CLASSIFICATION ANALYSIS OF THE BADMINTON FIVE DIRECTIONAL LUNGES

By:

HO ZHE WEI

(Matrix No.: 125014)

Supervisor:

Dr. Loh Wei Ping

June 2018

This dissertation is submitted to
Universiti Sains Malaysia
As partial fulfillment of the requirement to graduate with honors degree in
BACHELOR OF ENGINEERING (MECHANICAL ENGINEERING)

School of Mechanical Engineering
Engineering Campus
Universiti Sains Malaysia

Declaration

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed……………………………………………… (HO ZHE WEI)

Date…………………………………………………

Statement 1

This journal is the result of my own investigation, except where otherwise stated. Other sources are acknowledged by giving explicit references. Bibliography/ references are appended.

Signed……………………………………………… (HO ZHE WEI)

Date…………………………………………………

Statement 2

I hereby give consent for my journal, if accepted, to be available for photocopying and for interlibrary loan, and for the title and summary to be made available outside organizations.

Signed……………………………………………… (HO ZHE WEI)

Date…………………………………………………

Acknowledgments

First and foremost, I would like to offer my dedication to School of Mechanical Engineering, Universiti Sains Malaysia for implementing Final Year Project course in our curriculum. Final Year Project is given two semesters to identify, analyze and design solutions for real-life engineering problem. Knowledge and concept that I acquired for the past 4 years were able to apply in this project.

I wish express my sincerest gratitude to my supervisor, Dr. Loh Wei Ping who has guided me throughout the thesis with her patience and knowledge. She never feels bored in responding to my questions and queries and at the same time she willing to share her expertise and valuable guidance on my Final Year Project. Without her useful advice, my project would not be structured in proper and professional style.

Lastly, I am also grateful to have my course mates and seniors who have been supporting me throughout this venture no matter in terms of academic wise or mental support. We always share opinions in the journey to discover more knowledge and information..

Table of Contents

Declaration i

Acknowledgments ii

Table of Contents iii

List of Figures v

List of Tables vi

Abstrak viii

Abstract ix

CHAPTER 1: INTRODUCTION 1

1.0 Introduction 1

1.1 Project Background 2

1.2 Problem Statement 3

1.3 Objectives 3

1.4 Scope of Work 3

CHAPTER 2: LITERATURE REVIEW 4

2.0 Overview 4

2.1 Badminton Lunge Motion 4

2.2 Motion Analysis 7

2.3 Data Mining 7

2.4 Challenges and Issues 9

CHAPTER 3: METHODOLOGY 10

3.0 Overview 10

3.1 Data Collection 12

3.1.1 University-level Players 12

3.1.2 National-level Players 14

3.2 Data Pre-processing 14

3.3 Attributes 15

3.4 Data Classification 19

3.5 Significant Attribute Analysis 21

3.6 Knowledge Discovery 21

CHAPTER 4: RESULT AND DISCUSSION 22
4.0 Overview ... 22
4.1 Experimental Data-USM ... 22
4.1.1 Classification Analysis for Experimental Data-USM ... 23
4.1.2 Classification Accuracy Comparisons on Five Classifier Representatives ... 25
4.1.3 Significance of Attributes ... 26
4.1.4 Significant Attributes on REP Tree ... 28
4.2 Public Data-SEA ... 30
4.2.1 Classification Analysis for Public Data ... 31
4.2.2 Classification Accuracy Comparisons on Five Classifier Representatives ... 33
4.2.3 Significance of Attributes ... 34
4.2.4 Significant Attributes on REP Tree ... 36
4.2.5 Illustration of REP Tree Diagram ... 38
4.3 Knowledge Discovery ... 41
4.3.1 Accuracy of Original 27 attributes vs Selective attributes ... 41
4.3.2 Data count by Lunge Type (LT) ... 42
4.3.3 Lunge Type vs. Other Attributes ... 43
CHAPTER 5: CONCLUSION 47
5.0 Concluding Remark ... 47
5.1 Future Work ... 48
References 49

List of Figures

Figure 3.1: Flowchart of data mining approach for classifying the badminton lunge . 11

Figure 3.2: Experimental setup with five cameras in the badminton court [3]............ 13

Figure 3.3: The true leg length and apparent leg length [30]...................................... 15

Figure 3.4: Forward steps patterns (a) 2-step, (b) 3-step [17] 16

Figure 3.5: Boxplot with an Interquartile Range 19

Figure 4.1: Classification accuracy of pre-processed data on five classifiers with its built-in algorithms in WEKA... 24

Figure 4.2: Comparisons of classification accuracies on KStar, Decision Table, REP Tree, Multilayer Perceptron and Bayesnet classifier with ZeroR set as baseline 25

Figure 4.3: Comparisons of classification accuracies on KStar, Decision Table, REP Tree, Multilayer Perceptron and Bayesnet classifier with ZeroR set as baseline 28

Figure 4.4: Classification accuracy of pre-processed data on five classifiers with its built-in algorithms in WEKA... 32

Figure 4.5: Comparison of classification accuracies on Tree classifier algorithms using REP Tree as the baseline... 33

Figure 4.6: Comparisons of classification accuracies on Bayesnet, Multilayer Perceptron, KStar, PART and REP Tree classifier with ZeroR set as baseline........... 34

Figure 4.7: Comparison of classification accuracies using REP Tree as the baseline . 34

Figure 4.8: Absolute percentage difference impact on REP Tree classification accuracy against attribute removal... 36

Figure 4.9: Confusion matrix of REP Tree perfect classification into five lunges 38

Figure 4.10: REP Tree Diagram developed for public domain data-SEA................... 39

Figure 4.11: Classification accuracy of original and selective attribute (ID, GT, LT) 41

Figure 4.12: % of Lunge Type used by players in Experimental data-USM and Public data-SEA throughout the badminton game 43

Figure 4.13: Distribution of LT patterns against ID 44

Figure 4.14: Visualization of GT against ID.. 46

List of Tables

Table 3.1: Description of attributes measured (Raw data)... 17

Table 3.2: Classifier Algorithm Applied 20

Table 4.1: Description of attributes after IQR is applied... 22

Table 4.2: Description of attributes after removal of outlier and extreme value......... 22

Table 4.3: Significance of attribute to classification accuracy 27

Table 4.4: Classification accuracy of selective attribute on REP Tree 29

Table 4.5: Confusion matrix of REP Tree based on selective attributes (ID, GT, LT) 29

Table 4.6: Description of attributes after IQR is applied... 30

Table 4.7: Description of attributes after removal of outlier and extreme value......... 30

Table 4.8: % Classification accuracy of Tree algorithms on pre-processed public data 31

Table 4.9: Significance of attribute to classification accuracy 35

Table 4.10: Classification accuracy of selective attribute on REP Tree 37

Table 4.11: Accuracy of selective attributes.. 37

Table 4.12: If-else Rule of REP Tree Diagram.. 40

Table 4.13: Tabulation of LT against ID 44

Table 4.14: Summary of Lunge type performed at the end of GT............................... 45

Table 4.15: Summary of Lunge type performed at the end of GT............................... 46

