations in a three-zone reactor with 172.893 kg U^{235} in each zone, the actual and predicted changes, with relative standard error (RSE) for 1000 neutrons per cycle, 1000 cycles are shown in Table 2 for method 1 which includes first and second derivatives, and method 2 which uses only first derivatives.

Table 2: Predicted and actual Δk_{eff} vs Δu (RSE < 0.00005)

Δu	Predicted Δk_{eff} Method 1 Method 2	Re-run Δk_{eff}	k_{eff}
	Zone 1		
0.02	0.00366 0.00370	0.00693	1.00737 (0.00054)
0.05	0.00900 0.00925	0.01345	1.01389 (0.00052)
0.10	0.01749 0.01850	0.02504	1.02548 (0.00055)
	Zone 2		
0.02	0.00206 0.00208	0.00222	1.00266 (0.00053)
0.05	0.00506 0.00520	0.00553	1.00597 (0.00053)
0.10	0.00986 0.01040	0.0104	1.01084 (0.00053)
	Zone 3		
0.02	0.00072 0.00073	0.0007	1.00114 (0.00054)
0.05	0.00177 0.00182	0.00197	1.00241 (0.00241)
0.10	0.00345 0.00363	0.00348	1.00392 (0.00054)

The reference design in this section has enrichment u = 1.1%, for which k_{eff} = 1.00044 (0.00052). The perturbations are for fissile atomic densities N_5 = 7.6934 × 10^{19}, 7.8333 × 10^{19}, 8.0431 × 10^{19}, and 8.3928 × 10^{19} atoms cm^{-3} ($\frac{\delta N_5}{N_5}$ = 1.8%, 4.5%, 9.1%). The 'specific' change in k_{eff} for mass addition and mass removal, as well as the relative importance of each zone (relative to zone 1) are shown in Table 3.

Table 3: Δk_{eff} per unit mass U^{235} change

Zone	Mass addition		Mass removal	
	$\frac{\Delta k_{eff}}{\Delta m_i}$ (10^{-3} kg^{-1})	Importance	$\frac{\Delta k_{eff}}{\Delta m_i}$ (10^{-3} kg^{-1})	Importance
1	1.2163	1.0000	-1.2923	1.0000
2	0.5959	0.4899	-0.6692	0.5178
3	0.1874	0.1541	-0.2094	0.1620

Table 4: Optimal distribution from MC & GA

Zone	1	2	3	Total (kg)
x_i (kg)	14.755	1.306	-99.917	-83.856
Estimated optimal	187.655	174.206	72.983	434.844
Enrichment %	1.1939	1.1084	0.4643	-

From Table. 3, a random search was carried out using genetic algorithms (GA) and optimal values of Δm_i (for zone i) are obtained; results for mass vs. number of iterations (N) are shown in Table 4, Figure 5, and Figure 6. The convergence was obtained within 1000 iterations.

For mass changes in zone i specified as $0 \leq \Delta m_1, \Delta m_2 \leq 100$, and $-100 \leq \Delta m_3 \leq 100$, the MCM for the bare reactor is found as 434.844 kg with the maximum in the central zone and a progressive decrease toward the outer zone. Based on the zone importance alone, the mass additions x_i are 14.755 kg and 1.306 kg in zones 1 and 2 respectively, and mass removal 99.917 kg from zone 3 with a net savings of 83.856 kg (~16%) from the critical mass corresponding to uniform composition.

The convergence towards optimal MCM is shown in Figure 5 for the reactor and in Figure 6 for each of the zones.

For the distribution 1.1939%, 1.1%, and 0.5%, MCNP simulation gives $k_{eff} = 1.02165\ (0.00080)$ for a total fissile mass of 439.13 kg. System multiplication is estimated from re-runs with the following results:

$k_{eff} = 1.00884\ (0.00088)$ at enrichment 1.15%, 1%, 0.5%, and at enrichments 1.12%, 1% and 0.5%. Thus, a material composition change is accurately predicted for a single region perturbation.
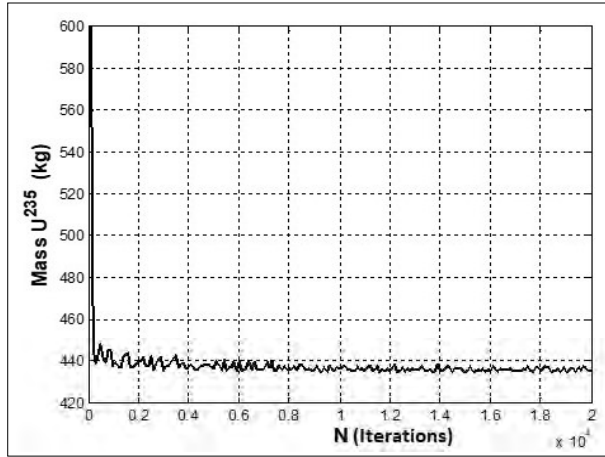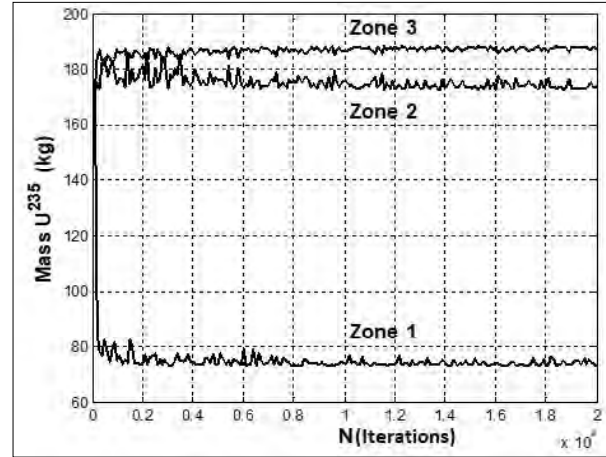

Figure 5: Minimum critical mass in a bare cylindrical reactor

Figure 6: Optimal fissile distribution in a three-zone bare cylindrical reactor

Criticality experiment

We now consider experimental results for a U^{235}-H_2O multi-region criticality assembly (Clark, 1966). In the experiment, a cylindrical configuration with five regions separated by thin aluminium partitions of dimensions listed in Table 5, was surrounded by a 'thick' reflector laterally and no reflector at the top or bottom.

Table 5: Criticality experiment for U^{235}-H_2O

Region	I	II	III	IV	V
Inner radius (cm)	0	5.54	8.14	10.27	12.55
Outer radius (cm)	5.48	8.08	10.21	12.49	15.10

Table 6: Accuracy of perturbation estimates for concentration change

No.	Concentration (gU^{235}/L)	% change in k (Predicted)	σ_k	% change in k (Actual)	σ_k
1	43.769	2.7740	0.00006	2.699	0.00084
2	35.811	-3.1946	0.00007	-2.974	0.00078)

Table 7: Accuracy of perturbation estimates for mass addition

Zone	Concentration (gU^{25}/L)	% change in k (Predicted)	σ_k	% change in k (Actual)	σ_k
1	43.769	0.9927	0.00005	1.502	0.00082
2	43.769	0.8563	0.00003	0.9350	0.00077
3	43.769	0.9987	0.00003	0.6612	0.00079

Table 8: Predicted and actual changes Δk_{eff} for fissile mass addition by zone

Zone	Volume (L)	Mass (kg)	MU^5 (g)	M U^5 added (g)	Δk_{eff} Re-run and Predicted		
					Re-run	1st-order	1st & 2nd-order
1	4.02846	4.19644	160.29	16.027	0.00821	0.00458	0.00501
2	4.64075	4.83427	184.66	18.462	0.00686	0.00476	0.00459
3	5.09547	5.30795	202.75	20.272	0.00660	0.00481	0.00465
4	6.77802	7.06066	269.70	26.965	0.00628	0.00607	0.00587
5	9.45830	9.85271	376.35	37.628	0.00591	0.01006	0.00963
			1193.75	119.354	0.03386	0.03029	0.02817

The MC simulations for the criticality experiment with uniform concentration 39.79 gU^{235}/L estimated k_{eff} as 1.00637 (0.00080) while the effect of ±10% changes in concentration (g U^{235}/L) was predicted as shown in Table 6.

To estimate the reliability of the derivatives in a multi-region configuration, a simplified three-zone model of the criticality experiment is considered; the second-order derivatives were estimated for a 10% increase in concentration one zone at a time, for which the predictions were acceptable for the first two zones, as shown in Table 7. The third zone, however, gave an unreliable prediction for k_{eff} of ~1% for an actual change of ~0.7%.

Similar trends were observed in the full five-region MC simulation for the experiment, as shown in Table 8.

The reference concentration in each zone was 39.79 gU^{235}/L with criticality at 1193.75 g U^{235} distributed as shown above. With a 10% mass addition, one zone at a time, Δk_{eff} from re-runs was found to vary from 0.00821 in the first zone, decreasing to 0.00591 in the last zone. The predictions by 1st- and 2nd–order perturbation are at variance with the actuals and thus need to be corrected before they are used to estimate the optimal distribution. It is important to note that the overall prediction of a 10% mass increase is ~2.8% compared with an actual change of about ~3.4%. An estimate of the accuracy of prediction is the difference between the sum effect of individual changes in zones and the overall change in the system. The former gives $\Delta k_{eff} = \sum_{i=1}^{5} \Delta k_{eff_i} = +0.03426$ while an overall change gives 0.02818. There is a difference of ~18% which is the best accuracy that can be predicted by perturbation. The specific changes $\frac{\Delta k_{eff}}{\Delta m_i}$ are shown in Table 9 for mass additions and removals from individual zones.

Table 9: Δk_{eff} per unit mass U^{235} change

Zone	Mass addition		Mass removal	
	$\frac{\Delta k_{eff}}{\Delta m_i}$ (kg^{-1})	Importance	$\frac{\Delta k_{eff}}{\Delta m_i}$ (kg^{-1})	Importance
1	0.51851	1.0000	-0.37864	1.0000
2	0.37698	0.7271	-0.32380	0.85518
3	0.33051	0.6374	-0.28998	0.76584
4	0.23660	0.4563	-0.17758	0.46901
5	0.15706	0.3029	-0.13948	0.36837

Once the 'importance' of each zone is determined, the fissile material is added to the most important zone and others are subsequently filled. As this is done, the system multiplication k_{eff} will increase; thus, to bring the system back to criticality, the material can be removed from less important zones. The overall fissile mass in the system will be reduced if more material is removed from the less important zones than is added in the more important zones. Using the specific values $\frac{\Delta k_{eff}}{\Delta m_i}$ from Table 9, the optimum distribution results in a change of k_{eff} from $k_{eff_u} = 0.9999$ to $k_{eff_o} = 1.0100$ which is an error of ~1%. Thus, the re-runs give a reliable prediction of the optimal configuration. The optimal distribution is obtained using the derivatives from Table 9. For these, the uniform and optimal k_{eff} values of 0.9999 and 1.0101, with the net mass change being 127.99 g U^{235}, are estimated from $\sum_{i=1}^{5} \Delta k_{eff_i}$ for experimental optimal mass additions/removals from each zone. Thus, atomic densities for U^{235} should be in the ratios of zone importance, *i.e.*, cell importance normalized to zone 3 are 1.60, 1.14, 1, 0.72, and 0.47; thus the importance provides a good reference point for MC derivative sampling and

Table 10: Optimal distribution from MC & GA

Zone	1	2	3	4	5	Total (g)
x_i	100.004	60.012	5.011	-79.99	-230.043	-145.006
Est. Opt.	260.29	244.672	207.761	189.71	146.307	1048.74
Exp. Opt.	269.1	249.67	210.44	193.17	143.39	1065.77
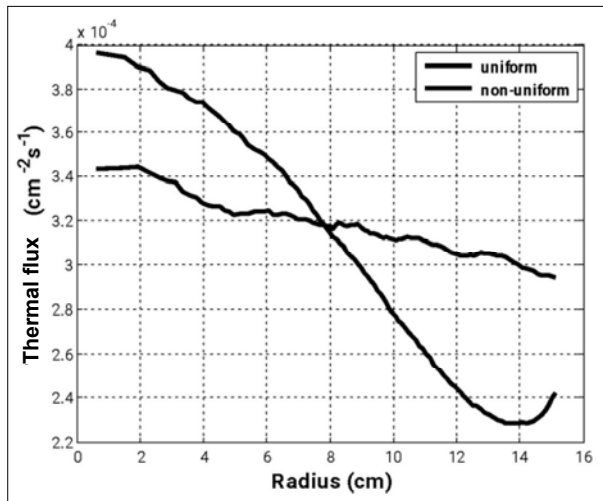
Figure 7: Thermal flux comparison uniform vs. optimal

perturbation estimates need to be corrected before being used for deciding which zone U^{235} should be added to or removed from.

Table 10 shows that a net removal of 145 g U^{235} is predicted by using correct derivatives to estimate an MCM of 1048.74 g U^{235} compared with the experimental MCM of 1065.77 g U^{235}, *i.e.*, a difference of 1.6%. The thermal flux for uniform and MCM (optimal) distributions are shown in Figure7 showing less peaking for the latter.

The correction factors are thus important for sampled derivatives to be used for estimating the zone importance and subsequently the optimal placement of fissile material.

CONCLUSIONS

The objective of this work was to estimate the reduction in critical mass in illustrative nuclear systems using results from a variational method and Monte Carlo (MC) simulation and particularly the derivatives obtained from a 'reference' MC computation. Optimization strategies were used for a bare cylindrical reactor with a restrictive condition giving 'critical' enrichment pairs. Reduced critical mass was obtained using derivative sampling for perturbation in a 'reference' design and Pontryagin's maximum principle in a variational formulation. Both these methods were applied and shown to estimate the minimum critical mass (MCM) for a bare cylindrical reactor and for a criticality experiment.

REFERENCES

Briesmeister J.F. (2003). *X-5 Monte Carlo Team: MCNP—A General Monte Carlo N-Particle Transport Code*. Los Alamos National Laboratory, Los Alamos, NM, USA.

Clark H.K. (1966). Effect of distribution of fissile material on critical mass. *Nuclear Science and Engineering* **24**(2): 133–141.
DOI: https://doi.org/10.13182/NSE66-A18298

Dam H. (2013). Minimum critical mass in a heterogeneous thermal system. *Annals of Nuclear Energy* **53**: 221–227.
DOI: https://doi.org/10.1016/j.anucene.2012.08.028

Dam H. (2015). On the anomaly in sizes of minimum critical mass cores. *Annals of Nuclear Energy* **86**: 23–28.
DOI: https://doi.org/10.1016/j.anucene.2014.11.042

Goertzel G. (1956). Minimum critical mass and flat flux. *Journal of Nuclear Energy* **2**(3–4): 193–201.
DOI: https://doi.org/10.1016/0891-3919(55)90034-6

Gosling F.G. (1999). *The Manhattan Project: making the atomic bomb*. Diane Publishing, Pennsylvania, USA.

Koreshi Z.U., Khan H., & Yaqub M. (2019). Variational methods and speed-up of Monte Carlo perturbation computations for optimal design in nuclear systems. *Nuclear Technology and Radiation Protection* **34**(3): 211–221.
DOI: https://doi.org/10.2298/NTRP190214032K

Koreshi Z.U. & Lewins J.D. (1990). Two-group Monte Carlo perturbation theory and applications in fixed-source problems. *Progress in Nuclear Energy* **24**(1–3): 27–38.
DOI: https://doi.org/10.1016/0149-1970(90)90020-6

Lamarsh J.R. & Baratta A.J. (2001). *Introduction to Nuclear Engineering*, 3rd edition, Prentice Hall Upper Saddle River, New Jersey, USA.

Lee C.K. (1973). Critical mass minimization of a cylindrical geometry reactor by two-group diffusion equation. *Nuclear Engineering and Technology* **5**(2): 115–131.

Lewins J. (2004). Minimum critical mass and flat flux in a 2-group model. *Annals of Nuclear Energy* **31**(5): 541–576.
DOI: https://doi.org/10.1016/j.anucene.2003.09.001

Otsuka M. (2017). Fuel-importance function and minimum critical mass. *Nuclear Science and Engineering* **18**(4): 514–517.
DOI: https://doi.org/10.13182/NSE64-A18770

Rief H. (1984). Generalized Monte Carlo perturbation algorithms for correlated sampling and a second-order Taylor series approach. *Annals of Nuclear Energy* **11**(9): 455–476.
DOI: https://doi.org/10.1016/0306-4549(84)90064-1

Rowinski M.K., White T.J. & Zhao J. (2015). Small and medium sized reactors (SMR): A review of technology. *Renewable and Sustainable Energy Reviews* **44**: 643–656.
DOI: https://doi.org/10.1016/j.rser.2015.01.006

Saracco P., Chentre N., Abrate N., Dulla S. & Ravetto P. (2019). Neutron multiplication and fissile material distribution in a nuclear reactor. *Annals of Nuclear Energy* **133**: 696–706.

DOI: https://doi.org/10.1016/J.ANUCENE.2019.06.044

Segrè E. (1970). *Enrico Fermi, Physicist*. University of Chicago Press, Illinois, USA.

Szasz F.M. (1984). *The Day the Sun Rose Twice: The Story of The Trinity Site Nuclear Explosion*, *July 16, 1945*. University of New Mexico Press, New Mexico, USA.

Wilkins Jr., J.E. & Srivastava K.N. (1982). Minimum critical mass nuclear reactors. part II. *Nuclear Science and Engineering* **82**(3): 316–324.
DOI: https://doi.org/10.13182/NSE82-A19392

Williams M.M.R. (2003). Some Further Considerations on Goertzel's Minimum Critical Mass Problem. *Annals of Nuclear Energy* **30**(10): 1075–1088.
DOI: https://doi.org/10.1016/S0306-4549(03)00014-8

Williams M.M.R. (2017). Conditions for a Minimum Critical Mass in a Nuclear Reactor and Considerations on Goertzel's Theorem in Transport Theory. *Nuclear Science and Engineering* **146**(2): 152–175.

DOI: https://doi.org/10.13182/NSE04-A2400

Annex 1

Notation	Description
$\hat{C}$	collision operator
D_i	diffusion coefficient, m
E	neutron energy, *MeV*
H	Hamiltonian
I_0	modified zero-order Bessel functions
$\tilde{J}$	objective function
J_0	zero-order Bessel function
K_0	modified zero-order Bessel function
$K(P' \rightarrow P)$	transfer kernel
N	atomic density
$P(\vec{r}, \mathbf{\Omega}, E, t)$	phase space
R	radius, m
S_{ij}	coupling coefficient
$\hat{T}$	transport operator
V	volume
Y_0	zero-order Bessel function
f	distribution function
k_{eff}	effective multiplication factor
k_∞	infinite multiplication factor
m	fuel mass
p_i	coefficient in polynomial
p	resonance capture
r	radius, m
t	time
$u(r)$	control parameter
u	lethargy; $\left(\ln {E_o}/{E} \right)$
x	position
y	state variable
α	alpha particle
β	beta particle
γ	gamma ray
λ	Lagrange multiplier
μ	linear attenuation coefficient
τ_i	neutron age
ϕ_1	group-1 flux
ϕ_2	group-2 flux
$\chi(\mathrm{P})$	collision denisty
φ_i	adjoint functions
ψ_i	Switching function
Σ_i	macroscopic cross-section (type *i*)
$\mathbf{\Omega}$	solid angle (steradians)
i	index for energy group or zone index
m	minimum
M	maximum

RESEARCH ARTICLE

Data Science

Two level neuro-functional forecaster: A novel dynamic hybridization for functional data forecasting

KAD Deshani[1*], DT Attygalle[1] and LL Hansen[2]

[1] *Department of Statistics, Faculty of Science, University of Colombo, Colombo 03, Sri Lanka.*
[2] *Western Sydney University, Locked Bag, Penrith NSW 2751, Australia.*

Abstract: With the advancement of technology, time series data are automatically collected without human intervention. As the data collection process becomes effortless, the next change encountered is to identify the best method to forecast time series data with high accuracies. In this regard, hybrid approaches have gained much attention where the strengths of two approaches can be combined to lessen the weaknesses of each individual approach. When exploring the features of time series data, some depict repetitive patterns and also data can be observed at several levels. The repeating curves can be considered as the higher level, whereas each individual observation can be considered as the lower level. Thus, in order to handle the data in a more effective way, the series can be handled at two levels by giving prominence to the features of each level separately. This paper proposes a novel algorithm named two level neuro-functional forecaster, which is capable of handling data at two levels, hybridizing a statistical approach with an artificial intelligence approach, to gain high accuracy levels. In addition, as this approach handles data at two levels, data sparsity at a particular level can be accommodated at the other level. To apply this algorithm to a real world dataset, electricity demand data in Sri Lanka was considered where the series consisted of daily load curves with repetitive pattern across the days. The proposed hybrid algorithm, outperforms the two approaches when individually used, with a MAPE of 3.324% for a year.

Keywords: Dynamic, forecasting, functional data analysis, hybrid, neural network.

INTRODUCTION

In most real-world situations, data are periodically collected and analyzed using various forecasting methods, and the structural aspect of such periodic data can be analyzed in different ways. One way of considering the structural aspect when forecasting, is through a sequence of curves that evolve over time. Alternatively, each point on the curve can be taken as a separate observation of the series forecast. As such, the forecasting can be done at a higher level in terms of curves or at a lower level in terms of observations.

A real-world example to illustrate the above-mentioned features of a time series can be commonly witnessed in the energy industry. A plot displaying the electricity demand within a day is usually named as a "daily load curve," where for a specific time period, the electricity demand series shows a clear periodic pattern, possibly with a few exceptions based on other external factors. When considering load data, it can be commonly experienced that such data consist of several levels such as hourly, daily, weekly, or seasonally. Thus, the load forecasting models can be considered as an ideal application to examine how the structural aspect of a series can be considered in many ways when implementing

* Corresponding author (deshani@stat.cmb.ac.lk; https://orcid.org/0000-0002-5489-3436)

forecasting models for improved accuracies. Currently, accurate and efficient short term load forecasting (STLF) models are gaining much attention, as even a minor improvement to such models can have a remarkable impact on the power and energy industry, and thereby the economy of a country.

In the recent past, much published work can be found utilizing hybrid models to forecast the load in a short term manner, showing promising results. When exploring such hybrid models in load forecasting, various approaches have been used. Some approaches are through combining two statistical techniques at two levels (Cho *et al*., 2013) and some other approaches use two artificial intelligence approaches at a single level (Kavousi *et al*., 2014; Bashir *et al*., 2022). Studies also reveal that there are approaches combining a statistical technique with an artificial intelligence approach but based on data at a single level (Zhang, 2003; Nie *et al*., 2012).

The model proposed by Bashir *et al*. have utilized a Prophet model and a long short-term memory (LSTM) model to predict load, in Belgium. They have used data at a single level where the hybridization is to use a LSTM model for the residuals resulting from the Prophet model. The results indicated that their hybrid models outperform the LSTM, Prophet model and ARIMA models when used on their own (Bashir *et al*., 2022). Fard and Akbari-Zadeh (2014) have proposed a hybrid forecasting method based on the wavelet transform, autoregressive integrated moving average (ARIMA), and artificial neural network (ANN) for short-term load forecasting of the Fars Electric Power Company in Iran. The three models used in this study had been combined in such a way that the linear nature of the series has been captured using the ARIMA model, whereas the nonlinear patterns have been captured by the ANN models, and the wavelet for transformation purposes to avoid overfitting problems. They have shown that when the methods are combined, the accuracy level increases substantially (Fard & Akbari-Zadeh, 2014). By utilizing two artificial intelligence approaches along with regional level temperature data, Barman, Choudhury and Sutradhar have shown better accuracies when forecasting the load. They have used support vector machine (SVM) and the grasshopper optimization algorithm (GOA) to implement the hybrid model (Barman *et al*., 2018).

As opposed to the above work where data are at a single level, Cho *et al*. (2013) have forecasted the electricity demand of a French energy company using a hybrid approach with data at two levels. Their approach combined a general additive model to capture the overall trend and seasonality, and a curve linear regression model

to capture the dependence structure, across consecutive daily loads. It was revealed that four versions of the hybrid models have been tested and the hybrid models have resulted in very low errors compared to using individual approaches.

The use of functional data analysis techniques has grown rapidly in relation to statistical analysis for clustering (Jacques & Preda, 2014), prediction (Antoch *et al*., 2010; Shan, 2013; Jornaz, 2016), and system investigation (Gubian *et al*., 2014), proving to generate promising results. However, there are only a few functional modelling approaches that have been applied to electricity demand on short term, medium term (Antoch *et al*., 2010) and very short term (Shan, 2013) for forecasting. Antoch *et al*. (2010) have also proved, through their research, that functional approaches provide promising results for short-term or medium-term forecasts of electricity demand. This is due to their capability of integrating the underlying functional nature of the data when modeling. It is also emphasized that such an approach allows a researcher to consider additional information such as smoothness of the functional parameters, that is not available when using standard time-series techniques. However, the researchers have not addressed the issue when forecasting a series with unusual or outlying observations but have stated that such situations can affect the model accuracy (Antoch *et al*., 2010). Considering demand data as functions, Shan (2013) has reduced the dimensionality of the curves using functional principal component analysis, where the forecasting was based on the principal component scores. It is to be emphasized that the high accuracies have been obtained eliminating the unusual observations, thus reducing the volatility of the demand series (Shan, 2013). The functional principal component analysis has been used to reduce the dimensionality and a probabilistic forecast has been made based on the derived components of generalized quantile curves. Moreover, it has been shown that this method outperforms many widely used non-functional approaches based on error analysis (Cabrera & Schulz, 2017).

It is evident that the deterministic behaviour of the curves can be better captured using functional data analysis techniques than any other approach. However, the unusual behaviours of the curves are weakly captured using functional data analysis approaches, which may result in low accuracy levels. On the other hand, machine learning approaches have gained much attention due to their ability to capture non-linear and volatile behaviours of data even in load forecasting. Thus, when forecasting a time series with a repetitive pattern, a hybrid approach

combining a functional data analysis approach with an artificial intelligent approach seems to be an ideal hybridization to improve the accuracy substantially. However, there are no forecasting hybridization methods that focusses on high accuracy levels, where such data are at two different levels and also utilizing a statistical approach and an artificial intelligence approach.

Apart from choosing the most appropriate technique to forecast, additional features are commonly used for model improvements, but such information may not always be available at the data level of interest. However, if the available data at different levels are integrated and used effectively for forecasting, the challenges faced with scarce data at specific levels can be overcome. More importantly, this will improve accuracy, as more structural information of the series is used when forecasting. Moreover, implementing dynamism is useful to accommodate real world scenarios as opposed to implementing static models.

Considering all the above-mentioned aspects, this research paper proposes a novel algorithm named as Two Level Neuro-Functional Forecaster that can be used to dynamically forecast functional data at two levels, combining a statistical approach and an artificial intelligence approach.

MATERIALS AND METHODS

The proposed hybrid Two-Level Neuro Functional Forecaster algorithm is an efficient algorithm that can be useful to forecast a time series with a repetitive pattern, where data are at two different levels, even when data are sparse. When considering the hybridization in this algorithm, it utilizes a rule-based approach combining a statistical approach and an artificial intelligence approach as illustrated in Figure 1. The approach will be explained under three subsections namely, required data, techniques used, and rule-based hybridization.
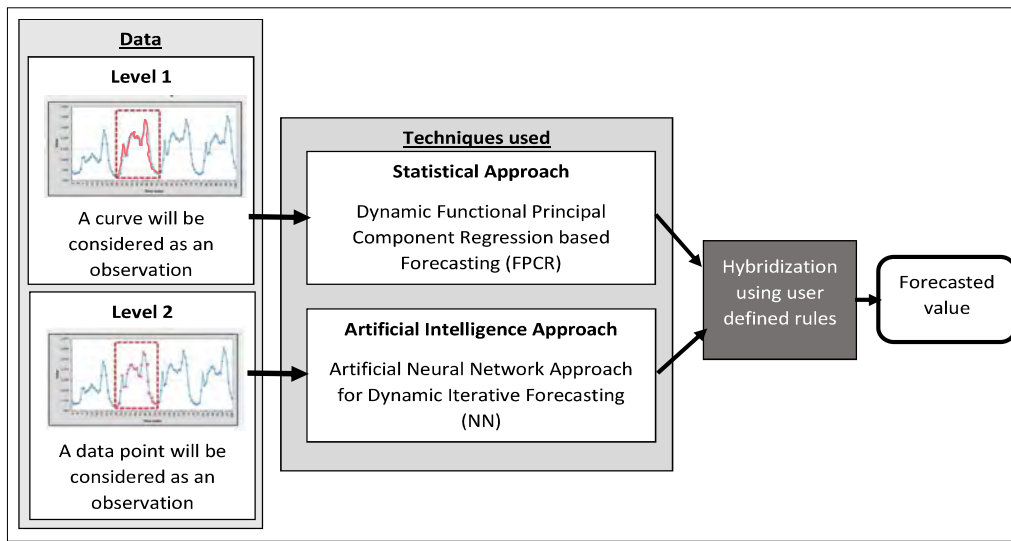

Figure 1: A diagrammatic representation of the proposed hybrid approach

Required data

Certain time series data with a repetitive pattern can be forecasted at two levels depending on the availability of data. Recall that at the higher level, observations are considered as curves, whereas at the lower level, observations are taken at each time point.

When considering the higher level, a statistical approach, developed as Dynamic Functional Principal Component Regression (FPCR), is used to forecast

the series considering the repetitive pattern (Deshani *et al*., 2017). At the lower level, an artificial intelligence approach, developed as Artificial Neural Network Approach for Dynamic Iterative Forecasting (NN) (Deshani *et al*., 2019), is utilized to forecast the series.

Techniques used

The two main approaches FPCR and NN used in this Two-Level Neuro Functional Forecaster algorithm are described below.

Dynamic functional principal component regression (FPCR) approach

The FPCR algorithm considers a curve, related to a repetitive pattern, as a single observation where the curves are presented as functional objects. This algorithm consists of three main steps.

Step 1: Reducing the dimensionality of the functional objects through functional principal component analysis.

Step 2: Forecasting the principal component scores through fitting time series models. A carefully chosen moving window is then used to dynamically forecast the values. That is, more recent data will be incorporated to fit the models while observations prior to the moving window will be disregarded.

Step 3: Calculating the future load using the forecasted principal component scores.

Artificial intelligence approach named as artificial neural network approach for dynamic iterative forecasting (NN)

The NN forecasting method, applied to lower-level data, considers many associative variables available at each time point. It utilizes an error backpropagated neural network for an iterative forecast. This algorithm is developed, so that the recently forecasted values are incorporated to the immediate next forecast. Moreover, a moving window is selected to make the forecasting process dynamic. This algorithm consists of four main steps

Step 1: Identifying the lagged terms and the features that are influential to the respective time series.

Step 2: Identifying the best network architecture for the model and use such multiple neural networks for the forecasting.

Step 3: Selecting the best moving window.

Step 4: Iteratively forecasting the series.

Rule based hybridization

In this proposed algorithm, the set of rules to be used for the hybridization are formulated by considering the most recent time frame and various characteristics of the errors in each of two approaches. Here, the error bands (first quartile of the error series, third quartile of the error series) of each time point are explored for the two distinct approaches, and the method having a narrow error band centred around zero or the average value of the two forecasts from the two methods is chosen for the final forecast, as appropriate.

Thus, the proposed two-level neuro functional forecaster algorithm can be applied for forecasting a time series that has a structure of a repetitive pattern and where the data are available at two levels.

RESULTS & DISCUSSION

To illustrate this forecasting algorithm, hourly total electricity demand for the period of 1st January 2008 to 31st December 2012 from Sri Lanka was considered as the study was initiated in 2013. The structure of the dataset was explored and is given in Figure 2, where the data are at two levels, namely daily and hourly.
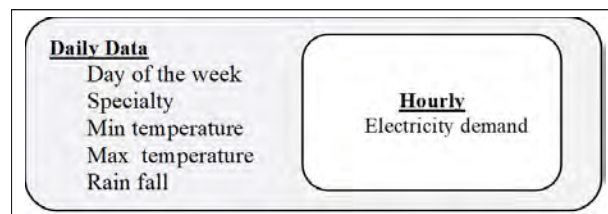
Figure 2: The structure of the dataset at different levels

Furthermore, it consisted of **day of the week** (Sunday, Monday, ..., Saturday), **specialty** (None, Public Holiday, Bank Holiday, Poya day, Saturday after a holiday, 1 day after New Year, 2 days after New Year), **minimum daily temperature**, **maximum daily temperature** and daily **rainfall** at the daily level. The only available hourly measurement was the hourly electricity demand.

Exploring the data

Figure 3 displays the load curves over the study period showing a clear repetitive pattern over the years with a gradual increase of demand, with a very few exceptions. The demand variations are shown through colours of the rainbow where the lowest demand being violet and the highest being red. Recall that a day is considered as the higher level, and an hour is the lower level.

Figure 4 illustrates load curves for a selected time period, separately drawn for the months of January 2008 and January 2012 to elicit the demand variations across days of the week. It can be clearly observed that the demand curve in general has shifted up from 2008 to 2012 but maintaining a similar shape between them. When comparing load curves across the days of the week, demands on Saturdays and Sundays were comparatively lower than other working days of the week, indicating
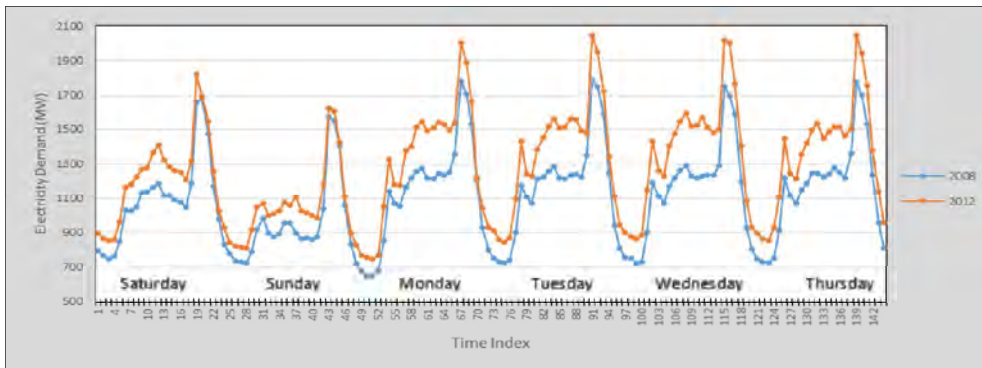
the importance of using the day effect. Thus, it was decided to perform a comprehensive cluster analysis and to incorporate the results as inputs to the forecasting models to enhance the accuracy and the efficiency by incorporating them as additional features (Deshani *et al.*, 2014a; 2014b; 2022).
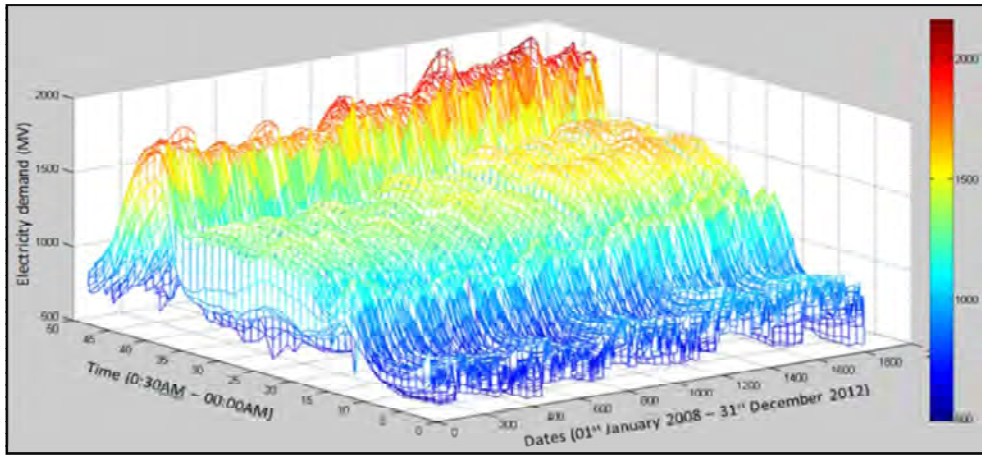
Figure 3:   Fluctuations of electricity load curves over the years

Figure 4:   Shape of few selected daily load curves in January 2008 and 2012

Hourly temperatures and rainfall measurements were not available in this data, and thus had to be estimated to be incorporated in this algorithm. The two estimations, respectively, were based on using the MinMax Cos-LEA estimation (Deshani *et al.*, 2016) algorithm and the rainfall adjusted temperature estimation (RATE) algorithm (Deshani *et al.*, 2016). Both algorithms utilize the available daily maximum and minimum temperatures and rainfall data to estimate the hourly temperature measurements.

As such, identifying the key variables separately at both data levels, and preparing data to suit the said data structure, are an integral part of the process, before applying the two-level neuro functional forecaster algorithm.

Application of the algorithm

The two-level neuro functional forecaster algorithm, as illustrated in Figure 5, accommodates data at two levels using 2 different approaches. First, at the higher level, considering daily load curves as observations, the FPCR approach was used to forecast the demand series using the daily measurements. The data used for this approach were the daily load curves, day of the week, speciality

of the day, daily maximum temperature, daily minimum temperature, curve clustering results and the month, based on this data set. The selected principal component scores were forecasted using the ARIMA model with regressors that best suited this data (Deshani *et al.*, 2017).

Then the NN approach was used to forecast the series considering hourly demand values as separate observations. Selected lagged demand data, demand clustering results and hourly temperatures were used as inputs to this approach. One day ahead hourly demand values were forecasted in an iterative manner and the

findings have been published by the same authors (Deshani *et al.*, 2019).

