

Efficient Data-Mining Classification Approach For Ecg Data In Health Care Application

Shana J (✉ shanajdrpf@gmail.com)

Coimbatore Institute of Technology <https://orcid.org/0000-0002-0649-2056>

Venkatachalam T

Coimbatore Institute of Technology

Research

Keywords: Data mining , Classification, ECG Signal, Power Line Interference, Adaptive Notch Filters

Posted Date: March 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-277577/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Efficient Data-Mining Classification Approach For Ecg Data In Health Care Application

Dr. J. Shana^{1*}, Dr. T. Venkatachalam²

¹Department of Computer Applications in Coimbatore Institute of Technology, India.
Correspondence Email – shanajdrpf@gmail.com
²Department of Physics at Coimbatore Institute of Technology, India

Abstract

Data mining enables classification of Electrocardiographic (ECG) signals of the heart for diagnosing many cardiac diseases. ECG signals often consist of unwanted noises, speckles and redundant features. An unwanted noise and redundant features always degrade the quality of ECG signal and may lead to loss of accuracy in classification technique. To overcome these challenges, we introduced Optimize Discrete Kernel Vector (ODKV) classifier with an impressive pre-processing in this paper. In order to remove the noises, image processing filter namely the Adaptive Notch Filters (ANF) are initially used to remove Power Line Interference from ECG Signals. Moreover, reducing the redundant features from the ECG signal plays a vital role in diagnosing the cardiac disease. So, Optimize Discrete Kernel Vector (ODKV) classifier is used to reduce the redundant features and also to enhance the classification accuracy of the input ECG signal. Thus, Optimize Discrete Kernel Vector (ODKV) classifier identifies the Q wave, R wave and S wave in the input ECG signal. Finally, performance metrics Sensitivity, Specificity, Accuracy and Mean Square Error (MSE) are calculated and compared with the existing method such as SVM-kNN, ANN-kNN, GB-SVNN, and CNN to prove the enhancement of the classification technique.

Keywords : Data mining , Classification, ECG Signal, Power Line Interference, Adaptive Notch Filters

1. Introduction

Along with stroke and heart diseases, almost 17 million cardiovascular diseases (CVD) deaths occur each year. The World Health Organization (WHO) demonstrates that about 16.7 million inhabitants die from CVD each year. By 2020, cardiovascular sickness, the estimate of mortality, 20 million deaths annually by 2020, and 24 million deaths annually by 2030, will be the primary cause of death and universal disability. A cardiac arrhythmia is a form of abnormal heartbeat. We require an initial diagnosis and prognosis to decrease deaths due to cardiovascular disease that demands a precise and reliable biomedical diagnostic technique [1]. Some main causes of heart attacks are lack of physical activity, Smoking, elevated cholesterol, hypertension, destructive use of liquor, and undesirable diet [2]. To measure the heart's electrical activity, an electrocardiogram (ECG) procedure is used to monitor many electrical potentials of the heart [3].

Considering its versatility and non-invasive nature, it has been commonly used for predicting heart diseases. It can be branched approximately through the Inside stages of the depolarization and repolarization of the heart-forming muscular tissue fibers. P wave (Depolarization in the atrium) and QRS-wave lead the depolarization for the phases (ventricles depolarization). T-wave and U-wave (ventricular repolarisation) [4] refer to the repolarisation processes. By observing the electrical signal of each pulse, the composition impulse-of-action frequency modulation is generated by various sensitive tissues of the cardiac positioned inside the heart [5], which can identify some of the anomalies. A significant amount of information is stored in the ECG signal and can be used in various ways. It makes it possible to evaluate heart health problems. Doctors advise the patient to have the ECG registered before the full diagnosis of the patient for all forms of cardiac disorders and problems indirectly associated with the heart [6]. The mechanism of identification and classification of the ECG signal can be very troublesome for a person because, for example, every heartbeat of records from the ECG collected by the holder display can often be analyzed over for hours, or even for days. Furthermore, there is a chance of human mistakes during the time of analysis of ECG data due to exhaustion. It's an idea to use methods for automatic categorization.

The continuous detection of irregular heartbeats from a large heartbeat volume data on ECG [7] is an effective and important task. Therefore, solutions are required in the management and exploration of such massive, diverse, and enormous sets of data with a fair complexity of time and storage space. Wide and complex databases play a key function in processing the large quantity of data on health care and improving the efficiency of healthcare services for the patients. In handling the vast amount of health-care data and enhancing the quality of health-care services delivered to patients, Big Data analytics plays a vital role. One, in this sense, the challenges are data classification, which relies on effectively distributed computing platforms, advanced data mining and techniques for machine learning. [8]. So data mining is a cognitive method of finding the big data set's secret approach patterns. It is widely used in applications like critical thinking, media distribution industry, retail, genetic data analysis, financial data, logical applications, health mind frameworks, etc [9].

Using Data Mining, we can not only substantially reduce this amount, as well as save doctors and patients' time and resources [10]. Due to the enormous data set size [11], it is important to provide faster and more cost-effective models while using data mining techniques to carry out the classification of ECG datasets. Moreover this data set consists of noise and redundant functions. For this function, to find a collection of appropriate features [12], efficient noise reduction and feature selection techniques are used. When constructing classification models, these selected characteristics are rated as the most essential.

2. Related Work

The author in [13] studied the methods like anisotropic diffusion, adaptive filters, wavelet transforms condensation of empirical mode, morphological filter, the algorithm of non-local means and denoising total variation was utilized for examination. It is concluded that the best denoising is generated by multivariate wavelet denoising or wavelet-PCA and is, therefore, the most suitable for DVP enhancement applications in the real world. Therefore the two-dimensional function of a Gaussian derivative was used in the images to obtain the spatiotemporal and most important spectral characteristics. The key purpose of the filters in two-dimensions was to manage and remove most of the significant characteristics from multi-scale images. Principal Components Analysis (PCA) was included to extract GD features from the output image produced from the two-dimensional image in this work filter. The efficiency of the proposed extraction method of the two-dimensional filter-based image function was examined with the support of different picture perspectives [14]. Although it is not sufficient to decrease the noise in Signal Processing, specifically signals from ECG.

The different stages of signal from ECG pre-processing, including the treating of imbalanced data, data normalization, and noise filtering through band pass filter and feature extraction method, were defined as the Random Forest machine learning method to analyze the ECG signal. The accuracy of this model is not properly indicated [15]. To enhance the accuracy, In this paper, the author suggest edit's not just an extraction technique of Multiple purpose vectors from ultrasound images for electrocardiogram signal Carotid Arteries (CAS) and Heart Rate Variability (HRV), but also an electrocardiogram signal effective and accurate prediction model in diagnosing the disease of Cardiovascular Disease (CVD) using SVM and showed about 89.51% after evaluating the diagnosis or prediction approaches in terms of diagnosing accuracy rate utilizing the multiple feature vectors are chosen [16]. Despite this, the presence of noise is affected by the accuracy of classification. To be classified into usual and abnormal subjects, delayed error normalized LMS filtering method using ECG signal pre-processing domain features of HRV Frequency and are adapted to SVM classifier-based Classification of the arrhythmic beat. Thus, noise is reduced insufficiently. The SVM-based classifier developed system provides maximum accuracy of 96% for the classification of normal and arrhythmic abnormal risk subjects [17].