Nomenclature/Symbols

CF – Centre-Forward Lunge

LF – Left-Forward Lunge

RF – Right-Forward Lunge

L – Left-sideward Lunge

R – Right-sideward Lunge

Kinovea – Kinovea Sports Video Analysis Software

WEKA – Waikato Environment for Knowledge Analysis

Abstrak

Gerakan lanj badminton adalah kemahiran penting untuk pemain supaya mempunyai gerak kaki asas semasa pertandingan. Kajian-kajian lampau telah melaporkan kajian gerakan lanj pada pemain lelaki. Kekangan-kekangannya adalah penemuan yang dijumpai kepada pemain lelaki tidak sesuai kepada pemain wanita. Tiada kajian yang dijalankan untuk melombong corak pergerakan lanj badminton berarah. Oleh itu, kajian ini berusaha untuk: i) mengkaji corak gerakan lanj semasa pertandingan badminton, (ii) mengklasifikasikan postur pemain badminton mengikut jenis gerakan lanj, dan (iii) membandingkan perbezaan dalam corak gerakan lanj antara pemain badminton daripada peringkat universiti dan kebangsaan. Kes kajian ini melibtkan 11 pemain peringkat universiti dan 2 pemain peringkat kebangsaan. Lima gerakan lanj berarah: tengah ke hadapan, kiri ke hadapan, kanan ke hadapan, kiri ke sisi dan kanan ke sisi dan parameter-parameter yang sepadan dikesan melalui perisian Kinovea. Konsep perlombongn data diterimapakai dalam empat peringkat: pra pemprosesan data, pengelasan data, analisis parameter bererti dan penemuan pengetahuan dengan menggunakan perisian WEKA. REP Tree merupakan pengelas yang terbaik dipilih atas kekuatan dan keupayaan pengelasannya. Data eksperimen-USM dan data awam-SEA mencecah ketepatan pengelasan yang tertinggi sebanyak 93.75% dan 93.01% masing-masing dengan pengelas REP Tree. Bagi konfigurasi terpilih yang terdiri daripada pengenalan (ID), masa tindak balas (GT) dan jenis gerakan lanj (LT), keertian pengelasan meningkat kepada 99.61% bagi data eksperimen-USM dan 100% bagi data awam-SEA . Corak gerakan lanj adalah berkait dengan ID dan GT. Kesimpulannya, penemuan ini menunjukkan bahawa identiti, masa tindak balas dan jenis gerakan lanj merupakan penentu-penentu utama bagi pengelasan lanj badminton bagi menghasilkan ketepatan pengelasan tertinggi dengan pengelas REP Tree.

Abstract

Badminton lunge motion is important skill for players in order to have a fundamental footwork in badminton. Majority previous badminton studies on lunge motions investigated male players. The gap was that the findings reported were not applicable to the female players. There are no works conducted to mine the patterns of directional badminton lunge motions. Therefore, this study attempted to (i) study the patterns of lunge motion in the badminton game, (ii) classify badminton players' postures by lunge type and (iii) compare the differences in the badminton lunge patterns between university and national level players. The case study involved 11 university level and 2 national level players in badminton singles captures. Five directional lunge motions: center-forward, left-forward, right-forward, left-sideward and right-sideward lunge and its corresponding attributes were tracked through Kinovea software. Data mining concept is adopted in four stages: data pre-processing, data classification, significant attribute analysis and knowledge discovery using the WEKA software. REP Tree classifier is the best selected classifier for its strength and classification capability. The highest classification accuracy obtained for experimental data-USM and public data-SEA, were 93.75% and 93.01% respectively on REP Tree classifier. On selective attribute configuration, the identity (ID), game reaction time (GT) and type of lunge (LT) significantly enhanced the classification accuracy to 99.61% for experimental data-USM and 100% for the public data-SEA. Lunge type patterns were related to ID and GT. Conclusively, the identity, game reaction time and type of lunge were found being the key determinants for badminton lunge classification accounting for highest classification accuracy in REP Tree algorithm.

CHAPTER 1
INTRODUCTION

1.0 Introduction

Badminton is the fastest racket sport in the world with smash speed around 320km/h. This game requires high physical demand and stable footwork in order to rapidly change posture and motion during the game [1]. Good footwork requires good posture for a shot execution while maintaining good body control.

Lunging is the fundamental footwork in badminton, in which, players move into secure base for shuttlecock hitting and rapidly move back into the court to prepare for the next shot [2]. Skillful badminton players will have strong fundamental in footwork, therefore they are able to reach shuttlecock as quick as possible with minimum lunge step and game reaction time. The lunging steps are commonly categorized into right-forward, left-forward, front-forward based on the lunging directions to compare the foot loading differences [3]. Lunge techniques are also analysed to understand the players' biomechanics movement interactions and forces as well as to predict the badminton lunge pattern. A video-based pilot study has confirmed that lunging contributes around 15% of all movements in any competitive singles games [2].

Previous studies considered with ground reaction force in order to improve badminton performance and prevent injury on players. Consequently, most of the findings concerned on injury prevention strategies because good physical condition of a badminton player impacts badminton performance the most. Nevertheless, few works have reported the badminton lunge pattern in badminton games. Data mining approach could be a useful tool in the badminton in order to discover interesting knowledge in the lunge pattern. Majority of sports motion recognition systems were based on traditional machine learning algorithms. However, there were no study applied data mining in the badminton lunge motion. For instances, Vijayakumar and Nedunchezhian [4] presented an overview of the data mining applications in sports. The potential applications of video mining also included such as video shot detection and pattern analysis. Yu et al. [5], whereas, multimedia and interactive data acquisition system and intelligent analysis system for the techniques and tactics in net sports (TTNS) developed decision support system based on data mining. Therefore, this study aims to apply the data mining concept into badminton game case studies focusing on five directional lunge motions: center-forward, left-forward, right-forward, left-sideward and right-sideward lunge. The specific goals include (i) to study the patterns of lunge motion in the badminton game (ii) to classify badminton players' postures by lunge type (iii) to compare the differences in the badminton lunge patterns between university and national level players.

In this study, WEKA software will be used to aid the data mining analysis in order to interpret and mine the badminton lunge data on video-image-numeric transformation. The video captures is initially transformed into a sequence of images represented in a time frame in Kinovea. The first level involved data pre-processing analysis in order to segregate outliers and extreme values from the informative data. Next, classification analysis is performed to assign the pre-processed data into five lunge classes. Data patterns are evaluated to identify the representative attributes for accurate classification. Significant attribute analysis is performed to investigate impact of attribute on badminton lunge pattern. Comparisons are performed between the university-level and national-level players. The interesting classified patterns are translated into relation and modelled into mathematical rules.

1.1 Project Background

This study focused on the effect of five directional lunge motions: center-forward, left-forward, right-forward, left-sideward and right-sideward lunge in the badminton game of university and national level players. The attributing factors include gender, game reaction time, distance of lunge step, speed of lunge, number of lunge steps and so on. The, game reaction time and speed of lunge are considered as the decisive factors of badminton lunge pattern[6]. Game reaction time is defined as the time interval to quickly complete a lunge and return to the start or move off in another direction. Besides, the distance and number of steps taken from original static stand of the player to all parts of the court are considered.

Data mining technology is widely used in sport for technical and tactical analysis, including badminton game. Though data mining is not reported in badminton lunge motion, it has been applied in other sports-related patterns such as soccer and tennis [7]. The WEKA, a machine learning tool was employed for data mining analysis in order to interpret and mine the badminton lunge data. The effect of lunge motion on badminton game can be predicted through knowledge discovery from the badminton game data mining analysis.

