Machine learning enabled ECG monitoring for early detection of hyperkalaemia

Farrell, Constance; Shakir, Muhammad Zeeshan

Published in:
AI for Emerging Verticals

Accepted/In press: 10/04/2020

Document Version
Peer reviewed version

Chapter 14

Machine Learning enabled ECG Monitoring for Early Detection of Hyperkalaemia

Constance Farrell[1] *and Muhammad Zeeshan Shakir*[1]

Hyperkalaemia is the medical terminology for a blood potassium level above normal parameters (greater than 5.5 millimoles). This can have a variety of causes, however, patients with significant renal impairment/disease are particularly at a high risk as their kidneys are compromised and unable to filter out excess potassium from the bloodstream. Hyperkalaemia is a medical emergency and requires urgent medical intervention, if left untreated there is an extremely high risk of cardiac arrest and death as high potassium levels directly affect the electrical activity of the heart. Cardiac activity can be observed by performing an Electrocardiogram (ECG) test which is routinely used in all clinical areas worldwide. This research investigates the use of ECG tests as a diagnostic tool for the early detection of hyperkalaemia, and the role of machine learning in the prediction of blood potassium levels from ECG data alone. Support Vector Machines, k-Nearest Neighbour, Decision Tree and Gaussian Naive Bayes classifiers were used comparatively to classify ECG data as either "normokalaemia" or "hyperkalaemia". Results showed that the Decision Tree model performed the best, achieving a 90.9 percent predictive accuracy.

14.1 Introduction

Current research from Kidney Care UK [1] approximates there to be around three million people in the United Kingdom (UK) living with renal/kidney disease at varying degrees of severity. All of which are considered to be at a high risk of developing hyperkalaemia and would benefit greatly from remote testing and early diagnosis. Patients with renal impairment also have a highly increased risk of developing other additional health problems in other anatomical systems, including; heart failure, high blood pressure, electrolyte imbalance, metabolic acidosis, among others [2]. Certain individuals may also develop hyperkalaemia without the presence of renal impairment, however, those with kidney disease are the most at risk due to the high incidence of occurrence and the difficulty of managing fluctuating potassium levels in renal patients [3].

[1]School of Computing, Engineering and Physical Sciences, University of the West of Scotland, Paisley PA1 2BE, Scotland, UK

Potassium plays a vital role in maintaining electrical charge within the heart which is essential for the heart to beat regularly. Fluctuations in blood potassium levels cause disruption to the timing and rhythm of contractions within the heart which has a detrimental effect, resulting in arrythmias and irregularities. The severity and risk of sudden cardiac arrest progressively worsens as the potassium level increases, hospital admission and urgent treatment is required for hyperkalaemia.

At present, the only way to detect and diagnose hyperkalaemia is by obtaining venous (or arterial) blood and performing a laboratory test which can only be done by suitably qualified professionals in a healthcare environment. By the time patients decide to seek medical attention, symptoms tend to have developed and progressed significantly with potassium levels dangerously high [4]. This explains the high number of instances of cardiac arrests associated with hyperkalaemia as symptoms are not always present, especially in mild-moderate stages. A method of remote testing potassium levels would enable earlier detection of hyperkalaemia which means patients could seek treatment sooner, maximising the chances of a good outcome and reducing the risk of sudden cardiac arrest [5].

Remote patient monitoring is an emerging field of technologies that are currently being developed to support the management of chronic health conditions in the community. ECG tests are traditionally performed in a clinical setting by a clinician who is able to interpret the findings of the test. Advancements in technology have enabled the development of mobile sensors that are capable of capturing an ECG from a handheld smartphone device [6]. The data from the ECG is then stored on the device awaiting review from a clinician. An application that is able to interpret the ECG results almost instantaneously through machine learning would enable the patient to seek treatment faster and improve patient safety.

Blood potassium levels cannot be diagnosed from an ECG test alone at present, however, with the aid of technology there is huge potential for this to be developed. Certain ECG trends and changes are exhibited and associated with varying stages of hyperkalaemia thus an estimate of blood potassium level can be derived from an ECG test [7].

This research investigates the possible uses of machine learning in the identification of hyperkalaemia related changes to electrical activity within the heart and its accuracy in predicting blood potassium values from ECG test output. Machine learning expands on existing statistical techniques and can accommodate a broad array of data types enabling the production of results in more complex situations [8]. It is a rapidly expanding field and is already being used in many areas within the healthcare sector. Modern electronic health records generate a vast amount of data and machine learning can assist in making sense of this data by identifying patterns and producing clinically valuable statistical information. This research will focus on exploring supervised learning techniques using existing ECG data which is publicly available online.

Early identification and diagnosis of hyperkalaemia is imperative for the timely initiation of treatment and is essential for the prevention of sudden cardiac arrest. The mortality rate is heightened in patients with pre-existing kidney disease and other co-morbidities. Long, Warix and Koyfman [9] emphasise the importance of rapid diagnosis and management of hyperkalaemia as the rate at which it accrues impacts the severity of negative effects on the physiological systems of the body.

A blood test is currently the gold standard of clinical test for blood potassium level detection [10]. However, this requires laboratory testing which is invasive for the patient, time consuming and can result in a delay in treatment. There is a definite need for remote-monitoring methodologies capable of detecting abnormal potassium levels before the patient becomes symptomatic [11]. The potential of ECG testing for potassium level detection in the community setting has attracted attention from researchers, from both technological and clinical backgrounds.

14.1.1 ECG for Chronic kidney disease

ECG tests are traditionally performed in a clinical setting by a suitably qualified professional using 12 adhesive electrodes and 12 leads which connect to a portable machine that displays and prints the output. Sabiullah et al. [12] found hyperkalaemia to be a major threat to patient safety in Chronic kidney disease (CKD), with the severity of threat increasing as kidney function deteriorates. They found that the mortality rate due to cardiac arrest was significantly higher in CKD patients. They examined the ECG changes that occur as blood potassium levels increase and found a distinctive correlating pattern. Research by Viera and Wouk [13] observed similar ECG patterns in patients as potassium levels increased. The earliest initial sign is peaked T waves, followed by loss of P waves and then widening of the QRS complex and sine waves [14]. Typical hyperkalaemia induced arrythmias include; sinus bradycardia, sinus arrest, ventricular tachycardia and ventricular fibrillation [15].