Cardiac disease diagnosis such as Arrhythmia using ECG recording was performed in [18] by wrapper-based feature selection technique and classification of multi-classes. In detecting the frequency and lack of

arrhythmias, Support Vector Machine (SVM) approaches based on it like One-Against-One (OAO), One Against-All (OAA), and Error-Correction Code (ECC) are used for multiclass classification. Accordingly, the OAO method of SVM provided an accuracy rate of 81.11%. Despite this, it is not presented sufficient accuracy in classification. Using a k-Nearest Neighbour algorithm and statistical features, 5-second ECG segments were classified into good-quality and bad-quality levels for signal quality classification in [19] and achieved a 96.87% average. Classifying blood pressure records obtained from the analysis of the Electrocardiogram (ECG) using the SVM classifier could reach an acceptable accuracy of 98.18% in [20]. As a portion of the filter-based feature selection techniques, an effective feature that the search algorithm for harmony can be altered and used in combination with other evaluators of function subsets. It is possible for the proposed expert systems to also be willing to be readily accessible and utilized by other biomedical indications, for instance, electrocardiography (ECG) and electromyography (EMG) signals, for classification tasks.

3. Problem Statement

The existence of PLI noise and redundant ECG signal features influences the precise classification of the recording of ECG signals that may aid in the diagnosis and care of patients with heart disease. It can find out the following validation metrics,

- Sensitivity
- Specificity
- Accuracy
- Mean Square Error

3.1 Sensitivity

The sensitivity is calculated based on the correct positive rate. It can be defined as the number of positive predictions correctly divided by the total number of positive predictions. It is called even as True Positive Rate (TPR). The best sensitivity is 1.0, whereas the worst is 0.0.

$$Sensitivity = \frac{TruePositive(TP)}{TruePositive(TP) + FalseNegative(FN)} \quad (1)$$

3.2 Specificity

The specificity metric is used to predict the exact prediction. It is also defined and divided by the total number of negative predictions as the number of accurate negative predictions. Often it is known as a true negative rate (TNR). The highest specificity is 1.0, while the worst specificity is 0.0.

$$Specificity = \frac{TrueNegative(TN)}{FalsePositive(FP) + TrueNegative(TN)} \quad (2)$$

3.3 Accuracy

Accuracy (ACC) is determined as the sum of all right predictions divided by the dataset's total number. Here, 1.0 is the highest accuracy, while the lowest is 0.0. Also, it can be estimated through 1.

$$Accuracy = \frac{TruePositive + TrueNegative}{TruePositive + TrueNegative + FalseNegative + FalsePositive} \quad (3)$$

3.4 Mean Square Error (MSE)

Mean Square Error (MSE) is computed and centered based on inequality between an estimator and the true value of the calculated quantity. MSE is determined by,

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

4. Methodology

To capture the ECG signals of patients, wearable medical devices were used. To classify the heartbeat we proposed an Optimize Discrete Kernel Vector (ODKV) classifier for the prediction of cardiac disease. The noise in the ECG signal interrupts the classification accuracy. Thus initially, the Adaptive Notch Filter (ANF) is used to delete the power line of interference present in the signal from the ECG input. To minimize the characteristics present in the ECG signal inputs and for classification, the Optimize Discrete Kernel Vector classifier is utilized. To identify the heartbeat stage as Left Bundle Branch Block (LBBB), Right Bundle Branch Block (RBBB), Premature Ventricular Contraction (PVC), and Premature Atrial Contractions (PACs) for the prediction of cardiac disease, the Optimize Discrete Kernel Vector classifier approach substantially distinguish the input ECG signal, Q wave, R wave, and S wave. The MHEALTH (Mobile HEALTH) dataset is collected from UCI machine learning repository. The number of instances in this dataset is 120 and number of attributes is 23. Shimmer2 [BUR10] wearable sensors were used for the recordings. Using ECG machine the dataset is collected. The proposed Optimize Discrete Kernel Vector classifier method for classification of ECG signal and Adaptive Notch Filters to remove the power line interference during the noise removal process.

Comparatively, with existing approaches, the efficiency of the proposed Optimize Discrete Kernel Vector classifier is evaluated. To predict cardiac disease, the classification of the other input ECG signal is categorized as Left Bundle Branch Block (LBBB), Right Bundle Branch Block (RBBB), Premature Ventricular Contraction (PVC), and Premature Atrial Contractions (PACs). The categorization of heartbeat outcomes is utilized to take the appropriate steps for the patients by the doctors. The workflow of the ECG signal classification system is represented in Fig. 1.

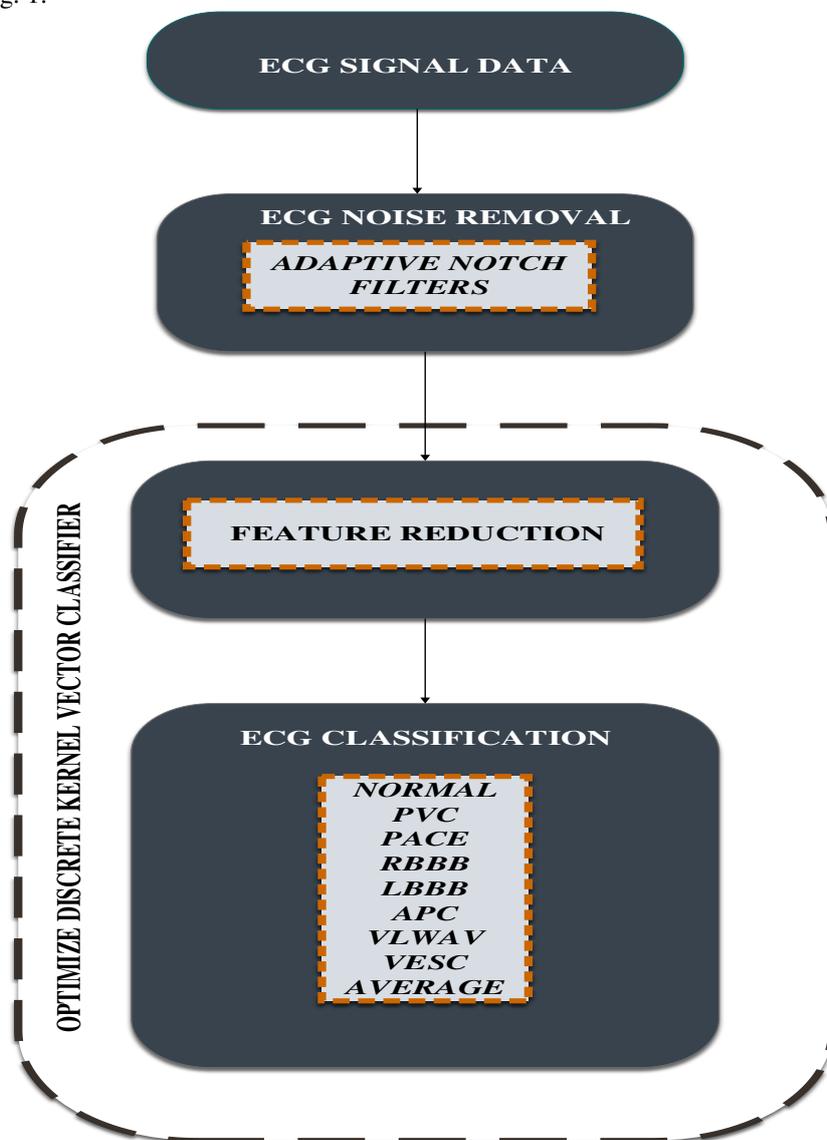


Figure 1 Architecture of proposed method

4.1 Image Processing

ECG recordings are typically tainted by noise and objects of various kinds. The aims are to minimize certain noise and artifacts in the pre-processing phase to assess the fiducial points and to prevent amplitude and offset effects to compare signals from various patients. The ECG classification is an important task because the signal contains an excessive number of unrestful noises. The different noise in different degrees in the classification of cardiovascular disease allows a physician to make inaccurate diagnoses of patients and decreases the correctness of the diagnosis. ECG signal denoising and pre-processing are now becoming a discriminatory necessity. During the recording of the ECG signal, matters are more complicated. Power line interference, baseline wandering, electrode touch noise, electrode motion artifacts, muscle contractions (electromyography noise, electrosurgical noise, and instrumentation noise are common types of noise. Among these, noise that is described as a signal at 50 or 60 Hz frequency, and below 1 Hz bandwidth, is discussed as power line interference (PLI). An Adaptive Notch Filter reducing/cancelling PLI has been suggested. Nonetheless, the frequency varies in the narrow band around the fundamental frequencies of 50 Hz in ECG Signal in real-time re-recording. It is due to differences in the power supplies available that comply with various requirements and thus result in wandering between 47-53 Hz. This contributes conceptually to the design of a relevant aim filter and can delete between the PLI frequencies and the 47-53 Hz band, but keep the valuable signal for ECG's signal frequencies in that range intact to prevent degradation of any type. The Architecture of Adaptive Notch Filter is shown in figure 2.