Consequently, as stated in section 'Rule based hybridization', the process of combining the results of the two approaches (FPCR approach and the NN approach) to implement the hybrid forecast was carried out based on carefully chosen rules. These rules were established using the forecasting patterns of the year 2011, the latest available data, through observing the mean absolute percentage error (MAPE) of the dynamic forecasts of the two approaches.
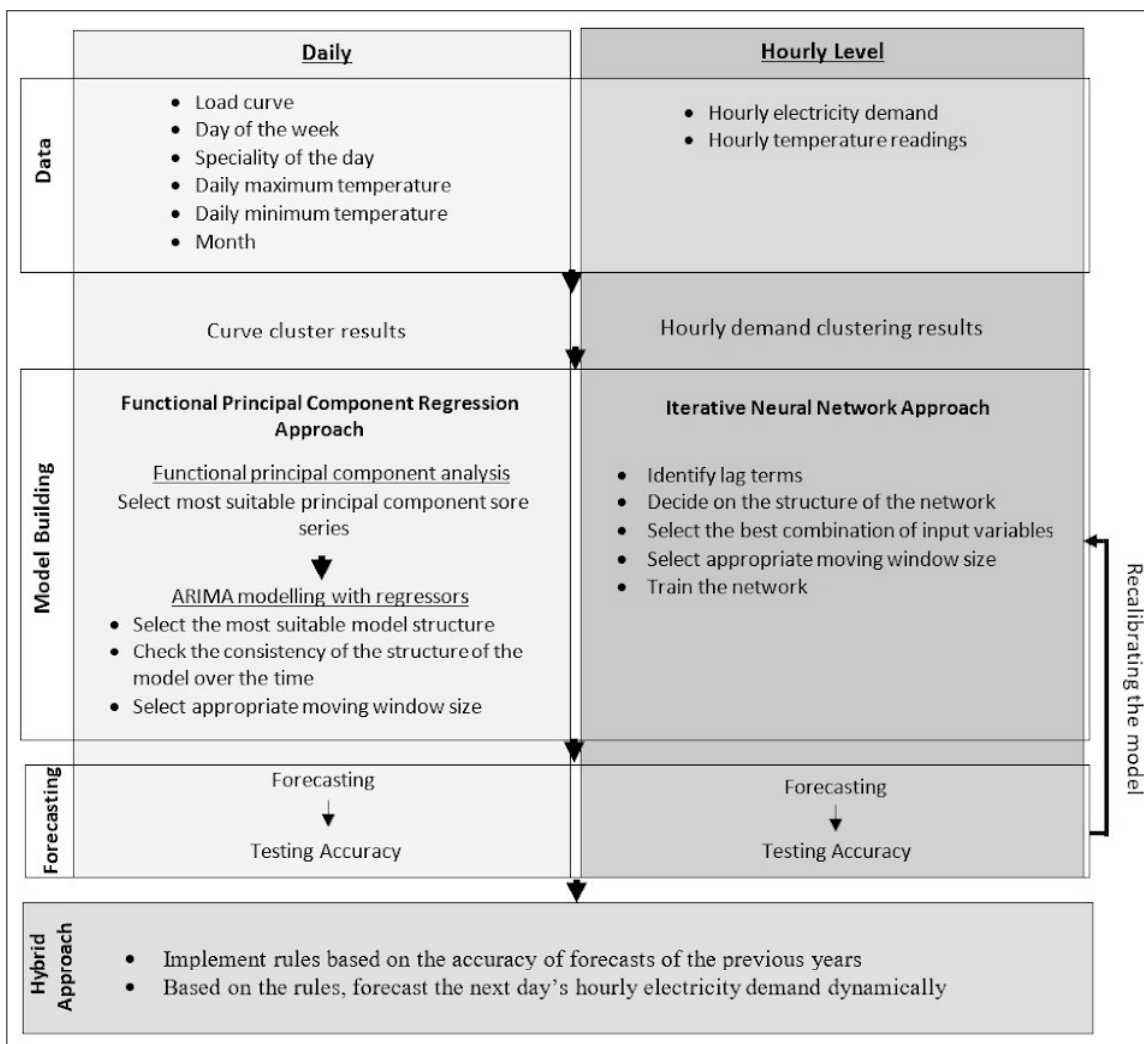

Figure 5:   Flow chart of the proposed decision-making algorithm

The outcome of the application of the two approaches was such that, for some day types, the forecast of one approach outperformed the forecast of the other approach. It should be highlighted that the day of the week and the specialty had a strong influence on the demand. Moreover, the time of the day also played a key role when selecting the most suitable approach for the forecast. Thus, it was decided to use a rule-based approach to do the hybridization, focusing on the time of the day, the day of the week and the specialty of the day. The hybridization took place differently for various time periods, choosing the better performing approach or their averages accordingly. Figure 6 clearly shows how the error bands (the first and third quartiles of the hourly errors) of the two approaches vary with respect to the time of the day. It can be clearly seen that the error bands obtained from one approach become very narrow and centred around zero during some time intervals different from the other approach. Thus, when combining the two approaches, for a particular time period, it was decided to select the better performing approach that had a narrow error band centred around zero to do the forecast. If such a prominent difference cannot be observed, the average forecasted value obtained from the two methods was to be selected. Applying the two approaches in this manner, promising results have been shown, as opposed to using a weighted average of the forecasted values, without any such considerations.
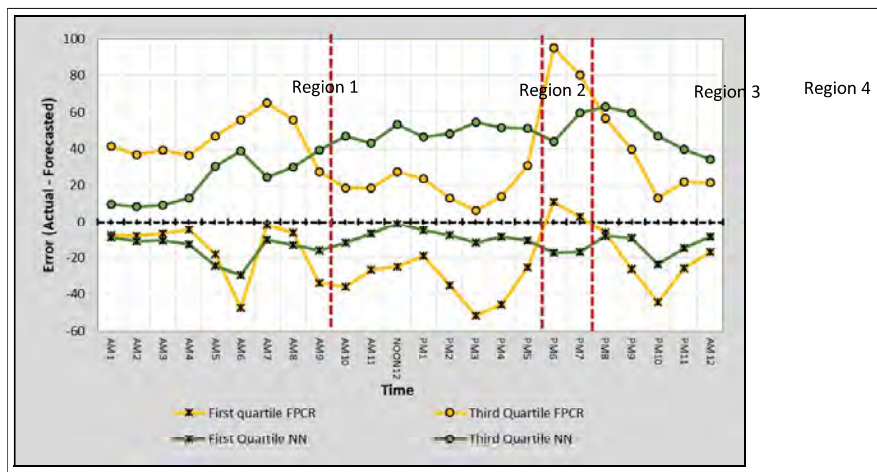

Figure 6: Variations of the first quartile and the third quartile in each hour during 2011

The implemented rules for forecasting were formulated for the preceding entire year paying attention to impacting features, which in this application were the time of the day, the day of the week and the specialty of the day.

After a comprehensive error analysis, the following rules were established and adopted when forecasting the hourly electricity demand.

Rule 1: For each day, forecast electricity demands during 1.00 a.m – 9.00 a.m using the NN approach

Rule 2: If the speciality of a day is a PBM holiday, Poya day, Saturday after a holiday, 1 day after New Year, or 2 days after New Year, use the NN approach to forecast all 24 hours.

Rule 3: If a day is having no specialty, forecast the electricity demands during 10.00 a.m – 5.00 p.m using the average of the forecasts using the FPCR approach and the NN approach.

Rule 4: Forecast the electricity demands during 6.00 p.m – 7.00 p.m using the NN approach.

Based on the above-mentioned rules, the final forecasts were made.

The accuracy of the dynamic forecast using two level neuro-functional forecaster

The hourly electricity demand in Sri Lanka for the year 2012 was forecasted using the two level neuro-functional forecaster algorithm considering the rules implemented to do the hybridization. Due to the dynamism in this forecasting algorithm the errors for an entire year were investigated. It was found that the MAPE values were not consistent over the year but gave an overall average of 3.324% for the entire year (Average RMSE for the entire year 55.05). As with much published work in relation to such forecasts, with this data too, the highest errors were reported on holidays where the consumption patterns were highly unpredictable. According to Figure 7, it can be clearly observed that the MAPE for the New Year season was comparatively higher than that of other days whereas the maximum error was observed for the day after the New Year. Figure 8 displays some selected forecasted daily load curves with different MAPEs.
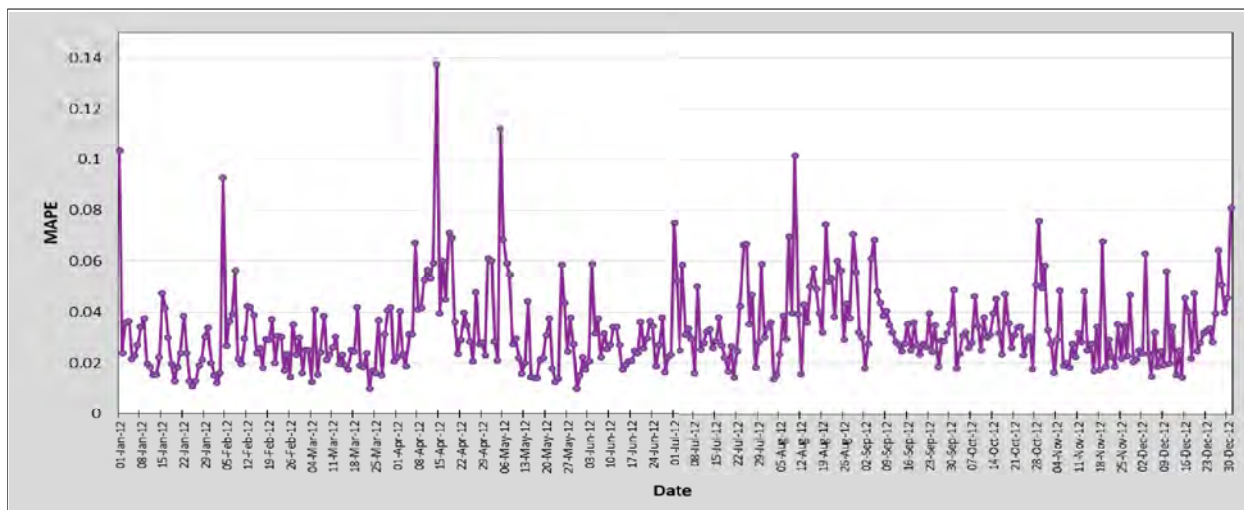

Figure 7: MAPE of the final dynamic forecasting model

A comparison of dynamic forecasts for the year 2012 using the FPCR approach, NN approach and the proposed two level neuro-functional load forecaster approach is depicted in Figure 9. It can be clearly seen that using the proposed hybrid approach, the best forecasts can be made, as opposed to when the approaches are used separately as two single approaches. Wilcoxon sign-rank test for paired samples were used to statistically assess whether the median MAPE of the two individual approaches are significantly higher than the median MAPE of the hybrid forecasts. Wilcoxon signed-rank test showed that the two approaches elicited a statistically significant increase in the MAPEs compared to the MAPE of the hybrid forecasting (p-value = 1.108e-12 when comparing with the NN approach and p-value < 2.2e-16 when comparing with the FPCR approach). Thus, the two level neuro-functional forecaster approach yields superior results compared to the individual forecasts from the two approaches.

Increasing the accuracy levels of the forecasting through hybrid approaches has gained much attention in the recent past. This research highlights the importance of exploring the structural aspects and other features of the time series and incorporating them when selecting the most appropriate forecasting models for hybridization. One such important aspect is to view the structure of the time series with a repetitive pattern, as data from two different levels, that will allow the capture of unique features from each level, effectively. There can be other types of time series where no such structural patterns can be easily identified. Then it would not be possible to enhance the accuracy of the forecasts, by incorporating level specific information to the models. Through the electricity demand data application presented in this paper, the structural behaviour can be easily captured at the daily level, while the unusual sudden behaviours can be easily accommodated at the hourly level. With the use of the features at two levels, the thorny issue of

data scarcity can also be addressed in an effective way as one level will facilitate the model with the required data, though the other level does not comprise the data as required.
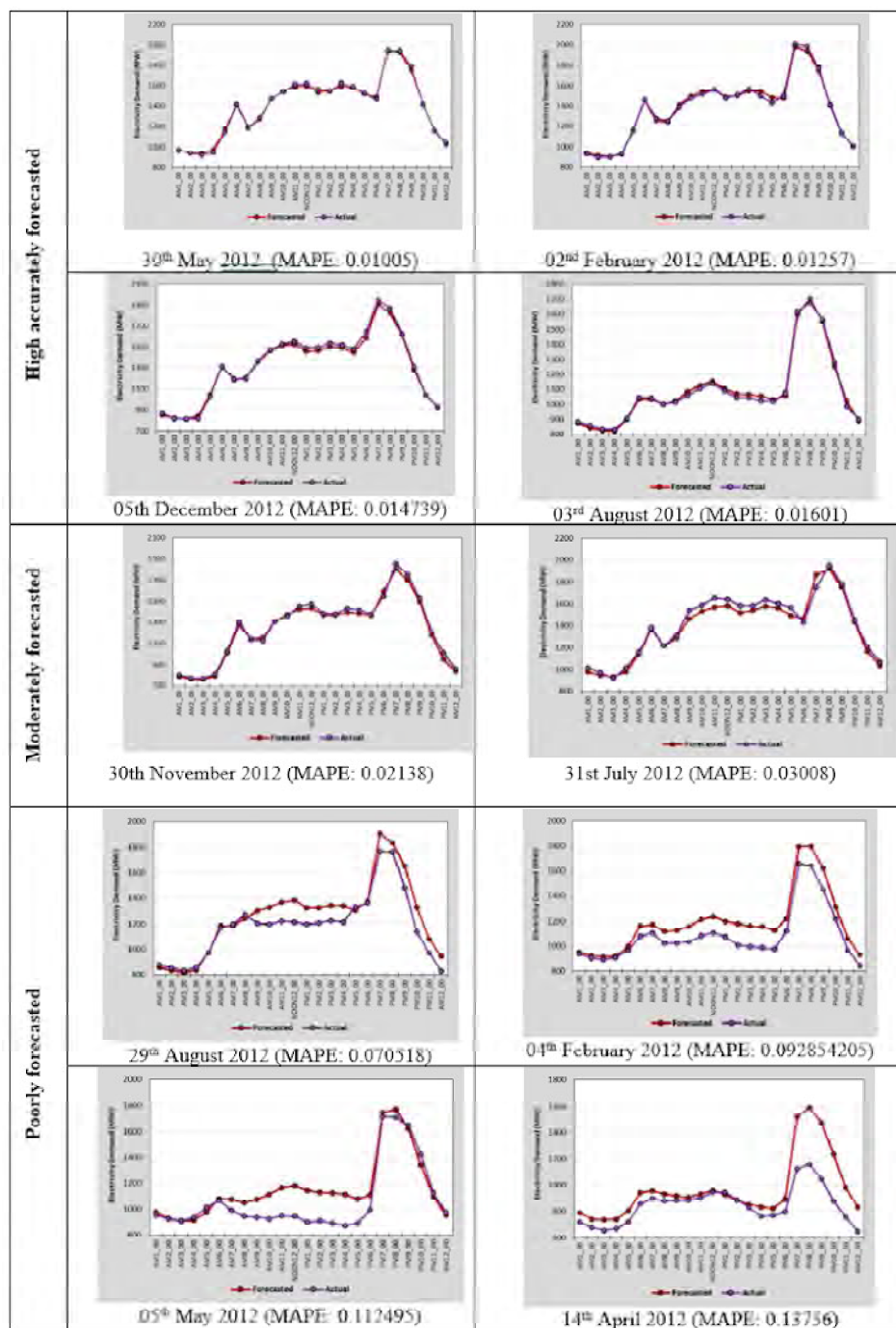

Figure 8: Some well, moderately and poorly forecasted days using the hybrid approach
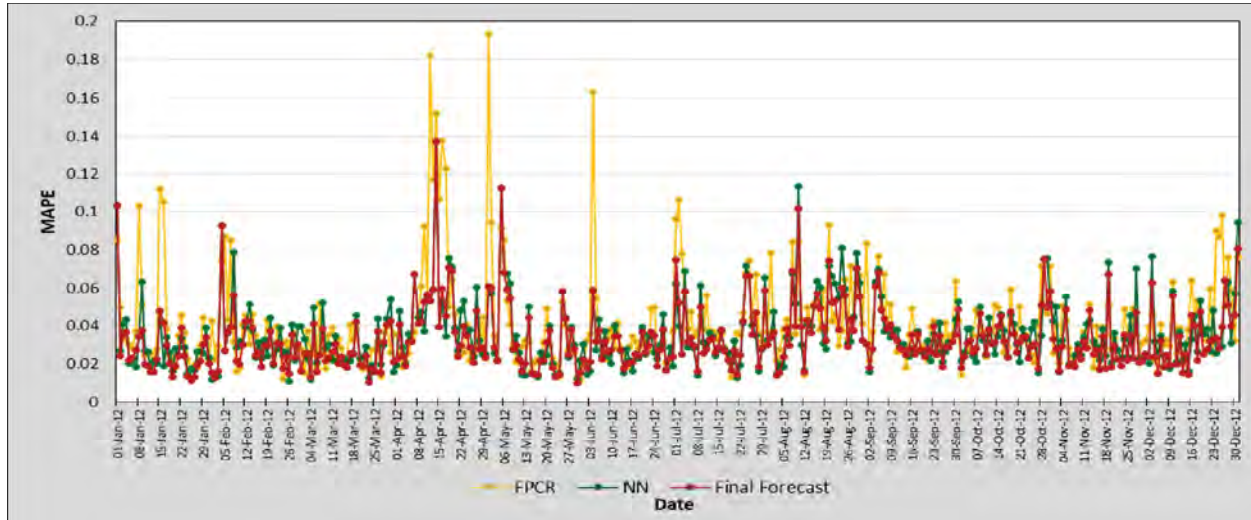
Figure 9:   MAPE of the 2012 dynamic forecasts from the three methods

CONCLUSION

When implementing hybrid models, it should be highlighted that approaches from two different fields can be fruitfully utilized to overcome the weaknesses of one approach compared to the other. Thus, considering all the above-mentioned aspects and also with proven promising results, the proposed two level neuro-functional forecaster algorithm can be recommended as an ideal forecasting algorithm for a time series with a repetitive pattern where the data are available at two levels. For the application presented in this paper, the dynamic forecast yielded an average MAPE of 3.32% for an entire year while the FPCR approach and the NN approach resulted in a MAPE of 4.08% and 3.96%, respectively, for the entire year when used separately. Thus, it can be concluded that the proposed dynamic hybrid algorithm can be used to gain high accuracies for time series data with repetitive patterns rather than using the FPCR approach or the NN approach with the use of the concept of data at two levels. This study sets the platform to explore various rule-based hybridization techniques to forecast complex time series data.

REFERENCES

Antoch J., Prchal L., De Rosa M.R. & Sarda P. (2010). Electricity consumption prediction with functional linear regression using spline estimators. *Journal of Applied Statistics* **37**(12): 2027–2041.
DOI: https://doi.org/10.1080/02664760903214395

Barman M., Choudhury N.D. & Sutradhar S. (2018). A regional hybrid GOA-SVM model based on similar day approach for short-term load forecasting in Assam, India. *Energy* **145**:710–720.
DOI: https://doi.org/10.1016/j.energy.2017.12.156

Bashir T., Haoyong C., Tahir M.F. & Liqiang Z. (2022). Short term electricity load forecasting using hybrid prophet-LSTM. *Energy Reports* **8**:1678–1686.

Cabrera B.L. & Schulz F. (2017). Forecasting generalized quantiles of electricity demand: a functional data approach. *Journal of the American Statistical Association* **112**(517): 127–136.
DOI: https://doi.org/10.1080/01621459.2016.1219259

Cho H., Goude Y., Brossat X. & Yao Q. (2013). Modelling and forecasting daily electricity load curves: a hybrid approach. *Journal of the American Statistical Association* **108**(501):7–21.
DOI: https://doi.org/10.1080/01621459.2012.722900

Deshani K.A.D, Attygalle M.D.T, Hansen L.L. & Karunarathne A. (2014). An exploratory analysis on half-hourly electricity load patterns leading to higher performances in neural network predictions. *International Journal of Artificial Intelligence & Applications* **5**(3):37–51.

Deshani K.A.D, Hansen L.L., Attygalle M.D.T. & Karunarathne A. (2014). Improved neural network prediction performances of electicity demand: modifying inputs through clustering. *Proceedings of 2nd International Conference on Computational Science and Engineering.* India: AIRCC, pp. 137–147.
DOI : https://doi.org/10.5121/csit.2014.4412

Deshani K., Attygalle D. & Liyanage Hansen L. (2016). Diurnal temperature modeling with sparse data and data integration. *Proceedings of 4th Annual Operational Research and Statistics.* Global Science and Technology

Forum, Singapore, pp. 28–32.

Deshani K.A.D, Attygalle D. & Liyanage Hansen L. (2016). Incorporating influential factors in diurnal temperature estimation with sparse data. *GSTF Journal of Mathematics, Statistics, Operations Research* 3(2):12–16.

Deshani K., Attygalle D., Liyanage-Hansen L. & Lakraj G. (2017). dynamic short term load forecasting using functional principal component regression. *Proceedings of the 1*st *International Conference on Machine Learning and Data Engineering,* 20-22 November. Sydney, Australia, pp. 76–82.

Deshani K.A.D, Liyanage L.H. & Attygalle D. (2019). Artificial neural network for dynamic iterative forecasting: forecasting hourly electricity demand. *American Journal of Applied Mathematics and Statistics* 7(1): 9–17.

Deshani K., Liyanage L.H. & Attygalle D.T. (2022). Clustering time related data: A regression tree approach. *American Journal of Applied Mathematics and Statistics* **10**(1): 22–27.
DOI: https://doi.org/10.12691/ajams-10-1-4

Fard A.K. & Akbari-Zadeh M. (2014). A hybrid method based on wavelet, ANN and ARIMA model for short-term load forecasting. *Journal of Experimental and Theoretical Artificial Intelligence* **26**(2): 167–182.
DOI: https://doi.org/10.1080/0952813X.2013.813976

Gubian M., Torreira F. & Boves L. (2014). Using functional data analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics* **49**: 16–40.
DOI: https://doi.org/10.1016/j.wocn.2014.10.001

Jacques J. & Preda C. (2014). Model-based clustering for multivariate functional data. *Computational Statistics and Data Analysis* **71**(C): 92–106.

Jornaz A.S. (2016). Modeling daily electricity load curve using cubic spline and functional principal components. *PhD thesis*, Missouri University of Science and Technology, USA.

Kavousi A.F., Samet H. & Marzbani F. (2014). A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting. *Expert Systems with Applications* **41**(13): 6047–6056.
DOI: https://doi.org/10.1016/j.eswa.2014.03.053

Nie H., Liu G., Liu X. & Wang Y. (2012). Hybrid of ARIMA and SVMs for short-term load forecasting. *Energy Procedia* **16**(C): 1455–1460.

Shan H.L. (2013). Functional time series approach for forecasting very short-term electricity demand. *Journal of Applied Statistics* **40**(1): 152–168.

Zhang G.P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **50**: 159–175.

RESEARCH ARTICLE

Plant Genetics

Path coefficient analysis using traditional and improved rice genotypes for trait effect on grain yield

HAPA Shyamalee[1] and AL Ranawake[2*]
[1] *Agriculture Research Station, Thelijjawila, Sri Lanka.*
[2] *Department of Agricultural Biology, Faculty of Agriculture, University of Ruhuna, Mapalana, Kamburupitiya 81100, Sri Lanka.*

Abstract: Accessions of traditional rice play a crucial role in preserving the genetic diversity of commercial rice. They contain specific and useful traits that can help to improve the overall yield, which cannot be achieved by relying solely on improved rice varieties. This is because traditional rice accessions display a wide range of morphological characteristics that contribute to their unique yield determinants. Fifty traditional and forty-five improved rice genotypes were evaluated for thirteen agronomic traits (days to 50% flowering, plant height, number of tillers per plant, number of effective tillers per plant, number of leaves per plant, panicle length, flag leaf length, flag leaf width, number of grains per panicle, filled grain percentage, 100-grain weight at maturity, grain length, and grain width) to understand the yield determination of traditional and improved rice genotypes. Path analysis was carried out using IBM SPSS AMOS statistical software to understand the difference in the direct and indirect effect of the studied parameters. The direct effect of effective tillers per plant ($\beta = 0.519$), filled grain percentage ($\beta = 0.496$), and hundred-seed weight ($\beta = 0.403$) on grain yield was high in improved rice varieties. In traditional rice accessions, the direct effect of effective tillers per plant ($\beta = 0.746$) and filled grain percentage ($\beta = 0.395$) on grain yield was high, but there was no direct effect of hundred seed weight. Filled grains per panicle recorded a moderate effect in both traditional ($\beta = 0.246$) and improved ($\beta = 0.266$) rice genotypes. None of the studied parameters directly and negatively affected the yield of improved rice varieties. Still, several traits (flag leaf length, days to 50% flowering, panicle length, seed length, and seed width) negatively affected the yield of traditional rice accessions. The contributions of yield determinants in traditional rice accessions deviated substantially from those observed in improved rice varieties, highlighting the critical role of plant architecture in determining the final yield. Fertile tiller number, filled grain percentage and grains per panicle are possible traits that can be used in direct selection criteria for improved and traditional rice genotypes for high yield.

Keywords: Direct effect, improved rice, indirect effect, path analysis, traditional rice.

INTRODUCTION

The traditional rice accessions maintain diversity within the rice gene pool. These traditional accessions often harbour specific and advantageous traits that serve as a buffer against potential vulnerabilities in improved varieties. The coexistence of traditional and improved rice genotypes, each characterised by a spectrum of morphological variations, presents a unique challenge in understanding the determinants of yield. Given the extensive diversity in morphological characteristics among these genotypes, it becomes essential to find how various agronomic traits contribute to yield in both traditional and improved rice genotypes.

* Corresponding author (lankaranawake@hotmail.com; https://orcid.org/0000-0003-0517-9911)

Path analysis is a statistical technique to study variables' direct and indirect relationship (Valenzuela & Bachmann, 2017). The direct and indirect effects explained by path analysis (path coefficient, β) can be considered as selection criteria for crop improvement (Sudeepthi *et al.,* 2020). Path analysis is a step-forward technique to correlation analysis. Correlation coefficients assess the degree of relationship between or among traits (Mecha *et al.,* 2017; Baye *et al.*, 2020). The correlation coefficient tables explain the interrelationship among yield and yield components, which is also helpful for developing an efficient selection strategy. Path analysis divides correlation coefficients into direct and indirect effects (Ulukan *et al.,* 2003), and it explains the hierarchical order of trait contribution to final grain yield (Samonte *et al.,* 1998), while the correlation coefficient interprets the association of traits, but not the comparative impact of each trait on the final yield. Path analysis facilitates indirect selection indicators, and the coefficient decides the selection efficiency between the considered trait and the yield (Li *et al.,* 2019). The promising results of path analysis have been reported in many studies on rice using different yield determinants as parameters (Agahi *et al.,* 2007; Singh *et al.,* 2018; Immanuel *et al.,* 2011). Path analysis separates the direct effect of a trait and indirect effects of the same through other traits for better interpretation of the total effect (Rasel *et al.*, 2018), and it reveals which trait should be considered as positive indicator for yield increment or which traits should be avoided as negative yield indicators.

The selection of traits for crop improvement is a complex decision-making process. The promising results of path analysis, as evidenced by previous studies on rice, emphasise its efficacy in guiding selection strategies (Shahidullah *et al.,* 2009; Oladosu *et al.,* 2018; Al-Musawi *et al.,* 2019; Laxmi *et al.,* 2023; Limbongan *et al.,* 2023; Reddy *et al.,* 2023). By presenting the direct and indirect effects of various traits, path analysis elucidates the hierarchical order of trait contributions, and it aids in identifying positive indicators for yield increment and potential negative yield indicators. This approach to trait selection contributes to the development of more efficient and targeted strategies for crop improvement (Chakraborty *et al.,* 2012; Lule *et al.,* 2012; Thangamani & Jansirani 2012; Owere *et al.,* 2015; Eric *et al.,* 2016; Janaki, 2018; Rekha *et al.,* 2019; Nagalakshmi *et al.,* 2020; Sivakumar *et al.,* 2020). The historical context of the Green Revolution emphasises the transformative impact of genetic improvements in rice cultivars.

After the green revolution, the yield potential of introduced rice cultivars was doubled due to genetic improvement for short stature in high-yielding varieties (Khush, 1999; Davies, 2003). The rice variety, IR8 introduced after the Green Revolution, which gave around eight tons per hectare, was short and stiff (Dalrymple, 1985). The introduction of dwarfing genes to rice and wheat was the turning point of the Green Revolution (Khush, 2003). Sd1 (Monna *et al.,* 2002; Spielmeyer *et al.,* 2002) was the dwarf gene in rice which accelerated the yield revolution. Architectural improvement in rice during the Green Revolution demarcated the high-yielding semidwarf varieties from traditional rice accessions. Due to plant architectural changes, improved rice cultivars and traditional rice accessions differ significantly in yield. Hence, yield determinants do not function or contribute at the same magnitudes in the two groups of rice genotypes, improved rice cultivars and traditional rice accessions. The present study examined the specificities of yield determinants in traditional rice accessions and improved rice cultivars. Recognizing that these determinants may not function at uniform magnitudes across the two groups of rice genotypes, path analysis emerges as a tool of choice to unravel the differential impacts of various traits on yield. By scrutinising key agronomic parameters' direct and indirect effects, this research provides a comprehensive understanding of the back-and-forth relationship between plant architecture and yield in traditional and improved rice genotypes.

The study evaluated fifty traditional and forty-five improved rice genotypes, focusing on thirteen key agronomic traits. These traits included essential parameters such as days to 50% flowering, plant height, tiller characteristics, leaf morphology, panicle attributes, and grain dimensions. By systematically analysing these traits, the research aimed to unravel the intricate interactions that ultimately determine the yield in traditional and improved rice genotypes. This exploration serves as an essential step towards enhancing our understanding of the factors governing yield in diverse rice populations. The study aims to bridge the existing knowledge gap regarding the contribution of different traits to yield in traditional and improved rice genotypes. The results from this research may help targeted breeding strategies, ensuring the preservation of beneficial traits in traditional varieties and the continued enhancement of yield in improved cultivars. By exploring the complex relationships within diverse rice populations, the study contributed to the broader discourse on sustainable agriculture and crop improvement strategies in the face of evolving environmental and economic challenges.

MATERIALS AND METHODS

Forty-five (45) improved rice varieties and fifty (50) traditional rice accessions collected from the Plant Genetic Resource Centre, Gannoruwa, Peradeniya, Sri Lanka, were used for the study (Table 1).

Dormancy broken seeds were planted in nursery beds and transplanted in the field according to a randomised complete block design with four replicates and twenty plants per replicate. The field was established under rain-fed conditions at the experimental field, Research Station, Thelijjawila in the *Maha* season 2021/2022. Crop management was done according to the recommendation of the Department of Agriculture for wet zone paddy field cultivation under rainfed conditions (DOA, 2013). Data was collected on twenty plants in each replicate (Table 2).

Table 1:   Traditional and improved rice genotypes were used in the study in the 2021/2022 *Maha* season.

Traditional rice accessions		Improved rice varieties	
Dahanala3917	*Mawee3683*	*Bg300*	*Bg314*
Dahanala3304	*Mawee5384*	*Bg352*	*Bg357*
Dikwee3741	*Murungakayan6285*	*Bg359*	*Bg366*
Dikwee3504	*Murungakayan6263*	*Bg369*	*Bg409*
Dikwee2203	*Murungakayan3921*	*Bg94-1*	*Bg358*
Heenati6402	*Murungakayan3900*	*At306*	*Bg310*
Heenati4935	*Murungakayan3490*	*At308*	*Bg379-2*
Heenati4618	*Murungakayan3489*	*At353*	*Bg406*
Heenati3998	*Pokkali3922*	*At354*	*Bg374*
Heenati3707	*Pokkali3881*	*At362*	*Bw267-3*
Kaluwee3876	*Pokkali3573*	*At405*	*Bw312*
Kaluwee3728	*Pokkali3567*	*Bw361*	*Bw302*
Kalubalawee5479	*Pokkali3562*	*Bw363*	*Bw452*
Kalubalawee3976	*Polayal3661*	*Bw364*	*Ld253*
Kalubalawee5480	*Ratawee4580*	*Bw367*	*At309*
Kaluheenati7802	*Ratawee3555*	*Bw372*	*At373*
Kaluheenati5191	*Ratawee3525*	*Bw272-6b*	*At311*
Kaluheenati4991	*Ratawee3466*	*Bw400*	*IR64*
Kaluheenati3471	*Rathuwee3905*	*Bw451*	*IRRI200*
Kuruwee3982	*Rathuwee3473*	*Bw453*	*IRRI192*
Kuruwee3898	*Rathuheenati6250*	*Ld365*	
Kuruwee3552	*Rathuheenati4992*	*Ld368*	
Kuruwee3465	*Suduheenati7799*	*Ld371*	
Mawee8497	*Sudurusamba3671*	*Ld408*	
Mawee3704	*Sudurusamba4362*	*Bg403*	

Table 2:   Quantitative traits of rice used for morphological evaluation.

Parameter	Description
Days to 50% flowering	Number of days from seedling to 50% flowering
Plant height (cm)	Measured from the base of the shoot to the tip of the tallest leaf blade from the soil surface to the tip of the tallest panicle (awns excluded)
Number of tillers per plant	Total number of grain-bearing and non-bearing tillers
Number of effective tillers per plant	Panicle bearing tillers were counted
Number of leaves per plant	The number of fully expanded leaves on the main culm on the day of measurement at the maturity stage
Panicle length (cm)	Measured in cm from the rachis's base to the panicle's tip, with an average of five panicles per plant
Flag leaf length (cm)	Distance from the leaf base to the tip of the flag leaf
Flag leaf width (cm)	Measured at the most comprehensive position of the leaf blade of five leaves which were used to measure the length
Number of grains per panicle	Number of total grains of sampled panicle
Filled grain percentage	The number of filled grains of a sampled panicle was calculated as a percentage value.
100-grain weight at maturity	A random sample of 100 seeds, well-developed grains dried to 13% moisture content
Grain length (mm)	The actual measurement of length in millimeters
Grain width (mm)	The actual measurement of width in millimeters is the distance across the fertile lemma and the palea at the widest point.

Source: (IRRI, 2013; PGRC, 1999)

Path analysis was done using IBM SPSS AMOS statistical software (SPSS Inc, 2020) separately for the improved rice varieties and traditional rice accessions to understand the effect of different traits on the final grain yield. Path coefficients were calculated, and the direct and indirect effects were ranked according to the scale developed by Lenka & Misra (1973); 0.00 - 0.09 : negligible, 0.10 - 0.19 : low, 0.20 - 0.29 : moderate, 0.30 - 0.99 : high, and >1.00 : very high. This scale has been used for data interpretation of path analysis by Lule *et al.* (2012), Ranawake *et al.*, (2014) Owere *et al.*, (2015), and Eric *et al.*, (2016) for finger millet, Thangamani & Jansirani (2012) for brinjal, Chakraborty *et al.*, (2012) for corn, Janaki (2018) and Rekha *et al.*, (2019) for chilli, Sivakumar *et al.*, (2020) for turmeric, and Nagalakshmi *et al.*, (2020) for cowpea.

RESULTS AND DISCUSSION

Direct and indirect effect of the yield components on the final grain yield of traditional and improved rice genotypes

Path analysis was carried out to determine the direct and indirect effects of traits based on the path coefficients. The path diagram explains the relationship of traits (days to 50% flowering, plant height, number of total tillers per plant, number of effective tillers per plant, number of leaves per plant, panicle length, flag leaf length, flag leaf width, number of grains per panicle, filled grain percentage, 100-grain weight at maturity, seed length, and seed width) to yield. Direct and possible indirect relationships were designed in the path diagram (Figure 1).

The direct and indirect effects of all the traits finally decide the grain yield of rice (Rajamadhan *et al.*, 2011). Several studies have used path coefficient values to describe the contribution of yield-attributing characteristics to the final yield of rice genotypes (Kishore *et al.,* 2015; Dhurai *et al.,* 2016; Gour *et al.,* 2017; Rashmi *et al.,* 2017; Saleh *et al.,* 2020). The trait effect is defined by a β value in the path analysis, and the total effect of any trait on final grain yield is a collective effort in many ways (Table 3).

The results revealed that several traits behaved similarly on final grain yield in both traditional and improved rice genotypes. In contrast, a few traits acted differently in traditional rice accessions and improved rice varieties (Table 4).
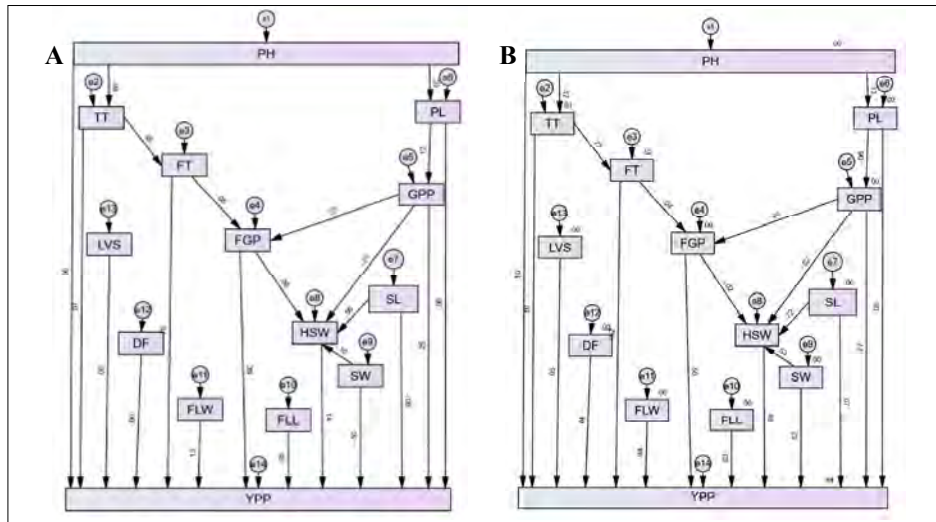
Figure 1: Path diagrams for traditional rice accessions (A) and improved rice varieties (B) obtained from SPSS AMOS software. PH: plant height, TT: number of total tillers per plant, FT: number of effective tillers per plant, LVS: number of leaves per plant, PL: panicle length, FLL: flag leaf length, FLW: flag leaf width, GPP: number of grains per panicle, FGP: filled grain percentage, HSW: hundred-grain weight, YPP: yield per plant, DF: number of days to 50% flowering, SL: grain length, SW: grain width

Table 3: The direct, indirect, and total effect of traits on single plant grain yield of traditional and improved rice genotypes as determined by path analysis.

	Yield per plant									
	Traditional rice β value					Improved rice β value				
Trait	D	I	T	DL	r	D	I	T	DL	r
FT	0.746	-0.02	0.726	High	.819**	0.519	-0.018	0.501	High	.569**
FGP	0.395	0	0.395	High	434**	0.496	-0.008	0.488	High	.446**
GPP	0.246	0.077	0.323	Moderate	.448**	0.266	-0.154	0.112	Moderate	.089
HSW	0.139	0	0.139	Low	-.022	0.403	0	0.403	High	.406**
FLW	0.127	0	0.127	Low	.185	0.037	0	0.037	Negligible	.065
TT	0.074	0.642	0.716	Negligible	.731**	0.095	0.361	0.456	Negligible	.411**
PH	0.063	-0.086	-0.023	Negligible	.019	0.101	-0.01	0.054	Low	.035
LVS	0.03	0	0.03	Negligible	.247	0.047	0	0.047	Negligible	.349*
FLL	-0.054	0	-0.054	Negligible	-.044	0.032	0	0.032	Negligible	.014
DF	-0.057	0	-0.057	Negligible	.105	0.037	0	0.037	Negligible	-.126
PL	-0.079	-0.039	-0.119	Negligible	-.079	0.053	-0.007	0.046	Negligible	-.043
SL	-0.085	0.078	-0.006	Negligible	-.015	0.066	0.289	0.355	Negligible	.270
SW	-0.095	-0.015	-0.11	Negligible	-.173	0.124	0.212	0.336	Low	.117

* Correlation is significant at the $\alpha = 0.05$ level (two-tailed)

** Correlation is significant at the $\alpha = 0.01$ level (two-tailed)

D: direct effects. I: indirect effects, T: total effects, DL: direct effect level, PH: plant height (cm), TT: number of total tillers per plant, FT: number of effective tillers per plant, LVS: number of leaves per plant, PL: panicle length (cm), FLL: flag leaf length (cm), FLW: flag leaf width (cm), GPP: number of grains per panicle, FGP: filled grain percentage, HSW: hundred-grain weight (g), YPP: yield per plant (g), DF: number of days to 50% flowering, SL: grain length (mm), SW: grain width (mm)

Table 4:   Direct effect traits on single plant grain yield of traditional and improved rice genotypes as determined by path analysis.

