[1]Achamma Thomas

[2]Dr Prasad Lokulwar

[3]Dr Vibha Bora

ECG Signal Classification for Detection of Hyperkalemia

Journal of Electrical Systems

Abstract: - Potassium Imbalance is a serious problem that is the leading cause of sudden cardiac deaths in chronic kidney patients (CKD). Currently blood test is the gold standard for detection of potassium imbalances. Electrocardiogram (ECG) signals provide a non- invasive way to view the cardiac activity of the heart. It can also be used to detect potassium imbalance in Chronic Kidney patients. However the unique characteristics and complexity of ECG signals make it a very challenging process. This paper presents a machine learning classifier for detection of hyperkalaemia in patients using features extracted from ECG. Feature extraction plays a very crucial role in the machine learning process as it helps to capture the essential information for the learning task which can be used for applications such as classification and detection. Ten statistical features were extracted from ECG signals of patients having potassium within normal range and those with elevated levels of potassium. Classification performance of four different classifiers namely Naïve Bayes Classifier, Support Vector Machine(SVM), K Nearest Neighbors(kNN) and Artificial Neural Networks(ANN) was compared using statistical features. kNN and ANN performed best with classification accuracy of 97.9%.The results we found are in-line with other state-of-art hyperkalaemia classification approaches

General Terms: *ECG Signal Processing, Machine Learning*

Keywords: Potassium imbalance, hyperkalemia, machine learning, chronic kidney disease.

I. INTRODUCTION

Cardiovascular vascular disease is the leading cause of morbidity and mortality world-wide. Early cardiac anomaly detection and diagnosis can be crucial for prompt treatment and better patient outcomes. Electrocardiography (ECG) has long been a primary tool for the diagnosis of heart disorders due to its non-invasive nature, affordability, and widespread availability. The ECG signal pro-vides a graphical representation of the heart's electrical activity, offering insights into various pathological conditions. Hyperkalaemia, characterized by elevated potassium levels (>5.3mEq/L) in the blood, is one such condition that leaves a dis-tinct footprint on the ECG waveform. High levels of potassium can lead to sud-den cardiac deaths in patients suffering from Chronic Kidney Disease [1]. Early detection of hyperkalaemia-induced ECG changes can aid in rapid clinical inter-vention, preventing potential life-threatening complications.

Traditional methods of ECG interpretation rely heavily on the expertise of clinicians but the emergence of machine learning and computational techniques has paved the way for automated ECG analysis systems. However, the success of these systems depends on the selection of appropriate features that capture the changes in the ECG waveform effectively.

This paper introduces a novel approach to ECG classification for hyperkalaemia using statistical features. We hypothesize that certain statistical attributes of the ECG signal can effectively differentiate between normal and hyperkalaemia conditions. By harnessing these features, we aim to develop an auto-mated classification system that offers high accuracy and can be integrated into routine clinical practice to aid in the early detection of hyperkalaemia. Our work contributes to the growing body of research on automated ECG analysis by focusing on a specific and clinically significant condition. By leveraging statistical features, we aim to create a model that is robust and bridges the gap between computational methodologies and clinical applicability.

[1] Research Scholar, Department of Computer Science & Engineering, G H Raisoni University, Amravati, India, achamma.thomas@raisoni.net

[2]Professor, Department of Computer Science & Engineering, G H Raisoni College of Engineering, Nagpur, India, prasad.lokulwar@raisoni.net

[3]Professor, Department of Electronics Engineering, G H Raisoni College of Engineering, Nagpur, India, vibha.bora@raisoni.net

*Correspondence: achamma.thomas@raisoni.net

The rest of this paper unfolds as follows: Section II explores the literature survey. Section III presents the dataset and provides details on the extraction methods used. In Section IV, we elaborate on the methodology employed for this study. Section V encompasses the results and ensuing discussion. Ultimately, Section VI presents our conclusions.

II. RELATED WORK

Hyperkalemia is a potential threat especially in patients suffering from Chronic Kidney Disease [1]. An automatic, non invasive method for detection hyperkalemia is crucial for timely intervention, appropriate treatment, and preventing potential complications. Previous studies have utilized deep learning models [2] [3] [4] [5] to classify between normal and hyperkalaemia classes. Deep learning models require vast amounts of data to train effectively without over-fitting. Given the limited size of our dataset, we opted for machine learning approach over deep learning techniques.