There have been recent developments in the design of ECG capturing technologies. Attia et al. [16] conducted research in to the abilities of a single-lead ECG to calculate blood potassium values without requiring a blood and laboratory test. Their study used a population of haemodialysis patients of which 66 percent were men and 34 percent were female, a fairly even distribution. A data-processing algorithm was created using the Matlab environment (MathWorks) to estimate potassium levels based on T wave characteristics present on the single-lead ECG test. Calculated potassium values from the ECG were then compared with actual values from patients' blood tests which was found to have a high level of accuracy with a mean absolute error of 10 percent. This research was highly significant in identifying the potential of remote ECG testing as a method of calculating blood potassium levels.

Yasin et al. [17] further expanded on this concept and tested the potassium detection abilities of a single-lead ECG using electrodes attached to a handheld smartphone. They also compared the results with potassium values from blood tests taken at the same time and achieved a high level of accuracy with a reduction in the mean absolute error to 9 percent. Although their algorithm can be considered as successful in their research, the population used to test it only consisted of 18 patients all with potassium levels in the normal range (3.5 – 5.5 mmols). Therefore, this cannot presently be applied in the detection of hypo/hyperkalaemia as there is no evidence to suggest or prove that the algorithm is capable of detecting values lower or greater than the normal range.

A breakthrough in the real-life implementation of mobile healthcare technologies was made in 2012 when the American Food and Drug Administration (FDA) approved the use of the Kardia smartphone case with ECG capturing sensors by medical device and artificial intelligence company, AliveCor. This works by the user placing their fingers on the sensors attached to the smartphone case which obtains an ECG, the data is then stored offline on an application installed on the mobile device [18]. Various tests showed the smartphone ECG performed excellently and had a strong correlation with the traditional 12-lead ECG used in hospitals making it a viable option for remote monitoring [19].

Since 2012, many developments have been made to AliveCor's mobile ECG. It is currently compatible with most iPhones and Android smartphones [20] ECG data is available for clinical practitioners to review at the users request with data being transmitted over WiFi, LTE or 4G connectivity [21]. The mobile ECG has been shown to produce highly accurate results but still requires analysis and interpretation by a clinician for the detection of abnormalities [22]. However, research into the interpretative and diagnostic potential of remote ECG applications is ongoing.

14.1.2 Non-invasive methods for Potassium Detection

Advances in healthcare designed to assist in the remote monitoring of physiological signs include wearable sensor technologies and mobile devices that generate large amounts of data [23]. Machine learning has the potential to revolutionise the way in which the data contained in these devices is analysed and used by clinicians. Several machine learning algorithms have been used for ECG classification such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Naïve Bayes and and k-Nearest Neighbour (k-NN) models [24, 25, 26, 27]. The vast majority of current research into ECG data analysis and classification tends to focus on the identification and/or prediction of cardiac abnormalities and disease.

Although indicative of the potential diagnostic role machine learning can play with ECG data, the technological techniques used for the detection of cardiac issues cannot automatically be applied to the detection of hyperkalaemia as parameters differ greatly. ECG changes specifically related to high potassium levels require a unique model designed exactly for that purpose such as the algorithm proposed by Attia et al. [16]. Their algorithm was able to detect potassium levels with a 12 percent mean error (approximately 0.5 mmols) in a small sample of stable haemodialysis patients. Their ECG analysis focused solely on T-wave peak and amplitude changes with varying potassium levels at different periods of time. Findings suggest that further research is required to improve the level of accuracy and generalisability, as the algorithm is unable to work for anyone with any form of cardiac disease which is very common among renal patients. Hadjem and Naït-Abdesselam [28] tested the accuracy of 7 different supervised learning classification models in detecting T-wave changes in ECG test data. Results showed that the decision tree model performed best with high accuracy (92.5 percent) and low error rate (7.4 percent).

Galloway et al. [29] investigated the abilities of ECG tests to detect hyperkalaemia but used a different method of testing, a deep-learning classification model and found results which contrasted the findings by Rafique et al. [30]. In this research, 1.5 million ECGs from around 450,000 patients were used to train a deep neural network. Results were promising as they showed that the deep learning model was able to detect hyperkalaemia with a high sensitivity and a high area under the curve (between 0.853 and 0.901). This highlights the potential effective use of applying machine learning and artificial intelligence in the identification of hyperkalaemia by ECG data analysis and requires further investigation.

Research by Velagapudi et al. [31] found that incorporation of the QRS complex as well as T wave analytics improved the diagnostic performance of their classification model in the identification of hyperkalaemia from raw ECG data. The data used in the study was obtained from 12-lead ECGs and whilst their model may be useful in a clinical setting it cannot be applied to the single-lead ECGs used in remote health monitoring without further research. However, other researchers such as Littmann and Gibbs have also found that QRS widening and axis shift is highly indicative of extreme hyperkalaemia and should be included in diagnostic modelling [32].

A non-invasive method for potassium detection was proposed by Dillon et al. [33] in which they developed a mathematical formula to calculate blood potassium levels from ECG data. The formula was based on signal-averaged T-wave characteristics identified from extracting features. The features extracted were all T-wave data and no other peaks in the signals were included in the development of the formula. A very small sample of participants was used (12) and the researchers found that even small changes were detectable. Yasin et al. [17] used a linear regression model to calculate blood potassium values from raw ECG data however, their model is person-specific and must be altered for each individual so therefore cannot be generalised. A summary of key research in this field is provided in table 14.1.

Pilia et al. highlighted the potential abilities of an artificial neural network to predict blood potassium concentrations from ECG features [34]. The neural network consisted of one layer and six neurons and the researchers state that a small predictive error margin was achieved. However, this research is completely theoretical with the ECG data simulated in a laboratory setting thus making it vulnerable to researcher bias. The model has not been tested with any real clinical data and lacks ecological validity. Neural networks have been found to perform well in the detection and diagnosis of cardiac abnormalities/arrythmias [25],[35] and could be a potentially useful method for the detection of hyperkalaemia from ECG data. This is currently an under-developed area and may benefit from further research in future.

This research differs from existing bloodless potassium-predicting technologies in that it incorporates all PQRST peak patterns and changes in the classification stage whereas the current methodologies are largely solely focused on T-waves only. Some of the current proposed methods involve deriving mathematical equations on an individual patient-to-patient basis [16],[17]. An algorithm can therefore not be generalised and individualised programmes for each patient would be largely time consuming and unrealistic. This research aims to develop a model that can be applied to all patients, is non-specific and can be used remotely by all.

