FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 11, Issue 2 | Pages 418-425 | e-ISSN: 2347-470X

Non-invasive and Automatic Identification of Diabetes Using ECG Signals

Anuja Jain[1], Anurag Verma[2] and Amit Kumar Verma[3]
[1]*College of Pharmacy, Teerthanker Mahaveer University, Moradabad, UP, 244001 India, anuja.scholar@tmu.ac.in*
[2]*College of Pharmacy, Teerthanker Mahaveer University, Moradabad, UP, 244001 India, principal.pharmacy@tmu.ac.in*
[3]*Department of Pharmacy, Mahatama Jyotiba Phule Rohilkhand University, Bareilly, UP, 243006 India, amit.verma@mjpru.ac.in*

***Correspondence:** anuja.scholar@tmu.ac.in

ABSTRACT- Diabetes Mellitus is a chronic medical condition in which the body is unable to properly regulate the amount of glucose (a type of sugar) in the blood. It can cause serious consequences like heart disease, nerve damage, and kidney illness. Diabetes causes cardiac autonomic neuropathy, which affects the pattern of electrocardiogram (ECG) signals. ECG measures electrical activity of the hearts. In this paper, the features extraction method is proposed for the classification of diabetic ECG and normal ECG signals. Ten features, namely, log energy, threshold, Shannon, sure entropy, root mean square value, kurtosis, skewness, maximum value, energy, and variance are extracted from the single-lead ECG signal. Fisher-score has been employed for features ranking methods the ranked features are used as input to the classifiers namely medium tree, coarse Tree, linear discriminant, quadratic discriminant, and Gaussian naive Bayes, classifiers. The five ranked features using medium tree classifier has produced an accuracy of 87.19%. The analysis of performance measurement shows the effectiveness of the proposed method in the classification of diabetic and non-diabetic ECG signals.

Keywords: Diabetes Mellitus, electrocardiography (ECG), features extraction, machine learning.

ARTICLE INFORMATION

Author(s); Anuja Jain, Anurag Verma and Amit Kumar Verma;
Special Issue Editor: Dr. Debnath Bhattacharyya
Received: 21/04/2023; **Accepted**: 13/06/2023; **Published**: 15/06/2023;
E- ISSN: 2347-470X;
Paper Id: IJEER_SP-DB_2023_03;
Citation: 10.37391/IJEER.110223
Webpage-link:
https://ijeer.forexjournal.co.in/archive/volume-11/ijeer-110223.html

This article belongs to the Special Issue on **Machine Intelligence and Deep Learning in Healthcare Decision Making**

1. INTRODUCTION

Diabetes is a chronic health problem, describe by hyperglycemia. Excess amounts of glucose level in the bloodstream are known as hyperglycemia. The increased glucose level in the blood may cause very serious destructive effects. This chronic disease affects a large population all over the world. According to the World Health Organization (WHO), about 536.6 million cases were reported in 2021, and it is expected to reach 783.2 million by 2045 [1]. Diabetes is increasing in middle- and low-income countries very rapidly, the 7th leading cause of death worldwide may be due to its chronic condition and hereditary linkages. In diabetes mellitus (DM), blood sugar is not utilized or consumed by the body cells. Body cells are not able to produce a sufficient amount of insulin. Diabetes is a non-curable disorder that can be controlled after an accurate diagnosis. The consequences of diabetes may lead to other disorders such as heart attacks, stroke, kidney failure, etc. [2].

Presently, the diagnosis of diabetes is performed by blood, plasma, and serum, which is an invasive and painful process. Sometimes the insertion of a needle cause serious infections and people are afraid of skin punctures during blood tests. In worst cases, people even faint during blood tests. In addition to that, blood glucose testing is done many times a day for effective monitoring of diabetes. Glucose concentration can be measured by blood, serum samples, and plasma, most of the cases plasma and serum samples are preferred because the measurement of the glucose from blood having 15% lower, which may be due to the presence of the extra water content in the blood cells [3-4]. Blood glucose concentrations can be measured by two methods namely Enzymaticamperometric and hexokinase in the laboratories: Enzymatic-Amperometric Method, the oxidation of glucose can be done presence of water (H2O), oxygen (O2), and GOx which provides hydrogen peroxide (H2O2) and gluconic acid.

The electrochemical oxidation of (H2O2) is performed in anode of an electrochemical probe which produces an amperometric (current) signal that is directly proportional to glucose concentration in the blood [5]. Hexokinase method contains a series of chemical reactions, also called as photometric method. Adenosine triphosphate (ATP) and magnesium ions are present during the initial step of this process, a glucose interaction with the enzyme hexokinase, which results in the production of ADP and G6P. In the following step, G6P dehydrogenase is used to oxidise nicotinamide adenine dinucleotide (NAD) and G6P until they are reduced to NADH and 6-phosphogluconate. The amount of NADH may be inversely related to the

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 11, Issue 2 | Pages 418-425 | e-ISSN: 2347-470X

concentration of glucose sample, which absorbs light at 340 nm. The NADH amount determines how much is absorbed, which can measure glucose from standard spectrophotometric techniques [6]. The main drawback of these methods is the inherent invasiveness that needs to be done in-vitro required trained personnel, having extra costs, and time consuming. The laboratory equipment is not well calibrated which affects the accuracy. Self-monitoring blood glucose typical glucometer used and capillary blood access by finger pricking. Glucose concentration measurement is performed by electrochemical technique. A glucose strip is connected to meter and it is used for chemical reaction and detection. This method consists of three electrodes namely the working, reference, and counter electrodes [7].

