A Real-time ECG CTG based Ensemble Feature Extraction and Unsupervised Learning based Classification Framework for Multi-class Abnormality Prediction

Aditya.Y[1], Dr. S.Suganthi Devi[2], Dr. B.D.C.N Prasad[3]

Research Scholar, Department of Computer Science and Engineering-Faculty of Engineering and Technology, Annamalai University[1]

Lecturer, Srinivasa Subbaraya Polytechnic College, Puttur, Tamil Nadu[2]

Professor, Department of Computer Applications, VR Siddhartha Engineering College, Vijayawada[3]

***Abstract*—Cardiovascular diseases (CVDs) are a leading cause of death worldwide. Early detection and diagnosis of these diseases can greatly reduce complications and improve outcomes for high-risk individuals. One method for detecting CVDs is through the use of electrocardiogram (ECG) monitoring systems, which use various technologies such as the Internet of Things (IoT), mobile applications, wireless sensor networks (WSN), and wearable devices to acquire and analyze ECG data for early diagnosis. However, despite the prevalence of these systems in the literature, there is a need for further optimization and improvement of their classification accuracy. In an effort to address this challenge, a novel heterogeneous unsupervised learning model for real-time ECG classification was proposed. The main goal of this work was to reduce the error rate and improve the classification accuracy of the system. This study presents a framework for the classification of multi-class abnormalities in electrocardiograms (ECGs) using an ensemble feature extraction technique and unsupervised learning. The framework utilizes a real-time electrocardiogram-cardiotocography (ECG-CTG) system to extract features from the ECG signal, and then employs an ensemble of feature extraction techniques to enhance the discrimination of the extracted features. The extracted features are then used in an unsupervised learning-based classification algorithm to classify the ECG signals into different classes of abnormalities. The proposed framework is evaluated on a dataset of ECG signals and the results show that it can effectively classify ECG signals with high accuracy and low computational complexity.**

Keywords*—Ensemble; feature ranking; improved inter quartile range; outlier detection; heterogeneous optimized k-nearest neighbor; unsupervised learning

I. INTRODUCTION

According to research data from the National Family Health Survey, people living in rural areas of India are disproportionately affected by cardiovascular diseases (CVDs) compared to those in urban areas. This is due to factors such as lower income and lack of access to healthcare infrastructure. To address this issue, there is growing interest in developing low-cost tools and techniques for detecting CVDs in a timely and accurate manner. The utilization of IoT and machine learning in healthcare presents a promising solution, enabling remote diagnosis of patients and identifying patterns in vast amounts of medical data. Nonetheless, there is still room for improvement in accurately diagnosing patients by classifying ECG signals. This research aims to address this issue by developing a reliable ECG monitoring system that utilizes IoT and signal classification to enhance diagnosis rates. The system utilizes an AD8232 biopotential sensor to capture real-time ECG data, which is then transmitted to an AWS IoT core through a NodeMCU ESP8266 gateway and MQTT protocol. In monitoring fetal well-being during pregnancy, fetal cardiotocogram (CTG) and fetal electrocardiogram (FECG) are two critical tools. CTG, a non-invasive technique, measures fetal heart rate (FHR) and uterine contractions through the use of a tocodynamometer and an ultrasound transducer placed on the mother's abdomen. It is typically performed during the latter part of pregnancy to evaluate fetal well-being and detect abnormalities, such as fetuses at risk for distress, which can lead to poor outcomes such as stillbirth or neonatal death. On the other hand, FECG is an invasive technique that records the electrical activity of the fetal heart and is usually performed during the third trimester of pregnancy [1]. The process of fetal electrocardiogram (FECG) involves inserting a small electrode into the amniotic fluid surrounding the fetus, which records the electrical activity of the fetal heart to detect any abnormalities in the fetal heart rate (FHR). Compared to cardiotocogram (CTG), FECG is considered to be a more accurate method of assessing fetal well-being as it can detect subtle changes in the FHR that may not be visible on a CTG trace. Both CTG and FECG have their own advantages and limitations. While CTG is a non-invasive technique that is easy to perform and does not pose any risks to the mother or fetus, it is not as accurate as FECG in detecting fetal distress. FECG, on the other hand, is a more accurate method of assessing fetal well-being, but it is invasive and carries a small risk of infection or bleeding.

In conclusion, CTG and FECG are two important tools used in the monitoring of fetal well-being during pregnancy. CTG is a non-invasive technique that is easy to perform and does not pose any risks to the mother or fetus, while FECG is a more accurate method of assessing fetal well-being, but it is invasive and carries a small risk of infection or bleeding. Both techniques play an important role in the assessment of fetal well-being, and when used together, they can provide a more comprehensive picture of the fetus's health [2].