Table 14.1 Normal ECG feature parameters

Signal Feature	Duration(s)	Cardiac Function
PR Interval	0.12 - 0.20	Atrial Depolarisation
QRS Complex	0.06 - 0.10	Ventricular Depolarisation
ST Segment	$0 < 0.43$	Complete Ventricular Depolarisation
QT Interval	$0 < 0.44$	Total duration of depolarisation and repolarisation
T Wave	0.10 - 0.25	Ventricular Repolarisation
R-R Interval	0.6 - 1.0	Ventricular Rhythm (duration per heartbeat)

14.2 ECG Signal Analysis

ECG tests produce sensor-based data which exists in waveforms, this is significantly more complex to analyse compared with standard numerical or image data. The typical signal produced by a standard heartbeat viewed as normal exhibits a waveform consisting of several peaks identified as, “PQRST” complex. The ECG signal depicted in Figure 14.1 displays the expected pattern of a normally functioning heart. Previous research has shown that every interval and duration of the signal produced by the ECG must be in a very specific range to be considered normal [36]. The typical duration and associated function of each interval is given in Table 14.1. Any disturbance to the electrical conduction of the heart will alter the heartbeat’s pathway and timing. Resultingly, any defect will cause a change in signal produced by the ECG with features out-with the normal PQRST pattern.

Matlab appears to be the most popular programming language and environment used in ECG analysis and modelling research as it has multiple application-programming interfaces (APIs) specifically designed for working with sensor-based data [37]. Azariadi et al. [38] successfully developed an algorithm in Matlab for ECG signal analysis and cardiac arrythmia detection. They used Discrete Wavelet Transform (DWT) for analysis and a Support Vector Machine (SVM) classifier, a binary classifier to label each heartbeat as ‘Normal’ or ‘Abnormal’. The model was trained and tested, producing a 98.9 percent accuracy which suggests this is a highly suitable method for ECG abnormality detection.

Feature extraction is an important aspect of machine learning in the predictive modelling of ECG data, as ‘normal’ peaks must first be identified, from which a rule base can then be established to determine boundaries between normal and abnormal outputs [36]. The detection of the PQRST peak is an important first step in ECG feature extraction as highlighted by Reddy, Vijaya and Suhasini [39]. In their research, Matlab functions are used to derive normal parameter values for PQRST waves which must be precise as any values out-with these are then classified as abnormal. Feature extraction provides the foundation for successful algorithm building in machine learning [40].

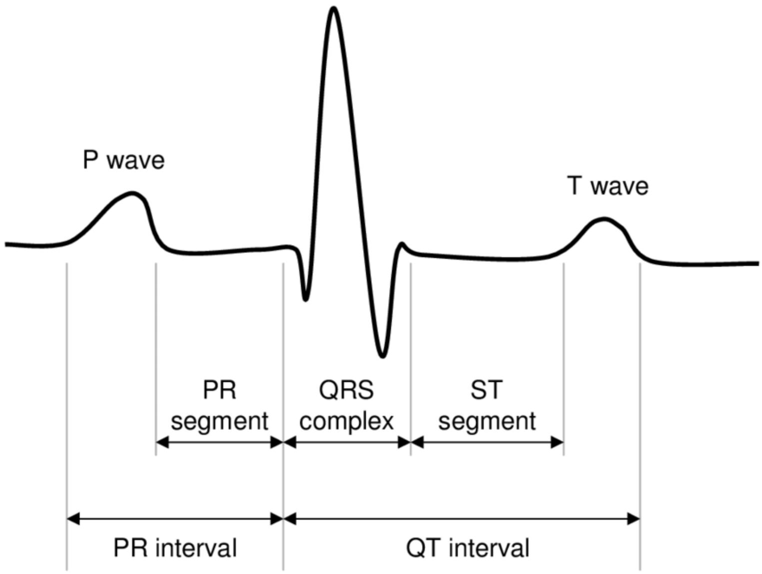

Figure 14.1: A regular ECG signal [41]

14.2.1 ECG Abnormalities caused by Hyperkalaemia

Potassium plays a vital role in maintaining regular electrical charge and contributing to calcium release within the heart, causing it to beat and pump blood around the body. A constant safe level of potassium is required in the blood serum in order to maintain homeostasis within the body. Disruptions to blood potassium levels affect the timing and rhythm of cardiac muscle contractions which can be observed in the signals emitted from an ECG test.

Research has found that hyperkalaemia causes a pattern of distinct ECG abnormalities to occur which progressively worsen in severity as potassium levels increase [12]. The first indication is peaked T-waves which are a clear indication of hyperkalaemia. Previous studies have solely used a change in T-waves as an estimation of potassium level and produced fairly accurate results [16]. However, as potassium levels become further elevated, more life-threatening abnormalities occur which should also be recognisable by any classification model. A prolonged P-R segment is also another indication of mild hyperkalaemia. P-wave flattening and a prolonged QRS complex is indicative that potassium levels are largely elevated. Progressive QRS widening can be seen as levels increase, signals resembling a mathematical sine wave are representative of extreme hyperkalaemia with heart failure imminent without urgent medical attention. These effects are depicted in Figure 14.2.

14.2.2 Feature Extraction and Peak Detection algorithm

The extraction of 'PQRST' features from raw ECG data is an important task as it provides the basis for quantification and modelling. Different feature extraction method-

