June 2017

Predictive Analytics in Cardiac Healthcare and 5G Cellular Networks

Dilranjan S. Wickramasuriya
University of South Florida, dswickramasuriya@gmail.com

Predictive Analytics in Cardiac Healthcare and 5G Cellular Networks

by

Dilranjan S. Wickramasuriya

ACKNOWLEDGMENTS

I would first like to express my gratitude to my supervisor Dr. Richard Gitlin for his support and guidance throughout the course of the past two years. I am also thankful to Dr. Kemal Davaslioglu for his help in one of the research projects described in this thesis. Finally, I am grateful to colleagues at the *i*nnovations in Wireless Information Networking Laboratory (*i*WINLAB) and my family for their encouragement during my time as a graduate student at the University of South Florida.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This thesis proposes the use of Machine Learning (ML) to two very distinct, yet compelling, applications – predicting cardiac arrhythmia episodes and predicting base station association in 5G networks comprising of virtual cells. In the first scenario, Support Vector Machines (SVMs) are used to classify features extracted from electrocardiogram (EKG) signals. The second problem requires a different formulation departing from traditional ML classification where the objective is to partition feature space into constituent class regions. Instead, the intention here is to identify temporal patterns in unequal-length sequences. Using Recurrent Neural Networks (RNNs), it is demonstrated that accurate predictions can be made as to the base station most likely to provide connectivity for a mobile device as it moves.

Atrial Fibrillation (AF) is a common cardiac arrhythmia affecting several million people in the United States. It is a condition in which the upper chambers of the heart are unable to contract effectively leading to inhibited blood flow to the ventricles. The stagnation of blood is one of the major risk factors for stroke. The Computers in Cardiology Challenge 2001 was organized to further research into the prediction of episodes of AF. This research revisits the problem with some modifications. Patient-specific classifiers are developed for AF prediction using a different dataset and employing shorter EKG signal epochs. SVM classification yielded an average accuracy of just above 95% in identifying EKG epochs appearing just prior to fibrillatory rhythms.

5G cellular networks were envisaged to provide enhanced data rates for mobile broadband, support low-latency communication, and enable the Internet of Things (IoT). Handovers contribute to latency as mobile devices are switched between base stations due to movements. Given that customers may not be willing to continuously share their exact locations due to privacy concerns and the establishment of a mobile network architecture with dynamically created virtual cells, this research presents a solution for proactive mobility management using RNNs. A RNN is trained to identify patterns in variable-length sequences of

iv

Received Signal Strength (RSS) values, where a mobile device is permitted to connect to more than a single base station at a time. A classification accuracy of over 98% was achieved in a simulation model that was set up emulating an urban environment.

CHAPTER 1: INTRODUCTION

The development of compact intelligent devices for long-term physiological monitoring has the potential to revolutionize the delivery of healthcare [1]. Patients may no longer have to make unnecessary hospital visits, previously unknown medical conditions might be diagnosed with precision early on and patients could potentially be provided with advanced warnings of life-threatening conditions enabling drug delivery, or other procedures, to avoid medical emergencies. This technology will be driven by data and seamless wireless connectivity, in which Machine Learning (ML) plays an integral role.

This thesis proposes the use of ML to two separate problems in different domains. The first is in predictive analytics for a type of cardiac disease. The electrocardiogram (EKG) is the primary method used to assess the proper functionality of the heart and different features of an EKG yield vital information regarding possible ailments. The main repetitive component in an EKG – the QRS complex – and accompanying waves are shown in Fig. 1.1. Secondly, although ML has been applied extensively in biomedical signal processing, it has not yet been used widely in wireless communications. The next part of
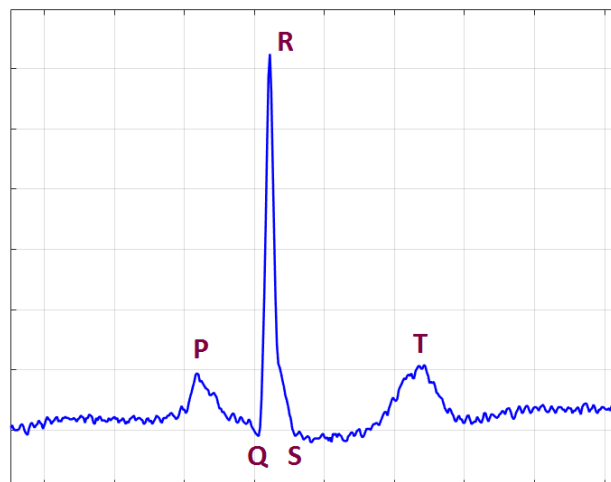
Figure 1.1 The QRS complex in an EKG along with P and T-waves.

1

this thesis proposes a solution to predict base station association for mobile users in emerging 5G networks employing ML classification. Next generation mobile network architectures will have vastly different "cell" structures compared to their predecessors.

1.1 Atrial Fibrillation

Atrial Fibrillation (AF) is a common cardiac arrhythmia affecting over 5 million people in the United States [2]. In normal sinus rhythm, the heart beats regularly as blood is pumped from the upper chambers of the heart (atria) to the lower ones (ventricles). Thereafter, blood is pumped to the rest of the body and a subject's EKG does not show any irregularities (Fig. 1.2). AF is a condition in which the atria are unable to contract effectively in response to the pacing activity of the sinoatrial node. This is caused by abnormalities in the electrical conduction pathways throughout the atria. Consequently, disorganized atrial contraction leads to an inhibited blood flow into the ventricles causing stagnation. This can result in the formation of a blood clot and makes AF to be considered one of the leading causes of embolic stroke [3]. Depending on the duration and nature of termination of the AF episodes, a patient can be diagnosed as having permanent, persistent or paroxysmal AF. A patient suffering from paroxysmal AF will experience
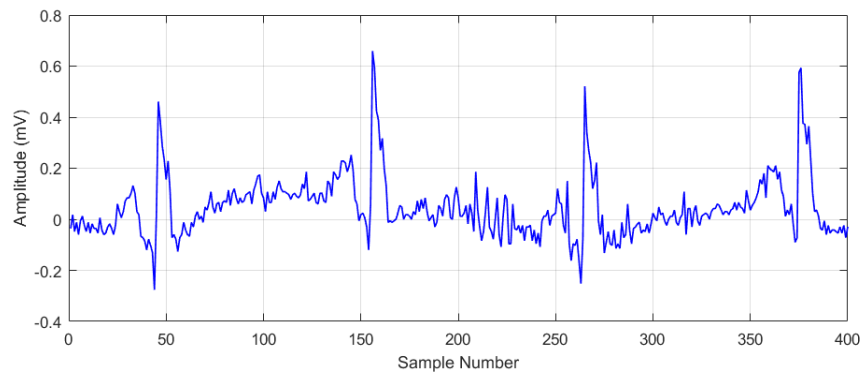
Figure 1.2 A noisy EKG recording from a subject when the heart is in normal sinus rhythm.

spontaneous AF episodes of relatively short duration that terminate on their own. In contrast, persistent AF can last for days. If normal sinus rhythm cannot be restored, by cardioversion for instance, then AF is classified as being permanent [4]. Cardioversion is a medical procedure performed to restore the heart's

2

normal rhythm. This may be done chemically, by prescribing drugs, or electrically, by providing a shock to the heart.

Certain medical conditions put patients at an increased risk of developing AF. These include congestive heart failure, hypertension, valve disease, myocardial infarction and diabetes [5]. Though sometimes asymptomatic, AF patients can experience shortness of breath, fatigue, palpitations and chest discomfort. AF can go undiagnosed as a patient may not necessarily experience a fibrillatory rhythm while undergoing an EKG during routine examination. This highlights the need for long-term EKG monitoring especially in cases such as infrequent paroxysmal AF where episodes are unaccompanied by noticeable symptoms [6]. AF is generally characterized by the absence of any clear P-waves, a much faster heartbeat and "irregularly irregular" QRS complexes (caused by chaotic atrial contractions) appearing on an EKG trace (Fig. 1.3). If diagnosed, a physician will typically prescribe an anticoagulant to minimize stroke risk and some other medication to control heart rate by slowing down electrical conduction. Other treatment types include catheter ablation – a minimally invasive procedure where some conduction pathways in the heart are electrically burned – and pacemaker implantation.
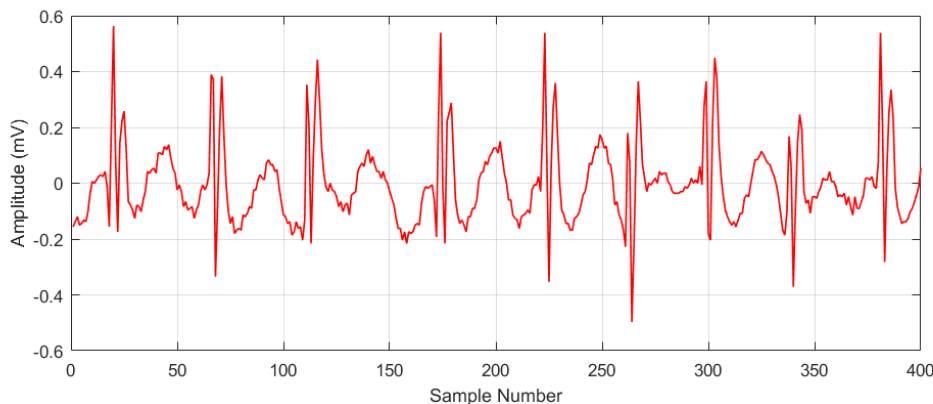
Figure 1.3 An EKG recording from a subject during AF. The heart rate is more rapid in comparison to the normal EKG shown in the previous figure.