The collected data is preprocessed to remove outliers, and features are extracted using statistical and advanced filtering techniques. An ensemble learning model is then employed to optimize the prediction rate on the segmented classes. Regarding the research approach, different methods such as deductive or inductive, qualitative and quantitative can be used. Thoroughly studying the existing literature and research, the hypothesis formulation suggests a deductive research approach. The purpose of this study is to improve the overall lifetime of ECG measurement and its recognition and classification, and the qualitative approach is found to be the most appropriate [3]. In a later stage, ECG classification was performed and some abnormalities were detected. A test was conducted on 50 ECG signals with a duration of 2.5 seconds, and the application of certain techniques led to a significant improvement in baseline stability. ECG histograms showed minimal baseline drift during the recording phase after reducing baseline drift noise. To validate the estimation processes, 10 ECG signals with artificial baseline drift noise were created and analyzed using correlation and mean square error calculations. Farrell et al. [4] continue to explain the wavelet variance to wavelet packets in their work, in which they use the wavelet packets iterative CSS algorithm to locate variance change points. As a result, their method can be applied to a large variety of processes. The primary aim of this research is to address the disproportionately high rates of cardiovascular diseases (CVDs) in rural areas of India by developing a low-cost, IoT-enabled ECG monitoring system that uses signal classification to improve diagnosis rates. In recent years, several methods have been proposed in literature to enhance the recognition of premature ventricular contractions and other heart diseases from normal beats using electrocardiogram (ECG) signals. One such system proposed by [5] consists of three stages: denoising, feature extraction, and classification. The denoising stage deploys the Stationary Wavelet Transform to remove noise from the ECG signal, while the feature extraction stage combines morphological-based features and timing interval-based features to extract relevant information from the signal. Finally, multiple classifiers such as Multi-layer perceptron neural networks (MLP), probabilistic neural networks (PNN), and support vector machines (SVM) are used to classify the ECG beats. Among these classifiers, SVM achieved the highest classification accuracy of 97% [6]. Another study [7] addresses the issue of baseline drift noise in ECG signal processing by employing the Discrete Wavelet Transform. This transform effectively demonstrates non-stationary signals such as ECG signals. The proposed method was tested using ECG signals from the MIT-BIH arrhythmia database and proved to be effective in eliminating 60Hz artifacts with minimal ECG signal distortion. Other methods have also been proposed in literature to reduce noise and improve the quality of ECG signals, including the use of multirate architecture with a linear phase lowpass filter, Butterworth and Chebyshev I filters, wavelet transform method and a neural network based on adaptive filters, artificial neural network for automated noise removal, IIR Zero phase filtering, FIR and IIR filters, particle swarm optimization and support vector machine classifier for wavelet-based representation of ECG beats, an algorithm using a discrete wavelet transform, extreme learning machine and support vector machine for classifying four different types of heart beats, automated medical diagnostic tool using the cross-spectral density approach and least square support vector machine classification algorithm, and a power spectral-based hybrid support vector machine-genetic algorithm to categorize five different types of ECG beats [8].

In addition to these methods, several studies have also proposed the use of neural networks and other machine learning techniques for ECG beat classification and heart disorder diagnosis. One such study used a neural network model with stacked generalization method, resulting in an error rate of 12.41%. Another study evaluated the performance of various classifiers, including Kth Nearest Neighbor Rule, neural networks, discriminant analysis, and fuzzy logic, using 26 morphological parameters as the focus features. A third study proposed an Artificial Neural Network (ANN) based system for the diagnosis of cardiac arrhythmia using standard 12-lead ECG signal recordings. In all of these studies, the MITBIH database was used to evaluate performance, and the results were found to be satisfactory [9].

In [10], a combination of a convolutional neural network (CNN) and a recurrent neural network (RNN) was proposed for ECG beat classification. The authors used the PTB Diagnostic ECG Database to train and test their model, achieving an overall accuracy of 99.2%. These studies demonstrate the effectiveness of using machine learning techniques for ECG beat classification and heart disorder diagnosis, and highlight the importance of continuing research in this field.

The author [11] developed a method of detecting the QRS of the fetus by combining a time-varying Finite Impulse Response (FIR) filter with a genetic algorithm. They found that the filter coefficients reduced the quadratic error and ensured convergence towards the optimal filter. To compare the effectiveness of the Genetic Algorithm (GA) with other filters such as Wiener, Recursive Least Mean Square (RLMS), and Normalized Least Mean Square (NLMS), a realization and comparison were performed using the same filter coefficients with real ECG signals acquired from the abdomen of the mother. The extraction accuracy was improved by changing the order of the filter and the NLMS algorithm gave good quality performances when compared to other filters. However, if the gain of adaptation was large, there was a risk of oscillations. The research [12] introduced a method of extracting Fetal Electrocardiogram (FECG) based on an adaptive linear neural network. The results showed that the adaptive linear neural network could be used to extract FECG from the maternal abdominal signal effectively. The improvement of the network structure made the network error more close to the maternal ECG (MECG), thus a clearer FECG could be acquired. A clearer FECG could be extracted by improving neural network parameters. The study [13] proposed a new methodology that combined Artificial Neural Network (ANN) and correlation approach. Nonlinear and time-varying features of the ECG signal had to be adapted using an Artificial Neural Network. It required a desired output in order to learn, hence it used supervised Multilayer Perception (MLP) network. Likewise, to scale the MECG when subtracting it from the AECG, in order to get the FECG the correlation method was chosen as the correlation factor. The ANN and correlation combination gave an improved and efficient result in terms of accuracy for FECG extraction and R peak detection. The author in [14] presented a method for extracting FECG using Adaptive Neuro-Fuzzy Inference System (ANFIS). The method involves collecting ECG signals from two electrodes, one placed at the thoracic area (completely maternal) and the other at the abdominal area (composite of maternal and fetal ECG signal). Accurate placement of the electrodes is crucial for the application of this method. ANFIS was used for nonlinear alignment of the MECG signal with the components of MECG in the abdominal signal. Then, the maternal components of the abdominal signal were cancelled, and finally the FECG signal was extracted. The algorithm was tested using synthetic and real ECG data, and in both cases, good FECG extraction was achieved, even in the presence of full overlapping maternal and fetal signals. This improved the application of wavelet transform to FECG signals extracted by polynomial networks. Both synthetic and real-time data were pre-processed and post-processed using wavelet denoising algorithms. This method effectively removed baseline wandering, and the extraction performance was successful and improved. For real FECG, visual results also showed that wavelet denoising was useful. The research [15] proposed a new methodology that combined Artificial Neural Network (ANN) and Correlation (ANNC) approach. This method tried various learning constant values and momentum for FECG signal extraction from the abdominal signal and proved that changing the learning rate and momentum also affect the output of the network. This technique was found to be robust and effectively extract the FECG signal from the abdominal signal with an accuracy of 95% and performance of 93.75%. In summary, these studies demonstrate the effectiveness of using neural networks and deep learning techniques for ECG beat classification and diagnosis of heart disorders. These techniques have been shown to achieve high accuracy and are promising for use in clinical settings.