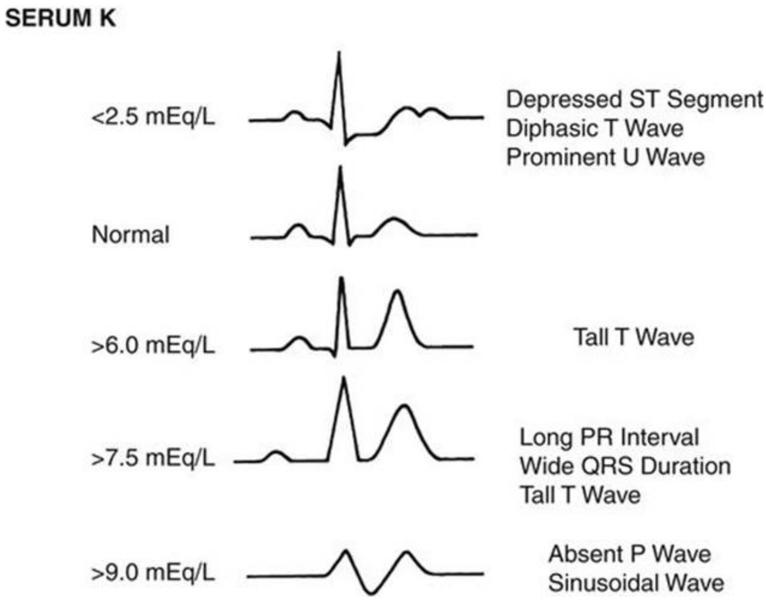

Figure 14.2: Hyperkalaemia related ECG changes [42]

ologies used in previous research as detailed in the literature review section include; genetic algorithms, wavelet transform and non-linear transformation. There have been many 'QRS' detection algorithms successfully developed and implemented in real-world applications. The creation of such algorithms tends to be very complex due to the variability of normal QRS intervals and the presence of different types of noise, such as; artifacts resulting from electrode motion, muscle noise and power-line interference.

An algorithm for feature extraction was developed in MATLAB computing environment due to its' capabilities in working with quantitative data and was deemed suitable for the analysis of raw ECG data. This was based upon the same basic principles as the Pan-Tompkin's algorithm for QRS detection [43]. This is a real-time algorithm that consists of band-pass filter, differentiator, integrator and moving-window [42]. Adaptations have been made and a prototype of the algorithm has been developed to include P and T wave detection which is highly important in the detection of hyperkalaemia.

A simplified prototype algorithm has been developed to detect PQRST peaks as accurately as possible from raw ECG signal data. Band-pass filtering is used for pre-processing, including high-pass and low-pass filters which only allow a certain frequency band to pass through thus helping to eliminate the effects of movement induced noise. MATLAB code is used to calculate and display original signal, signal with low pass filter, signal with high pass filter and the derivative base filter. The output is displayed in Figure. 14.3. Signal filtering at varying frequencies is used to produce a windowed estimate of energy in QRS frequency band and optimise accuracy of peak selection. The imaginary parts in Figure 14.3 illustrate the process of Fourier Analysis - breaking down the signal into individual sine waves for each of the filters used.

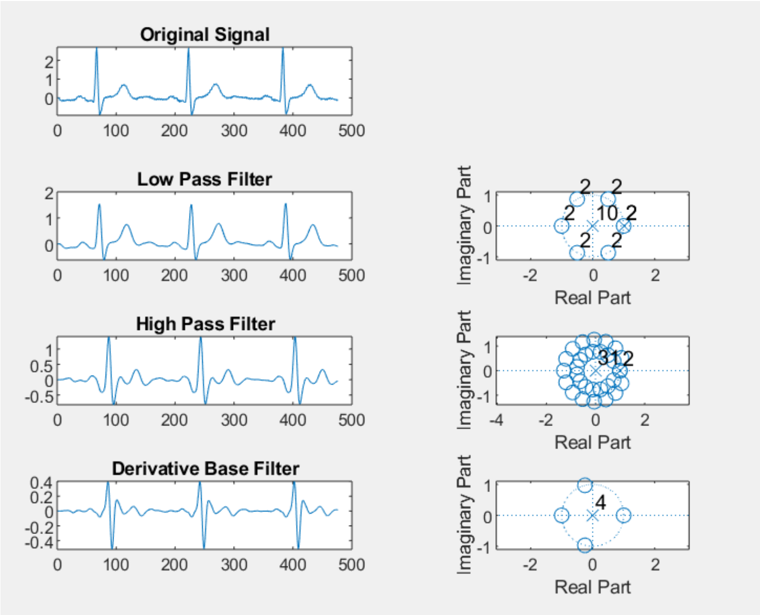

Figure 14.3: ECG signal pre-processing in MATLAB

The filtered signal is derived using a derivative base filter to highlight the PQRST features. The decision rule uses a first derivative based squaring function of the filtered ECG signal to extract features P,Q,R,S and T as can be observed in Figure 14.4. Heart rate can also be calculated by measuring the R-R intervals.

14.3 ECG Data Collection and Pre-processing

To investigate the performance of machine learning models in the prediction of blood potassium levels, a suitable dataset was identified and used as further detailed below. This was done using Python, a high-level programming language and a suitable IDE chosen for its statistical tools and good graphical/visual tools. The data was explored to check for outliers, missing values and irregularities which is fundamental in any data science task to avoid obscure results being produced [44].

Due to ethical reasons and the timescale for the research, it was unfeasible to physically perform ECGs and obtain blood samples from the population. Instead, relevant data that had already been collected for research purposes was used. The dataset used was obtained from the “Electrocardiographic Abnormalities and QTc Interval in Patients Undergoing Haemodialysis” study by Nie et al. and is freely available online from the public library of science [45].

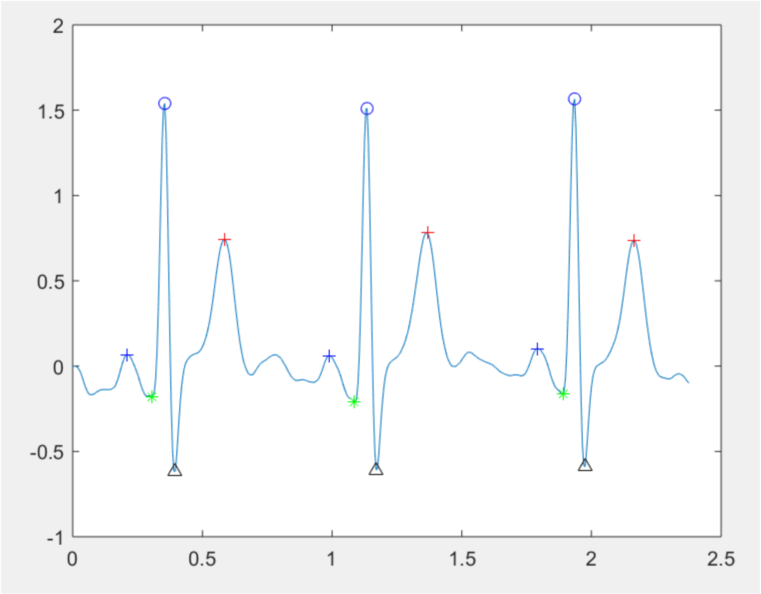

Figure 14.4: PQRST features

The participants consisted of 141 haemodialysis patients of mixed gender. Various clinical information was obtained including ECG test data and blood potassium results which were of particular interest for the purpose of this research. After some participants were excluded from the original study due to known severe cardiac disease, the final dataset published contained 109 rows and 73 columns. A new subset of this dataset was created for the purpose of this research in which the columns were narrowed down to 19 and excluded irrelevant information, the number of rows remined the same as in the original dataset. The numerical data of ECG components after feature extraction had been performed such as; P wave, QRS complex, T wave and the duration between each, were present in the columns as well as; gender, age, heart rate, blood potassium values both pre and post dialysis and the presence of known cardiac defects (RBBB and LBBB).

