BIOMETRICS-BASED USER IDENTIFICATION WITH OPTIMAL FEATURE EVALUATION AND SELECTION

by

Namrata Kayastha, BS

THESIS

Presented to the Faculty of

The University of Houston-Clear Lake

In Partial Fulfillment

Of the Requirements

For the Degree

MASTER OF SCIENCE

in Computer Information Systems

THE UNIVERSITY OF HOUSTON-CLEAR LAKE

DECEMBER, 2019

BIOMETRICS-BASED USER IDENTIFICATION WITH OPTIMAL FEATURE EVALUATION AND SELECTION

by

Namrata Kayastha

APPROVED BY

_____
Kewei Sha, PhD, Chair

_____
Wei Wei, PhD, Committee Member

_____
Kwok-Bun Yue, PhD, Committee Member

RECEIVED/APPROVED BY THE COLLEGE OF SCIENCE AND ENGINEERING:

_____
David Garrison, PhD, Associate Dean

_____
Miguel A. Gonzalez, PhD, Dean

Dedication

I dedicate this thesis to my beloved family in Nepal – my parents Nabin and Bina, my big brothers Bipin and Sachin, my sisters-in-law Jyotsna and Sharmila, my nephews Meezan and Sayun, and my entire Kayastha family. I would not be where I am today without your immense love and support. Thank you all for encouraging me to pursue my dreams and boosting my morale through all these years of studying abroad.

Acknowledgements

First, I would like to express my sincere gratitude to my thesis advisor, Dr. Kewei Sha for recognizing my potential and giving me an incredible opportunity to work on this research project. I cannot thank him enough for his patience, valuable feedback, continuous encouragement and guidance, and always willing to help attitude throughout my thesis journey. And, I am proud to share that our conference paper on this research was published in Proceedings of The Fourth IEEE/ACM Conference on Connected Health: Applications, Systems, Engineering Technologies (IEEE/ACM CHASE 2019).

Beside my advisor, I would like to thank the rest of my thesis committee: Dr. Wei Wei and Dr. Kwok-Bun Yue for their precious time, constructive discussions, insightful comments, and support.

I am also thankful to all the volunteers who took their valuable time to participate in this research by wearing the activity sensor and walking with me in the hallway of Delta building several times.

Very special thanks to my soulmate, Manish Dhaubhadel for his unwavering love, support, encouragement, and belief in me. He has been my biggest support system and a shoulder to cry on when I needed. Because of him, I am one step closer to my goal.

Last but not the least, my appreciation goes to all my family and friends near and far, who understood that as a grad student I was busy most of the times, and constantly motivated me to stay focused and work harder towards my master's degree.

Thank you all from the bottom of my heart.

ABSTRACT

BIOMETRICS-BASED USER IDENTIFICATION WITH OPTIMAL FEATURE EVALUATION AND SELECTION

Namrata Kayastha
University of Houston-Clear Lake, 2019

Thesis Chair: Kewei Sha, PhD

Recently, biometrics-based identification algorithms have gained popularity as a means of identifying a person using their unique behavioral characteristics such as gait or hand movement pattern. Classifications based on biometric features are broadly used in modern healthcare applications, including user identification, authentication, and tracking. The complexity and accuracy of classification algorithms largely depend on the size and the quality of the feature set used to build classifiers. In this thesis, we mostly focus on feature evaluation and selection as these are the essential steps to decide a small set of high-quality features to build accurate and efficient classifiers in user identification. We propose a novel and efficient approach to evaluate and select biometric features for user identification based on activity sensor data collected from the users' wrists while they are walking. For each feature, we first generate an NRMSD matrix, each entry of which represents the similarity level of any two users. Based on the observations from NRMSD matrices, we define two heuristics, Farness Value and Farness Ratio to evaluate

the quality of the feature. We evaluated a total of 72 features and selected 18 high-quality features based on our evaluation results. Finally, we train our data with different classifiers and select KNN as the best classification model. Compared to other feature evaluation and selection techniques, this approach is more efficient and yields a higher accuracy of 98.3%.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I:

INTRODUCTION

1.1 Background and Significance

User authentication is an effective mechanism to protect malicious access to sensitive resources. Identification is a crucial component in the authentication protocol design as the purpose of the authentication is to verify the identity of the user [1-4]. Over the past few decades, several identification technologies have been developed that can uniquely identify users and prevent impersonation. It is important that these identification solutions provide practical and cost-effective approach to easily identify the user as well as offer a smooth user experience. Username/password based identity is widely adopted in the digital world [5]; yet they are susceptible to hacking, theft, and fraud. A digital signature based on cryptographic algorithms is another popular approach for building a verifiable identity [6]. It is an effective solution, but it requires a powerful processor to generate digital signatures; therefore, resource-constrained devices have difficulty in creating such an identity. Recently, a hardware-based solution, Physical Unclonable Function (PUF) [7], has evolved as a way to identify users and many authentication protocols are built based on it. PUF provides a strong identity solution but it requires extra hardware support. Similarly, tokens and access cards [8] provide a hardware-based solution for identity.

Biometrics-based identity solutions are the next frontier of identification and verification [9]. They are considered more effective than the aforementioned digital identities because of the following reasons. First, biometrics are natural part of the user. Unlike other traditional means of identity verification like usernames/passwords, PINs, tokens, etc., biometrics cannot be forgotten, lost or stolen [10]. Second, biometrics are unique for each individual, therefore they are hard to be forged. Third, the biometrics-

based identities are easily verifiable by measuring the biometric characteristics [11]. Most of the biometrics solutions require special hardware to measure the biometrics. This can be expensive, inconvenient as well as very intrusive to the user's experience.

1.2 Motivation and Research Challenges

As smartphones, smartwatches, and wristbands become pervasively available, many sensors, such as accelerometers and gyroscopes embedded on these devices can be used as measuring devices for biometrics. Therefore, we can design solutions that construct and verify digital identity for users in a cost-effective and convenient way, by using these sensors to measure biometrics. A number of activity sensor-based user identification approaches have been proposed in the past few years. Previous designs have used different activities such as walking, running, jumping, and arm gestures for identities [4, 12, 13]. However, we have not yet seen large-scale deployments of these technologies because of the following concerns. First, we have observed the deployment of sensors on different body parts, including waist [1, 2], leg [3], sternum [14], wrist [15] and on different parts of the body [16]. Many of them are not practical in real-life scenario. Considering the rising popularity of smart watches (e.g., Apple Watch) and activity bands (e.g., Fitbit), we believe it is more practical to make use of activity data collected to construct identity with the help of the activity sensors installed on these devices. In this way, we do not need to add any extra sensors to the human body. Second, the accuracy of the identification needs to be improved. Third, many existing algorithms are too heavy to be deployed in resource constrained smart devices such as a smart lock. Hence, we aim to design a lightweight and high-accurate identification algorithm based on activity data collected by the sensors embedded on these devices.

Despite many research efforts have been made in developing an efficient algorithm, there are still challenges that need to be addressed. First, the big size of the

feature set increases the complexity of the identification algorithm. We need to keep the feature set size as small as possible. On the other hand, we do not want to miss important features that work well to produce the uniqueness of identity. It is a challenge to identify a user accurately based on a small set of high-quality features. Second, many user identification applications have real-time requirements, but many embedded devices like smart lockers, smart wristbands are heavily resource constrained, including a slow processor and a small-size memory. Therefore, the user identification algorithms need to be lightweight and can be executed in many types of smart devices. Third, to provide a smooth user experience and to satisfy real-time requirements, the identification should be completed in a very short period of time, like less than a minute. Hence, only a small set of data should be collected. Finally, it is difficult to achieve both efficiency and accuracy in a lightweight algorithm. Thus, the trade-off needs to be investigated for a better balance between the efficiency and the accuracy.

1.3 Novelty of the Research

We tackle the challenges presented in the last section by designing the ActID framework, which consists of a feature evaluation and selection mechanism, a set of high-quality features from multiple perspectives, and a sliding-window based identity modeling algorithm which are discussed in detail in Chapter III. The novelty of our proposed method is three-fold. Firstly, we design a novel feature evaluation and selection method which evaluates the extracted features, selects useful features and removes irrelevant ones. Therefore, we can keep the size of feature set as small as possible. It also reduces the algorithm complexity. Secondly, we identify a set of high-quality features which can distinctly identify individuals. Thirdly, we provide a smooth user experience with our proposed framework. Unlike other research methods where the users have to wear multiple sensors on different parts of the body, our experiment only requires the

users to wear one wrist sensor and walk normally like they do on a plain surface for about a minute each for two sessions.

1.4 Research Design and Results

1.4.1 Research Design

First, we design our ActID Framework that consists of the identity modeling phase and the identification phase. For our experiment, we construct the activity model based on a small set of data, i.e., 15 seconds of user activity data from the wrist sensor to construct user identity. The collected activity data is preprocessed to ensure high quality data. Then, we define three novel measures to evaluate and select biometric features, namely, NRMSD, Farness Value, and Farness Ratio. Finally, we implement a system prototype and deliver a comprehensive evaluation of the proposed algorithms comparing our approach with other widely used techniques. Our detailed research methodology is covered in Chapter 3.

1.4.2 Research Results

Based on our feature evaluation method, we discarded 75% of the total features that we evaluated and selected the remaining 25% features as a set of high-quality features for classification. Comparing our approach with two other previous efforts by Kumar et al. [15] and Damaševičius et al. [49, 50], we discarded more low-quality features and achieved a higher identification accuracy of 98.3% using KNN.

1.5 Organization of Thesis

This chapter introduces the significance of research in biometrics-based user identification and presents our motivation behind this study. The remainder of the thesis is organized as followed: Chapter II overviews our extensive literature review that covers the current state-of-the-art in biometric identification and system design principles related to this thesis. Chapter III illustrates our methodological approach along with the rationale

of our design. Chapter IV discusses the experimental results of our research. Finally, we conclude this thesis and identify future work in Chapter IV.

CHAPTER II:

LITERATURE REVIEW

2.1 Biometric User Identification

Biometrics-based user identification is an effective solution to identify or verify individuals based on their unique physiological or behavioral characteristics [17]. Physiological biometrics is associated with the precise measurements, dimensions, and physical traits of an individual. Example of physiological characteristics include fingerprints, hand geometry, ECG/EEG patterns, facial features, and iris patterns. On the other hand, behavioral biometrics is based on the behavioral patterns of an individual. Behavioral characteristics include the dynamics of signatures and keystrokes, voice, gait pattern, full body motion, head movement, and hand movement patterns [18].