1.2 The Computers in Cardiology Challenge 2001

The Computers in Cardiology Challenge 2001 was held with the objective of furthering research into the detection and prediction of paroxysmal AF [7]. The competition comprised of two separate parts. In the first part, contestants were provided a set of EKG recordings, each being half an hour in duration,

3

acquired from both healthy subjects and patients who were diagnosed as having paroxysmal AF. These recordings were to be used to identify features or descriptors in the EKGs that best distinguished the people in both classes with the help of ML algorithms or other statistical models. Once entries to the competition had been submitted, they were evaluated by checking their performance in classifying a previously unseen set of EKGs belonging to patients and healthy subjects. Accuracy was measured by the percentage of correct predictions.

The competition consisted of a second component that was more challenging. Contestants were provided with pairs of EKG recordings, with both EKGs in a pair belonging to the same patient. One of the half hour EKGs in each pair terminated right at the commencement of an AF episode and the other EKG came from a time interval far away from any such episode. The problem was to predict, in each pair, which of the two EKGs just preceded the AF episode. Different methodologies were proposed by the contestants and these have been reviewed in more detail in the following chapter. The types of features that were extracted could be classified into two – features related to Heart Rate Variability (HRV) and features related to P-wave morphology. Accuracy values achieved in the AF prediction part of the challenge were in the vicinity of 60-80%. Later attempts have been made to develop better models that are able to perform this classification with more precision.

1.3 Mobility Management and 5G Cellular Networks

An efficient handover scheme is critical to providing uninterrupted service to mobile users who are streaming multimedia content [8]. Currently, a user receives data from a single base station or access point when connected. This is expected to change in a 5G cellular network where a user may be connected to multiple base stations simultaneously. In order to minimize interference and maximize channel capacity, it is essential that the network determines which base stations are the best ones for a particular user to be connected to. Moreover, 5G networks also need to have the ability to support low latency communications as would typically be required for self-driving cars. This has given rise to the concept that handover, or mobility management in general, needs to be proactive with the network having to anticipate which base

station a user is going to be connected to next. Predicting the next base station accurately would enable the user's information to be forwarded to it, enabling a smooth handover in a timely manner.

One of the ways in which this problem has been addressed is by utilizing the GPS (Global Positioning System) coordinates of a mobile device and building up prediction models based on previous user trajectories. Individual histories of visited locations and dominant paths can be extracted from the moving patterns of a large number of users. This information can be leveraged to predict the next probable location of a user in the short term and also the base station or access point which would pick up the connection. However, continuously enabling GPS signal transmission can reduce battery life and invade privacy [9]. Mobile users can therefore be reluctant to let the network know their exact location all the time.

Currently, base station association in LTE-A (Long Term Evolution-Advanced) is performed using a mechanism known as range extension. Here, constant bias values are added to Received Signal Strengths (RSS) in order to address the uplink-downlink imbalance. This enables, for instance, a mobile user in a heterogeneous network to transmit to a nearby picocell base station in the uplink so that the picocell base station receives at a higher power level, while in the downlink, the mobile user can receive data from a macrocell base station at a higher throughput. Next generation cellular networks are anticipated to have virtual cells where base stations will have to jointly make decisions regarding user association in a dynamic environment. This adds to the complexity of providing anticipatory mobility management.

1.4 Motivation

There has been recent interest in the field of personalized medicine [10]. Clinicians acknowledge that patients respond to medications differently based on their genetic makeup. Therefore, there is an ongoing effort to customize therapeutic procedures for patients in order to produce better results. The work presented in this thesis related to the prediction of oncoming episodes of AF reflects this personalized view. While AF may be classified universally using distinct EKG markers that occur during fibrillatory rhythms, it is likely that there are unique EKG indicators, specific to each individual, that are precursors to these AF episodes. Based on a study of Holter recordings [11] in which it was sought to determine whether a common mode of AF initiation existed, Hnatkova *et al.* concluded that there was no consistent mode of AF onset

between patients, but rather that there was marked variation in rhythm and rate prior to AF rhythms within the population. This motivated the development of patient-specific ML models to predict AF occurrence. Chapter 2 reviews the relevant literature where AF prediction methods have been presented and describes the methodology used for binary classification.

This thesis also presents a novel means of predicting base station association using sequences of RSS values. While base station prediction algorithms have been proposed in the literature, none of them have considered prediction in the case where a single user is connected to multiple base stations simultaneously. Chapter 3 describes the model used to collect synthetic data, the feature extraction process and classification. Chapter 4 provides some concluding remarks noting directions for future research.

CHAPTER 2: ATRIAL FIBRILLATION PREDICTION

This chapter describes the methodology used to predict AF onset and reviews some of the prior art in the area. Much of the work cited here relates to the competition entries that were submitted to the original Computers in Cardiology Challenge 2001. Algorithms subsequently proposed have occasionally been able to achieve higher accuracy values in AF prediction than what was originally achieved, but it should be noted that later researchers had access to some of the test datasets that were originally unavailable to the contestants in 2001. The work described in this chapter was presented at the IEEE International Conference on Biomedical and Health Informatics 2017 [12].

2.1 Literature Review

The AF prediction part of the Computers in Cardiology Challenge 2001 had 28 EKG pairs in the test set with one recording in each pair immediately preceding an AF rhythm. An entry was given a score based on the number of pairs that were correctly classified.

State-space domain features of Heart Rate Variability (HRV) were used by Lynn and Chiang [13]. They first considered the difference map of the HRV signals where x_i was plotted against x_{i+1}. Next they obtained the HRV return plot that was constructed by plotting $(x_{i+1} - x_i)$ vs. $(x_{i+2} - x_{i+1})$. They divided the plots into patches and took the fraction of points within each patch as a feature. Classification was performed using a *k*-Nearest Neighbors (*k*NN) algorithm and they obtained an accuracy of 64% on the prediction challenge. Langley *et al.* [14] used an automated QRS detector to obtain the sequence of RR-intervals. They maintained a moving average of these intervals and identified an ectopic beat as one for which the interval was less than that of the moving average by more than 20%. These suspected ectopic beats were then further classified as either having atrial or ventricular origin based on the beat that immediately followed it. Their method selected the EKG record with the higher number of atrial ectopic beats as the one that immediately preceded AF onset. They achieved an accuracy of 61%. Zong *et al.* [15]

used a method they had developed previously to detect Premature Atrial Complexes (PACs). The method was based on calculating certain parameters such as the width, amplitude and area of a QRS complex as well as how prematurely the beat appeared. They further developed a weighting scheme to favor PACs that had appeared most recently and used a rule-based classification algorithm. They obtained 79% – the highest accuracy by an entry for the second part of the challenge. Maier *et al.* [16] explored several different kinds of features based on RR-intervals and utilized a first order polynomial classifier. The features included statistical measures of RR-intervals, measures of RR-intervals of only normal heartbeats and wavelet coefficients. They correctly classified 19 of the 28 EKG records in the prediction challenge and obtained 68% accuracy. Schreier *et al.* [17] applied a beat detection and classification methodology to extract a template for a regular/normal P-wave. They next calculated the correlation between this representative P-wave and P-waves from both normal and premature beats. A statistical test was used to determine whether both samples came from the same distribution. They developed a metric attaching more weight to premature beats that had occurred most recently and used this to identify which EKG record just preceded AF onset. They correctly classified 20 pairs and obtained an accuracy of 71%. De Chazal and Heneghan [18] considered four different groups of features for PAF prediction. These included both time and frequency domain features of RR-intervals and P-waves. They obtained superior results based on the RR-interval features using a linear discriminant classifier. Their score was 68%.

Later works using the same competition dataset include that of Thong *et al.* [19] who developed a rule-based system that depended on the appearance of PACs, atrial trigeminy and bigeminy rhythms and short durations of paroxysmal atrial tachycardia. Changes in any of these were taken as indicators of oncoming AF and their algorithm outperformed the entry that had achieved the highest score in 2001. Correctly classifying 25 out of the 28 pairs in the PAF prediction challenge, they obtained an accuracy of 89%. Hilavin and Kuntalp [20] divided the half hour EKG records into six equal parts. They calculated a complex correlation measure based on the Poincare plot of HRV signals and used the standard deviation of these values to classify EKGs far away from an AF rhythm from EKGs just prior to onset. They achieved an accuracy of 72%. Arotaritei and Rotariu [21] used Teager entropy (an indication of the oscillatory nature

of a time series), the appearances of PACs and the Root Mean Square (RMS) of successive differences in RR-intervals for classification and obtained a sensitivity of 52%. Chesnokov *et al.* [22] used the power spectral density of the HRV signals and demonstrated statistically significant differences in the 0-0.5 Hz frequency range between EKGs preceding AF and those distant to onset. They trained a neural network for classification and reported 72.7% sensitivity on a test set.

Due to the limited amount of data made available in this competition, none of the contestants and subsequent researchers were able to adequately develop patient-specific classification techniques to identify whether or not AF was about to occur. This is the primary difference in the approach used here. Additionally, the contestants were given two EKGs at a time and had to determine which of them preceded AF. This is easier in comparison to the more general approach proposed in this thesis. Given a set consisting of both EKGs located far away from AF episodes and EKGs just prior to onset, a binary classification is attempted to separate the records into two classes. Moreover, prediction was performed using training data comprising of half hour EKG signals in the 2001 competition. In this thesis, EKGs of two minutes in duration are used for prediction.

2.2 Dataset

PhysioBank's Long-Term Atrial Fibrillation Database (LTAFDB) consists of 84 EKG recordings acquired from patients with a form of AF. They have been digitized at a sampling frequency of 128 Hz and the duration of each recording is approximately one day. Annotations accompany each recording indicating the occurrence of abnormal beats and rhythms. In this thesis, patient-specific classifiers are developed to distinguish between EKG signals leading up to episodes of AF and those far away or distant from such episodes. First, EKGs of two minutes in duration were extracted from each record, where each of them terminated right at the commencement of an AF episode. These EKG intervals are called epochs. If an epoch contained yet another AF episode, it was discarded. Thereafter, EKG epochs were extracted, where they terminated a minute, two minutes and three minutes prior to the start of AF episodes. This gap between the termination of an epoch and the start of a fibrillatory rhythm was named as the prediction horizon. In this manner, four sets of EKG epochs per record were extracted with prediction horizon equal to zero, one,

two and three minutes. These epochs formed the pre-AF class. A record needed to contain at least 20 episodes of AF within it to be considered. Otherwise, there would not be sufficient data to train a patient-specific model.