Data was checked for outliers using scatter plots and any missing values were removed using .dropna() function. The .info() and .describe() functions were used to gain a further insight into the dataset and identify the type of data (e.g. integers, floats etc.). A subset of the original dataset was created as outlined above, containing only data which was directly relevant to this research. Columns were renamed accordingly, and a correlation table was used to identify significant relationships between variables.

14.4 Machine Learning Classification Models

A comparison of machine learning supervised classification models was made to investigate the suitability of their use in predicting hyperkalaemia from ECG data alone. Python programming language and PyCharm IDE were used due to its extensive selection of libraries and evidenced proficiency in machine learning [46]. Models used include; SVM, kNN, Decision Tree and Gaussian Naïve Bayes. These models were chosen as they have all been successfully used in previous research involving the prediction of cardiac disease from ECG data. Although a different medical condition, the findings are transferable due to similarities in identifying/classifying patterns and trends in ECG data.

After data pre-processing and visualisation has been performed, the data was then split into training and testing data. This is a method of cross-validation, a statistical technique that is used to test the performance of the machine learning model and is important in the reduction of selection bias and avoidance of overfitting. The function, 'Xtrain,Xtest,ytrain,ytest=traintestsplit', was used to split data into training and testing sets. The X variable consisted of a combination of attributes attained from ECG analysis including; heart rate, P wave height, P-R duration, QRS complex, QT duration and QTc. Data was then normalised to enable comparison of data from multiple columns within the dataset. The Y variable was the blood potassium level which the model is aiming to predict based on the independent variables listed above. The data was split into 80 percent training and 20 percent for testing.

14.4.1 Support Vector Machines

Support vector machines (SVM) are a supervised learning technique that are capable of both classification and regression. It attempts to classify data by finding a hyperplane that linearly separates data from different classes [47]. The algorithm is given labelled training data and then outputs an optimal plane which categorises new examples. It finds a line/hyperplane in multidimensional space that separates out classes. Linear algebra is used to transform the problem and establish a suitably positioned hyperplane. Different kernel functions use algebraic equations for predicting new inputs. For the linear kernel which was used in this research, the prediction of a new input is made by using the dot product between the input x and each support vector x_i and is calculated with the following equation:

$$f(x) = B(0) + sum(a_i * (x, x_i)) \tag{14.1}$$

This equation involves calculating the inner products of a new input vector x with all of the support vectors in the training dataset. $B(0)$ and a_i are coefficients that must be estimated by the learning algorithm, for each input from the training data. The 'C' parameter from the sklearn library is also known as the regularisation parameter. This informs the SVM optimisation how much to avoid misclassifying each training sample [48]. Higher values tend to produce a decreased occurrence in the instances of misclassification. The SVM algorithm from sklearn library was used with the training and testing data to predict a value for the blood potassium level.

14.4.2 k-Nearest Neighbour

K-Nearest Neighbours (k-NN) is another, simpler supervised machine learning algorithm that can be used in both classification and regression problems [49]. This algorithm involves the assumption that similar things exist in close proximity to one another and bases its classification method around the mathematical calculation of distance between points. It is a non-parametric learning algorithm that separates data points into different classes with the aim of predicting the class of a new sample point. A simplified step-by-step process of the k-NN algorithm is as follows:

Algorithm 1 kNN Classifier

Require: training data (T) and testing data (D)
Ensure: $k \geq 1$ where $k \leftarrow 1, 2, 3 \cdots n$
  Initialise the value of k
  for each data point d in D **do**
    calculate distance (Euclidean/Manhattan) (e) between d and all data point $t \in T$
    store e in list and sort in ascending order based on the distance values
    select k-nearest points according to k minimum distances.
  end for
  take the commonest class
  return the anticipated class

Hu et al. [50] found k-NN to be useful in their research with EEG signal data, their findings can be applied in this research as the raw data is similar. The k-NN algorithm was imported from the sklearn library and training/testing data used to classify new input values as either being hypokalaemia, normokalaemia or hyperkalaemia.

14.4.3 Decision Tree Model

In computational complexity theories, the decision tree model is a model that essentially incorporates a sequence of branching operations based upon comparisons of some quantities which are assigned the unit computational cost. They develop classification systems that are capable of forecasting future observations based on a set of pre-determined decision rules. The decision tree model is composed of nodes and edges, starting from a root node with no incoming edge. Instances are navigated from the root of the tree according to the decisions made by internal nodes and then classified [51].

Verma, Srivastava and Negi [53] successfully used a decision tree classifier in the prediction of cardiac arrythmias using ECG signal data similar to that used in this research. Results showed the model to have an average predictive accuracy of 88.4 percent which is significant enough to suggest potential usage in real-life implementation. Kasar and Joshi [54] also found that characteristics extracted from raw ECG signal data and feature vectors were compatible for use with their decision tree algorithm in the identification of cardiac disease. Evaluation showed their algorithm