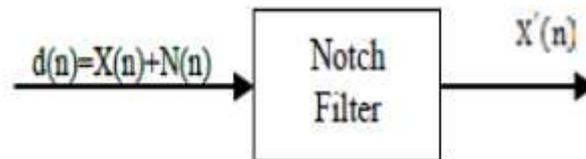


Figure 2 Architecture of Adaptive Notch Filter

The perverted ECG signal is composed of the 50 Hz noise and pure ECG signals are introduced. An ECG model corrupted by interference from the power line can be generated by

$$d(n) = x(n) + N(n)$$

Although, the 50 Hz sinusoidal signal is emitted like noise here. The $d(n)$ corrupted ECG signal is $x(n)$ then the pure signal of the ECG and $N(n)$ is the echo. Adaptive Notch Filters (ANF) is used to eradicate the Power Line Interference noise in the ECG signal.

4.2 Proposed Optimize Discrete Kernel Vector Classifier

The proposed Optimize Discrete Kernel Vector classifier is the second stage in this for the enhancement of the classification accuracy. The classification of the ECG signal helps to predict cardiac disease using heartbeat level.

The three main components consist of a single beat of an ECG signal: the P wave, the complex QRS and the T wave. Variations of these elements are linked to different heart features and disorders. In making an assessment or inquiry, cardiologists commonly use factors as well as other derived positions, magnitudes and modes of the waves and characteristics for example, QT interval, ST segment, PR interval and PR segment. ECG has been used to provide important insights into the prevention, care and prevention of heart disease and diagnosis, such as arrhythmias. In the proper heart diagnosis disease, the accurate representation of the signal of ECG plays a crucial role, since electrocardiography is an examination of the electrical activity of the heart. We can optimize different signal features, such as position, length, shape, altitude, peak points, etc. The significant step is the feature reduction of specific properties from biomedical signals (ECG, EEG, etc.) during the method of classification. Subsequent to the pre-processing stage, the need is to obtain the basic assets for use in the last one stage. A mathematical model that is supervised is Discriminant. It is characterised by high computational power, robust performance, and easy implementation. Discriminant is a widely used technique for ECG signal reduction of features. Discriminant aim the intention is to classify the low-dimensional subspace into which the scattering of the intra-class reconstruction is minimised although maximising scatter for inter-class

reconstruction. Suppose we have the perfect forecasts $P = \{\varphi_1, \varphi_2 \dots \varphi_d\}$ the best representation of all of their intra-class samples can be projected on which samples. Project each point of data x_i^j into the subspace:

$$y_i^j = P^T x_i^j \quad (6)$$

In the subspace, the sample dispersion of the intra-class reconstruction is

$$\sum_i \sum_j \|y_i^j - Y_i \beta_i^j\|^2 = \text{tr}(P^T S_W^R P) \quad (7)$$

$$\text{Where, } S_W^R = \sum_i \sum_j (x_i^j - X_i \beta_i^j)(x_i^j - X_i \beta_i^j)^T \quad (8)$$

The equation (8) above is labelled the scatter matrix of reconstruction within the intra-class.

In the subspace, the sample dispersion for inter-class reconstruction is

$$\sum_i \sum_j \sum_m \|y_i^j - Y_i \beta_m^j\|^2 = \text{tr}(P^T S_b^R P) \quad (9)$$

$$\text{Where, } S_b^R = \sum_i \sum_j \sum_m (x_i^j - X_m \beta_m^j)(x_i^j - X_m \beta_m^j)^T \quad (10)$$

The above equation (10) is labelled the scatter matrix of inter-class reconstruction.

In our method, the gap from a sample of x to a sample of x i.e., the i th class is defined as the error of reconstruction by the i th class,

$$d_i = \|x - X_i \beta_i\|^2 \quad (11)$$

In our method, the function we can obtain focuses scatter on the ranks of intra-class reconstruction and inter-class scatter reconstruction. Usually, the intra-class reconstruction dispersion and the inter-class reconstruction dispersion are both in a high-dimensional subspace of total rank. Therefore at most, Discriminant can reduce n characteristics.

After feature reduction ECG classification is done. Automatic ECG signal classification is a difficult problem for many reasons. Inconsistencies in the temporal and morphologic features of different patients' patterns of simple ECG can be seen in the waveform of the ECG. The waveforms of ECGs can be equivalent to different patients with various heart rhythms and can vary at various times with the same patient. Heart rate variability is also a concern involved in the ECG Signals Classification. Cardiac rate is dependent on physiological and behavioural problems factors such as pressure, arousal, and workout can induce changes in ECG characteristics such as RR interval, PR interval, etc. In addition to such problems, the absence of consistency of variations of ECG, the complexity signals of an ECG, the non-existence of optimal rules of classification, and the variation of the beat in a single ECG is the main problems complicating the classification of ECG signal. Several algorithms for ECG heartbeat detection and classification have been evolved. Most of these ECG beat methods of classification work is well done on the data for preparation, but offer poor output and the ECG signal of various patients as a result of the above difficulties. Owing to the absence of standardization of the classification algorithm in the development and assessment criteria, analysis of the results across most of these works could not be carried out.

The development of scalable classifier for data mining is the subject of several scientific studies. ODKV classifier is an effective technique widely used to address supervised classification issues due to its generalization ability. It is a binary classifier that tries to find a maximum margin hyper plane to represent the decision boundary, i.e., it defines a decision boundary with the greatest possible margin that can still distinguish the two classes. This not only decreases the chances of prediction errors but also reduces the high over-fitting possibility of limited margins inherent in decision limits. This is a classification algorithm that plots each data object as a point where n has become several characteristics with a particular coordinate value in the space n -dimensional. In this paper, we utilize the ODKV method to distinguish more ECG input signal. In the input ECG signal, the Optimize Discrete Kernel Vector process effectively distinguishes the Q wave, R wave, and S waves to identify the heartbeat stage, such as LBBB, RBBB, PVC, and PACs.