The same amounts of C-peptide measure and insulin are released into the bloodstream by the pancreas. It is eliminated through the urine instead than being absorbed by the body like insulin. C-peptide levels reveal the precise amount of insulin the body is producing [8-9]. Measure the quantity of glucose that is bound to the haemoglobin in the red blood cells using the glucoseheamoglobin (HbA1c) blood test. Hemoglobin develops a sugary coating when haemoglobin and glucose bind. When there is more sugar in the blood, that coat thickens. A1c tests assess the thickness of that coat during the previous three months. [10-11]. One way to use saliva to measure diabetes is by measuring glucose levels in saliva. In diabetes, the body's ability to produce or use insulin, a hormone that regulates glucose metabolism, is impaired. This leads to high levels of glucose in the blood and can also be found in saliva. Measuring glucose levels in saliva can be done by using a non-invasive method such as glucose oxidase method (GOD-POD), which uses an enzymatic kit glucose oxidase-peroxidase. This method utilizes the enzyme GOD to catalyze the oxidation of glucose to gluconic acid. This method allows for the non-invasive and easy measurement of glucose levels in both blood and saliva, making it a valuable tool for the diagnosis and monitoring of diabetes [12-13]. Saliva is a complex mixture of biological fluids that contains not only enzymes and antibodies but also other important components such as glucose, amino acids, lipids, and mono- and diglycerides of fatty acids. These additional components found in saliva can be used to diagnose various medical conditions diabetes, autoimmune diseases, genetic problems, and other infectious diseases. One commonly used method for measuring glucose levels in saliva is the GOD-POD, which utilises an enzymatic kit glucose oxidase peroxidase [14].

The fluorescence technique for measuring glucose concentration in saliva has some major disadvantages associated with it. One of the main disadvantages is the short lifespan of the fluorophore, which limits the amount of time that the technique can be used for measuring glucose. Additionally, this technique is highly susceptible to interference caused by pH fluctuations, changes in oxygen levels, and biocompatibility issues due to local tissue trauma [15-16]. Optical polarimetry principle of optical polarimetry based on the rotation polarized light. This method is used on the aqueous humor in the anterior chamber of aneye [17-18]. The facial key-based color features were extracted and used as input to a probabilistic collaborative based classifier for the diagnosis of diabetes. The method does not require the extraction of any bodily fluids, it also reduces testing time [19]. This type of method suffers the image angle, and intensity; the face can be influenced intentionally by the person. The major drawback in laboratory methods is inherent invasiveness due to the need for an in-vitro procedure which requires blood samples extracted from patients; is expensive due to additional costs, and is also time-consuming. All laboratory equipment is not highly accurate shown in the study, this equipment also requires continuous calibration. In this condition regular testing of blood sugar levels is necessary, it is a very painful procedure. Electrochemical skin conductance has been explored and ESC is high in case of diabetic patient [20]. Machine learning approach has been employed for detection of diabetes [21-22]. In [23], non-linear characteristics from the R-R interval signal have been retrieved and identified in this study utilising the AdaBoost classifier. Poincare plot based features has been used as input to different machine learning classifier using heart rate variability (HRV) [24].

Overall, the literature on diabetes detection indicates that there is a need for more convenient and accessible methods for detecting diabetes and that there are promising developments in the use of non-invasive techniques such as ultrasound, infrared spectroscopy, and biosensors. Nowadays, electrocardiography (ECG) is the most emerging tool to continuously monitor diabetic patients. An ECG records the electrical activity of the heart using a non-invasive approach. It is a widely used and essential tool for the diagnosis, management, and treatment diabetes. Diabetes affects the electrical activity of the heart, which can be seen in ECG signals.

The changes in ECG can be caused by several factors, including damage to the heart's blood vessels and nerves due to high blood sugar levels. In some cases, these changes may indicate an increased risk of heart disease or other complications associated with diabetes. It is important for people with diabetes to regularly monitor their heart health, including regular ECG testing, to help identify and manage any potential problems. In this paper, features extracted from the ECG signals are explored for the classification of diabetics and normal ECG signals. The rest of this paper is separated into four sections: A brief explanation of the strategy, which includes data collection, extraction of features, ranking techniques, and classifier, is provided in *section 2*. *Section 3* presents the findings and related discussion. The summary and future directions of the work that has been presented are found in *Section 4*.

2. PROPOSED METHOD

This section comprises a dataset of ECG recording, features extraction, Fisher score, and classifier. Figure 1 presents a summary of the steps that make up the proposed architecture.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 11, Issue 2 | Pages 418-425 | e-ISSN: 2347-470X

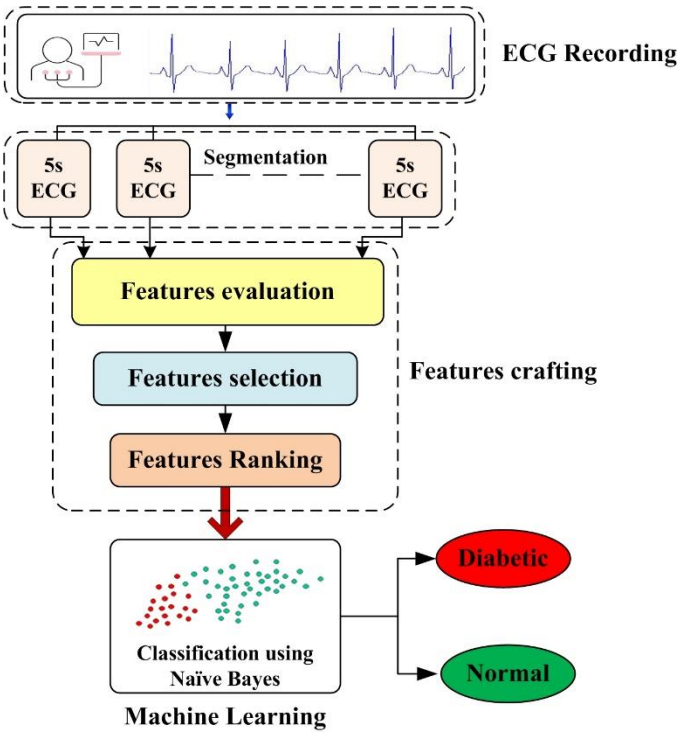

Figure 1: Block representation of the proposed method for classification of diabetes and normal ECG signals