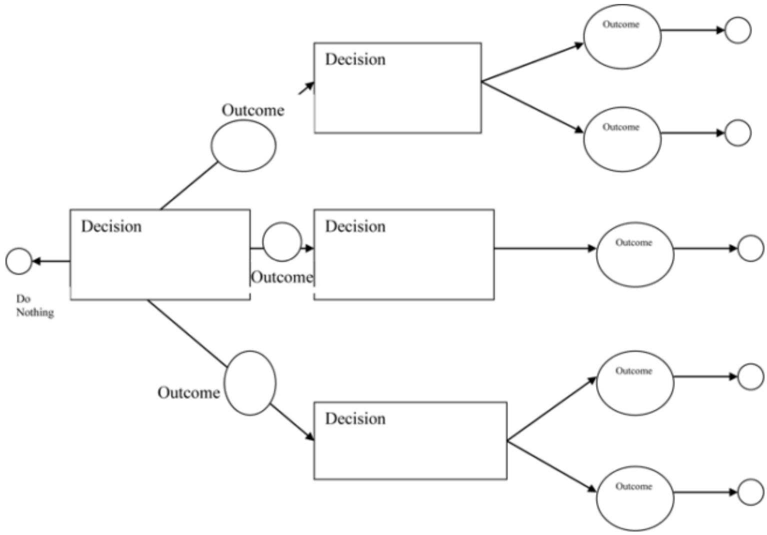

Figure 14.5: Decision Tree Model[52]

to have an average classification accuracy of 92.5 percent which is relatively high, especially when compared with other classification models.

Existing research that uses decision tree models to classify ECG signal data is largely focused on the identification/prediction of cardiac disease/abnormalities, there is little application in the diagnosis of hyperkalaemia from ECG data alone. As discussed in the literature review, Hadjem and Naït-Abdesselam [28] used a decision tree technique to identify T-wave ECG abnormalities from ECG data after feature extraction had been performed. Results showed the decision tree model to have an overall accuracy of 92.5 percent and a 7.5 percent error rate. These findings are applicable to this research as the detection of T-waves is highly significant in the identification of hyperkalaemia. Expanding on this, the pre-designed decision tree classifier was imported from the sklearn library using Python and used with multiple ECG features to enable a more precise prediction.

The use of T-waves alone as an indicator of blood potassium level can lead to error and misdiagnosis as there are other possible causes for T-wave distortion/elevation [55]. Additional ECG changes indicative of hyperkalaemia were included in this research as well as T-wave peak height to enable a more accurate prediction and avoid misdiagnosis. The decision tree model was able to accommodate for the additional variables.

14.4.4 Gaussian Naïve Bayes model

Naïve Bayes is a probabilistic machine learning algorithm that is based on Bayes Theorem (eqn. 14.2) and can be used in multi-class classification problems. This approach is based the simplistic hypothesis in that it assumes the presence/absence of a particular feature of a class is unrelated to the presence/absence of any of the other features involved, it follows the principle of conditional probability.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{14.2}$$

where $P(A|B)$ denotes the probability of A given B, $P(B|A)$ denotes the probability of B given A, $P(A)$ denotes the probability of A occurring and $P(B)$ denotes the probability of A occurring. The Gaussian Naïve Bayes model is the simplest Naïve Bayes classifier with the assumption that the features in the dataset are normally distributed. The mean μ and standard deviation σ are estimated from the training data input values x for each class to summarise the distribution. Probabilities of new input x values are then calculated by using the Gaussian Probability Density Function (PDF):

$$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp - \frac{x-\mu^2}{2\sigma^2} \tag{14.3}$$

Naïve Bayes classifiers have been used in previous research involving automated ECG data interpretation and abnormality detection like that used by Bayasi et al. [56] when they looked at predicting ventricular arrythmias from ECG signal data. They used a Naïve Bayes classifier after feature selection had been performed, results showed the model to have an overall accuracy of 86 percent when trained and tested. Sannino and De Pietro [57] also found a Naïve Bayes classifier to perform well in their research which was heartbeat detection and classification from ECG data which could be transferable in the diagnosis of hyperkalaemia due to similarities. As with the previous algorithms, Python was used again with the Gaussian Naïve Bayes model from the sklearn library.

14.5 Results

The prototype algorithm developed in MATLAB for PQRST feature extraction from raw ECG data successfully identified all peaks from the signals successfully. After the moving average filter, threshold signals and derivative base filter were applied, peaks were accurately found at the relevant points in the sample data. The algorithm was tested with the sample ECG data and performed well for all of the samples, an example is given of the output from one of the dataset samples in Figure 14.6.

Multiple machine learning models (SVM, kNN, Decision Tree and Gaussian Naïve Bayes) were used with training and testing data, and accuracy used as a measure of classifier performance. Based solely on ECG data, the aim was for the models to distinguish between normokalaemia and hyperkalaemia. There were some instances of hypokalaemia in the sample data set but not enough to eliminate the issue of overfitting, this is something that could be explored further in future research. The Decision Tree model performed the best with 90.9 percent predictive accuracy, kNN achieved 77.3 percent accuracy, SVM and Gaussian Naïve Bayes performed the worst as they both had approximately 72 percent accuracy. Kasar and Joshi [54] found the Decision Tree model to have 92.5 percent accuracy in a similar study,

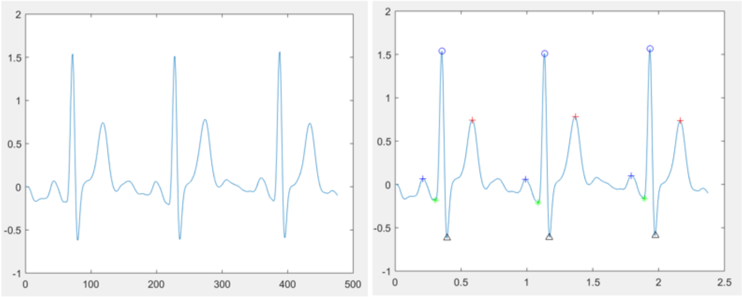

Figure 14.6: Sample ECG data before and after feature extraction

however, their goal was to identify cardiac disease from ECG data, specifically myocardial infarction in which ECG feature changes are much more distinguishable and intensified. Hyperkalaemia ECG changes vary between levels of serum potassium elevation which explains why a lower predictive accuracy was achieved in this instance.

Normal ECG parameters are given in Figure 14.1. Features extracted during peak analysis within these ranges can be classified as normokalaemia. Specific parameters leading to the classification of hyperkalaemia were identified as follows:

- T wave amplitude greater than 0.53mV
- QRS duration greater than 0.12 seconds
- PR interval greater than 0.22 seconds
- P wave amplitude less than 0.04mv

The presence of one or a combination of features found to be in the threshold/s listed above during peak analysis is highly likely to be classified as hyperkalaemia. This is very generalised and does not account for any cardiac abnormalities which should be considered and thresholds adapted accordingly.