Feature extraction plays a very important role in Machine Learning. In [6] authors have used morphological features such as P and T wave amplitudes, QRS and PR intervals as features for classifying between normal and hyperkalaemia class which were fed to four classifiers namely Gaussian Naïve Bayes, Decision Tree, SVM and kNN, of which decision tree performed the best with classification accuracy of 90.9%. T wave, T wave amplitude, T wave duration was used to estimate blood potassium in patients undergoing haemodialysis [7] [8]. These morphological features, derived from the intricate shapes and patterns within the ECG waveform, can be highly indicative of various cardiac anomalies. However, the accurate extraction of these morphological features is dependent upon the quality and consistency of the ECG recordings and the algorithms being used to extract it. Any slight distortion, noise, or variability in the waveform can lead to inaccuracies in feature extraction, which can subsequently mislead and negatively influence the performance of the classifier. In this research, our aim is to extract statistical features from ECG signals, as they are inherently more robust to noise and various artifacts. Our ultimate goal is to leverage these features for the accurate detection of hyperkalemia.

III. DATASET

The "Medical Information Mart for Intensive Care" or MIMIC-III is a large single-center database that contains data on patients admitted to critical care units at large tertiary care hospitals. The dataset used in this work is extracted from MIMICS III Waveform Database Matched Subset [9] [10] [11]. The records in this database, which is a subset of the MIMICS III Waveform Database, are those for which the patient has been identified. Their accompanying clinical rec-ords are available in the MIMIC-III Clinical Database. The MIMICS III Wave-form Database Matched Subset contains 22,317 waveform records for 10,282 distinct patients. However not all signals are of adequate quality for rigorous analysis. Several factors including noise, artifacts and other irregularities com-promises the quality of the signals. Given the critical nature of our study and the necessity for high-quality data, we embarked on a meticulous process of manual signal verification. Each ECG recording was visually inspected by our team. This step ensured that only signals meeting our predefined quality criteria, devoid of significant artifacts, and offering clear representation of the heart's electrical activity, were retained for subsequent analysis. We selected those ECG wave-forms from MIMICS III ECG matched subset whose corresponding clinical rec-ords suggested that the patient had normal or high levels of serum potassium. To ensure the relevance and accuracy of the correlation between potassium levels and ECG characteristics, only those ECG recordings were selected where the time difference between the associated potassium test and the ECG acquisition was less than 4 hours. The ECG signals extracted from this database were sampled at 125 Hz. The available signals were of varying lengths hence ECG segments of 20 seconds duration were extracted to ensure consistent length. A total of 207 sam-ples of patients having normal potassium level (3.5-5.2 mEq/L) and 280 samples of patients having elevated potassium levels (5.3-8 mEq/L) were considered for this study. Thus we have a 487 samples with 2500 data points forming an input matrix of size 487 x 2500.

IV. METHODOLOGY

This section describes the overall methods used for the classification task. Figure 1 presents the block diagram of the methodology used in this paper.

4.1 Pre-processing

In the course of data examination, we identified instances where certain ECG sig-nals exhibited missing values, potentially due to artifacts, or data recording issues. We employed interpolation as our primary pre-processing technique to fill in the missing data points within the ECG signals. Interpolation offers a mathematically sound approach to estimate missing values based on the neighbouring data points, ensuring that the imputed values align well with the inherent characteristics of the ECG waveform. By using interpolation, we were able to reconstruct the af-fected segments of the ECG signals, ensuring they remained suitable for subsequent analysis without introducing artificial biases

4.2 Feature Extraction

Feature extraction techniques play a very important role in machine learning classification task. ECG Signal is a time series representation of the electrical activity of the heart. Subtle changes in the ECG due to mild/moderate hyperkalemia may not be visible to the naked eye. As the level of hyperkalaemia intensifies, its impact on the ECG becomes progressively more pronounced. The alterations in the ECG waveform commence with tall, peaked T-waves, followed by a prolongation of the PR interval and a flattening of the P-wave. As hyperkalaemia advances, the QRS complex broadens, and in severe cases, it may even merge with the T-wave, resulting in a sine-wave pattern [9]. These changes are gradual, aligning with the severity of hyperkalaemia; the more elevated the potassium levels, the more distinct the ECG alterations become.