In order to collect EKG epochs far away from AF episodes, 100 locations were randomly selected from each record such that it was located at least 10 minutes away from a fibrillatory rhythm. The assumption being that there was no indication that AF was about to occur 10 minutes before actual onset. Then, EKG epochs of two-minute durations were extracted from these 100 locations. These epochs formed the distant class. Four separate classifiers were developed for each record to distinguish between the distant class and the set of epochs belonging to the pre-AF class for a particular prediction horizon. The purpose was to investigate how much forewarning could be given prior to the actual occurrence of a cardiac event.

2.3 Feature Extraction

The rate of atrial contraction increases rapidly when a patient experiences AF. This causes the ventricles to contract faster than usual as well. However, the activity of the atrioventricular node prevents the ventricles from contracting as fast as the atria. Consequently, heart rate typically increases up to about 140 beats per minute during AF. This change in heart rate during AF motivated the extraction of RR-interval related features in the lead up to fibrillatory episodes as a possible indicator of AF onset. Secondly, a number of abnormal beat and rhythm annotations are provided in the database. The appearance of these was also incorporated into the features to feed into a classifier.

2.3.1 RR-interval Outliers

Annotations provided in the database were used to extract the duration in-between consecutive heartbeats for each EKG epoch. As heart rate can fluctuate with physical activity, a cubic-spline interpolated trend line was subtracted from this sequence of RR-intervals (i.e., the sequence of values was centered around zero) and thereafter normalized by dividing by the maximum absolute within the sequence.

Next, statistical outliers were defined as the points within the normalized sequence that were greater than three standard deviations away from zero. The first feature extracted was the number of these outliers. Then, the minimum, maximum, mean and median of the outliers were calculated. These five features

characterized the outliers of the RR-interval sequence within each epoch. Thereafter, the outliers were removed from the sequence and the median and RMS value was calculated. Finally, the number of data points within the sequence (including the outliers) was counted. This is a measure of the number of heartbeats within an epoch. These eight statistics helped capture some of the heart rate variations that were taking place within each EKG epoch.

2.3.2 Autoregressive (AR) Modeling

Consider a white noise process $w[n]$ being filtered by an IIR (Infinite Impulse Response) filter of order m. The output $x[n]$ satisfies the difference equation,

$$x[n] + a_1 x[n-1] + \cdots + a_m x[n-m] = w[n] \tag{2.1}$$

If a known random process $x[n]$ is modeled in this manner, the filter coefficients a_i can be determined using the Yule-Walker equations [23], [24]. This method was used to extract four filter coefficients for the RR-interval sequence of each epoch. The filter coefficient values also help capture some of the heart rate fluctuations that may occur just prior to an AF episode. AR modeling has previously been used in AF detection [25].

2.3.3 Ectopic Beats

Different abnormal heartbeats and rhythms have been annotated in the database. The abnormal beat types include PACs, premature ventricular contractions etc. Some of the annotated rhythm abnormalities include sinus bradycardia, ventricular tachycardia and atrial bigeminy. The number of times each of these abnormal beat and rhythm types occur in an EKG epoch was also included in the features. With all three types of features being included, a total of 27 different values were selected to characterize each two-minute EKG epoch.

2.4 Feature Selection

Patient-specific classifiers were trained for each prediction horizon. Therefore, for each patient, the best set of features that distinguish between the distant and pre-AF class for a particular horizon needed to be selected. This was done using a Mann-Whitney U test [26]. This hypothesis test checks whether two different samples of values come from populations with different medians. For instance, consider the

number of PACs in an epoch when classification needs to be performed for patient ID 121 with a prediction horizon of one minute. In order to decide if this would be a good feature to use for classification, the sample comprising of all such values in the distant class and the sample comprising of all such values in the pre-AF class need to checked if they come from distributions with different medians or not. If the Mann-Whitney test reveals that there is sufficient evidence to believe that the samples come from different distributions, then the feature is selected. Therefore, a best set of features exists to perform binary classification for each patient, for each prediction horizon.

2.5 Visualization

The feature vectors belonging to both classes are distributed in a high-dimensional space and need to be projected onto a 2D plane for visualization. Principal Component Analysis (PCA) [27] was employed for this purpose and required the first two eigenvalues of the feature covariance matrix. The PCs (Principal Components) are linear combinations of the original features and do not have a physical interpretation.
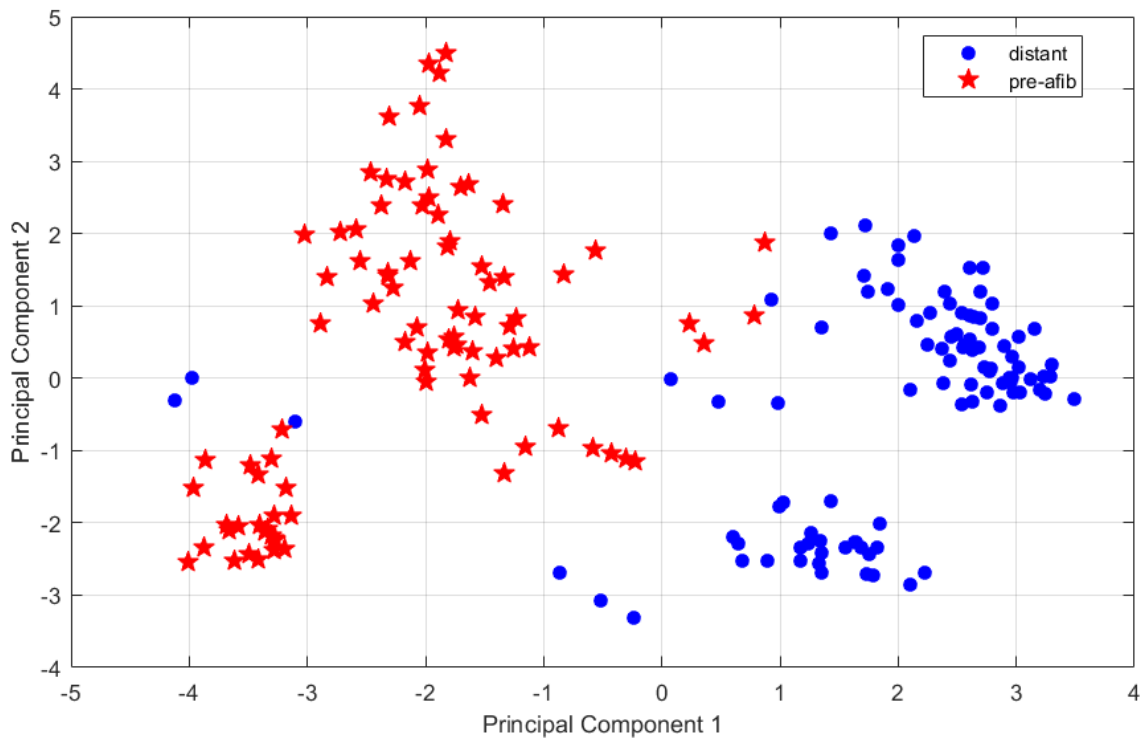
Figure 2.1 Feature distribution for record 33 for zero-minute prediction horizon.

Several feature distributions for different patients for distinct prediction horizons are shown in Fig. 2.1 – 2.4.
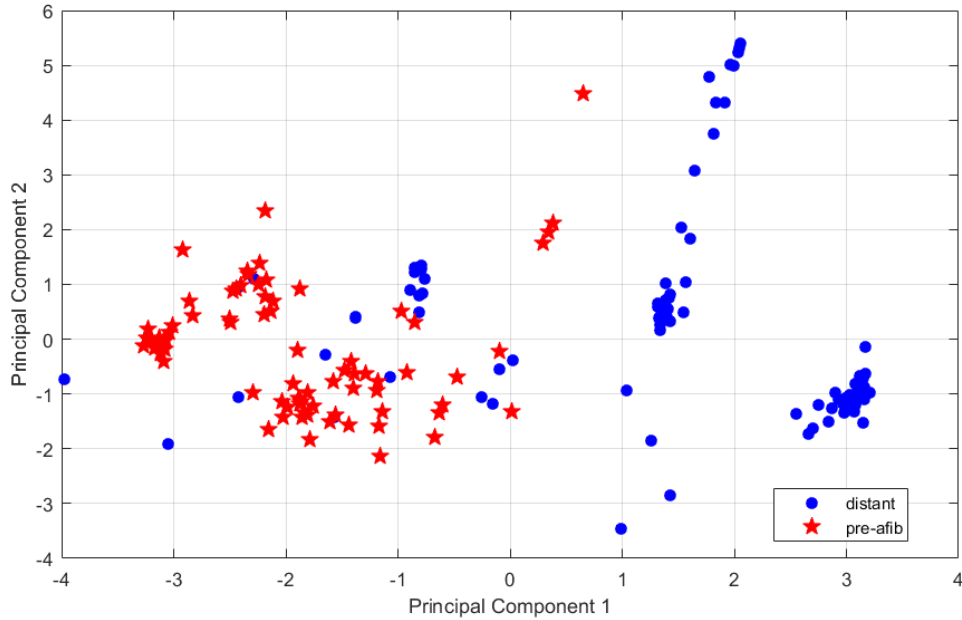