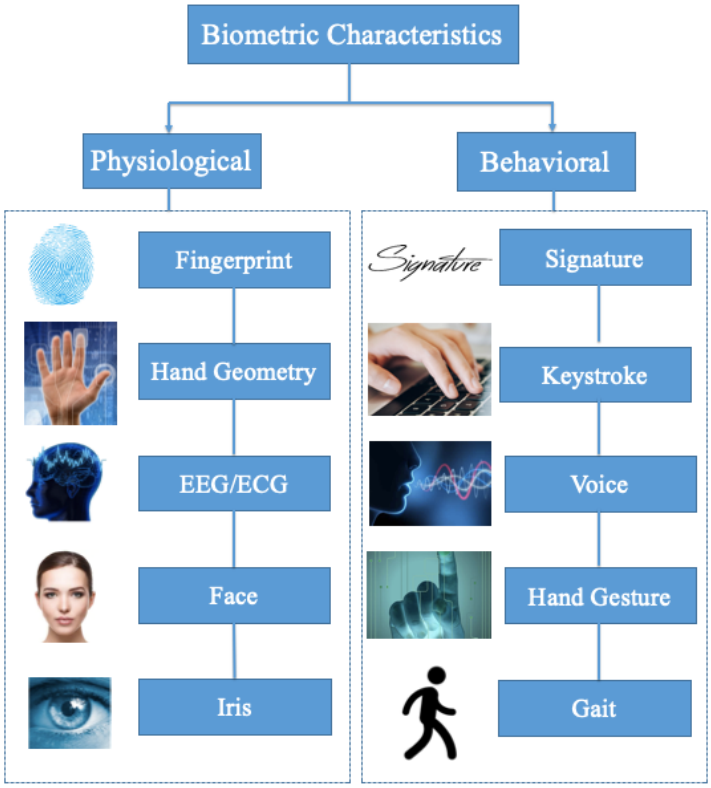


Figure 2.1. Types of Biometrics

Biometrics-based user identification has attracted significant research attention in recent years. Many biometrics-based identities have been developed and applied in modern computing systems. Based on the above mentioned two categories, some of the most popular biometric identification are discussed below [19]:

2.1.1 Physiological Biometrics

- ***Fingerprint recognition***: Fingerprint recognition is the most widely adopted biometrics-based identity which uses the ridges and valleys found on the surface tips of a human finger to identify an individual [20, 21]. It is becoming more common on smartphones and PCs, mainly to access the system or log in to applications.
- ***Hand Geometry Recognition***: Hand geometry recognition uses the geometric features of the hand such as the lengths of fingers and the width of the hand to identify an individual [22]. Hand geometry recognition systems are commonly used for attendance tracking, physical access, and personal verification.
- ***ECG/EEG Signals***: The studies have shown that the electrical activity of the brain recorded in Electrocardiogram (ECG) and Electroencephalogram (EEG) signals have unique features among individuals. ECG is the recordings of electrical activity of the heart over a period of time, and EEG is the recording that is the result of electrical activity across the scalp. Biometrics using ECG/EEG signals provide several advantages to an identification system because of these signals are unique, confidential, and secure to an individual [23-25].
- ***Face Recognition*** : Face recognition is another popular biometrics-based identity which uses the facial features or patterns for the authentication or recognition of an individual's identity. Most of the new phones in the market from Apple, Samsung, LG, etc. have face recognition feature to unlock phones. For example,

Apple even replaced it's Touch ID fingerprint scanner with Face ID on some devices like iPhone X series and iPhone 11 [26].

- ***Iris Recognition***: Iris recognition is another alternative for noninvasive verification and identification of people that uses the patterns of veins in the back of the eye to accomplish recognition. [27]

2.1.2 Behavioral Biometrics

- ***Signature Recognition***: Signature recognition is the authentication of an individual by the analysis of handwriting style, in particular the signature. It is one of the most successful behavioral biometric recognition methods which is socially and legally accepted all over the world [28].
- ***Keystroke Recognition***: Keystroke recognition is establishing identity based on the unique characteristics of a person's typing or finding habitual rhythm patterns in the way they type [29, 30].
- ***Voice Recognition***: Voice recognition uses voice as a method of determining the identity of a speaker for access control. For example, if the speaker claims to be of a certain identity and the voice is used to verify this claim [31].
- ***Hand Gesture Recognition***: Hand gesture recognition has evolved immensely in the recent few years because of its ability to interact with machine efficiently [32, 33]. There have been many researches done to interpret sign language using cameras and computer vision algorithms.
- ***Gait Recognition***: Gait recognition establishes identity based on an individual's walking style or gait. Comparing to other biometrics-based identification like fingerprint and iris recognition, gait recognition does not require the user's interaction, and it can be done at a visible distance [34].

In contrast to physical biometrics, behavioral biometrics are easily gathered with existing hardware or wearable sensors that require less power consumption, requiring only software for analysis purpose. Hence, it makes behavioral biometrics cost-effective and easy to implement. Our study falls into the category of behavioral biometrics, in particular to hand movement while walking, captured through a wrist wearable sensor that contains built-in accelerometer and a gyroscope.

2.2 Activity Sensor-Based Identification

In behavioral biometrics, activity sensor-based user identification have shown a great research potential in the last few years. Based on the data collected by activity sensors, such as accelerometers and gyroscopes, researchers have analyzed the activity patterns of humans and have found unique traits that can be used as the identity. In literary, activity sensors have been used in identifying users based on their keystroke dynamics [30], hand movements [32, 33], and gait patterns [1-4, 10, 16, 34-50].

One of the most popular activity-based biometric characteristics is gait because researchers have shown it to be feasible means for authentication. Table 2.1 summarizes some of the recent studies on gait recognition based on activity sensor. Ailisto et al. [1] were the first to propose sensor-based gait authentication. Their gait authentication was based on the acceleration sensor that was attached to the user's waist. They applied cross-correlation as a measure of similarity achieving 6.4% of EER. Their approach was further developed and analyzed by Gafurov et al. [3]. Some designs have used sensors attached to different parts of the body (e.g., leg, waist, hip, arm, and all over the body) for gait authentication [16], which is not practical in real-life scenario. Therefore, we have not yet seen large-scale deployments of these technologies.

Table 2.1

A summary of gait recognition based on activity sensor

Study	Subjects	Sensor Location	Results
Ailisto et al. [1]	36	Waist	EER: 6.4%
Mantyjarvi et al. [2]	36	Waist	EER: 7% - 19%
Gafurov et al. [3]	21	Lower leg	EER: 5%, 9%
Al Kork et al. [16]	50	Leg, hand, wrist, pant pocket, shirt pocket and bag (left and right side)	EER: 0.17% - 2.27%
	23	Hand (holding smartphone)	EER: 1.23% - 4.07%
Derawi et al. [10]	51	Pocket attached to the belt (right-hand side of the hip)	EER: 20.1%
Rong et al. [35]	21	Waist	EER: 5.6%, 21.1%
Sun et al. [36]	22	ankle	EER: 3.03%
Kwapisz et al. [37]	36	Front pants leg pocket	Accuracy: 82.1%, 92.9%
Thang et al. [38]	11	Trouser pocket position	Accuracy: 92.7% (SVM)
Johnston et al. [39]	59	Wrist (smartwatch)	EER: 2.6% - 8.1%

Modern smartphones and wrist-wearables are equipped with powerful sensors that capture activity sensor data of individuals who carry them. Hence, these devices are unobtrusive, easier to carry, and convenient to collect activity data for user identification compared to other technologies. Studies related to activity recognition with smartphone sensors and wrist-wearable sensors are examined in the sections below.

2.2.1 Smartphone Sensor-Based Activity Recognition

The ubiquity and multi-purpose functionality of smartphones have revolutionized the method of capturing biometric features. Biometric features such as fingerprint scanning, facial recognition, voice recognition and iris scanning are used as security features to unlock smartphones. Smartphones have become a rich data source to measure human activities such as walking, jogging, sitting, climbing stairs, and so on [40].

Nickel et al. [4] developed a method to extract gait features using k-nearest neighborhood algorithm and demonstrated its feasibility on smartphone achieving EER of 8.24%. Al Kork et al. [16] developed a multi-model biometric database for human gait using wearable sensors and a smartphone. They achieved very low EER of 0.17% to 2.27%. At the same time, it can be noted that they have used five sensor nodes on different body locations in addition to a smartphone with built-in accelerometer and gyroscope sensors held in hand. We, on the other hand, have used a single sensor node in our method. Also, their data collection time is 4.5 minutes while our data collection time is 60 seconds out of which we use only 15 seconds of data.

Garcia et al. [33] were the first to consider hand dynamics for authentication based on hand movement while opening a door. They used sensors, namely, accelerometer, gyroscope, and magnetometer embedded in Google Nexus 4 smartphone to collect sensor data. For classification, they proposed a machine learning based approach, consisting of various statistical and physical features and Support Vector Machine (SVM). With their approach, they achieved an accuracy of 92%. However, it would be more convenient to use wrist-wearable sensor instead of attaching the smartphone to the outside of the glove.

Most studies on smartphone-based gait recognition assume that the phone is placed at a fixed location (e.g., waist, pocket, or hand) so that they can disregard the variations introduced in the walking pattern captured by motion sensors due to changes in the placement of the phone (e.g. from pocket to hand) [41]. However, in a real situation, there is no precise location of the phone on user's body and no proper framework that can locate position of the phone automatically exists currently [15]. A better approach would be to use a wrist wearable sensor because it actually has a fixed location i.e., wrist. For our study, we focus on capturing activity data though wrist wearable sensors.

2.2.2 Wrist Wearable Sensor- Based Activity Recognition

Wrist-wearables like smartwatches and wristbands provide great advantages over smartphones, particularly in gait authentication because users usually wear their smartwatches or wristbands in the same location and orientation. Compared to the most common location for smartphones such as pockets or handbags, the wrist location provide more accurate information about a user's movements [39]. Wearable sensors-based activity recognition have several useful applications in health care, patient or elderly monitoring, rehabilitation training, and many other areas of human interaction [42]. Due of its rising popularity, location consistency, and wide applicability, it is more practical to collect activity data from wrist-wearables for user identification.

In [39], Johnston et al. used smartwatch to collect gait data and achieved the EER of 2.6% using features derived from the accelerometer data and EER of 8.1% using data derived from gyroscope data. They showed their result using 6 types of features, namely, average, standard deviation, average absolute difference, the time between peaks, binned distribution and average resultant acceleration by training five minutes of the dataset with a maximum accuracy of 98.3%.

Kumar et al. [15] proposed four continuous authentication designs by using the characteristics of arm movements while individuals walk. They collected motion data with smartwatch's sensor. Their first design uses accelerometer sensor to capture acceleration of arms, the second design uses gyroscope sensor to collect rotation of arms, third uses the combination of both accelerometer and rotation at the feature level and the fourth design uses the fusion at score-level.

A recent study done by Liu et al. [43], illustrated an approach for authentication using 20 different features from time and frequency domain. They adopted C4.5 decision tree in their proposed scheme and achieved accuracy of 86.7%. The author concluded

with the need for feature selection strategy to improve the performance of the model and reduce computational complexity.

Our study in this thesis act as a follow up of Liu et al.'s [43] work. Our experiment only use 15 seconds of data and we tackle the challenges faced by the previous studies. With our proposed framework, we intend to keep the size of feature as small as possible, identify a set of high-quality features that can help distinctly identify individuals, provide smooth user experience, as well as provide a promising result.

2.3 Feature Analysis in Biometric User Identification

Feature analysis is considered as an essential preprocessing step to biometrics-based user identification. For better classification results, it is important to analyze the discriminability of these features as they directly affect the classification results. Previous work on biometric literature has mostly given priority to feature extraction and classification while ignoring the feature evaluation and selection process. Some of the common feature selection and extraction approaches are ReliefF, Principal Component Analysis (PCA), Correlation Based Feature Selection (CFS or CFSS), Information Gain Feature Ranking (IGFR), and Random Projections [51].

ReliefF is one of the popular feature selection methods that ranks importance of features by weighing them based on their quality or relevance [52]. The feature relevance is determined by how well data instances are separated. For each data instance, Relief algorithm finds the nearest data point from the same class and nearest data points from different classes. This method is considered to be efficient because of low computational complexity. However, one of the drawbacks of this method is that it is sensitive to noise.

Principal Component Analysis (PCA) is a technique used to reduce data dimensionality. It simplifies the complexity in high-dimensional data and reduces noise while maintaining trends and patterns by transforming the data into fewer dimensions due

to orthogonality of components [53]. PCA helps to easily interpret the data, but it will not always find the important patterns. Also, with PCA, the covariance matrix is difficult to be evaluated accurately, and even the simplest invariance could not be captured by the PCA unless the training data explicitly provides this information [49].

Researchers have proposed various feature evaluation and selection algorithms to improve the quality of feature set. Among all the existing literature, [15, 49, 50] did an outstanding job in providing comprehensive work on feature evaluation and selection, which are relevant to our research study. This paper examines three feature evaluation methods which are discussed in the following subsections.

2.3.1 Correlation Based Feature Subset Selection (CFSS)

CFSS is a method that ranks subsets of features by a correlation based heuristic evaluation function [54]. A feature is considered as a good one if it is relevant to the target concept, however not redundant to any other relevant features. The measure of goodness is determined by a correlation between features. CFS selects the subset of features which has the highest measure. Hence, the chosen subset of features have high correlation with the class and unrelated to each other [51].