14.6 Conclusions and Recommendations

This research has successfully produced a prototype of an algorithm for the feature extraction of PQRST peaks in MATLAB. Alternative existing algorithms such as the Pan Tompkins algorithm [43] focus on identifying the QRS complex which alone is unsuitable in the diagnosis of hyperkalaemia as P and T waves are fundamentally affected by alterations in blood potassium levels. Discrete Wavelet Transform (DWT) is another frequently used method in feature extraction from ECG signals [38]. However, DWT uses kernels which can bias the shape of the signal to be similar to that of an arbitrary, pre-selected shape which could negatively impact on the accuracy of estimated potassium level.

Results showed that out of the four machine learning classification models tested (SVM, kNN, Decision Tree and Gaussian Naïve Bayes), the Decision Tree model performed the best with 90.9 percent accuracy. This indicates that the Decision Tree model is the most suitable out of all of the models tested, in the identification of hyperkalaemia from ECG data. There is definite potential for the use of machine learning in recognising ECG changes associated with raised blood potassium levels.

This research has highlighted that machine learning has the potential to be hugely beneficial to people living with kidney disease and/or on dialysis. The early identification of hyperkalaemia will prevent symptoms, reduce the negative effects and decrease the risk of mortality as seeking treatment sooner is linked to a highly improved outcome for the individual [58]. The proposed prototype algorithms for hyperkalaemia detection have been designed separately on different platforms (MATLAB and Python). Future recommendations would be to merge and fully automate the algorithms enabling compatibility with mobile devices.

Further testing is also recommended with real-world testing involving a sample of patients from a population of people with kidney disease is suggested. Testing should be carried out thoroughly with ECGs from all participants obtained from the same mobile sensors in a controlled environment to exclude any bias or interference and maintain consistency. Blood samples should be collected from participants by a suitable qualified clinician to compare the actual value with that predicted by the machine learning model.

Existing literature suggests that mixed models which incorporate multiple algorithms into a combined model (e.g. Deep Learning and SVM) have performed very well in the identification of cardiac disease from raw ECG data. This has the potential to be useful in the detection of hyperkalaemia due to similarities in identifying trends and patterns in ECG signals. Combined classifiers should be further explored to ensure the best possible model is developed.

References

[1] UK KC. Facts and Stats; 2019. Available from: https://www.kidneycareuk.org/news-and-campaigns/facts-and-stats.

[2] Kurniawan A, Hsu C. Association of kidney function-related dietary pattern, weight status, and cardiovascular risk factors with severity of impaired kidney function in middle-aged and older adults with chronic kidney disease: a cross-sectional population study. Nutrition Journal. 2019;18(1).

[3] Montford J, Linus S. How Dangerous Is Hyperkalemia? Journal of the American Society of Nephrology. 2017;28(11):3155–3165.

[4] Albright. Water and Electrolyte Disturbances in Acute Renal Failure. Critical Care Nephrology. 2019;3:645–649.

[5] Noah B, Keller M, Stein L, et al. Impact of remote patient monitoring on clinical outcomes: an updated meta-analysis of randomized controlled trials. Digital Medicine. 2018;1(1).

[6] Vashist SK, Luong JH. Wearable Technologies for Personalized Mobile Healthcare Monitoring and Management. Wearable Technology in Medicine and Health Care. 2018;235.

[7] Corsi C, Cortesi M, Callisesi G, et al. Noninvasive quantification of blood potassium concentration from ECG in hemodialysis patients. Scientific Reports. 2017;7(1).

[8] Beam A, Kohane I. Big Data and Machine Learning in Health Care. JAMA. 2018;319(13):1317.

[9] Long B, Warix JR, Koyfman A. Controversies in management of hyperkalemia. The journal of emergency medicine. 2018;55(2):192–205.

[10] Opoku-Okrah C, Safo B, Dogbe E. Changes in potassium and sodium concentrations in stored blood. Pan African Medical Journal. 2015;20.

[11] Núñez J, Bayés-Genís A, Zannad F, et al. Long-Term Potassium Monitoring and Dynamics in Heart Failure and Risk of Mortality. Circulation. 2018;137(13):1320–1330.

[12] Sabiullah M, Rahmathunnisa R, Veeramalla V. FATAL EFFECTS OF HYPERKALEMIA. Journal of Evolution of Medical and Dental Sciences. 2015;4(76):13267–13273.

[13] Viera AJ, Wouk N. Potassium Disorders: Hypokalemia and Hyperkalemia. American family physician. 2015;92(6).

[14] Panneerselvam A. ECG in hyperkalaemia. Postgraduate Medical Journal. 2018;94(1115):537.

[15] Varga C, Kálmán Z, Szakáll A, et al. ECG alterations suggestive of hyperkalemia in normokalemic versus hyperkalemic patients. BMC Emergency Medicine. 2019;19(1).

[16] Attia Z, DeSimone C, Dillon J, et al. Novel Bloodless Potassium Determination Using a Signal-Processed Single-Lead ECG. Journal of the American Heart Association. 2016;5(1).

[17] Yasin O, Attia Z, Dillon J, et al. Noninvasive blood potassium measurement using signal-processed, single-lead ecg acquired from a handheld smartphone. Journal of Electrocardiology. 2017;50(5):620–625.

[18] Walker A, Muhlestein J. Smartphone electrocardiogram monitoring: current perspectives. Advanced Health Care Technologies. 2018;4:15–24.

[19] Waks J, Fein A, Das S. Wide Complex Tachycardia Recorded With a Smartphone Cardiac Rhythm Monitor. JAMA Internal Medicine. 2015;175(3):437.

[20] Rahman T, Amirfar V. AliveCor Mobile ECG: A smartphone with heart. Pharmacy Today. 2016;22(3):56.

[21] Kanchi S, Sabela M, Mdluli P, et al. Smartphone based bioanalytical and diagnosis applications:A review. Biosensors and Bioelectronics. 2018;102:136–149.

[22] Guzik P, Malik M. ECG by mobile technologies. Journal of Electrocardiology. 2016;49(6):894–901.

[23] Majumder S, Mondal T, Deen M. Wearable sensors for remote health monitoring. Sensors. 2017;17(1):130.