Figure 2.2 Feature distribution for record 39 for zero-minute prediction horizon.
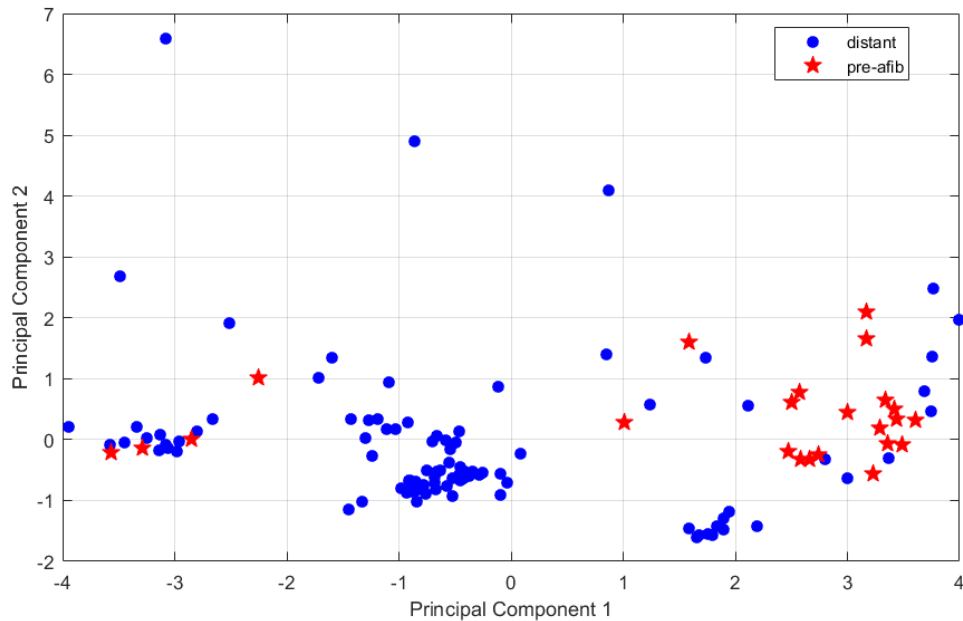
Figure 2.3 Feature distribution for record 45 for one-minute prediction horizon.

2.6 Classification

Support Vector Machines (SVMs) [28] were used for classification using 10-fold cross validation. This was performed using the WEKA software [29]. SVMs belong to the class of large-margin classifiers

13
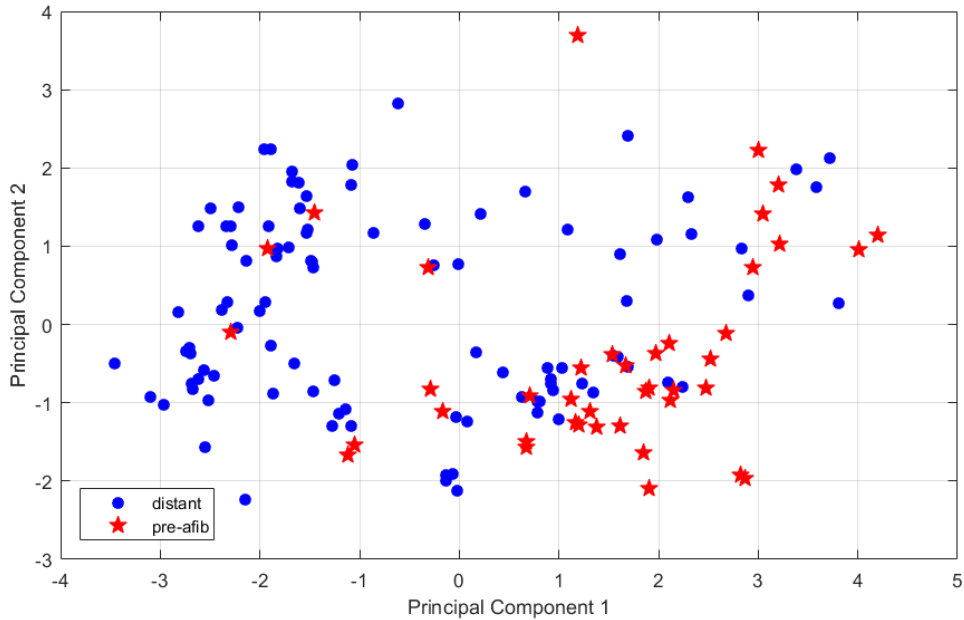
Figure 2.4 Feature distribution for record 51 for one-minute prediction horizon.

and attempt to separate out data points in different groups by mapping the points to a much higher (possibly infinite) dimensional space. Classification is more robust because the separating hyperplane only depends on a few data points that are known as support vectors. Four separate classifiers were trained for each patient (i.e., a classifier for four different prediction horizons) and sensitivity, specificity and overall accuracy were calculated. Sensitivity is the ratio between the number of correctly classified feature vectors belonging to the pre-AF class and the actual number of feature vectors in that class. Specificity measures the same quantity, but is calculated for the distant class instead. Overall accuracy is the ratio between the number of correctly classified feature vectors and their total number. Accuracy values for every record for three different prediction horizons appear in Table 2.1 and the variation of mean and median sensitivity with prediction horizon has been shown in Fig. 2.5.

2.7 Discussion

Different features best distinguish between the distant and pre-AF class for a particular prediction horizon for every patient. For some of them, it is the features related to RR-intervals that are most suitable for classification while for others it is the descriptors related to the appearance of ectopic beats that are the best. Very few statistically significant features are present for certain patients as prediction horizon

14

increases. For instance, patient ID 102 did not have any suitable features for classification for a prediction

horizon beyond a minute.

Table 2.1 Variation of classification specificity, sensitivity and accuracy for three separate prediction horizons for each patient/record.

Patient/ Record ID	Total AF Episodes	0 minutes			1 minute			2 minutes		
		Spec.	Sens.	Acc.	Spec.	Sens.	Acc.	Spec.	Sens.	Acc.
00	21	0.99	1.00	0.992	0.98	0.647	0.932	0.98	0.643	0.939
01	35	0.96	0.971	0.963	0.92	0.889	0.913	0.94	0.739	0.902
10	31	0.98	1.00	0.985	0.97	0.65	0.917	0.94	0.588	0.889
15	91	0.94	0.989	0.963	0.90	0.727	0.839	0.93	0.714	0.866
23	37	0.94	0.973	0.949	0.90	0.481	0.811	0.94	0.25	0.806
25	69	0.88	0.957	0.911	1.00	0.00	0.645	1.00	0.00	0.694
26	89	0.80	0.966	0.878	0.75	0.739	0.746	0.83	0.627	0.762
32	32	1.00	1.00	1.00	1.00	0.647	0.949	1.00	0.40	0.945
33	90	0.97	1.00	0.984	0.84	0.26	0.588	0.79	0.352	0.608
39	78	0.94	0.987	0.961	0.85	0.403	0.679	0.91	0.212	0.671
42	24	1.00	0.958	0.992	0.94	0.143	0.842	0.92	0.154	0.832
45	32	0.93	0.875	0.917	0.92	0.696	0.878	0.96	0.75	0.938
51	45	0.98	0.956	0.972	0.90	0.561	0.801	0.91	0.432	0.781
53	47	0.87	0.702	0.816	0.93	0.313	0.78	0.99	0.083	0.815
62	38	0.98	0.974	0.978	0.90	0.00	0.652	0.91	0.105	0.688
74	96	0.86	0.927	0.893	0.82	0.50	0.708	0.97	0.025	0.70
100	89	0.98	1.00	0.989	0.90	0.439	0.732	0.99	0.22	0.766
101	90	0.98	1.00	0.989	0.88	0.671	0.79	0.93	0.607	0.814
102	23	1.00	0.957	0.992	0.99	0.00	0.839	-	-	-
105	20	0.98	0.80	0.95	1.00	0.00	0.885	0.97	0.20	0.90
112	146	0.844	0.979	0.938	0.766	0.96	0.884	0.797	0.935	0.872
115	44	1.00	1.00	1.00	0.98	0.25	0.82	0.97	0.05	0.817
119	26	0.98	0.885	0.96	0.95	0.632	0.899	0.95	0.632	0.899
120	21	0.99	0.905	0.975	1.00	0.00	0.833	1.00	0.00	0.847
121	70	0.84	1.00	0.906	0.87	0.298	0.662	1.00	0.00	0.69
204	102	0.81	1.00	0.906	0.89	0.754	0.836	0.86	0.63	0.788
Mean	57.15	0.939	0.952	0.952	0.913	0.448	0.802	0.935	0.374	0.809

It can also be seen that classification accuracy is highest for a zero-minute prediction horizon. This

is intuitive as well, since it is likely that prior to actual AF onset, there is enough indication in the EKGs

that AF is just about to occur. As the prediction horizon increases and there is a larger gap between AF
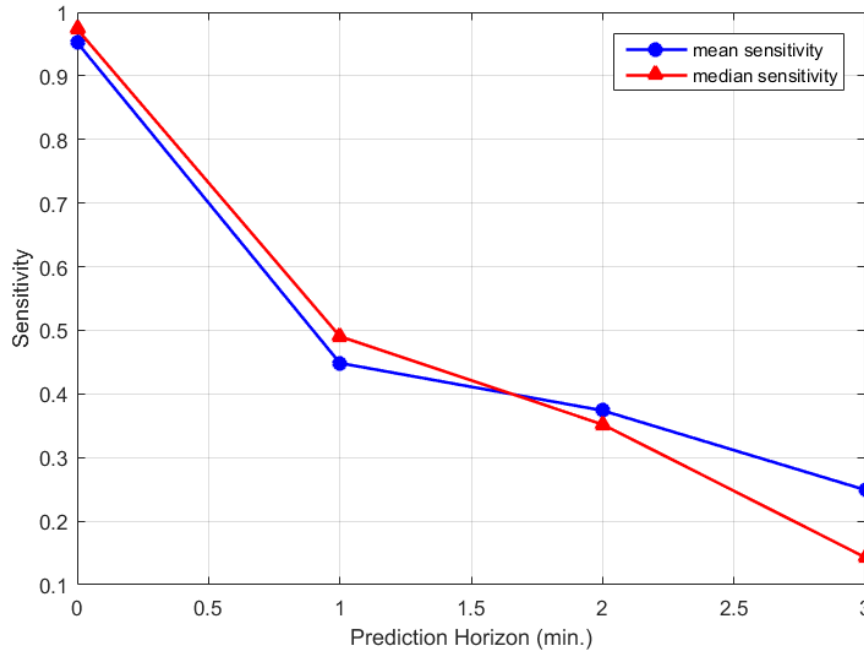
Figure 2.5 Variation of mean and median classification sensitivity with prediction horizon.

onset and the pre-AF EKG epochs, less predictive information is present and hence classification tends to perform poorly. For most patients, the number of feature vectors in the distant class is higher than in the pre-AF class for a particular prediction horizon. This distinction is more pronounced as the prediction horizon increases. Consequently, a statistical classifier tends to favor the majority class and hence specificity can be seen to remain relatively high throughout. However, this is of not much value since it is the AF episodes that interest lies in. To illustrate, suppose an intelligent monitor keeps on predicting that no cardiac event will occur. Since events do not occur frequently, the specificity (and overall accuracy) will be high. This is hardly of use to a patient, as it is the medical conditions of clinical importance that need to be averted.