Kumar et al. [15] applied CFSS method implemented in Weka, with five search methods such as the best first(BF), genetic search (GS), greedy stepwise (GRS), linear forward selection (LFS), and subset size forward selection (SSFS). These search methods use somewhat different mechanisms to choose the resulting subset. The authors also showed the correlation among the subset of features from the set of acceleration and rotation based features by using the CFSS method. However, the CFSS method fails to provide any comparison between the features with respect to the users. On the contrary, our design of feature evaluation shows the evaluation of each feature with respect to the users.

2.3.2 Information Gain Feature Ranking (IGFR)

Another popular feature selection technique is to calculate the information gain. In this method, the information gain (also called entropy) for each attribute for the output variable is calculated. The values range from 0 (no information) to 1 (maximum information). The attributes that contribute more information will have a higher information gain value and can be selected, whereas the attributes that do not add much information will have a lower value and can be removed.

In [15], along with CFSS method, Kumar et al. used IGFR method to compensate for the drawback of CFSS by providing a measure of discriminability for each feature separately and ranking them accordingly with respect to the class label [26]. They relied on attribute evaluator InfoGainAttributeEval, which is a WEKA implementation for information gain-based feature selection that measures how each feature contributes in decreasing the overall entropy. Hence, this method evaluates the rank of the feature by computing the information gain of each feature with respect to the class. Conversely, this method does not work well for attributes with large number of distinct values because of overfitting issue, and it fails to accurately discriminate among the attributes.

2.3.3 Random Projection

Random projection is a simple technique used to reduce the dimensionality of a set of points that lie in Euclidean space. It is well-known for its power and low error rates when compared to other methods. Damasevicius et al. [49, 50] proposed a method for human activity recognition and user identification using random projections to reduce the dimension of features so that they improve the efficiency of classification by lowering dimensionality feature space [29, 30]. In their proposed model, the best random projection with the smallest overlapping area is selected. Then, they selected the top three [29] and top ten [30] best features ranked by the Matlab Rankfeatures function using the

entropy function. Rankfeatures in Matlab ranks key features by class separability criteria. The good thing about this function is features that are highly correlated with already selected features are less likely to be included. However, the disadvantage is that it assumes that data classes are normally distributed which is not always the case. Therefore, their feature evaluation approach is not quite stable and efficient, and there also lacks a mechanism to decide the number of features, which results in missing of many high-quality features.

2.4 Summary of Literature Review

From the analysis of our literature review, we find that the above mentioned feature selection algorithms are mostly common algorithms used in feature evaluation and selection for user identification. These methods have some disadvantages and are not efficient enough. Table 2.2 provides the summary of feature selection methods discussed in the previous section. Considering that the goal of biometrics-based user identification is to efficiently differentiate users, we believe the feature selection algorithms should be designed and optimized to align with the application goal, i.e., to select a set of high-quality features that distinctly differentiate one user from others. Therefore, this paper proposes a novel and efficient approach to evaluate biometric features focusing on better user identification. The contribution of this paper is to use Normalized Root Square Mean Difference (NRMSD) [55] to measure the similarity level of any two users based on the user's activity data for each feature. Then, we define the Farness Value and Farness Ratio to evaluate the quality of the feature. Our fundamental goal is to find a minimum set of high-quality features by eliminating low-quality features based on the feature evaluation results and use the selected features for user identification.

Table 2.2

Summary of feature selection methods. Adapted from [49]

Method	**Advantages**	**Disadvantages**	**Complexity**
ReliefF	Low computational complexity	Unstable due to random selection of instances	$O(p \cdot n \cdot \log n)$, where n are data points, each represented with p features
PCA	High dimensionality reduction; reduction of noise; lack of redundancy of data due to orthogonality of components	The covariance matrix is difficult to be evaluated accurately; even the simplest invariance could not be captured by the PCA unless the training data explicitly provides for it	$O(p^2 n + p^3)$
CFSS	It evaluates a subset of features are less likely to be included	It fails to select locally predictive features when they are overshadowed by strong, globally predictive features	$O\left(n\left(\frac{p^2 - p}{2}\right)\right)$
IGFS	Determines the relevance of an attribute and its order in the decision-tree	Fails to accurately discriminate among the attributes; does not work well for attributes	$O(p^2 \cdot n)$
Random Projections	Simplicity of implementation and scalability; robustness to noise; low computational complexity	Highly unstable – different random projections may lead to radically different clustering results	$O(dkN)$, where k is the resulting dimensionality

CHAPTER III:

METHODOLOGY

The purpose of this research is to propose an optimal solution for feature evaluation and selection in biometrics-based user identification based on activity sensor data collected from the user's wrist. For this, we first designed our ActID Framework, then we proceeded with the data collection and data analysis process. Next, we defined new measures to evaluate and select biometric features. In the following sections, we explain our methodological approach in detail.

3.1 Design of the ActID Framework

The steps of the ActID framework is depicted in Figure 3.1. The framework consists of two phases, the *identity modeling phase* and the *identification phase*.

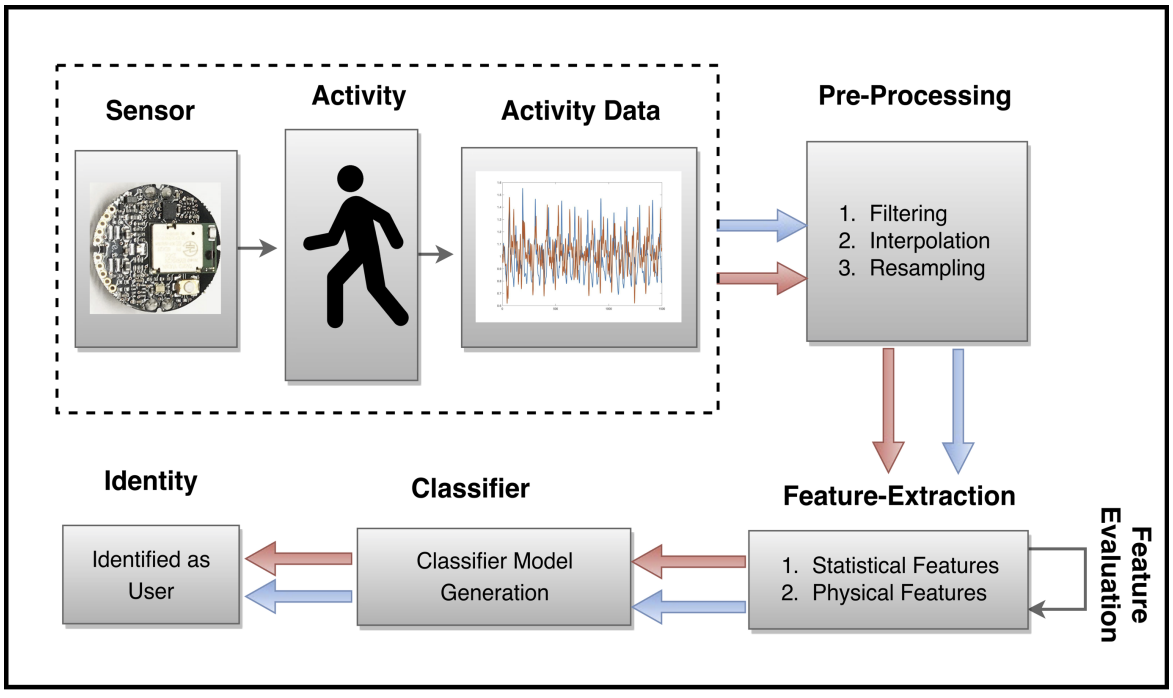


Figure 3.1. The ActID Framework

In the Figure 3.1, the *identity modeling phase* is shown by path of blue arrows and the *identification phase* is depicted by the path of red arrows. In the first phase, when the user walks around, the changes in motion are captured by activity sensor consisting of accelerometer and gyroscope, which is placed on the wrist of the user. The sensing data is then transferred to a smart device via a Bluetooth channel. Next, the received data is filtered, resampled and interpolated to improve the quality of the data. A set of features are extracted from the processed data. These features consist of both statistical attributes such as mean, standard deviation, and variance, as well as physical attributes like peak value for acceleration of hand. With extracted features, we establish a classifier as the identity model using a sliding-window based algorithm.

In the second phase, user activity data is collected as in the first phase. Then, the collected user data is given as input to the classifier and classifier identifies the user. In addition to the two phases, the ActID framework consist of a feature evaluation algorithm, which is used to select a set of high-quality features.

3.2 Data Acquisition

3.2.1 MetaWearC Board

In our experiment, we collect activity data from 14 users in two sessions. In each session, the users walk as they usually walk on a plain surface for 60 seconds. We use MetaWear C board as shown in Figure 3.2 to collect the activity data. The MetaWear C board comes in a very small round form-factor equipped with two sensors including an accelerometer and a gyroscope. There is an LED and push button switch on the board, which is powered by a CR2032 coin-cell battery [56]. The sensor is placed on the wrist of the user as shown in Figure 3.3. The sensor captures the hand movement of users as they walk. Sensor readings consist of both accelerometer and gyroscope readings along x, y, and z-axes. Therefore, each data point is a 6-tuple, (Ax, Ay, Az, Gx, Gy, Gz), where A_i

and G_i specify the accelerometer and gyroscope on the i axis respectively. Each session collects 60 seconds of data sampled at a frequency of 100Hz. Table 3.1 shows the specifications of the accelerometer and gyroscope sensors of MetaWearC board.

Table 3.1

Key specifications of MetaWearC Sensor

Sampling Rate	100 Hz
Accelerometer (m/s^2)	± 8 G's
Gyroscope (rad/s)	± 500 dps

Among the two sessions, the data collected in the first session is used to construct the classifier as illustrated in the phase one. The data collected in the second session will be used to test the classifier in the phase two. The output of the classifier will be the identity of the user.

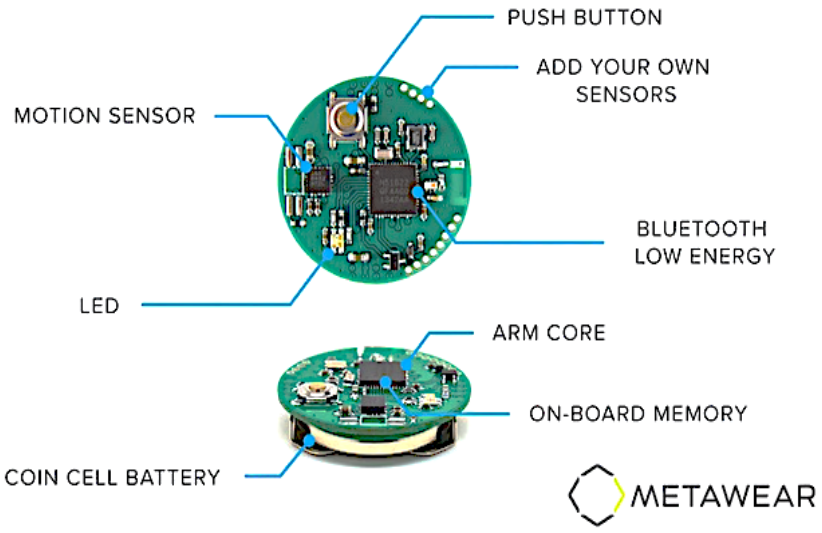


Figure 3.2. Features of MetaWearC Board

Figure 3.3. Location of Sensor

Figure 3.4 displays the sample of accelerometer and gyroscope data of a user in X, Y and Z dimensions for two sessions (S1and S2). Several studies [3, 18] have used a combined signal of all three dimensions by using a vector summation method. These approaches have the advantage of reducing computation time by reducing dimensions; however, if the amplitude of the signal in a particular dimension is much higher than others, dimensions with smaller amplitude signal become ignored. In our study, we use data in all three dimensions separately for feature computation and comparison, because this strategy helps identify high-quality features.