1.2 Problem Statement

Most recent studies focused on the kinematics and dynamics badminton motion via complex experimental or computational methods. In badminton game, vast amounts of data collected for each player by training session and seasonal games stored by the Badminton World Federation (BWF). However, no efforts done to translate the vast amount of data to study the effects of lunge motion patterns. At the same time, no study has applied data mining approach to classify the lunge motions patterns in the game.

1.3 Objectives

This project aims to:

(i) study the patterns of lunge motion in the badminton game

(ii) classify badminton players' postures by lunge type

(iii) compare the differences in the badminton lunge patterns between university and national level players.

1.4 Scope of Work

The study applied data mining approach to classify badminton lunge pattern based on five directional lunge maneuver: center-forward, left-forward, right-forward, left-sideward and right-sideward lunge. There were two case studies involved: (i) An experimental case study involving 11 (five males, six females) university-level players in badminton singles of 21 points (experimental data-USM) and (ii) A public domain case study involving two (one male, one female) national-level players in 29th SEA Games Badminton singles. The demographic (age, gender), anthropometric (players' height, weight, apparent leg length, true leg length) and characteristics of badminton players were measured and compared. The hypothesis was that the quantitative demographic and anthropometric attributes could distinguish categories of five lunge type. The lunge motion postural data in the badminton singles is analyzed on data mining approach, specifically on classification analysis using Tree, Function, Rules, Bayes and Lazy algorithms to compare classification accuracy between university-level and national-level players. Classification accuracies will be considered on percentage correctly classified instances and time needed to build the model. Significant attribute analysis is performed in order to distinguish lunge type patterns based on the main attributes.

CHAPTER 2
LITERATURE REVIEW

2.0 Overview

This chapter presents the state-of-the-art review on previous studies related to badminton lunge motion, motion analysis and data mining. The focus of past works particularly on the key attributes were discussed. The strengths and weaknesses from reported works were addressed.

2.1 Badminton lunge motion

Previous studies on badminton game motions put emphasis on the kinetics and kinematics aspects of the highly dynamic movement techniques such as smash, backhand overhead strokes and the drop motion. Lunge motion is not the most salient kinematic analysis in badminton game but it contributes around 15% of all movements in the court during any competitive badminton singles [2].

Kuntze et al. [2] investigated nine professional male badminton players on their lunge performance from three perspectives which are kick, step-in, and hop lunge. However, their study only involved male players. Their study found lower mean horizontal reaction force at drive-off and lower mean peak hip joint power observed during the step-in lunge as compared to the kick lunge. Therefore, the step-in lunge is concerned in this study.

Mohammad and Chinnasee [8] compared the effects of step (SFL) versus jump forward lunge (JFL) on muscle architecture before and after training intervention among badminton players. Their results showed SFL caused significant changes in muscle thickness of fascicle length of vastuslateralis and pennation angle of vastusmedialis after the test.

Mei et al. [9] conducted kinematics analysis on eight national-level and university-level badminton players. However, similar to Kuntze et al. [2], their study were only targeted on male players. In fact, there were proven difference in the badminton performances between the male and female players [10]. Mei et al. highlighted that the national level players acquired higher peak pressure and force-time integral in medial forefoot during lunge as compared to the university level players. At the same time, Light and McLellan [11] emphasized on the experienced players whom required lesser vertical and horizontal force magnitude (significant difference at $p<0.01$) during impact loading when compared to the less experienced players. The skill level difference motivates this project to compare the lunges for both national and university level badminton players.

Hong et al. [12] had identified the left-forward lunge experiencing high loading magnitude during the heel impact, thus being a critical maneuver for badminton. However, the authors' study only concerned forward and backward lunge. From a different perspective, Hu et al. [3] indicated that the left and right maximum forward lunges induced greater plantar loads on the great toe region of the dominant leg of badminton players as compared to the front-forward lunge. However, similar to Mei et al. and Kuntze et al. there were limitations in their study whereby only male players were recruited. Thus, the results of study may not represent the female badminton players.

Electromyography (EMG) is a method used to detect the level of voluntary activation in a muscle. Nadzalan et al. [13] revealed that all EMG data during high load forward lunge (70FL) were significantly higher compared to those recorded during low load (30FL). Furthermore, muscle of dominant limb experienced greater activation as compared to non-dominant limb during loading. Despite imbalances of muscle adaptation in lower extremity, dominant and non-dominant limb are considered in this study to examine their relationship with lunge patterns. Nadzalan et al. [14] further supported that the dominant limb acquired lesser game reaction time and greater step distance when compared to the non-dominant limb during step forward lunge (SFL) and jump forward lunge (JFL). Slower movement in the non-dominant leg reflects the lack of strength compared to the dominant leg.

Marcos et al. [15] implied that the game reaction time and accumulated time frame increased with uncertainty. Meanwhile, speed of the lunge reduced the effect of uncertainty. Prior prediction of opponent's action which means uncertainty reduced, as a result, player can execute faster, reduce movement time and increase the success of lunging in a game. This finding motivates the current study to explore speed of lunge, game reaction time and accumulated time frame.

Lin et al. [16] expressed that the three-step movement was significantly faster than two-step when both biomechanical variables were compared. Moreover, there was a greater hip adduction torque in three-step compared to the two-step footwork movement.

Gavkare et al. [17] indicated that faster reaction time among athletes over non-athletes can improve the concentration and alertness, better muscular coordination as well as improved performance at speed. Therefore, game reaction time was also emphasized in this study as study attributes. In relation to reaction time, the athletes portrayed faster reaction time scores and higher consistency but fewer errors in anticipation time compared to non-athletes [18]. Both visual and perceptual skills could be taken as attributes to study their effects on the badminton lunge motion.

Lam et al. [19] denoted that no significant difference was found between repetitive and single movement lunge trials on the approaching speed. The approaching speed was defined as the averaged speed from the starting position to the initial contact of the force plate.

Majority previous badminton studies on lunge motions investigated male players. The gap was that the findings reported were not applicable to the female players. This study considers both male and female players to avoid the limitation. We considered left and right forward lunges following Hu et al. [3] and the step forward lunge following Mohammad and Chinnasee [8]. Considering the findings from Lin et al. [16] the number of steps was concerned in this study. Unlike Lam et al.[19], this study utilized to identify the lunging speed of the players.

2.2 Motion Analysis

Subordinate leg lunges exercise (SLLE) exhibited homogeneity in the angle variable which is maximum knee extension repetition wise (MKER). On the other hand, heterogeneity is considered in SLLE for maximum knee flexion repetition wise (MKFR) and total time taken repetition wise (TTTR). It was found that these angle and temporal variables differed by gender [20].

In response to motion analysis, Huang and Cham [21] revealed that the cost function is a tracking algorithm that can be implemented in C++ without code optimization and is suitable for real-time processing. The shuttlecock can be tracked efficiently from different views based on the algorithm's performance.

The methodology of this study is inspired by the study of Taha et al [22] whom evaluated real time motion tracking of badminton in six degree of freedom. A motion capture video camera is set up to record motion of player and analysed the parameters through Kinovea software.