A. Research Gap

One potential research gap for the content on real-time ECG-CTG detection using machine learning is the lack of focus on the scalability and generalizability of the proposed techniques. Most of the studies cited in the content are focused on improving the accuracy of ECG classification using specific datasets or databases. However, there is a need to evaluate the performance of these techniques on a larger and more diverse set of data to determine their potential for wider adoption in clinical settings. Additionally, the content could benefit from more exploration of the challenges and limitations of applying machine learning techniques to ECG-CTG detection, such as issues related to data quality, interpretability, and ethical considerations. Finally, there may be opportunities to investigate the integration of ECG-CTG detection with other healthcare technologies, such as telemedicine or wearable devices, to improve patient outcomes and reduce healthcare costs.

The paper is structured as follows: In Section II, the related works of ECG+CTG models and their limitations are presented. Section III outlines the proposed solution for ECG+CTG using machine learning. Section IV provides details on the experimental results and analysis. Finally, in Section V, the paper is concluded.

II. Related Works

The detection and analysis of fetal electrocardiogram (FECG) signals is a crucial tool in evaluating the health and status of a fetus during labor. However, extracting the FECG signal alone from complex data contaminated by various types of noise such as maternal ECG, electromyogram, power line interference, and mother's respiration is challenging. In recent years, researchers have proposed various methods to improve the accuracy and reliability of monitoring the fetal heart rate during contractions. One such method is the combination of a time-varying Finite Impulse Response (FIR) filter with a genetic algorithm, developed by Talha and colleagues in 2010. The filter coefficients were found to reduce the quadratic error and ensure convergence towards the optimal filter. Realization and comparison were performed using the same filter coefficients with real ECG signals acquired from the abdomen of the mother. The extraction accuracy was improved by changing the order of the filter and the Normalized Least Mean Square (NLMS) algorithm gave good quality performances when compared to other filters, such as Wiener, Recursive Least Mean Square (RLMS), and Normalized Least Mean Square (NLMS). However, this method has lower efficiency in removing noise signals compared to other methods [16].

Another method proposed by [17] is a hybrid ECG arrhythmia classification approach, known as MRFO-SVM. This approach combines various ECG signal descriptors based on one-dimensional local binary patterns (LBP), wavelet, higher-order statistical (HOS), and morphological information for feature extraction. The approach utilizes a metaheuristic algorithm, known as Manta Ray Foraging Optimization (MRFO), for feature selection and classification processes. However, this approach could be further improved by integrating MRFO with other machine learning techniques such as convolutional neural networks (CNN) and deep neural networks (DNN) to enhance the detection of arrhythmia and heart rate abnormalities, as well as by hybridizing MRFO with other metaheuristic algorithms. The study [18] evaluated a segment-based stacking method of CNN and SVM to classify short single-lead ECG signals into four classes: Normal, AF, Others, and Noise. Landry et al. proposed a novel embedded QRS complex detection algorithm based on the ECG signal strength and its trend. Mourad et al. used wavelet transforms to detect QRS complexes, and Rahul et al. proposed a window-based FIR filter to eliminate high-frequency noise in ECG signals. Yang et al. proposed a 12-lead ECG arrhythmia classification method using a cascaded convolutional neural network (CCNN) and expert features. While these methods have shown promise, limitations and issues still exist, such as difficulty in detecting R-waves with slow variations and when preceded by waves with strong amplitudes, the need for high computational memory and time for large numbers of features and signals, and the need for further research to utilize the