OPTIMIZE DISCRETE KERNEL VECTOR CLASSIFIER ALGORITHM:

1. *Data: ECG signal collected from the patients*
2. *Input: Health data table*
3. *Output: A weight vector \bar{w}*
4. *Step1: $R \leftarrow$ number of reducts in D*
5. *Step2: $F_d \leftarrow$ number of features in D*
6. *Step2: $F_r \leftarrow$ number of reducts in R*
7. *for ($x \leftarrow 0$ to F_d) do*
8. $w_x \leftarrow 0$;
9. *end*
10. *for ($x \leftarrow 0$ to F_d) do*
11. *for ($y \leftarrow 0$ to F_r) do*
12. *if (feature x in the y^{th} reducts R_y) then*
13. $v \leftarrow$ number of feature in R_y ;
14. $w_x \leftarrow w_x + \frac{1}{m}$;
15. *end*
16. *end*
17. *Range w_x into ($0 \leftarrow n$), where $n \leftarrow 1, 2, \dots, 100$*
18. *end*
19. *return \bar{w}*
20. *Compute $f(\bar{x}) = \text{sgn} (\sum_{i=1}^1 \alpha_i y_i K(\bar{w} \bar{x}_i, \bar{w} x_i) + b)$*

For the cases positioned in the borderline between two classes, the ODKV classifier uses a hyper plane that produces the greatest separation of values calculated from the decision function. For a labelled data $M = \{(x_{p,q}, y_p)\}$, where $p = 1, 2, \dots, n$; n stands for the complete number of samples of results, and ODKV classifier mapped the input vectors using a nonlinear kernel function to the desired value φ_x . If the mapping function associated with it is defined as $f(x)$ the judgment on the product of classification or mapping depends on the following equation:

$$f(x) = w^T \varphi(x) + b \quad (12)$$

Here, w is the vector of weight and b is the value of bias, and $f(x): R^n \rightarrow R$ is a decision function that produces the product of the classification for each input vector by linear classification x_p . So the product of the classification is:

$$z = \begin{cases} +1 & \text{if } f(x) \geq \Delta \\ -1 & \text{if } f(x) < \Delta \end{cases} \quad (13)$$

The parameters of W and b is the training data considering minimization of the cost function are decided by the training data. The cost function associated with this can be written as:

$$J(w, \xi) = \frac{1}{2} \|W\|^2 + C \sum_{p=1}^m \zeta_p \quad (14)$$

C stands for positive meaning and Δ means the threshold Value Specified by the User. Under the restriction determined under

$$y_i (W^T x_i + b) \geq 1 - \xi_i \quad (15)$$

Where, $i = 1, 2, \dots, m$ and C, ξ_i are two parameters specified by the user. Linear, polynomial, and sigmoidal are the popular functions used in ODKVC. Since Optimize Discrete Kernel Vector classifier can determine the degree of the pulse, such as PVC, PAC, LBBB, and RBBBs from the ECG signals. More efficient real-time discrimination could be possible after the extent of classification steps has been recombined. The ECG signal having features may directly or indirectly influence the complexity of the Optimize Discrete Kernel Vector

model while the network is being trained. The ODKV method identifies the heart disease in the ECG signal input by classifying the Amount of Pulse.

Here, the MSE is measured based on the differentiation between estimators and the real value of the measured quantity. The Mean Square Error Calculated for the proposed Optimize Discrete Kernel Vector form of help is roughly analyzed with other classification methods that they are SVM-kNN, ANN-kNN, GB-SVNN, and CNN. In contrast to other current deep learning methods, the proposed approach generates fewer MSE. The MSE is calculated by

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Where, n = Observation number, y_i = Values observed, \hat{y}_i = RMSE expected values are used to calculate the variation sin-between the values obtained through a model or an estimator and the observed values. Root Mean Squared Error is MSE's root square.

$$RMSE = \sqrt{MSE} \quad (16)$$

The R^2 (R-squared) is a mathematical issue degree for how near the knowledge to fit the regression line. R-squared is likewise called because of the determinative coefficient. The R^2 as follows is defined:

$$R^2 = 1 - \frac{SSE}{TSS} \quad (17)$$

i.e., SSE =Amount of Errors Squared, TSS = Complete Number of Errors Squared. SSE is the number of the square variations among the mean of every single observation and its party.

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$SSE = n \times MSE \quad (18)$$

TSS is described as the combination of every square difference between each observation and the sum of average over all observations. TSS is defined by

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (19)$$

Where, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, Hence, R^2 is defined by,

$$R^2 = 1 - \frac{n \times MSE}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (20)$$

Q^2 (Q-squared),it's the ratio of MSE about the difference in response. Q^2 is represented as,

$$Q^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (21)$$

The Coefficient of determination R^2 is represented as

$$R^2 \sim 1 - Q^2 \quad (22)$$

Mean Absolute Error tests the variations among prediction and actual observation average over the test sample, where all individual variations occur areequally equal in weight use. MAE is represented as,

$$MAE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (23)$$

The Mean Absolute Percent Error (MAPE) finds in percentage terms, the magnitude of the error. MAPE is characterized by,

$$MAPE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|^2 \quad (24)$$

For the proposed Kernel Functional Support Vector Machine method, RMSE, MAPE, MAE, R^2 , and Q^2 can be measured and compared with other methods, such as SVM-kNN, ANN-kNN, GB-SVNN, and CNN.

5. Result

The UCI's machine learning library gathers the MHEALTH (Mobile HEALTH) database. The number of occurrences is 120 in this dataset, and the set of attributes is 23. The sensors were positioned on the chest, right wrist, and left ankle of the person respectively and connected using elastic straps. The ANF filter is used to eliminate the interruption of the power line available in the ECG input signals. Predominantly, ECG signals with a noise like power line interference affect the performance of the classification method. From the given input ECG signal, the proposed Optimize Discrete Kernel Vector classifier significantly determines the Q, R, and S wave to characterize the heartbeat frequency for the prediction of heart disease, include LBBB, RBBB, PVC, and PAC. The assessment metrics were used to determine the ability of the qualified classifier to generalize. This classifier is reduced redundant data which is the cause of inaccurate classification. In this case, when evaluated with unseen data, the assessment metric is used to calculate and summarise the output of the qualified classifier. With regards to, Mean Square Error (MSE) is used to analyze the inaccurate classification based on error reduction i.e redundant data reduction. Accuracy is among the most common criteria used to assess the ability of classifiers to generalize by many researchers. So that, the following metrics such as Sensitivity, Specificity and Accuracy are used to finding the performance of classification using equation (1, 2, 3,)

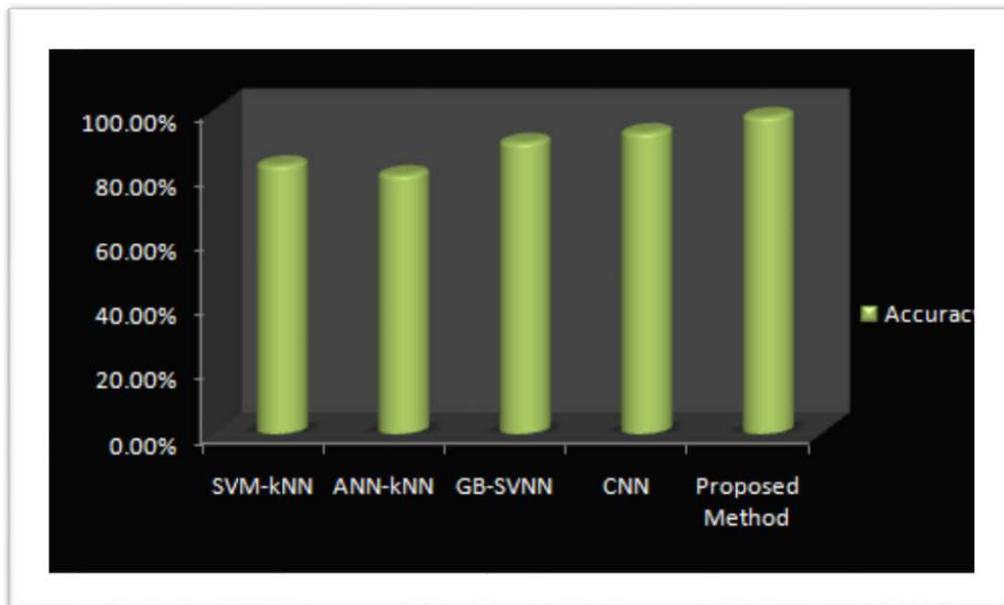


Figure 3 Comparison of accuracy

Here, the accuracy of the proposed Optimize Discrete Kernel Vector classifier is compared with the other classification methods of ECG signals like SVM-kNN, ANN-kNN, GB-SVNN, and CNN. It is obvious from the above figure 3 that the highest accuracy is obtained by our proposed classifier because the noise of power line interference in the ECG signal that disturbs the accuracy is eliminated in the pre-processing stage.