CHAPTER 3: BASE STATION PREDICTION

A methodology used to predict base station association for a mobile device is described in this chapter and was presented at the IEEE Wireless and Microwave Technology Conference 2017 [30]. The relevant prior art, where different approaches have been used for this purpose, is reviewed first and is followed by the details of the simulation model utilized in this thesis. Chapter 3 goes on to describe some of the unique features of Recurrent Neural Networks (RNNs). The chapter concludes with a discussion of the results obtained and an outline of directions for future work on base station prediction in a 5G cellular network.

3.1 Literature Review

Different ML algorithms for predicting the base station a user will be associated with next have been proposed in the literature. Michaelis *et al.* [31] explored a few types of features for next cell ID prediction using SVMs. They used short sequences of previously visited cell IDs and obtained a classification accuracy of close to 78%. Accuracy did not improve significantly after the length of the sequences was increased beyond three. When additional features such as the time duration that a user was within a particular cell, GPS coordinates and user IDs were incorporated into the classification model, accuracy values ranged between 77-78%. While this approach was able to predict where a user might move, it was unable to predict when the transition to another cell would occur. Hence, they further used undirected graphical models to come up with the most probable cell a user would be associated with at a specific time instance in the future. They presented the accuracy of this second type of generative model for three different users. It was highest for the user having the largest amount of data and was a little over 78%. Tkacik and Kordik [32] used a Neural Turing Machine to predict the mobility of suspects. A Neural Turing Machine is comprised of a memory and a controller and is similar to a RNN. It utilizes a memory matrix with selective write and read operations in order to remember data and have dynamic state. The authors

validated their proposed method both with synthetic data generated for seven people and real data recorded from six people over a period of three months. They defined their own accuracy metric and showed that it outperformed a *k*NN classifier. The highest accuracy they obtained in predicting locations was just over 82%, which was for one specific user. Luo *et al.* [33] used a neural network based on a Nonlinear Autoregressive Exogenous Controller (NARX) to make an optimal handover prediction. They attempted the use of a ML algorithm to minimize the "ping-pong effect" that could take place when a mobile device makes a connection from one access point (AP) to another. This minimizes the cost associated with the procedure. They considered two APs in a situation where a mobile device moves from the coverage area of one AP to another. They fed the neural network with RSS values and delays and trained it using optimal handover decisions that had been obtained by solving an Integer Linear Programming (ILP) problem. On test data, their algorithm was able to predict handover close to the actual optimal location. Their proposed method had the advantage of minimizing multiple handovers within a very short period of time, which can occur in a threshold-based approach. Javed *et al.* [34] used four different features to predict the occurrence of a handover in a future one-minute time interval. Prior to actual classification, they first examined the correlation that each of these features had with future handover rates. An active set was defined as the cells with signal strength above a certain level to which a mobile device remains connected. Although there may be up to four cells in an active set, only one of them actually transmits data to the device. The features they used for classification included the past handover rate, size of the active set, active set update rate and signal strength fluctuation. They used Decision Stumps [35] for classification along with boosting. Accuracy was between 75-80% when individual features were used for classification and just over 80% when all features were combined together.

Liou and Huang [36] suggested that although the motion trajectory of a mobile device appeared random, movement was actually a function of position, direction, speed and acceleration. They developed a mobility model and a location predictor based on a neural network to predict base station association. Successful prediction would eliminate the need for unnecessary bandwidth reservation in neighboring cells where the device was unlikely to travel to. The three previous location coordinates of a user were input to

the neural network and the output consisted of the next visited coordinates and cells. They evaluated prediction performance on three different roads and reported an accuracy of 78% on a mountainous road, 86% on a circular road in the city and 90% on a freeway with fewer sharp turns. Laasonen [37] proposed a graph-based approach for future user location prediction. Each cell was taken to be a node in a graph with an edge existing only if there was a possible transition between those two cells. Occasionally, signal strength from different base stations can be almost equal and handover can occur even if the mobile device is not moving. In this case, a cluster of cells within which a large number of transitions occur is considered to be a single node in the graph rather than each of those cells being separate nodes. These nodes are called locations and locations in which a user spends a significant amount of time are known as bases. An algorithm was presented to predict the next base to be visited by a user given the current base. Performance was evaluated on a dataset recorded from three volunteers over a six-month period. Accuracy varied for each person and was between 80-90% for more than three learning trips. Anagnostopoulos *et al.* [38] evaluated several different classifiers in predicting the next cell a user would move to. They used Naïve Bayes, Decision Trees, rules, ensemble learning and kNN. The sequence of previously visited cell IDs made up the input feature vector. They generated synthetic user movements by varying a particular parameter (δ) that controlled the degree of randomness. A δ value close to one represented highly randomly behavior and a value close to zero exhibited more regularity. A classification rule based on voting on the results of the other classifiers yielded the highest accuracy over 90% with $\delta = 0$ and approximately 70% with $\delta = 0.25$. Accuracy was best when they used a sequence comprising of the last four-visited cell IDs. Anagnostopoulos *et al.* [39] also evaluated the performance of Naïve Bayes, rules and Decision Trees when a temporal context (i.e., the time of day) was included. They used a similar approach to the one mentioned above and reported similar accuracies although the spatiotemporal models outperformed the models based solely on spatial context by a slight amount.

Despite the use of many ML algorithms for predicting base station association, none of them have considered learning patterns in variable-length sequences of RSS values for classification. Moreover, none of the methods have considered the mobile device to be receiving data from multiple base stations at the

same time. Simultaneous data reception is something that would be common in a 5G network due to Coordinated Multipoint. This thesis proposes the use of a RNN for classifying such sequences to proactively make decisions about mobility management.

3.2 Model

A 6 km x 6 km road network was defined as shown in Fig. 3.1 with eight base stations placed at arbitrary locations within it. Next, moving vehicles and pedestrians were generated at random locations on the roads. A pedestrian had an initial speed drawn from a uniform distribution between 8-12 km/h and a vehicle had an initial speed between 55-65 km/h. Each pedestrian or vehicle, termed a mobile node, continues to maintain its constant speed until leaving the road network. When a mobile node comes to an intersection, it decides to go straight ahead with probability 0.5 or make a left or right turn with an equal probability of 0.25. No U-turns were allowed in this simulation.
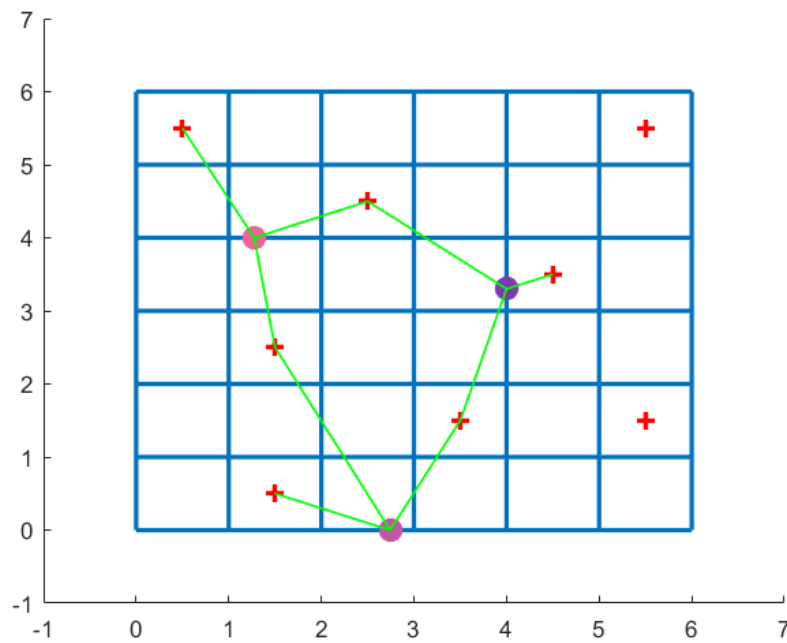
Figure 3.1 The simulated road network section with eight base stations (crosses) and three mobile nodes (circles). Each mobile node is connected to the three base stations closest to it.

While a mobile node is within the road network, it connects to the three base stations closest to it and maintains connectivity until completely leaving the grid. More complex base stations assignments are possible, but these have been left for future work. As the mobile node moves, it undergoes handover

20

whenever the distance to a new base station becomes smaller than the distance to the furthest base station it is connected to. The primary purpose of this research was to demonstrate that a RNN can be trained to identify patterns in variable-length RSS sequences to predict this new base station.