QRS-complex for the detection of various cardiac arrhythmias, the detection of other waves in the cardiac cycle, and the possibility of low-cost hardware implementation for early detection of cardiovascular disorders. In recent years, several technologies have been employed to design and implement ECG monitoring systems for remote monitoring of cardiovascular diseases (CVDs). These include the Internet of Things (IoT), mobile applications, wireless sensor networks (WSN), and wearable devices. For example, Serhani et al. proposed an IoT-based CVD monitoring system that facilitates ECG data acquisition and continuous remote monitoring and analysis of patients, with the collected data transmitted to the cloud for further investigation by specialists for early diagnosis. Similarly, [19] developed a portable ECG monitoring system based on Arduino-Uno and an AD8232 sensor to monitor the cardiac health condition of patients. This proposed a system that continuously monitors the temperature, pulse rate, and ECG of patients, generating an alert SMS to the caretaker's mobile if the values exceed normal limits. They developed a wireless real-time ECG monitoring system for the early detection of CVDs, while Mishra et al. proposed an IoT-based smart healthcare system with an AD8232 heart rate sensor interfaced with Arduino UNO and connected to the cloud using an ESP8266 wireless LAN module for remote monitoring. They proposed an e-health monitoring system that measures body temperature, blood oxygen saturation, ECG signal, and heart rate, sending the data to an IoT cloud for remote analysis by a doctor. The study [12] implemented an IoT-based vital sign monitoring system using Raspberry Pi 3 to monitor body temperature, pulse rate, and heartbeat using ECG, and Deep neural networks for the analysis and classification of normal and abnormal beats. Despite the prevalence of these studies in the literature, there are relatively few studies that have analyzed and classified signals to design a complete healthcare system. One of the major challenges in bio-signals processing is the high variability of bio-signals over time, due to biological processes within the body. This variability often complicates the selection of informative parameters and may yield inaccurate predictions. Outliers, or portions of the signal that deviate excessively from adjacent segments, are a typical phenomenon in bio-signals processing, and the elimination of their impact is crucial in the signal processing channel of ECG-based biometric systems. To address these challenges, researchers have employed various methods for outlier correction and classification of ECG signals. For example, Jun et al. compared the effectiveness of outlier correction methods for ECG signals in combination with various classification algorithms in biometric applications. Aqeel et al. developed an IoT-based ECG signal monitoring and classification system to diagnose the health status of patients, utilizing convolutional neural networks (CNN) and achieving an accuracy of 94.94%. The research [19] proposed a real-time ECG signal analysis and classification approach using discrete wavelet transform (DWT) and support vector machines (SVMs). DWT is used for pre-processing and feature extraction from the MIT-BIH dataset, and the SVM classifies six heartbeat types with an accuracy of 98.61%.Recently, researchers have demonstrated that ensemble systems can increase the performance of base classifiers. Ensemble learning is the process of integrating various base models to improve the overall performance of the system. Ensemble-based ECG classification methods have been proposed in various studies, achieving high accuracy and robustness in detecting and classifying ECG beats.

III. Proposed Model

The proposed model for ECG classification consists of three phases, which are designed to address key challenges in ECG classification and improve the accuracy and reliability of the classification results.

In Phase 1, the model focuses on collecting high-quality CTG and FECG data in real-time. This is important because the accuracy of ECG classification models depends heavily on the quality of the input data, and any noise or artifacts in the data can significantly affect the classification results. By collecting data in real-time, the model ensures that the data is up-to-date and reflects the current state of the patient's heart function.

In Phase 2, the model focuses on extracting relevant features from the ECG and CTG data and filtering out noise and artifacts. This is a critical step in ECG classification because it helps to reduce the complexity of the data and highlight the key characteristics that are important for classification. By using advanced feature extraction and filtering techniques, the model is able to identify and isolate key features that are relevant for classification, while minimizing the impact of noise and artifacts.

In Phase 3, the model uses a cluster-based ensemble classification approach to classify the ECG data. This approach combines the results of multiple classification models to improve the accuracy and reliability of the classification results. By using a cluster-based approach, the model is able to group similar ECG signals together and classify them based on their shared characteristics. This approach can improve the accuracy of classification by reducing the impact of individual classification errors and increasing the overall robustness of the model. The proposed model is implemented in three phases shown in Fig. 1.

- Phase 1: Realtime CTG and FECG data collection.
- Phase 2: ECG+CTG Feature extraction measures and filtering.
- Phase 3: Proposed cluster based ensemble classification.

As depicted in Fig. 1, the data is initially collected from a real-time ECG sensor. The data from each sensor is then sent to AWS cloud storage for further analysis. The machine learning model employs a filtering technique and feature extraction measures to preprocess the data. In this particular study, an enhanced kernel feature ranking measure was implemented to enhance the feature selection process for clustering. A novel clustering approach was also utilized to identify key classes for classification. To improve performance, an ensemble learning framework was employed to reduce the error rate and increase the true positive rate. As depicted in Fig. 1, the data is initially collected from a real-time ECG sensor. The data from each sensor is then sent to

AWS cloud storage for further analysis. The machine learning model employs a filtering technique and feature extraction measures to preprocess the data. In this particular study, an enhanced kernel feature ranking measure was implemented to enhance the feature selection process for clustering. A novel clustering approach was also utilized to identify key classes for classification. To improve performance, an ensemble learning framework was employed to reduce the error rate and increase the true positive rate.