A critical review of the measured specificity of the proposed Optimize Discrete Kernel Vector classifier is performed using classification approaches such as SVM-kNN, ANN-kNN, GB-SVNN, and CNN. Figure 4 shows that, relative to other current methods, the proposed method produces a high sensitivity score.

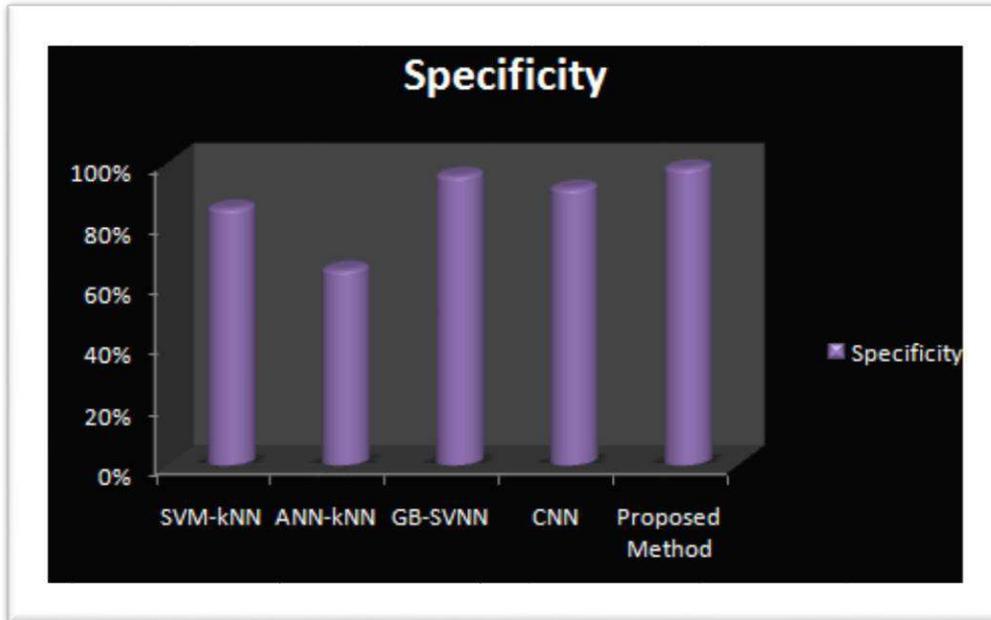


Figure 4 Comparison of Specificity

In comparison with classification methods such as SVM-kNN, ANN-kNN, GB-SVNN, and CNN, the measured sensitivity for the proposed Optimize Discrete Kernel Vector method is evaluated. Figure 5 shows that as compared to any other current methods, the proposed method produces a high sensitivity score.

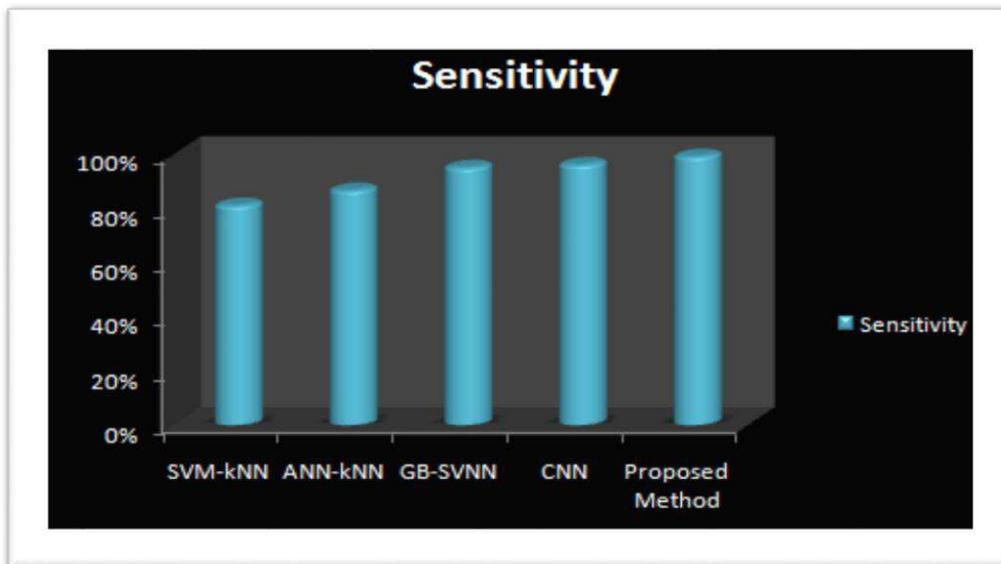


Figure 5 Comparison of sensitivity

As per the above discussion, to show the redundant data reduction is analyzed by the following metrics from table 1 for classification methods.

Table 1 Comparison of MAPE, RMSE, MAE, R^2 and Q^2 for the proposed and existing methods

Classification Methods	RMSE	MSE	SSE	TSS	R^2	Q^2	MAE	MAPE	MSD
SVM-kNN	0.631	0.123	29.324	33.234	0.234	0.755	0.4	5.435	8.66
ANN-kNN	0.123	0.121	8.323	12.011	0.145	0.6545	0.086	2.0125	4.14
GB-SVNN	0.342	0.129	13.123	16.643	0.2231	0.665	0.239	4.1154	7.17
CNN	0.432	0.125	29.234	33.444	0.344	0.854	0.5	5.532	8.56
ODKV	0.234	0.065	11.112	14.123	0.2123	0.712	0.123	4.0233	6.235

For the proposed Optimize Discrete Kernel Vector method, the measured MAPE, RMSE, MAE, R^2 and Q^2 are low compared to other methods, such as SVM-kNN, ANN-kNN, GB-SVNN, and CNN.

Table 2 Comparison of Proposed and Existing Classifiers

Existing Method	MSE	Sensitivity	Specificity	Accuracy
SVM-kNN	0.123	81%	85%	83.25%
ANN-kNN	0.121	86.58%	64.71%	80.60%
GB-SVNN	0.129	99%	95.83%	96.96%
CNN	0.125	96.01%	91.64%	93.47%
ODKV	0.065	99.13%	98.33%	98.06%

The standard of the presented work and is compared to four other classification methods. These were namely SVM-kNN, ANN-kNN, GB-SVNN, and CNN. The overall performance for the proposed and the existing method is shown in below figure 8. Moreover, the power line interference that exists in the ECG signal is withdrawn from the device using the Adaptive Notch Filter; the features in the signals are reduced using Discriminant method. Hence the Mean Square error is reduced and the performance of the classifier is get enhanced in terms of sensitivity, specificity, and accuracy.