Direct effect	Traditional rice	Improved rice
High	FT (0.746), FGP (0.395)	FT (0.519), FGP (0.496), HSW (0.403)
Moderate	GPP (0.246)	GPP (0.266)
Low	HSW (0.139), FLW (0.127)	PH (0.101), SW (0.124)
Negligible	TT (0.074), PH (0.063), LVS (0.03), FLL (-0.054), DF (-0.057), PL (-0.079), SL (-0.085), SW (-0.095)	FLW (0.037), TT (0.095), LVS (0.047), FLL (0.032), DF (0.037), PL (0.053), SL (0.066)

PH: plant height (cm), TT: number of total tillers per plant, FT: number of effective tillers per plant, LVS: number of leaves per plant, PL: panicle length (cm), FLL: flag leaf length (cm), FLW: flag leaf width (cm), GPP: number of grains per panicle, FGP: filled grain percentage, HSW: hundred-grain weight (g), YPP: yield per plant (g), DF: number of days to 50% flowering, SL: grain length (mm), SW: grain width (mm)

Traits that similarly and directly affected the grain yield of traditional and improved rice genotypes

Effective tillers and filled grain percentage directly affected grain yield similarly in traditional and improved rice genotypes, and the magnitude of the effect was high. Grains per panicle also similarly affected the grain yield of both rice genotypes, and the effect was moderate. In a similar study, Kishore *et al.* (2015) reported that the number of grains per panicle, hundred seed weight, and the number of effective tillers per plant showed high direct effects on the grain yield in a positive direction. In addition, Roy *et al.* (2015) reported the positive impact of effective tiller number per plant on the final yield. Similar findings have been obtained in many studies in rice (Oladosu *et al.,* 2018; Limbongan *et al*., 2023; Reddy *et al*., 2023).

The effect of the total tiller number, number of leaves per plant, flag leaf length, days to 50% flowering, panicle length and grain length on the final grain yield was negligible for both traditional and improved rice genotypes (Table 4).

The direct effect of traits that differently affected grain yield of traditional and improved rice genotypes

The direct effect of hundred-grain weight on grain yield in traditional rice accessions was low. However, the direct effect of hundred-grain weight on improved rice varieties was high. The reason would be that improved rice varieties have already been modified for desirable seed characteristics such as plumpness or fullness. The traditional rice seeds have yet to be enhanced for such characteristics. Further, the effect of the flag leaf width on final grain yield was negligible in improved rice varieties. At the same time, it was low (effect on yield was more than in improved rice varieties) in traditional rice accessions. This is because improved rice varieties have been architecturally designed for optimum yield so that the canopy, including flag leaf, contributes to a higher yield. Still, traditional rice accessions still need to be architecturally designed for optimum yield, and the contribution of flag leaf to the grain yield would be naturally more significant than the effect of flag leaf on improved rice varieties. The effect of seed width on the final grain yield of traditional and improved rice genotypes was negligible and low, respectively. Desirable seed characteristics, such as plumpness, could be better in traditional rice accessions. Even though the seed width is broader in some traditional rice accessions, their contribution to the seed weight may less if seed plumpness is poor. Here, the effect of seed width on the final yield of traditional rice accessions is negligible. At the same time, it is low (effect is more than for traditional rice accessions) in improved rice varieties. The effect of hundred seed weight was low on the final yield of traditional rice accessions, and high for improved rice accessions (Table 4). This may be due to the same phenomenon explained for seed width.

The effect of plant height on final grain yield in traditional rice accessions was negligible and there was a low effect of the plant height on the final grain yield of improved rice varieties (Table 4). This difference can be justified by the difference in their plant architecture. The plant height of the improved rice varieties was designed for the optimum yield during the green revolution. As a result, the sink-source relationship of improved rice varieties has been optimized. Still, the sink-source relationship of traditional rice cultivars has yet to be optimized, and partitioning dry matter for seed production is inefficient.

Table 5: Traits whose direct effect was high, moderate, or low on yield, and the effect of other traits on them (indirect effects on yield)

Traits with indirect effect (Traditional)	Trait	D	I	T	DL	Traits with indirect effect (improved)		D	I	T	DL
*FT	PH	-	-0.068	-0.068		*FT	PH	-0.083	-0.083	0	Negligible
(High)	TT	0.884	0	0.884	High	(High)	TT	0.72	0	0.72	High
*FGP	PH	-	-0.005	-0.005		*FGP	PH		0.005	0.005	
(High)	PL	-	-0.033	-0.033		(High)	PL	-	0.019	0.019	
	TT	-	-0.045	-0.045			TT	-	-0.026	-0.026	
	GPP	0.27	0	0.27	Moderate		GPP	-0.296	0	-0.296	Moderate
	FT	-0.051	0	-0.051	Negligible		FT	-0.037	0	-0.037	Negligible
*GPP	PH	-	-0.031	-0.031		*GPP	PH		0.008	-0.008	Negligible
(Moderate)	PL	-0.121	0	-0.121	Low	(Moderate)	PL	-0.064	0	-0.064	Negligible
						*HGW	PL	-	0.001	0.001	
						(High)	GPP	-0.024	0.006	-0.018	Negligible
							FT	-	0.001	0.001	
							SW	0.526	0	0.526	High
							SL	0.716	0	0.716	High
							FGP	-0.019	0	-0.019	

D: direct effects. I: indirect effects, T: total effects, DL: direct effect level, PH: plant height (cm), TT: number of total tillers per plant, FT: number of fertile tillers per plant, LVS: number of leaves per plant, PL: panicle length (cm), FLL: flag leaf length (cm), FLW: flag leaf width (cm), GPP: number of grains per panicle, FGP: filled grain percentage, HSW: hundred seed weight (g), YPP: yield per plant (g), SL: grain length (mm), SW: grain width (mm)

* Traits with direct effect on yield

Hence, according to the results obtained by this study, considering the direct effect of yield attributing traits, effective tillering ability, hundred seed weight, filled grain percentage, and grains per panicle should be considered as selection criteria for breeding for higher yield in improved rice varieties.

The traits that recorded a high, moderate, or low indirect effect on the yield via other traits (effective tiller number, filled grain percentage, grains per panicle, hundred seed weight) with direct effect.

Total tiller number, grains per panicle, panicle length, grain width and grain length indirectly affected the yield via another trait of traditional or improved rice genotypes (Table 5). According to Shahidullah *et al.* (2009), the direct effect of the number of tillers on the final yield was insignificant. However, the number of panicles recorded an indirect effect through the number of tillers on final yield, and the greatest direct effect on the yield was recorded from the total number of tillers. Oladosu *et al.,* (2018) also reported a maximum indirect effect on yield by the tillers per hill through the number of panicles per

plant. Grain-related characteristics directly or indirectly affect the final yield of rice.

Total tiller number per plant

The total tiller number per plant affected the effective tiller number equally in both traditional and improved rice genotypes, and the magnitude of the effect was high.

Grains per panicle

Grains per panicle moderately and positively affected filled grain percentage in traditional rice genotypes. Contrarily, grains per panicle moderately and negatively affected the filled grain percentage in improved rice varieties. Since improved rice varieties fill the seed to its maximum plumpness, filling becomes difficult when the number of seeds increases. However, in traditional rice accessions, since they do not fill the seeds to their maximum capacity, they can still fill many grains at least partially. The number of grains per panicle recorded a positive indirect effect through panicle number in the

study reported by Al-Musawi *et al.* (2019). However, Laxmi *et al.* (2023) reported a positive direct impact of grains of panicle on final yield.

Panicle length

There was a low effect of panicle length on grains per panicle in traditional rice accessions, while the effect was negligible in improved rice varieties. However, Reddy *et al.,* (2023) reported a positive direct effect of panicle length on rice yield. The panicle length of traditional rice recorded a broad variation, while it was comparatively narrow in improved rice varieties (unpublished data by the author). A limited variation in panicle length may not affect grains per panicle in improved rice varieties. The number of spikelets within a panicle exhibited variability in traditional rice accessions corresponding to variations in panicle length. This may create such a difference in the effect of panicle length on grains per panicle in traditional and improved rice genotypes (Table 5).

Grain width and grain length

Grain width and length highly affected hundred grain weight in improved rice varieties, and the effect was negligible in traditional rice accessions. This difference may be due to the plumpness in improved rice varieties, as opposed to the partial filling nature in traditional rice accessions.

Days to 50% flowering showed a negligible effect on the final yield of traditional and improved rice genotypes (Table 4). However, Al-Musawi *et al.,* (2019) found a weak negative direct effect of days to 50% flowering on final yield, and a significant indirect effect on days to 50% flowering on final yield through plant height.

CONCLUSIONS

The direct and indirect effects of traits on the final grain yield of traditional rice accessions and improved rice varieties were different. The number of effective tillers per plant, filled grain percentage and grains per panicle were the traits that directly affected grain yield of both traditional and improved rice genotypes. Hundred seed weight had a high direct effect on the grain yield of improved rice varieties. Total tillers per plant and grains per panicle were the common traits that recorded a high indirect effect on yield in traditional and improved rice genotypes. Grain length and width indirectly but highly affect the yield of improved rice varieties, The direct

and indirect effect of traits in traditional and improved rice genotypes on yield can be utilized as selection for breeding materials for higher yield.

REFERENCES

Agahi K., Fotokian M.H. & Farshadfar E. (2007). Correlation and path coefficient analysis for some yield-related traits in rice genotypes (*Oryza sativa* L.). *Asian Journal of Plant Sciences* **6**: 513–517.
DOI: https://doi.org/10.3923/ajps.2007.513.517

Al-Musaw B. & Al-Anbari M. (2019). Estimating path coefficient analysis for some quantitative traits in Rice (*Oryza sativa* L.) genotypes. *Revista Bionatura 2022* **7**(4): 39.
DOI: https://doi.org/10.21931/RB/2022.07.04.39

Baye A., Berihun B., Bantayehu M. & Derebe B. (2020). Genotypic and phenotypic correlation and path coefficient analysis for yield and yield-related traits in advanced bread wheat (*Triticum aestivum* L.) lines. *Cogent Food and Agriculture* **6**(1): 1752603.
DOI: https://doi.org/10.1080/23311932.2020.1752603

Chakraborty M. & Sah R.P. (2012). Genetic component in baby corn (*Zea mays* L.). *Plant Archives* **12**(1)**:** 291–294.

Dalrymple D.G. (1985). The development and adoption of high-yield varieties of wheat and rice in developing countries. *Agricultural and Applied Economics Association* **1**: 1–15.
DOI: https://doi.org/10.22004/ag.econ.278578

Davies W.P. (2003). An historical perspective from the green revolution to the gene revolution. *Nutrition Reviews* **61**(6):124–134.
DOI: https://doi.org/10.1301/nr.2003.jun.s124-s134

DOA Sri Lanka (2013). *New Fertilizer Recommendations*. Available at *https://doa.gov.lk/rrdi_fertilizerrecomendation/*, Accessed 30 November 2023.

Dhurai S., Reddy D.M. & Ravi S. (2016). Correlation and path analysis for yield and quality characters in rice (*Oryza sativa* L.). *Rice Genomics and Genetics* **7**(4): 1–6.

Eric M.O., Pangirayi T., Paul S., Mwangi G. & Abhishek R. (2016). Correlations, path coefficient analysis and heritability for quantitative traits in finger millet landraces. *Philippine Journal of Science* **145**(02): 197–208.

Gour L., Koutu G.K., Singh S.K., Patel D.D., Shrivastava A. & Singh Y. (2017). Genetic variability, correlation and path analyses for selection in elite breeding materials of rice (*Oryza sativa* L.) genotypes in Madhya Pradesh. *The*

Pharma Innovation Journal **6**(11)**:** 693–696.

Immanuel S.C., Nagarajan P., Thiyagarajan K., Bharathi M. & Rabindran R. (2011). Genetic parameters of variability, correlation and path coefficient studies for grain yield and other yield Attributes among rice blast disease resistant genotypes of rice (*Oryza sativa* L.). *African Journal of Biotechnology* **10**: 3322–3334.
DOI: https://doi.org/10.5897/AJB10.2575

IRRI (2013). *Standard Evaluation System For Rice.* Available at *http://www.knowledgebank.irri.org/images/docs/rice-standard-evaluation.*

Janaki M., Ramana C.V., Naidu L.N., Babu J.D., Rao K.K. & Krishna K.U. (2018). Character association and path co-efficient analysis studies on yield and yield attributing characters in chilli (*Capsicum annuum* L.). *Electronic Journal of Plant Breeding* **9**(2): 558–563.

Khush G. (2003). Productivity Improvements in Rice. *Nutrition Reviews* **61**: 114–116.
DOI: https://doi.org/10.1301/nr.2003.jun.s114-s116

Khush G.S. (1999). Green revolution: Preparing for the 21st century. *Genome* **42**(4): 646–655.
DOI: https://doi.org/10.1139/g99-044

Kishore N., Srinivas T., Nagabhushanam U., Pallavi M. & Sameera S. (2015). Genetic variability, correlation and path analysis for yield and yield components in promising rice (*Oryza sativa* L.) genotypes. *SAARC Journal of Agriculture* **13**: 99–108.
DOI: https://doi.org/10.3329/sja.v13i1.24184

Laxm T.G., La G.M. & Bara B.M. (2023). Direct and indirect effects of yield contributing traits in rice (*Oryza sativa* L.). *International Journal of Plant & Soil Science* **35**(19): 2091–2099.
DOI: https://doi.org/10.9734/ijpss/2023/v35i193760

Lenka D. & Mishra B., (1973). Path coefficient analysis of yield in rice varieties. *Indian Journal of Agricultural Science* **43**(4):376.

Li R., Li M., Ashraf U., Liu S. & Zhang J. (2019). Exploring the relationships between yield and yield-related traits for rice varieties released in china from 1978 to 2017. *Frontiers in Plant Science* **10**: 00543.
DOI: https://doi.org/10.3389/fpls.2019.00543

Limbongan Y.L., Parari T.Y. & Sjahril R. (2023). Performance of agronomic characteristics and path analysis of the highland new plant types of rice. *IOP Conference Series: Earth and Environmental Science* **1230**(1):012122.
DOI: https://doi.org/10.1088/1755-1315/1230/1/012122

Lule D., Tesfaye K., Fetene M. & De Villiers S. (2012). Inheritance and association of quantitative traits in finger millet (*Eleusine coracana* Subsp. Coracana) landraces collected from eastern and south eastern Africa. *International Journal of Genetics* **2**(2): 12-21.

Mecha B., Alamerew S. & Assefa E. (2017). Correlation and path coefficient studies of yield and yield associated traits in bread wheat (*Triticum aestivum* L .) genotypes. *Advances in Plants and Agriculture Research* **6**(5):128–136.
DOI: https://doi.org/10.15406/apar.2017.06.00226

Monna L., Kitazawa N., Yoshino R., Suzuki J., Masuda H., Maehara Y., Tanji M., Sato M., Nasu S. & Minobe Y. (2002). Positional cloning of rice semi dwarfing gene, sd-1: Rice "green revolution gene" encodes a mutant enzyme involved in gibberellin synthesis. *DNA Research* **9**:11–17.
DOI: https://doi.org/10.1093/dnares/9.1.11

Nagalakshmi R., Kumari R.U. & Kumar R.A. (2020). Correlation and path analysis in cowpea (*Vigna unguiculata unguiculata* (L.) Walp). *The Bioscan* **15**(3): 397–401.

Oladosu Y., Rafii M.Y., Magaji U., Abdullah N., Miah G., Chukwu S.C., Hussin G., Ramli A. & Kareem I. (2018). Genotypic and phenotypic relationship among yield components in rice under tropical conditions. *BioMed Research International* **15**(2018): 8936767.
DOI: 10.1155/2018/8936767

Owere L., Tongoona P., Derera J. & Wanyera N. (2015). Variability and trait relationships among finger millet accessions in Uganda. *Uganda Journal of Agricultural Sciences* **16**(2):161–176.

PGRC (1999). *Characterization Catalogue On Rice (Oryza sativa L.) Germplasm*. Plant Genetic Resources Centre, Sri Lanka.

Rajamadhan R., Eswaran R. & Anandan A. (2011). Investigation of correlation between traits and path analysis of rice (*Oryza sativa* L.) grain yield under coastal salinity. *Electronic Journal of Plant Breeding* **2**:538–542.

Ranawake A.L., Amarasinghe U.G.S. and Senanayake S.G.J.N. (2014). Ranking traditional rice cultivars based on yield and some morphological traits using path analysis, multi-criteria decision making model and compromise programming. *American Journal of Experimental Agriculture* **4**(12):1579–1592.
DOI: https://doi.org/10.9734/AJEA/2014/11325

Rasel M., Hassan L., Hoque M.I.U. & Saha S.R. (2018). Estimation of genetic variability, correlation and path coefficient analysis in local landraces of rice *(Oryza sativa* L.) for the improvement of salinity tolerance. *Journal of Bangladesh Agriculture University* **16**: 41–46.
DOI: https://doi.org/10.3329/jbau.v16i1.36479

Rashmi D., Saha S., Loitongbam B., Singh S. & Singh P.K. (2017). Genetic variability study for yield and yield components in rice (*Oryza sativa* L.). *International Journal of Agriculture Environment and Biotechnology* **10**(2): 171–176.
DOI: https://doi.org/10.5958/2230-732x.2017.00020.1

Reddy V.R., Rani T.S., Kumari C.A. & Soundharya B. (2023). Correlation and path analysis for yield and yield attributes in maintainer lines of rice (*Oryza sativa* L.). *International Journal of Bio-Resource & Stress Management* **6**(10): 2369–2376.
DOI: https://doi.org/10.20546/ijcmas.2017.610.279

Rekha G.K., Ramana C.V., Naidu L.N., Umajyothi K., Paratpararao M. & Sasikala K. (2019). Character association studies for yield and quality traits in chilli. *International Journal of Current Microbiological & Applied Sciences* **8**(8): 569–574.
DOI: https://doi.org/10.20546/ijcmas.2019.808.068

Roy R.K., Majumder R.R., Sultana S., Hoque M.E. & Ali, M.S. (2015). Genetic variability, correlation and path coefficient analysis for yield and yield components in transplant aman rice (*Oryza sativa* L.). *Bangladesh Journal of Botany* **44**: 529–535.
DOI: https://doi.org/10.3329/bjb.v44i4.38566

Saleh M.M., Salem K.F.M. & Elabd A.B. (2020). Definition of selection criterion using correlation and path coefficient analysis in rice (*Oryza sativa* L.) genotypes. *Bulletin of the National Research Center* **44**: 143.
DOI: https://doi.org/10.1186/s42269-020-00403-y

Samonte S.O.P., Wilson L.T. & McClung A.M. (1998). Path analyses of yield and yield-related traits of fifteen diverse rice genotypes. *Crop Science* **38**(5):1130–1136.
DOI: https://doi.org/10.2135/cropsci1998.0011183x003800050004x

Shahidullah S.M., Hanafi M.M., Ashrafuzzaman M., Ismail M.R. & Salam M.A. (2009). Tillering dynamics in aromatic rice genotypes. *International Journal of Agriculture and Biology* **11(5)**: 509–514.

Singh R., Yadav V., Mishra D.N. & Yadav A. (2018). Correlation and path analysis in rice (*Oryza sativa* L.). *Journal of Pharmacognosy and Phytochemistry* **2018** (7):2084–2090

Sivakumar V., Ravindrakuma, C.C., Chandrasekhar C. & Bhagavan B.V.K. (2020). Character association and path coefficient analysis studies on yield and yield attributing characters in Turmeric (*Curcuma longa* L.) *Inernational Journal of Chemical Studies* **8**: 2587–2589.

Spielmeyer W., Ellis M.H. & Chandler P.M. (2002). Semidwarf (sd-1), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. *Proceedings of National Academiy of Science of U.S.A.* **99**: 9043–9048.
DOI: https://doi.org/10.1073/pnas.132266399

SPSS Inc, (2020). Software, IBM SPSS Statistics for Windows (Version 27.0). SPSS Inc. Chicago, USA.

Sudeepthi K., Srinivas T., Kumar B.N.V.S.R.R., Jyothula D.P.B. & Umar S.N. (2020). Assessment of genetic variability, character association and path analysis for yield and yield component traits in rice (*Oryza sativa* L.). *Electronic Journal of Plant Breeding* **11**: 65–69.
DOI: https://doi.org/10.37992/2020.1101.026

Thangamani C. & Jansirani P. (2012). Correlation and path coefficient analysis studies on yield and attributing characters in brinjal (*Solanum melongena* L.). *Electronic Journal of Plant Breeding* **3**: 939–944.

Ulukan H., Guler M. & Keskin S. (2003). A Path Coefficient Analysis Some Yield and Yield Components in Faba Bean (*Vicia faba L.*) Genotypes. *Pakistan Journal of Biological Sciences* **6**:1951–1955.
DOI: https://doi.org/10.3923/pjbs.2003.1951.1955

Valenzuela S. & Bachmann I. (2017). Path Analysis, In: *The International Encyclopedia of Communication Research Methods,* pp.1–9. Wiley, USA.
DOI: https://doi.org/10.1002/9781118901731.iecrm0181

RESEARCH ARTICLE

Dairy Science

A preliminary study on milk composition of three buffalo breeds located in Polonnaruwa, Sri Lanka

SP Gunathilake, DA Gayanjalee, DAS Prathiraja, PRM Buddhika, WADV Weerathilake and DMD Rasika[*]

Department of Livestock and Avian Sciences, Faculty of Livestock, Fisheries and Nutrition, Wayamba University of Sri Lanka,Makandura, Gonawila, Sri Lanka.

Abstract: Buffaloes have great potential to be developed as one of the main dairy animals in Sri Lanka because of their higher adaptability to local conditions compared to European dairy cattle breeds. Processing dairy products such as curd, cheese, and yoghurt from buffalo milk has several advantages due to its greater quantities of total solids, solid-nonfat, and protein. However, milk composition and coagulation properties largely depend on the breed. Therefore, this research aimed to investigate the milk composition of three major buffalo breeds in Sri Lanka. The milk of the Lanka buffalo contained 18.1% total solids, 10.7% solid non-fat (SNF), 7.5% fat, 4.9% protein, 4.9% lactose, and 0.9% minerals. The milk of the Murrah crossbreed contained 15.4% total solids, 10.1% SNF, 5.3% fat, 4.2% protein, 5.1% lactose, and 1.1% minerals. The milk of the Nili-Ravi crossbreed contained 16.6% total solids, 10.2% SNF, 6.4% fat, 4.1% protein, 4.9% lactose, and 1.1% minerals. The results indicate that milk composition varies among the breeds in which the milk of the Lanka buffalo has significantly higher amounts of total solids, SNF, and fat. Protein, lactose, and mineral contents do not vary signiticantly.

Keywords: Buffalo milk, Lanka buffalo, milk fat, milk protein, Murrah, Nili-Ravi.

INTRODUCTION

The milk production of the world is supplied by only a few species of dairy animals. Among them, buffalo (*Bubalus bubalis*) is the second largest milk-producing species after cattle. Buffaloes produce approximately 15% of the world's milk production. In developing countries, one-third of the total milk production comes from species other than cattle (*e.g.*, 40% in Asia and 23% in Africa), mostly from buffaloes (FAO, 2023). South Asia is inhabited by approximately 78% of the global buffalo population and currently produces about 100 million metric tons of buffalo milk which accounts for about 93% of the world's buffalo milk production (Siddiky & Faruque, 2018). The share of buffalo milk is approximately 51% of the total milk production in South Asia. The contribution of buffalo milk to the total milk production of Nepal, Pakistan, and India is 67%, 60%, and 51%, respectively (Siddiky & Faruque, 2018).

Buffaloes are generally distributed in Asia, the Middle East, and Europe and are mainly divided into two sub-species: River-type and Swamp-type (Abdel-Hamid *et al*., 2023). River-type buffaloes have been developed mainly in the Indian subcontinent and are used for milk production. Swamp-type buffaloes are mainly used for draught and meat production but are poor milk producers (Han *et al*., 2007). Buffaloes can be used as dairy animals mainly in wet tropical regions in the world including South Asia (Bittante *et al*., 2022; FAO, 2023). They

[*] Corresponding author (diland@wyb.ac.lk; https://orcid.org/0000-0001-5486-1744)

are unique in their way of surviving during very hard nutritional conditions as well as under less beneficial management. This makes them well suited not only to tropical environments but also to where specialized dairy cows have difficulties in adapting to the conditions (Khan, 2002; Hallqvist, 2019; Bittante *et al*., 2022). Due to these beneficial traits, buffaloes are considered the dairy animal for the 21st century. There is a large diversity in buffalo genetic resources and South Asia is home to many high-yielding buffalo breeds such as Murrah and Nili-Ravi (Siddiky & Faruque, 2018).

Buffalo milk has a growing demand and popularity because of its flavour, and high content of protein, fat, vitamins, lactose, total solids, and other nutrients compared to cow milk (Han *et al*., 2007; Bittante *et al*., 2022; Abdel-Hamid *et al*., 2023). Therefore, buffalo milk receives increasing research interest and investment opportunities owing mainly to its attractive nutrient content (Han *et al*., 2007). Milk processing allows milk to be preserved for extended periods ranging from days to years while helping reduce the incidence of food-borne diseases and making it consumable for people with lactose intolerance, which affects about two-thirds of the human population (Lomer *et al.,* 2008*;* Bittante *et al*., 2022; FAO, 2023). Buffalo milk has been utilized to produce a large variety of commercial dairy products and numerous novel dairy-based products are also being researched. These products include yoghurts (Hekmat & Reid, 2006; Akgun *et al*., 2016; 2018), probiotic and symbiotic yoghurts (Han *et al*., 2012; Nguyen *et al*., 2014; Ehsani *et al*., 2016; Yapa *et al*., 2023), fortified yoghurts with different additives such as mango pulp and soymilk (Kumar & Mishra, 2007), apple fruit and honey (Ghadge *et al*., 2008), and bael fruit pulp (Yapa *et al*., 2023), dairy beverages (Silva *et al*., 2020), ice cream (Roy *et al*., 2021; Sert *et al*., 2021) and curd (Priyashantha *et al*., 2021). Differences found between buffalo milk and cow milk provide unique characteristics and processing capabilities to the dairy products processed from buffalo milk. Milk produced by buffaloes has the highest total solids and fat content, as well as the greatest casein/protein ratio of the six major dairy species (Roy *et al*., 2021; Bittante *et al*., 2022). Because of this, buffalo milk was reported to yield more fresh cheese (25%) than ewe (22.9%), cow (15.4%), dromedary (13.8%), and goat (11.9%) milk. This is largely due to the greater fat and protein contents of buffalo milk in addition to the greater recovery of fat (88.2%) in the curd (Bittante *et al*., 2022). Moreover, the coagulation and curd-firmness patterns of buffalo milk were reported to be excellent. Buffalo milk had the highest recoveries in the curd of fat and energy, and also had the highest cheese-making efficiency among

the 6 major dairy species, retaining as much as 76% of milk energy in cheese (Bittante *et al*., 2022). More importantly, unlike other species, the cheese-making efficiency of buffaloes was found to be unaffected by the parity but increased with the advancing stage of lactation (Bittante *et al*., 2022). Higher cheese yield from buffalo milk partly compensates lower daily milk yield of buffaloes compared to specialized dairy cattle breeds (Sun *et al*., 2014).

The composition of milk from dairy buffalos is of major interest to milk producers, processors, and consumers, since the concentration of certain milk components influences the pricing policy in the market and directly affects the economy of milk production as well as the economic conditions of the farmers (Boro *et al*., 2018). Knowing the composition is also useful for the manufacture of a wide variety of specialty dairy products (Han *et al*., 2012). Milk composition affects milk gelation characteristics, yield and quality of cheese (Glantz *et al*., 2010), foaming properties of milk (Huppertz, 2010), and milk processability (O'Brien *et al*., 2002). The general composition of milk is an essential consideration for variation in milk coagulation properties and is important for selecting desired dairy breeds for manufacturing cheese and yoghurt-like products (Abeykoon *et al*., 2016). Although Sri Lanka has favorable climatic conditions and other necessary resources for buffalo production, its true potential has not yet been realized. Buffaloes can effectively be utilized to uplift milk production in the country where the current production level can only be sufficient to fulfill about 40% of the demand. When the other countries in the region (e.g., Nepal, Pakistan, and India) managed to produce more than 50% of their total milk production from buffaloes, Sri Lanka only managed to produce just over 15% accounting for approximately 56 million liters in 2021 (DAPH, 2021), which is far below its true potential. Although the detailed chemical composition of milk from different buffalo breeds has been studied in many countries considering its importance from an industrial point of view, such a detailed study is absent in Sri Lanka to the best of our knowledge. One comprehensive report is available on milk coagulation properties and milk protein genetic variants of three cattle breeds/types in Sri Lanka (Abeykoon *et al*., 2016). In relation to buffaloes, milk yields and lactation performances of some major buffalo breeds in Sri Lanka have been documented (Ibrahim & Jayatileka, 2000; Nafees & Jeyamalar, 2005; Dematawewa & Dekkers, 2014; Charlini & Sinniah, 2015). In one study, only the milk composition of the local/Lanka buffalo has been documented (Horadagoda, 1990) and two more studies reported the milk composition of buffalo without

specifying the breeds (Mahanama, 2008; Randiwela *et al*., 2018). In this context, the current study aimed to analyze the milk composition of three major buffalo breeds in Sri Lanka namely local/Lanka buffalo, and the crossbreeds of Murrah and Nili-Ravi.

MATERIALS AND METHODS

Sample collection

Milk samples from the indigenous breed/Lanka buffalo and Murrah (Murrah × Local) and Nili-Ravi (Nili-Ravi × Local) crosses were collected from a semi-intensively managed buffalo farm located in the Wadigawewa area in the Polonnaruwa District of Sri Lanka. This farm was selected since it reared all the breeds and crosses considered in the current study. Collecting all the milk samples from the same farm allowed us to make fair comparisons among the breeds as feeding and other management practices are the same. If the samples were collected from several farms, feeding, and management practices may be different from farm to farm leading to greater compositional changes in milk obtained from different farms. Since the milk composition is also influenced by the season, parity number, and the stage of lactation, the milk samples for the current study were obtained from the buffaloes in the same parity number (3 or 4) and stage of lactation (mid-lactation, between 100 – 200 d of milking) during the month of October 2022. Accordingly, milk samples for analysis were collected from 10 lactating buffaloes from each breed (altogether 30 buffaloes) during morning milking. All the samples were collected on the same day. Milk samples were collected into plastic vials, sealed, and stored under refrigerated conditions (~4°C) until analysis. All the steps were conducted hygienically to avoid any contaminations. Buffalo breeds were identified based on the morphological characteristics unique to each breed and confirmed by the breeding records of the animals.

Determination of milk components

The fat, protein, and solid non-fat (SNF) contents of the milk samples were determined by an automatic milk analyser (MilcoScope, Julie Z9 Automatic, Scope Electric) established at the Nestle Lanka Kurunegala Factory at Pannala. The values obtained by the equipment were validated against the standard AOAC methods to determine protein (method 939.02) and fat (method 2000.18) contents to confirm the accuracy. Milk lactose content was determined by the iodometric method. Briefly, 10 mL of milk portion was transferred to a pre-weighed graduated flask and weighed. The milk sample was then diluted with 50 mL of distilled water followed by the addition of 10 mL of Mayer's Reagent (freshly prepared by dissolving a mixture of 1.36 g of mercuric chloride and 5 g of potassium iodide in 100 mL of distilled water) to precipitate proteins and fats. Then 2 mL of 1 N sulphuric acid was added, shaken well, and the solution was filtered. Twenty-five millilitres (25 mL) of this filtrate was then mixed with 20 mL of 0.1 N Iodine solution and 30 mL of 0.1 N sodium hydroxide. The solution was left in the dark for 20 min. Then 4 mL of sulphuric acid was added and the excess iodine was titrated with a 0.1 N sodium thiosulphate solution. The starch solution was added at the point of light yellow colour and the titration continued until the solution turned from deep blue to colourless. The amount of lactose was determined by multiplying the expended volume of 0.1 N iodine by 0.01705. The mineral content was determined by AOAC Method 930.30 (AOAC, 2019). The total solids content was determined by summing up the values for fat and SNF.

Statistical analysis

Statistical differences in milk components among different breeds are detected by performing a series of one-way analyses of variance (ANOVA). Means were separated by the Tukey Test. All statistical analyses were performed using SPSS version 22.0 (SPSS Inc., Chicago, IL) with a significance level of $P < 0.05$.

RESULTS AND DISCUSSION

Physicochemical characteristics such as milk components and pH are important parameters for the physicochemical and nutritional attributes of milk (Çinar *et al*., 2019). Milk is composed of water, proteins, fat, lactose, vitamins, and minerals. Milk solids are composed of the four major milk components namely, protein, fat, lactose, and minerals. On the other hand, the milk solid non-fat (SNF) portion comprises protein, lactose, and minerals. These components perform different functions and physiological roles. Out of these, fat, protein, SNF, lactose, and mineral contents are the most economically important milk constituents (Boro *et al*., 2018). In the current study, we examined the differences in milk constituents among three major dairy buffalo breeds in Sri Lanka. The total solids, SNF, fat, protein, lactose and mineral contents of milk of Lanka buffalo and crosses of Murrah and Nili-Ravi are summarized in Table 1.

Table 1:   Major milk components of Lanka buffalo, and crossbreeds of Murrah and Nili-Ravi in Sri Lanka

Milk Component	Murrah cross		Nili-Ravi cross		Lanka Buffalo	
	Average (%)	Range (%)	Average (%)	Range (%)	Average (%)	Range (%)
Total solids	15.4 ± 1.9^a	12.9 - 17.6	16.6 ± 1.0^{ab}	14.8 - 18.3	18.1 ± 1.9^b	15.4 - 20.5
Solid non-fat (SNF)	10.1 ± 0.3^a	9.6 - 10.7	10.2 ± 0.3^a	9.9 - 10.7	10.7 ± 0.3^b	10.1 - 11.1
Fat	5.3 ± 1.6^a	3.2 - 6.9	6.4 ± 0.9^{ab}	4.9 - 7.5	7.5 ± 1.6^b	5.2 - 9.4
Protein	4.2 ± 0.5^a	3.7 - 5.1	4.1 ± 0.2^a	3.9 - 4.6	4.9 ± 0.6^a	4.0 - 5.5
Lactose	5.1 ± 0.3^a	4.5 - 5.5	4.9 ± 0.3^a	4.6 - 5.5	4.9 ± 0.2^a	4.5 - 5.2
Minerals	1.1 ± 0.3^a	0.3 - 1.5	1.1 ± 0.4^a	0.7 - 1.4	0.9 ± 0.4^a	0.4 - 1.7

Data are mean $\pm$ SD (n = 10 for each breed)

Averages with the same superscript letter in the same row did not show any significant difference (P > 0.05).

Total solid contents

The total solid content represents the solid portion of milk that is composed of milk fat, protein, lactose, minerals, and vitamins. The total solid content of milk varies according to the dairy species. For example, the total solid content of milk from buffalo, cattle, and goats is approximately 18%, 13%, and 12%, respectively (Bittante *et al.*, 2022). There is a positive correlation between total solid and fat contents in which higher fat contents always result in higher total solid contents (Tekelioglu *et al.*, 2010). The total solid content in milk has a crucial role in dairy products manufactured by removing water content (*e.g.*, whole milk powder) where higher total solid content results in proportionately greater quantities of the end product. In this sense, buffalo milk is advantageous over bovine milk since buffalo milk contains a greater content of total solids.

In the current study, the total solid content in the milk of individual animals ranged from 12.9% to 20.5%. The highest average total solid content (18.1 $\pm$ 1.9%) was found in the milk of Lanka buffalo. Milk from Murrah and Nili-Ravi crosses were 15.4 $\pm$ 1.9% and 16.6 $\pm$ 1.0%, respectively. The average total solid contents observed among the three breeds were within the range reported in the previous studies conducted in various geographical areas in the world including China, the USA, the Mediterranean region, and the Indian sub-continent (Han *et al.*, 2007; Sun *et al.*, 2014; Çinar *et al.*, 2019; Bittante *et al.*, 2022).

SNF content

The solid non-fat (SNF) portion of milk consists of all solids in milk (protein, lactose, minerals) other than fat. Milk with high SNF is valuable to the consumer for its flavour and nutritional value and to the manufacturer of dairy products, especially related to cheese yield (Roberts, 1987).

In our study, the SNF contents in the milk of the Murrah crosses and Nili-Ravi crosses were comparable (10.1 and 10.2, respectively). However, the milk of Lanka buffalo showed significantly higher SNF content (10.7 $\pm$ 0.3%) than that of the crossbreeds. These results suggest that the milk of the Lanka buffalo is more advantageous than the milk of the Murrah and Nili-Ravi crossbreeds in cheese making as it may provide a comparatively higher cheese yield, due to higher SNF content. However, to the best of our knowledge, no study has been conducted to examine the milk coagulation properties of the buffalo breeds in Sri Lanka. More interestingly, the milk of the all buffalo breeds tested in the current study showed considerably higher SNF contents compared to that of the indigenous cattle breeds of Sri Lanka, Thamankaduwa white (9.45%) and Lanka cattle (9.44%) as well as the major dairy breed Friesian (8.87%) (Abeykoon *et al.*, 2016).

Previous studies conducted elsewhere in the world showed SNF contents ranging from 8.8% to 11.7% depending on the breed. In our study, the SNF contents varied from 9.63% to 11.05% in the individual animals used in the study and this range was in agreement with the previous studies (Table 2). The SNF contents of the milk from pure Murrah and Nili-Ravi were 9.34 – 11.21% and 9.70 – 11.10%, respectively. If crossbreds are concerned, this ranged from 10.66% to 11.70%. Our results obtained for the crossbreeds were within the range of the pure breeds but slightly less than that of the crossbreeds. Most probably, these variations in SNF content may be due to the differences in the stage of lactation and the parity/ lactation number reported in different studies (Dubey *et al.*, 1997; Yadav *et al.*, 2013). However, the milk of the three breeds concerned in our study showed considerably

higher SNF content than that of the two other prominent buffalo breeds in the Indian sub-continent, Mehsana (9.13 ± 0.06%) and Surti (8.80 ± 0.07%) (Misra *et al*., 2008).

Milk fat content

Milk fat is often a major component of dairy products such as butter and related products, cream products, cheese, ice cream, infant milk formulas, and milk chocolates. Milk fat contributes to the structure, flavour, colour, mouth feel, texture, and functional behaviour of dairy products depending on the product category (Waldron *et al*., 2020). Milk fat percentage shows a greater variation than any other constituent in milk (Roberts, 1987). The fat content in buffalo milk varies depending on the stage of lactation and parity/lactation number (Shah *et al*., 1983).

In the current study, the milk fat content of the individual buffaloes ranged from 3.2% to 9.4%. Milk from Lanka buffalo showed the highest average milk fat content (7.5 ± 1.6 %) and the lowest was observed in the milk of the Murrah crossbreed (5.3 ± 1.6%). The average milk fat contents for the Murrah crossbreed and Nili-Ravi crossbreed reported in the current study were lower than the values reported for pure Murrah (6.57 – 7.82%) and Nili-Ravi (6.53 – 9.22%) buffaloes in the available literature (Table 2). More interestingly, the values observed for the Murrah crossbreed were also considerably lower than the Murrah crossbreeds reported in the other studies. For example, Han and colleagues reported milk fat contents of approximately 6.5 – 10.6% for first- and second-generation Murrah × Guangxi crossbreeds (Han *et al*., 2007). A similar observation was made for the Nili-Ravi crossbreed as well, while the milk fat content reported in the current study was considerably lower than the values reported for various other crossbreeds (Table 2). Perhaps this variation may be due to poor nutrition, and differences in diet, stage of lactation, parity/lactation number, and geographical location. For instance, the fat content in buffalo milk tends to get higher progressively from 5.5% in the first month to 7.5% in the tenth month of lactation, coinciding with the decreasing milk yield with the progression of lactation (Shah *et al*., 1983; Yadav *et al*., 2013). On the other hand, the first lactation is said to be superior in terms of milk fat than that of successive lactations (Sundaram & Harharan, 2013).