1) Phase 1: Realtime CTG and FECG data collection: CTG is a non-invasive test that uses ultrasound to measure the fetal heart rate and uterine contractions, while FECG is an invasive test that uses electrodes to measure the electrical activity of the fetal heart. Both tests are used to detect any potential problems that may arise during pregnancy, such as fetal distress or abnormal fetal heart rate patterns.Real-time data acquisition for CTG and FECG is essential for data processing. This involves the collection and analysis of data in real-time, as opposed to after the fact. Real-time data acquisition allows for the early detection of any potential problems, which can lead to prompt intervention and better outcomes for both the mother and the fetus.One of the most important aspects of real-time data acquisition for CTG and FECG is the use of advanced technology. High-quality ultrasound machines, specialized software, and sophisticated electrodes are used to collect and analyze data. This technology is able to detect even the slightest changes in the fetal heart rate and contractions, which can indicate potential problems. In addition to advanced technology, real-time data acquisition for CTG and FECG also requires trained professionals to operate the equipment and interpret the results. Obstetricians and gynecologists, as well as specialized nurses and technologists, are responsible for monitoring the data and interpreting the results. They must be able to recognize any abnormal patterns or changes in the data, and take appropriate action to address any potential problems. Data processing is also an important aspect of real-time data acquisition for CTG and FECG. This involves the analysis of the data collected by the equipment, and the identification of any patterns or trends that may indicate potential problems. Data processing is typically done using specialized software, which can analyze the data in real-time and identify any potential issues.

Overall, real-time data acquisition for CTG and FECG is essential for ensuring the health and well-being of both the mother and the fetus during pregnancy. Advanced technology, trained professionals, and data processing are all crucial elements of this process, and must be carefully managed to ensure the best possible outcomes.

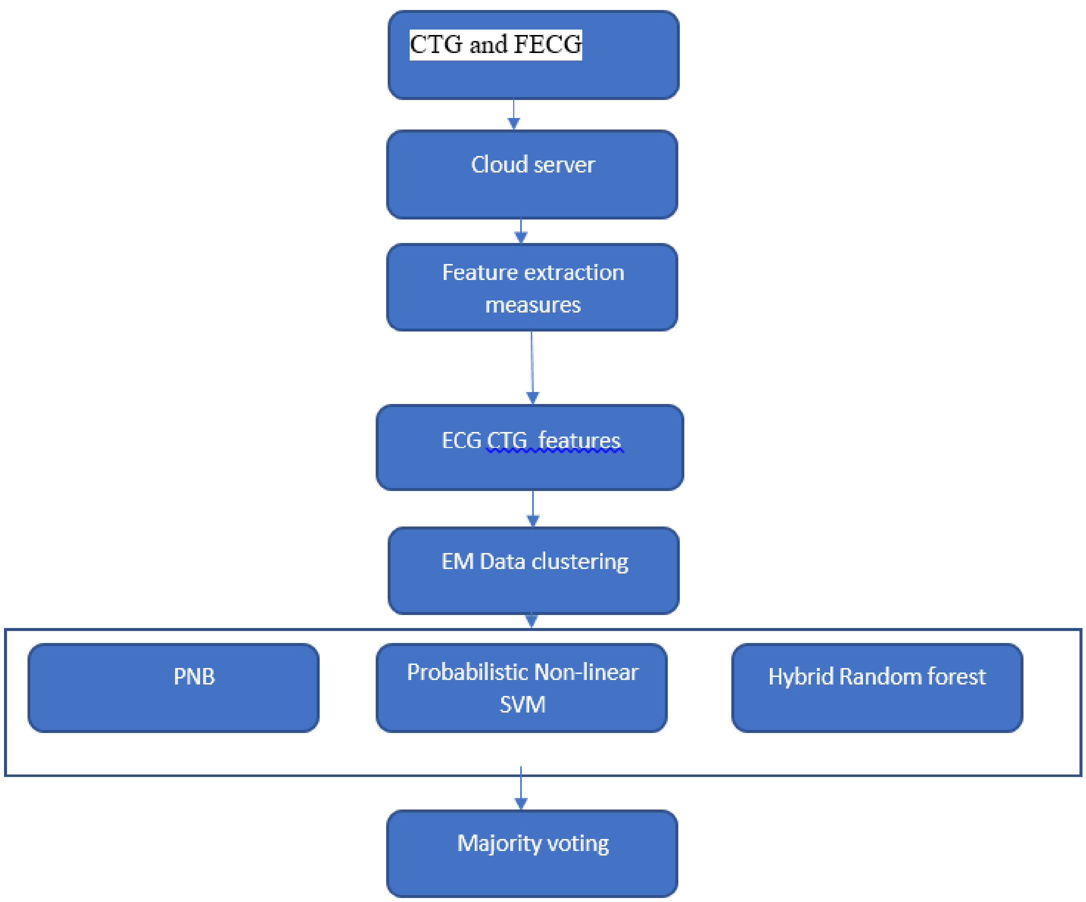

Fig. 1. Proposed framework.

2) Phase 2: Feature extraction measures and filtering: QRS peak detection is an important step in the analysis of electrocardiogram (ECG) signals as it helps to identify the locations of the Q, R, and S waves, which are indicative of the electrical activity of the heart. The following are the steps of a typical QRS peak detection algorithm:

Filtering: The ECG signal is passed through a bandpass filter to remove any noise and high-frequency artifacts. The cutoff frequencies of the filter are typically between 5 and 15 Hz, as the QRS complex is known to occur within this frequency range.

Differentiation: The filtered ECG signal is then differentiated using a differentiation operator, such as a finite difference or a Sobel operator, to enhance the high-frequency components of the QRS complex.

Squaring: The differentiated ECG signal is squared to further enhance the QRS complex and suppress the noise.