2.1 Dataset

In the ECG dataset, 51 normal (21 females and 30 males) and 35 diabetic patients (12 females and 23 males) are involved with the age in between 20-70 years, 35 diabetic patients are suffering from diabetes from type-2 diabetes. The ECG data has been taken from Biomedical Signal Acquisition Laboratory (BSAL), PDPM-IIITDM, Jabalpur, M.P., India, by submitting the End-User License Agreement (EULA) form. The blood glucose values greater than 160 mg/dL is treated as unhealthy ECG signals. The total segregated ECG fragments/signals are 24630 (9560 DM and 15070 normal) [25]. These signals are recorded BT traveler acquisition software. The data were collected using an ECG montage arrangement, which involved inserting two gel conductive electrodes on the wrist using a high-quality electrodes gel using a 24-bit ADC at a sampling frequency of 256 Hz. Participants with and without diabetes both have a normal BP and no cardiovascular abnormalities. Figure 2 shows the typical ECG fragments of diabetic and normal subjects.

2.2 Features Extraction

Features extraction is a method of identifying and extracting important and relevant information from raw biological signals. This information can be used to learn about the underlying processes and mechanisms that are responsible for the signals. Once the features have been extracted from the raw signals, they can be used for a wide range of applications [26]. Features extraction is an important step in the analysis of biological signals. It allows researchers to identify and analyze the most relevant information in the signals, and to better understand the underlying processes and mechanisms responsible for the signals. In this study, fourteen features namely, log energy entropy, Shannon, threshold, sure, log energy entropy, root mean square value (RMS), kurtosis, skewness, maximum value, minimum value, mean value, standard deviation, variance, and signal energy are computed from each ECG fragments. They are briefly summarized below:

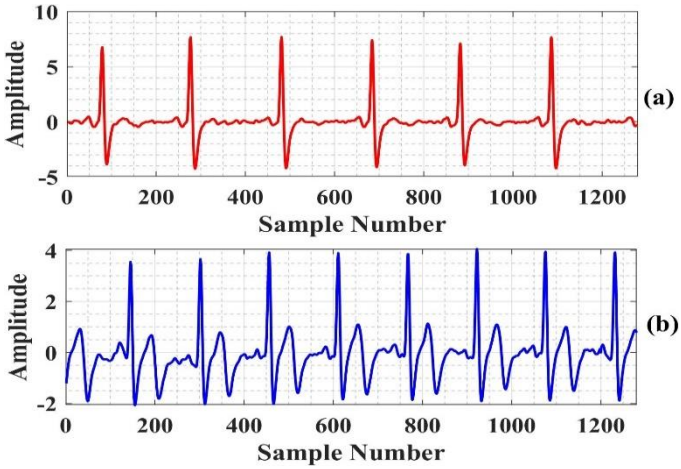

Figure 2: Typical ECG fragments of diabetic and normal subjects

2.2.1 *Log energy entropy (LEE)*

It represents a signal's degree of ambiguity or unpredictability. It is calculated by taking the logarithm of the energy of the signal and then calculating the entropy of the resulting distribution. The log energy entropy is calculated using the following formula:

$$LEE = -\sum(E_i) \times log_2 E_i \quad (1)$$

where E_i is the energy of the signal at each time step i. Log energy entropy is typically used in combination with other features, such as spectral features, to provide a more comprehensive description of a signal. It is a useful measure for evaluating the structure and complexity of a signal.

2.2.2 *Shannon entropy (ShE)*

It is a measure of the uncertainty or randomness of a system. It is calculated using the following formula:

$$ShE = -\sum Pi \times \log Pi \quad (2)$$

where P_i is the probability of each possible outcome in the system.

2.2.3 *Threshold entropy (TE)*

It is a measurement of a system's randomness or uncertainty that is based on the thresholding idea. Thresholding is a technique that is used to separate a signal into different regions or segments, based on a predetermined threshold value. Threshold entropy is calculated using the following formula:

$$TE = -\sum(P_i) \times log_2 P_i \quad (3)$$

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 11, Issue 2 | Pages 418-425 | e-ISSN: 2347-470X

where P_i is the probability of each possible outcome in the system, and the summation is taken over all possible threshold values.

2.2.4 *Sure entropy (SE)*

It is a measurement of randomness of a system that is based on the concept of the generalized likelihood ratio test (GLRT). The GLRT is a statistical test that is used to compare the performance of two or more models for a given data set. Sure entropy is calculated using the following formula:

$$SE = -\sum(P_i) \times log_2 P_i \quad (4)$$

where P_i is the probability of each possible outcome in the system, and the summation is taken over all possible models.

2.2.5 *Root mean square value (RMS)*

The RMS value of a signal is a measurement of the magnitude of the signal. It is figured out by taking the square root of the square root of the signal sample average. The RMS value of a signal is commonly used in signal processing, as it provides a measure of the overall strength or loudness of the signal. The RMS value of a signal is calculated using the following formula:

$$RMS = \sqrt{\frac{1}{n}\sum(x_i)^2} \quad (5)$$

Where x_i is the signal sample at each time step i, and total number of time samples in the signal is given by n.