RSS values in decibels were calculated according to the path loss model (shown below) in [40] for an urban environment. An additional term (σ) was included to account for shadow fading. σ is normally distributed with mean zero and variance 9 dB and d is measured in kilometers in the following formula.

$$PL(d) = 128.1 + 37.6 \log(d) + \sigma \tag{3.1}$$

3.3 Recurrent Neural Networks (RNNs)

A typical neural network has a drawback in that it is unable to learn time-dependent patterns in input data. Feature vectors have constant size and are typically assumed to be independent. RNNs represent a variation to typical neural networks and can learn patterns in sequences, even if they are of arbitrary length. This is accomplished by adding a feedback path so that the present output is not only dependent on the current input but also on the state. This addition of memory allows time dependencies in the data to affect the predictions made. A simple RNN with a layer of neurons A, an input vector X_t and an output vector h_t is shown in Fig. 3.2. Shown to the right of it in the same figure is the "unrolled" version of this structure. The subscript t represents the time instant. The unrolled version of the figure shows how the output at time t (h_t), depends not only on the current input X_t, but also on the previous state of the layer A at time t-1. The feedback path has been unrolled into horizontal arrows depicting the flow of state/memory.
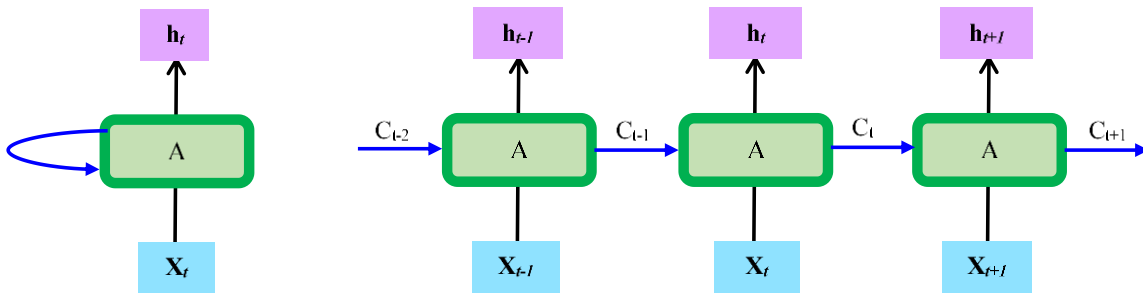

Figure 3.2 A simple RNN (left) and its "unrolled" version (right).

Unfortunately, these RNNs have a very serious issue. Despite their ability to learn patterns in a time series, they are difficult to train using back propagation and can only handle short time dependencies

in the input sequences. The issue lies with the derivatives of the sigmoid functions or hyperbolic tangents that are used as activation functions in neural networks. These functions have gradients that are small for a very large range of values. As long chains of these derivatives comprising of values less than one get multiplied, the result approaches zero quickly, i.e., it vanishes [41]. This is why only short temporal patterns in data sequences can be captured with typical RNNs.

The solution to this problem lies in having specialized units, instead of typical neurons, where the state is allowed to propagate to the next time instant adaptively. The LSTM (Long Short Term Memory) module was developed for this purpose [41]. A basic LSTM unit is shown in Fig. 3.3 and C_t represents the current state. "Gates" are used to control how much of the previous state is allowed inside in order to influence the current state, how much of the current state is to be remembered or forgotten and how much the current state is allowed to influence the final output. These gates comprise of multiplication operators along with sigmoid activation functions (labeled σ in the figure) whose inputs are the current inputs and previous output. Therefore, each LSTM module contains multiple neurons whose weights need to be learned during the training phase.

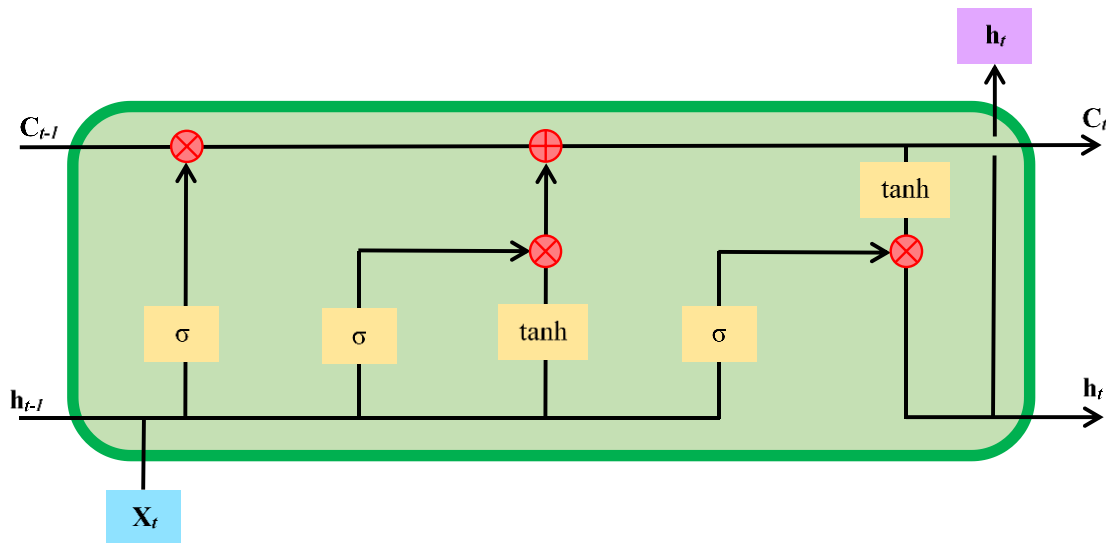Figure 3.3 A LSTM module.

The LSTM module depicted above is a single unit in the unrolled version of a RNN. A portion of the previous state C_{t-1} is allowed to flow in depending on the previous output h_{t-1} and current input X_t. The

22

current state C_t is also calculated by allowing the previous state to add to it. The current output h_t along with the next input X_{t+1} would then determine how much of C_t is allowed in at the next time instant $t+1$.

3.4 Sequence Extraction

As a mobile node moves along the roads, the RSS values from the three base stations it is connected to were recorded every 500 ms. Since a total of eight base stations were present in the network, a vector with eight values was created at each such time instant. Only three elements of the vector had actual RSS values while the other elements were set to zero. These RSS vectors were stored in a queue data structure. Since RSS values from a long time in the past do not contain a lot of information regarding where a mobile node might possibly be headed to, the oldest vectors in the queue are removed once its size exceeds 150. The RNN also requires a reasonable amount of data in order to make predictions and hence very short queues were discarded.

Once handover for a particular device occurred, the RSS sequence was extracted and labeled, at which point the queue was also reset. The label given to the sequence corresponds to the new base station that the mobile node got connected to. The very last RSS vector containing a value from this new base station was not added to the queue. A total of 100,000 RSS sequences were extracted in this manner. 70% of the sequences were used to train the RNN classifier and the remaining sequences were retained for testing accuracy. Some of the RSS sequences have been shown in Fig. 3.4.

3.5 Classification

Tensorflow [42] is a Python library developed by Google for training neural networks. This software library was used to develop a classification model for predicting base station association in this simulation. The code itself was based on [43] that serves as a general example of how Tensorflow can be used to separate out unequal-length sequences into distinct categories.

The RNN comprised of 640 LSTM units with a learning rate set to 0.0003 during training. Different combinations were tried for the model parameters and the ones selected finally provided the best results. One of the technologies that has fueled recent interest in neural networks has been the use of NVIDIA
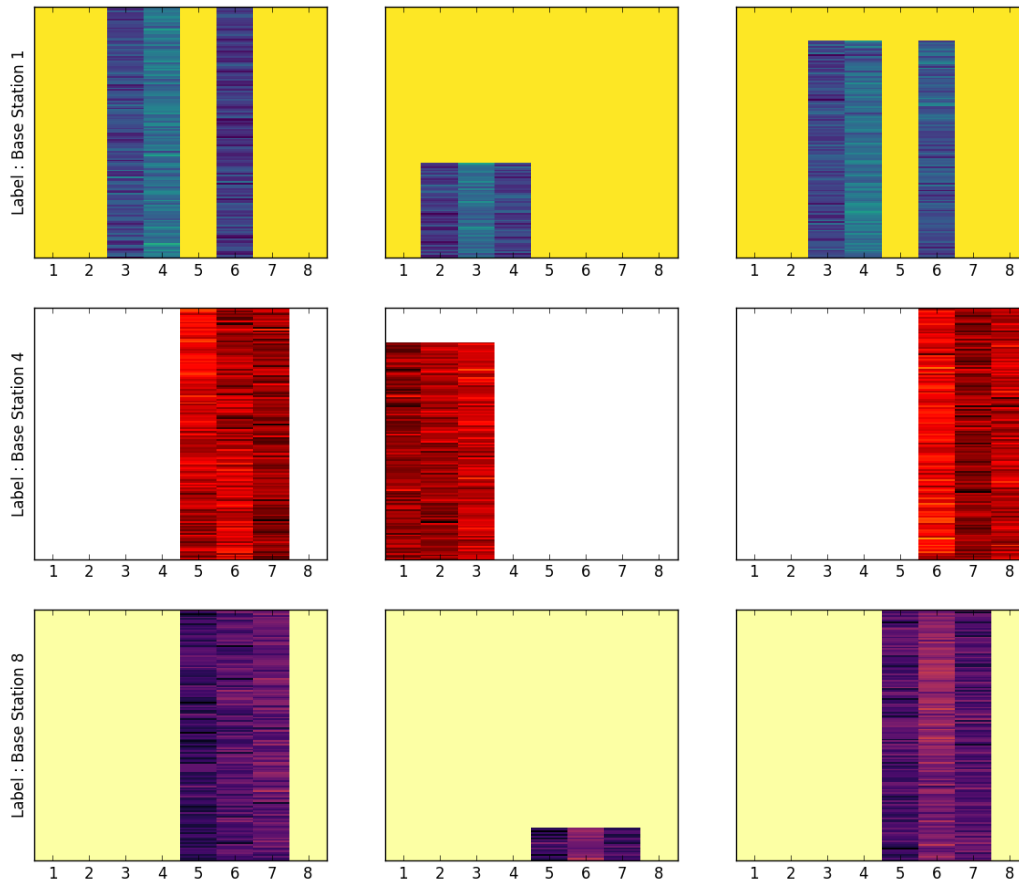
Figure 3.4 Sequences of RSS values leading up to handover to a particular base station. A mobile node is connected to the three closest base stations at any given time instance. The color intensity along a strip indicates signal strength. The label corresponds to the new base station that the mobile device connected to.

graphics cards for general purpose computing. Parallel processing capabilities dramatically speed up computation and enable large networks to be trained in much less time. Here, a CUDA-enabled laptop with an i7 processor, a GeForce GTX 950M graphics card and 16 GB RAM took approximately 35 minutes to train the neural network with the number of epochs set to 100. Training the model needs to be conducted offline and, once trained, predictions can be made very rapidly.