The Kinovea software was used in this study following Ahlawat et al. [20] whom applied the same software in the analysis of two-dimensional motion video. Kinovea is a sports video motion analysis software is used to analyze recorded video. In order to obtain accurate motion data from the video, the pixel distance in the video must be appropriately calibrated based on the known length of a real-world object reference present in the video. The lengths of the court lines were used to calibrate the pixel distance in the video for the X-Y plane as well as the Y-Z plane.

2.3 Data Mining

While most studies had considered the physics and biomechanics aspects of the badminton game, data mining application was only used to focus on the effects of overall badminton technique and tactics analysis instead of lunge motion. Any technical and tactical actions in badminton game that can be defined were considered an object of data mining [23]. Relationship between frequent itemsets can be obtained through technical and tactical statistical data.

Rassem et al. [24] introduced a recognition method for sports with single-handed swings like badminton from track motion data. The produced features are recognized by a support vector machine and resulted in precision rate of 95.67%. However, triaxial accelerometer is not applicable in this study because the gadget is a little bulky and thus might affect lunge performance of player during a real badminton game.

Huang and Shi [23] researched on the optimization of badminton training based on association rule in data mining. The national badminton game scene is retrieved from China Super League. This algorithm reduced the complexity of training mode and improved the learning speed. However, the difference of statistical results that measured rate of hitting shuttlecock was very small, thus the evaluation of the performance is basically the same.

Anik et al. [25] identified activities that were common in badminton game, including smash, serve and backhand considering the upper extremity. Their results showed that the support vector machines (SVM) classifier has more decent recognition rate of badminton game activities than K-Nearest Neighbours (k-NN). However, the dataset of 180 instances in data mining was considered small. Therefore, it was learnt that the size of dataset in this study need to go beyond 180 instances in order to increase stability of accuracy. .

Pernek et al. [26] discovered that acceleration information provides useful information for inferring heart rate during badminton game and training intensity. Selected machine learning algorithms had to deal with numerical class data. Their findings showed that the Multilayer Perceptron (MP) yielded the similar correlation coefficient as linear regression (LR) but with lower mean absolute error. Although LR is most feasible model, however only numerical data is applicable.

Lin et al. [27] implemented the machine learning technique in cell phone APP in order to classify types of badminton stroke strategy, including clear, smash and drop. Their results showed that the random forest showed an accuracy of 79.32% for general model testing and 95.91% for personal model validation. However, the accuracy gap was significant.

From a different viewpoint, Huynh and Bedford [28] suggested visual based training method to be implemented in badminton game. This is because it was found that the neural networks showed higher accuracy (100%) in prediction on different skill level groups when compared to discriminant analysis (80.5%).

Li et al. [29] employed the K-nearest neighbor (KNN) to identify badminton event for 2004 all-British Open Tournament. This rule-based detection algorithm judged the positions of the video event according to the prior structural knowledge based on the shot classification and labeling results.

2.4 Challenges and Issues

Overall reported studies in the badminton lunge motion analyses were focused on biomechanical analysis in terms plantar loads characteristics and ground reaction force. The lunge tasks commonly performed better on the dominant leg. The step-in lunge technique as mentioned in [2], [3], [8], [9] had inspired the current project analysis on badminton singles game performance of national-level and university-level players. At the same time, few works have also considered multi-directional lunge in badminton using data mining approach to classify the badminton performance. Thus, the intention of this study is to further explore the effects of five directional lunge motions during the badminton singles.

CHAPTER 3
RESEARCH METHODOLOGY

3.0 Overview

This chapter involved interpretation of lunge motion effects on the badminton patterns. In section 3.1, data collection were retrieved from experimental and public domain case studies for validation purposes and comparison of lunge skills. For section 3.2, WEKA was involved in data pre-processing analysis in order to segregate outliers and extreme values from the informative data. For section 3.3, 27 attributes measured in raw data are tabulated and described. In section 3.4, classification analysis was performed to assign the pre-processed data into five lunge classes: center-forward, left-forward, right-forward, left-sideward and right-sideward. In section 3.5, significant attribute analysis was conducted to study contributions of the attributes towards data classification accuracies. For section 3.6, several study attributes such as identity, game reaction time and lunge type were highlighted for knowledge discovery analysis.

The research implementation was generally divided into four stages: Data Collection, Pre-processing, Data Classification and Knowledge Discovery (Figure 3.1).

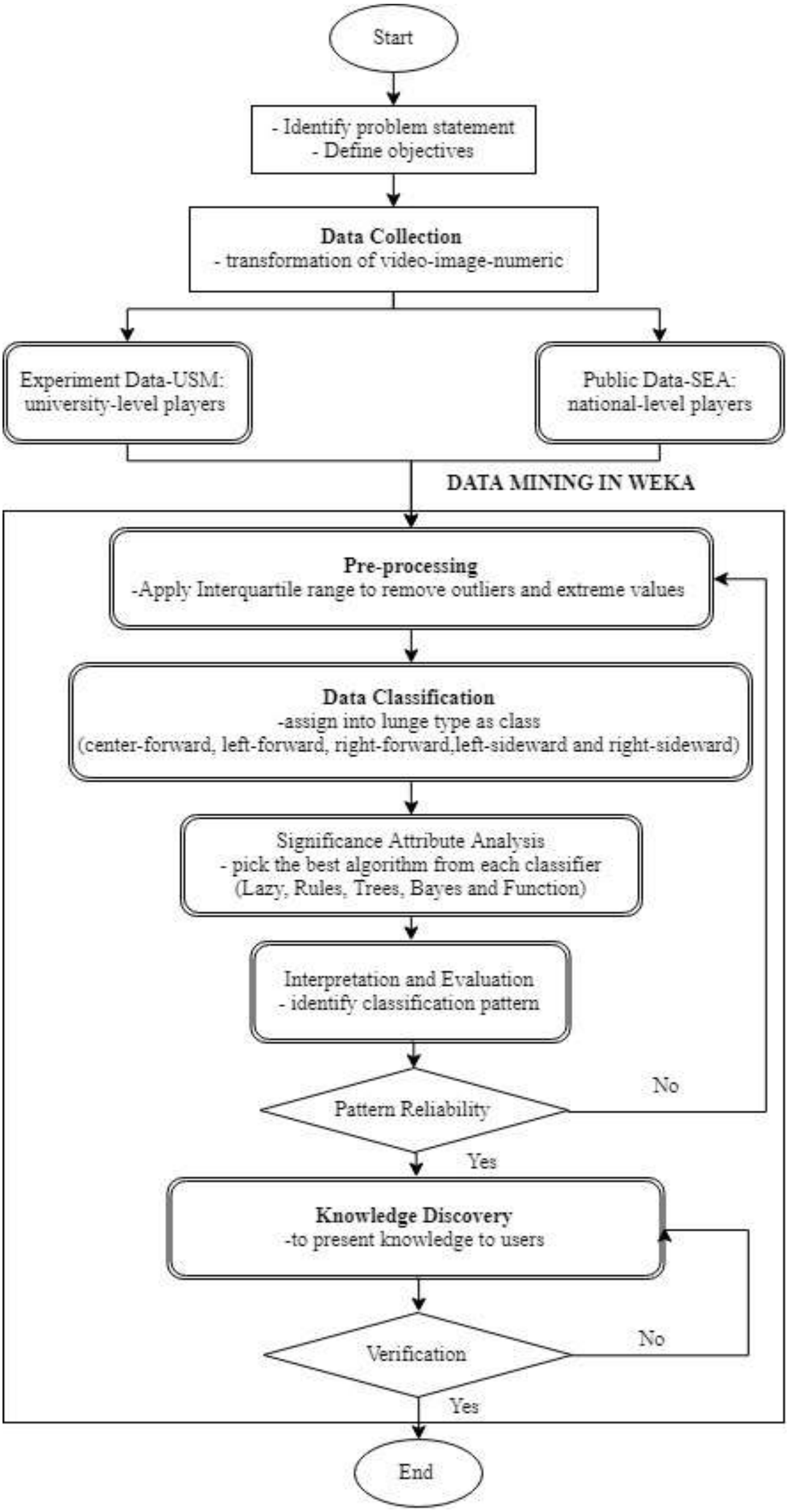

Figure 3. 1: Flowchart of data mining approach for classifying the badminton lunge