2.2.6 *Kurtosis (K)*

Kurtosis is a statistical measure that describes the shape of a probability distribution. It is a measure of the peaked ness or flatness of the distribution. Kurtosis is evaluated using the following formula:

$$K = \frac{\frac{1}{n}\sum(x_i - mean(x))^4}{\left[\frac{1}{n}\sum(x_i - mean(x))^4\right]^2} \quad (6)$$

where x_i is the data point at each time step i, x_mean is the mean of the data, and the total number of data points is n.

2.2.7 *Skewness (S)*

It is a statistical measure that describes the symmetry of a probability distribution.

$$S = \frac{\frac{\sum(x_i - mean(x))^3}{n}}{\left[\frac{\sum(x_i - mean(x))^3}{n}\right]^{1.5}} \quad (7)$$

2.2.8 *Maximum value*

It is computed as

$$MV = Max(x_i) \quad (8)$$

2.2.9 *Energy*

In signal processing, signal energy is a measure of the amount of power contained in a signal. It is calculated by squaring the amplitude of the signal and then taking the integral over time. It is computed as

$$E = \sum(x_i)^2 \quad (9)$$

2.2.10 *Variance*

In statistics, the spread or dispersion of a set of data is measured by variance. It is computed as

$$var = \frac{\sum(x_i - mean(x_i))^2}{n-1} \quad (10)$$

2.3 Fisher (F)-Score for Feature Ranking

Fisher (F)-score is a measure used to evaluate the importance or relevance of features in a dataset. It is generally employed in feature selection and ranking and is used to identify the most relevant and useful features in a dataset. The F-score measures the separation between the means of the two classes, normalized by the standard deviation of the feature. Features with a high F-score are considered to be more important or relevant because they have a large separation between the means of the two classes, relative to the spread of the data. F-score is often used in conjunction with other feature selection and ranking methods, such as mutual information and chi-squared. It is a useful measure for evaluating the relevance of features in a dataset, and it can help identify the most important and useful features for a particular problem or application. The Fisher-score (F_w) can be expressed as [27]:

$$F_w = \frac{(x_w{}^{(+)} - x_w)^2 + (x_w{}^{(-)} - x_w)^2}{\frac{1}{n^+-1}\sum_{k=1}^{n^+}(x_{k,w}{}^{(+)} - x_w)^2 + \frac{1}{n^--1}\sum_{k=1}^{n^+}(x_{k,w}{}^{(-)} - x_w)^2} \quad (11)$$

Where $x_w{}^{(+)}, x_w{}^{(+)}$ and x_w represent the mean of wth feature of feature of healthy class, diabetic class, and both classes ECG signals, respectively; $x_{k,w}{}^{(+)}$ and $x_{k,w}{}^{(-)}$ represent the wth feature of the kth ECG observation of healthy class and diabetic class classes, respectively; n^+ and n^-represent the number of ECG observations of healthy and diabetic classes, respectively.

2.4 10-Fold Cross Validation Scheme *(10-FCV)*

This work utilizes a 10-FCV scheme to fed the features in to machine learning classifiers. Cross-validation is a method employed to develop a robust system. The complete data is divided into three parts, for training, validation, and testing the network performance. In the 10-FCV technique, the complete dataset is divided into ten sections randomly. In this study, complete spectral images are split into 10 equal sections, one section is hold-out for network testing, one section is retained for validation, and the remaining sections are used for network training. The cross-validation procedure is repeated ten times by altering the training, validation, and test data. *Figure 3* represents the schematic diagram of the data partition scheme.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 11, Issue 2 | Pages 418-425 | e-ISSN: 2347-470X

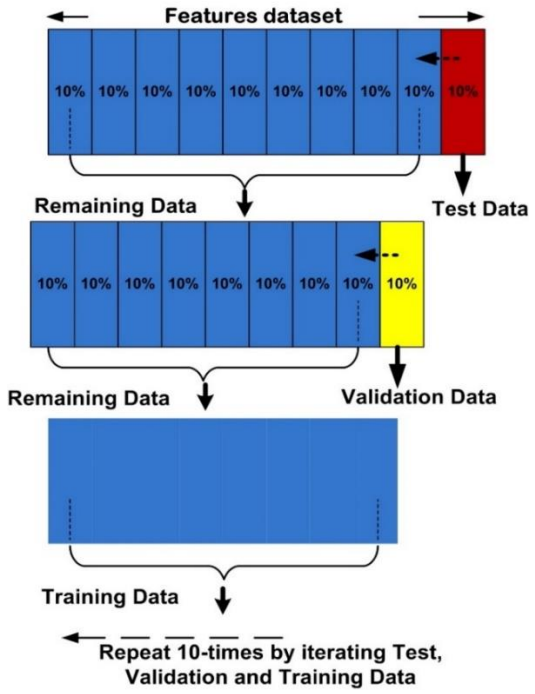

Figure 3: Schematic diagram of the data partition scheme

2.5 Machine Learning Classifier

A machine learning classifier is a type of algorithm that is used to automatically classify data. A classifier is trained using a labeled training dataset, which contains examples of the data that the classifier will be asked to classify. The classifier uses the training dataset to learn the patterns and relationships between the data and the corresponding labels. Once the classifier is trained, it can be used to automatically classify new data. To classify a new sample, the classifier applies the patterns and relationships it learned from the training dataset to the sample being classified, and assigns it a label based on the most likely match. There are many different types of machine learning classifiers, including decision trees, Naıve Bayes classifiers. The choice of the classifier depends on the specific problem and dataset being used. Each of these is described below:

2.5.1 Decision Tree Classifier

The decision tree classifier is a type of machine learning algorithm that is used for classification tasks. It is a supervised learning algorithm, which means that it is trained using labeled training data [28]. The training data consists of examples of the data that the classifier will be asked to classify, along with the correct class labels for each example. The decision tree classifier works by building a model of the training data in the form of a tree structure. Each node in the tree represents a decision or a rule that is used to classify the data. The decision tree is built by starting at the root node and splitting the data into smaller and smaller subsets based on the values of the features in the data. Once the decision tree is built, it can be used to classify new samples by following the decisions and rules in the tree. To classify a new sample, the classifier starts at the root node and follows the decisions in the tree until it reaches a leaf node, which contains the predicted class label for the sample.

2.5.2 Naive Bayes Classifier

The Naive Bayes classifier is a type of machine learning algorithm that is used for classification tasks. It is based on the idea that the presence or absence of a particular feature in a class is independent of the presence or absence of any other feature. This assumption, which is called the "naive" ssumption, allows the algorithm to make predictions based on a relatively small amount of data. The Naive Bayes classifier works by using Bayes' theorem to calculate the probability that a given sample belongs to each possible class [29]. Bayes' theorem states that the probability of an event A given another event B is equal to the probability of B given A times the probability of A, divided by the probability of B. To classify a new sample, the Naive Bayes classifier uses Bayes' theorem to calculate the probability that the sample belongs to each possible class, based on the features of the sample and the training data. The class with the highest probability is chosen as the predicted label for the sample.

3. RESULTS AND DISCUSSION

This study employed ECG data as input for the automated and accurate detection of DM patients. 24,630 (15070-normal and 9560-diabetic) labeled ECG segments are used to extract the ten statistically significant features, namely, log energy entropy, Shannon entropy, threshold entropy, sure entropy, root mean square (RMS) value, kurtosis, skewness, maximum value, signal energy, and Variance from each ECG fragments for classification of diabetic and normal subject ECG. The Kruskal-Wallis (KW test) is employed to check the statistical significance of the extracted features.

The K-W test is a non-parametric statistical test, used to determine whether there are significant differences between the medians of two or more groups. It is often used when the assumptions of an analysis of variance (ANOVA) test cannot be met, such as when the data are not normally distributed or have unequal variances. The p-value is used to measure the statistical significance of the test. It represents the probability of obtaining the observed results, or something more extreme, under the assumption that the null hypothesis is true. If the p-value is below a certain threshold, typically 0.05, the null hypothesis can be rejected and conclude that there is a statistically significant difference between the groups. The lower the p-value indicates the more suitability of the features. Table 1 shows the probability values (p) of each extracted feature with a mean and standard deviation of diabetic and normal ECG signals. It can be noted from Table 1, all the extracted features have a *p*-value less than 0.05, indicating that all features are statistically significant and suitable for classification. In machine learning, feature mapping is a necessary but crucial task for adequate classification. Therefore, all the extracted features are ranked and mapped based on their F-score. The higher F-score indicates a higher separation ability and the lower F-score depicts a lower separation ability. Consequently, features are ranked first for

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 11, Issue 2 | Pages 418-425 | e-ISSN: 2347-470X

higher F-score and ranked last for lower F-score. *Table 2* depicts the F-score and ranking of the features.

It can be noticed from *table 2*, the highest F-score for Log energy entropy is 0.032 and the lowest F-score for Kurtosis is 1:50 1008. Consequently, the first rank is assigned to Log energy entropy, and the lower/last rank is given to Kurtosis. The features are mapped based on the ranking of features from 1 to 10 and formed a feature matrix. Obtained feature matrix fed as an input to machine learning classifier with different kernel functions such as Medium Tree (MT), Coarse Tree (CT), Linear Discriminant (LD), Quadratic Discriminant (QD). The robustness of our developed system is evaluated by calculating the different performance parameters namely, true negative rate (*TNR*) or specificity (*SPE*), true positive rate (*TPR*) or sensitivity (*SEN*), Positive predicted value (P*PV*), negative predicted value (*NPV*), and *F*-1 score. The mathematical expression of all these performance characteristics is given as,