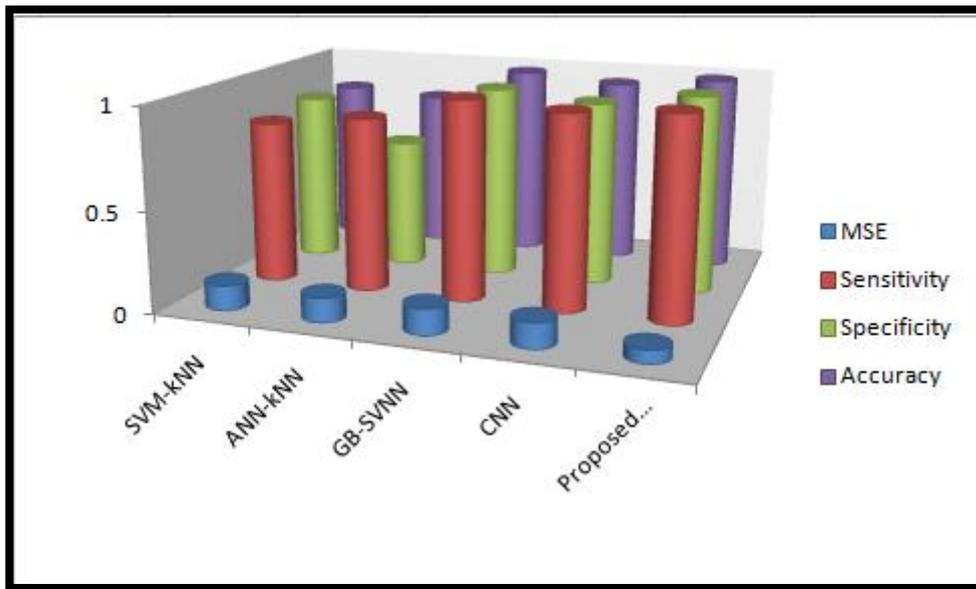


Figure 6 Overall output of the proposed and existing system

It is evident from Figure 6 that the accuracy of our classification is improved by eliminating the interference of the power line from the ECG signal and using an influential feature redundant method.

6. Conclusion

Electrocardiographic signals mostly consist of unwanted speckles and sounds. Various filters for Image Processing are used in different experiments to eliminate the noises. In this paper, the ANF filter is initially used to remove the power line interference that is present in the input ECG signal. To minimize the features present in the input ECG signal, Discriminant method is utilized. In this paper, to classify more of the input ECG signal characteristics, we use Optimize Discrete Kernel Vector classifier. The ODKV approach significantly distinguishes the Q wave, R wave, and S wave in the ECG input signal to identify the pulse stage for heart disease prediction includes LBBB, RBBB, PVC, and PACs. In comparison with other existing approaches such as SVM-kNN, ANN-kNN, GB-SVNN, and CNN, the performance of the proposed Optimize Discrete Kernel Vector classifier with effective noise removal and is evaluated. In comparison with other methods, such as SVM-kNN, ANN-kNN, GB-SVNN, and CNN, the measured MAPE, Q^2 , RMSE, R^2 , and MAE for the proposed Optimize Discrete Kernel Vector classifier method is low. Finally, to demonstrate the efficiency of the proposed ODKV classifier, sensitivity, specificity, and Mean Square Error (MSE) is measured.

Funding

Not Applicable

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Reference

- [1] Kumar, S. Udhaya, and H. Hannah Inbarani. "Neighborhood rough set based ECG signal classification for diagnosis of cardiac diseases." *Soft Computing* 21.16 (2017): 4721-4733.
- [2] Mathan, K., et al. "A novel Gini index decision tree data mining method with neural network classifiers for prediction of heart disease." *Design Automation for Embedded Systems* 22.3 (2018): 225-242.
- [3] Alarsan, Fajrlbrahem, and Mamoon Younes. "Analysis and classification of heart diseases using heartbeat features and machine learning algorithms." *Journal of Big Data* 6.1 (2019): 1-15.
- [4] Gupta, K. O., and P. N. Chatur. "Ecg signal analysis and classification using data mining and artificial neural networks 1." (2012).
- [5] Luz, Eduardo José da S., et al. "ECG-based heartbeat classification for arrhythmia detection: A survey." *Computer methods and programs in biomedicine* 127 (2016): 144-164.
- [6] Alfaouri, Mikhled, and Khaled Daqrouq. "ECG signal denoising by wavelet transform thresholding." *American Journal of applied sciences* 5.3 (2008): 276-281.
- [7] Sannino, Giovanna, and Giuseppe De Pietro. "A deep learning approach for ECG-based heartbeat classification for arrhythmia detection." *Future Generation Computer Systems* 86 (2018): 446-455.
- [8] Alarsan, Fajrlbrahem, and Mamoon Younes. "Analysis and classification of heart diseases using heartbeat features and machine learning algorithms." *Journal of Big Data* 6.1 (2019): 1-15.
- [9] Kinge, Durga, and S. K. Gaikwad. "Survey on data mining techniques for disease prediction." *International Research Journal of Engineering and Technology (IRJET)* 5.01 (2018): 630-636.
- [10] Kaushik, Divyansh, and Karamjit Kaur. "Application of Data Mining for high accuracy prediction of breast tissue biopsy results." 2016 Third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC). IEEE, 2016.
- [11] Alizadehsani, Roohallah, et al. "A database for using machine learning and data mining techniques for coronary artery disease diagnosis." *Scientific data* 6.1 (2019): 1-13.
- [12] Visalakshi, S., and V. Radha. "A literature review of feature selection techniques and applications: Review of feature selection in data mining." 2014 IEEE International Conference on Computational Intelligence and Computing Research. IEEE, 2014.
- [13] Tanantong, Tanatorn. "A KNN approach for ECG signal quality classification." *International Journal of Information and Electronics Engineering* 6.4 (2016): 269.
- [14] Mustaqeem, Anam, Syed Muhammad Anwar, and Muahammad Majid. "Multiclass classification of cardiac arrhythmia using improved feature selection and SVM invariants." *Computational and mathematical methods in medicine* 2018 (2018).
- [15] Zainuddin, Zarita, Kee Huong Lai, and Pauline Ong. "An enhanced harmony search based algorithm for feature selection: Applications in epileptic seizure detection and prediction." *Computers & Electrical Engineering* 53 (2016): 143-162.
- [16] SHABAAN, AR, and AA SHARAWI. "Machine learning for blood pressure classification using only the ecg signal." *Journal of engineering and applied science* 67.1 (2020): 257-274.
- [17] Sundar, Aditya, et al. "A comprehensive assessment of the performance of modern algorithms for enhancement of digital volume pulse signals." *International Journal of Pharma Medicine and Biological Sciences* 5.1 (2016): 91.
- [18] Kim, Hyeongsoo, et al. "A data mining approach for cardiovascular disease diagnosis using heart rate variability and images of carotid arteries." *Symmetry* 8.6 (2016): 47.
- [19] Mykoliuk, Iryna, et al. "Machine learning methods in electrocardiography classification." *ACIT 2018* (2018): 1-10.
- [20] Venkatesan, C., et al. "ECG signal preprocessing and SVM classifier-based abnormality detection in remote healthcare applications." *IEEE Access* 6 (2018): 9767-9773.