3.1 Data Collection

The project involved experimental data (experimental data-USM) that was collected from 11 university level players from Universiti Sains Malaysia (USM), whereas public domain data (Public data-SEA) was collected from two national-level players from 29^{th} SEA Game. Both available domains case studies focused on the lunge motion during the badminton singles.

3.1.1 University-level Players

There was 11 university level (USM representative with at least five years' experience) badminton players participated in this study, involving 5 males and 6 females (ages: 22.3±1.1 years old; height: 166.6±7.5 cm; weight: 56.1±6.7 kg). Prior permission was obtained from the Sports and Recreation Center, USM. All participants are right-handed players and free from any lower extremity injuries prior to experiment. All players were well informed on the experimental procedures for the research purpose and were asked to provide their consents to participate in the experiment.

The experiments were conducted indoor at the badminton court of Azman Hashim USM Sports Arena, Universiti Sains Malaysia. The recording system consisted of five cameras set up at five positions around the badminton court as shown in Figure 3.2. Camera dslr2 focused on opponent's side view, phone2 focused on player's front view, sports camera (SC) focused on player's and the opponent's side view, dslr1 focused on player's side view and the phone1 focused on opponent's front view. The players were required to play the standard 21-point badminton singles spontaneously without specific restrictions. The natural spontaneous lunge motions performed in five different directions are illustrated in Figure 3.2, including center-forward, left-forward, right-forward, left-sideward and right-sideward based on the player's foot orientations. Center-forward lunge is considered when lunge was performed at 0°, left-forward and right-forward are considered when lunge was performed in the range of $0° <$ lunge degree $\leq 45°$ and lastly left-sideward and right-sideward are considered when lunge was performed in the range of $45° <$ lunge degree $\leq 90°$.

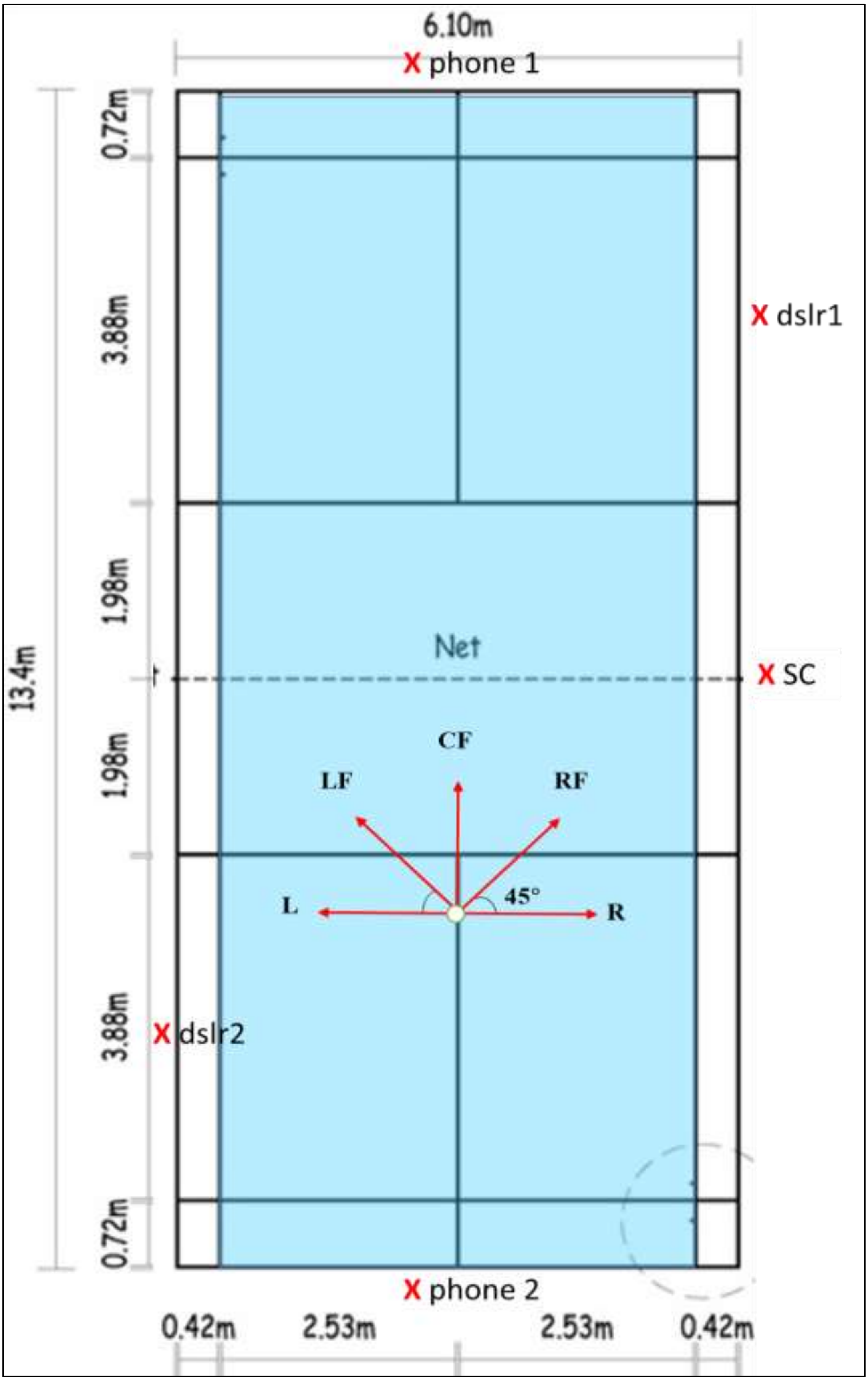

Figure 3. 2: Experimental setup with five cameras in the badminton court [3]

3.1.2 National-level Players

The public domain case study involved two national-level public domain videos on 21-point badminton singles. KL2017 29th SEA Games Women's Singles Finals [34] and KL2017 29th SEA Games Men's Singles Finals [35] were retrieved respectively from the official YouTube due to its high definition quality video. This case study was meant for compatible benchmarking with the university player experiments for validation purposes and to compare lunge skills with the university-level players. The two national-level players to be studied consisted of a male and a female (ages: 19.314±1.489 years old; height: 167.195±10.423 cm; weight: 67.692±6.452 kg). Both players were the gold medallist for badminton singles in 29^{th} SEA Games. Based on the videos, the relevant lunge maneuver of national-level players is identified in five different directions: center-forward, left-forward, right-forward, left-sideward and right-sideward. However, there were some limitations of obtaining the side view as the public domain videos were only available in front view shots, either the lunge posture of the elite-national level player or the opponent view can be determined.