Among the three breeds tested in the current study, only the milk fat content of the Lanka buffalo is in agreement with the values (6.57 – 9.22%) in published literature (Table 2). Moreover, the fat content of the Lanka buffalo was found to be more than double that

of the Thamankaduwa white (3.84%) and Lanka/Local cattle (3.31%), two indigenous cattle breeds of Sri Lanka (Abeykoon *et al*., 2016).

Milk protein

Milk coagulation is the primary step involved in the development of most dairy products and the coagulation properties of milk are largely influenced by protein content and composition (Abeykoon *et al*., 2016). Proteins are among the least variable milk constituents and do not vary significantly during stages of lactation (Yadav *et al*., 2013). Casein accounts for approximately 80 - 82% of the proteins found in buffalo milk (Dubey *et al*., 1997; Bittante *et al*., 2022).

The protein content in the milk of individual buffaloes used in the current study ranged from 3.7% to 5.5% which is in agreement with the range of protein contents reported in the literature (3.46 – 5.78%, Table 2). In the current study, all three breeds had comparable protein contents (P > 0.05) although that in the Lanka buffalo's milk was slightly higher (4.9 ± 0.6%). The protein content in the milk of the Murrah crossbreed (4.2 ± 0.5%) was within the range (3.47 – 4.92%) reported for the Murrah pure breeds in other studies. However, it was lower compared to the Murrah crossbreeds (5.51 – 5.78%) and other crossbreeds (4.75 – 5.23%). Values obtained for Nili-Ravi crosses were also within the range reported for Nili-Ravi purebreds (3.89 – 5.14%) but lower than that of the crossbreeds. As the protein content is least varied depending on the stage of lactation and diet, these differences may be due to the breed of buffalo or lactation/parity number. For instance, protein levels are reported to be high in advanced parities (Yadav *et al*., 2013) and the third lactation is superior in terms of milk proteins (Sundaram & Harharan, 2013).

More interestingly, the average protein contents observed in the current study were higher than that reported for the local cattle breeds Thamankaduwa white (3.48%) and Local/Lanka cattle (3.47%) suggesting that buffalo milk contains more milk proteins than bovine milk (Abeykoon *et al*., 2016).

Lactose

Lactose is the main carbohydrate in milk and drives the milk yield of the animal being a main determinant of the osmotic pressure of milk. Physical and chemical properties of lactose such as comparatively low solubility and specific crystallization behavior play a major role in the properties and quality of many dairy products and dairy-based ingredients.

Table 2: Composition of buffalo milk from different breeds and crossbreeds around the world

Breed	Total solids (%)	Fat (%)	SNF (%)	Protein (%)	Lactose (%)	Minerals (%)	Country	Reference
Murrah	17.01 ± 0.05	7.65 ± 0.05	9.36 ± 0.02	3.81 ± 0.02	4.83 ± 0.01		India	Dubey et al., 1997
		7.19 – 8.63		3.46 – 3.56	4.36 – 4.60		India	Yadav et al., 2013
	16.99 ± 0.12	6.99 ± 0.10	10.01 ± 0.06	3.78 ± 0.03	5.37 ± 0.04		India	Sarkar et al., 2006
	17.0 ± 0.1	7.7 ± 0.1	9.4 ± 0.1	3.81 ± 0.02	4.83 ± 0.01		India	Sodhi et al., 2008
	19.03	7.82	11.21*	4.92	5.18		China	Zhou et al., 2018
	18.23 ± 1.8	7.17 ± 1.7	10.34 ± 0.2	4.91 ± 0.2	5.25 ± 0.2		China	Abdel-Hamid et al., 2023
	16.69 ± 1.22	6.57 ± 1.21	10.12*	4.27 ± 0.43	5.07 ± 0.13	0.79 ± 0.05	China	Han et al., 2007
	17.07 ± 1.22	6.86 ± 1.28	10.21*	4.75 ± 0.36	4.60 ± 0.52		China	Sun et al., 2014
Nili-Ravi	17.44	6.77	10.67*	4.54	5.28		China	Zhou et al., 2018
	17.29 ± 1.6	6.84 ± 1.2	9.70 ± 0.5	4.38 ± 0.2	5.27 ± 0.3		China	Abdel-Hamid et al., 2023
	16.85 – 17.84	9.22 – 9.65	10.14 – 10.35	3.89 – 4.70	4.83 – 5.17	0.79 – 0.89	Pakistan	Faraz et al., 2021
	17.14 ± 1.34	6.53 ± 1.28	10.61*	4.16 ± 0.20	4.56 ± 0.10	0.81 ± 0.03	China	Han et al., 2007
	19.80 ± 3.05	7.99 ± 1.57	11.18*	5.14 ± 0.73	4.74 ± 0.81			Sun et al., 2014
Mehsana	15.59 ± 0.18	6.46 ± 0.17	9.13 ± 0.06	3.87 ± 0.05			India	Misra et al., 2008
Surti	14.96 ± 0.21	6.17 ± 0.20	8.80 ± 0.07	3.93 ± 0.05			India	Misra et al., 2008
Mediterranean	17.93	7.71	10.22*	4.38	5.01	0.8	Italy	Bittante et al., 2022
	20.18 ± 1.2	8.98 ± 1.2	10.17 ± 0.5	5.08 ± 0.5	4.98 ± 0.5		China	Abdel-Hamid et al., 2023
Local/Lanka buffalo		7.45 ± 0.06	9.62 ± 0.46	5.16 ± 0.15	4.3 ± 0.02	0.72 ± 0.02	Sri Lanka	Horadagoda, 1990
F1 crossbreed	19.75 ± 2.29	8.81 ± 1.89	10.94*	5.23 ± 0.45	4.80 ± 0.22	0.88 ± 0.07	China	Han et al., 2007
F2 crossbreed	19.21 ± 1.56	7.90 ± 1.30	11.31*	5.10 ± 0.45	4.64 ± 0.54	0.85 ± 0.05	China	Han et al., 2007
Multi-crossbreed	18.22 ± 1.24	7.56 ± 0.90	10.66*	4.75 ± 0.53	4.61 ± 0.20	0.84 ± 0.06	China	Han et al., 2007
F1 crossbreed (Murrah × Guangxi)	20.05 ± 1.84	8.35 ± 1.72	11.7*	5.78 ± 1.14	4.59 ± 0.87		China	Sun et al., 2014
F2 crossbreed (Murrah × Guangxi)	20.22 ± 2.54	8.69 ± 1.86	11.53*	5.51 ± 0.46	4.45 ± 1.00		China	Sun et al., 2014

*Estimated using the available data

There are two main roles of lactose: it acts as an energy source for the lactic acid bacteria that are used for fermented dairy products and as a precursor for specific flavour components (Hettinga, 2019). The stage of lactation significantly affects the lactose content in milk where it is significantly increased up to 4 – 6 months of lactation and decreased thereafter in buffaloes (Shah *et al.*, 1983; Yadav *et al.*, 2013).

The lactose content of milk of individual buffaloes used in the current study ranged from 4.5% to 5.5%, which is in agreement with the lactose content reported for buffalo milk in previous studies (4.36 – 5.37%, Table 2). The average lactose content was the same between the Nili-Ravi cross and Lanka buffalo (4.9%). Slightly higher, average lactose content was observed in the milk of Murrah crossbreeds (5.1 ± 0.3%). However, no statistically significant difference was found among the three breeds. The lactose content observed for Murrah and Nili-Ravi crossbreeds were compatible with the values reported for corresponding pure breeds (4.36 – 5.37% for Murrah and 4.56 – 5.28% for Nili-Ravi) and crosses (4.45 – 4.80%) in the previous studies.

Mineral content

The mineral content of milk is important nutritionally. Milk is a good source of minerals including calcium, phosphorus, magnesium, iron, iodine, and sodium. The mineral contents of the milk from Murrah and Nili-Ravi crossbreeds, and Lanka buffalo were comparable to each other (P > 0.05) although that in the Lanka buffalo (0.9 ± 0.4 %) was slightly lower than the crossbreds (1.1 %). Unlike other milk constituents, the mineral content of buffalo milk has seldom been determined in previous studies. However, available data showed that the mineral content of the Murrah purebreds, Nili-Ravi purebreds, and crossbreds were in the range 0.66 – 0.79 %, 0.79 – 0.81 %, and 0.84 – 0.88 %, respectively. Therefore, the mineral contents of the milk from three different breeds tested in our study showed slightly higher average mineral contents. This may be due to the differences in animal feeding, period of sample collection during the year, and environmental conditions (Coni *et al.*, 1995).

CONCLUSION

Milk composition varies among the breeds. Milk of the Lanka/local buffalo contains significantly greater quantities of total solids, solid non-fat, and fat contents. Milk protein, lactose, and mineral contents are comparable among Lanka buffalo and crossbreeds of Murrah and Nili-Ravi buffaloes. Based on the greater

total solid, solid non-fat, and fat contents of the milk of the local buffalo, it may be more advantageous compared to the milk of the two crossbreeds in cheese, yoghurt, and curd production, due to higher curd yields.

REFERENCES

Abdel-Hamid M., Huang L., Huang Z., Romeih E., Yang P., Zeng Q. & Li L. (2023). Effect of buffalo breed on the detailed milk composition in Guangxi, China. *Foods* **12**(08): 1603.
DOI: https://doi.org/10.3390/foods12081603

Abeykoon C.D., Rathnayake R.M.C., Johansson M., Silva G.L.L.P., Ranadheera C.S., Lundh Å. & Vidanarachchi J.K. (2016). Milk coagulation properties and milk protein genetic variants of three cattle breeds/types in Sri Lanka. *Procedia Food Science* **6**: 348–351.
DOI: https://doi.org/10.1016/j.profoo.2016.02.070

Akgun A., Yazici F. & Gulec H. A. (2016). Effect of reduced fat content on the physicochemical and microbiological properties of buffalo milk yoghurt. *LWT - Food Science and Technology* **74**: 521–527.
DOI: https://doi.org/10.1016/j.lwt.2016.08.015

Akgun A., Yazici F. & Gulec H.A. (2018). The combined effect of probiotic cultures and incubation final pH on the quality of buffalo milk yogurt during cold storage. *Food Science and Nutrition* **6**(2): 492–502.
DOI: https://doi.org/10.1002/fsn3.580

AOAC. (2019). *Official Methods of Analysis of AOAC International*, 21st Edition. Oxford University Press, Oxford, UK.

Bittante G., Amalfitano N., Bergamaschi M., Patel N., Haddi M.L., Benabid H., Pazzola M., Vacca G.M., Tagliapietra F. & Schiavon S. (2022). Composition and aptitude for cheese-making of milk from cows, buffaloes, goats, sheep, dromedary camels, and donkeys. *Journal of Dairy Science* **105**(03): 2132–2152.
DOI: https://doi.org/10.3168/jds.2021-20961

Boro P., Debnath J., Kumar Das T., Chandra Naha B., Debbarma N., Debbarma P., Debbarma C., Suniti Bala Devi L. & Gynashwari Devi T. (2018). Milk composition and factors affecting it in dairy Buffaloes: A review. *Journal of Entomology and Zoology Studies* **340**(63): 340–343.

Charlini B.C. & Sinniah J. (2015). Performance of Murrah, Surti, Nili-Ravi buffaloes and their crosses in the intermediate zone of Sri Lanka. *Livestock Research for Rural Development*, **27**(03). 1–17.

Çinar M.U., Özsoy T., Beyzi S.B., Kaliber M. & Konca Y. (2019). Milk and fatty acid composition of anatolian water buffalo (*Bubalus bubalis*) from different provinces. *Buffalo Bulletin* **38**(01): 107–118.

Coni E., Bocca A., Ianni D. & Caroli S. (1995). Preliminary evaluation of the factors influencing the trace element content of milk and dairy products. *Food Chemistry* **52**(02): 123–130.
DOI: https://doi.org/10.1016/0308-8146(94)P4190-Q

Dematawewa C.M.B. & Dekkers J.C.M. (2014). Lactation curve modeling for Murrah and Surti buffalo breeds in Sri Lanka. *Proceedings of the 10th World Congress on Genetics Applied to Livestock Production*, 17-22 August. Vancouver, BC, Canada, pp. 449.

DAPH (2021). *Livestock Statistical Bulletin 2021*. Department of Animal Production & Health, Sri Lanka. Availble at *www.daph.gov.lk/web/images/content_image/news_bulletins/livestock_statistical/2021/Stat_Bulletin_2021.pdf*, Accessed 19 May 2023.

Dubey P.C., Suman C.L., Sanyal M.K., Pandey H.S., Saxena M.M. & Yadav P.L. (1997). Factors affecting composition of milk of buffaloes. *Indian Journal of Animal Sciences* **67**(09): 802–804.

Ehsani A., Banihabib E.K., Hashemi M., Saravani M., & Yarahmadi E. (2016). Evaluation of various properties of symbiotic yoghurt of buffalo milk. *Journal of Food Processing and Preservation* **40**(06): 1466–1473.
DOI: https://doi.org/10.1111/jfpp.12732

FAO. (2023). *Gateway to dairy Production and Products*. Food & Agriculture Organization of the United Nations, Rome, Italy. Available at *https://www.fao.org/dairy-production-products/production/dairy-animals/en/*, Accessed 20 May 2023.

Faraz A., Tauqir N.A., Waheed A. & Hameed A. (2021). Effect of exogenous oxytocin administration on the performance of lactating Nili Ravi Buffalo. *Iranian Journal of Applied Animal Science* **11**(03): 517–525.

Ghadge P.N., Prasad K. & Kadam P.S. (2008). Effect of fortification on the physico-chemical and sensory properties of buffalo milk yoghurt. *Electronic Journal of Environmental, Agricultural and Food Chemistry* **7**(5): 2890–2899.

Glantz M., Devold T.G., Vegarud G.E., Lindmark Månsson H., Stålhammar H., & Paulsson M. (2010). Importance of casein micelle size and milk composition for milk gelation. *Journal of Dairy Science* **93**(04): 1444–1451.
DOI: https://doi.org/10.3168/jds.2009-2856

Hallqvist J. (2019) The importance of buffalo milk in the curd manufacture of Sri Lanka. *Thesis*, Swedish University of Agricultural Sciences, Uppsala, Sweden. Available at h*ttps://Stud.Epsilon.Slu.Se*. http://urn.kb.se/resolve?urn=nbn:se:slu:epsilon-s-10693, Accessed 18 May 2023.

Han B.Z., Meng Y., Li M., Yang Y.X., Ren F.Z., Zeng Q.K. & Robert Nout M.J. (2007). A survey on the microbiological and chemical composition of buffalo milk in China. *Food Control* **18**(06): 742–746.
DOI: https://doi.org/10.1016/j.foodcont.2006.03.011

Han X., Lee F.L., Zhang L. & Guo M.R. (2012). Chemical composition of water buffalo milk and its low-fat symbiotic yogurt development. *Functional Foods in Health and Disease* **2**(04): 86–106.
DOI: https://doi.org/10.31989/ffhd.v2i4.96

Hekmat S. & Reid G. (2006). Sensory properties of probiotic yogurt is comparable to standard yogurt. *Nutrition Research* **26**(04): 163–166.
DOI: https://doi.org/10.1016/j.nutres.2006.04.004

Hettinga K.A. (2019). Lactose in the dairy production chain. In: *Lactose: Evoloutionary Role, Health Effects, and Applications* (eds. M. Paques & C. Linder), pp. 231–266. Academic Press, Cambridge, UK.
DOI: https://doi.org/10.1016/B978-0-12-811720-0.00006-4

Horadagoda A. (1990). Studies on the composition of indigenous buffalo milk in Sri Lanka. *M.Phil. thesis*, University of Peradeniya, Peradeniya, Sri Lanka.

Huppertz T. (2010). Foaming properties of milk: A review of the influence of composition and processing. *International Journal of Dairy Technology* **63**(04): 477–488.
DOI: https://doi.org/10.1111/j.1471-0307.2010.00629.x

Ibrahim M.N.M. & Jayatileka T.N. (2000). Livestock production under coconut plantations in sri lanka: cattle and buffalo production systems. *Asian-Australasian Journal of Animal Sciences* **13**(01): 60–67.
DOI: https://doi.org/10.5713/ajas.2000.60

Khan M.S. (2002). Animals that produce dairy foods: Water buffalo. In: *Encyclopedia of Dairy Science* (eds. H. Roginski, J.W. Fuquay & P.F. Fox), pp. 340–342. Academic Press, Cambridge, UK.

Kumar P. & Mishra H.N. (2007). Effect of mango pulp and soymilk fortification on the texture profile of set yoghurt made from buffalo milk. *Journal of Texture Studies* **34**(03): 249–269.
DOI: https://doi.org/10.1111/j.1745-4603.2003.tb01060.x

Lomer M.C.E., Parkes G.C., & Sanderson J.D. (2008). Review article: Lactose intolerance in clinical practice - Myths and realities. *Alimentary Pharmacology and Therapeutics* **27**(02): 93–103.
DOI: https://doi.org/10.1111/j.1365-2036.2007.03557.x

Mahanama N. (2008). Manufacture of fruit cheese from buffalo milk and papaya pulp manufacture of fruit cheese from buffalo milk and papaya pulp. *Thesis*, University of Sri Jayewardenepura, Nugegoda, Sri Lanka.

Misra S.S., Sharma A., Bhattacharya T.K., Kumar P., & Roy S.S. (2008). Association of breed and polymorphism of α- and α-casein genes with milk quality and daily milk and constituent yield traits of buffaloes (*Bubalus bubalis*). *Buffalo Bulletin* **27**(04): 294–301.

Nafees M.S.M. & Jeyamalar R. (2005). Performance analysis of dairy cattle and socio-economic analysis of dairy cattle farmers in Trincomalee District. *Proceedings of the* 5th Annual *Research Session of Eastern University.*

Chenkalady, Sri Lanka, pp. 230–240.

Nguyen H.T.H., Ong L., Lefèvre C., Kentish S.E., & Gras S.L. (2014). The microstructure and physicochemical properties of probiotic buffalo yoghurt during fermentation and storage: A comparison with bovine yoghurt. *Food and Bioprocess Technology* **7**(4): 937–953.
DOI: https://doi.org/10.1007/s11947-013-1082-z

O'Brien B., Ryan G., Meaney W.J., McDonagh D. & Kelly A. (2002). Effect of frequency of milking on yield, composition and processing quality of milk. *Journal of Dairy Research* **69**(3): 367–374.
DOI: https://doi.org/10.1017/S0022029902005605

Priyashantha H., Ranadheera C.S., Rasika D.M.D. & Vidanarachchi J.K. (2021). Traditional Sri Lankan fermented buffalo (*Bubalus bubalis*) milk gel (Meekiri): Technology, microbiology and quality characteristics. *Journal of Ethnic Foods* **8**(1): 1–17.
DOI: https://doi.org/10.1186/s42779-021-00105-4

Randiwela R., Mangalika U., Adikari A., Pathirana A. & Weeragalla W. (2018). PCR based assay for the detection of cow's milk adulteration in buffalo curd. *International Journal of Livestock Research* **8**(5): 67–79.
DOI: https://doi.org/10.5455/ijlr.20180424043903

Roberts H.A. (1987). How important is extra solids-not-fat in your milk? *Kansas Agricultural Experiment Station Research Reports* **0**(2): 74–76.
DOI: https://doi.org/10.4148/2378-5977.3015

Roy S., Hussain S.A., Prasad W.G. & Khetra Y. (2021). Effect of emulsifier blend on quality attributes and storage of high protein buffalo milk ice cream. *LWT-Food Science and Technology* **150**: 111903.
DOI: https://doi.org/10.1016/j.lwt.2021.111903

Sarkar U., Gupta A.K., Sarkar V., Mohanty T.K., Raina V.S. & Prasad S. (2006). Factors affecting test day milk yield and milk composition in dairy animals. *Journal of Dairying, Foods and Home Sciences* **25**(2): 129–132.

Sert D., Mercan E. & Kılınç M. (2021). Development of buffalo milk ice-cream by high pressure-homogenisation of mix: Physicochemical, textural and microstructural characterisation. *LWT-Food Science and Technology* **150**: 112013.
DOI: https://doi.org/10.1016/j.lwt.2021.112013

Shah S.K., Schermerhorn E.C., Cady R.A. & McDowell R.E. (1983). Factors affecting milk fat percent of Nili-Ravi buffaloes in Pakistan. *Journal of Dairy Science* **66**(3): 573–577.
DOI: https://doi.org/10.3168/jds.S0022-0302(83)81827-X

Siddiky M. & Faruque M. (2018). Buffaloes for dairying in South Asia: Potential, challenges and way forward. *SAARC Journal of Agriculture* **15**(2): 227–239.

DOI: https://doi.org/10.3329/sja.v15i2.35167

Silva T.M.S., Piazentin A.C.M., Mendonça C.M.N., Converti A., Bogsan C.S.B., Mora D. & de Souza Oliveira R.P. (2020). Buffalo milk increases viability and resistance of probiotic bacteria in dairy beverages under in vitro simulated gastrointestinal conditions. *Journal of Dairy Science* **103**(9): 7890–7897.
DOI: https://doi.org/10.3168/jds.2019-18078

Sodhi S.S., Mehra M.L., Jain A.K. & Trehan P.K. (2008). Effect of non-genetic factors on the composition of milk of Murrah buffaloes. *Indian Veterinary Journal* **85**(09): 950–952.

Sun Q., Lv J.P., Liu L., Zhang S.W., Liang X. & Lu J. (2014). Comparison of milk samples collected from some buffalo breeds and crossbreeds in China. *Dairy Science and Technology* **94**(4): 387–395.
DOI: https://doi.org/10.1007/s13594-013-0159-9

Sundaram M. & Harharan G. (2013). Preliminary study on evaluation of effect of lactation number on milk yield and milk composition in Murrah (*Bubalus bubalis*) Buffaloes. *Research Journal of Animal, Veterinary and Fishery Sciences* **1**(7): 21–23.

Tekelioglu O., Cimen M. & Bayril T. (2010). The milk biochemical parameters having economic importance in machine milked cows. *Journal of Animal and Veterinary Advances* **9**(3): 519–521.
DOI: https://doi.org/10.3923/javaa.2010.519.521

Waldron D.S., Hoffmann W., Buchheim W., McMahon D.J., Goff H.D., Crowley S.V., Moloney C., O'regan J., Giuffrida F., Torres I.C. & Siong P. (2020). Role of milk fat in dairy products. In: *Advanced Dairy Chemistry,* (volume 2): *Lipids* (eds. P.L.H. McSweeney, P.F. Fox & J.A. O'Mahony), pp. 245–305. Aspen Publishers, New York, USA.
DOI: https://doi.org/10.1007/978-3-030-48686-0_9

Yadav S.P., Sikka P., Kumar D., Sarkar S., Pandey A.K., Yadav P.S. & Sethi R.K. (2013). Variation in milk constituents during different parity and seasons in Murrah buffaloes. *Indian Journal of Animal Sciences* **83**(07): 747–751.

Yapa D., Rasika D.M.D., Weerathilake W.A.D.V., Siriwardhana J. & Priyashantha H. (2023). Effects of fermenting with *Lacticaseibacillus rhamnosus* GG on quality attributes and storage stability of buffalo milk yogurt incorporated with bael (*Aegle marmelos*) fruit pulp. *NSF Journal* **31**: 102–109.
DOI: https://doi.org/10.1016/j.nfs.2023.04.002

Zhou L., Tang Q., Wasim I.M., Xia Z., Huang F., Li L., Liang M., Lin B., Qin G. & Zou C. (2018). A comparison of milk protein, fat, lactose, total solids and amino acid profiles of three different buffalo breeds in Guangxi. China. *Italian Journal of Animal Science* **17**(04): 873–878.
DOI: https://doi.org/10.1080/1828051X.2018.1443288

RESEARCH ARTICLE

Graph Theory

All Ramsey critical graphs for large cycles vs a complete graph of order six

CJ Jayawardene[1*] WCW Navaratna[2] and JN Senadheera[2]

[1] *Department of Mathematics, Faculty of Sciences, University of Colombo, Colombo, Sri Lanka.*

[2] *Department of Mathematics, Faculty of Natural Sciences, The Open University of Sri Lanka, Nawala, Sri Lanka.*

Abstract : A new area of graph theory emerged in the last few decades is the calculation of star critical Ramsey numbers related to different classes of graphs. Formally, we will say that $K_n \rightarrow (G, H)$ if given any coloring of K_n there is a copy of G in the first color, red, or a copy of H in the second color, blue. The Ramsey number $r(G, H)$ is defined as the smallest positive integer n such that $K_n \rightarrow (G, H)$. A closely related concept of Ramsey number is the star-critical Ramsey number $r_*(G, H)$ defined as the largest value of k such that $K_{r(G,H)-1} \sqcup K_{1,k} \rightarrow (G, H)$. A two-coloring of $K_{r(G,H)-1}$ such that $K_{r(G,H)-1} \nrightarrow (G, H)$ is called a Ramsey critical coloring. A Ramsey critical $r(G, H)$ graph is a graph induced by the first color of a Ramsey critical coloring. Lower bounds for star critical Ramsey numbers are usually found with the aid on Ramsey critical graphs. The particular problem we handle in this paper, on star critical Ramsey numbers, is based on a conjecture posed in 1973 by Bondy and Erdos relating to Ramsey numbers for large cycles versus complete graphs. Based on certain lemmas we present with proof, furthermore we show that there exist exactly sixty eight non-isomorphic Ramsey critical $r(C_n, K_6)$ graphs, when $n \geq 15$.

Keywords: Graph theory, Ramsey theory, Ramsey critical graphs.

INTRODUCTION

For any two graphs G and H, the Ramsey number $r(G,H)$ is the smallest positive integer n such that $K_n \rightarrow (G, H)$. The Ramsey number $r(G,H)$ is defined as the smallest positive integer n such that $K_n \rightarrow (G, H)$. Many interesting variations of the basic problem of finding classical Ramsey numbers have emerged. One such variation is the calculation of the number of Ramsey critical (G,H) graphs, where (G,H) represents any pair of graphs (Yali *et al*., 2015; Hook & Isaak, 2011; Budden, 2023). Another variation is the size Ramsey Numbers (Zhang *et al*., 2023). In this paper, we show that there are exactly 68 Ramsey critical $r(C_n, K_6)$ graphs, for all n exceeding fourteen.

Notations

Given a graph G and a vertex $v \in V(G)$, we define the neighbourhood of v in G, $\Gamma(v)$, as the set of vertices adjacent to v in G. The degree of a vertex v, $d(v)$, is defined as the cardinality of $\Gamma(v)$, i.e. $d(v) = |\Gamma(v)|$. The minimum degree of a graph $G(V,E)$ denoted by $\delta(G)$ is defined as $\min\{d(v)|v \in V\}$. Given a graph G, we say $I \subseteq V(G)$ is an *independent set*, if no pair of vertices of I is adjacent to each other in G. Equivalently, I forms a clique in G^c. Given a graph $G = (V, E)$ we define the independence number, $\alpha(G)$, as the size of the largest independent set. Thus, $\alpha(G) = \max\{|I| : I$ is an independent set of $G\}$. In the special case of $H = K_m$, alternatively r(G,Km) can be viewed as the smallest positive integer n such that every graph complete of order n either contains G as a subgraph or else satisfies $\alpha(G) \geq m$. For two disjoint subgraphs H and K of G, we denote the set of edges between H and K by $E(H, K)$.

* Corresponding author (c_jayawardene@yahoo.com; https://orcid.org/0000-0002-6951-6983)

LEMMAS USED TO GENERATE ALL RAMSEY CRITICAL (C_n, K_6) **GRAPHS FOR** $n \geq 15$

In an attempt to prove Bondy and Erdös conjecture $r(C_n, K_m) = (n-1)(m-1)+1$, for all $(n, m) \neq (3, 3)$ satisfying $n \geq m \geq 3$ under certain restrictions, Schiermeyer has proved that $r(C_n, K_6) = 5(n-1)+1$, for $n \geq 6$ (Bondy *et al.*, 1973; Schiermeyer 2003; Radziszowsk 2021). Characterizing all Ramsey critical (C_n, K_6) graphs boils down to finding all (red/blue) colorings of $K_{r(C_n, K_6)-1}$ such that there is no red C_n or a blue K_6. This is achieved by finding all C_n-free graphs G on $K_{r(C_n, K_6)-1}$ vertices such that $\alpha(G) < 6$. We first prove that any C_n-free graph (where $n \geq 15$) of order $5(n-1)$ with $\alpha(G) \leq 5$ contains a $5K_{n-1}$. To prove this, we use seven lemmas of which the first three are already proven results. For ease of reference, we reiterate Lemma 1 (Jayawardena & Samarasekara, 2017), Lemma 2 (Jayawardena, 2019) and Lemma 3 (Bollobás *et al.*, 2013).

Lemma 1: A C_n- free graph G of order N with independent number less than or equal to m has minimal degree greater than or equal to $N - r(C_n, K_m)$.

Lemma 2: A C_n- free graph (where $n \geq 7$) of order $4(n-1)$ with no independent set of 5 vertices contains a $4K_{n-1}$.

Lemma 3: Suppose G contains the cycle $(u_1, u_2, ..., u_{n-1}, u_1)$ of length $n-1$ but no cycle of length n. Let $Y = V(G) \setminus \{u_1, u_2, ..., u_{n-1}\}$. Then,

(a) No vertex $x \in Y$ is adjacent to two consecutive vertices on the cycle.

(b) If $x \in Y$ is adjacent to u_i and u_j then $u_{i+1}u_{j+1} \notin E(G)$.

(c) If $x \in Y$ is adjacent to u_i and u_j then no vertex $x' \in Y$ is adjacent to both u_{i+1} and u_{j+2}.

(d) Suppose $\alpha(G) = m-1$ where $m \leq \frac{n+2}{2}$ and $\{x_1, x_2, ..., x_{m-1}\} \subseteq Y$ is an $(m-1)$-element independent set. Then, no member of this set is adjacent to $m-2$ or more vertices on the cycle (We have taken the liberty of making a slight correction to the inequality $m \leq \frac{n+2}{2}$ of the original [1], Lemma 5(d)).

The next lemma plays a pivotal role in proving the main results of this paper.

Lemma 4: A C_n -free graph (where $n \geq 15$) of order $5(n-1)$ with no independent set of 6 vertices contains a $5K_{n-1}$.
Proof. We shall assume that in each of the three cases $n = 15$, $n = 16$ and $n \geq 17$ we consider, G as a graph on $5(n-1)$ vertices satisfying $C_n \nsubseteq G$ and $\alpha(G) \leq 5$. Since $r(C_{n-1}, K_6) = 5n-9 \leq 5(n-1)$ (see [1, 6]), there exists a cycle $C = (u_1, u_2, ..., u_{n-1}, u_1)$ of length $n-1$ in G. In consistent with the notation of [1], define H as the induced subgraph of G not containing the vertices of the cycle C. Then, $|V(C)| = n-1$ and $|V(H)| = 4(n-1)$.

Suppose there exists an independent set $Y = \{y_1, y_2, y_3, y_4, y_5\}$ of size 5 in H, so that $\alpha(G) = 5$. From Lemma 3 (as $5 \leq \frac{n+2}{2}$), it follows that no vertex of Y is adjacent to four or more vertices of the C_{n-1}. Thus, $|E(Y, V(C))| \leq 15$. For ease of reference, we define such a graph structure as a **Standard Configuration** (n).

Case 1: $n \geq 17$

Now, $|E(Y, V(C))| \leq 15 < n - 1$. Thus, there exists a vertex $x \in V(C)$ adjacent to no vertex of Y. This gives, an independent set $Y \cup \{x\}$ of size 6, a contradiction.

Case 2: $n = 16$

In this case as $n - 1 = 15$, in order to avoid an independent set of size 6, each vertex of $V(C)$ must be adjacent to at least one vertex of Y. Thus, we get that for each $1 \leq i \leq 5$, $|\Gamma(y_i) \cap V(C)| = 3$ and for each $1 \leq j < j' \leq 5$, $\Gamma(y_j) \cap \Gamma(y_{j'}) \cap V(C) = \phi$.

By Lemma 1, as $\delta(G) \geq 14$, $|\Gamma(y_i) \cap V(H \setminus Y)| \geq 11$ for $i = 1, 2$. Since $r(P_3, K_6) = 11$ and $\alpha(G) < 6$, each of $G[\Gamma(y_1) \cap V(H \setminus Y)]$ and $G[\Gamma(y_2) \cap V(H \setminus Y)]$ contains a copy of P_3. Thus, $P_3 \subseteq \Gamma(y_1) \cap V(H \setminus Y)$, where the P_3 is induced by $\{x, y, z\}$ such that $(x, y), (y, z) \in E(G)$ and $P_3 \subseteq \Gamma(y_2) \cap V(H \setminus Y)$, where this P_3 is induced by $\{p, q, r\}$ such that $(p, q), (q, r) \in E(G)$.

Suppose that x is not adjacent to any vertex of $\{y_2, y_3, y_4, y_5\}$ and p is not adjacent to any vertex of $\{y_1, y_3, y_4, y_5\}$. Re-order the vertices of the cycle such that $y_1 \in Y$ is adjacent to u_1. In this ordering, let y_1 be also adjacent to u_t where $2 \leq t \leq 15$.
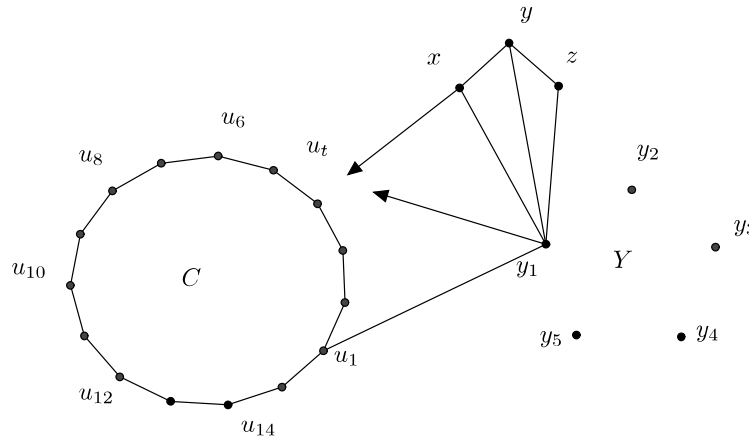
Figure 1: Configuration for $n = 16$

By Lemma 3(a), $t \neq 2$. In order to avoid an independent set of size 6, induced by $\{x, u_t, y_2, y_3, y_4, y_5\}$, we get that $(x, u_t) \in E(G)$. However, $t \neq 3$, in order to avoid a C_{16} comprising $(u_1, y_1, x, u_3, ..., u_{15}, u_1)$. Also, $t \neq 4$ in order to avoid a C_{16} comprising $(u_1, y_1, y, x, u_4, ..., u_{15}, u_1)$ and $t \neq 5$ in order to avoid a C_{16} comprising $(u_1, y_1, z, y, x, u_5, ..., u_{15}, u_1)$.

Thus, any pair of vertices adjacent to y_1 in C cannot be separated by a path of length 1, 2, 3 or 4 along C. Thus, $\Gamma(y_1) \cap C = \{u_1, u_6, u_{11}\}$. In this scenario, we use the prerogative that $(y_2, u_2) \in E(G)$. Then, by the previous argument $\Gamma(y_2) \cap C = \{u_2, u_7, u_{12}\}$. But by Lemma 3(b), $(u_2, u_7) \notin E(G)$. Henceforth, we will get that $\{u_2, u_7, y_1, y_3, y_4, y_6\}$ is an independent set of size 6, a contradiction.

This implies that there is a vertex of $X = \{x, y, z\}$ adjacent to some vertex of $\{y_2, y_3, y_4, y_5\}$ or there is a vertex of $\{p, q, r\}$ adjacent to some vertex of $\{y_1, y_3, y_4, y_5\}$. Therefore, without loss of generality, we may assume that y_1 is adjacent to $X = \{x, y, z\} \subseteq V(H \setminus Y)$ and y_1 is adjacent to $X' = \{x', y', z'\} \subseteq V(C)$ where X' induces a P_3 and y_2 is adjacent to x. Next since C has 15 points without loss of generality, $\{x', y', z'\} = \{u_1, u_3, u_5\}$ or $\{x', y', z'\} = \{u_1, u_3, u_6\}$ or $\{x', y', z'\} = \{u_1, u_3, u_7\}$ or $\{x', y', z'\} = \{u_1, u_3, u_8\}$ or $\{x', y', z'\} = \{u_1, u_3, u_9\}$ or $\{x', y', z'\} = \{u_1, u_4, u_7\}$ or

$\{x', y', z'\} = \{u_1, u_4, u_8\}$ or $\{x', y', z'\} = \{u_1, u_4, u_9\}$ or $\{x', y', z'\} = \{u_1, u_4, u_{10}\}$ or $\{x', y', z'\} = \{u_1, u_5, u_9\}$ or $\{x', y', z'\} = \{u_1, u_5, u_{10}\}$ or $\{x', y', z'\} = \{u_1, u_6, u_{11}\}$. Moreover, as y_1 and y_2 are connected by paths of lengths 2, 3 and 4 in H, no pair of vertices selected from $\Gamma(y_1) \cap V(C)$ and $\Gamma(y_2) \cap V(C)$ can be separated by a path of length 3, 4 or 5 along the cycle C. Using this we argue that when $\{x', y', z'\} = \{u_1, u_3, u_5\}$, $\Gamma(y_2) \cap V(C) = \phi$, when $\{x', y', z'\} = \{u_1, u_3, u_6\}$, $\Gamma(y_2) \cap V(C) = \phi$ and when $\{x', y', z'\} = \{u_1, u_3, u_7\}$, $\Gamma(y_2) \cap V(C) \subseteq \{u_9\}$. Similarly, when $\{x', y', z'\} = \{u_1, u_3, u_8\}$, $\Gamma(y_2) \cap V(C) \subseteq \{u_2, u_9, u_{10}\}$, when $\{x', y', z'\} = \{u_1, u_3, u_9\}$, $\Gamma(y_2) \cap V(C) \subseteq \{u_2, u_{10}\}$, when $\{x', y', z'\} = \{u_1, u_4, u_7\}$, $\Gamma(y_2) \cap V(C) = \phi$ and when $\{x', y', z'\} = \{u_1, u_4, u_8\}$, $\Gamma(y_2) \cap V(C) \subseteq \{u_2, u_{10}\}$. When $\{x', y', z'\} = \{u_1, u_4, u_9\}$, $\Gamma(y_2) \cap V(C) \subseteq \{u_2, u_3, u_{10}\}$, when $\{x', y', z'\} = \{u_1, u_4, u_{10}\}$, $\Gamma(y_2) \cap V(C) \subseteq \{u_2, u_3\}$, when $\{x', y', z'\} = \{u_1, u_5, u_9\}$, $\Gamma(y_2) \cap V(C) \subseteq \{u_3, u_7\}$ and when $\{x', y', z'\} = \{u_1, u_5, u_{10}\}$, $\Gamma(y_2) \cap V(C) \subseteq \{u_3\}$ and when $\{x', y', z'\} = \{u_1, u_6, u_{11}\}$, $\Gamma(y_2) \cap V(C) = \phi$. Since none of these give a viable configuration, we get a contradiction.