$$ACC\ (\%) = \frac{T_P+T_N}{T_P+T_N+F_P+F_N} \times 100 \quad (12)$$

$$SPE/TNR(\%) = \frac{T_N}{T_N+F_P} \times 100 \quad (13)$$

$$SEN/TPR\ (\%) = \frac{T_P}{T_P+F_N} \times 100 \quad (14)$$

$$NPV = \frac{T_N}{T_N+F_N} \times 100 \quad (15)$$

$$F-1\ Score = \frac{2\times Prc\times TPR}{Prc+TPR} \quad (16)$$

$$PPV = \frac{T_P}{T_P+F_P} \times 100 \quad (17)$$

Where *TP* and *TN* represent the count of correctly identified positive and negative episodes, respectively. Similar to this, the number of positive and negative events that were incorrectly detected is indicated as *FP* and *FN*, respectively. Tables 3 and 4 show the evaluated parameters for the one to ten ranked features with different classifiers. It can be noted from Tables 3 and 4, the maximum classification *ACC* of 87.19%, SEP of 84.74%, and F1-score 0.9 are yielded by the decision tree classifier with MT kernel for the first five ranked features. However, the highest *SEN* of 93.85% is achieved by the CT kernel of the decision tree for all ten ranked features. Figure 4 shows the ROC plot achieved for the classification of the first five ranked features using a medium decision tree classifier. An area under the curve (AUC) of 0.92 is obtained by the medium decision tree classifier. Then several methods has been developed for detection of diabetes using biological signals. These methods made use of various signal processing methods, signal types, and approaches. For instance, In [23], Acharya et al. developed employed non-linear features of R-R interval signals with AdaBoost classification scheme for diagnosis of diabetic subjects with ACC of 90% in 32 subjects. In [30] presented a diabetes index-based method using HRV features to determine diabetic neuropathy. They used Ad-Boost scheme for identification of diabetic subject and reported an ACC of 86%. Jian et al. [31] explored higher-order statistics (HOS) of HRV with SVM an ACC of 79.93% for detection of diabetes. In another study, authors of [32] diagnosed the diabetic patients by applying several artificial intelligence methods and attained a PPV value of 77%. Furthermore, Gupta et al. [25] utilized the ITD method to extract PPV and evaluated features from the obtained components. They used DTC and reported an ACC of 86.9%. The advantages of the presented system are listed below.

1. The described strategy is relatively simple to implement and can be used to enhance medical diagnosis performance.
2. The proposed framework is very effective and achieves higher classification performance by mapping the statistical significance features.
3. The presented system is accurate and robust because it is designed using the 10-fold validation scheme.

The proposed system's weaknesses are listed below.

1. The suggested characteristics of ECG signals have been determined at a predetermined sampling frequency. It is necessary to investigate the effect of different sample frequencies on categorization accuracy.
2. Only Type-2 diabetic patients have been identified in this study.

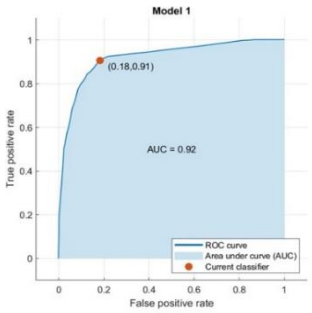

Figure 4: Receiver operating characteristic curve plots

Table 1: The Features mean ± STD for healthy and diabetics ECG signals

S. No	Features	Normal (mean ± std)	Diabetics (mean ± std)	p-Value
1	Log energy entropy	7.63×10^{03} ± 960.4026	7.29×10^{03} ± 935.0185	4.07×10^{-198}
2	Shannon entropy	-1.14×10^{08} ± 5.45×10^{08}	-1.06×10^{08} ± 3.59×10^{08}	0.0374
3	Threshold entropy	1.26×10^{03} ± 38.0873	1.25×10^{03} ± 42.6626	1.15×10^{-80}
4	Sure entropy	1.16×10^{04} ± 642.4346	1.14×10^{04} ± 736.5119	7.66×10^{-90}
5	Root mean square value (RMS)	78.9859 ± 39.4865	76.8561 ± 36.3737	0.0015
6	Kurtosis	9.3031 ± 4.8946	9.3018 ± 5.2883	1.24×10^{-13}
7	Skewness	0.6705 ± 1.3031	0.8495 ± 0.9368	3.89×10^{-4}
8	Maximum value	349.5088 ± 245.3226	355.5921 ± 252.9814	1.98×10^{-9}
9	Signal energy	9.98×10^{06} ± 3.25×10^{07}	9.25×10^{06} ± 2.30×10^{07}	0.0015
10	Variance	7.77×10^{03} ± 2.53×10^{04}	7.19×10^{03} ± 1.79×10^{04}	4.97×10^{-4}

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 11, Issue 2 | Pages 418-425 | e-ISSN: 2347-470X

Table 2: Features F-score and ranking

S. No	Features	F-Score	Ranking
1	Log energy entropy	0.032	1
2	Shannon entropy	6.02×10^{-05}	9
3	Threshold entropy	2.15×10^{-03}	4
4	Sure entropy	1.44×10^{-02}	2
5	RMS value	7.86×10^{-04}	5
6	Kurtosis	1.50×10^{-08}	10
7	Skewness	6.18×10^{-03}	3
8	Maximum value	1.49×10^{-04}	8
9	Signal energy	1.66×10^{-04}	7
10	Variance	1.75×10^{-04}	6

4. CONCLUSIONS

Diabetes is a serious health condition that is becoming more common around the world. Many people with diabetes have died from COVID-19, highlighting the importance of managing the disease. Currently, there is no cure for diabetes, so it is important for people with the condition to undergo regular monitoring and take their prescribed medications.

Table 3: Performance parameters for first 1 to 5 ranked features with different classifier

Nf	Classifier	Acc	SEN	SPE	NPV	PPV	F-score
1	MT	63.20	62.80	71.03	8.74	97.74	0.76
	CT	63.12	62.59	76.97	7.34	98.60	0.77
	LD	62.76	62.99	60.02	12.12	94.88	0.76
	QD	62.83	62.93	61.53	11.31	95.51	0.76
	GNB	62.83	62.93	61.53	11.31	95.51	0.76
2	MT	64.57	66.37	57.91	31.69	85.41	0.75
	CT	63.50	62.77	84.15	7.33	99.12	0.77
	LD	62.41	62.82	57.66	11.88	94.47	0.75
	QD	62.85	63.14	59.70	13.17	94.36	0.76
	GNB	61.75	63.46	51.93	19.75	88.40	0.74
3	MT	79.60	79.26	80.37	62.77	90.27	0.84
	CT	70.85	92.68	57.74	92.92	56.85	0.70
	LD	65.83	64.87	75.39	17.78	96.32	0.78
	QD	68.29	74.00	59.20	58.87	74.27	0.74
	GNB	64.23	66.48	56.64	33.40	83.78	0.74
4	MT	79.57	79.20	80.43	62.59	90.34	0.84
	CT	70.86	92.65	57.75	92.89	56.89	0.70
	LD	66.16	65.02	77.64	17.98	96.72	0.78
	QD	66.96	71.04	58.72	50.08	77.66	0.74
	GNB	63.73	65.32	56.81	27.36	86.80	0.75
5	**MT**	**87.19**	**88.66**	**84.74**	**81.73**	**90.66**	**0.90**
	CT	70.85	92.68	57.74	92.92	56.85	0.70
	LD	66.74	65.31	81.59	18.49	97.35	0.78
	QD	71.26	73.16	66.82	51.56	83.76	0.78
	GNB	64.02	66.07	56.58	31.44	84.69	0.74