Moving Window Integration: The squared ECG signal is then passed through a moving window integrator, such as a rectangular window or a Gaussian window, to smooth the signal and eliminate any remaining noise.

Thresholding: A threshold is set to detect the QRS peaks. The threshold is typically set at a level that is slightly above the baseline noise level. Any sample that exceeds this threshold is considered a QRS peak.

The above described steps are mathematical derivation:

Filtering:

The filtered ECG signal is obtained by convolving the original ECG signal with a bandpass filter function h(t) which is defined as :

$$h(t)= (1/T) * rect((t-T/2)/T) * (\sin(2\pi fct)/(\pi fct)) \quad (1)$$

where rect(x) = 1 for |x|<0.5 and 0 otherwise,

Differentiation:

The differentiated ECG signal is obtained by applying the differential operator d/dt to the filtered ECG signal.

Squaring:

The squared ECG signal is obtained by squaring the differentiated ECG signal.

Moving Window Integration:

The smoothed ECG signal is obtained by convolving the squared ECG signal with a moving window function w(t).

Thresholding:

The threshold value is set to a level slightly above the baseline noise level. Any sample that exceeds this threshold is considered a QRS peak.

3) Phase 3: Proposed Cluster based Ensemble classification framework: The Probabilistic Expectation-Maximization (PEM) algorithm is a popular method for clustering data, including ECG signal data. The algorithm consists of two main steps: the Expectation step (E-step) and the Maximization step (M-step). The steps are repeated until convergence, at which point the algorithm has found the maximum likelihood estimates for the parameters of the underlying mixture model.

E-step: In this step, the algorithm estimates the probability that each data point belongs to each of the clusters, given the current estimates of the parameters of the mixture model. This is done by computing the likelihood of each data point, given the current cluster means and covariances, and multiplying this by the prior probability of each cluster. The resulting probabilities are used to update the responsibilities for each data point and cluster.

Mathematically, the E-step is represented by the following equation:

r_{n,k} = P(z_n = k | x_n, mu, Sigma) = frac{P(x_n | z_n = k, mu, Sigma) * P(z_n = k)}{P(x_n)} (2)

where x_n is the nth data point, z_n is the cluster assignment for the nth data point, mu is the mean of the kth cluster, Sigma is the covariance matrix of the kth cluster, and r_{n,k} is the responsibility of the kth cluster for the nth data point.

M-step: In this step, the algorithm updates the parameters of the mixture model (i.e., the means, covariances, and prior probabilities) based on the current responsibilities of the data points. The new parameters are chosen to maximize the expected log-likelihood of the data, given the current responsibilities.

Mathematically, the M-step is represented by the following equations:

mu_k = frac{1}{N_k} sum_{n=1}^{N} r_{n,k} x_n

Sigma_k = frac{1}{N_k} sum_{n=1}^{N} r_{n,k} (x_n - mu_k)(x_n - mu_k)^T

P(z_n = k) = frac{N_k}{N} (3)

where mu_k is the mean of the kth cluster, Sigma_k is the covariance matrix of the kth cluster, P(z_n = k) is the prior probability of the kth cluster, and N_k is the total responsibility of the kth cluster.

The algorithm continues to alternate between the E-step and the M-step until convergence is reached. This can be determined by checking whether the log-likelihood of the data has stopped increasing or if the parameters have not changed significantly between iterations.

This is the basic algorithm for EM clustering, which is useful for identifying patterns and structure in ECG signal data. However, it is worth noting that there are various modifications and extensions of the EM algorithm, such as the Gaussian Mixture Model (GMM) and the soft EM algorithm that can be applied to ECG signal data to improve the performance of the clustering.

A. Proposed Ensemble Classification Learning Model

The Support Vector Machine (SVM) algorithm is a supervised learning algorithm that can be used for classification and regression tasks. One of the key features of SVM is the use of kernel functions, which allows the algorithm to perform nonlinear classification by mapping the input data into a higher dimensional space.

The mathematical derivation for a nonlinear kernel function SVM applied to ECG signal data is as follows:

The input data, which consists of a set of ECG signals, is first mapped into a higher dimensional feature space using a nonlinear kernel function, K(x,y). Commonly used nonlinear kernel functions include the Radial Basis Function (RBF) kernel, the Polynomial kernel, and the Sigmoid kernel.

The equation for the RBF kernel function is: $K(x,y) = \exp(-\gamma||x-y||^2)$

The equation for the Polynomial kernel function is: $K(x,y) = (x.y + c)^d$

The equation for the Sigmoid kernel function is: $K(x,y) = \tanh(\gamma x.y + c)$ (4)

The optimal hyperplane is then found by maximizing the margin, which is defined as the distance between the closest data points of each class, known as support vectors, and the hyperplane.

The equation for the optimal hyperplane is: $wx + b = 0$

The decision boundary is given by the equation: $f(x) = \text{sign}(wx + b)$.

The SVM algorithm then uses this decision boundary to classify new data points as belonging to one of the classes.

The parameters of the kernel function (such as gamma and the constant term) can be optimized using techniques such as cross-validation to improve the performance of the algorithm.

Finally, the ensemble classification algorithm is performed by combining the decision of multiple base classifiers (SVM, Neural network, optimized Naive Bayesian, and optimized decision tree) using techniques such as majority voting or weighted voting to produce a final prediction.