Case 3: $n = 15$

To deal with the case $n = 15$, we first prove three Lemmas. Lemma 5, deals with the possible scenarios generated by the Standard Configuration (15). Lemmas 6 and 7 deal with showing that none of the scenarios generated by Lemma 5 give viable configurations.
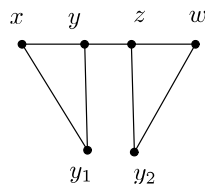
Lemma 5: In the Standard Configuration ($n = 15$), one of the following three scenarios (a), (b) and (c) will occur:

(a) $y_1 \in Y$ is a vertex of the subgraph K_4 (see Figure 2(a)) in H.

(b) $y_1, y_2 \in Y$ are vertices of the subgraph K (see Figure 2(b1)) or subgraph K' (see Figure 2(b2)) in H.

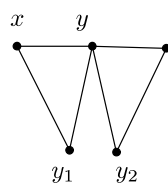**(c)** $y_1, y_2 \in Y$ are vertices of the subgraph L (see Figure 2(c)) in H.
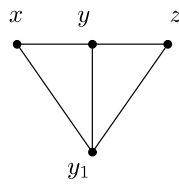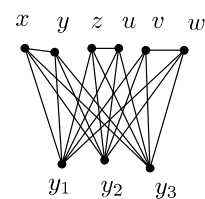

$(a) : graph \quad K_4 \qquad (b1) : graph \quad K \qquad (b2) : graph \quad K' \qquad (c) : graph \quad L$

scenario (a)   scenario (b)   scenario (c)

Figure 2: (a),(b1), (b2) and (c). The first three scenarios, scenario (a), scenario (b) and scenario (c) of the Standard Configuration ($n = 15$).

Proof. As in the case of $n = 16$, we get that without loss of generality $1 \leq i \leq 4$, $|\Gamma(y_i) \cap V(C)| = 3$ and for each $1 \leq j < j' \leq 4$, $\Gamma(y_j) \cap \Gamma(y_{j'}) \cap V(C) = \phi$. Also $|\Gamma(y_5) \cap V(C)| \in \{2, 3\}$. In particular, if $|\Gamma(y_5) \cap V(C)| = 2$ then, for each $1 \leq j < j' \leq 5$, $\Gamma(y_j) \cap \Gamma(y_{j'}) \cap V(C) = \phi$ and if $|\Gamma(y_5) \cap V(C)| = 3$

then, for each $1 \le j < j' \le 5$,

$$|\Gamma(y_j) \cap \Gamma(y_{j'}) \cap V(C)| = \begin{cases} 1 & \text{if } j, j' \in \{4, 5\}, \\ \\ 0 & \text{otherwise.} \end{cases}$$

By Lemma 1, as $\delta(G) \ge 13$, we get that $|\Gamma(y_i) \cap V(H \setminus Y)| \ge 10$. Suppose that there is some $1 \le i \le 3$ (say $i = 1$) such that $|\Gamma(y_i) \cap V(H \setminus Y)| \ge 11$. Then as $r(P_3, K_6) = 11$ we get scenario (a). Next, assume that for all $1 \le i \le 3$, $|\Gamma(y_i) \cap V(H \setminus Y)| = 10$. By the classification of the Ramsey critical (P_3, K_6) graphs, we get that for all $1 \le i \le 3$, $G[\Gamma(y_i) \cap V(H \setminus Y)] \supseteq 5K_2$.

This gives two possibilities. The first possibility is $\left| \cup_{i=1}^{3} \Gamma(y_i) \cap V(H \setminus Y) \right| = 10$. In this case, as for all $1 \le i \le 3$, $5K_2 \subseteq G[\Gamma(y_i) \cap V(H \setminus Y)]$ we get scenario (c). The second possibility if $\left| \cup_{i=1}^{3} \Gamma(y_i) \cup V(H \setminus Y) \right| \ge 11$.
Without loss of generality, we may assume that $|\Gamma(y_1) \cup \Gamma(y_2) \cup V(H \setminus Y)| \ge 11$. Let x_{11} be any vertex of $\Gamma(y_2) \cap (\Gamma(y_1))^c \cup V(H \setminus Y)$. Since $r(P_3, K_6) = 11$, we get that $G[\Gamma(y_1) \cup V(H \setminus Y) \cup \{x_{11}\}]$ contains a subgraph P isomorphic to a P_3. If P is contained in $G[\Gamma(y_1) \cup V(H \setminus Y)]$ we get scenario (a). Otherwise, $x_{11} \in P$. However, x_{11} is an element of $5K_2 \subseteq G[\Gamma(y_2) \cap V(H \setminus Y)]$ and therefore, x_{11} is adjacent to some other vertex say w in $G[\Gamma(y_2) \cap V(H \setminus Y)]$. Depending on whether or not w belongs to $V(P)$, we get scenarios $(b2)$ or $(b1)$ respectively. Hence the Lemma. $\blacksquare$

Lemma 6: In the Standard Configuration ($n = 15$), $y_1 \in Y$ can not be a vertex of a K_4 in H (see Figure 2(a)).
Proof. As indicated in Figure 2(a), let $x \in \Gamma(y_1) \cap V(H \setminus Y)$. Then we get two possibilities depending on whether or not x is adjacent to a vertex of $\{y_2, y_3, y_4, y_5\}$. In the first possibility, x is adjacent to some vertex of Y (say y_2). Then, as y_1 and y_2 are connected by paths of lengths 2, 3 and 4 in H, no pair of vertices selected from of $\Gamma(y_1) \cap V(C)$ and $\Gamma(y_2) \cap V(C)$ can be separated by a path of length 3, 4 or 5 along the cycle C. However, as argued in $n = 16$, we get that $|\Gamma(y_2) \cap V(C)| \le 2$ and $|\Gamma(y_2) \cap V(C)| = 2$ only when $\{x', y', z'\} = \{u_1, u_3, u_8\}$, $\{x', y', z'\} = \{u_1, u_4, u_9\}$ or $\{x', y', z'\} = \{u_1, u_5, u_9\}$. This gives a contradiction. In the second possibility, re order the vertices of the cycle such that $y_1 \in X$ is adjacent to u_1. In this ordering, suppose further that y_1 is also adjacent to u_t where $2 \le t \le 14$. By the argument used in $n = 16$, we get that any pair of vertices adjacent to y_1 in C cannot be separated by a path of length 1, 2, 3 or 4 along C. However, this again leads to a contradiction. $\blacksquare$

Lemma 7: In the Standard Configuration ($n = 15$), the vertices $y_1, y_2 \in Y$ can not be vertices of the subgraph K, K' or L in H (see Figure 2).
Proof. In the case $y_1, y_2 \in K$, since y_1 and y_2 are connected by paths of lengths 3, 4 and 5 in H, no pair of vertices selected from of $\Gamma(y_1) \cap V(C)$ and $\Gamma(y_2) \cap V(C)$ can be separated by a path of length 4, 5 or 6 along the cycle C. The cardinality of the possible vertex sets of $\Gamma(y_i) \cap V(C)$ ($i = 1, 2$), subject to this condition, are presented in Table 1.

Because $\Gamma(y_2) \cap V(C) = 3$, we are only left to deal with the last two possibilities of Table 1 for $\Gamma(y_2) \cap V(C)$. In both possibilities, $\Gamma(y_1) \cap V(C)$ will induce a C_3 by Lemma 3(b). In the first possibility, $\Gamma(y_1) \cap V(C) = \{u_1, u_5, u_9\}$ gives rise to the 15-cycle given by $(u_1, u_5, ..., u_8, y_2, y, x, y_1, u_9, ..., u_{14}, u_1)$, a contradiction. In the second possibility, $\Gamma(y_1) \cap V(C) = \{u_1, u_5, u_{10}\}$ gives rise to the 15-cycle given by $(u_1, u_5, u_6, ..., u_8, y_2, z, y, x, y_1, u_{10}, ..., u_{14}, u_1)$, a contradiction.
In the case $y_1, y_2 \in K'$, since y_1 and y_2 are connected by paths of lengths 2, 3 or 4 in H, no pair of vertices selected from of $\Gamma(y_1) \cap V(C)$ and $\Gamma(y_2) \cap V(C)$ can be separated by a path of length 3, 4 or 5 along the cycle C.

Table 1: Cardinality of $\Gamma(y_2) \cap V(C)$: Graph K.

$\{x', y', z'\}$ equals	$\Gamma(y_2) \cap V(C)$ is contained in	Cardinality of $\Gamma(y_2) \cap V(C)$
$\{u_1, u_3, u_5\}$	$\{u_2, u_4\}$	$|\Gamma(y_2) \cap V(C)| \leq 2$
$\{u_1, u_3, u_6\}$	$\{u_4\}$	$|\Gamma(y_2) \cap V(C)| \leq 1$
$\{u_1, u_3, u_7\}$	$\{u_4, u_{14}\}$	$|\Gamma(y_2) \cap V(C)| \leq 2$
$\{u_1, u_3, u_8\}$	ϕ	$|\Gamma(y_2) \cap V(C)| = 0$
$\{u_1, u_3, u_9\}$	$\{u_2\}$	$|\Gamma(y_2) \cap V(C)| \leq 1$
$\{u_1, u_4, u_7\}$	ϕ	$|\Gamma(y_2) \cap V(C)| = 0$
$\{u_1, u_4, u_8\}$	ϕ	$|\Gamma(y_2) \cap V(C)| = 0$
$\{u_1, u_4, u_9\}$	$\{u_2\}$	$|\Gamma(y_2) \cap V(C)| \leq 1$
$\{u_1, u_5, u_9\}$	$\{u_2, u_8, u_{12}\}$	$|\Gamma(y_2) \cap V(C)| \leq 3$
$\{u_1, u_5, u_{10}\}$	$\{u_3, u_8, u_{12}\}$	$|\Gamma(y_2) \cap V(C)| \leq 3$

The cardinality of the possible vertex set of $\Gamma(y_1) \cap V(C)$ is presented in Table 2 and each of these leads to a contradiction as $\Gamma(y_2) \cap V(C) < 3$.

In the case $y_1, y_2 \in K'$, since y_1 and y_2 are connected by paths of lengths 2, 3, 4 and 5 in H, no pair of vertices selected from of $\Gamma(y_1) \cap V(C)$ and $\Gamma(y_2) \cap V(C)$ can be separated by paths of length 3, 4, 5 or 6 along the cycle C. Thus, Table 2 will give us the required contradiction.

Table 2: Cardinality of $\Gamma(y_2) \cap V(C)$: Graph K'

$\{x', y', z'\}$ equals	$\Gamma(y_2) \cap V(C)$ is contained in	Cardinality of $\Gamma(y_2) \cap V(C)$
$\{u_1, u_3, u_5\}$	ϕ	$|\Gamma(y_2) \cap V(C)| = 0$
$\{u_1, u_3, u_6\}$	ϕ	$|\Gamma(y_2) \cap V(C)| = 0$
$\{u_1, u_3, u_7\}$	$\{u_9\}$	$|\Gamma(y_2) \cap V(C)| \leq 1$
$\{u_1, u_3, u_8\}$	$\{u_2, u_9\}$	$|\Gamma(y_2) \cap V(C)| \leq 2$
$\{u_1, u_3, u_9\}$	$\{u_2\}$	$|\Gamma(y_2) \cap V(C)| \leq 1$
$\{u_1, u_4, u_7\}$	ϕ	$|\Gamma(y_2) \cap V(C)| = 0$
$\{u_1, u_4, u_8\}$	$\{u_2\}$	$|\Gamma(y_2) \cap V(C)| \leq 1$
$\{u_1, u_4, u_9\}$	$\{u_2, u_3\}$	$|\Gamma(y_2) \cap V(C)| \leq 2$
$\{u_1, u_5, u_9\}$	$\{u_3, u_7\}$	$|\Gamma(y_2) \cap V(C)| \leq 2$
$\{u_1, u_5, u_{10}\}$	$\{u_3\}$	$|\Gamma(y_2) \cap V(C)| \leq 1$

Similarly, in the case $y_1, y_2 \in L$, since y_1 and y_2 are connected by paths of lengths 2, 3, 4, 5 and 6 in H, no pair of vertices selected from of $\Gamma(y_1) \cap V(C)$ and $\Gamma(y_2) \cap V(C)$ can be separated by paths of length 3, 4, 5, 6 or 7 along the cycle C. As before, for all possibilities $|\Gamma(y_2) \cap V(C)| < 3$, a contradiction. Thus, lemmas 5, 6 and 7 imply that H cannot have an independent set of size 5. Having proved that H cannot have an independent set of size 5 in all three cases $n = 15, 16$ and 17, we next continue with the main proof. Since, H satisfies all conditions of Lemma 2, H contains a $4K_{n-1}$.

Next we show that $V(C_{n-1})$ induced a K_{n-1}. Suppose that there exists two vertices of $V(C)$, say v and w, such that $(v, w) \notin E(G)$. In order to avoid a C_n both v and w will have to be adjacent

to at most one vertex of each of the four copies of K_{n-1} in H. Moreover, any vertex of any copy of K_{n-1} in H will have to be adjacent to at most one vertex of another copy of a K_{n-1} in H. Thus, each copy of a K_{n-1} will have at most 5 vertices adjacent to some vertex outside that of K_{n-1}, in $V(H) \cup \{v, w\}$. Since $(n-1) - 5 \geq 1$, we can select x_1 in the first K_{n-1}, x_2 in the second K_{n-1}, x_3 in the third K_{n-1} and x_4 in the fourth K_{n-1} such that $\{x_1, x_2, x_3, x_4\}$ is an independent set of size four and no vertex of $\{x_1, x_2, x_3, x_4\}$ is adjacent to any vertex of $\{v, w\}$. Hence $\{x_1, x_2, x_3, x_4, v, w\}$ is an independent set of size 6, a contradiction. Therefore, we get that any two pair of vertices of $V(C)$ are connected by an edge. Hence, $G[V(C_{n-1})] = K_{n-1}$ as required. This K_{n-1} along with the $4K_{n-1}$ contained in H gives the required result.

ALL RAMSEY (C_n, K_6) CRITICAL GRAPHS FOR $n \geq 15$

We have already observed that any Ramsey (C_n, K_6) critical graph will consist of a red graph containing $5K_{n-1}$, with respect to the red/blue coloring. Let $\{V_i : i \in \{1, 2, ..., 5\}\}$ be the vertex set of the five K_{n-1} graphs. We notice that there are two types of Ramsey (C_n, K_6) critical graphs. The first type (Type1) of Ramsey (C_n, K_6) critical graphs will satisfy the condition that at most one vertex of each V_i is adjacent to any other vertex in V_i^c. The second type (Type2) of Ramsey (C_n, K_6) critical graphs will satisfy the condition that there exists a V_k for some $1 \leq k \leq 5$ such that at least two vertices of V_k have neighbors in V_k^c. Moreover, it is worth noting that a Type1 critical graph is completely determined by the structure of the external edges between V_i's and not by the $\binom{n}{2}$ edges inside each of the five V_i's. This fact is taken into consideration when representing the Ramsey (C_n, K_6) critical graphs.

Each subgraph of K_5 generates a unique Ramsey (C_n, K_6) critical graph of Type1. Thus, as illustrated in the following figure, there are 34 critical graphs (R_i, $1 \leq i \leq 34$) of Type1 generated by the 34 subgraphs of K_5.
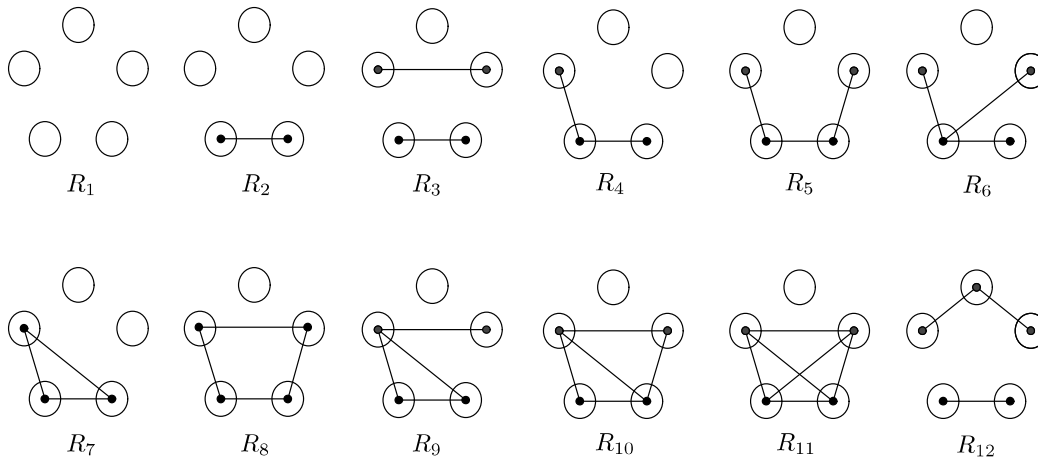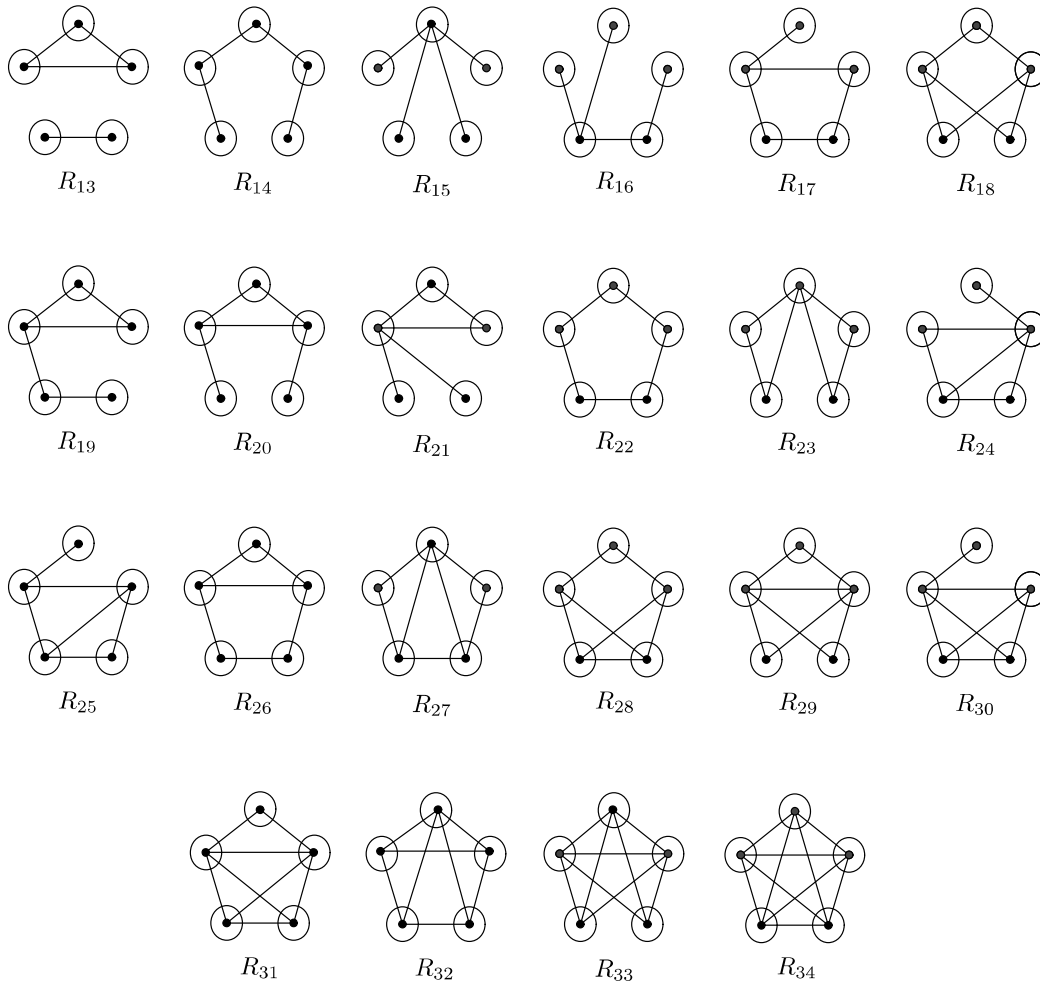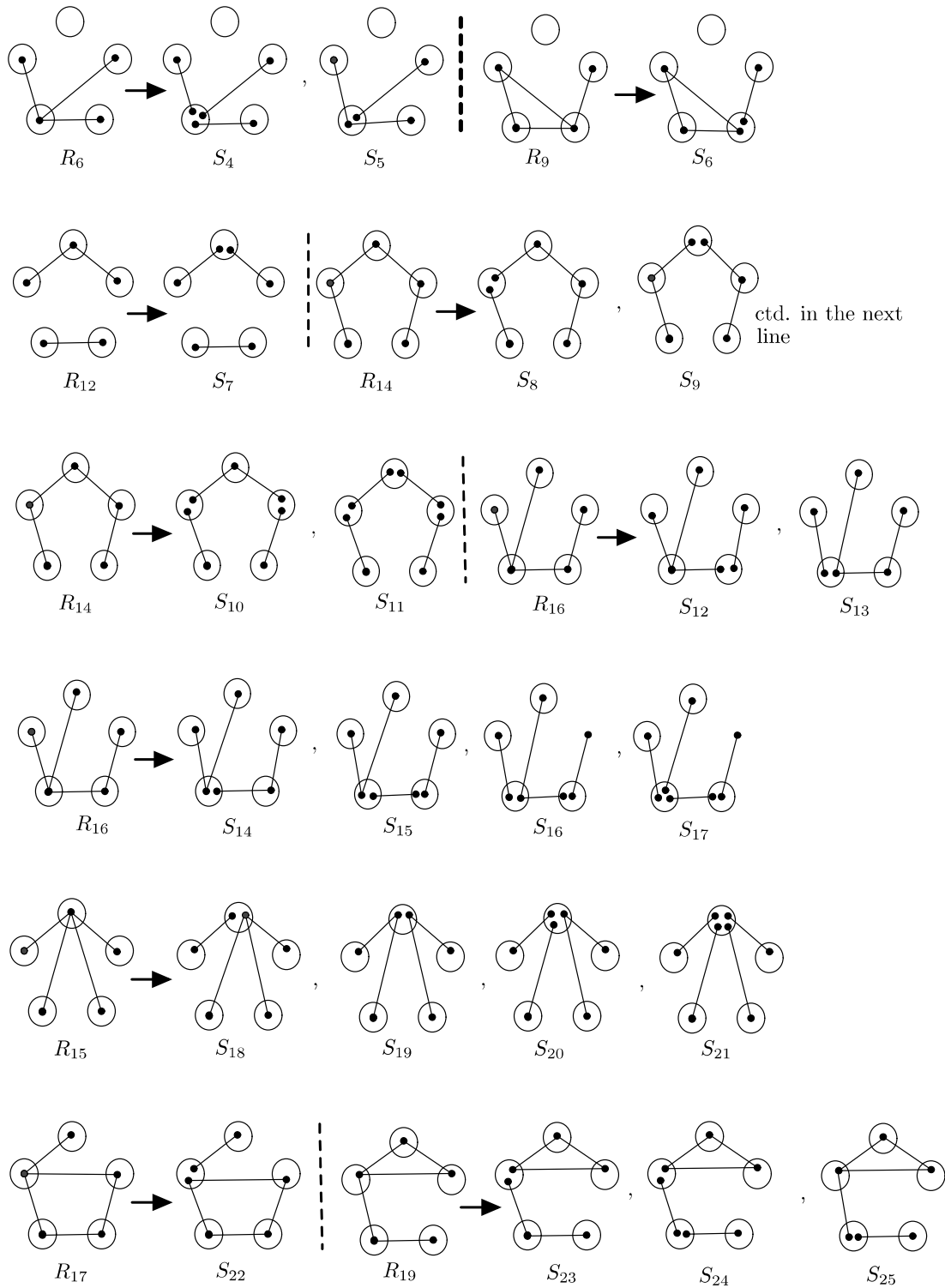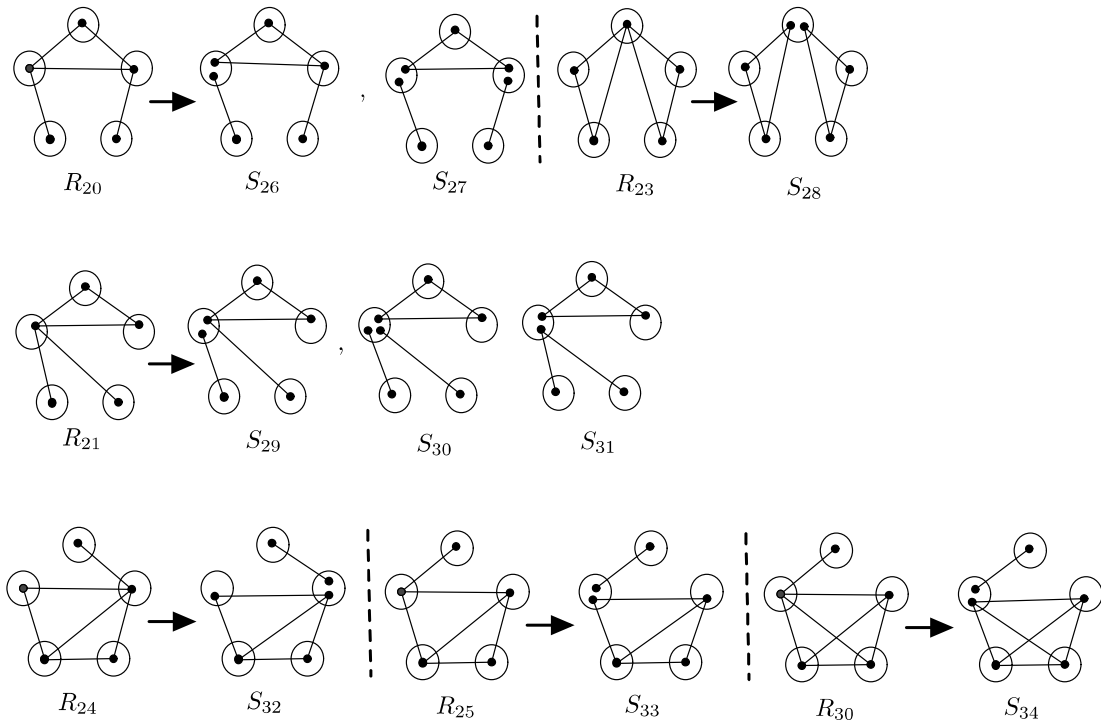
Figure 3: Ramsey (C_n, K_6) critical graphs of Type1, R_i $(1 \leq i \leq 34)$

First note that each and every Type2 critical graph is obtained by an appropriate vertex splitting of some Type1 critical graph. As illustrated in the Figure 4, there are exactly 34 Type2 critical graphs (labeled S_i where $1 \leq i \leq 34$) generated by 18 critical graphs of Type1, since exactly sixteen Type1 critical graphs do not generate Type2 critical graphs.

ctd. in the next line
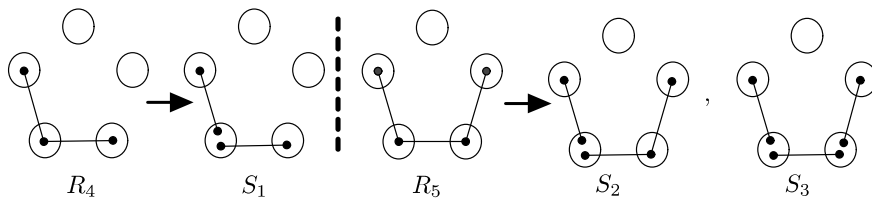
Figure 4: Ramsey (C_n, K_6) critical graphs of Type2 (S_i, $1 \leq i \leq 34$)

Henceforth, we conclude that there are exactly 68 Ramsey (C_n, K_6) critical graphs out of which 34 are categorized as Type1 critical graphs (labeled R_i, $1 \leq i \leq 34$) and the balance 34 are categorized as Type2 critical graphs (labeled S_i, $1 \leq i \leq 34$).

CONCLUSION

From this paper we see that the number of Ramsey Critical graphs of (C_n, K_6) is 68 and that the number of Ramsey Critical graphs of (C_n, K_m) is growing steadily in number as m increases up to 6. Moreover, this paper investigates the relationship between Ramsey Critial graphs of $m = 5$ and $m = 6$. This technique should play a pivotal role in finding all Ramsey Critical graphs for (C_n, K_m)f or $m \geq 7$.

REFERENCES

Bollobas B., Jayawardene C.J., Sheng Y.J., Ru H.Y., Rousseau C.C. & Min Z.K. (2000). On a conjecture involving cycle-complete graph Ramsey numbers. *The Australasian Journal of Combinatorics* **22**: 63–71.

Budden M.R. (2023). *Star-Critical Ramsey Numbers for Graphs*. Springer International Publishing, Cham, Switzerland.

Hook J. & Isaak G. (2011). Star-critical Ramsey numbers. *Discrete Applied Mathematics* **159**:328–334.
DOI: https://doi.org/10.1016/j.dam.2010.11.007

Jayawardene C.J. (2019). The Star-critical Ramsey Number for any Cycle vs. a K5.arXiv:1901.04802.

Jayawardene C.J. & Samarasekara B.L. (2017). The size multipartite Ramsey numbers for stars vs cycles. *Journal of the national Science Foundation of Sri Lanka* **45**(1): 67–72.
DOI: http://dx.doi.org/10.4038/jnsfsr.v45i1.8039

Radziszowski S.P. (2021). Small Ramsey numbers. *Electronic Journal of Combinatorics* **16**:DS1.
DOI: https://doi.org/10.37236/21

Schiermeyer I. (2003). All cycle-complete graph Ramsey number r(Cm,K6). *Journal of Graph Theory* **44**: 251–260.
DOI: https://doi.org/10.1002/jgt.10145

Yali W., Yongqi S. & Radziszowski S.P. (2015). Wheel and star-critical Ramsey numbers for quadrilaterals. *Discrete Applied Mathematics* **185**: 260–271.
DOI: https://doi.org/10.1016/j.dam.2015.01.003

Zhang Y., Zhang Y. & Zhi H. (2023) A proof of a conjecture on matching. *AIMS Mathematics* **8**(4): 8027 –8033 DOI: https://doi.org/10.3934/math.2023406

RESEARCH ARTICLE

Deep Learning

Skeletal point analysis to determine the accuracy of forehand smash shots played by badminton players [†]

N Krishnaram, J Ahamed, N Sathyamoorthy, KD Sandaruwan and AMKB Athapaththu
University of Colombo School of Computing, UCSC Building Complex, 35 Reid Ave, Colombo 00700, Sri Lanka.

Abstract: This study aims to address the scarcity of scientific research on badminton performance analysis, specifically the accuracy of forehand smash shots. The authors propose the use of a skeletal coordinates-based technology to analyze a badminton player's biomechanics. To achieve this, specific techniques, such as formulating a quantitative description of badminton smash biomechanics based on the available literature, collecting video footage of badminton rallies and processing them using a MediaPipe-powered Python program, were followed. Three main approaches were considered for the analysis, defining a dynamic mathematical model, creating a player-to-player comparison model, and developing a machine-learning model. Preliminary results suggest that the use of three-dimensional points in comparison to two-dimensional points provides more accuracy in detecting the angle between three skeletal points from any camera perspective. This research also proposes a novel approach to compare two players and evaluate their skills based on a set of key parameters. The study explores the integration of machine learning algorithms to classify and predict player performance accurately. All three proposed methods enable coaches and players to identify and improve upon their weaknesses, enhancing their overall performance, as these findings have the potential to reduce subjectivity in measuring shot accuracy during training and to provide players with a more objective means of evaluating their performance. The proposed methodology and results contribute to a better understanding of badminton biomechanics and have implications for future research in this field.

Keywords: Badminton forehand smash, biomechanics, machine learning, mathematical model, mediapipe, shot analysis.

INTRODUCTION

Badminton is a popular racquet sport (He & Gong, 2022) in which players experience multiple intense actions and specific movement patterns, including numerous explosive shifts over short distances (Rusdiana, 2021). In badminton, actions such as 'smash' and 'drop' must be mastered as they can heavily influence in changing the pace of the game in favour of the player (He & Gong, 2022). Out of various smash action styles in badminton, this study is about 'forehand smash' as it is an overhead shot performed at full power by attacking the opponent in a dive movement, thus making it one of the game-changer actions.

According to the literature, a successful smash shot can be divided into four distinct phases (Rusdiana, 2021) as shown in Figure 1: preparation, acceleration, contact point and follow-through.

Recognizing body posture and motion is a key physical function for maintaining body balance (Zeng & Zhao, 2011). The integration of sports science into coaching has been pivotal in the significant advancements in international sports performance over the past two decades, particularly in analyzing players' biomechanics to identify and rectify shot-related mistakes (Vora, 2018). Several studies have been conducted to investigate the
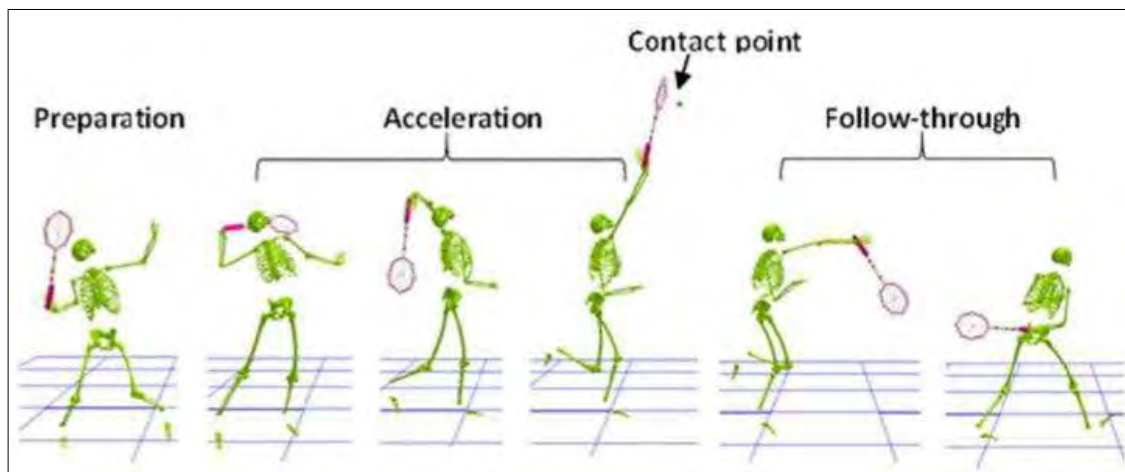
biomechanics of badminton, using various sensors, cameras, reflective markers, and motion analysis software

to analyze different shots and techniques (Salim *et al*., 2010; Hsueh *et al*.. 2012; Rusdiana *et al*., 2020).

Figure 1:   Phases of badminton forehand smash (Zhang *et al*., 2016)

Analyzing a player's playing style, performance indicators, and so on has been observed in many popular sports, including badminton. However, the majority of such performance analysis methods such as skeletal point based technologies, require a player to be surrounded by multiple pieces of equipment, or to wear external peripherals that can help track their performance based on their body movements, heart rate, and so on. Although these technologies have been shown to be effective (Ananth *et al*., 2019), it is challenging for a player to perform at their peak while wearing these peripherals.

However, there is relatively less research on pose recognition and video analysis (He & Gong, 2022). In comparison to other racquet sports, badminton has received very little scientific research attention (Tan *et al*., 2016). There is a demand for a plug and play technology that can help track the performance of a badminton player without relying on any external peripherals, especially when playing a sport like badminton, where a player's weight and accessories such as the racquet and shuttlecock are very important in affecting their optimal gameplay. Furthermore, based on the market conditions at the time of this research, these external peripherals could cost the players a fortune.

The aims of this research are twofold: to reduce subjectivity in measuring the accuracy of badminton shots during training sessions and to empower players to evaluate their shots independently. This will be

achieved by developing a simplified dynamic model that can effectively identify instances where a shot is played inaccurately.

Related work

The smash shot is a crucial technique in badminton, particularly the forehand smash. Mastering this shot is essential for players to gain points effectively (Kurnia *et al.*, 2020; Rusdiana *et al.*, 2021). Putra & Lumintuarso (2020) have investigated the biomechanical principle of the forehand badminton smash. Biomechanical studies have shown its significance and effectiveness in producing a fast shuttlecock rate (Vora, 2018).

In any sport, the main target of the players and coaches is to achieve the maximum level of success, and the aid of science and technology can be used to achieve it. What they expect is a way to evaluate the skill movements and the ability to correct themselves (Bartlett, 2021). It is discovered that this kind of biomechanical analysis is useful to improve the efficacy of approaches as well as the development of new and sustainable motions in technological development (Lu & Chang, 2012).

A recent study conducted by Ghazali *et al.* (2022) focused on stroke classification in badminton, employing an inertial sensor and a machine learning approach, whereas Kurnia *et al.* (2020) explored the movement patterns of the drop shot and smash in badminton using

motion sensors such as accelerometers and gyroscopes. Inertial Measurement Unit (IMU) sensors have also been used in several research projects (Ananth *et al*. 2019). Moreover, reflective markers were employed in several studies (Salim *et al*. 2010; Hsueh *et al*., 2012; Rusdiana, *et al*. 2020) to capture the biomechanical points of badminton players. These markers enable precise tracking and analysis of specific body movements and joint angles during gameplay. According to a review (Adesida *et al.*, 2019), there is growing interest in using wearable technology to quantify kinetic and kinematic characteristics in sports, to better understand movement and discern skill levels.

However, Alderson (2015) claims that the developed marker-less system is capable of estimating accurate joint kinematics in a variety of blind body pose estimation scenarios (*i.e.*, sporting activities) and provides an exciting and promising foundation for the non-invasive on-field measurement of athletes during match play. Using deep learning models (GoogleNet, Vgg-16 and RestNet 18) Ying *et al*. (2022) have performed research to recognise badminton smashing through video performance, where they were able to achieve the best accuracy of 97.51% and 98.86% on training and testing data sets, respectively, on ResNet-18. Tsai *et al.* (1998) used Direct Linear Transformation (DLT) to calculate the 19 3D coordinates for the segment endpoints and racquet from the recorded video. Many studies (Chu & Situmeang, 2017; Rahmad *et al*., 2020; Ying *et al*., 2022) have used video footage from the Badminton World TV YouTube channel with a camera angle which focuses in a top-down manner. Also, video-based analysis has been used not only to detect players, but also to detect the court (Chu & Situmeang, 2017), identify badminton shot events using shuttlecock trajectory data from recorded videos (Ju *et al*., 2020) and provide game statistics.

If we consider other racquet sports such as Tennis, Squash, Table Tennis, and Racquetball, studies have been done to identify shots using IMU sensors (Torres-Luque *et al*., 2011; Sharma *et al*., 2018; Williams *et al*., 2021), compare tennis serves between professional and elite players using MediaPipe (Liu & Sun, 2022), analyse backhand strokes of 3 different levels of players (Ghani *et al.*, 2019), determine the number of shots in a rally through recorded video footage (Torres-Luque *et al*. 2011), etc.

Human pose estimation (HPE) is a popular research topic in computer vision, with applications in video surveillance, medical assistance, and sports motion analysis. Numerous HPE libraries have been developed to detect and extract motion from real-world, and the usage of depth cameras and pose estimation models have shown high reliability for such marker-less motion-capturing applications. There are various skeleton-based HPE libraries available, each with its own set of advantages and disadvantages (Chung *et al.*, 2022).

MATERIALS AND METHODS

Figure 2 shows a top-down block diagram that provides a high-level overview of the research methodology employed in this study.