Table 4: Performance parameters for first 6 to 10 ranked features with different classifier

Nf	Classifier	Acc	SEN	SPE	NPV	PPV	F-score
6	MT	84.43	86.09	81.56	77.36	88.91	0.87
	CT	71.71	93.80	58.42	94.01	57.56	0.71
	LD	67.26	65.73	81.76	20.16	97.15	0.78
	QD	73.10	72.35	75.56	45.37	90.69	0.80
	GNB	63.87	65.79	56.51	30.04	85.34	0.74
7	MT	84.57	86.69	81.05	78.59	88.36	0.88
	CT	71.69	93.75	58.44	93.96	57.55	0.71
	LD	67.33	65.71	82.83	20.24	97.33	0.78
	QD	77.73	88.47	66.73	84.98	73.13	0.80
	GNB	63.60	70.04	53.17	52.27	70.79	0.70
8	MT	84.18	86.64	80.19	78.68	87.67	0.87
	CT	71.73	93.83	58.44	94.03	57.58	0.71
	LD	67.26	65.81	80.26	20.75	96.76	0.78
	QD	78.08	85.59	68.83	79.53	77.15	0.81
	GNB	61.35	75.94	50.14	73.09	53.90	0.63
9	MT	84.27	87.05	79.86	79.53	87.27	0.87
	CT	71.75	93.73	58.47	93.92	57.68	0.71
	LD	67.28	65.79	80.88	20.58	96.91	0.78
	QD	77.65	89.33	66.26	86.42	72.08	0.80
	GNB	55.73	80.87	46.24	86.50	36.22	0.50
10	MT	84.91	87.83	80.35	80.90	87.45	0.88
	CT	71.73	93.85	58.44	94.05	57.57	0.71
	LD	69.62	67.79	81.44	28.14	95.93	0.79
	QD	77.58	92.35	65.11	90.98	69.07	0.79
	GNB	56.67	81.80	46.86	86.84	37.52	0.51

One way to monitor diabetes is through the use of ECG signals, which can detect changes in the heart's activity. However, visual inspection of these signals can be difficult due to their non- linear and non-stationary nature. In this study, ten statistical significance features have been extracted and mapped according to their F-score value and given to different machine learning classifiers. The proposed framework achieved an accuracy of 87.19% in detecting diabetes using ECG signals.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 11, Issue 2 | Pages 418-425 | e-ISSN: 2347-470X

The benefits of this approach include the ability to diagnose diabetes non-invasively and reduce the workload for clinicians. In the future, the model will be validated with different sampling frequencies and applied to diagnose other diseases and diabetic groups.

REFERENCES

[1] Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, Stein C, Basit A, Chan JC, Mbanya JC, Pavkov ME. IDF diabetes atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes research and clinical practice. 2021 Dec 6:109119

[2] Hippisley-Cox J, Coupland C. Diabetes treatments and risk of amputation, blindness, severe kidney failure, hyperglycaemia, and hypoglycaemia: open cohort study in primary care. BMJ. 2016 Mar 30; 352: i1450. doi: 10.1136/bmj.i1450. PMID: 27029547; PMCID: PMC4816603.

[3] Liang, Y.M.D.;Wanderer, J.M.D.M.; Nichols, J.H.P.; Klonoff, D.M.D.F.; Rice, M.J.M.D. Blood Gas Analyzer Accuracy of Glucose Measurements. Mayo Clin. Proc. 2017, 92, 1030^a¿\1041.

[4] McMillin, J.M. Blood Glucose. In Clinical Methods: The History, Physical and Laboratory Examinations, 3rd ed.; Walker, H.K., Hall, W.D., Hurst, J.W., Eds.; Butterworth-Heinemann: Boston, MA, USA, 1990.

[5] Delost, M.E. Blood Gas and Critical Care Analyte Analysis. In Equipment for Respiratory Care; Volsko, T.A., Chatburn, R.L., El-Khatib, M.F., Eds.; Jones and Bartlett Publishers, Inc.: Sudbury, MA, USA, 2014.

[6] Burrin, J.M.; Price, C.P. Measurement of Blood Glucose. Ann. Clin. Biochem.1985, 22, 327-342.

[7] Dalvi, N. Glucose meter reference design. In Application Note Nr. 1560; Microchip Technology Inc.: Chandler, AZ, USA, 2013.

[8] Gjessing, H. J., L. E. Matzen, O. K. Faber, and A. FrA¸land. Fasting ~ plasma C-peptide, glucagon stimulated plasma C-peptide, and urinary Cpeptide in relation to clinical type of diabetes Diabetologia 32 (5);1989: 305-311.

[9] Foteinopoulou, Evgenia, Catriona AL Clarke, Rebecca J. Pattenden, Stuart A. Ritchie, Emily M. McMurray, Rebecca M. Reynolds, Ganesan Arunagirinathan, Fraser W. Gibb, John A. McKnight, and Mark WJ Strachan. Impact of routine clinic measurement of serum C-peptide in people with a clinician diagnosis of type 1 diabetes. Diabetic Medicine 38, no. 7; 2021; e14449.