3.2 Data Pre-processing

Captured video data for the experimental and public available domain undergo data transformation video-image-numeric under Kinovea. The video-image transformation is performed using "Track Path" function in Kinovea. Number of frames per second (fps) at capture time for every video data was according to Kinovea's default system. Fps is a unit that measures display device performance, whereby number of consecutive images that can be handled by a display device each second (equation (3.1)).

$$Frame\ per\ second\ (fps) = \frac{1}{time\ frame\ (s)} \tag{3.1}$$

There were 24 fps (0.04167s) for experimental data-USM and 25 fps (0.04s) for public data-SEA respectively.

Attributes extracted from the video files included game reaction time, x and y-coordinate of leg, distance of lunge and speed of lunge. The score point and number of step are recorded. The game reaction time of the player was directly obtained from the time frame. Meanwhile, the distance of lunge are determined from the x and y-coordinates of the leg based on the Pythagorean Theorem calculation (equation (3.2)), whereas the speed of lunge can be determined from the distance of lunge by dividing with time frame (equation (3.3)).

$$Distance\ of\ lunge\ (m) = \sqrt{(dx)^2 + (dy)^2} \quad (3.2)$$

where

$dx = x_2 - x_1$

$dy = y_2 - y_1$

$$Speed\ of\ lunge\ (ms^{-1}) = \frac{Distance\ of\ lunge}{Time\ frame} \quad (3.3)$$

Based on the motion-tracked images, a subsequent image-numeric transformation followed by exporting the data to Microsoft Excel spreadsheet. The recorded numeric data is tabulated in .csv format readable by WEKA tool (Table 3.1).

3.3 Attributes

The type of lunge which consisted of center-forward, left-forward, right-forward, left-sideward and right-sideward are focused as the class attributes. In order to further explore the badminton game performance the influence of the demographic and anthropometric characteristics of players were essential. The demographic factors are age and gender. Anthropometric measurements considered including the players' height, weight, apparent leg length and true leg length (Figure 3.3). Apparent leg length is defined as unilateral asymmetry of the lower extremity without any concomitant shortening of the osseous components of the lower limb. Meanwhile, the true leg length is defined as differences in leg length resulting from inequalities in the bony structure.

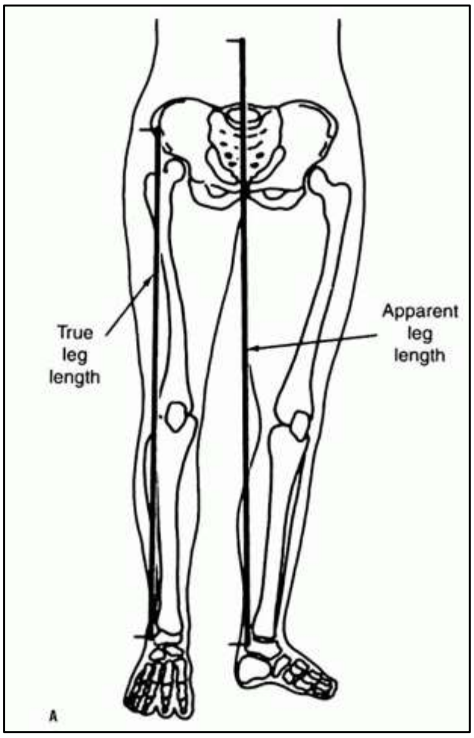

Figure 3. 3: The true leg length and apparent leg length [30]

Other physical body parameters that affect the performances of players (Table 3.1) include NOS, x-l, y-l, dx-l, dy-l, DL, SL, x-r, y-r, dx-r, dy-r, DR and SR. NOS measured number of step performed by the player from starting position during lunge. For an example, the number of step for forward-directional lunge (Figure 3.4) can either be one, two or three steps.

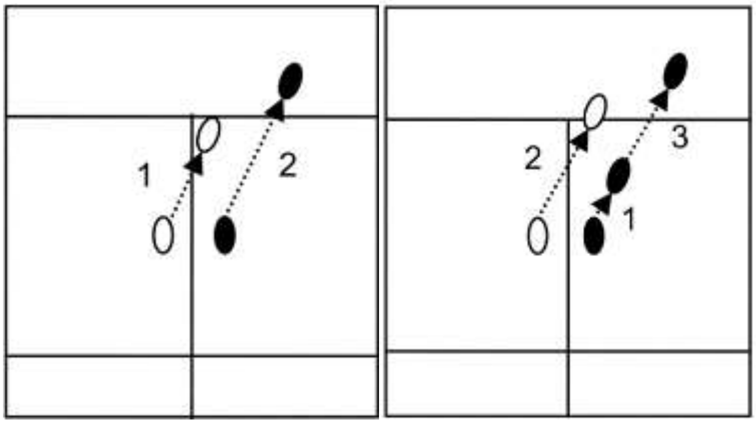

Figure 3. 4: Forward steps patterns (a) 2-step, (b) 3-step [16]

DL that measured the distance of lunge for left leg can be obtained from Pythagorean Theorem (Equation 3.2). x-l and y-l measured the x and y coordinate respectively from image-numeric transformation in Kinovea. In the meantime, dx-l and dy-l measured the difference between two coordinates for x and y-axis respectively. SL measured the speed of lunge for left leg based on Equation 3.3. The speed of lunge is defined as the average speed from the starting position towards the shuttlecock hit position. For right leg, similar steps could be performed by x-r and y-r to obtain DR and derived to get SR.

Other relevant parameters to distinguish the performances of the players are T, AT, GT and Sc. T referred as the time frame that is set by the default of Kinovea. AT referred as the accumulated time frame, which means addition of previous and current time frame. GT is defined as the length of time taken for the player to hit the shuttlecock and recover to the starting position. The game reaction time is analyzed starting from the moment of player's leg begin to move from starting position to hit the incoming shuttlecock until the player return back to the starting position, then the lunge step is considered as successful. Sc measured the score point of player for the whole game. The calculation of score is based on standard league method, player had to win the game by at least two points. Table 3.1 described all the 27 attributes measured in raw data.