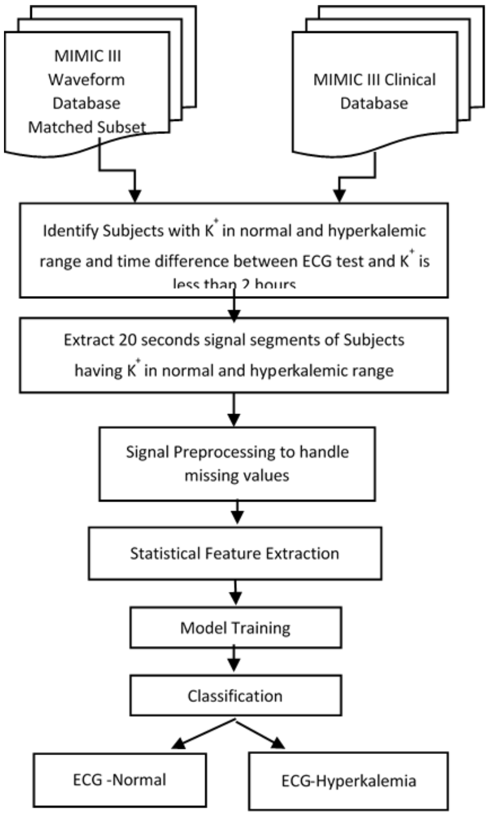

Figure. 1: Methodology

Statistical features of the signal can capture the underlying characteristics of changes in the ECG. They are less sensitive to noise and other interference making the classification more robust. Moreover they are interpretable and hence can provide clinicians a rationale to understand the machines decision. The statistical features used are described below.

4.2.1 ***Mean(μ)*** **:** The average value of the signal, computed as:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (1)$$

where xi represents each data point in the signal and N is the total number of data points.

4.2.2 ***Median(Med) :*** The median of the ECG Signal is the value that divides the higher half from the lower half of the data sample.

$$\boldsymbol{Med(X)} = \begin{cases} \boldsymbol{X\left[\frac{n+1}{2}\right]} & \boldsymbol{if\ n\ is\ odd} \\ \frac{\boldsymbol{X\left[\frac{n}{2}\right]+X\left[\frac{n}{2}+1\right]}}{\boldsymbol{2}} & \boldsymbol{if\ n\ is\ even} \end{cases} \quad (2)$$

Where X represents the signal.

4.2.3 ***Variance (σ^2):*** A measure of the signal's dispersion, calculated as:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \quad (3)$$

Where μ is the mean of the signal.

4.2.4 ***Standard Deviation (σ):*** Represents the amount of variation or dispersion in the signal, given by

$$\sigma = \sqrt{\sigma^2} \quad (4)$$

4.2.5 ***Skewness:*** Skewness quantifies the lack of symmetry in the probability distribution of a real-valued random variable with respect to its mean**.**

$$Skewness(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] \quad (5)$$

Where E denotes the expected value, μ denotes the mea and σ denotes the standard deviation.

4.2.6 ***Kurtosis:*** Kurtosis gauges the extent of "tailedness" in the probability distribution of a real-valued random variable.

$$\boldsymbol{Kurtosis(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] - 3} \quad (6)$$

4.2.7 ***Interquartile range(IQR):*** The IQR measures the statistical spread, or the range between the first and **third** quartiles, effectively capturing the middle 50% of the signal**.**

$$IQR = Q_3 - Q_1 \quad (7)$$

Where Q_3 is the third quartile (or 75th percentile) and Q_1 is the first quartile (or 25th percentile).

4.2.8 ***Energy (E):*** Energy represent the power of a signal segment.

$$E = \sum_{n=0}^{N-1}|x[n]|^2 \quad (8)$$

where N is the length of the segment

4.2.9 ***Minimum Value*****:** It is the lowest value of the signal

4.2.10 ***Maximum Value*****:** It is the highest value of the signal

Ten features mentioned above were extracted from the input data matrix of size 487x2500 to obtain a reduced feature matrix of size 487x10. The features were normalized using Z-score.