Tensorflow gradually learns the weights of the neurons by penalizing a measure based on cross entropy. The convergence of cross entropy with epoch number is shown in Fig. 3.5. A maximum classification accuracy of 98.2% was obtained in this particular classification problem having eight classes.

24

At the conclusion of each epoch, classes were predicted for the 30,000 unseen sequences in approximately 2.15 s.

3.6 Discussion

Classification accuracy was high for the base station prediction problem described here. Although a relatively simple base station association rule was used in this simulation, more complex assignments are also possible. Additional constraints such as load balancing and different quality of service levels for different customers can be incorporated in future to simulate a more realistic 5G scenario. Predictions can also be made very quickly and this demonstrates how a RNN can be used for proactive mobility management even with strict time requirements, as would be the case for self-driving cars.
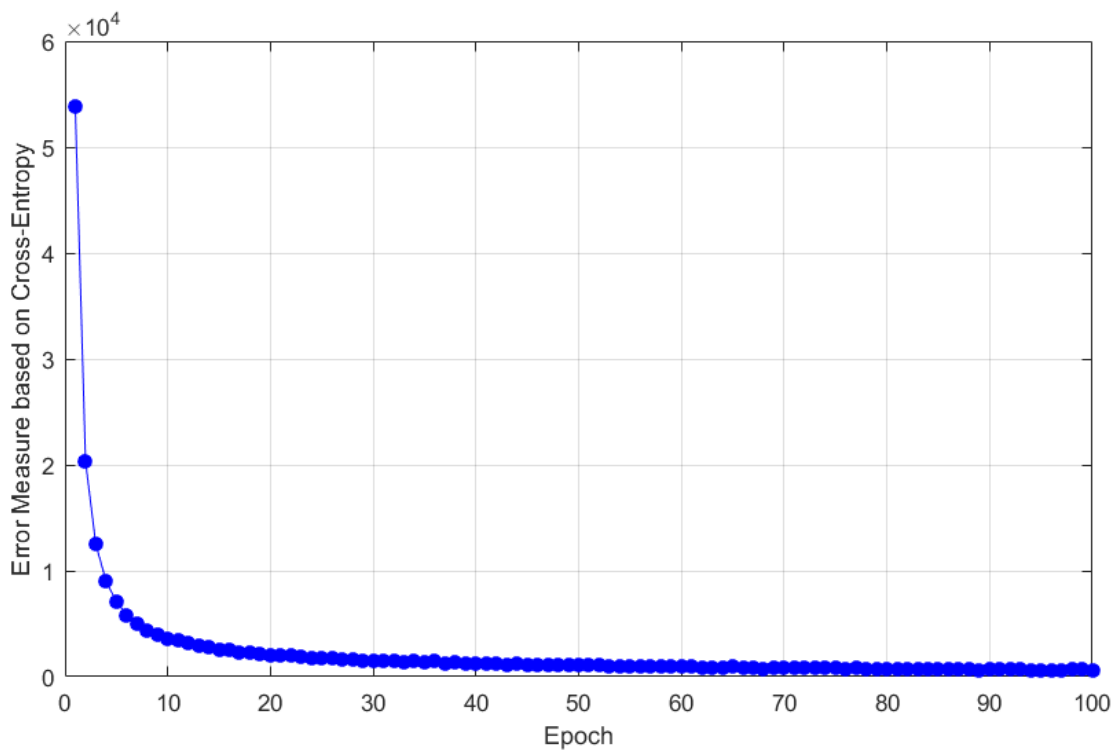

Figure 3.5 Convergence of model training cost calculated based on cross entropy.

CHAPTER 4: CONCLUSION

This thesis presented techniques for applying predictive analytics to cardiac healthcare and 5G cellular networks. The specific details related to the methodology and prior art were reviewed in previous chapters. This chapter briefly summarizes what was described earlier and provides some concluding remarks.

4.1 Atrial Fibrillation Prediction

The AF prediction problem was first posed in the Computers in Cardiology Challenge 2001 in order to further research in predictive analytics and develop better pacing therapies. As part of the second half of the challenge, competitors had to come up with different types of features and ML algorithms to classify EKG epochs that were presented to them in pairs, one of which just preceded an AF episode and the other came from a location far away from such an episode. The types of features explored could broadly be categorized into two – features related to heart rate and features related to P-wave morphology. Subsequent researchers have, since then, revisited the same problem with the objective of improving on the classification accuracy that was achieved by the winner in 2001.

In this thesis, a patient-specific approach was developed for classifying 2-minute epochs of EKG into the same two classes. A different dataset from the one provided in 2001 was used for this purpose. Features based on heart rate variability, AR coefficients and ectopic beats were extracted from EKG epochs located just prior to AF episodes (for different prediction horizons) and far away from those abnormal rhythms. After selecting the best set of features for each patient, SVMs were trained for classification. Sensitivity, specificity and accuracy were calculated using 10-fold cross-validation. Sensitivity was highest for a 0-minute prediction horizon and tended to decrease as the prediction horizon increased. A class imbalance favored the majority category and caused specificity to remain high. It is likely that as the gap

26

between the actual fibrillatory rhythm and the termination of the EKG epoch under consideration increases, there is less predictive information that an AF episode is about to occur.

The high accuracy achieved using this patient-specific approach demonstrates the disruptive market potential that wearable medical monitors might have in the future. Patients can avoid frequent hospital visits due to remote monitoring and could potentially be provided with warnings of oncoming cardiac events. This might include a condition such as a heart attack. A challenge remains for engineers to develop commercial low-power devices that are as unobtrusive as possible to the daily routine of a patient, while at the same time acquiring diagnostic quality signals for subsequent ML analysis. A possible future direction of work would include investigating P-wave features and incorporating them into the algorithm presented here for an improvement in accuracy.

4.2 Base Station Prediction

5G cellular networks sought to address three major concerns – a much higher data rate, low latency communication, and serving the Internet of Things. The time taken for handover has been one of the factors contributing to latency. Among solutions, such as better protocol design, researchers have also sought to use ML to predict future base station association for a mobile device. Some of the approaches have relied on extracting the path traveled by a customer using GPS coordinates in order to make these predictions. Others have relied on the sequence of previously visited cells and the time spent in them for developing classification algorithms or graph-based approaches for making a prediction. However, users can be reluctant to continuously transmit their GPS coordinates due to privacy concerns and device power consumption. Moreover, the idea of a cell might no longer be clearly defined in a 5G network, since a user might be receiving data from multiple base stations or access points simultaneously. Given the unwillingness of users to share their GPS trajectories, the fuzzy definition of a cell and the severe time constraints a future cellular network will have to operate under, the question still remains – how can decisions regarding handover be made proactively? Is it possible to predict base station association using some other means?

27

This thesis presented a method to predict future base station association by employing a RNN to learn patterns in sequences of RSS values. A MATLAB simulation was set up and data was recorded from users as they moved around within a grid-like road network. A RNN was trained using 70,000 RSS sequences and tested on 30,000 sequences. Accuracy was shown to be above 98% in this eight-class classification problem. Future work might include an extension to solving a similar problem when users are also able to move above the ground in aerial vehicles.

REFERENCES

[1] G. Appelboom, E. Camacho, M. E. Abraham, S. S. Bruce, E. L. P. Dumont, B. E. Zacharia, R. D'Amico, J. Slomian, J. Y. Reginster, O. Bruyère and E. S. Connolly Jr, "Smart Wearable Body Sensors for Patient Self-assessment and Monitoring," *Archives of Public Health,* vol. 72, no. 28, 2014.

[2] E. J. Benjamin, M. J. Blaha, S. E. Chiuve, M. Cushman, S. R. Das, R. Deo, S. D. de Ferranti, J. Floyd, M. Fornage, C. Gillespie, C. R. Isasi, M. C. Jiménez, L. C. Jordan, S. E. Judd and D. Lackland, J. H. Lichtman, L. Lisabeth, S. Liu, C. T. Longenecker, R. H. Mackey, K. Matsushita, D. Mozaffarian, M. E. Mussolino, K. Nasir, R. W. Neumar, L. Palaniappan, D. K. Pandey, R. R. Thiagarajan, M. J. Reeves, M. Ritchey, C. J. Rodriguez, G. A. Roth, W. D. Rosamond, C. Sasson, A. Towfighi, C. W. Tsao, M. B. Turner, S. S. Virani, J. H. Voeks, J. Z. Willey, J. T. Wilkins, J. H. Y. Wu, H. M. Alger, S. S. Wong, P. Muntner, "Heart Disease and Stroke Statistics - 2017 Update: A Report from the American Heart Association," *Circulation,* vol. 135, pp. e1-e458, 2017.