Table 3. 1: Description of attributes measured (Raw data)

No.	Attribute	Description	Scale type	Experimental Data-USM	Public Data-SEA
1	ID	Identity	Nominal	{tn-TST, tn-NYL, sn-Sab, sn-NJL, tl-TYA, tl-LPY, ct-CTS, ct-TCH, jf-Joa, jf-FCL, ot-OKH}	{goh, chris}
2	A	Age	Numeric	[21 – 25]	[18 – 21]
3	G	Gender	Nominal	{ F, M }	{ F, M }
4	H	Height	Numeric	[154 – 178] (cm)	[158 – 179] (cm)
5	W	Weight	Numeric	[44.7 – 73.1] (kg)	[62 – 75] (kg)
6	TLL	True Leg Length	Numeric	[80.33 – 98.0] (cm)	[85 – 96] (cm)
7	ALL	Apparent Leg Length	Numeric	[85.0 – 99.33] (cm)	[89.7 – 98.2] (cm)
8	T	Time Frame	Numeric	[0 – 0.05] (s)	[0 – 0.04] (s)
9	AT	Accumulated Time Frame	Numeric	[0 – 2.95] (s)	[0 – 2.64] (s)
10	GT	Game Reaction Time	Numeric	[14.3 – 707.45] (s)	[58.16 – 1329.96] (s)
11	x-l	x-coordinate of leg(left)	Numeric	[-13.45 – 4.31] (m)	[-6.65 – 6.56] (m)
12	y-l	y-coordinate of leg(left)	Numeric	[-1.78 – 1.51] (m)	[-1.39 – 2.56] (m)
13	dx-l	Change in x-coordinate of leg(left)	Numeric	[-1.33 – 1.07] (m)	[-0.58 – 0.91] (m)
14	dy-l	Change in y-coordinate of leg(left)	Numeric	[-0.7 – 0.38] (m)	[-0.45 – 0.44] (m)
15	DL	Distance of lunge step (left)	Numeric	[0 – 1.332] (m)	[0 – 0.922] (m)

Table 3.1: Description of attributes measured (Raw data)

No.	Attribute	Description	Scale type	Experimental Data-USM	Public Data-SEA
16	SL	Speed of lunge (left)	Numeric	[0 – 33.3] (ms^{-1})	[0 – 23.057] (ms^{-1})
17	x-r	x-coordinate of leg(right)	Numeric	[-11.98 – 5.22] (m)	[-7.33 – 8.07] (m)
18	y-r	y-coordinate of leg(right)	Numeric	[-1.54 – 1.03] (m)	[-1.49 – 2.95] (m)
19	dx-r	Change in x-coordinate of leg(right)	Numeric	[-1.37 – 1.08] (m)	[-2.18 – 0.92] (m)
20	dy-r	Change in y-coordinate of leg(right)	Numeric	[-0.53 – 0.35] (m)	[-0.42 – 1.37] (m)
21	DR	Distance of lunge step (right)	Numeric	[0 – 1.396] (m)	[0 – 2.336] (m)
22	SR	Speed of lunge (right)	Numeric	[0 – 31.9] (ms^{-1})	[0 – 58.406] (ms^{-1})
23	NOS	Number of Step	Numeric	{1, 2, 3}	{1, 2, 3}
24	Sc	Score point	Numeric	[0 – 28]	[0 – 21]
25	LT	Type of Lunge	Nominal	{CF, LF, RF, L, R}	{CF, LF, RF, L, R}

The raw data undergo qualitative inspection through data pre-processing process. An unsupervised attributes filter, Interquartile Range (IQR) of WEKA tool is adopted to aid in identifying potential outliers and extreme values that might affect the accuracy of classification when LT is set as class of data (equation (3.4)). IQR is a measurement of how spread out the data points in a set are from the mean of the dataset as shown in Figure 3.5 [31].

$$Interquartile\ Range\ (IQR) = Q_3 - Q_1 \tag{3.4}$$

where Q3 is the upper quartile and Q1 is the lower quartile

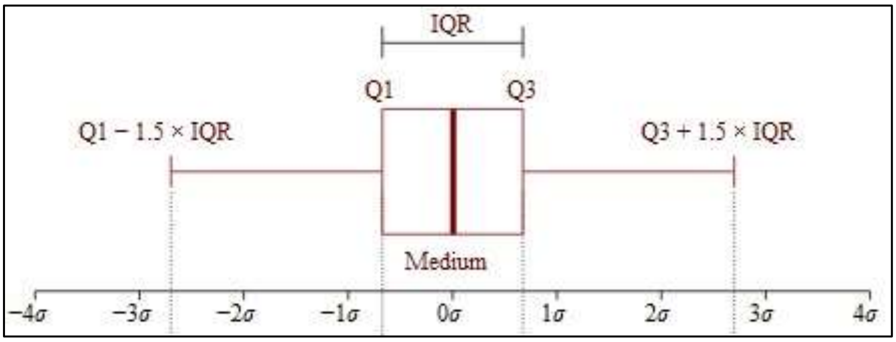

Figure 3. 5: Boxplot with an Interquartile Range

The second level involved data qualitative inspection which involved data cleaning to segregate outliers and extreme values from the informative data.

Potential outliers are identified from IQR based on the equation (3.5) and (3.6) [31]:

$$Outlier < Q_1 - 1.5(IQR) \quad (3.5)$$

$$Outlier > Q_3 + 1.5(IQR) \quad (3.6)$$

Extreme values referred to the maximum and minimum values of a function [32].

3.4 Data Classification

Data classification stage sorted and categorized data into several distinct classes aided by the WEKA tool. Data were initially segregated by players (ID) followed by demographic factors (A and G), anthropometric measurements H, W, ALL and TLL), physical body parameters (NOS, x-l, y-l, dx-l, dy-l, DL, SL, x-r, y-r, dx-r, dy-r, DR and SR) and other relevant parameters (T, AT, GT and Sc). The LT was regarded as the class attribute for experimental and public domain data. Several classification algorithms built in the WEKA tool were adopted at preliminary classification analysis as listed in Table 3.2. These algorithms were used to group data into five lunge classes: center-forward, left-forward, right-forward, left-sideward and right-sideward.

Table 3. 2: Classifier Algorithm Applied

Classifier	Algorithms	Advantages	Disadvantages
Lazy	IBK Kstar LWL	- Ability to construct infinite data structures - Does not suffer from data interference	- Large amount of memory to store the data
Rules	Decision table Jrip OneR PART ZeroR	- Easy to generate - Performance comparable to decision trees - Easily handle missing values and numeric attributes	- High computational cost - Memory and computationally intensive
Trees	Decision Stump Hoeffding Tree J48 LMT Random Forest Random Tree REP Tree	- Simpler presentation - Easy to create and understand - Can be represented as rules - Fast and scalable	- Easily overfit - Accuracy depends a lot on the data presented
Function	Logistic Multilayer perception Simple logistic SMO	- Easily updated with new data - Use for the regression and mapping	Time-consuming require a very large amount of data
Bayes	Bayesnet Naive Bayes Multinomial text Updateable	- Less training data needed - Can handle continuous and discrete data. - Insensitive to irrelevant features.	- Loss of accuracy due to class conditional independence - Unable to learn interactions between features

Selection of algorithm from each classifier was based on several considerations. First of all, accuracy of algorithm was prioritized, thus algorithm with highest accuracy from each classifier was considered. However, algorithm with 100% accuracy was considered because it unable to perform significant attribute analysis since maximum accuracy was achieved. Next, time to build model was the second consideration. Therefore, algorithm with highest accuracy but time-consuming will be replaced by algorithm with second highest accuracy but took shorter time to build model.

3.5 Significant Attribute Analysis

In order to examine the contributions of the attributes towards data classification accuracies, the significant attribute analysis was performed. The significance of every single attribute from pre-processed data was used to classify with the five classifier representatives. This was to investigate its role and impact on the classification accuracy. Then, the best classifier representative was picked based on several criteria considerations such as accuracy of classification, time to build model and easiness of presentation for further interpretation in terms of selective attributes. If tree classifier was picked, the result should be portrayed in tree diagram and detailed explanation of each section of tree was needed respectively to investigate their accuracy and relationship.