V. CLASSIFICATION

In this study, our primary objective is to classify ECG signals into one of two distinct categories: ‘Normal’ and ‘Hyperkalemic’ class. In our analysis, we employed four distinct classifiers to train on the normalized features extracted from the ECG signals. Specifically, we utilized the following classifiers:

5.1 Naive Bayes [13]:

This probabilistic classifier is based on applying Bayes' theorem, assuming independence between the features. Given the continuous nature of the ECG signals with their characteristic peak patterns, we employed the Kernel

Naive Bayes classifier as implemented in MATLAB. This approach is advantageous for our dataset since it does not make strong parametric assumptions and estimates the probability density function of the continuous data using kernel smoothing techniques. This ability to adapt to the actual distribution of the data, without being confined to a specific shape like the Gaussian distribution, enhances the classifier's capability to discern intricate patterns in the ECG signals.

5.2 Support Vector Machine (SVM) [14]:

SVM seeks to find the optimal hyperplane that best divides the dataset into classes. For our study we employed SVM with cubic kernel also known as polynomial kernel with a degree of three, as it is capable of capturing non-linear relationships in the data. ECG signals with their subtle variations between 'Normal' and 'Hyperkalemic' patterns, can exhibit complex non-linear characteristics. The cubic kernel allows our SVM model to operate in an expanded feature space, enabling it to find hyperplanes that can effectively differentiate between the two classes even if they aren't linearly separable in the original space.

5.3 **K-Nearest Neighbors (KNN)** [15]:

This non-parametric method classifies data points based on how their neighbors are categorized. We employed the K-Nearest Neighbors (KNN) algorithm for ECG signal classification due to its inherent strengths. KNN's instance-based learning approach captures individual signal nuances, making it apt for the varied waveforms in ECG data. Being non-parametric, KNN doesn't impose rigid assumptions on data distribution, ensuring flexibility in handling the intricate patterns of ECG signals. Its sensitivity to local patterns makes it adept at identifying subtle ECG variations. The transparent decision-making process of KNN also offers interpretable results, a valuable asset in clinical contexts

5.4 **Neural Networks (NN) [16]:**

A multi-layer perceptron model can capture complex non-linear relationships in the data. A wide NN architecture was used for classification as it effectively models feature interactions inherent in temporal ECG data and offers a rich set of transformations for nuanced differentiation.

Each classifier was trained using the same dataset to ensure consistency in the evaluation of their performance. 10 fold cross validation was used and 20% of the data was set aside for testing

VI. RESULTS & DISCUSSION

All the experiments in this study was carried out in MATLAB. The dataset extracted from MIMICS III Waveform matched subset database were of 20 seconds sampled at 125 Hertz. Linear interpolation was employed to address the limited missing values within our ECG dataset. The sparse nature of missing data in our ECG signals meant that only short segments were absent. Given the brief duration of these gaps, the signal's characteristics between two adjacent known points were reasonably approximated as linear. After pre-processing 10 statistical features were extracted from the signal which reduced the size of the dataset from 487x2500 to 487x10. Then we compared the performance of four classifiers on the reduced feature set. The results of the classification is presented in Table 1. K Nearest Neighbour classifier and Artificial Neural Network achieved maximum accuracy of 97.9% in classifying normal and hyperkalemic classes.

Table 1: Classification Result

Classifier	***Sensitivity***	***Specificity***	***Accuracy***
SVM Cubic Kernel	100.0	93.2	96.9
Neural Network	98.2	97.6	97.9

K Nearest Neighbors	98.2	97.6	97.9
Naïve Bayes	98.0	87.0	92.8

The AUC of the Receiver Operating Characteristics (ROC) curve suggest that ANN classifier(Figure 2) has high discriminating power with AUC 99.4% followed by SVM with cubic kernel with an AUC of 98.5%, KNN with AUC of 97.8% and Naïve Bayes with AUC of 94.8%. This suggests the models ability to distinguish between the normal and hyperkalemia classes.