[3] D. E. Singer, G. W. Albers, J. E. Dalen, A. S. Go, J. L. Halperin and W. J. Manning, "Antithrombotic Therapy in Atrial Fibrillation: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy," *Chest,* vol. 126, no. 3, pp. 429S-456S, 2004.

[4] C. T. January, L. S. Wann, J. S. Alpert, H. Calkins, J. C. Cleveland, J. E. Cigarroa, J. B. Conti, P. T. Ellinor, M. D. Ezekowitz, M. E. Field, K. T. Murray, R. L. Sacco, W. G. Stevenson, P. J. Tchou, C. M. Tracy and C. W. Yancy, "2014 AHA/ACC/HRS Guideline for the Management of Patients with Atrial Fibrillation: a Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society," *Circulation,* vol. 129, 2014.

[5] E. J. Benjamin, D. Levy, S. M. Vaziri, R. B. D'Agostino, A. J. Belanger and P. A. Wolf, "Independent Risk Factors for Atrial Fibrillation in a Population-Based Cohort: The Framingham Heart Study," *JAMA,* vol. 271, no. 11, pp. 840-844, 1994.

[6] P. J. Lindsberg, L. Toivonen and H.-C. Diener, "The Atrial Fibrillation Epidemic is Approaching the Physician's Door: Will Mobile Technology Improve Detection?," *BMC Medicine,* vol. 12, no. 180, 2014.

[7] G. B. Moody, A. L. Goldberger, S. McClennen and S. P. Swiryn, "Predicting the Onset of Paroxysmal Atrial Fibrillation: The Computers in Cardiology Challenge 2001," *Computers in Cardiology,* vol. 28, pp. 113-116, 2001.

[8] E. Shim, H.-y. Wei, Y. Chang and R. D. Gitlin, "Low Latency Handoff for Wireless IP QoS with NeighborCasting," in *IEEE International Conference on Communications*, 2002.

[9] Q. Lv, Y. Qiao, N. Ansari, J. Liu and J. Yang, "Big Data Driven Hidden Markov model based Individual Mobility Prediction at Points of Interest," *IEEE Transactions on Vehicular Technology,* to be published.

[10] M. A. Hamburg and F. S. Collins, "The Path to Personalized Medicine," *New England Journal of Medicine,* vol. 363, no. 4, pp. 301-304, 2010.

[11] K. Hnatkova, J. E. P. Waktare, F. D. Murgatroyd, X. Guo, X. Baiyan, A. J. Camm and M. Malik, "Analysis of the Cardiac Rhythm Preceding Episodes of Paroxysmal Atrial Fibrillation," *American Heart Journal,* vol. 135, no. 6, p. 1010–1019, 1998.

[12] D. S. Wickramasuriya, C. A. Perumalla and R. D. Gitlin, "Predicting Episodes of Atrial Fibrillation using RR-Intervals and Ectopic Beats," in *IEEE International Conference on Biomedical and Health Informatics*, 2017.

[13] K. S. Lynn and H. D. Chiang, "A Two-stage Solution Algorithm for Paroxysmal Atrial Fibrillation Prediction," *Computers in Cardiology,* vol. 28, p. 405–407, 2001.

[14] P. Langley, D. di Bernado, J. Allen, E. Bowers, F. E. Smith, S. Vecchietti and A. Murray, "Can Paroxysmal Atrial Fibrillation be Predicted?," *Computers in Cardiology ,* vol. 28, p. 121–124, 2001.

[15] W. Zong, R. Mukkamala and R. G. Mark, "A Methodology for Predicting Paroxysmal Atrial Fibrillation based on ECG Arrhythmia Feature Analysis," *Computers in Cardiology,* vol. 28, p. 125–128, 2001.

[16] C. Maier, M. Bauch and H. Dickhaus, "Screening and Prediction of Paroxysmal Atrial Fibrillation by Analysis of Heart Rate Variability Parameters," *Computers in Cardiology,* vol. 28, p. 129–132, 2001.

[17] G. Schreier, P. Kastner and W. Marko, "An Automatic ECG Processing Algorithm to Identify Patients Prone to Paroxysmal Atrial Fibrillation," *Computers in Cardiology,* vol. 28, p. 133–135, 2001.

[18] P. de Chazal and C. Heneghan, "Automated Assessment of Atrial Fibrillation," *Computers in Cardiology,* vol. 28, p. 117–120, 2001.

[19] T. Thong, J. McNames, M. Aboy and B. Goldstein, "Prediction of Atrial Fibrillation by Analysis of Atrial Premature Complexes," *IEEE Transactions on Biomedical Engineering,* vol. 51, no. 4, p. 561–569, 2004.

[20] I. Hilavin and M. Kuntalp, "Prediction of Paroxysmal Atrial Fibrillation Onset by Using ECG," in *International Symposium on Innovations in Intelligent Systems and Applications*, 2012.

[21] D. Arotaritei and C. Rotariu, "Automatic Prediction of Paroxysmal Atrial Fibrillation in Patients with Heart Arrhythmia," in *International Conference and Exposition on Electrical and Power Engineering*, 2014.

[22] Y. V. Chesnokov, A. V. Holden and H. Zhang, "Distant Prediction of Paroxysmal Atrial Fibrillation using HRV Data Analysis," *Computers in Cardiology,* vol. 34, p. 455–458, 2007.

[23] G. U. Yule, "On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences,* vol. 226, no. 636-646, pp. 267-298, 1927.

[24] G. Walker, "On Periodicity in Series of Related Terms," *Proceedings of the Royal Society of London: Series A, Containing Papers of a Mathematical and Physical Character,* vol. 131, no. 818, p. 518–532, 1931.

[25] K. Padmavathi and K. S. Ramakrishna, "Classification of ECG Signal during Atrial Fibrillation using Autoregressive Modeling," *Procedia Computer Science,* vol. 46, pp. 53-59, 2015.

[26] H. B. Mann and D. R. Whitney, "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other," *The Annals of Mathematical Statistics,* vol. 18, no. 1, pp. 50-60, 1947.

[27] K. Pearson, "On Lines and Planes of Closest Fit to Systems of Points in Space," *Philosophical Magazine Series 6,* vol. 2, no. 11, pp. 559-572, 1901.

[28] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning,* vol. 20, no. 3, p. 273–297, 1995 .

[29] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, "The WEKA Data Mining Software: an Update," *ACM SIGKDD Explorations Newsletter,* vol. 11, no. 1, pp. 10-18, 2009.

[30] D. S. Wickramasuriya, C. A. Perumalla, K. Davaslioglu and R. D. Gitlin, "Base Station Prediction and Proactive Mobility Management in Virtual Cells using Recurrent Neural Networks," in *IEEE Wireless and Microwave Technology Conference*, 2017.

[31] S. Michaelis, N. Piatkowski and K. Morik, "Predicting Next Network Cell IDs for Moving Users with Discriminative and Generative Models," in *Mobile Data Challenge Workshop (Nokia) in conjunction with International Conference on Pervasive Computing*, 2012.

[32] J. Tkacik and P. Kordik, "Neural Turing Machine for Sequential Learning of Human Mobility Patterns," in *International Joint Conference on Neural Networks*, 2016.

[33] Y. Luo, P. N. Tran, C. An, J. Eymann, L. Kreft and A. Timm-Giel, "A Novel Handover Prediction Scheme in Content Centric Networking using Nonlinear Autoregressive Exogenous Model," in *IEEE Vehicular Technology Conference*, 2013.

[34] U. Javed, D. Han, R. Caceres, J. Pang, S. Seshan and A. Varshavsky, "Predicting Handoffs in 3G Networks," *ACM SIGOPS Operating Systems Review,* vol. 45, no. 3, pp. 65-70, 2011.

[35] W. Iba and P. Langley, "Induction of One-Level Decision Trees," in *International Conference on Machine Learning*, 1992.

[36] S. Liou and Y. Huang, "Trajectory Predictions in Mobile Networks," *International Journal of Information Technology,* vol. 11, no. 11, pp. 109-122, 2005.

[37] K. Laasonen, "Clustering and Prediction of Mobile User Routes from Cellular Data," in *9th European Conference on Principles and Practice of Knowledge Discovery in Databases*, 2005.

[38] T. Anagnostopoulos, C. Anagnostopoulos, S. Hadjiefthymiades, M. Kyriakakos and A. Kalousis, "Predicting the Location of Mobile Users: a Machine Learning Approach," in *International Conference on Pervasive Services*, 2009.

[39] T. Anagnostopoulos, C. B. Anagnostopoulos, S. Hadjiefthymiades, A. Kalousis and M. Kyriakakos, "Path Prediction through Data Mining," in *IEEE International Conference on Pervasive Services*, 2007.

[40] K. Davaslioglu and E. Ayanoglu, "Interference-based Cell Selection in Heterogeneous Networks," in *Information Theory and Applications Workshop*, San Diego, 2013.

[41] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation,* vol. 9, no. 8, pp. 1735-1780, 1997.

[42] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, R. Jozefowicz, Y. Jia, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, M. Schuster, R. Monga, S. Moore, D. Murray, C. Olah, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, *TensorFlow: Large-scale Machine Learning on Heterogeneous Systems,* 2015.

[43] D. Hafner, "Introduction to Recurrent Networks in TensorFlow," 5 May 2016. [Online]. Available: https://danijar.com/introduction-to-recurrent-networks-in-tensorflow/. [Accessed 21 April 2017].

APPENDIX A: COPYRIGHT PERMSSION

Shown below is the permission to use material for chapter 2 and 3 in this thesis.

- **Does IEEE require individuals working on a thesis or dissertation to obtain formal permission for reuse?**

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you must follow the requirements listed below:

Textual Material

Using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.

In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.

If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Full-Text Article

If you are using the entire IEEE copyright owned article, the following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]

Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.

In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.