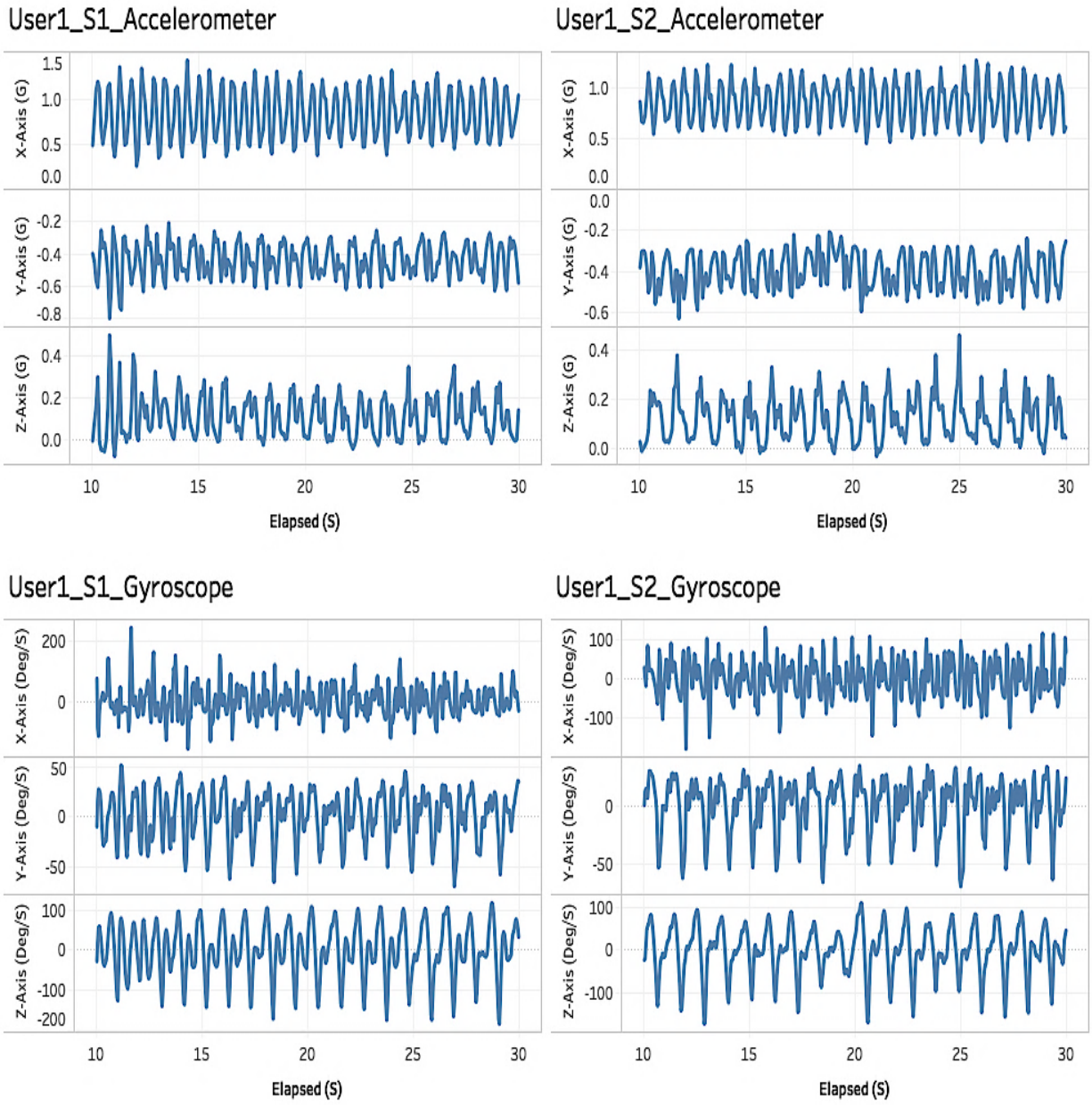


Figure 3.4. Sample of Accelerometer and Gyroscope readings for two sessions

3.2.2 MetaBase App

We use MetaBase App, developed by Mbientlab Inc. to configure MetaWearC Board and retrieve sensor data in logging or streaming mode via the Bluetooth channel. MetaBase is freely available on Windows, iOS, and Android App stores. The sensor data can be sent to PCs, tablets, smartphones, and laptops. Figure 3.5 shows the sensor configuration in MetaBase App for data collection.

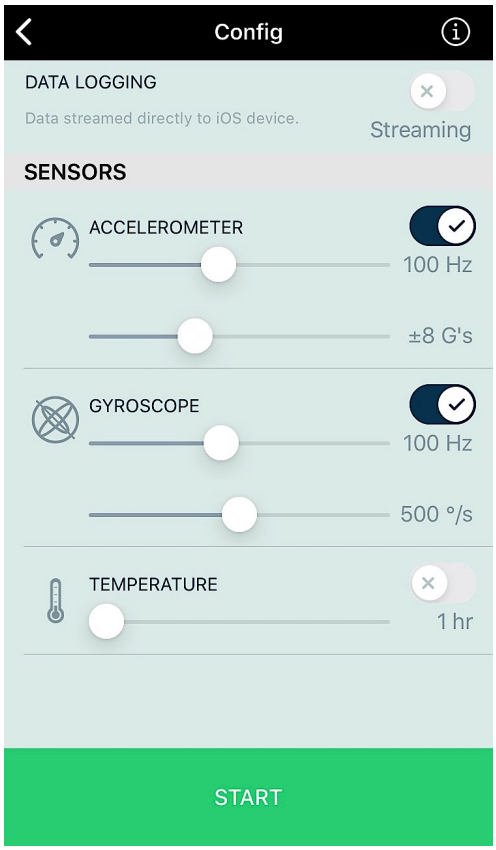


Figure 3.5. Sensor Configuration

3.3 Data Pre-Processing

The sensor data collected through MetaBase app is transferred to a laptop for data pre-processing. Figure 3.6 presents the sample consisting of 1 second of raw data. The raw data is then cleaned and calibrated.

epoc (ms)	timestamp (-0500)	elapsed (s)	x-axis (g)	y-axis (g)	z-axis (g)
1563477690259	2019-07-18T14.21.30.259	0.000	-0.956	0.432	0.082
1563477690269	2019-07-18T14.21.30.269	0.010	-0.968	0.441	0.081
1563477690279	2019-07-18T14.21.30.279	0.020	-0.968	0.439	0.083
1563477690289	2019-07-18T14.21.30.289	0.030	-0.964	0.434	0.083
1563477690299	2019-07-18T14.21.30.299	0.040	-0.963	0.433	0.084
1563477690309	2019-07-18T14.21.30.309	0.050	-0.963	0.431	0.085
1563477690319	2019-07-18T14.21.30.319	0.060	-0.961	0.431	0.088
1563477690329	2019-07-18T14.21.30.329	0.070	-0.958	0.428	0.085
1563477690339	2019-07-18T14.21.30.339	0.080	-0.958	0.428	0.081
1563477690349	2019-07-18T14.21.30.349	0.090	-0.957	0.433	0.084
1563477690359	2019-07-18T14.21.30.359	0.100	-0.961	0.429	0.079

Figure 3.6. Sample of Raw Data

3.3.1 Interpolation and Resampling

Accelerometers record data whenever there is a change in acceleration, thus the intervals of two continuously sampled accelerometer data may vary. In order to build better classifiers as identification models, we need to resample the data and make the interval of any two neighboring accelerometer readings always be 0.01 second as our sampling frequency is 100Hz. The resampling process handles two cases. In the first case, if there is a data recorded at the interval of 0.01, the resampled data will be the same as the originally recorded data. For example, at time points of 0.19 and 0.20 second, if the recorded data is 0.1049 and 0.1193 then the resampled data for these two time points will remain same. In the second case, when there is no value recorded at a specific time point, we use linear interpolation to estimate the most possible data value at that time point. For example, based on the sampled data value of 0.1176 and 0.1137 at time points of 0.38 and 0.41, the linear interpolation estimates the data value at time points of 0.39 and 0.40

second to be 0.1163 and 0.1150 respectively. Figure 3.7 shows the scatter plot of a set of sampled data before and after interpolation. In the figure, the blue dots represents the data before interpolation and the orange dots depicts the data after interpolation. We can find that several missing data points have been added after interpolation. It helps building better identification models.

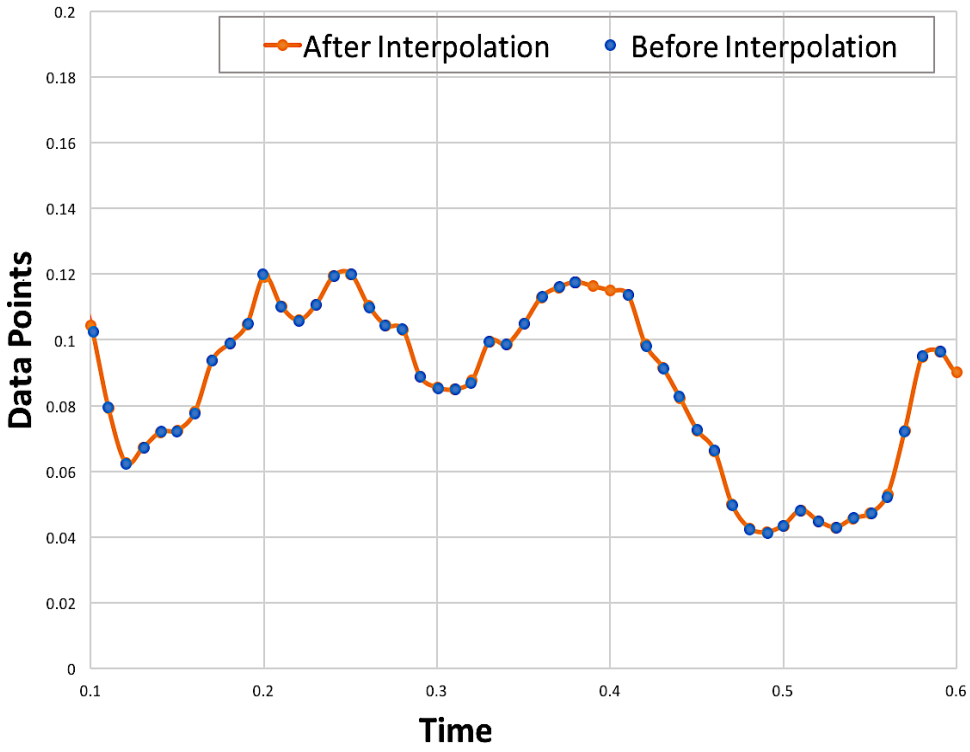


Figure 3.7. Results of Resampling and Interpolation

3.3.2 Filtering

Since the raw data may contain noise, we applied a smoothing filter to remove any high-frequency abnormal value. We adopt the moving average filter for this purpose. This filter takes seven data points at a time, finds the average of these points, and replaces the central data point with the average value. This approach effectively removes outliers in the sampled data set.

3.3.3 Data Selection

To better satisfy the real-time requirements in identification applications, we try to build identity models using a small set of sensing data as the training set. Reducing the amount of training data not only reduces the time needed to collect more data, but also decreases the complexity of the identification algorithm; however, a bigger training dataset usually improves the accuracy of the identification algorithms. Balancing the accuracy and the efficiency in identification, in this paper, we pick 1500 samples (15 seconds of data with 100Hz sampling frequency) from the processed dataset of 6000 samples (60 seconds of data) for our further analysis. Because the first few and last few data points may contain more noise, we eliminate the first and the last 2250 data points in the dataset and select 1500 data points.

3.4 Rationale of Feature Evaluation

Identifying unique features from the biometric dataset is one of the most important steps of designing any biometrics-based identification algorithm [10]. On one hand, the quality of selected features mainly decides the accuracy of identification results. In order to achieve highly accurate user identification, we will need high-quality features that can distinctly differentiate any two users. The difference between a user and others is not significant when they are compared using a poor feature. On the other hand, the number of features impact the efficiency of the identification algorithm. Usually, a single feature may not be sufficient to identify a person. Most previous studies use a vector of features in their algorithms [16, 21], but increasing the number of features increases both the dimension and the size of the dataset, which complicates the identification algorithm. Thus, we aim to find a minimum set of high-quality features by eliminating poor features based on the feature evaluation results.