Figures

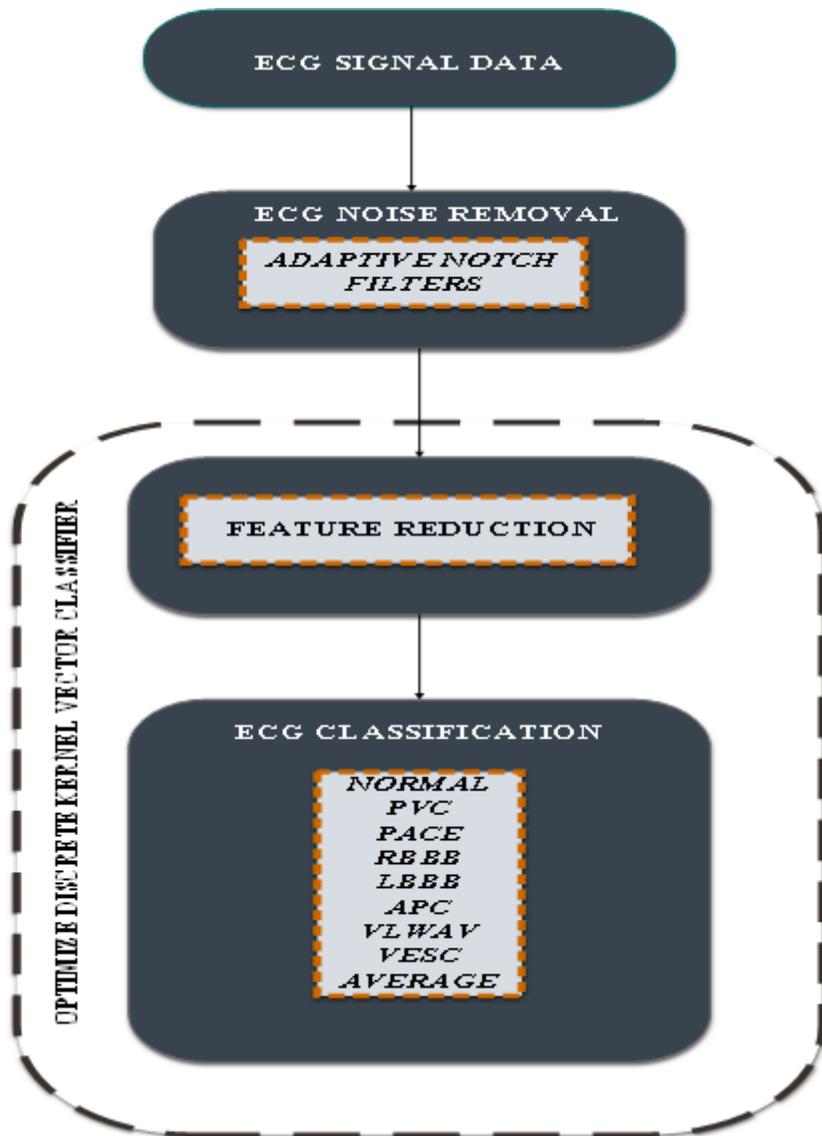


Figure 1

Architecture of proposed method

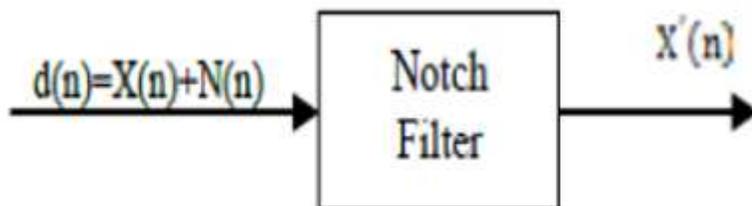


Figure 2

Architecture of Adaptive Notch Filter

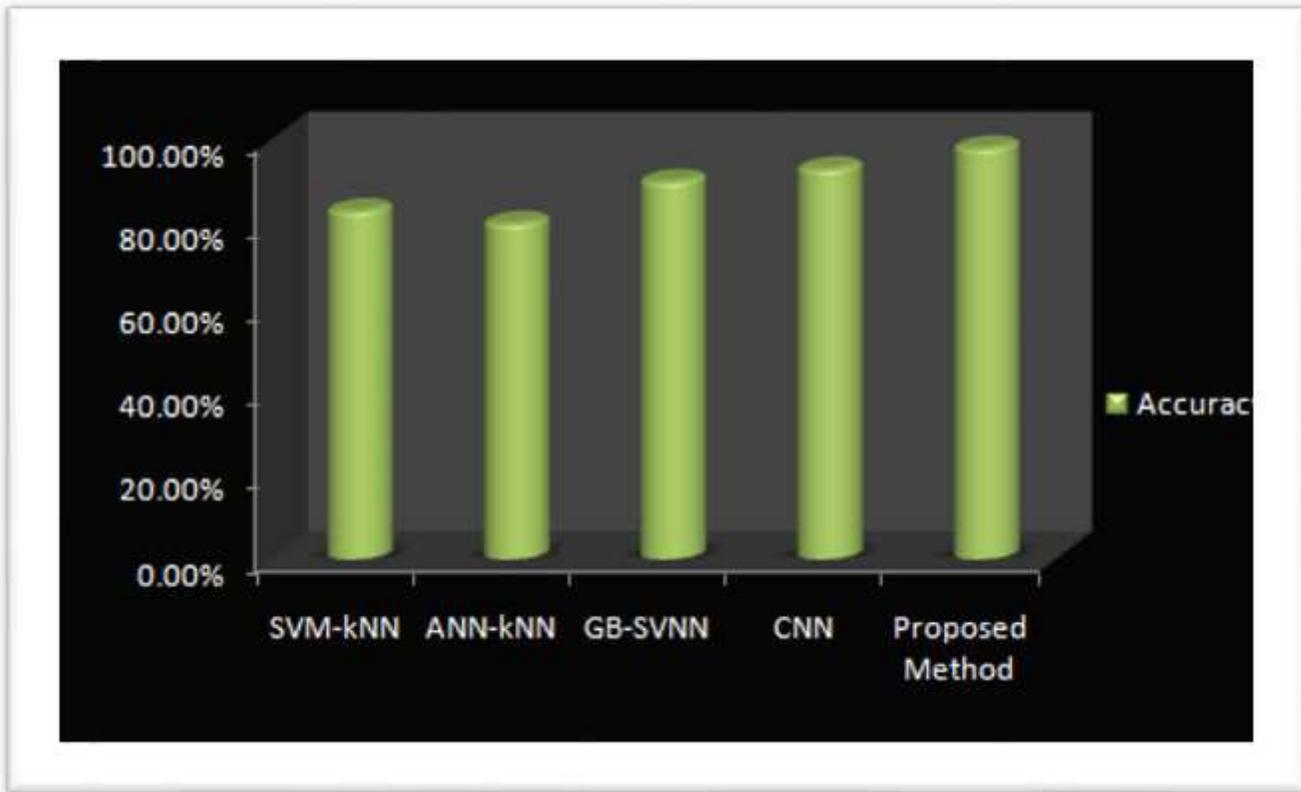


Figure 3

Comparison of accuracy