[24] Luz E, Schwartz W, Cámara-Chávez G, et al. ECG-based heartbeat classification for arrhythmia detection: A survey. Computer Methods and Programs in Biomedicine. 2016;127:144–164.

[25] Patro KK, Kumar PR. Effective feature extraction of ECG for biometric application. Procedia computer science. 2017;115(1):296–306.

[26] Diker A, Avci E, Cömert Z, et al. Classification of ECG signal by using machine learning methods. In 2018 26th Signal Processing and Communications Applications Conference. 2018;p. 1–4.

[27] Celin S, Vasanth K. ECG Signal Classification Using Various Machine Learning Techniques. Journal of Medical Systems. 2018;42(12).

[28] Hadjem M, Naït-Abdesselam F. An ECG T-wave anomalies detection using a lightweight classification model for wireless body sensors. In 2015 IEEE International Conference on Communication Workshop (ICCW). 2015;p. 278–283.

[29] Galloway C, Valys A, Shreibati J, et al. Development and Validation of a Deep-Learning Model to Screen for Hyperkalemia From the Electrocardiogram. JAMA Cardiology. 2019;4(5):428.

[30] Rafique Z, Aceves J, Espina I, et al. Can physicians detect hyperkalemia based on the electrocardiogram? The American Journal of Emergency Medicine. 2019;.

[31] Velagapudi V, O'Horo JC, Vellanki A, et al. Computer-assisted image processing 12 lead ECG model to diagnose hyperkalemia. Journal of electrocardiology. 2017;50(1):131–138.

[32] Littmann L, Gibbs M. Electrocardiographic manifestations of severe hyperkalemia. Journal of Electrocardiology. 2018;51(5):814–817.

[33] Dillon J, DeSimone C, Sapir Y, et al. Noninvasive potassium determination using a mathematically processed ECG: Proof of concept for a novel "bloodless, blood test". Journal of Electrocardiology. 2015;48(1):12–18.

[34] Pilia N, Dössel O, Lenis G, et al. ECG as a tool to estimate potassium and calcium concentrations in the extracellular space. In 2017 Computing in Cardiology. 2017;p. 1–4.

[35] Abdalla F, Wu L, Ullah H, et al. ECG arrhythmia classification using artificial intelligence and nonlinear and nonstationary decomposition. Signal, Image and Video Processing. 2019;13(7):1283–1291.

[36] Gandham S, Bhuma A. ECG FEATURE EXTRACTION AND PARAMETER EVALUATION FOR DETECTION OF HEART ARRHYTHMIAS. Journal on Digital Signal Processing. 2017;5(1):29.

[37] Demski A, Soria ML. ecg-kit: a Matlab toolbox for cardiovascular signal processing. Journal of open research software. 2016;4(1).

[38] Azariadi D, Tsoutsouras V, Xydis S, et al. ECG signal analysis and arrhythmia detection on IoT wearable medical devices. In 2016 5th International conference on modern circuits and systems technologies (MOCAST). 2016;p. 1–4.

[39] Reddy KG, Vijaya PA, Suhasini S. ECG Signal Characterization and Correlation To Heart Abnormalities. International Research Journal of Engineering and Technology. 2017;4(5):1212–1216.

[40] Savalia S, Acosta E, Emamian V. Classification of cardiovascular disease using feature extraction and artificial neural networks. Journal of Biosciences and Medicines. 2017;5(11):64.

[41] Malasri K, Wang L. Design and Implementation of a Secure Wireless Mote-Based Medical Sensor Network. Sensors(Basel, Switzerland). 2009;9:6273–6297. Available from: https://creativecommons.org/licenses/by/3.0/legalcode.

[42] Serum Potassium effects on ECG;. Available from: https://doctorlib.info/cardiology/park-pediatric-cardiology-practitioners/park-pediatric-cardiology-practitioners.files/image064.jpg.

[43] Pan J, Tompkins W. A Real-Time QRS Detection Algorithm. IEEE Transactions on Biomedical Engineering. 1985;32(3):230–236.

[44] Lee C, Yoon H. Medical big data: promise and challenges. Kidney Research and Clinical Practice. 2017;36(1):3–11.

[45] Nie Y, Zou J, Liang Y, et al. Electrocardiographic Abnormalities and QTc Interval in Patients Undergoing Hemodialysis. PLOS ONE. 2016;11(5).

[46] Raschka S, Mirjalili V. Python Machine Learning. Packt Publishing Ltd. 2017;.

[47] Suthaharan S. Support Vector Machine. Machine learning models and algorithms for big data classification. 2016;p. 207–235.

[48] Chapter 2 : SVM (Support Vector Machine);. Available from: https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72.

[49] Zhang S, Li X, Zong M, et al. Efficient knn classification with different numbers of nearest neighbors. IEEE transactions on neural networks and learning systems. 2017;29(5):1774–1785.

[50] Hu B, Li X, Sun S, et al. Attention recognition in EEG-based affective learning research using CFS+ KNN algorithm. IEEE/ACM transactions on computational biology and bioinformatics. 2016;15(1):38–45.

[51] Ghorbani, Afkhami R, Azarnia G, et al. Cardiac arrhythmia classification using statistical and mixture modeling features of ECG signals. Pattern Recognition Letters. 2016;70:45–51.

[52] Decision Tree Model Template;. Available from: https://www.wordtemplatesdocs.org/decision-tree-templates.html.

[53] Verma L, Srivastava S, Negi PC. A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data. Journal of medical systems. 2016;40(7):178.

[54] Kasar SL, Joshi MS. Analysis of multi-lead ecg signals using decision tree algorithms. International Journal of Computer Applications. 2016;134(16).

[55] Jalanko M, Heliö T, Mustonen P, et al. Novel electrocardiographic features in carriers of hypertrophic cardiomyopathy causing sarcomeric mutations. Journal of electrocardiology. 2018;51(6):983–989.

[56] Bayasi N, Tekeste T, Saleh H, et al. Low-power ECG-based processor for predicting ventricular arrhythmia. IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 1974;24(5):1962–1974.

[57] Sannino G, De Pietro G. A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. Future Generation Computer Systems. 2018;86:446–455.

[58] Sterns RH, Grieff M, Bernstein PL. Treatment of hyperkalemia: something old, something new. Kidney international. 2016;89(3):546–554.