The research team conducted interviews with domain experts and external advisors to investigate the biomechanics of a badminton player while playing a forehand smash shot. It was revealed that coaches often assess angles between body parts such as elbow, shoulder joint, wrist and so on, yet a quantified description of these biomechanics was lacking in existing badminton resources. Thus, the research team bridged this gap by translating theoretical knowledge into mathematical conditions, taking the distances between joints of the body into account, and calculate the angles between them. Choosing suitable technology for detecting skeletal points on a player's body is crucial for obtaining accurate and reliable data for analysis, considering factors such as the nature of detected skeletal points (2D or 3D coordinates), its performance with minimal hardware specifications, and so on.

The study introduces the simple dynamic model, which involves evaluating calculated angles against predefined conditions. If these conditions are met, the shot is deemed to meet the required standards, offering a precise quantitative approach to assessing forehand smash shot biomechanics. Additionally, the angles obtained from one player's analysis can be compared with those of another, enabling a player-to-player comparison model for self-evaluation and technique improvement. This player-to-player comparison model can be used by coaches and players alike as a tool to improve their technique by referencing the ideal biomechanics of a badminton smash shot. Alternatively, the extracted skeletal points can serve as a dataset for machine learning algorithms, which can predict shot accuracy without relying on mathematical formulas.

However, to develop these models, it is crucial to select a suitable technology to extract 3D skeletal points from a data set comprising forehand smash shots played

by professional players. The selection of an appropriate perspective to record the shots, a suitable aspect ratio, preprocessing techniques such as cutting the videos into separate shots, and normalization of the shots are some of the data collection and preprocessing methodologies that should be followed. Furthermore, the generated results from all three approaches will be evaluated with expert opinion to determine their accuracy and feasibility in analyzing a badminton forehand smash shot.
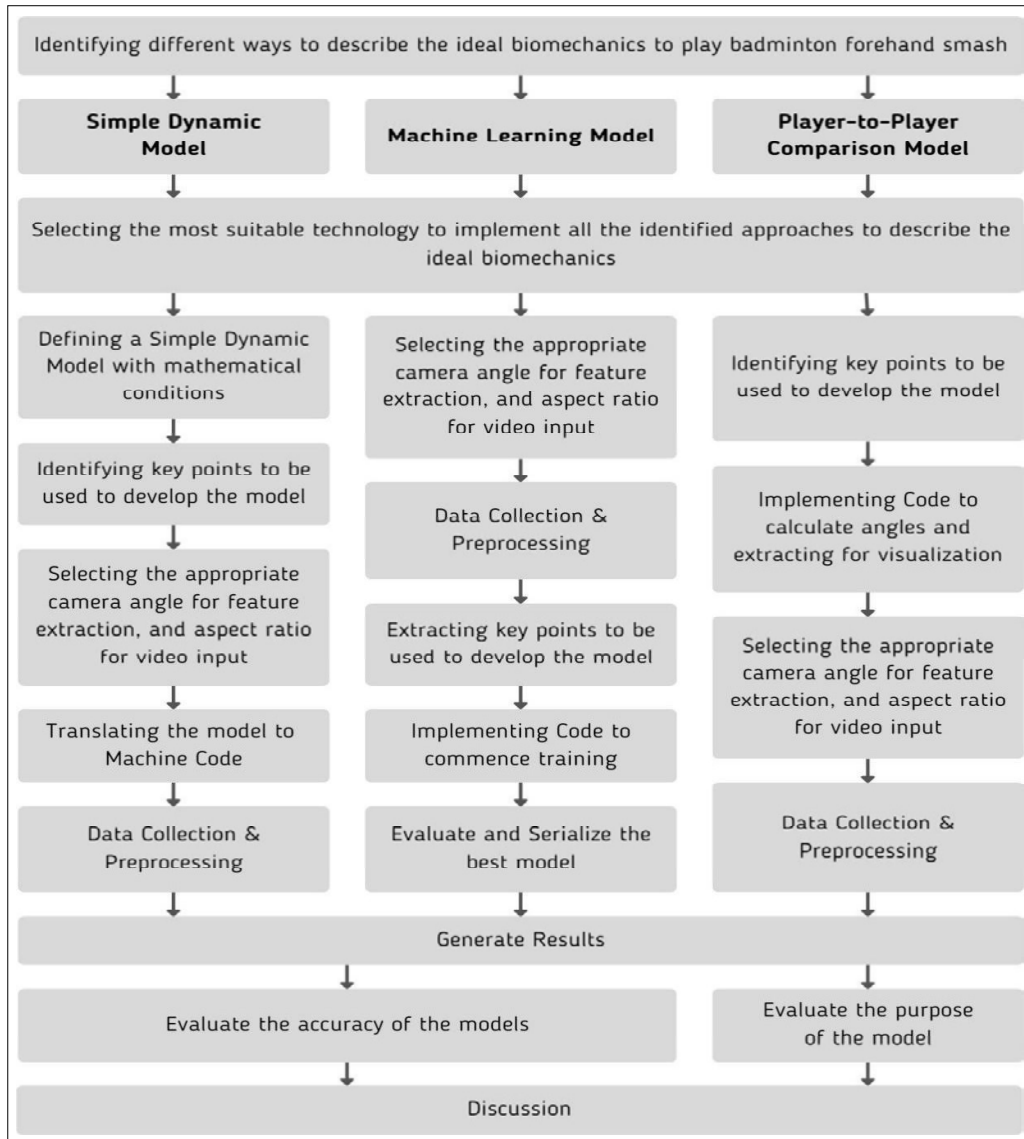

Figure 2: High-level overview of the research methodology

Implementation

Authors have proposed three approaches to describe the ideal biomechanics of a badminton smash shot: defining a simple dynamic model, developing a player-to-player comparison model, and building a machine-learning model.

Selecting the suitable technology

This study reviewed the official documentation of some of the Human pose detection libraries, and Table 1 depicts the overall comparison among the popular technologies. Among the 4 most popular computer vision technologies used for human pose estimation, BlazePose excels in multiple areas, including optimized performance for both single and multiple-person detection, low flickering, CPU-edged device compatibility, and high processing speed, thus making it the suitable technology for human pose estimation for this study. MediaPipe is a technology that is built using BlazePose. Throughout the research MediaPipe was used to extract the skeletal points.

Definition and development of a simple dynamic model

To achieve a simple dynamic model in which body movements are mathematically described, the authors translated theoretical knowledge from available literature and YouTube explanatory videos into a set of mathematical conditions, and these conditions were converted into a MediaPipe-powered Python Program.

**Table 1:** Technology Comparison Matrix

Technology	Single/Multiple	2D/3D	Approach	CPU/GPU	Flicker/Jitter	Key points	Speed	Optimized
BlazePose (Bazarevsky *et al.*, 2020)	Single	2D 3D	Bottom-up	CPU	Low	33	High	Yes
OpenPose 2D (Cao *et al.*, 2018)	Multiple	2D	Bottom-up	GPU	Low	135	Low	No
OpenPose 3D (Nakano *et al.*, 2020)	Single	3D	Bottom-up	GPU	Low	135	Low	No
AlphaPose 3D (Fang *et al.*, 2022)	Multiple	2D 3D	Top-down	GPU	High	26	Low	No
YOLOv7 3D (Wang *et al*, 2022)	Multiple	2D 3D	Top-down Bottom-up	GPU	High	17	Low	No

Defining a simple dynamic model

To identify a forward smash attempt as successful throughout the four stages of the shot, the following set of conditions must be satisfied (note: all angles are presented in degrees):
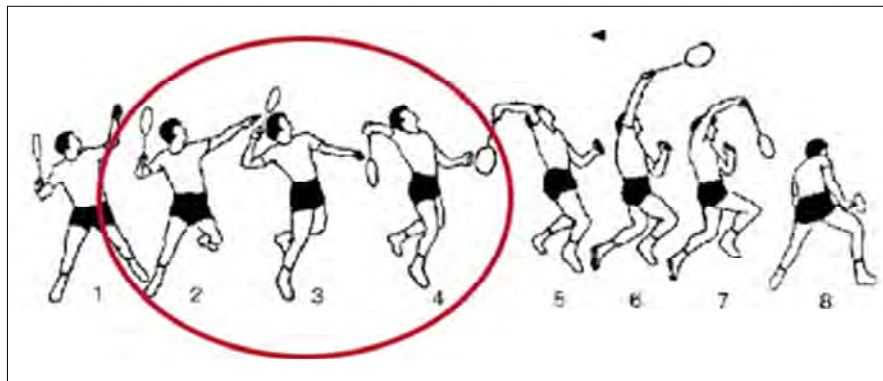
Preparation phase
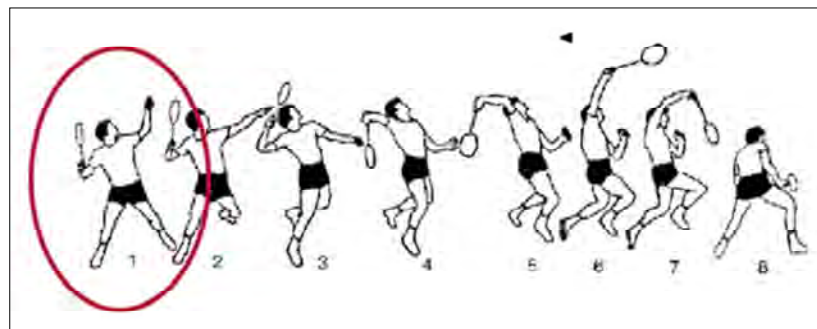
Figure 3: Smash phase - Preparation phase (Available at: https://jordynhealy.wordpress.com/biomechanical-analysis/)

Angle in between the dominant elbow -
(let wrist = W, elbow = E, shoulder = S)

 = Angle *WES* $\leq 75°$ and

 = Angle WES $\geq 30°$

Angle in between dominant underarm -
(let elbow = E, shoulder = S, hip = H)

 = Angle *ESH* $> 90°$ and

 = Angle *ESH* $\geq 45°$

Angle in between non-dominant underarm -
(let elbow = E', shoulder = S', hip = H')

 = Angle E'S'H' $> 90°$

Dominant legs should be behind the body. The non-dominant leg should be in front. -
(let ankle = A, hip = H, non dominant hip = H')

 = Angle AHH' $> 90°$

Angle between dominant arm and chest -
(let elbow = E, shoulder = S, non-dominant shoulder = S')
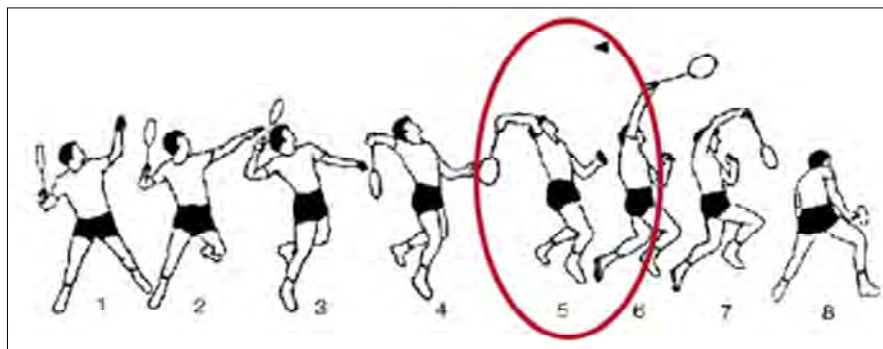
 = Angle ESS' $> 135°$ (State 01)

Record the angle in between the dominant wrist -
(let elbow = E, wrist = W, index = I)

 = Angle EWI   (State 01)

Acceleration phase

Hip should move forward, giving a force
Elbow should come forward, following the hip -
(let elbow = E, shoulder = S, non-dominant shoulder = S')

 = Angle ESS' $< 135°$ (*State* 02)

Difference between State 01 and State 02

 = |State 01 - State 02| $\geq 30°$ and

 = |State 01 − State 02| $\leq 60°$

The racquet should go behind, *i.e.*, Wrist angle should change

Record the angle in between the dominant wrist -
(let elbow = E, wrist = W, index = I)

 = Angle EWI (State 02)

Difference between State 01 and State 02 of Angle EWI

 = |State 01 - State 02| $\geq 15°$

Figure 4:   Smash phase - Acceleration phase (Available at: https://jordynhealy.wordpress.com/biomechanical-analysis/)

Figure 5:   Smash phase - Contact point (Available at: https://jordynhealy.wordpress.com/biomechanical-analysis/)

Contact point

Angle between the dominant elbow -
(let wrist = W, elbow = E, shoulder = S)
    = Angle WES > 135˚
The racquet should come forward. *i.e.*, Wrist angle should change. Record the angle in between the dominant wrist
(let elbow = E, wrist = W, index = I)
    = Angle EWI (State 03)
Difference between State 01 and State 02 of Angle EWI
   = | State 02 - State 03 | ≥ 15

Follow through

The dominant arm should make it to the very front, completing the rotation. Angle Between Dominant underarm -
(let elbow = E, shoulder = S, hip = H)
    = Angle ESH < 30
The dominant leg should be in front of the body. The non-dominant leg should be behind. -
(let ankle = A, hip = H, non dominant hip = H')
    = Angle AHH' ≥ 90

These mathematical conditions, which satisfy the ideal biomechanics of the smash shot, are based on the theoretical definition given by Putra & Lumintuarso (2020) on the biomechanical principle of the forehand smash in badminton, as well as a YouTube tutorial on smash and clear by Wadenka (2019), a professional badminton player from Germany.
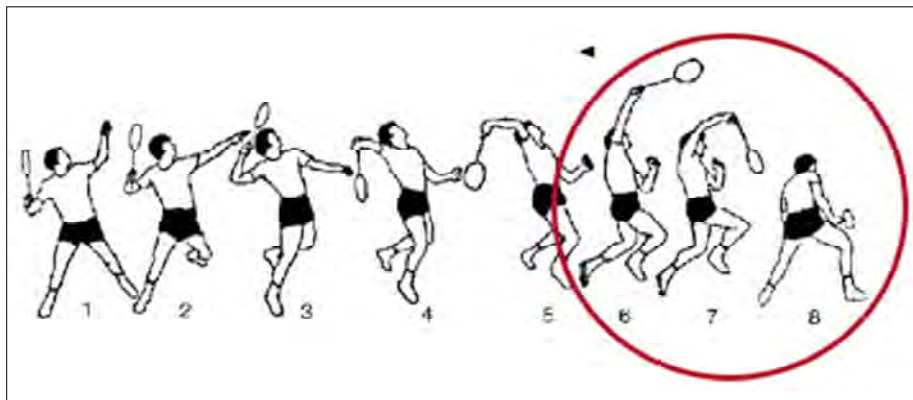
Figure 6: Smash phase - Follow through (Available at: https://jordynhealy.wordpress.com/biomechanical-analysis/)

Key points used for the features of the simple dynamic model

To calculate the mathematical relationships between the body joints and identify the ideal biomechanics for a forehand smash, the researchers have decided to focus on the 12 key points (pose landmarks) highlighted in Figure 7 based on the theoretical description of the ideal biomechanics.

Ten different angles based on the 12 skeletal points, including the right and left elbow angles, right and left underarm angles, right and left leg angles, right and left chest angles, and right and left wrist angles will be calculated.

Deriving the angle between coordinates in degrees using the law of cosines
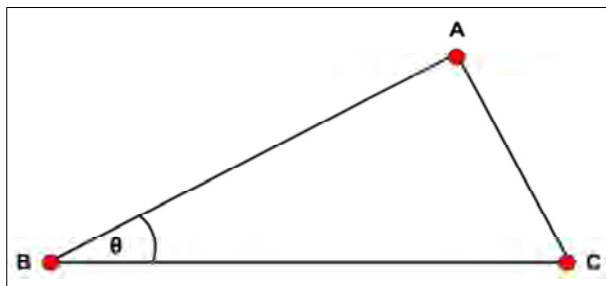
The Law of Cosines, as shown in Figure 8, is a mathematical formula used to find the lengths of the sides and the measures of the angles of a non-right triangle.

In mathematical notation, the Law of Cosines can be expressed as equation (1):

$$(Distance(A, C))^2 = (Distance(A, B))^2 + (Distance(B, C))^2$$
$$- 2(Distance(A, B))(Distance(B, C)) \cos(\theta)$$

$$...(1)$$
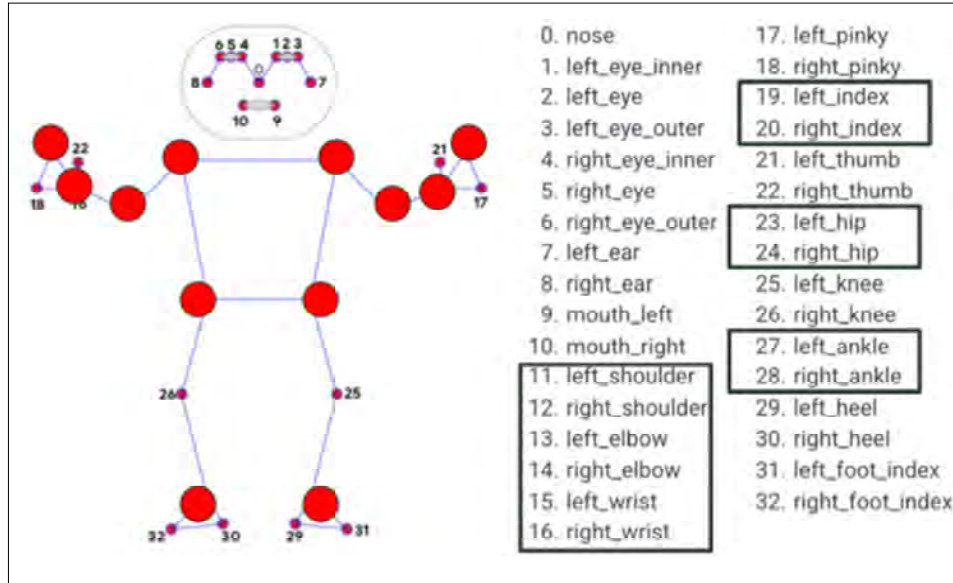
Figure 7:   12 Key points mainly focused on this research (Available at: https://github.com/google/mediapipe/blob/master/docs/solutions/pose.md)

Figure 8:   A triangle depicting the Law of Cosines

where Distance(A,C) is the length of the side opposite the angle, and Distance(A,B) and Distance(B,C) are the lengths of the other two sides of the triangle, and θ is the angle opposite to the side Distance(A,C).

In the study, the researchers utilize the Law of Cosines to compute the angle between three three-dimensional skeletal points. This calculation is intended to assess the degree of deviation between the skeletal points obtained from two distinct perspectives. The equation will be suitably adapted to account for this purpose, as following equation (2):

$$\cos(\theta) = \frac{(Distance(A,B))^2 + (Distance(B,C))^2 - (Distance(A,C))^2}{2 \times Distance(A,B) \times Distance(B,C)} \quad ...(2)$$

Now, the distance between two points in a three-dimensional space can be calculated using the Pythagorean Theorem, which states that the square of the hypotenuse of a right-angled triangle is equal to the sum of the squares of its two other sides. In a 3D space, the distance between two points (x_1, y_1, z_1) and (x_2, y_2, z_2) can be found using the following formula (3):

$$Distance = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad ...(3)$$

This formula involves finding the difference between the coordinates of the two points in each dimension, squaring them, adding them together, and then taking the square root of the sum to obtain the distance. Using this formula, we can derive the distances in between three 3D coordinates A, B and C as following equations:

$$Distance(A,B) = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2}$$

$$Distance(B,C) = \sqrt{(x_C - x_B)^2 + (y_C - y_B)^2 + (z_C - z_B)^2}$$

$$Distance(A, C) = \sqrt{(x_C - x_A)^2 + (y_C - y_A)^2 + (z_C - z_A)^2}$$

After substituting the values from the distance formula into the modified formula of the law of cosines, the angle between the three 3D points can be obtained by applying the following formula (4):

Angle in between three 3D points = $\cos^{-1}(\cos(\theta))$ ...(4)

It is important to note that this angle will be represented in radians. To convert the angle value from radians to degrees, the following formula will be applied as in (5):

Angle in degrees = $\left| \theta \times \dfrac{180}{\pi} \right|$, where θ is in radians. ...(5)

By applying the above formula, the angle in degrees can be calculated, which can be used to determine the deviation between the skeletal points extracted from two different perspectives.

Data collection and preprocessing techniques: on-site data collection

For data collection, the researchers approached a segment of university-level badminton players consisting of 12 right-handed male players, as recommended by the badminton coach of the University of Colombo (UoC). After obtaining consent, each player was requested to play a rally of forehand smash shots, which were then recorded using a Redmi Note 10 Pro smartphone camera. The rallies were recorded in 1080p resolution and at a rate of 30 frames per second.
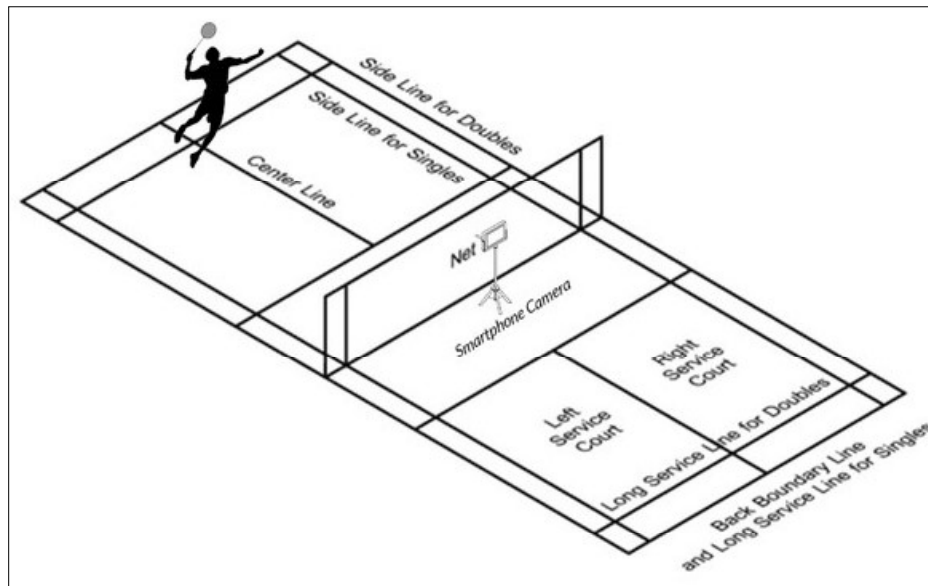
Figure 9: Field setup (Available at: https://commons.wikimedia.org/wiki/File:Badminton_court_3d_small.png)

Field setup and training protocol followed prior to on-site data collection

The smartphone camera, securely mounted to a tripod, was strategically positioned at the center and in front of the badminton court net to record the player's full-body movements as shown in figure 9. After comparing skeletal coordinates extracted from various angles, the decision was made to front-face the camera. This orientation captures all necessary skeletal points of the player, as MediaPipe relies on face recognition to map the remaining coordinates in the body. Next, every right-handed player faced the smartphone camera while another player in the opposite end of the court shot 15-20 shuttlecocks one after the other, allowing the first player to perform forehand smashes towards the incoming shuttlecocks. Imperfect smash shots in the rallies would then be evaluated by the badminton coach. The coach viewed recorded shots individually, assisting the research team in categorizing perfect smashes and discarding the rest.

Data preprocessing

The recorded videos were cropped to a 1:1 aspect ratio and the 12 rallies were segmented into 227 short clips of individual smash shots using Adobe Premiere Pro Creative Cloud (CC) 2019. These were then exported as separate videos. These videos were later reviewed by the domain expert, who assisted in accurately classifying the shots as either 'smash' or 'non-smash.' Some of the 'smash' shots that were identified by the badminton coach as perfect were used to evaluate the accuracy of the proposed Simple Dynamic model.

Normalizing the video clips to 100 frames

After careful observation, it was noted that while the duration of a badminton forehand smash shot may vary among players, it was determined that normalizing the video clip of a shot to 100 frames would be appropriate. This would ensure that the processed video would have a length of 3 seconds plus 10 frames, given a frame rate of 30 frames per second. This approach would allow for a more accurate evaluation of the biomechanics of the forehand smash shot, as each stage of the shot would be represented proportionally across all shots.

Development of a player-to-player comparison model

A player-to-player comparison model is developed by comparing angle values obtained from the biomechanical analysis of two different players' forehand smash shots. This model calculates and visually represents the deviation between these angle sets, providing a valuable tool for coaches to assess player performance or for players to self-evaluate.

To achieve this, angle values from a reference video showcasing the ideal forehand smash were extracted and saved in a CSV file, as were angle values from another player's body parts. Both videos underwent preprocessing and normalization to ensure consistent dimensions in the CSV files. The angle values from these files were imported into separate numpy arrays, and the deviation between them was calculated and stored in another array. This deviation data was then used to create a heatmap, which graphically illustrates differences between the angle sets and pinpoints areas for potential technique improvement in the player.

Building a machine learning model

By installing Mediapipe, OpenCV-python, Pandas Scikit-learn, the machine learning model will have access to the necessary tools to train and test models on data.

Detect skeletal points using Mediapipe

Computer vision and machine learning approaches were used to detect and track a person's body movements in real-time through a video feed. The Model first captures video passed, then initializes the Mediapipe Holistic model, which detects and tracks pose. Next, the model reads each frame from the captured video feed, and applies the Holistic model to it. The Holistic model then makes detections on the recoloured frame and identifies the position of each body part. The model then renders the body parts on it.

Capture landmarks and export to CSV

The pose landmarks are represented by a list of landmarks, each containing x, y, and z coordinates of the landmark, as well as a visibility measure for the landmark. The first column is named 'class' and the following columns are named in a sequential manner using a loop. The loop runs from 1 to 33, i.e the number of total skeletal coordinates (inclusive) and for each iteration, it appends four strings to the landmarks list: 'x', 'y', 'z', and 'v', along with the iteration number as a suffix. These four strings represent the x, y, z coordinates and the visibility of each landmark point. A total of 9017 frames extracted from the video footage dataset were divided into two classes, 'smash' and 'non smash', which have 5591 frames and 3426 frames each, respectively. Each frame has coordinates for 33 skeletal points. Each skeletal point has 3-dimensional coordinates and a visibility measure. So the total features in the data set are 33 x 4 = 132 features per frame. The target variable, called 'class', contains 2 classes as mentioned above.

Train custom model using Scikit Learn

The dataset is being preprocessed to create x and y, which will be used for training and testing the machine learning model. x is created by dropping the 'class' column from the dataset, while y is created by assigning the 'class' column of the dataset to the variable y, which will be used as the target variable for training and testing the machine

learning model, using the train test split method from the scikit learn library to split the data set into training and testing sets. In this case, the test data will contain 30% of the original data set, while the remaining 70% will be used for training.

Logistic regression and Ridge classifier are linear models, while the Random Forest classifier and Gradient Boosting classifier are tree-based models. The selection of machine learning models depends on this research problem which is to identify and evaluate the forehand standing smash shot, the nature of the data set which is very huge in the number of features to evaluate (132

features) and the specific objective of the research. In this case, since the target variable is categorical (*i.e.*, Smash or Non Smash), classification models would be more suitable than regression. It then fits the models using the training data and stores the models in the dictionary 'fit models'.

Make detections with model

After successfully training the model, new video footage was provided to the model as an input to check if the model accurately shows the probability of the played shot being a smash, as shown in Figure 10.
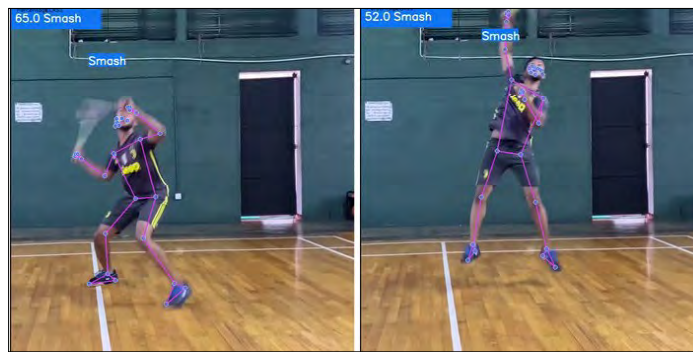

Figure 10: Trained machine learning model making detections

RESULTS AND DISCUSSION

Simple dynamic model

In this study, the biomechanical accuracy of the forehand smash shot in badminton was evaluated using two video clips of different players performing a perfect forehand smash. The clips were normalized to 100 frames and processed using a MediaPipe-powered Python

program, which calculated the angles between body parts. These angles were compared against predefined mathematical conditions mentioned earlier. The program outputted values of 1 or 0 based on whether the conditions were met or not, and the results were saved in a CSV file. The CSV file was then used to generate a graph as shown in Figure 11, where green indicated a value of 1 and red indicated a value of 0, representing the accuracy of the shot performed by player 1.

Figure 11: Colour-coded graph depicting the simple dynamic model of player 1 performing the shot

To assess the terminal pose accuracy of each phase, only the last 5 frames of each phase were considered. The accuracy of the terminal pose was calculated, and these values were cumulated to determine the overall accuracy of the entire forehand smash shot. The results were discussed with the domain experts, who confirmed the accuracy of the players' terminal poses based on their expertise.

The effectiveness of the simple dynamic model was further verified by testing it on player 2's video clip as shown in Figure 12, which also successfully aligned with the evaluation provided by the domain expert. Additionally, when compared with a video clip from a web source called Badminton TV, as shown in Figure 13, the model highlighted areas where the player needed to improve their technique at the follow-through stage.

Figure 12: Colour-coded graph depicting the simple dynamic model of player 2, to further confirm the accuracy of the model

Figure 13: Colour-coded graph depicting the simple dynamic model of a player footage that was extracted from the web

Player-to-player comparison model

The player-to-player comparison model for evaluating the biomechanical accuracy of the forehand smash shot in badminton involved selecting two video clips of different players. One video served as a reference for a perfect smash shot, while the other video was used to evaluate a player's technique by comparing it with the reference player. Both videos were normalized to 100 frames and processed using a MediaPipe-powered Python program to calculate the angles between body parts. The angle

values were saved in separate CSV files, which were then used to generate a graph.

The graph, depicted in Figure 14, visually represented the deviation between the angle values of the two players. The graph was colour-coded, with a range of colours between black and white, indicating the deviation in each cell. White represented negative deviation, while black represented positive deviation. Additionally, the graph was improved by further colour-coding to represent the four phases of the shot, enhancing the user experience.
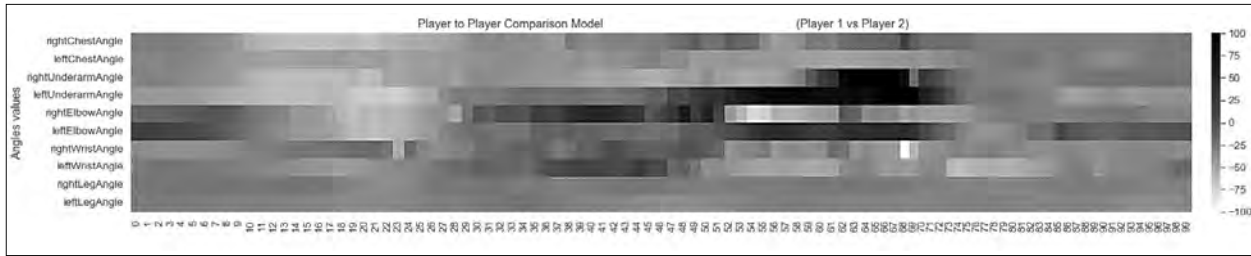
Figure 14: Colour-coded graph depicting the player-to-player comparison model
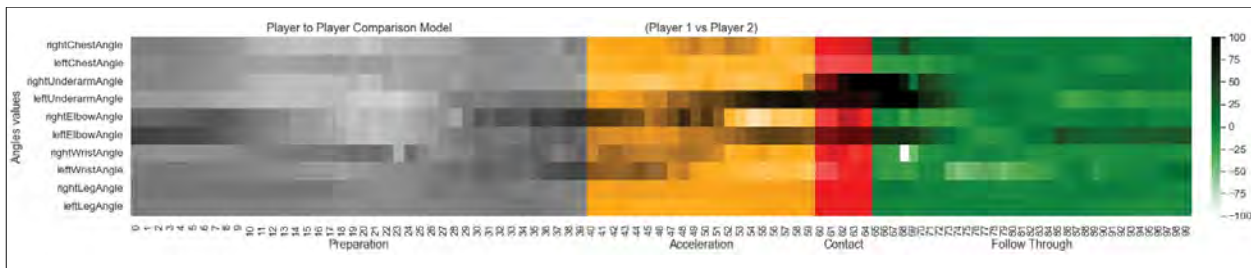

Figure 15: Colour-coded graph depicting the player-to-player comparison model, separated by phases

Figure 15 displays the graph showing the variation of deviation between angle values recorded by the two players at each frame of the video. This analysis demonstrated promising potential for assessing a player's biomechanics by comparing them to another player, providing valuable insights into their technique and performance.

Machine learning model

This research employed four types of machine learning (ML) models, namely logistic regression (LR), ridge classifier (RC), random forest classifier (RF), and gradient boosting classifier (GB), to analyze the data. To determine the best-performing model for this research, all the ML models were trained using a pipeline approach. Accuracy, F1 score, and the confusion matrix were utilized as evaluation metrics. The Confusion Matrix, derived from the results of all four algorithms, provided a comprehensive overview of the model's performance. Figure 16 represents the confusion matrix for each algorithm, where true positive (TP), false positive (FP), false negative (FN) and true negative (TN) values for respective models are shown.

		Predictions (System) - Logistic Regression				**Predictions (System) - Ridge Classifier**	
		Positive (Smash)	Negative (Smash)			Positive (Smash)	Negative (Smash)
Actual (Expert)	Positive (Smash)	TP = 1480	FN = 200	**Actual (Expert)**	Positive (Smash)	TP = 1532	FN = 148
	Negative (Smash)	FP = 457	TN = 569		Negative (Smash)	FP = 478	TN = 548
		Predictions (System) - Gradient Boosting				**Predictions (System) - Random Forest**	
		Positive (Smash)	Negative (Smash)			Positive (Smash)	Negative (Smash)
Actual (Expert)	Positive (Smash)	TP = 1611	FN = 69	**Actual (Expert)**	Positive (Smash)	TP = 1677	FN = 3
	Negative (Smash)	FP = 337	TN = 689		Negative (Smash)	FP = 71	TN = 955

Figure 16: Confusion matrices for LR, RC,GB and RF algorithms

Hyper parameter tuning for the selected classifiers

The purpose of hyper-parameter tuning is to enhance the model's performance on unseen data by avoiding over-fitting or under-fitting and improving its generalization ability. By optimizing the hyper-parameters, the model becomes better equipped to learn patterns from the training data and apply those patterns to new, unseen data.

Various methods can be employed for hyper-parameter tuning. In this research, hyper-parameter tuning was conducted using the random search method, and the results of the training process are presented in Figure 17.

The Simple Dynamic Model revealed that evaluating smash shots from a side view offers clearer insights into dominant arm movements, while a front view provides a better understanding of body rotation at the end of the shot. However, this research focuses on analyzing forehand smash shots from the front view since MediaPipe requires face detection to identify skeletal points. The study suggests a comprehensive methodology that incorporates multiple angles to assess and record player body movements, contributing to analysing the shot better.

The player-to-player comparison model serves multiple purposes. Firstly, it enables the comparison of a
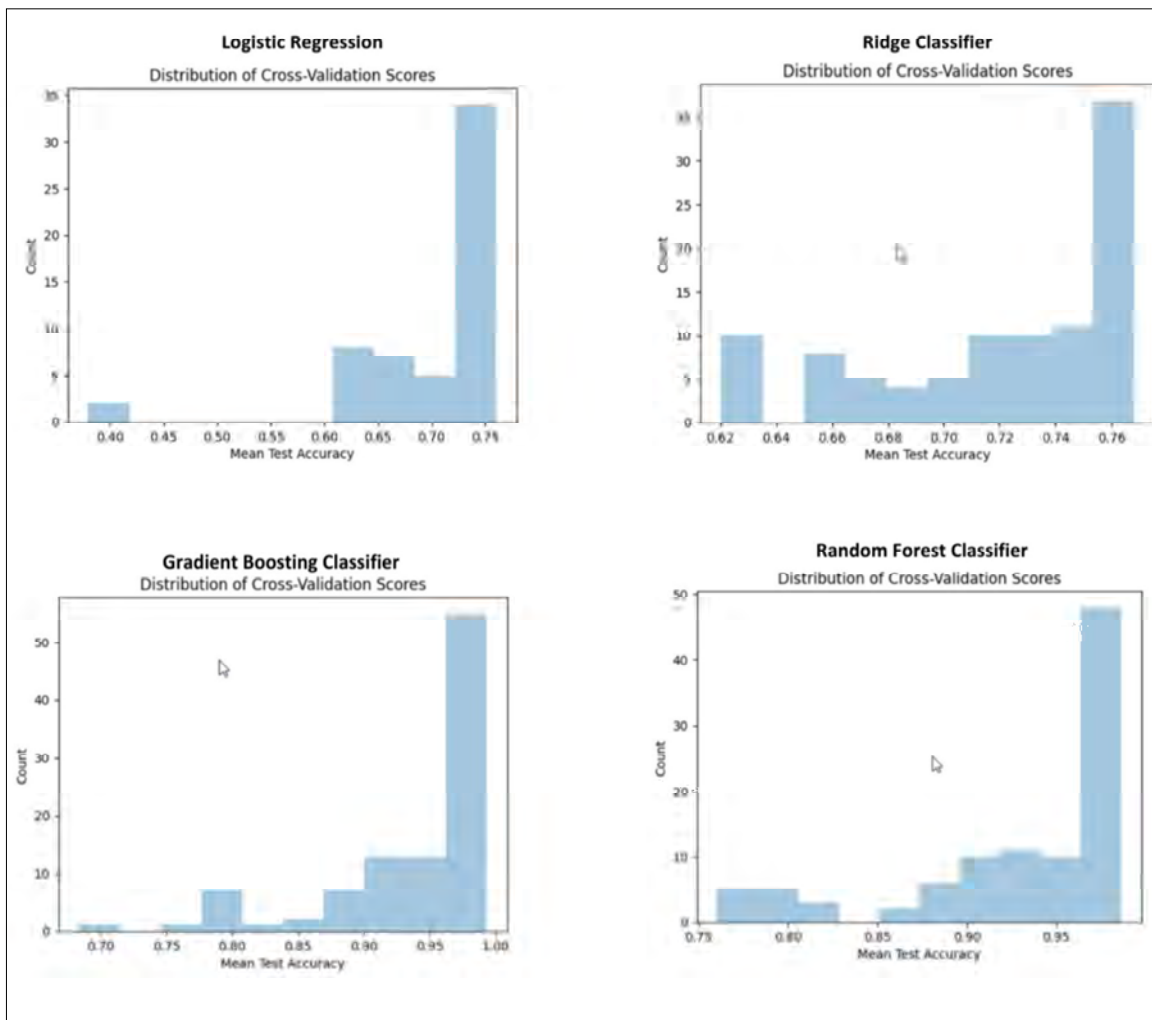

Figure 17: Distribution of cross-validation scores

player's performance by using an ideal performance as a reference, allowing for a qualitative assessment of their skills. Additionally, it facilitates the improvement of individual players by comparing their current performance with their previous ones, helping them identify areas of improvement. The model also provides valuable insights by analyzing the deviations between performances, allowing trainers and players to pinpoint specific areas that require correction or adjustment. Lastly, the model accommodates the evaluation of unorthodox approaches to playing the forehand smash, expanding the scope of analysis to include other playing styles.