3.4.1 Normalized Root Mean Squared Difference (NRMSD)

We use Normalized Root Mean Squared Difference (NRMSD) [27] to measure the quality of each identified feature when it is used for user identification purposes. The NRMSD is calculated based on two sets (set A and set B) of feature data, each representing a user. In this paper, we use the session 1 data from user 1 as set A and the session 2 data from user 2 as set B. NRMSD is defined as in Equation1, where a_i is the i^{th} data point in set A and b_i is the i^{th} data point of in set B, and w is window size the number of data points in both sets.

$$NRMSD = \frac{\sqrt{\sum_{i=1}^{w}(a_i - b_i)^2}}{\sqrt{\sum_{i=1}^{w} a_i{}^2} + \sqrt{\sum_{i=1}^{w} b_i{}^2}} \quad (1)$$

NRMSD is similar to Euclidian distance but is normalized by magnitude. The value of NRMSD ranges from zero to one. If the two datasets are exactly the same, then the NRMSD value will be zero. Otherwise, the NRMSD value increases with the difference between two datasets. Calculating the NRMSD for every pair of users, we can build a 2-dimension $n \times n$ NRMSD matrix for n users. The diagonal elements (element *i, i*) of the matrix indicate the NRMSD of the two sessions of each user. The values of diagonal elements are expected to be low for high quality features. Non-diagonal elements *(i, j, where i≠j)* in the matrix represent the NRMSD values of user *i* and user *j*. The values of these elements are expected to be high for high quality features.

Since NRMSD is normalized by the magnitude of the features, it is very sensitive to changes in amplitude and phase. Thus, two factors are important to consider when we use NRMSD to evaluate the feature, including the alignment of the data points in two datasets and the number of data points in the dataset used in NRMSD calculation. The first consideration finds a best match between two datasets, and the second guarantees there is enough data to be used in feature evaluation. For example, although a small part

of the data may be enough to evaluate a feature, it is important to have at least a full cycle of body motion is recorded.

In this paper, we use a sliding window algorithm in order to perform operations such as construction of NRMSD matrix on specific window size of our large dataset. We use a window of 500 data points which corresponds to 5 seconds of motion data. The objective is to find the minimum NRMSD in all the windows. The details of the NRMSD matrix construction algorithm for n users is depicted below.

Algorithm 1 Construction of NRMSD Matrix

```
1: procedure CALCULATE _NRMSD
2:   for i=1 to i=NumberOfUsers in dataset A do
3:     for j=1 to j=NumberOfUsers in dataset B do
4:       for k=1 to k=NumberOfWindows in dataset B do
5:           Step 1: → Find the best match between dataset A and dataset B based on
6:                     Euclidean distance.
7:           Step 2: → Compute the evaluating features for both dataset A and dataset B.
8:           Step 3: → Compute the NRMSD using equation 1.
9:           Step 4: → Select the minimum NRMSD in all the windows
10:      end for
11:    end for
12:  end for
13: end procedure
```

In the above algorithm, Step 1 tries to find a match between two windows of data points in dataset A and dataset B. This is accomplished by searching the minimum Euclidian distance between two windows of data. In Step 2, the evaluating feature is computed based on the matched windows. Computed features may contain various number of values as per the definition of different features. For a better statistical stability of NRMSD, Step 2 can be further segmented into different segments in the calculation. In Step 3, we simply compute the NRMSD value for each window. Finally, in Step 4, we select the minimum NRMSD in all the windows. This computation results in a two-dimensional matrix of NRMSD as defined by the following equation.

$$NRMSD_{u,v;1\le u\le n;1\le v\le n} = min\left(\frac{\sqrt{\sum_{i=1}^{w}(a_{ui}-b_{vj})^2}}{\sqrt{\sum_{i=1}^{w} a_{ui}{}^2}+\sqrt{\sum_{j=1}^{w} b_{vj}{}^2}}\right) \quad (2)$$

where,

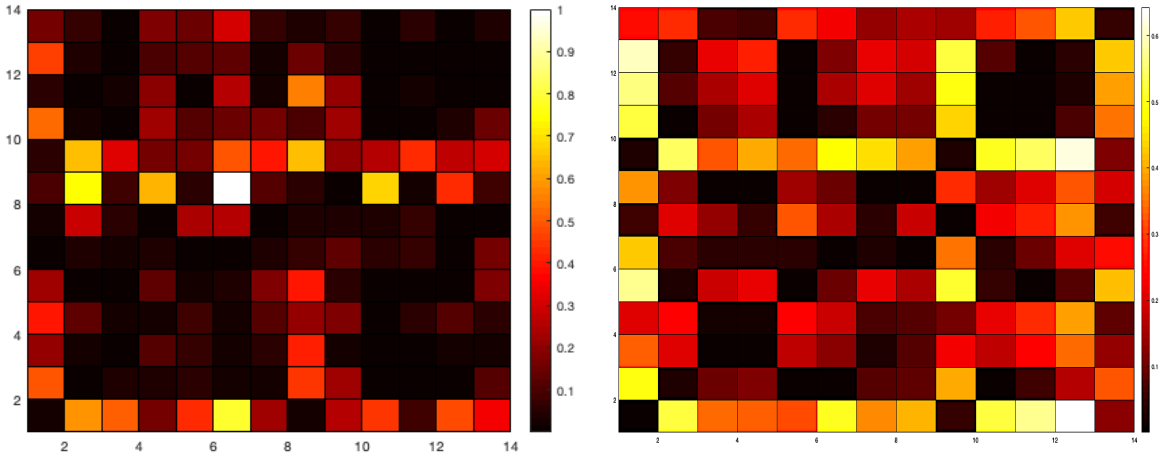
u, v : *User u and User v*

w : *window size*

i, j : data points

n : number of users

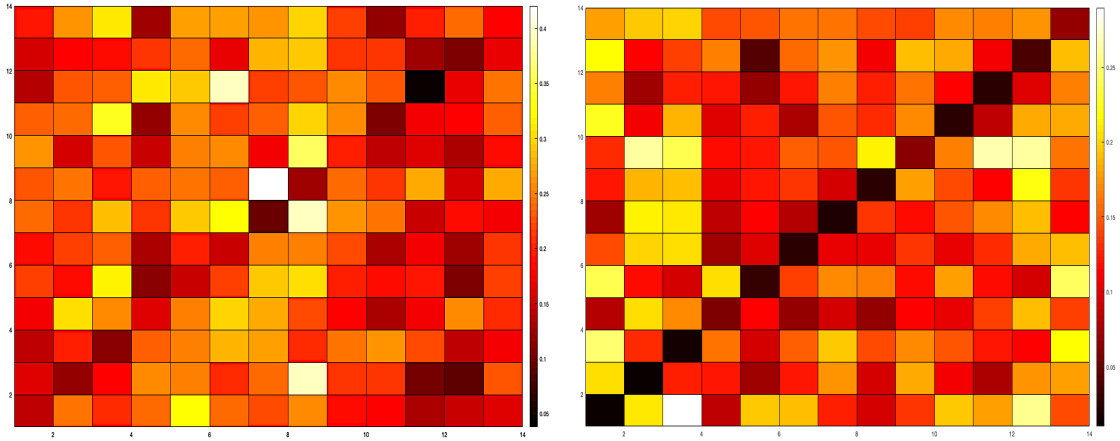
Figure 3.8 illustrates an example of visualized NRMSD matrix of 14 users for two features, the Mean (Figure 3.8(a)) and the Standard Deviation (Figure 3.8(b)) on the z-axis data of the gyroscope. In the figure, each square represents an entry in the NRMSD matrix. The diagonal squares represent the NRMSD value of the same user, and other squares represent the NRMSD value of different users. The darker color represents a relatively smaller NRMSD value, which indicates a high level of similarity, while the lighter color represents a relatively bigger NRMSD value, which indicates a low level of similarity. Therefore, a high-quality feature will result in a darker color for diagonal squares in comparison to a lighter color for other squares in the figure.



(a) (b)

Figure 3.8. NRMSD map of (a)Mean on gz-axis and (b)Standard Deviation on gz-axis

Examining the two sub-figures, Figure 3.8(a) and Figure 3.8(b), we have two observations. First, if the color of other squares in the figure is very similar to the diagonal squares, the feature may not be good enough to differentiate users. Therefore, Standard Deviation(b) is a better feature than Mean(a). Second, a single feature is usually not sufficient to distinguish other users from one specific user, because other users may exhibit a high level of similarity on that feature. For example, observing from the figure, although Standard Deviation is a good feature, user 1 and user 9 exhibit a high level of similarity on Standard Deviation. For a better identification result, we need to consider multiple features together to distinctly differentiate users.



(a) (b)

Figure 3.9. NRMSD map of combined features: (a)Kurtosis+Skewness (b)Mean+Std

Figure 3.9 shows the results of two NRMSD matrix when features are aggregated together, the combination of Kurtosis and Skewness (Figure 3.9(a)) and the combination of Mean and Standard Deviation (Figure 3.9(b)). From both figures, Figure 3.8 and Figure 3.9, we have following observations. First, even though Mean alone is not a better feature compared to Standard Deviation in Figure 3.8, when they are aggregated together, they make a better feature in Figure 3.9(b). Second, some features are not good

individually as well as when combined with other features, for instance the combination of Kurtosis and Skewness in Fig. 3.9(a). Hence, we need to carefully evaluate which features are high-quality features and discard the poor-quality features.

Observation Summary from NRMSD Maps

- *Observation 1*: A high-quality feature results in a darker color for diagonal elements in comparison to lighter color for other elements. E.g., Figure 3.8(b)
- *Observation 2*: If the color of other elements is close to or darker than that of the diagonal elements, the feature may not be good for user identification. E.g., Figure 3.8(a)
- *Observation 3***:** Aggregating two or more features may yield better results in most cases. E.g., Figure 3.9(b)
- *Observation 4***:** If a feature is of significant low-quality, aggregating it with other features may not yield better result. E.g., Figure 3.9(a)

3.5 Feature Evaluation

The NRMSD matrix figures mentioned in the above section, provide us an intuitive way of finding high-quality features through visualization. However, we still need to define a measure that can quantitatively determine if a feature is a high-quality feature. Here, we define two heuristics, Farness Value and Farness Ratio, as the measures to evaluate features based on NRMSD matrix.

3.5.1 Farness Value (FV)

The first approach, Farness Value measures the average of the differences between other elements (other users) and the diagonal elements (same user) in the NRMSD matrix. The formula to calculate the Farness Value is given below.

$$Farness\ Value = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|c_{ij} - c_{ii}|}{(n-1)*n} \quad (3)$$

where,

i : a candidate user (fixed)

j : other users (variable)

n : the number of users (i.e., n = 14)

C_{ii} : the NRMSD value of the same user, calculated based on two sessions of data for any one user

C_{ij} : the NRMSD value of user i and user j, calculated based on user i's 1st session data & user j's 2nd session data

The Farness Value is calculated as the difference between the NRMSD value calculated based on a user's two sessions of dataset and the average of NRMSD value calculated based on the user's dataset A and other user's dataset B. The Farness Value is a good heuristic because it considers both the average NRMSD value of two sessions of the same user and the average NRMSD value of different users. Algorithm 2 gives details on how to calculate the Farness Value in four steps. Firstly, the two-dimensional NRMSD matrix is computed based on the above equation. Then, we make each element subtract the diagonal element at the same column in Step 2 and 3. Finally, we calculate the Farness Value as the average of the non-diagonal elements. It ranges from 0 to 1.The bigger the Farness Value, the higher the quality of the feature, because a bigger Farness Value indicates significant difference between a user and other users.