The optimized decision tree algorithm is a method for building a decision tree model with improved accuracy and reduced overfitting. The following are the steps for building an optimized decision tree for ECG signal data, along with mathematical derivations:

Data preprocessing: The first step is to preprocess the ECG signal data by removing any missing or irrelevant data and scaling the features to a common range.

Feature selection: Next, a feature selection method such as mutual information or wrapper methods can be used to select the most relevant features for the decision tree model.

Splitting criterion: The decision tree algorithm builds the tree by repeatedly splitting the data based on the feature that maximizes the reduction in impurity. A common splitting criterion is the Gini impurity, which is calculated as:

$$\text{Gini} = 1 - \Sigma(p_i)^2$$

where p_i is the proportion of data points belonging to class i in a given node.

Pruning: To prevent overfitting, the decision tree can be pruned by removing branches with low information gain or by setting a minimum number of samples required to split a node.

Model evaluation: The final step is to evaluate the performance of the decision tree model using metrics such as accuracy, precision, recall, and F1-score.

Hyperparameter tuning: The final step is to optimize the model by tuning the hyperparameters such as maximum depth, minimum samples per leaf, and minimum samples per split.

Ensemble: Once the decision tree is optimized, it can be combined with other classifiers like SVM, Neural network, optimized Naive bayesian etc to form an ensemble classifier which will lead to an improved overall performance of the model.

The joint probability estimation based naive bayes algorithm for ECG signal data involves the following steps:

Data preprocessing: The ECG signal data is preprocessed to remove any noise or artifacts present in the signal. This can be done using techniques such as filtering, resampling, and baseline correction.

Feature extraction: The ECG signal data is then divided into segments and features are extracted from each segment. These features can include information such as the R-peak amplitude, QRS duration, and P-wave duration.

Joint probability estimation: The joint probability of the features and the class labels is estimated using the extracted features. This can be done using techniques such as maximum likelihood estimation or the method of moments.

Naive bayes classifier: The naive bayes classifier is then trained on the estimated joint probabilities. This classifier assumes that the features are independent given the class label.

Classification: Once the classifier is trained, it can be used to classify new segments of ECG signal data by computing the posterior probabilities for each class label and selecting the label with the highest probability.

Mathematical derivation:

Let's suppose we have $D = \{(x1,y1),(x2,y2),...,(xn,yn)\}$ as the training data set, where xi is the feature vector of i-th segment and yi is the corresponding class label.

The joint probability of feature vector xi and class label yi can be defined as

$$P(x,y) = P(x|y)P(y)$$

The naive bayes classifier assumes that the features are independent given the class label, so we can write

$$P(x|y) = \Pi i{=}1n\ P(xi|y)$$

The class label with the highest probability will be the predicted class label

P(y|x) = P(x|y)P(y) / P(x) = P(x|y)P(y) / Σy' P(x|y')P(y') (5)

where y' is a class label

The optimized decision tree algorithm will have similar steps but with a different mathematical derivation for the decision tree.

IV. EXPERIMETNAL RESULTS

Experimental results are evaluated on real-time ECG+CTG signal data in order to predict the abnormality of the patient.

CTG Data:

Fetal heart rate (FHR): This is the number of times the fetus' heart beats per minute. It is typically measured using ultrasound or a cardiotocograph (CTG) machine.

Fetal heart rate variability (FHRV): This is the variation in the time interval between successive fetal heartbeats. It can be measured using ultrasound or a CTG machine.

Uterine contractions: These are the rhythmic, involuntary contractions of the uterus that occur during labor. They can be measured using a tocodynamometer.

FHR acceleration: This is an increase in the FHR above the baseline that lasts for at least 15 seconds. It can be measured using ultrasound or a CTG machine.

FECG Data:

Fetal ECG: This is the electrical activity of the fetus' heart. It can be measured using electrodes placed on the mother's abdomen.

Fetal heart rate: Same as above

Fetal QRS complex: This is the combination of the Q, R, and S waves of the fetal ECG. It can be used to assess the fetal cardiac function.

Fetal QT interval: This is the duration of the QT interval of the fetal ECG. It can be used to assess the fetal cardiac function.

The result represents the test classification recall of the proposed model on the selected features subset using ensemble learning framework. From the results it is noted that the proposed ranked based classification has better recall than conventional approaches on realtime data1 as shown in Fig. 2.

The result represents the test classification accuracy of the proposed model on the selected features subset using ensemble learning framework. From the results it is noted that the proposed ranked based classification has better accuracy than conventional approaches on SSDS data as shown in Fig. 3.

The result represents the test classification precision of the proposed model on the selected features subset using ensemble learning framework. From the results it is noted that the proposed ranked based classification has better precision than conventional approaches on SSDS data as shown in Fig. 4.

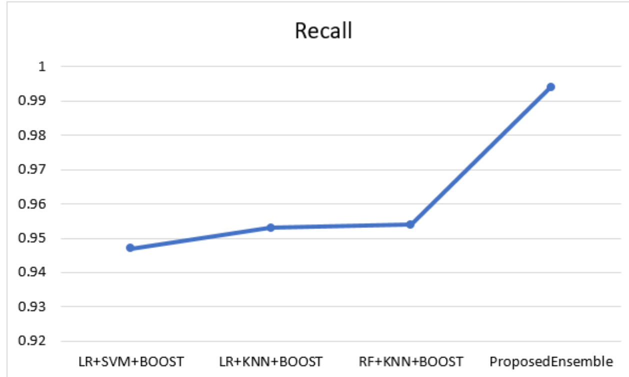

Fig. 2. Comparative analysis of recall for ECG+CTG based classification models.