After training all 4 models to determine the best performing classifier for the Machine Learning Model, various measures were found to identify the best model for this research purpose and the dataset we have used. Table 2 shows the overall comparison matrix.

Table 2:   Overall Performance metrics

Classifier	Accuracy	Error rate	Precision	Recall	Specificity	FP rate	FN rate	F1 score
Logistic regression	0.757	0.243	0.764	0.881	0.556	0.445	0.119	0.818
Ridge classifier	0.769	0.231	0.762	0.912	0.534	0.466	0.088	0.830
Gradient boosting	0.849	0.150	0.827	0.959	0.672	0.328	0.041	0.887
Random forest	0.974	0.027	0.959	0.998	0.931	0.069	0.001	0.980

As shown in table 2, Random Forest Classifier outperformed the other 3 models by comparatively high margin. This is one of the reasons to choose a Random Forest Classifier as the best model for this research.

The cross-validation scores across the hyper-parameter search space for all 4 classifiers are listed below.
- Logistic Regression: 75.11%
- Ridge Classifier: 77.11%
- Gradient Boosting Classifier: 99.33%
- Random Forest Classifier : 98.73%

Approximate time taken for hyper-parameter tuning process is as follows.

- Logistic Regression : 2.5 hours
- Ridge Classifier: 1.5 hours
- Gradient Boosting Classifier: 9 hours
- Random Forest Classifier : 4.5 Hours

The results indicate that Logistic Regression and Ridge Classifier performed relatively poorly on the dataset used in this research. On the other hand, both Random Forest Classifier and Gradient Boosting Classifier showed high accuracy scores, with little difference between them. However, it is important to note that the hyper-parameter tuning process took nearly double the time for the Gradient Boosting Classifier compared to the Random Forest Classifier. Considering these factors, the Random Forest Classifier was chosen as the best-performing algorithm for this study.

In the initial stage of the research, videos from university-level badminton players and "Badminton World TV" YouTube videos were used to train the model. This resulted in a 100% accuracy for all four models. However, further investigation revealed that the dataset contained both smash and non-smash shots from different sources, leading to this perfect accuracy. To address this, the dataset was reproduced using only manually recorded videos, including non-smash shots recorded with the help of university-level badminton players. Initially, the model provided an output for each frame, indicating whether it was a smash or non-smash shot and the corresponding probabilities. Later, it was modified to only display the probability of being a smash shot.

There were issues with extracting skeletal points using Mediapipe, as some frames had low visibility or flickering, resulting in missing or incorrect coordinate records in the CSV file. Also, training the model by dividing the smash shot into phases was unsuccessful due to the clip duration being too small for MediaPipe.

Recommendations, limitations, and future directions

Based on the research findings, several recommendations are proposed to enhance the evaluation of badminton forehand smash shots and improve player performance assessment. Firstly, it is recommended to use the developed Simple Dynamic Model and Player-to-player Comparison Model for evaluating single-player rallies, offering systematic approaches for assessing shot

accuracy, particularly under orthodox conditions or with players of unorthodox styles.

The recording of shots at a frame rate of 30 fps may not capture all the intricate details of the shot, resulting in blurred body images and potentially impacting the visibility of the detected skeletal points by MediaPipe, leading to inaccurate results. To address this, using a higher frame rate and normalizing frames to a larger number can be considered, although it may affect the processing speed, particularly on hardware with limited specifications. Therefore, a balance must be struck between accuracy and processing speed when selecting parameters for the program. At the moment, to ensure compatibility with these models and Mediapipe technology, it is advised to use video clips with a frame rate of 30 fps or lower and shots recorded in the front view.

The research also has other limitations, including the focus on the front angle, visibility issues, and challenges in machine learning model training. Future research can be expanded to include diverse player demographics, other types of smash shots, and additional badminton shot evaluations. Improvements in technology, accurate shot phase prediction, and the inclusion of shot intensity as an assessment feature are suggested.

Furthermore, evaluating two-player rallies to assess team performance and doubles play dynamics is recommended to provide a more comprehensive understanding of badminton performance. These recommendations and future research directions aim to broaden the scope of analysis and contribute to comprehensive assessment frameworks in badminton.

CONCLUSION

This research introduces a cost-effective method for athletes to independently assess their badminton smash shot performance, reducing reliance on external resources or expertise. Simultaneously, it offers coaches a streamlined, data-driven approach to evaluate players during forehand smash shots, minimizing subjectivity for more accurate feedback. The proposed methodology eliminates the need for extensive manual analysis and external peripherals, thereby reducing the resources required for performance assessment. The simple dynamic model described in this research is the first of its kind, as this is the first attempt to quantitatively describe a badminton shot. The study makes a significant contribution to the badminton performance analysis domain.

REFERENCES

Adesida Y., Papi E. & McGregor A.H. (2019). Exploring the role of wearable technology in sport kinematics and kinetics: A systematic review. *Sensors* **19**(7): 1597.
DOI: https://doi.org/10.3390/s19071597

Alderson J. (2015). A markerless motion capture technique for sport performance analysis and injury prevention: Toward a 'big data', machine learning future. *Journal of Science and Medicine in Sport* **19**: e79.
DOI: https://doi.org/10.1016/j.jsams.2015.12.192

Ananth S.V., Patawari A., Sundari K., Raghavi, Likitha S., & Arivalagan S. (2019). Biomechanical Sports Analysis Using Inertial Measurement Unit. *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)* 29-30 March. Pondicherry, India, pp. 1–5.
DOI: 10.1109/ICSCAN.2019.8878691

Bartlett R. (2021). *Introduction to Sports Biomechanics: Analysing Human Movement Patterns,* 2nd edition. Routledge, London, United Kingdom.
DOI: https://doi.org/10.4324/9781315889504

Bazarevsky V., Grishchenko I., Raveendran K., Zhu T., Zhang F. & Grundmann M. (2020). BlazePose: On-device real-time body pose tracking. *arXiv preprint arXiv:2006.10204.*
DOI: https://doi.org/10.48550/arXiv.2006.10204

Cao Z., Hidalgo G., Simon T., Wei S. & Sheikh Y. (2021). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **43**(1): 172–186.
DOI: 10.1109/TPAMI.2019.2929257

Chu W. & Situmeang S. (2017). Badminton video analysis based on spatiotemporal and stroke features. *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, Bucharest, Romania, 6-9 June. Association for Computing Machinery, New York, NY, USA, pp. 448–451.
DOI: https://doi.org/10.1145/3078971.3079032

Chung J., Ong L. & Leow M. (2022). Comparative analysis of skeleton-based human pose estimation. *Future Internet* **14**(12): 380.
DOI: https://doi.org/10.3390/fi14120380

Fang H.S., Li J., Tang H., Xu C., Zhu H., Xiu Y., Li Y.L. and Lu C. (2022). AlphaPose: Whole-body regional multi-person pose estimation and tracking in real-time. *IEEE*

Transactions on Pattern Analysis and Machine Intelligence **45**(6): 7157–7173.
DOI: 10.1109/TPAMI.2022.3222784

Ghani D.Z.A., Zainuddin Z.A. & Ibrahim H. (2019). Squash backhand stroke analyses for three different playing levels in Malaysia. *Proceedings of the 3rd International Colloquium on Sports Science, Exercise, Engineering and Technology* (eds. Sulaiman N., Ismail S. & Adnan R.), Singapore. Springer, Singapore, pp. 17–23.
DOI: https://doi.org/10.1007/978-981-10-6772-3_3

Ghazali N. F., Shahar N. & As'Ari M.A. (2022). Badminton strokes recognition using inertial sensor and machine learning approach. *2022 2nd International Conference on Intelligent Cybernetics Technology and Applications* (ICICyTA), Bandung, Indonesia. IEEE, pp. 1–5.
DOI: 10.1109/ICICyTA57421.2022.10037897

He S. & Gong R. (2022). Recognition and prediction of badminton attitude based on video image analysis. *Mobile Information Systems* **2022**: 1–9.
DOI: https://doi.org/10.1155/2022/6960343

Hsueh Y., Chen Y., Pan K. and Tsai C. (2012). Biomechanical analysis of badminton forehand net shots. *30th Annual Conference of Biomechanics in Sports*, Melbourne, Australia. pp. 256–259.

Ju N., Yu D., Ik T. & Peng W. (2020). Trajectory-based badminton shots detection. *2020 International Conference on Pervasive Artificial Intelligence (ICPAI)*, 3-5 December. Taipei, Taiwan, IEEE, pp. 64–71.
DOI: 10.1109/ICPAI51961.2020.00020

Kurnia R., Pasmai A. & Elfitri I. (2020). Analysis of the smash and dropshot movement patterns in badminton sports using Pearson correlation. *Proceedings of the 2020 10th International Conference on Biomedical Engineering and Technology (ICBET '20)*, Tokyo, Japan. Association for Computing Machinery, New York, NY, USA, pp. 164–169.
DOI: https://doi.org/10.1145/3397391.3397434

Liu F. & Sun Y. (2022). A Tracing-based Tennis coaching and smart training platform using artificial intelligence and computer vision. Artificial intelligence and fuzzy logic system. *8th International Conference on Artificial Intelligence and Fuzzy Logic System (AIFZ 2022)* (eds. Wyld D.C. & Nagamalai D.), Toronto, Canada, 24–25 September. Academy and Industry Research Collaboration Center (AIRCC), pp. 89–100.
DOI: https://doi.org/10.5121/csit.2022.121608

Lu T. & Chang C. (2012). Biomechanics of human movement and its clinical applications. *The Kaohsiung Journal of Medical Sciences* **28**(2): S13–S25.
DOI: https://doi.org/10.1016/j.kjms.2011.08.004

Nakano N., Sakura T., Ueda K., Omura L., Kimura A., Iino Y., Fukashiro S. & Yoshioka S. (2020) Evaluation of 3D markerless motion capture accuracy using openpose with multiple video cameras. *Frontiers in Sports and Active Living* **2**: 50.
DOI: 10.3389/fspor.2020.00050

Putra G.E. & Lumintuarso R. (2020). What is the Biomechanical principle of forehand smash in badminton? *Physical Education and Sport Through the Centuries* **7**: 221–229.
DOI: https://doi.org/10.2478/spes-2020-0018

Rahmad N.A., As'ari M.A., Soeed K. & Zulkapri I. (2020). Automated badminton smash recognition using convolutional neural network on the vision based data. In: IOP Conference Series: Materials Science and Engineering. *Sustainable and Integrated Engineering International Conference 2019 (SIE 2019)* 8–9 December. The Everly Putrajaya, Putrajaya, Malaysia, pp. 1–8.
DOI: 10.1088/1757-899X/884/1/012009

Rusdiana A., Abdullah M.R.B., Syahid A.M., Haryono T. & Kurniawan T. (2021a). Badminton overhead backhand and forehand smashes: a biomechanical analysis approach. *Journal of Physical Education and Sport* **21**(4): 1722–1727.
DOI: 10.7752/jpes.2021.04218

Rusdiana A., Subarjah H., Badruzaman B., Budiman D., Wibowo R., Nurjaya D.R., Pramutadi A., Mustari A., Kusdinar Y. & Syahid A. (2021b). Kinetics analysis of overhead standing smash in badminton. *Jurnal Pendidikan Jasmani dan Olahraga* **6**(1): 81–88.
DOI: https://doi.org/10.17509/jpjo.v6i1.32577

Rusdiana A., Subarjah H., Imanudin I., Kusdinar Y., Syahid A.M. & Kurniawan T. (2020). Effect of fatigue on biomechanical variable changes in overhead badminton jump smash. *Annals of Applied Sport Science* **8**(s1): e895.
DOI: http://dx.doi.org/10.29252/aassjournal.895

Salim M.S., Lim H.N., Salim M.S.M. & Baharuddin M.Y. (2010). Motion analysis of arm movement during badminton smash. *2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)* 30 November - 02 December. Kuala Lumpur, Malaysia, IEEE, pp. 111–114.
DOI: 10.1109/IECBES.2010.5742210.

Sharma M., Anand A., Srivastava R. & Kaligounder L. (2018). Wearable Audio and IMU based shot detection in racquet sports. *arXiv preprint arXiv:1805.05456.*
DOI: https://doi.org/10.48550/arXiv.1805.05456

Tan D.Y.W., Ting H.Y. & Lau S.B.Y. (2016). A review on badminton motion analysis. *2016 International Conference on Robotics, Automation and Sciences (ICORAS)* 5-6 November. Melaka, Malaysia, IEEE, pp. 1–4.
DOI: 10.1109/ICORAS.2016.7872604

Torres-Luque G., Cabello-Manrique D., Hernández-Garcia R. & Garatachea N. (2011). An analysis of competition in young tennis players. *European Journal of Sports Science* **11**(1): 39–43.
DOI: https://doi.org/10.1080/17461391003770533

Tsai C., Chang S. & Huang C. (1998). Biomechanical analysis of differences in the badminton smash and jump smash between Taiwan elite and collegiate players. *XVI International Symposium on Biomechanics in Sports* (eds. Riehle H. & Vieten M.M.), University of Konstanz, Germany. UVK-Universitätsverlag Konstanz, Konstanz, Germany, pp., 259–262.

Vora V., Arora M. and Ranawat D. (2018). Biomechanics in badminton - A review. *Orthopedics and Sports Medicine: Open Access Journal* **2**(1): 114–116.

Wadenka T. (2019). Smash and clear tutorial: Get more power

in your overheadshots. Available at *https://www.youtube.com/watch?v=HvAOMnoT3zQ*, Accessed 4 January 2023.

Wang C., Bochkovskiy A. & Liao H.M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, 17-24 June. IEEE, pp. 7464–7475.

Williams B. K., Sanders R. H., Ryu J. H., Graham-Smith P. & Sinclair P. J. (2021). Racket orientation angle differences between accurate and inaccurate squash shots, as determined by a racket embedded magnetic-inertial measurement unit. *Sports Biomechanics* 1–13.
DOI: https://doi.org/10.1080/14763141.2021.1892175

Ying Y.Z., Khairuddin I.M., Isa W.H.M., Majeed A.P.P.A., Abdullah M.A. & Razman M.A.M. (2022). Badminton smashing recognition through video performance by using deep learning. *Mekatronika* **4**(1): 70–79.
DOI: https://doi.org/10.15282/mekatronika.v4i1.8607

Zeng H. & Zhao Y. (2011). Sensing movement: microsensors for body motion measurement. *Sensors* **11**(1): 638–660.
DOI: https://doi.org/10.3390/s110100638

Zhang Z., Li S., Wan B., Visentin P., Jiang Q., Dyck M., Li H. & Shan G. (2016). The influence of X-factor (trunk rotation) and experience on the quality of the badminton forehand smash. *Journal of Human Kinetics* **53**(1): 9–22.
DOI: 10.1515/hukin-2016-0006

RESEARCH ARTICLE

Deep Learning

Evaluation of Taekwondo Poomsae movements using skeleton points [†]

M Fernando[*], KD Sandaruwan and AMKB Athapaththu
University of Colombo School of Computing, UCSC Building Complex, 35 Reid Ave, Colombo 00700, Sri Lanka.

Abstract: Taekwondo is a widely practised martial art and an Olympic sport. In Taekwondo, Poomsae movements are essential, as they form the foundation of the sport and are fundamental for success in competitions. The evaluation of Poomsae movements in Taekwondo has been a subjective process, relying heavily on human judgments. This study addresses the above issue by developing a systematic approach to evaluate Poomsae movements using computer vision. A long short-term memory-based (LSTM-based) machine learning (ML) model was developed and evaluated for its effectiveness in Poomsae movement evaluation. The study also aimed to develop this model as an assistant for self-evaluation, that enables Taekwondo players to enhance their skills at their own pace. For this study, a dataset was created specially by recording Poomsae movements of Taekwondo players from the University of Colombo. The technical infrastructure used to capture skeleton point data was cost-effective and easily replicable in other settings. Small video clips containing Taekwondo movements were recorded using a mobile phone camera and the skeleton point data was extracted using the MediaPipe Python library. The model was able to achieve 61% of accuracy when compared with the domain experts' results. Overall, the study successfully achieved its objectives of defining a self-paced approach to evaluate Poomsae while overcoming human subjectivity otherwise unavoidable in manual evaluation processes. The feedback of domain experts was also considered to finetune the model for better performance.

Keywords: Long short-term memory (LSTM), machine learning (ML), skeleton points, taekwondo, video classification.

INTRODUCTION

The integration of technology to enhance athletes' performance has become a prominent research area. However, movement evaluation in sports remains challenging due to limited technological involvement and the high costs associated with accessing required high-end technological infrastructure. Computer vision has also emerged as a promising technology also for evaluating movements in sports training.

Taekwondo is a globally popular martial art which has also been recognized as an Olympic sport. In Taekwondo, Poomsae movements are essential, as they form the foundation of the sport and are fundamental for success in competitions. Therefore, there is a growing demand for effective and objective evaluation methods to assess the quality and accuracy of players' Poomsae movement. Despite its popularity, less attention has been given to using technologies such as computer vision to assess Taekwondo movements.

There is no systematic way to track players' performance and traditional evaluation methods for Taekwondo movements rely heavily on subjective visual observation, which can lead to errors and biasedness in judgments. Accurate execution of these movements is
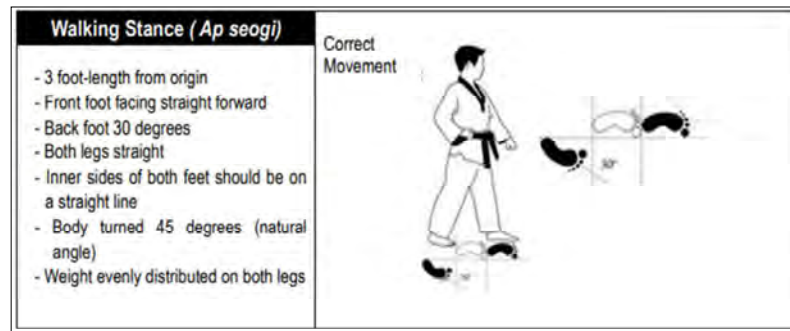
[*] Corresponding author (michelleufernando@gmail.com; https://orcid.org/0009-0000-4911-6272)

also crucial for success in competitions. Figure 1 is an illustration of some movement guidelines given by the

World Taekwondo Federation (WTF, 2014) for Poomsae movements.

Figure 1: WTF guidelines for walking stance in Poomsae (WTF, 2014)

Taekwondo training typically involves multiple sessions per week, each lasting for several hours. With each session coach dealing with more than ten students at a time, it can be challenging to provide individualized attention to each player. This can result in less time devoted to mastering specific movements, and slower the overall progress in skill development. Therefore, self-training is a must to sharpen the skills, and unavailability of a self-paced movement evaluation method is a serious problem in Taekwondo.

Contribution of the paper

In response to the above challenges, this paper presents a method that involves the skeleton data points of players to assess the quality and accuracy of the movements, using supervised deep learning (DL). It utilizes variations in skeleton points data to evaluate movements and distinguish correct from incorrect movements. This approach leverages the power of machine learning (ML) algorithms to accurately identify and classify the quality of movements based on the observed patterns in the skeleton data.

Organization of the paper

The remaining sections of the paper are organized as follows: the next sections delve into existing literature, providing a comprehensive overview of prior research in the field. Following this, the Methodology section details the approach to handling data, outlining the design and development of the classification model employing LSTM networks. Subsequently, the Results and Discussion section presents experimental findings,

including model evaluation metrics, and discusses results in conjunction with domain experts' assessments. Finally, the Conclusion summarizes key findings of the study, referring insights presented throughout the paper.

Player performance evaluation

When it comes to player performance evaluation using computerized systems, Human action/activity recognition (HAR) plays a major role. HAR can be referred to as the art of identifying and naming activities from a video containing complete action execution (Kong & Fu, 2022), using artificial intelligence (AI) from the gathered activity raw data, by utilizing various kinds of hardware devices such as wearable sensors, electronic device sensors, camera devices (Kinect, Orbbec, CCTV), and also some commercial off-the-shelf (COTS) equipment like a webcam with an open-source software solution to detect movements.

HAR can be done using various data modalities like Red-Green-Blue (RGB) depth images, skeleton points, infrared, radar, and WiFi signals. Skeleton point detection is more cost-effective compared to other data modalities (Sun *et al.*, 2022). Insensitivity to the background, use of COTS hardware devices, and the availability of open-source software are some of the advantages.

There are many different approaches to HAR, including rule-based dynamic models and machine-learning methods (Vrigkas *et al.*, 2015). Each approach has its strengths and weaknesses, and the choice of method depends on the specific application. In recent years, deep learning methods such as convolutional neural networks

(CNNs) and Long-Short Term Memory (LSTM) networks have shown great promise in HAR tasks, achieving state-of-the-art results on many benchmark datasets (Sun *et al*., 2022). HAR has been widely researched in the computer science field, with applications ranging from surveillance and robotics to healthcare and sports. In sports, it has been used to analyze and evaluate various athletic performances, such as gymnastics, dance, and martial arts (Emad *et al*., 2020; Host & Ivašić-Kos, 2022).

Use of skeleton joint tracking technology in HAR

In the field of HAR, the use of skeleton joint tracking technology has gained significant attention in recent years. This technology involves tracking the movement of human joints using various sensors and cameras. The resulting data can be used to analyse and understand the biomechanics of human motion, which can be applied in fields such as sports, healthcare, and rehabilitation as described by Vrigkas *et al*., (2015). It offers non-intrusive, cost-effective, and convenient solutions for evaluating body posture and movement. The use of skeleton point data obtained from pose estimation algorithms and motion capture systems has the potential to revolutionize action recognition and motion analysis.

A paper by Chung *et al*., (2022) provides a comprehensive overview and analysis of various state-of-the-art techniques for skeleton-based human pose estimation algorithms. The authors review several algorithms including OpenPose, PoseNet, MoveNet, and MediaPipe-Pose that use different methods, such as DL, graphical models, and optimization techniques, to estimate the human body's 2D or 3D pose from images or videos.

The techniques are evaluated using several benchmark datasets, and their strengths and limitations are discussed in detail. The paper also presents a comparative analysis of these techniques, including their accuracy, computational complexity, and suitability for real-time applications. The authors highlight the importance of considering factors such as pose variability, occlusion, and image resolution when selecting a technique for a specific application. The authors indicate that MediaPipe has performed well in video data.

ML models for player performance evaluation

A paper by Zhang *et al*. (2012) proposes a system to recognize and segment the postures of golf swings consisting of two main components: a Gaussian mixture model (GMM) and a support vector machine (SVM).

The GMM was used to detect the golfer's body parts. The detected body parts were then used to calculate the angles and positions of the golfer's body during the swing.

The SVM was trained on the extracted features to classify the golf swing into one of four categories: perfect, good, average, or poor. The experimental results showed that the proposed system achieved an accuracy of 85% in classifying the golf swing into one of the four categories. It can provide a quantitative evaluation of golf swings and identify areas for improvement. One of the weaknesses of this system is its use of the Kinect sensor, which is relatively expensive for general-purpose usage.

In a paper by Piergiovanni & Ryoo (2018), the authors proposed a model for recognizing fine-grained actions in baseball videos using a combination of convolutional neural networks (CNNs) and long short-term memory (LSTM) networks. The authors demonstrated that their model outperformed several baseline models. However, further research is needed to explore the generalizability of the model to other types of video data. Specifically, the model consists of two main parts: a feature extractor and a classifier.

The classifier is an LSTM network that takes as input the sequence of feature vectors, and outputs a probability distribution over the fine-grained action labels. To train the model, the authors used a large dataset of baseball videos that had been annotated with fine-grained action labels. They trained the model using a cross-entropy loss function and the Adam optimizer. However, it has not used skeleton points for the classification.

In a research paper, Zhao *et al*., (2019) propose a DL model for recognizing human actions using skeleton point data. The paper addresses the challenge of recognizing human actions from skeleton point data by exploiting Bayesian Graph convolutional long short-term memory (GC-LSTM) networks. The proposed model consists of two main components: a graph convolutional layer and a Bayesian GC-LSTM layer that applies graph convolutions on the skeleton data to capture the spatial relationships between the joints. The Bayesian GC-LSTM layer combines the temporal dependencies of the skeleton data with uncertainty modelling using Bayesian inference. The authors evaluated the proposed model on three publicly available datasets: the NTU RGB+D dataset, the Kinetics-Skeleton dataset, and the SBU Interaction dataset. The experimental results showed that the proposed model performed well on all three datasets.

Liu *et al.*, (2017) proposed a skeleton-based human action recognition approach utilizing global context-aware attention LSTM networks. The method addresses the challenge of capturing spatial and temporal information from skeleton data by incorporating a context-aware attention mechanism into LSTM networks.

This enables the network to focus on relevant joints while considering the overall context of the action. The approach achieves improved accuracy in action recognition, as demonstrated through experiments on benchmark datasets. However, limitations include the reliance on accurate skeleton joint tracking and the potential sensitivity to noise or missing joint information, which may impact performance in real-world scenarios. Further research is needed to address these limitations and enhance the method's applicability.

According to the literature review based on HAR for martial arts performance evaluation, there are some existing approaches for classifying taekwondo and other martial arts movements using videos.

The Taekwondo unit technique human action dataset with key frame-based CNN action recognition (TUHAD) paper (Lee & Jung, 2020) defined a reliable Taekwondo Poomsae movement dataset called TUHAD and proposed a key-frame-based Convolutional Neural Networks (CNN) architecture to recognize Taekwondo actions using their dataset. TUHAD has 1936 action samples of eight unique actions performed by ten individuals and recorded from front and side views. The suggested model achieved recognition accuracy of up to 95.83%, according to a correlation analysis of the input configuration and precision.

A paper by Barbosa *et al.*, (2021) compared 4 different deep-learning approaches to classify Taekwondo movements. To determine which methodology or technique yields the greatest results, it was tested using a dataset that had already been created. It was determined that convolution layer models, such as CNN plus long short-term memory (LSTM) and convolutional long short-term memory (ConvLSTM) DL models, give more than 90% accuracy. It produced the best results given the movements, which were typically fast and frequently high-leg movements.

The paper by Liang & Zuo (2022) proved that movement classifications from videos and mapping them to score a Taekwondo sparring (fighting) match are also possible. They introduced a graph convolution framework to recognize, segment, and evaluate Taekwondo actions with a specific part of the perception structure. The identified Taekwondo movements are marked in a time series using LSTM, and feature extraction is done at the graph convolution level to acquire the spatial and temporal correlations between skeleton joints. Using the manually labelled database, it predicted the action class and then matched the score for each movement. Finally, it is validated using Taekwondo competition data videos. This approach has an average action identification accuracy of 90% and an average action score matching rate of 74.6%.

Emad *et al.* (2020) proposed a smart coaching system called iKarate for Karate training. It offers a system that will track the players' movements using an IR camera sensor and classify the data (after a pre-processing phase) using the fast dynamic time warping algorithm. As an output, it will generate an accurate report that includes every action the player performed, mistakes made in each movement, and suggestions to improve movements.

A paper by Cunha *et al.* (2021) proposed a user-friendly and affordable approach for quickly evaluating the performance of Taekwondo competitors. They have created a mobile application to monitor the movements made by the player throughout a training session. Sometimes, occlusion occurs due to the rotation of the player or the overlapping of a limb. To overcome this issue, they have used motion sensors and an inertial measurement unit for data collection. Then it will be transmitted to the mobile app through Wi-Fi. Hardware components will be fixed to the players through Velcro strips. They used a special lab environment to conduct the study with the relevant hardware devices.

Identified limitations of existing systems

Upon careful analysis of the literature review, it is discerned that the following limitations exist within current systems: using relatively expensive hardware devices (sensors, IR trackers, Kinect camera), lack of skeleton point data usage for movement evaluation, classifying only one movement without any way to detect a sequence of actions, and the need to set up a special lab environment. Currently there is no systematic way to evaluate a sequence of Poomsae movements.

MATERIALS AND METHODS

The study followed the design science research methodology (vom Brocke *et al.*, 2020). Design science methodology is a systematic approach that aims to develop and evaluate innovative solutions to practical

problems. The process involves several stages starting with problem identification, followed by the defined objectives, design and development, demonstration, evaluation, and communication of a solution. Figure 2 illustrates the alignment of each step in the research with the design science approach adopted throughout the study. The methodology emphasizes the importance of iteration and feedback in the design and development process, as well as the need for iterative and rigorous evaluation of the solution's effectiveness.
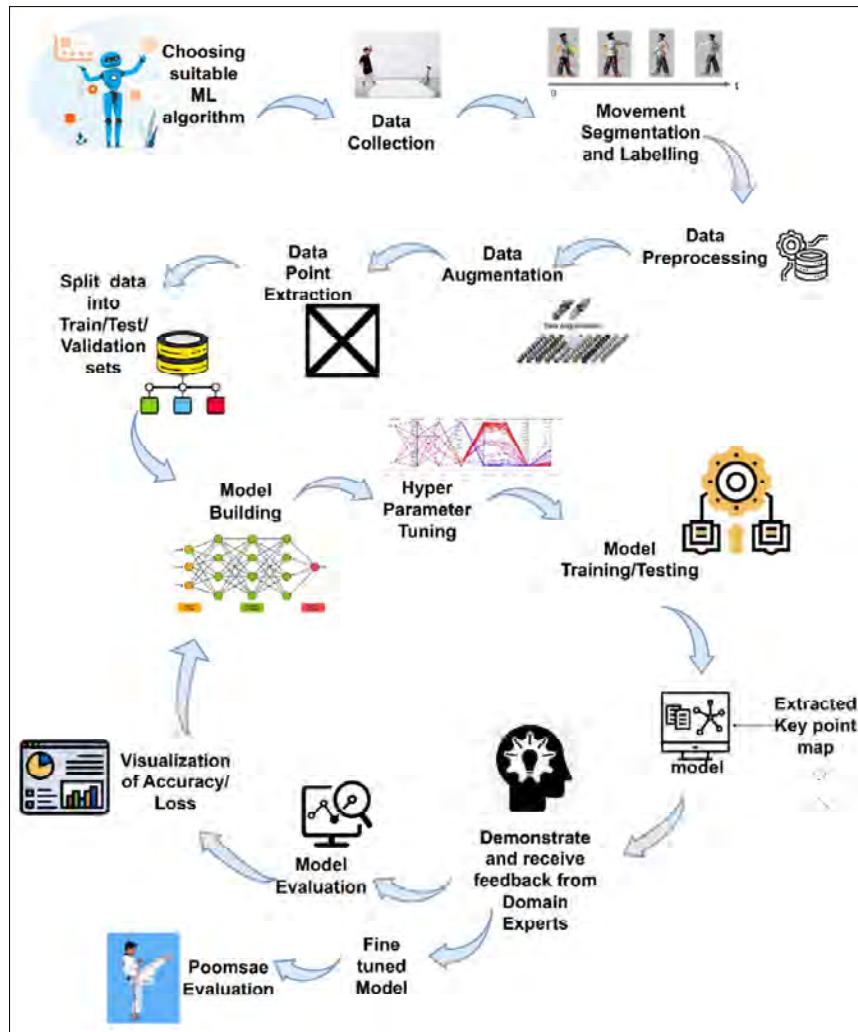

Figure 2: Designing and evaluation of the solution

Choosing suitable ML/DL algorithm

The problem of Taekwondo movement classification using skeleton point data can be mapped to a classification problem, which is a widely researched area in human action recognition. However, due to the sequential and temporal nature of the movement data, simple ML models are not suitable for accurately classifying the movements. Therefore, there is a need for more complex models such as artificial neural networks (ANNs) to handle this video classification task.

Convolutional neural networks (CNNs) are commonly used for image recognition tasks, but they are not well-suited for temporal data analysis, because the model has to store the previous state of the frame to process the current state of the frame. Recurrent neural networks (RNN) are designed to handle sequential

data by processing the data sequentially, and they can remember previous inputs using hidden states. However, traditional RNNs may suffer from the problem of vanishing gradients during training, which can affect their performance in video kind of data classification. (Olah, 2015; Mittal, 2019)

Long short-term memory (LSTM) is a type of RNN that addresses the problem of vanishing gradients by introducing memory cells and gates. LSTM has been shown to be effective in dealing with dynamic temporal datasets, making it a suitable model for movement classification problems using skeleton point data (Barbosa *et al*., 2021). Therefore, in this research, LSTM was selected as the best classification technique for accurately classifying Taekwondo movements based on skeleton point data.

Data collection

A collection of recorded videos was obtained from players of the University of Colombo (UOC) Taekwondo team. Before recording videos of each player, the research objectives were explained to the players, and their consent obtained to participate in the study. Fifteen members of the UOC Taekwondo team volunteered, comprising seven senior players and eight junior players. A senior player means he or she should have membership in the Sri Lanka Taekwondo Federation (SLTF), should have a blue or above belt (till 6th GUP), have experience in participating in Poomsae competitions for more than two years, and have won a medal from at least one Poomsae event. A Junior player means he or she should have membership in SLTF, should have a junior green or above belt (8th GUP or below), have experience in participating in one poomsae competition, but may or may not have won a medal from a Poomsae event. Therefore, it was confirmed that all participants were aware of performing a Poomsae.

Both correctly and incorrectly performed movements were captured, providing a comprehensive dataset for analysis. Initially, over 750 videos were collected, encompassing a wide range of skill levels. A 1 m tripod stand and an iPhone XS max mobile phone camera were used for recording videos. (Camera specifications: iPhone XS max, 12-megapixel (f/1.8, 1.4-micron),1920 * 1080 resolution, 30 fps).

Movement segmentation and labelling

The whole performance of the series of Poomsae movements by a player was segmented into short video clips that contained only one movement. The segmentation part was done manually, and, in this phase, each video clip was labelled as to whether it was corrected or not with the help of domain experts (okankop, 2020). The 'incorrect' labelled data pertains to the commonly occurring mistakes made by Poomsae players. Collecting data from different players with different skill levels helped to increase the diversity of the dataset and made the model more robust and generalizable.

Figure 3 shows a player performing 4th movement. In the left side image, the stance is incorrect, and the punch is correct. Therefore, the movement is labelled as i4 (incorrect movement 4). Figure 3 right side image shows a player performing the same movement. The stance is correct, and the punch is correct. Therefore, the movement is labelled as c4 (correct movement 4).
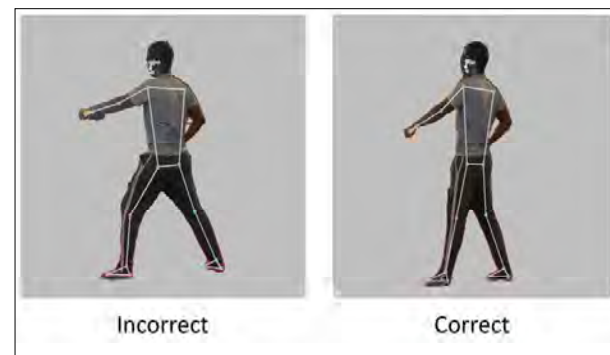
Figure 3: Player performing the fourth movement in two different ways

Data pre-processing

Segmented videos contained different frame numbers. However, the LSTM model requires the same sample size when it comes to the model training process. Therefore, bringing all the video clips into the same frame rate was necessary.

The approach involved using a video processing library to standardize the frame rate and duration of the video clip. Specifically, the frame rate was adjusted to 30 frames per second (fps) with a consistent one-second duration per video clip. This process ensured temporal uniformity, facilitating precise and systematic analysis of the video content. The method was chosen to provide a reliable foundation for research and evaluation. A simple Python script written with the 'moviepy' library was used.

Data augmentation

The data augmentation technique was utilized to increase the size of the dataset and improve the accuracy of the model as described by Park & Sohn (2020). To achieve this, salt and pepper noise and colour inverter techniques were applied. These techniques helped to diversify the dataset by adding additional variations to the original images (Nida *et al*., 2022). The resulting augmented dataset was then mixed with the original dataset and increased to 1500+ video clips. That dataset was used for training the DL model. The 'Vidaug' (okankop, 2020) Python library was used for the augmentation process.

Skeleton point data extraction

Skeleton point data was extracted using the MediaPipe Pose (Google, n.d.) solution. It is a Python library that provides highly accurate landmark (skeleton joint) detection. The extracted coordinates were stored as binary files in a well-defined folder structure, considering read/write efficiency. Figure 3 shows how the MediaPipe Pose detection library identifies human skeleton points.

The dataset contained 8 different action classes with 188 videos for each class, resulting in a total of 1504 videos. Each video has a one-second duration and consists of 30 frames. When extracting key points, one frame had 132 coordinates, which corresponded to 33 joints and 4 coordinates (x and y coordinates of the point in the image frame, the depth or z coordinate, and a visibility value that indicates whether the landmark is visible in the image or not) for each joint. Therefore, each video clip contained 132 * 30 key points. Therefore, the total number of key points in the dataset can be calculated as:

1504 videos * 30 frames per video * 33 joints per frame * 4 coordinates per joint = 5,955,840 key points.

The collected data was divided into two sets of datasets: training and testing. Random seed was set to 42, to ensure that the data would be split into the same training and testing sets each time. The number 42 is often used as a random seed in ML because it is a reference to the book "The Hitchhiker's Guide to the Galaxy," in which 42 is famously referred to as the "Answer to the Ultimate Question of Life, the Universe, and Everything."

Hyper parameter tuning

Hyperparameter tuning was done for the Long Short-Term Memory (LSTM) network by adjusting the number of layers, the number of neurons per layer, dropout rate, activation function, and learning rate for different model architectures and checking the validation accuracy increment in several attempts.

Selected hyperparameters for the tuning process are listed below.

- Number of layers - deep neural networks (DNN) should use a minimum number of Layers to keep the generalized behaviour as mentioned in the papers by Hobs (2015) and Eckhardt (2018).
- Number of neurons per layer – The number of neurons in the input layer is the size of the input data (features), and the number of neurons in the output layer should be equal to the number of classes, The number of neurons in the hidden layers is determined by the complexity of the problem as discussed by Heaton (2008).
- Activation function - Tanh (hyperbolic tangent) and Relu (rectified linear unit) are widely used for classification tasks when using neural networks. Tanh maps the input to the range of -1 to 1 while Relu maps the input to the range of 0 to infinity (Lee & Jung, 2020).
- Dropout rate – Adding dropout layers prevents overfitting. Typically, the dropout rate value between 0.1 and 0.5 performs well in video data classification (Brownlee, 2017).
- Learning rate - controls the step size at each iteration and smoothens the convergence (Eckhardt, 2018).
- Optimizer - Adam optimizer adjusts the learning rate adaptively during training, allowing it to converge faster than other optimization algorithms.

Model building

Initially, the model was designed with an LSTM input layer and a dense output layer. The number of hidden layers was decided with the hyperparameter tuning as described in the previous section. Two hidden LSTM layers and two hidden dense layers were used. Each layer contains a dropout layer to avoid overfitting. The final model was designed with the hyperparameters chosen in the above section. Figure 4 indicates how the hyperparameter tuning process was carried out using a parallel coordinate graph.