Algorithm 2 Computing Farness Value

```
1: procedure CALCULATE _FARNESS_VALUE
2:         Step 1: → Compute the two-dimensional (n*n) matrix 1 using equation 2.
3:         Step 2: → Define a two-dimensional (n*n) matrix 2 by setting the value of
4:                   elements at each column to the value of the diagonal element at that
5:                   column.
6:         Step 3: → Subtract matrix 2 from matrix 1.
7:         Step 4: → Sum all the elements in the updated matrix from Step 3 and divide by
8:                   (n-1)*n.
9: end procedure
```

3.5.2 Farness Ratio (FR)

The second approach, Farness Ratio is defined as the average of the geometric average of column element to diagonal element ratio and the geometric average of row element to diagonal element ratio. The formula to calculate Farness Ratio is as follows:

$$Farness\ Ratio = \frac{\sqrt[n-1]{\prod_{j=1}^{n} \frac{C_{ij,\, i\neq j, 1\leq i\leq n}}{C_{ii}}} + \sqrt[n-1]{\prod_{j=1}^{n} \frac{C_{ji,\, i\neq j, 1\leq i\leq n}}{C_{ii}}}}{2} \qquad (4)$$

$$= \frac{\sqrt[n-1]{\prod_{j=1}^{n} C_{ij,\, i\neq j, 1\leq i\leq n}} + \sqrt[n-1]{\prod_{j=1}^{n} C_{ji,\, i\neq j, 1\leq i\leq n}}}{2 * C_{ii}}$$

where,

- i : a candidate user (fixed)
- j : other users (variable)
- n : the number of users (i.e., n = 14)
- C_{ij} : the NRMSD value of user i and user j, calculated based on user i's 1st session data & user j's 2nd session data
- C_{ii} : the NRMSD value of the same user, calculated based on two sessions of data for any one user

Algorithm 3 presents the detailed steps to calculate Farness Ratio. The first step is similar to Algorithm 2. In Step 2, we take the ratio of sum of all non-diagonal elements in the NRMSD matrix. Then, we divide each value from Step 2 by 2(n-1)*diagonal element. In Step 3, we sum up all the non-diagonal values from Step2. Similarly, in Step 4, we sum up all the diagonal values. Finally, we take the Farness Ratio by dividing the total sum of non-diagonal value in Step 4 by the total sum of diagonal value in Step 5. Likewise, the bigger the Farness Ratio, the higher the quality of the feature.

Algorithm 3 Computing Farness Ratio

procedure CALCULATE _FARNESS_RATIO
    Step 1: → Compute the two-dimensional (n*n) matrix 1 using NRMSD equation.
    Step 2: → For each cell in a column (C_{ij}), divide by C_{ii}.
    Step 3: → Take $(n-1)^{th}$ root of Step 2.
    Step 4: → Take the average of all the values from Step 3.
    Step 5: → For each cell in a column (C_{ji}), divide by C_{ii}.
    Step 6: → Take $(n-1)^{th}$ root of Step 5.
    Step 7: → Take the average of all the values from Step 6.
    Step 8: → Take the average of the values from Step 4 and Step 7.
end procedure

3.6 Feature Selection

Extracting high-quality features from the raw dataset not only reduces the dimension of the data, but also helps prepare a set of more meaningful data. Based on the results of our feature evaluation method, we select a set of high-quality features for user identification. We use two approaches for Feature Selection, compare the results and select the most appropriate one.

3.6.1 Based on the special values of Farness Value and Farness Ratio

First, we observe the range of Farness Value of all the extracted features. Then, we set the threshold to FV and analyze which features meet the criteria. Second, we follow the same procedure for Farness Ratio and set the threshold to FR. Finally, we

consider both FV and FR together, and select features that meet the threshold for both FV and FR as high quality features because the bigger FV and FR implies significant difference between two users. Our goal here is to attain a smaller set of high-quality features. Hence, when we consider both measures together, we will achieve more efficient results.

3.6.2 Based on ranking of Farness Value and Farness Ratio

For this approach, first we rank FV of all the extracted features from highest to lowest. Similarly, we rank FR for all the extracted features and calculate the average between FV and FR ranks for each feature. Then, we rank the value of averages and arrange the ranks in ascending order. Based on the ranking of both FV and FR, we select the top 12 features from all the extracted features and compare the ranking result with the above-mentioned feature selection approach.

3.7 Classification Algorithm

In order to build a classifier, we use the users' session 1 dataset as our training dataset and users' session 2 dataset as our test dataset. Our labels correspond to the number of users we have, i.e., User ID from 1 to 14. Training dataset reveals the "ground truth" as it contains the correct labels for given data points. On the contrary, to see how accurately a classifier can identify the user, we remove the labels from the test dataset. We use the training dataset to train our classifier model. The trained model is then used for the classification task or prediction of users on the test dataset.

3.8 Evaluation Metrics

In order to evaluate the effectiveness of our proposed method, we use four metrics widely used in user identification studies [49,50].

- *False Accept Rate (FAR)* is the probability that the identity verification system incorrectly identifies the imposter as the genuine user.

- *False Reject Rate (FRR)* is the probability that the identity verification system incorrectly rejects the genuine user.
- *Equal Error Rate (EER)* is the rate at which both FAR and FRR are equal. The lower the value of ERR, the higher is the accuracy of the biometric system.
- *Accuracy* (also known as *True Positive Rate, TPR*) is a proportion if all recognition attempts where users are correctly identified.

CHAPTER IV:
EXPERIMENTAL RESULTS

4.1 Feature Evaluation Results

Based on our literature search [4, 15, 33, 37, 49, 50, 60] , we considered 12 different features (details are listed in Table 4.1) that are mostly used in biometrics identification for our feature evaluation and selection approach. Our activity sensor readings include both accelerometer and gyroscope readings along x, y, and z-axes consisting of 6-tuples (Ax, Ay, Az, Gx, Gy, Gz). We extracted all together 72 features (i.e., 12 features X 6-tuples), comprising 36 features from the accelerometer readings and 36 features from the gyroscope readings. Subsequently, we evaluated each of the 72 features individually with our feature evaluation approaches based on NRMSD computation, Farness Value, and Farness Ratio.

Our features can be classified into two categories: time-domain features and frequency-domain features. Time-domain refers to the analysis of mathematical or statistical functions, physical signals or time series data, with respect to time. Examples of time-domain features include mean, median, standard deviation, root mean square (rms), variance, kurtosis, skewness, peak to peak, and peak to rms. On the other hand, the frequency domain refers to the analysis of mathematical functions or signals with respect to frequency instead of time [58]. Examples of frequency-domain features include mean frequency, median frequency, and fast fourier transform (fft).

Table 4.1

A list of extracted features from both accelerometer and gyroscope sensor

No.	Feature	Description	Formula
1.	Mean	Mean is a measure of central tendency calculated as the normalized sum of all the elements.	$Mean\ (\mu) = \frac{\sqrt{\sum_{i=1}^{n} a_i}}{n}$ Where, a = ax, ay, az, gx, gy, gz; n = the total number of data points
2.	Standard Deviation (Std)	Standard deviation is a measure that is used to quantify the amount of variation or dispersion of a set of data values.	$Std\ (\sigma) = \frac{\sqrt{\sum_{i=1}^{n}(a_i - \mu)^2}}{n}$
3.	Root Mean Square (RMS)	RMS is defined as the square root of the mean square.	$RMS = \sqrt{\frac{\sum_{i=1}^{n}(a_i)^2}{n}}$
4.	Mean Frequency (MNF)	The mean frequency of a spectrum is calculated as the sum of the product of the spectrogram intensity and the frequency, divided by the total sum of spectrogram intensity.	$MNF = \frac{\sum_{i=1}^{M} f_i P_i}{\sum_{i=1}^{M} P_i}$ Where, f_i = *frequency of spectrum,* P_i = ith line of power spectral density, and M = length of the frequencies
5.	Median	The median is the value separating the upper half of a sample data set from the lower half. In our case, n is even (i.e. n=1500)	$Median = \frac{\left(\frac{n}{2}\right)^{th} term + \left(\frac{n+1}{2}\right)^{th} term}{2}$
6.	Median Frequency (MDF)	Median Frequency is a frequency that divides the power spectrum into two regions with equal amplitude.	$MDF = \frac{1}{2}\sum_{i=1}^{M} P_i$
7.	Kurtosis	Kurtosis measures the tail-heaviness of the distribution.	$Kurtosis = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{(a_i - \mu}{\sigma}\right]^4$
8.	Skewness	Skewness is a measure of the symmetry in a distribution.	$Skewness = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{(a_i - \mu}{\sigma}\right]^3$
9.	Peak to Peak	Peak to Peak computes the maximum minus minimum value in the time series.	$Peak\ to\ Peak = Max(a_i) - Min(a_i)$

No.	Feature	Description	Formula
10.	Peak to RMS	Peak to RMS results the ratio of maximum amplitude in the time series to RMS of the time series.	$Peak\ to\ RMS = \frac{(Max(\lvert a_i \rvert))}{RMS}$
11.	Variance	Variance is the average of the squared distances from each point to the mean.	$\sigma^2 = \frac{\sum_{i=1}^{n}(a_i - \mu)^2}{n}$
12.	Fast Fourier Transform (FFT)	FFT converts time domain vector signal to frequency domain vector signal.	$X(k) = \sum_{n=0}^{N-1} X(n) * e^{-i*2\pi*n*k/N}$ Where, X(n) = n^{th} input sample (n=0..N-1) X(k) = Freq. domain samples N = FFT size k = 0,1,2,…….N-1

4.2 Feature Selection Results Based on Specific Values of FV and FR

Based on our feature evaluation method, we select features that have a FV>=0.15 and FR>=15 as high-quality features because the bigger FV and FR implies significant difference between two users. Appendix A shows the complete list of the Farness Value and Farness Ratio for 72 extracted features calculated based on the two sessions of dataset. The highlighted part on Appendix A are the features that met the threshold of FV>= 0.15 and FR>=15. We select these specific threshold values for FV and FR because based on our calculation and analysis, these values appear to be the most preferable. As a result, we achieve 18 features out of total 72 features for classification. Table 4.2 displays the summary of the selected and discarded features for our study. The features that did not meet the threshold for any dimensions are Kurtosis, Peak to Peak, Peak to RMS, and FFT. We also observed that the selected features only included some x and z-axes and discarded all y-axis dimensions.

Table 4.2

Summary of selected and discarded features

Selected Features	**Discarded Features**
Mean (ax, az)	Mean (ay, gx, gy, gz)
Std (gx, gz)	Std (ax, ay, az, gy)
RMS (az, gx, gz)	RMS (ax, ay, gy)
MNF (ax, az)	MNF (ay, gx, gy, gz)
Median (ax, az, gx)	Median (ay, gy, gz)
MDF (ax, az)	MDF (ay, gx, gy, gz)
Skewness (gz)	Skewness (ax, ay, az, gx, gy)
Variance (az, gx, gz)	Variance (ax, ay, gy)
	Kurtosis (ax, ay, az, gx, gy, gz)
	Peak to Peak (ax, ay, az, gx, gy, gz)
	Peak to RMS (ax, ay, az, gx, gy, gz)
	FFT (ax, ay, az, gx, gy, gz)

Figure 4.2 shows the graphs representing the comparison of FV and FR in selected features (e.g., Mean and Root Mean Square). High-quality features seem to have similar trend for both FV and FR. On the contrary, Figure 4.2 shows the graph representing the same comparison in discarded features (e.g., Peak to Peak and Peak to RMS). Low-quality features seem to have an inconsistent trend for FV and FR.