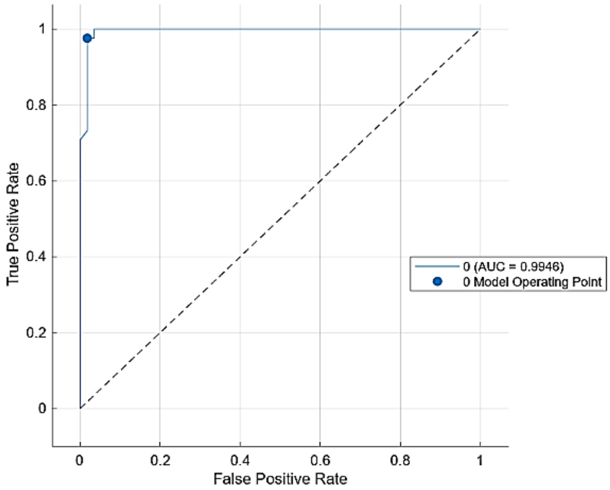

Figure2. AUC of ANN Model

VII. CONCLUSION

In this study data from MIMICS III Waveform matched subset of subjects suffering from hyperkalaemia were considered. The corresponding clinical records were taken from MIMICS III Clinical database. Ten statistical features were extracted from the 487 signals. The extracted features could detect hyperkalaemia with an accuracy of 97.9%. Future studies will incorporate a broader dataset and explore a more diverse set of features to enhance the detection and interpretation of hyperkalemia in ECG signals

REFERENCES

[1] Rodrigues, Ana Santos, et al., "Noninvasive monitoring of potassium fluctuations during the long interdialytic interval," IEEE Access, pp. 188488-188502, 2020.

[2] Urtnasan, Erdenebayar, et al., "Noninvasive Screening Tool for Hyperkalemia Using a Single-Lead Electrocardiogram and Deep Learning: Development and Usability Study," JMIR Med Inform, vol. 10, no. 6, p. e34724, 2022.

[3] Kwon, Joon-myoung, et al., "Artificial intelligence for detecting electrolyte imbalance using electrocardiography.," Annals of Noninvasive Electrocardiology, vol. 26, no. 3, p. e12839, 2021.

[4] Lin, Chin-Sheng, et al., "A Deep-Learning Algorithm (ECG12Net) for Detecting Hypokalemia and Hyperkalemia by Electrocardiography: Algorithm Development," JMIR Med Inform., vol. 8, no. 3, p. e15931., 2020.

[5] Galloway, Conner D., et al., "Development and Validation of a Deep-Learning Model to Screen for Hyperkalemia From the Electrocardiogram," JAMA Cardiology, vol. 4, no. 5, pp. 428-436, 2019.

[6] C. Farrell and M. Z. Shakir, Machine learning enabled ECG monitoring for early detection of hyperkalaemia, AI for Emerging Verticals: Human-Robot Computing, Sensing and Networking (2020 ed.), 2020.

[7] Yasin, Omar Z., et al., "Noninvasive blood potassium measurement using signal-processed, single-lead ecg acquired from a handheld smartphone.," J Electrocardiology, vol. 50, no. 5, pp. 620-625, 2017.

[8] Regolisti, Giuseppe, et al., "Electrocardiographic T wave alterations and prediction of hyperkalemia in patients with acute kidney injury," Intern Emerg Med, vol. 15, no. 3, pp. 463-472, 2020.

[9] B. M. Elliott TL, "Electrolytes: Potassium Disorders," FP Essent., vol. 459, pp. 21-28, 2017.

A. P. T. &. M. R. Johnson, "MIMIC-III Clinical Database (version 1.4)," Scientific Data, p. DOI: 10.1038/sdata.2016.3, 2016.

[10] G. Webb, Naïve Bayes. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning., Boston, MA: Springer, 2011.

[11] S. Suthaharan, Support Vector Machine. In: Machine Learning Models and Algorithms for Big Data Classification., Boston, MA: Springer, 2016.

[12] Johnson, Alistair EW, et al., "MIMIC-III, a freely accessible critical care database," Scientific Data 3, vol. https://doi.org/10.1038/sdata.2016.35, p. 160035, 2016.

[13] Goldberger, Ary L., et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals.," Circulation [Online]., vol. 101, no. 23, p. e215–e220., 2000.

[14] Guo, Gongde, et al., "KNN Model-Based Approach in Classification," in Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2003.

[15] Braspenning, Petrus J., Artificial Neural Networks, Berlin, Heidelberg: Springer, 1995.