3.6 Knowledge Discovery

Highest accuracy attained by selective attribute configuration from both experimental and public domain data were compared to identify their percentage difference. Not only that, count of lunge type for experimental data-USM and public data-SEA were compared in the window of WEKA Explorer in order to examine type of lunge perform the most and least by university and national level players respectively. Not only that, relationship of LT with other attributes, such as demographic factors, anthropometric measurements, physical body parameters and other relevant parameters are studied in the next section.

CHAPTER 4
RESULTS AND DISCUSSION

4.0 Overview

The data mining analyses from the experimental data-USM and public data-SEA are executed using WEKA Explorer and presented in section 4.1.1 and 4.2.1. Data are classified based on lunge type (LT) being the class attribute. Both case study data are segregated into raw data, pre-processed data and pre-processed data with combination of significant attributes for better classification accuracy. The accuracy of pre-processed data is verified on WEKA Experimenter in section 4.1.2 and 4.2.2. WEKA Experimenter ran differently from WEKA Explorer. WEKA Explorer studied relationship of data with different attribute, visualized data in interface and analyzed type of algorithms to run in experiments, whereas WEKA Experimenter designed experiments with selection of algorithms and datasets, ran experiments and analyzed the results.

4.1 Experimental Data-USM

On video-image-numeric transformation using the Kinovea software, the raw experimental data consists of 27 attributes and 10894 instances without missing values. The LT: center-forward lunge (CF), left-sideward lunge (L), left-forward lunge (LF), right-sideward lunge (R) and right-forward lunge (RF) is defined as the class attribute for data classification analysis. At data pre-processing level, 2902 are identified as outliers and 2588 are identified as extreme values after unsupervised attribute filter (Interquartile Range) is applied (Table 4.1). After removal of outlier and extreme value, remaining 6382 instances as pre-processed data (Table 4.2).

Table 4. 1: Description of attributes after IQR is applied

No.	Parameter	Scale type	Experimental Data-USM
25	Outlier	Numeric	No (7992); Yes (2902)
26	Extreme Value	Numeric	No (8306); Yes (2588)

Table 4. 2: Description of attributes after removal of outlier and extreme value

No.	Parameter	Scale type	Experimental Data-USM
25	Outlier	Numeric	No (6382); Yes (0)
26	Extreme Value	Numeric	No (6382); Yes (0)

4.1.1 Classification Analysis for Experimental Data-USM

Classification performances are considered on the raw data followed by pre-processed data using five classifiers with 23 algorithms. The lunge patterns are measured by percentage accuracies of correct classification into five lunge types: CF, L, LF, R and RF. Classification accuracies on raw data showed Decision Table the highest accuracy (98.8985%), followed by Random Forest (98.0173%) and PART (96.0896%) based on Appendix 3. More than half of the accuracies on pre-processed data decreased within the range of 5%, except IBK dropped the most by -6.96%. On the other hand, the accuracies of pre-processed data in Bayes classifier increased the most, for instance, Naive Bayes Updateable increased the most (19.81%) and followed by Naive Bayes (19.49%).

The classification accuracies of pre-processed data on 23 algorithms of five classifiers are shown in Figure 4.1. According to Appendix 3, the Tree classifier resulted in the highest average accuracy (82.4484%) whereby 5 out of 7 algorithms indicated accuracies above 90%. The remaining classifiers showed average accuracies of 79.4485% for Rules classifier, followed by 76.3606% for Lazy classifier and then 73.0688% for Function classifier. On the other hand, Bayes classifier gave the lowest average accuracy (59.4171%) which is considered relatively not reliable.

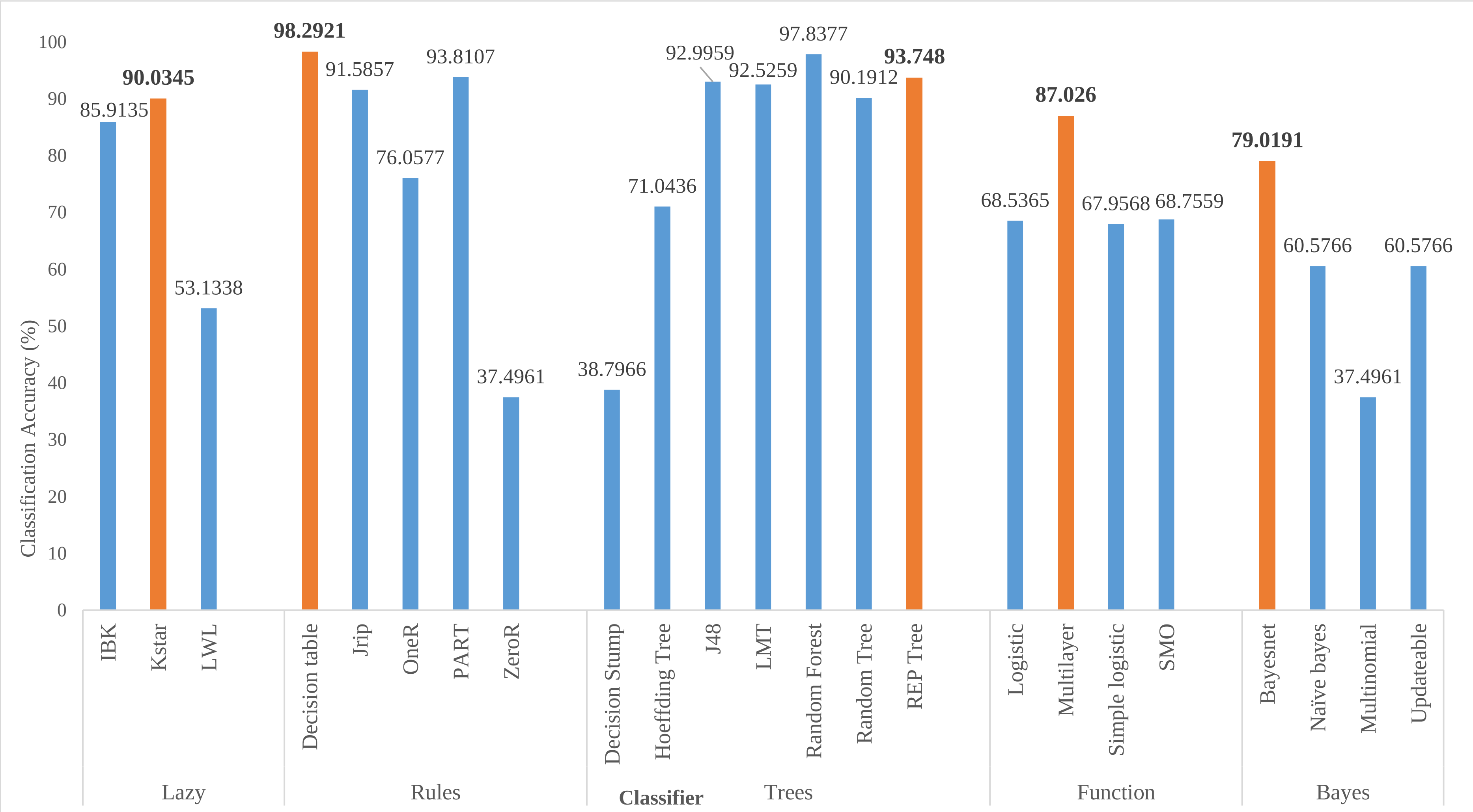

Figure 4. 1: Classification accuracy of pre-processed data on five classifiers with its built-in algorithms in WEKA