Figure 4.1. Comparison of FV and FR in selected features (Mean and RMS)

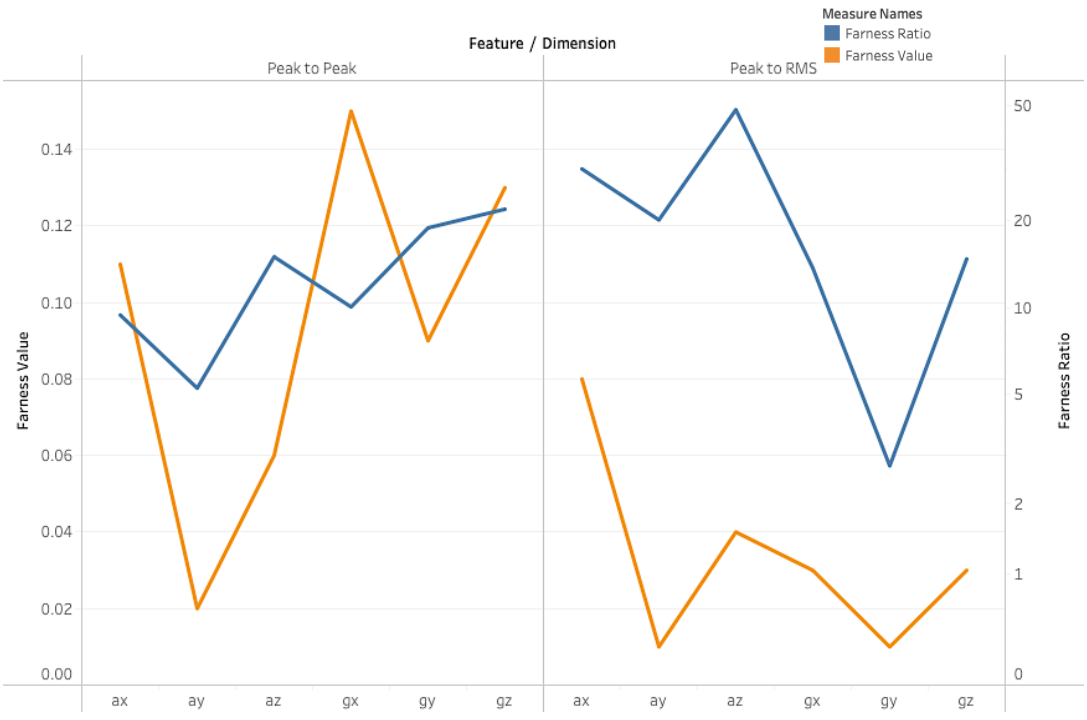


Figure 4.2. Comparison of FV and FR in discarded features (Peak to Peak and Peak to RMS)

From the figures, we observe the correlation between Farness Value and Farness Ratio for the given features. Higher the Farness Value and the Farness Ratio, higher is the quality of the feature because higher value for Farness Value and Farness Ratio means there is significant difference between a user and other users. As a result, the classification result will be better.

4.2 Feature Selection Results Based on Ranking

For all 72 extracted features, we first ranked them from highest to lowest based on FV and FR separately. Then, we ranked them again considering both FV and FR and get the top 12 out of 72 features. We also ranked the 18 selected features that we achieved from the previous feature selection method and selected the top 12 features. Next, we compared the two lists of top 12 features that we achieved from our two feature selection approaches in Table 4.3.

Table 4.3

Comparison of top 12 features from our two feature selection approaches

Rank	**Top 12 features based on specific values of FV and FR**	**Top 12 features based on ranking of FV and FR**
1	Variance_gx	Variance_gx
2	Median Frequency_ax	Median Frequency_ax
3	Mean_az	Mean_az
4	Mean Frequency_ax	Mean Frequency_ax
5	Skewness_gz	Skewness_gz
6	Mean Frequency_az	Mean Frequency_az
7	Median_az	Mean_ax
8	Mean_ax	Median_az
9	Standard Deviation_gx	Median_ax
10	Variance_gz	Standard Deviation_gx
11	Median_ax	Root Mean Square_gx
12	Root Mean Square_az	Root Mean Square_az

From the comparison, we determine that our two feature selection approaches are similar. Most of the features have same rank. Only few are shifted up and down by few ranks. For example, Median_az and Mean_ax are on 7th and 8th position for the first approach, whereas they switched their positions for the second approach. Hence, our second feature selection approach verifies that the thresholds that we selected for FV and FR on the first approach are reasonable and the 18 selected features are qualified for classification.

4.3 Classification Result

For classification, we used *WEKA* (Waikato Environment for Knowledge Analysis), a free software consisting of a collection of machine learning algorithms for data mining tasks. First, we trained the model using users' session 1 dataset. Then, we applied the trained model to users' session 2 dataset as our test data. However, we did not get good accuracy with this approach. Therefore, we split each session dataset into training and test dataset with 70:30 ratio and calculated the mean accuracy. Finally, we were able to get some good accuracy. We computed the accuracy for different classification algorithms and selected the best algorithm for classification. From the evaluation, K-Nearest Neighbor (KNN) gave us the best and consistent results. A comparison between different classification algorithms can be found in Table 4.4.

Table 4.4

Comparison of different classification algorithms

No.	Classification Method	Mean Accuracy
1	K-Nearest Neighbor (KNN)	**98.3%**
2	Naïve Bayes	97.8%
3	Random Forest	97.2%
4	Support Vector Machine (SVM)	91%

4.4 Comparison of our approach versus other approaches

We compared our proposed feature evaluation and selection algorithm with two previous efforts by Kumar et al. [15] and Damaševičius et al. [49, 50]. In [15], a total of 76 features (32 features from the accelerometer readings and 44 features from the gyroscope readings) are evaluated and the feature evaluation approach discards about 25% of the features. The authors in [49, 50] have extracted 99 features from the collected data based on the extensive analysis of the literature and features used by other authors. They used Rankfeatures selection method to select the top three and top ten features respectively. Table 4.5 shows the comparison of our approach with the above two approaches. Our approach discards more quality features than [15] and keeps more high-quality features than [49, 50]. With our approach, we discarded 75% of the total features and selected the remaining 25% for classification.

Table 4.5

Comparison of different classification algorithms

Paper	**Features**	**Classification method**	**Accuracy**
[15]	76 features (32 from accelerometer, 44 from gyroscope)	K-NN	95%
[49, 50]	99 time, frequency and physical features	Heuristic (random projections + PDFs + Jaccard distance)	95.52%
Our approach	72 features (time and frequency-domain)	K-NN	**98.3%**

CHAPTER V:
CONCLUSION AND FUTURE WORK

5.1 Conclusion

In this study, we proposed a novel and efficient feature evaluation and selection approach for biometrics-based user identification. We designed a lightweight identification framework based on activity sensor data collected from the users' wrists while they are walking providing a smooth user experience. Our proposed method discussed an approach for preliminary evaluation of different features before model generation. We defined three measures: NRMSD, Farness Value, and Farness Ratio to quantitatively measure the quality of features. Then, we evaluated a total of 72 features from both accelerometer and gyroscope readings in x, y and z-axes. Based on our feature evaluation, we discarded 75% of the total features, and selected the remaining 25% features as a set of high-quality features for classification. As a result, we selected 18 features for classification based on the specific values and ranking of Farness Value and the Farness Ratio. By selecting a set of high-quality features, we reduced the number of features to improve the efficiency of the algorithm. An accuracy of 98.3% was achieved with the minimum number of features. This suggests that dimension reduction is important, and a good balance can be achieved between efficiency and accuracy while designing an identification protocol. Hence, the experimental results demonstrate the effectiveness of our proposed method.

5.3 Future Work

Although the accuracy of our proposed method is high in our study, the algorithm needs to be evaluated over a number of factors such as number of data points, classification method. Due to the time limitation, we had a small sample size. In the future work, our method should be verified with a larger sample size. One of the

drawbacks of this method is relatively common to all gait based methods i.e., effects of changes in the speed of walking, shoes, walking surface, with or without luggage, human factors (injuries, illness, drunkenness, tiredness, and laziness) affect gait and the hand movement while walking [1]. Thus, these issues need to be further investigated. This study focuses on single feature individually. In future, the analysis on combination of multiple features should be considered as we observed the great potential in combining features from our NRMSD maps . Other future work may include applying this method to evaluate more features and test the feature selection results with other classification algorithms to improve the authentication performance. The efficiency of this method also needs to be verified against spoofing. In addition, our method can be added as an additional level of security in a multi-modal system.

REFERENCES

[1] H. J. Ailisto, M. Lindholm, J. Mantyjarvi, E. Vildjiounaite, and S.-M. Makela, “Identifying people from gait pattern with accelerometers,” in Proc. SPIE, vol. 5779, 2005, pp. 7–14.

[2] J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S.-M. Mäkelä, and H. Ailisto, "Identifying users of portable devices from gait pattern with accelerometers," IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, pp. ii/973 - ii/976, 2005.

[3] D. Gafurov, K. Helkala, and T. Søndrol, “Biometric gait authentication using accelerometer sensor.” JCP, vol. 1, no. 7, pp. 51–59, 2006.

[4] C. Nickel, T. Wirtl, and C. Busch, “Authentication of smartphone users based on the way they walk using k-nn algorithm,” in Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on. IEEE, 2012, pp. 16–20.

[5] C.-L. Lin and T. Hwang, “A password authentication scheme with secure password updating,” Computers & Security, vol. 22, no. 1, pp. 68–72, 2003.

[6] R. C. Merkle, “A digital signature based on a conventional encryption function,” in Conference on the Theory and Application of Cryptographic Techniques. Springer, 1987, pp. 369–378.

[7] G. E. Suh and S. Devadas, “Physical unclonable functions for device authentication and secret key generation,” in Proceedings of the 44th annual design automation conference. ACM, 2007, pp. 9–14.

[8] W. Tan, J. Hsu, and F. Pinn, “Method and system for token-based authentication,” Feb. 23, 2001, US Patent App. 09/792,785.

[9] M. Alizadeh, S. Abolfazli, M. Zamani, S. Baharun, and K. Sakurai, "Authentication in mobile cloud computing: A survey." Journal of Network and Computer Applications, 61 (2016): 59-80.

[10] M. O. Derawi, C. Nickel, P. Bours, and C. Busch, “Unobtrusive user-authentication on mobile phones using biometric gait recognition,” in Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on. IEEE, 2010, pp. 306–311.

[11] K. Sha and M. Kumari, “Patient identification based on wrist activity data,” in 2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). IEEE, 2018, pp. 29–30.

[12] J. P. Gupta, N. Singh, P. Dixit, V. B. Semwal, and S. R. Dubey, "Human activity recognition using gait pattern," International Journal of Computer Vision and Image Processing (IJCVIP), vol. 3, no. 3, pp. 31-53, 2013.

[13] F. Abate, M. Nappi, and S. Ricciardi, “I-am: Implicitly authenticate me person authentication on mobile devices through ear shape and arm gesture,” IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2017.

[14] E. Vural, S. Simske, and S. Schuckers, “Verification of individuals from accelerometer measures of cardiac chest movements,” in Biometrics Special Interest Group (BIOSIG), 2013 International Conference of the. IEEE, 2013, pp. 1–8.

[15] R. Kumar, V. V. Phoha, and R. Raina, “Authenticating users through their arm movement patterns,”arXiv preprintarXiv:1603.02211, 2016.

[16] S. K. Al Kork, I. Gowthami, X. Savatier, T. Beyrouthy, J. A. Korbane, and S. Roshdi, “Biometric database for human gait recognition using wearable sensors and a smartphone,” in Bioengineering for Smart Technologies (BioSMART), 2017 2nd International Conference on. IEEE, 2017, pp. 1–4.

[17] J. R. Vacca, Biometric technologies and verification systems. Elsevier, 2007.

[18] W. Yang, S. Wang, J. Hu, Z. Guanglou, J. Chaudhry, E. Adi, and C. Valli, “Securing mobile healthcare data: A smart card based cancelable finger-vein bio-cryptosystem,” IEEE Access, vol. PP, pp. 1–1, 06 2018.

[19] Biometrics Institute. (2019). Types of Biometrics [Online]. Available: https://www.biometricsinstitute.org/what-is-biometrics/types-of-biometrics/

[20] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, Handbook of fingerprint recognition. Springer Science & Business Media, 2009.