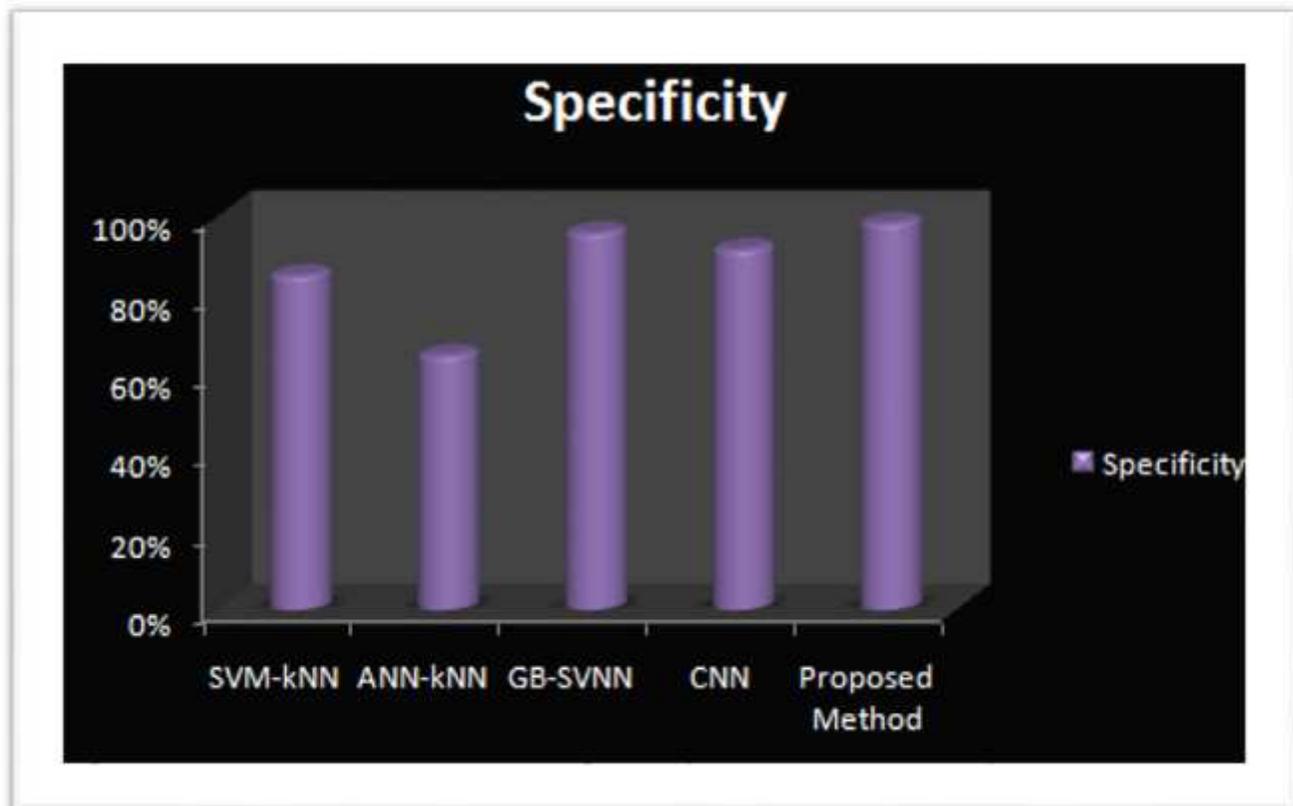


Figure 4

Comparison of Specificity

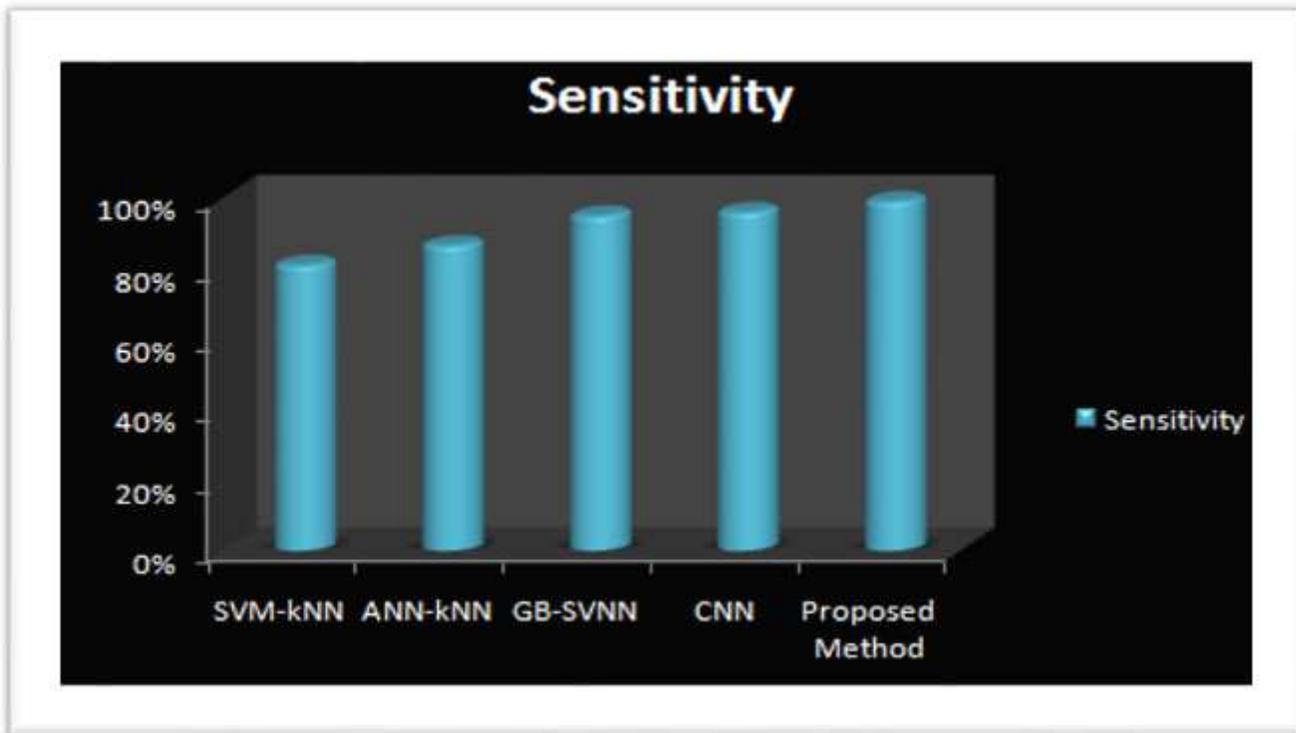


Figure 5

Comparison of sensitivity

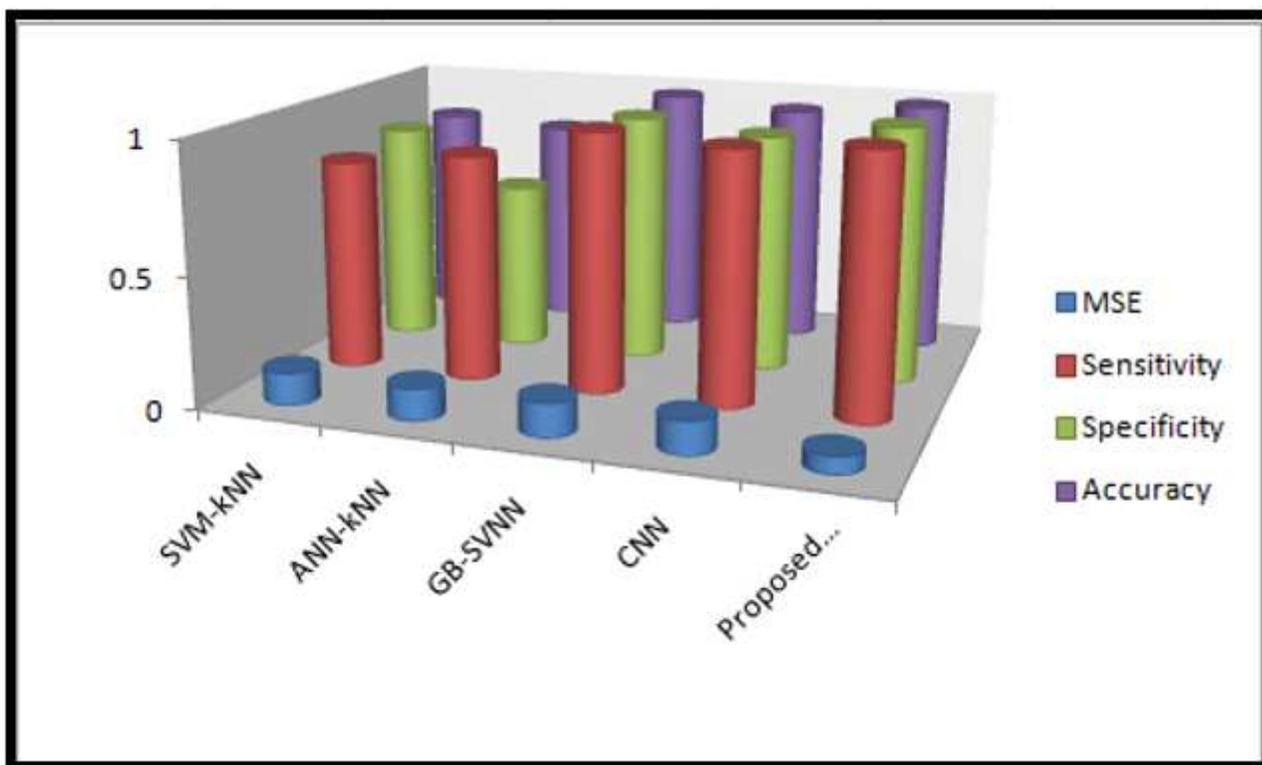


Figure 6

Overall output of the proposed and existing system