Table 1 and Figure 5 visualize the deep learning model architecture. The input layer of the neural network comprises 132 neurons, which correspond to the 132 coordinates utilized as features for the classification task. Notably, these coordinates serve as the fundamental input attributes for the network's classification model. It

is important to highlight that the sample size is precisely 30. This sample size is representative of a single video clip, each of which contains 30 frames. Each frame, in turn, encapsulates 132 distinct features. The utilization of this input structure is pivotal in enabling the network to process and classify video data effectively.
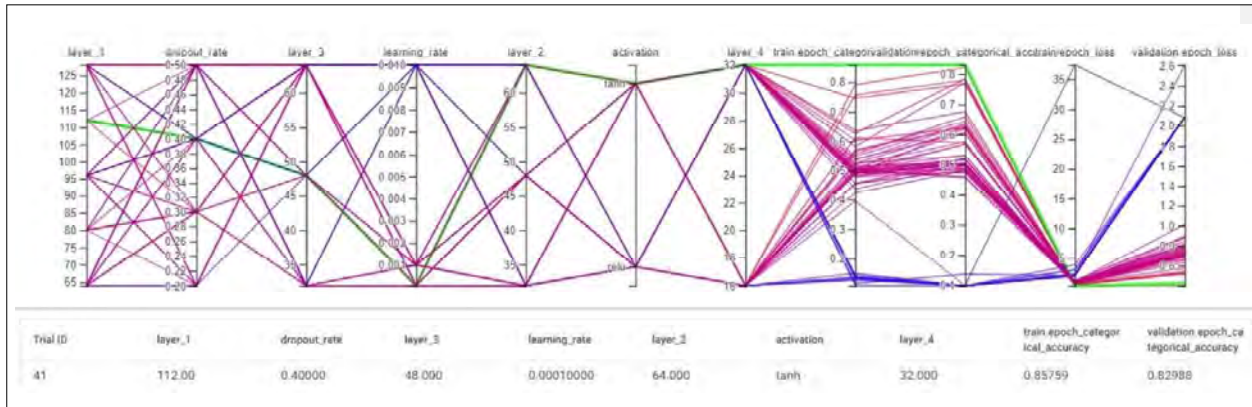
Figure 4: Hyperparameter tuning visualized in a parallel coordinate view

Table 1: Model architecture

Layers	Type & no. of neurons	Hyper parameters
Input	LSTM Layer with 132 Neurons	Input size = 33 landmarks * 4 coordinates = 132 Sample size = 30 frames
Hidden	LSTM layer with 112 Neurons Dense layer with 64 Neurons LSTM layer with 48 Neurons Dense layer with 32 Neurons	dropout rate = 0.4 activation function = tanh
Output	Dense Layer with 8 neutrons	SoftMax activation function (8 movement categories)

Figure 5: Deep learning architecture diagram

The output layer of the neural network was configured as a dense layer, employing the SoftMax activation function. This choice of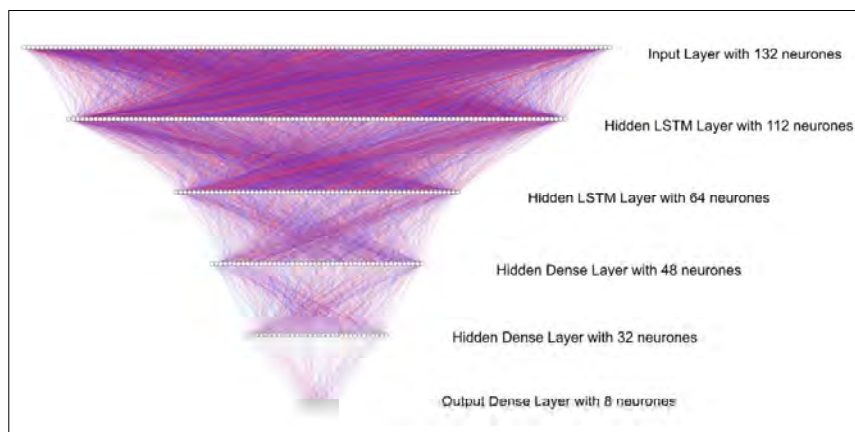 activation function was deliberate, as it offers the capacity to provide action class probabilities in the context of multi-class classification. In this movement classification task, there are a total of 8 distinct movements, and accordingly, the output layer was composed of 8 neurons. Each neuron in the output layer is responsible for generating a probability estimate for its corresponding class, thereby facilitating the classification of input data into one of the 8 defined movement classes.

The input data for the model consisted of preprocessed skeleton point data, as described in the data preprocessing section. Therefore, the first input layer has 132 neurons and has a [30, 132] input data vector. In the end, it was expected to have a multi-class classification of 8 classes, therefore the output dense layer had 8 neurons.

Model training / testing

Model training involves feeding the preprocessed and extracted skeleton point data into the ML model. The models are then trained to recognize patterns in the data. In this case, it could be defined as identifying the relevant pose within 30 frames of skeleton point data. Finally, the models are tested using a separate dataset to evaluate their accuracy and generalization ability.

The model was trained using a categorical cross-entropy loss function since this is a multi-class classification model and optimized using the Adam optimizer due to its ability towards fast convergence and adaptive learning rate. The training result is described in Table 2.

Table 2: Model training summary after 333 epochs

Result description	Score
Categorical accuracy for the training dataset	0.995
Loss for the training dataset	0.03722
Categorical accuracy for the testing dataset	0.96013
Loss for the testing dataset	0.1730

The model training and testing process was conducted using a local machine that solely utilized a CPU for training. The average duration for training the LSTM model with a minimum of four layers, using the collected dataset, was 30 to 40 minutes. However, hyperparameter tuning required more time, as it involved evaluating various potential combinations to improve accuracy. This task took an average of five hours to complete 60 trials and output a good hyperparameter combination.

Classification algorithm

In the classification algorithm, the initial step is to break down a sequence of movements into individual movements for evaluation. This means the algorithm gets an initial 30 frames as a movement and extracts key points for each frame. Then all the coordinates with respective frame numbers will be fed into the LSTM classification model. Then the model will perform the classification task. If the input coordinates lack the usual characteristics needed for classification, the algorithm follows a specific protocol. It dequeues the initial frame that was initially selected as the input frame sequence and enqueues the subsequent frame from the video file. Then again, the newly selected input frame sequence is fed to the LSTM classification model. This process is repeated for the entire video.

RESULTS AND DISCUSSION

The final movement evaluation algorithm was designed in a way that users could upload a video file of a player's Poomsae performance. The output would be the classified movement sequence, indicating whether each movement is correct or not. As shown in Figure 6 it detects the first movement as correct (c1) and in Figure 7 it detects the second movement as incorrect (i2) for the given video. The classified action sequence result is displayed in the top left corner of the window.
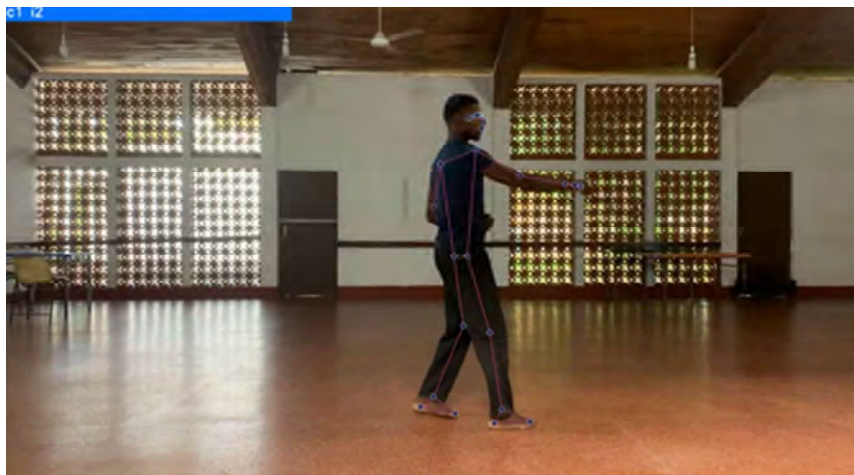
Model evaluation

The evaluation of the ML model can be done in two different ways. The first approach involves the use of standard classification metrics to measure the accuracy and performance of the model on a test data set. The second approach is to evaluate the model's performance with the help of a domain expert to ensure that it is addressing the identified research problem.

During the training process, the model was optimized using various hyperparameters to ensure that it was accurately detecting correct and incorrect patterns. The model achieved an average accuracy of 96% for the test data set. After training, the model was applied to a new validation data set of 11 videos and the results compared with a domain expert. The comparison resulted in the ML model classifying the correct movements with 66% accuracy as shown in the Table 4.

Figure 6:   Screen capture obtained from the ML model output video - first movement evaluation.

Figure 7:   Screen capture obtained from the ML model output video - second movement evaluation.
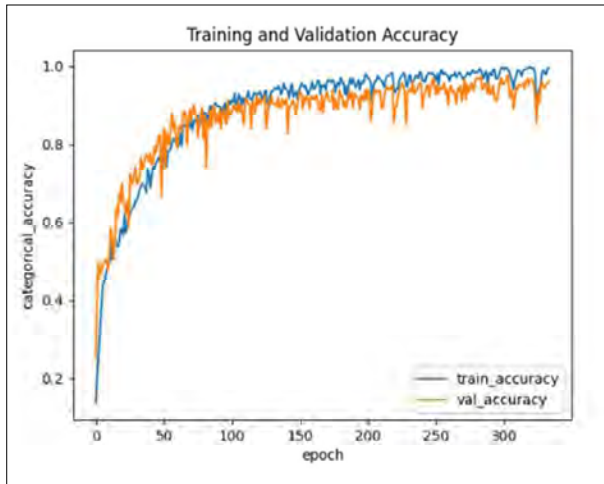
Classification metrics

Accuracy is defined as the fraction of all correct predictions, which includes true positives and true negatives, out of all the predictions made by the model for a certain class of movements. However, it is important to note that high accuracy alone does not necessarily imply good model performance, as it may not reflect a class imbalance in the data. To account for this, additional evaluation measures such as precision, recall, and the F1 score are commonly used.

These measures provide insights into the model's performance concerning correctly identifying the positive class (precision) and capturing all positive instances in the data (recall), and the harmonic mean of these two measures (F1 score) provides a more balanced view of the model's performance.

The utilization of a neural network architecture with four hidden layers resulted in the highest achieved accuracy of 96%. This configuration also demonstrated a consistent and smooth convergence in the train/test accuracy, as visually depicted in Figure 8. Whereas, Figure 9 illustrates a progressive reduction in both training and testing loss over the course of more than 300 epochs.

Figure 8: Training and testing (validation) accuracy for the LSTM Model
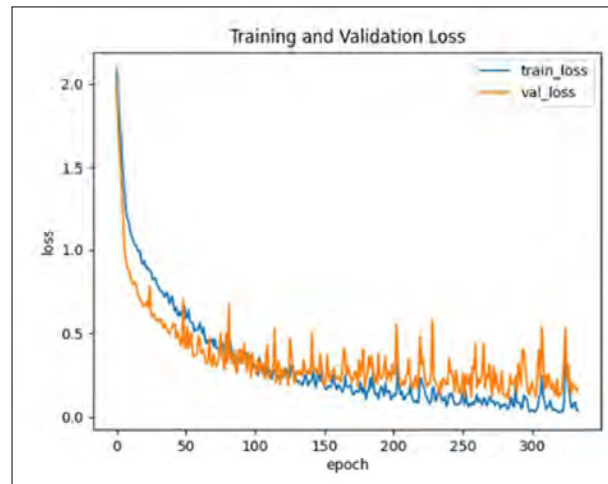
Figure 9: Training and testing (validation) Loss for the LSTM Model

The test dataset comprises a total of 301 video clips. The classification report depicted in Table 3 provides an overview of the performance metrics of the trained model on the test dataset. The report includes accuracy, precision, recall, and f1-score measures, which are essential for evaluating the effectiveness of the model.

Table 3: Classification report containing precision, recall, and f1-score for each movement class

	Precision	Recall	F1-score	Support
c1	0.83	0.94	0.88	32
c2	1.00	0.95	0.98	43
c3	0.97	1.00	0.99	37
c4	0.97	1.00	0.99	37
i1	0.94	0.85	0.89	39
i2	0.95	1.00	0.98	40
i3	1.00	0.98	0.99	42
i4	1.00	0.97	0.98	31
Accuracy			0.96	301
Macro Average	0.96	0.96	0.96	301
Weighted Average	0.96	0.96	0.96	301

Model evaluation with the domain experts

Table 4 simply presents the evaluation results of a Poomsae movement by the domain expert and the ML model. The experiments were done using 11 validation video files from several Poomsae players. Domain experts' results are considered the 'ground truth' when evaluating the model performance.

Table 4: ML model evaluation results with the domain experts

Video No.	Domain expert's * result sequence				ML model classification sequence				No. of correctly detected movements by the ML model	Detection percentage
1	c1	c2	i3	c4	i1	i2	c3	c4	1	25%
2	i1	i2	c3	i4	i1	i2	i4	i4	3	75%
3	i1	c2	i3	i4	i1	c2	c3	c4	2	50%
4	c1	c2	c3	c4	c1	c2	c3	c4	4	100%
5	i1	c2	c3	i4	i1	c2	c3	c4	3	75%
6	c1	c2	i3	i4	c1	i2	i3	i4	3	75%
7	i1	i2	c3	i4	i1	i2	i4	i4	3	75%
8	c1	i2	c3	i4	i1	c2	c3	c4	1	25%
9	i1	c2	i3	c4	i1	i2	c3	i4	1	25%
10	i1	c2	c3	c4	i1	c2	c3	c4	4	100%
11	i1	i2	i3	i4	i1	i2	i3	i4	4	100%
					Average Accuracy					66%

* Domain expert: an experienced individual in Poomsae evaluation, and their assessments (holding a Black Belt 6th Dan (Kukkiwon) and also an International Referee of the World Taekwondo Federation.) participated in this study.

The letter 'c' indicates that the movement is correct, and 'i' indicates that the movement is incorrect. The numbers 1 to 4 indicate what the movement was that the player performed. The last two columns of the above table list how many of the movements were correctly evaluated by the ML model and then the classification percentage for a given video by the ML model. Finally, the accuracy of the model for the validation data set was presented as a percentage. The ML model achieved 66% of the average accuracy score for the given dataset of 11 videos.

CONCLUSION

In conclusion, this study aimed to evaluate Taekwondo movements using an LSTM-based ML model in terms of its classification performance. In the evaluation of the LSTM model on the test dataset, it achieved a noteworthy accuracy of 96%. When benchmarked against domain experts, the model maintained an average accuracy of 61%. This comparative performance highlights the model's robustness and potential applicability in real-world scenarios. The substantial accuracy suggests that the proposed classification model could serve as a valuable tool for assessing player performance.

The model was able to classify the performance of the movement by observing a set of frames and considering the behaviour of the movement, rather than solely focusing on the terminal pose. It was noticed that this aspect of the ML model could be useful in Poomsae movement evaluation, as it can capture not only the accuracy of the movement but also the strength and expression of energy. While this study focused only on the accuracy of the Poomsae movements, future work could consider these additional aspects in the evaluation process. Furthermore, the model was tested on various video data to ensure its generalizability to real-world scenarios. The feedback of domain experts was also considered to fine-tune the models for better performance.

A major research problem addressed in this study was the issue of subjectivity in the domain of Poomsae evaluation using traditional methods. The ML model performed well in the context without human intervention, indicating that improving these models could lead to a systematic approach to evaluating Poomsae by eliminating human subjectivity. By providing a reliable and objective evaluation tool, this research project contributes significantly to the advancement of the Poomsae evaluation domain.

The lack of self-evaluating methods was identified as another problem that this study aimed to address. The research project successfully developed a model that could be easily integrated with mobile apps or web apps as a self-paced approach for Poomsae evaluation, utilizing commercial off-the-shelf hardware devices and open-source software solutions.

Additionally, the model was insensitive to background distractions, ensuring that the focus remains on the subject of interest. Skeleton point data optimized train/test data storage by requiring less space compared to traditional video file preservation methods, reducing storage costs and resource requirements.

Looking ahead, several exciting avenues exist for extending and enhancing the impact of this research. One promising direction involves the implementation of models tailored for real-time performance evaluation, potentially revolutionizing how movements are assessed as they occur. Additionally, a diverse range of datasets will be employed to further test and refine the models, ensuring applicability and robustness across varied contexts. Furthermore, this research can be extended for a deeper exploration into the capabilities of the ML model for evaluating not only basic movement accuracy but also other criteria such as the expression of energy and presentation skills. These efforts not only aim to improve current research, but also lead to new uses and understandings in movement analysis.

Acknowledgments

REFERENCES

Barbosa P., Cunha P., Carvalho V. & Soares F. (2021). Classification of taekwondo techniques using deep learning methods: First insights. *Proceedings of the BIODEVICES 2021 - 14th International Conference on Biomedical Electronics and Devices;* Part of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2021, 11–13 January, Vienna, Austria, pp. 201–208.

Brownlee J. (2017). Dropout with LSTM networks for time series forecasting. In: *Machine Learning Mastery*. Available at *https://machinelearningmastery.com/use-dropout-lstm-networks-time-series-forecasting/*

Chung J.-L., Ong L.-Y. & Leow M.-C. (2022). Comparative analysis of skeleton-based human pose estimation. *Future Internet* **14**(12): 380.
DOI: https://doi.org/https://doi.org/10.3390/fi14120380

Cunha P., Barbosa P., Ferreira F., Fitas C., Carvalho V. & Soares F. (2021). Real-time evaluation system for top taekwondo athletes: project overview. *Proceedings of the BIODEVICES 2021 - 14th International Conference*

on Biomedical Electronics and Devices; Part of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2021, 11–13 January, Vienna, Austria, pp. 209–220.
DOI: https://doi.org/10.5220/0010414202090216

Eckhardt K. (2018). *Choosing the right Hyperparameters for a simple LSTM using Keras*. Available at *https://towardsdatascience.com/choosing-the-right-hyperparameters-for-a-simple-lstm-using-keras-f8e9ed76f046*

Emad B., Atef O., Shams Y., El-Kerdany A., Shorim N., Nabil A. & Atia A. (2020). Ikarate: Karate Kata guidance system. *Procedia Computer Science* **175**(2019): 149–156.
DOI: https://doi.org/10.1016/j.procs.2020.07.024

Google for Developers. (2023). *Pose landmark detection guide*. Available at *https://developers.google.com/mediapipe/solutions/vision/pose_landmarker*

Heaton J. (2008). *Introduction to Neural Networks with Java*. Heaton Research, Cop.

Hobs. (2015). How to choose the number of hidden layers and nodes in a feedforward neural network? In *Stackexchange*. Available at *https://stats.stackexchange.com/questions/181/how-to-choose-the-number-of-hidden-layers-and-nodes-in-a-feedforward-neural-netw/136542#136542*

Host K. & Ivašić-Kos M. (2022). An overview of human action recognition in sports based on computer vision. *Heliyon* **8**(6): e09633.
DOI: https://doi.org/https://doi.org/10.1016/j.heliyon.2022.e09633

Kong Y. & Fu Y. (2022). Human action recognition and prediction: a survey. *International Journal of Computer Vision* **130**(5): 1366–1401.
DOI: https://doi.org/10.1007/s11263-022-01594-9

Lee J. & Jung H. (2020). TUHAD: Taekwondo unit technique human action dataset with key frame-based CNN action recognition. *Sensors* **20**(17): 4871.
DOI: https://doi.org/https://doi.org/10.3390/s20174871

Liang J. & Zuo G. (2022). Taekwondo action recognition method based on partial perception structure graph convolution framework. *Scientific Programming* **2022**: 1838468.
DOI: https://doi.org/10.1155/2022/1838468

Liu J., Wang G., Hu P., Duan L. Y. & Kot A. C. (2017). Global context-aware attention LSTM networks for 3D action recognition. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 21–26 July, Honolulu, HI, USA, 3671–3680.
DOI: https://doi.org/10.1109/CVPR.2017.391

Mittal A. (2019). Understanding RNN and LSTM. Available at *https://aditi-mittal.medium.com/understanding-rnn-and-lstm-f7cdf6dfc14e*

Nida N., Yousaf M. H., Irtaza A. & Velastin S. A. (2022). Video augmentation technique for human action recognition using genetic algorithm. *ETRI Journal* **44**(2): 327–338.
DOI: https://doi.org/https://doi.org/10.4218/etrij.2019-0510

okankop. (2020 May). *okankop/vidaug*. GitHub. Available at *https://github.com/okankop/vidaug*

Olah C. (2015, August). *Understanding LSTM Networks – Colah's Blog*. Available at *https://colah.github.io/posts/2015-08-Understanding-LSTMs/*

Park C.-I. & Sohn C.-B. (2020). Data augmentation for human keypoint estimation deep learning based sign language translation. *Electronics* **9**(8): 1257.
DOI: https://doi.org/https://doi.org/10.3390/electronics9081257

Piergiovanni A. J. & Ryoo M. S. (2018). Fine-grained activity recognition in baseball videos. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 18–22 June, Salt Lake City, UT, USA, pp. 1821–18218.
DOI: https://doi.org/10.1109/CVPRW.2018.00226

Sun Z., Ke Q., Rahmani H., Bennamoun M., Wang G. & Liu J. (2022). Human action recognition from various data modalities: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**(3): 3200–3225.
DOI: https://doi.org/https://doi.org/10.1109/tpami.2022.3183112

Vrigkas M., Nikou C. & Kakadiaris I. (2015). A review of human activity recognition methods. *Frontiers in Robotics and Artificial Intelligence* **2**: 28.
DOI: https://doi.org/10.3389/frobt.2015.00028

Vom Brocke J., Hevner A. & Maedche A. (2020). Introduction to design science research. In: *Design Science Research* (eds. J. vom Brocke, A. Hevner & A Maedche), pp. 1–13. Springer International Publishing, New York, USA.
DOI: https://doi.org/10.1007/978-3-030-46781-4_1

Zhang L., Hsieh J.-C., Ting T.-T., Huang Y.-C., Ho Y.-C. & Ku L.-K. (2012). A Kinet based golf swing score and grade system using GMM and SVM. *2012 5th International Congress on Image and Signal Processing*, 16–18 October, Chongqing, China, pp. 711–715.
DOI: https://doi.org/10.1109/CISP.2012.6469827

Zhao R., Wang K., Su H. & Ji Q. (2019). Bayesian graph convolution LSTM for skeleton based action recognition. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 27 October–02 November, Seoul, South Korea, pp. 6881–6891.
DOI: https://doi.org/10.1109/ICCV.2019.00698

WTF (2014). *Poomsae Scoring Guidelines for International Referees.* World Taekwondo Federation. Available at *https://d17nlwiklbtu7t.cloudfront.net/983/document/ Poomsae_scoring_guidelines.pdf*

GUIDANCE TO CONTRIBUTORS

GENERAL INFORMATION

Scope

The Journal of the National Science Foundation of Sri Lanka publishes the results of research in all aspects of Science and Technology. It is open for publication of Research Articles, Reviews, Research Communications and Correspondence.

IT related and other non-empirical articles

The JNSF is a journal primarily devoted to natural sciences. It also considers for publication significant and novel contributions from formal sciences. Authors of emerging sub-disciplines of Computing and related areas such as Machine Learning, Artificial Intelligence and Data Sciences are requested to carefully adhere to the following guidelines when submitting manuscripts for this journal.

- Clear formulation of outcome-oriented **Research Objective/s** for targeted knowledge (sub)domain/s or (sub)discipline/s.
- Selection and comprehensive summarization of **appropriate Research Method/s** adopted to achieve the stated Research Objective/s.
- Reporting a sound **(Empirical) Evaluation** of the research finding/s thereby arguing reliability, validity, and generalizability of research claim/s.

Categories of manuscripts

Research Articles: Research Articles are papers that present complete descriptions of original research. Research Articles should include an Abstract, Keywords, Introduction, Methodology, Results and Discussion, Conclusion and Recommendations where relevant. References should be prepared according to the "Guidelines for the preparation of manuscripts". Maximum length of the article should be limited to 25 pages with a word count of 10,000 including references, figures and tables. Any articles above this limit will be returned.

Reviews: Reviews are critical presentations on selected topics of Science or Technology. They should be well focused and organized and avoid general "textbook" style. As reviews are intended to be critical presentations on selected topics, reviewers need to have had substantial leadership in research supported by a publication track record in the areas covered by the review. A person/s wishing to submit a Review Article should obtain prior approval from the Editorial Board by submitting a concise summary of the intended article, along with a list of the author's publications in the related area (jnsf@nsf.gov.lk). Maximum length of the article should be limited to 40 pages with a word count of 12,000 including references, figures and tables. Any articles above this limit will be returned.

Research Communications: Research Communications are intended to communicate important new findings in a specific area of limited scope that are worthy of rapid dissemination among the scientific community. Authors are required to provide a statement justifying the suitability of the submission for a Research Communication. The article should include an Abstract, Keywords, Introduction, Methodology, Results & Discussion, Conclusion and References. Maximum length of the article should be limited to 10 pages with a word count of 2,500 including references, figures and tables. Any articles above this limit will be returned.

Correspondence: Correspondence will be accepted regarding one or more articles in the preceding four issues of the Journal, as well as Letters to the Editor. Articles covering important scientific events or any other news of interest to scientists, reviews of books of scientific nature, articles presenting views on issues related to science and scientific activity will also be considered. Publication will be made at the discretion of the Editor-in Chief. Maximum length of the article should be limited to 05 pages with a word count of 1,500 including references, figures and tables. Any articles above this limit will be returned.

SUBMISSION OF MANUSCRIPT

Authors submitting articles to the JNSF should first create an account in the Sri Lanka Journals Online System (https://jnsfsl.sljol.info/). All manuscripts in MS Word format must be electronically submitted to the journal's online platform at https://jnsfsl.sljol.info/submit/start/. Submissions *via* emails are not encouraged. Please make sure that no author information is mentioned in the article submitted. The names and details of affiliations of all authors and contact information of the corresponding author must be fed into the system during the online submission process. Authors (at least the corresponding author) are required to provide their personal, validated ORCID ID (by obtaining an ORCID ID from https://orcid.org/) when submitting the manuscript. No change to the authors or order of authors will be accepted after the submission. All those who have made significant contributions should be listed as co-authors. The corresponding author should ensure that all contributing co-authors are included in the author list and have approved the final version of the paper and have agreed to its submission for publication.

All submissions should be in English. If the manuscript conforms to the guidelines specified, the date received will be the date that the manuscript was submitted to the online system.

Submissions are accepted for processing on the understanding that they will be reviewed and that they have not been submitted for publication elsewhere (including publication as a full paper or extended abstract as a part of Conference Proceedings). The JNSF does not accept manuscripts that have already been submitted to pre-print servers.

Suggesting potential reviewers by authors

The authors may suggest up to three names of referees when submitting their manuscript, in the Cover Letter space provided at the bottom of the page in the first stage of online submission. Referees should not be from the institution where the work was carried out and should not have been co-authors in previous publications. The address, institutional affiliation and e-mail of the suggested referees should be supplied. Please note that the JNSF is not bound to select all or any of the suggested referees for sending the manuscript for reviewing

Authorship

All authors designated as authors should be eligible for authorship. Those who have made a substantial contribution to the concept or design of the work; or acquisition, analysis or interpretation of data are recognized as Authors.

The corresponding author should be prompt and ensure adherence to timelines when responding to requests, queries and recommendation of reviewers conveyed by or on behalf of the Editor-in Chief and Editorial Board.

Supplementary materials

Any experimental data necessary to evaluate the claims made in the paper but not included in the paper should be provided as supplementary materials. Supplementary materials will be sent to the reviewers and published online with the manuscript if accepted. The supplementary materials should conform to Journal guidelines and should be uploaded as separate files. Authors should number Supplementary Tables and Figures as, for example, 'Supplementary Table S1'. Refer to each piece of supplementary material at the appropriate point(s) in the main article. Supplementary Materials may include description of the materials and methods, controls, or tabulated data presented in Tables or Figures, and programming codes.

Peer review

The manuscripts submitted to the JNSF will initially be screened by the Editorial Board and, if suitable, will be referred to at least two subject experts in the relevant field. The peer-review process of the JNSF is double-blind.

When revision of a manuscript has been requested, the revised manuscript should be submitted on or before the stated deadline. If the revised manuscript is not received on time, the manuscript will not be processed further. The authors' response to the comments of referees should be tabulated with the comment, response and the line number/s for reference. The decision of the Editorial Board shall be final.

Accepted papers are subject to editing. The date of acceptance will be the date when the Editorial Board has decided it to be acceptable for publication.

Article publication fee

A total of US$ 250 will be levied for each accepted manuscript for publication, except when the corresponding author is affiliated to a Sri Lankan Institute, in two stages as explained below.

- A processing fee of US$ 20 will be levied for each manuscript at peer-review stage and the remaining US$ 230 will be charged for accepted manuscripts at the time of publication.

Payments can be made online via NSF Payment Portal (http://pg.nsf.gov.lk/)

Authors' declaration

When an article is accepted for publication, the authors are required to submit the Authors' Declaration signed by all the authors.

Copyright

Articles in JNSF are published under the Creative Commons License CC-BY-ND. This license permits use, distribution and reproduction of articles for commercial and non-commercial purposes, provided that the original work is properly cited and is not changed in anyway. The copyright of the article is with the National Science Foundation of Sri Lanka. Therefore, authors are requested to check with institution's copyright and publication policy before submitting an article to the JNSF. Authors secure the right to reproduce any material that has already been published or copyrighted elsewhere. When an article is accepted for publication, the authors are required to submit the Transfer of Copyright document signed by all the authors.

Post-publication corrections

The Editorial Board reserves the right to take action on publishing an erratum or corrigendum. If serious errors are identified in a published article, the Journal may consider a retraction or publishing a correction.

STRUCTURE OF MANUSCRIPT

Manuscript

The manuscript should be free of errors and prepared in single column, using double-spaced text of Times New Roman 12 font throughout with line numbers, leaving at least 2 cm margins on both sides, and liberal spacing at the top and bottom of each page. Pages should be numbered consecutively.

a. Style

The paper should be written clearly and concisely. The style of writing should conform to scholarly writing. Slang, jargon, unauthorized abbreviations, abbreviated phrasings should not be used. In general, the impersonal form should be used. Poor usage of language will result in rejection of the manuscript during initial screening.

b. Layout

Manuscripts other than review articles should be generally organized as follows: Title, Abstract, Keywords, Introduction, Methodology, Results and Discussion, Conclusions and Recommendations (where relevant), Acknowledgements and References. Pages should be arranged in the following order:

Title page should include the title of manuscript, and no author information should be mentioned in the title page. If a major part of the research has been published as an abstract in conference proceedings, it should be cited as a footnote on the title page. Authors must also indicate the **general and specific research area** of the manuscript in the title page. In order to highlight the significance of the manuscript, authors are required to provide the following highlights in brief. (1) Why was this study conducted? (2) What are the new findings? (3) Possible applications of the findings. Please limit your answers to 25-30 words for each.

Title: Should accurately and concisely reflect the contents of the article.

Running title: Should be a shortened title (limited to a maximum of 50 characters) that could be printed at the top of every other page of the Journal article.

Abstract: Should be between 200 - 250 words for full length articles and written as a single paragraph. It should not contain any references and should be able to stand on its own. It should outline objectives and methodology together with important results and conclusions. A Review Article should carry a summary of not more than 300 words.

Keywords: Include a maximum of six keywords, which may include the names of organisms (common or scientific), methods or other important words or phrases relevant to the study.

Introduction: This should state the reasons for performing the work with a brief review of related research studies in the context of the work described in the paper. Objectives of the study should be clearly stated.

Materials and Methods: This section should give the details of how you conducted your study. New methods may be described in detail with an indication of their limitations. Established methods can be mentioned with appropriate references. Sufficient details should be included to allow direct repetition of the work by others. Where human subjects are involved, they should be referred to by numbers or fictitious names. A paper reporting the results of investigations on human subjects or on animals must include a statement to the effect that the relevant national or other administrative and ethical guidelines have been adhered to, and a copy of the ethical clearance certificate should be submitted. Methods of statistical analyses used should be mentioned where relevant.

Results and Discussion: Results: the results should be concisely and logically presented. Repetition of the same results in figures, tables or text should be avoided.
Discussion: data essential for the conclusions emerging from the study should be discussed. Long, rambling discussions should be avoided. The discussion should deal with the interpretation of results. It should logically relate new findings to earlier ones. Unqualified statements and conclusions not completely supported by data should be avoided.
Molecular sequence data, such as gene or rDNA sequences, genome sequences, metagenomic sequences etc. must be deposited in a public molecular sequence repository, such as GenBank, that is part of the International Nucleotide Sequence Database Collaboration (INSDC). The accession numbers obtained must be cited in the text, Table or on Figures of phylogenetic trees of the manuscript.

Conclusion: The conclusion should be brief, highlight the outcomes of the study and should be aligned with the objectives of the study. It should not contain references.

Conflict of interest statement: All authors should include a statement on conflict of interest disclosing any financial or other substantive conflicts of interest that may be construed to influence the results or interpretation of their research. All sources of financial support for the project should be disclosed.

Acknowledgement: Should be brief and made for specific scientific, financial and technical assistance only. If a significant part of the research was performed in an institution other than in those indicated by the authors' affiliations given in the title page, this fact should be acknowledged. All those who have made substantial contribution to the research but do not qualify to be authors should be acknowledged.

References :
All research work of other authors, when used or referred to or cited, should be correctly acknowledged in the text and in the References.

Citing references in the text:
- References to the literature must be indicated in the text and tables as per the Author-Year System, by the author's last name and year, inparenthesis (i.e. Able, 1997) or (Able & Thompson, 1998).
- Citation to work by more than two authors should be abbreviated with the use of *et al*. (i.e. Able *et al*., 1997).
- Multiple publications by the same first author in the same year should be coded by letters, (i.e. Thompson, 1991a; b).
- Multiple citations should be made in chronological order and separated by a semi-colon, (i.e. Zimmerman *et al*., 1986; Able *et al*., 1997).
- Reference to unpublished work, work in preparation or work under review should be cited in italics as (*unpublished data*) or, with the author's initials and surname given; such works should not be included in the Reference section.
- Personal communications may be mentioned in the text with the date of communication as (*Personal communication, 2 June 2000*).

List of references:
- The list of References should be arranged in alphabetical order based on the last name of the first author.
- Names of all the authors should be given except when there are more than 10 authors. When there are more than 10 authors, only the name of
  the first author can be given followed by *et al*.
- All the initials of the author must be given after the last name and the year of publication should follow in parentheses.
- This should be followed by the full title of the referred publication.
- When journal articles are listed, the journal name should be given in full and in italics and followed by the volume number in bold type, issue number in parentheses and then the inclusive pages.
- Where there are several publications by the same author(s) and published in the same year they should be differentiated by adding a lower-case letter after the year. When books are listed, the order should be: author(s), year, book title, volume number, edition, pagination/ inclusive pages, publisher and place of publication. The book title should be in italics. When sections of a book are listed, the order should be: author(s) of chapter, year, title of the section, title of the book, edition, inclusive pages, publisher and place of publication.
- Digital object identifiers (DOIs) should be included for all references where available.
- References should only be cited as 'in press' if the paper has been accepted for publication.

Examples of correct forms of references are given below.

Journal Articles
Boutin C. & Harper J.L. (1991). A comparative study of the population dynamics of five species of *Veronica* in natural habitats. *Journal of Ecology* **79**(01): 199 – 221.
DOI: https://doi.org/10.2307/2260793

Books
Burnham K.P. & Anderson D.R. (2002). *Model Selection and Multimodal Inference*, 2nd edition, pp. 488. Springer Science and Business Media, Inc., New York, USA.

Book Chapters
Hinrichsen R.A. & Holmes E.E. (2009). Using multivariate state-space models to study spatial structure and dynamics. In: *Spatial Ecology* (eds. R.S. Cantrell, C. Cosner & S. Ruan), pp. 145 – 166. CRC/ Chapman Hall, Florida, USA.
DOI: https://doi.org/10.1201/9781420059861.ch8

Edited Books

KimatI H., Amorim L., Rezende J.A.M., Bergamin Filho A. & Camargo L.E.A. (eds.) (2005). *Manual de Fitopatologia*, volume 2. Doenças das Plantas Cultivadas, 4th edition. Ceres, São Paulo, Brazil.

Conference Papers

Weaver D. (2002). Implementation of a learning management system using an integrated approach to professional development. In: Winds of change in the sea of learning. *Proceedings of the 19th Annual Conference of the Australasian Society for Computers in Learning and Tertiary Education (ASCILITE)* (eds. A. Williamson, C. Gunn, A. Young & T. Clear), volume 2, Auckland, New Zealand, 8-11 December. Unitec Institute of Technology, Auckland, New Zealand, pp. 711-720.

Agency Publications

U.S. Census Bureau (2009). *World Population: 1950 – 2050*. U.S. Census Bureau, Washington DC, USA.

Department of Health (2008). *Health Inequalities: Progress and Next Step* (pdf). Department of Health, London, UK. Available at *http:// PublicationsPolicyAndGuidance/ DH _ 08 5307* , Accessed 9 June 2008.

Other

Robinson L.J. (2003) Spatial scale and depletion models of farmland birds in a fragmented landscape. *PhD thesis*, University of Reading, Reading, UK.

Efford M.G. (2008). Density 4.3: software for spatially explicit capture-recapture. Available at *http://www.otago.ac.nz/density* , Accessed 15 March 2009.

Abbreviations and Symbols: Unless common, these should be defined when first used, and not included in the abstract. The SI System of units should be used wherever possible. lf measurements were made in units other than SI, the data should be reported in the same units followed by SI units in brackets, e.g. 5290 ft (1610 m).

Formulae and Equations: Equations should be typewritten and quadruple spaced. They should be started on the left margin and the number placed in parentheses to the right of the equation.

Nomenclature: Scientific names of plants and animals should be printed in italics. In the first citation, genus, species and authority must be given. e.g. *Borassus flabellifer* Linn. In latter citations, the generic name may be abbreviated, for example, *B. flabellifer* L.

Tables and figures: Tables and Figures should be clear and intelligible and kept to a minimum, and should not repeat data available elsewhere in the paper. Any reproduction of illustrations, tabulations, pictures etc. in the manuscript should be acknowledged.

Tables: Tables should be numbered consecutively with Arabic numerals and placed at the appropriate position in the manuscript. If a Table must be continued, a second sheet should be used and all the headings repeated. The number of columns or rows in each Table should be minimized. Each Table should have a title, which makes its general meaning clear, without reference to the text. All Table columns should have explanatory headings. Units of measurement, if any, should be indicated in parentheses in the heading of each column. Vertical lines should not be used and horizontal lines should be used only in the heading and at the bottom of the table. Footnotes to Tables should be placed directly below the Table and should be indicated by superscript lower case italic letters (*a*, *b*, *c*, etc.).

Figures: All illustrations are considered as figures, and each graph, drawing or photograph should be numbered consecutively with Arabic numerals and placed at the appropriate position in the manuscript. Any lettering to appear on the illustrations should be of a suitable size for reproduction and uniform lettering should be used in all the Figures of the manuscript. Scanned figures or photographs should be of high quality (**300 dpi**), to fit the proportions of the printed page (12 × 17 cm). Each figure should carry a legend so that the general meaning of the figure can be understood without reference to the text. Where magnifications are used, they should be stated.

Units of measurement

Length: km, m, mm, μm, nm

Area: ha, km^2, m^2
Capacity: kL, L, mL, μL
Volume: km^3, m^3, cm^3
Mass: t, kg, g, mg, μg

Time: year(s), month(s), wk(s),

d(s), h, min, s
Concentration: M, mM, N, %,
 g/L, mg/L, ppm
Temperature: °C, K
Gravity: x g

Molecular weight: mol wt

Others: Radio-isotopes: 32P
Radiation dose: Bq
Oxidation-reduction potential: rH
Hydrogen ion concentration: pH