[21] D. J. Ohana, L. Phillips, and L. Chen, “Preventing cell phone intrusion and theft using biometrics,” in Security and Privacy Workshops (SPW), 2013 IEEE. IEEE, 2013, pp. 173–180.

[22] R.Sanchez-Reillo,C.Sanchez-Avila,andA.Gonzalez-Marcos,“Biometric identification through hand geometry measurements, ” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 10, pp.1168–1171, Oct 2000

[23] A. Pal, A. K. Gautam, and Y. N. Singh, “Evaluation of bioelectric signals for human recognition,” Procedia Computer Science, vol. 48, pp. 746 – 752, 2015, international

Conference on Computer, Communication and Convergence (ICCC 2015). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050915007206

[24] M. Dey, N. Dey, S. K. Mahata, S. Chakraborty, S. Acharjee, and A. Das, "Electrocardiogram feature based inter-human biometric authentication system," in Electronic Systems, Signal Processing and Computing Technologies (ICESC), 2014 International Conference on. IEEE, 2014, pp. 300–304.

[25] A. S. Chatra, "Cognitive biometrics based on eeg signal," in Contemporary Computing and Informatics (IC3I), 2014 International Conference on. IEEE, 2014, pp. 374–376.

[26] Apple Inc. (2017, November). Face ID Security [Online]. Available: https://www.apple.com/ca/business-docs/FaceID_Security_Guide.pdf.

[27] R. P. Wildes, "Iris recognition: an emerging biometric technology," Proceedings of the IEEE, vol. 85, no. 9, pp. 1348–1363, 1997.

[28] M. Faundez-Zanuy, "Signature recognition state-of-the-art," IEEE Aerospace and Electronic Systems Magazine, vol. 20, no. 7, pp. 28–32, July 2005.

[29] F. Monrose and A. D. Rubin, "Keystroke dynamics as a biometric for authentication," Future Generation computer systems, vol. 16, no. 4, pp. 351–359, 2000.

[30] H. Lee, J. Y. Hwang, D. I. Kim, S. Lee, S.-H. Lee, and J. S. Shin, "Understanding keystroke dynamics for smartphone users authentication and keystroke dynamic son smartphones built-in motion sensors, "Security and Communication Networks, vol. 2018, 2018.

[31] R. Tadeusiewicz and G. Demenko, “Voice as a key,” in 2009 International Conference on Biometrics and Kansei Engineering, June 2009, pp. 28–33.

[32] J. Guerra-Casanova, C. Sánchez-Ávila, G. Bailador, and A. de Santos Sierra, “Authentication in mobile devices through hand gesture recognition,” International Journal of Information Security, vol. 11, no. 2, pp. 65–83, 2012.

[33] F. T. Garcia, K. Krombholz, R. Mayer, and E. Weippl, “Hand dynamics for behavioral user authentication,” in 2016 11th International Conference on Availability, Reliability and Security (ARES). IEEE, 2016, pp. 389–398.

[34] J. Chen, “Gait correlation analysis based human identification,” The Scientific World Journal, vol. 2014, 2014.

[35] L. Rong, Z. Jianzhong, L. Ming, and H. Xiangfeng, "A wearable acceleration sensor system for gait recognition," in 2007 2nd IEEE Conference on Industrial Electronics and Applications, 2007: IEEE, pp. 2654-2659.

[36] H. Sun and T. Yuao, "Curve aligning approach for gait authentication based on a wearable accelerometer," Physiol Meas, vol. 33, no. 6, pp. 1111-20, Jun 2012, doi: 10.1088/0967-3334/33/6/1111.

[37] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, “Cell phone based biometric identification,” in Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on. IEEE, 2010, pp. 1–7.

[38] H. M. Thang, V. Q. Viet, N. D. Thuc, and D. Choi, “Gait identification using accelerometer on mobile phone,” in 2012 International Conference on Control, Automation and Information Sciences (ICCAIS). IEEE, 2012, pp. 344–348.

[39] A. H. Johnston and G. M. Weiss, "Smartwatch-based biometric gait recognition," in 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS), 2015: IEEE, pp. 1-6.

[40] X. Su, H. Tong, P. Ji, "Activity recognition with smartphone sensors", Tsinghua Science and Technology, vol. 19, no. 3, pp. 235-249, 2014.

[41] A. Primo, V. Phoha, R. Kumar, and A. Serwadda, "Context-aware active authentication using smartphone accelerometer measurements," in CVPRW 2014, June 2014.

[42] H. Xu, J. Liu, H. Hu, and Y. Zhang, "Wearable sensor-based human activity recognition method with multi-features extracted from Hilbert-Huang transform," Sensors, vol. 16, no. 12, p. 2048, 2016.

[43] Liu, H. Luo, and C. W. Chen, “A novel authentication scheme based on acceleration data in wban,” in Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2017 IEEE/ACM International Conference on. IEEE, 2017, pp. 120–126.

[44] D. Sugimori, T. Iwamoto, and M. Matsumoto, “A study about identification of pedestrian by using 3-axis accelerometer,” in Embedded and Real-Time Computing Systems and Applications (RTCSA), 2011 IEEE 17th International Conference on, vol. 2. IEEE, 2011, pp. 134–137.

[45] D. Guan et al, "Review of Sensor-based Activity Recognition Systems," IETE Tech. Rev., vol. 28, (5), pp. 418-433, 2011. Available:

https://libproxy.uhcl.edu/login?url=https://search.proquest.com/docview/900724079?accountid=7108.

[46] J. Blasco, T. M. Chen, J. Tapiador, and P. Peris-Lopez, “A survey of wearable biometric recognition systems,” ACM Computing Surveys (CSUR), vol. 49, no. 3, p. 43, 2016.

[47] M. D. Marsico and A. Mecca, “A survey on gait recognition via wearable sensors,” ACM Comput. Surv., vol. 52, no. 4, pp. 86:1–86:39, Aug. 2019. [Online]. Available: http://doi.acm.org.libproxy.uhcl.edu/10.1145/3340293

[48] C.-H. Yang, D. Liang, and C.-C. Chang, “A novel driver identification method using wearables,” in Consumer Communications & Networking Conference (CCNC), 2016 13th IEEE Annual. IEEE, 2016, pp. 1–5.

[49] R. Damaševičius, M. Vasiljevas, J. Šalkevičius, and M. Woźniak2, "Human Activity Recognition in AAL Environments Using Random Projections," Computational and Mathematical Methods in Medicine, vol. 2016, p. 17, 2016, Art no. 4073584, doi: 10.1155/2016/4073584.

[50] R. Damaševičius, R. Maskeliūnas, A. Venčkauskas, and M. Woźniak, "Smartphone User Identity Verification Using Gait Characteristics," Symmetry, vol. 8, no. 10, 2016, doi: 10.3390/sym8100100.

[51] S. Khalid, T. Khalil, and S. Nasreen, “A survey of feature selection and feature extraction techniques in machine learning,” in 2014 Science and Information Conference. IEEE, 2014, pp. 372–378.

[52] R. Urbanowicz, M. Meeker, W. LaCava, R. Olson, and J. Moore, “Relief-based feature selection: Introduction and review,”Journal of Biomedical Informatics, vol. 85, 11 2017.

[53] J. Lever, M. Krzywinski, and N. Altman, N, “Principal component analysis,” *Nat Methods*, vol. 14, p. 641–642, 2017, doi:10.1038/nmeth.4346

[54] M. A. Hall, Correlation-based Feature Selection for Machine Learning, 1999.

[55] O. Carugo and S. Pongor, "A normalized root-mean-square distance for comparing protein three-dimensional structures." Protein science 10, no. 7 (2001): 1470-1473.

[56] Mbientlab Inc. MetaWear C Product Specification v1.0 [Online]. Available: https://mbientlab.com/documents/MetaWearC-CPRO-PS.pdf.

[57] N. Kayastha and K. Sha, “A novel and efficient approach to evaluate biometric features for user identification,” in 2019 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). IEEE, 2019

[58] C. Altın and O. Er, "Comparison of different time and frequency domain feature extraction methods on elbow gesture’s EMG," European journal of interdisciplinary studies, vol. 2, no. 3, pp. 35-44, 2016.

[59] P. Heckbert, "Fourier Transforms and the Fast Fourier Transform (FFT) algorithm," Computer Graphics, vol. 2, pp. 15-456, Feb. 1995.

[60] W. -C. Hsu, T. Sugiarto, Y. -J. Lin, F. -C. Yang, Z. -Y. Lin, C. -T. Sun, C. -L. Hsu, and K. -N. Chou,“Multiple-wearable-sensor-based gait classification and analysis in patients with neurological disorders,”Sensors,vol. 18, no. 10, p. 3397, 2018.

[61] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.

[62] MATLAB version 9.7 (R2019b), The Mathworks, Inc., Natick, Massachusetts, 2019.

APPENDIX A:

FARNESS VALUE AND FARNESS RATIO

Feature	Dimension	Farness Value (FV)	Farness Ratio (FR)
Mean	ax	0.28	46.72
Mean	ay	0.03	11.71
Mean	az	0.22	221.20
Mean	gx	0.15	6.85
Mean	gy	0.15	11.04
Mean	gz	0.10	5.15
Standard Deviation	ax	0.12	10.98
Standard Deviation	ay	0.04	35.33
Standard Deviation	az	0.08	91.60
Standard Deviation	gx	0.17	233.04
Standard Deviation	gy	0.11	8.70
Standard Deviation	gz	0.19	18.15
Root Mean Square	ax	0.14	34.84
Root Mean Square	ay	0.02	27.34
Root Mean Square	az	0.17	124.09
Root Mean Square	gx	0.18	89.75
Root Mean Square	gy	0.11	8.57
Root Mean Square	gz	0.19	15.13
Mean Frequency	ax	0.30	63.02
Mean Frequency	ay	0.09	8.90
Mean Frequency	az	0.25	75.77
Mean Frequency	gx	0.04	5.22
Mean Frequency	gy	0.03	8.14
Mean Frequency	gz	0.04	10.34
Median	ax	0.28	32.37
Median	ay	0.03	82.06
Median	az	0.21	90.92
Median	gx	0.16	31.60
Median	gy	0.24	9.15
Median	gz	0.18	10.44
Median Frequency	ax	0.31	64.18

Feature	Dimension	Farness Value (FV)	Farness Ratio (FR)
Median Frequency	ay	0.00	9.42
Median Frequency	az	0.18	16.90
Median Frequency	gx	0.05	175.19
Median Frequency	gy	0.09	75.54
Median Frequency	gz	0.06	107.02
Kurtosis	ax	0.01	4.66
Kurtosis	ay	0.04	16.74
Kurtosis	az	0.04	36.69
Kurtosis	gx	0.05	10.03
Kurtosis	gy	0.01	2.26
Kurtosis	gz	0.04	17.48
Skewness	ax	0.14	9.03
Skewness	ay	0.28	5.40
Skewness	az	0.23	3.68
Skewness	gx	0.18	8.18
Skewness	gy	0.23	7.47
Skewness	gz	0.31	42.03
Peak to Peak	ax	0.11	9.45
Peak to Peak	ay	0.02	5.25
Peak to Peak	az	0.06	15.03
Peak to Peak	gx	0.15	10.08
Peak to Peak	gy	0.09	18.93
Peak to Peak	gz	0.13	21.97
Peak to RMS	ax	0.08	30.25
Peak to RMS	ay	0.01	20.16
Peak to RMS	az	0.04	48.44
Peak to RMS	gx	0.03	13.75
Peak to RMS	gy	0.01	2.77
Peak to RMS	gz	0.03	14.78
Variance	ax	0.20	10.50
Variance	ay	0.08	34.91
Variance	az	0.15	130.71
Variance	gx	0.30	344.16
Variance	gy	0.19	8.13

Feature	Dimension	Farness Value (FV)	Farness Ratio (FR)
Variance	gz	0.33	16.45
FFT	ax	0.22	3.03
FFT	ay	0.04	2.30
FFT	az	0.20	3.73
FFT	gx	0.24	3.58
FFT	gy	0.18	2.14
FFT	gz	0.27	3.96