[10] Goldstein, David E., Randie R. Little, Hsiao-Mei Wiedmeyer, Jack D. England, Curt L. Rohlfing, and Alethea L. Wilke. "Is glycohemoglobin testing useful in diabetes mellitus? Lessons from the Diabetes Control and Complications Trial." Clinical Chemistry 40, no. 8;1994; 1637-1640.

[11] Mitchai M, Suwansaksri N, Seanseeha S, Saenboonsiri J, Kraitree P, Piyapromdee J, Silsirivanit A. Misleading HbA1c Measurement in Diabetic Patients with Hemoglobin Variants. Medical Sciences. 2021 Jun;9(2):43.

[12] Chiappin S, Antonelli G, Gatti R, De Palo EF Saliva specimen: a new laboratory tool for diagnostic and basic investigation. Clin Chim Acta 383: 2007; 30-40.

[13] Kaufman E, Lamster IB The diagnostic applications of saliva{a review. Crit Rev Oral Biol Med 13: 2002; 197-212

[14] Agrawal, R.P., Sharma, N., Rathore, M.S., Gupta, V.B., Jain, S., Agarwal, V. and Goyal, S., 2013. Noninvasive method for glucose level estimation by saliva. J Diabetes Metab, 4(5), p.266.

[15] Amer, S., Yousuf, M.U.N.I., Siddqiui, P.Q. and Alam, J., 2001. Salivary glucose concentrations in patients with diabetes mellitus^a¿\a minimally invasive technique for monitoring blood glucose levels.Pak J Pharm Sci, 14(1), pp.33-7.

[16] McShane, M.; Stein, E. Fluorescence-Based Glucose Sensors. In In Vivo Glucose Sensing; Cunningham, D.D.,Stenken, J.A., Eds.; John Wiley& Sons, Inc.: Hoboken, NJ, USA, 2009; Volume 174, pp. 269-316.

[17] Malik, B.H.; CotA~©, G.L. Real-time, closed-loop dual-wavelength optical polarimetry for glucose monitoring.J. Biomed.Opt. 2010, 15, 017002.

[18] Rawer, R.; Stork,W.; Kreiner, C.F. Non-invasive polarimetric measurement of glucose concentration in theanterior chamber of the eye. Graefes Arch. Clin. Exp. Ophthalmol. 2004, 242, 1017-1023

[19] Shu T, Zhang B, Tang YY. An improved noninvasive method to detect Diabetes Mellitus using the Probabilistic Collaborative Representation based Classifier. Information Sciences. 2018 Oct 1; 467: 477-88.

[20] Kim, S., Cho, J., Ku, B., Jun, M., Kim, G., Yoo, H., Park, S. and Kim, J.U., 2019. Variability of electrochemical skin conductance for screening diabetes mellitus. Biomedical Engineering Letters, 9(2), pp.267-274.

[21] Howlader, K.C., Satu, M.S., Awal, M.A., Islam, M.R., Islam, S.M.S., Quinn, J.M. and Moni, M.A., 2022. Machine learning models for classification and identification of significant attributes to detect type 2 diabetes. Health information science and systems, 10(1), p.2.

[22] Maniruzzaman, M., Rahman, M.J., Ahammed, B. and Abedin, M.M., 2020. Classification and prediction of diabetes disease using machine learning paradigm. Health information science and systems, 8, pp.1-14.

[23] Acharya U.R., Faust, O., Kadri, N.A., Suri, J.S., Yu, W. Automated identification of normal and diabetes heart rate signals using nonlinear measures, Comput. Biol. Med. 43 (10) (2013) 1523-1529 2013

[24] Shashikant, R., Chaskar, U., Phadke, L. and Patil, C., 2021. Gaussian process-based kernel as a diagnostic model for prediction of type 2 diabetes mellitus risk using non-linear heart rate variability features. Biomedical Engineering Letters, 11(3), pp.273-286.

[25] K. Gupta, V. Bajaj, A robust framework for automated screening of diabetic patient using ecg signals, IEEE Sensors Journal 22 (24) (2022) 24222-24229

[26] Patel, P. R, Annavarapu, R.N. EEG-based human emotion recognition using entropy as a feature extraction measure. Brain Inf. 8, 20 (2021).

[27] Lin Y.P. et al., EEG-based emotion recognition in music listening, IEEE Trans. Biomed. Eng., vol. 57, no. 7, pp. 1798-1806, Jul. 2010

[28] Swain P.H. and Hauska H., The decision tree classifier: Design and potential, IEEE Trans. Geosci. Electron., vol. GE-15, no. 3, pp. 142{147, Jul. 1977.

[29] Rish, Irina. An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence, vol. 3, no. 22, pp. 41-46. 2001.

[30] Acharya, UR, Faust, O. Sree, S. V., Ghista, D. N. Dua, S. Joseph, P. Ahamed, V. T., Janarthanan, N., Tamura T., An integrated diabetic index using heart rate variability signal features for diagnosis of diabetes, Computer methods in biomechanics and biomedical engineering 16 (2) (2013) 222{234}.

[31] Jian, L. W, Lim T.C., Automated detection of diabetes by means of higher order spectral features obtained from heart rate signals, Journal of medical imaging and health informatics 3 (3) (2013) 440-447.

[32] Mercaldo F., Nardone V., Santone A., Diabetes mellitus affected patients classification and diagnosis through machine learning techniques, Procedia computer science 112 (2017) 2519-2528