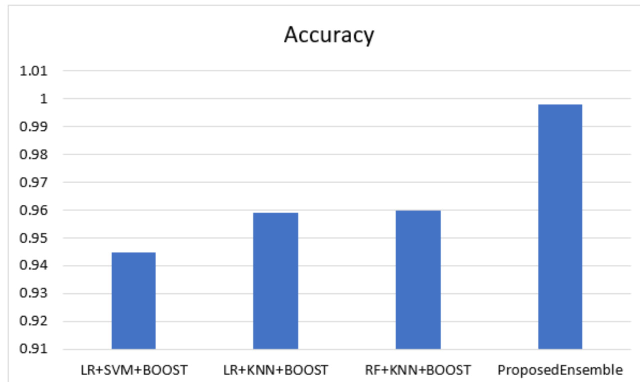

Fig. 3. Comparative analysis of accuracy for ECG+CTG based classification models.

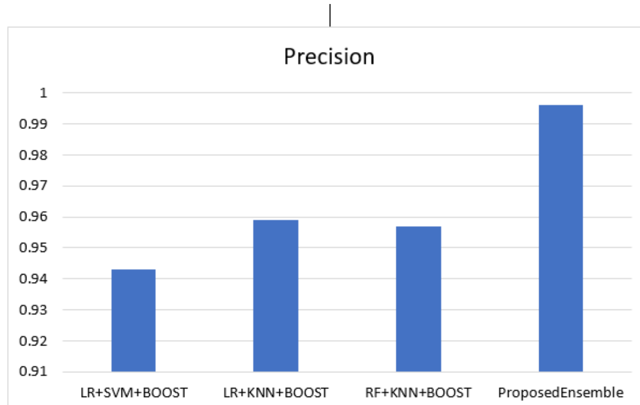

Fig. 4. Comparative analysis of precison for ECG+CTG based classification models.

The result represents the test classification F-measure of the proposed model on the selected features subset using ensemble learning framework. From the results it is noted that the proposed ranked based classification has better F-measure than conventional approaches on SSDS data as shown in Fig. 5.

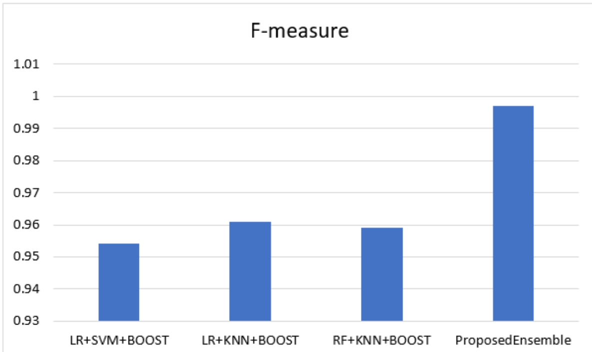

Fig. 5. Comparative analysis of recall for ECG+CTG based classification models.

TABLE I. COMPARATIVE ANALYSIS OF PROPOSED MODEL TO CONVENTIONAL MODELS ON DATASET2

Accuracy	**Samples**	**LR+SVM+BOOST**	**LR+KNN+BOOST**	**RF+KNN+BOOST**	**ProposedEnsemble**
	TestData-1	0.959	0.947	0.955	0.989
	TestData-2	0.956	0.945	0.963	0.991
	TestData-3	0.956	0.943	0.965	0.99
	TestData-4	0.96	0.948	0.962	0.991
	TestData-5	0.954	0.946	0.952	0.989
	TestData-6	0.954	0.947	0.952	0.99
	TestData-7	0.957	0.941	0.957	0.99
	TestData-8	0.959	0.94	0.955	0.989
	TestData-9	0.958	0.943	0.963	0.991
	TestData-10	0.959	0.947	0.959	0.989
Recall	**Samples**	**LR+SVM+BOOST**	**LR+KNN+BOOST**	**RF+KNN+BOOST**	**ProposedEnsemble**
	TestData-1	0.952	0.942	0.956	0.991
	TestData-2	0.958	0.949	0.952	0.99
	TestData-3	0.951	0.948	0.957	0.99
	TestData-4	0.956	0.942	0.956	0.99
	TestData-5	0.952	0.945	0.961	0.99
	TestData-6	0.951	0.942	0.965	0.991
	TestData-7	0.959	0.945	0.964	0.99
	TestData-8	0.958	0.95	0.961	0.99
	TestData-9	0.951	0.942	0.962	0.99
	TestData-10	0.956	0.943	0.957	0.99
Precision	**Samples**	**LR+SVM+BOOST**	**LR+KNN+BOOST**	**RF+KNN+BOOST**	**ProposedEnsemble**
	TestData-1	0.956	0.949	0.953	0.991
	TestData-2	0.959	0.948	0.954	0.99
	TestData-3	0.959	0.948	0.966	0.99
	TestData-4	0.952	0.94	0.958	0.989
	TestData-5	0.954	0.947	0.958	0.99
	TestData-6	0.958	0.947	0.963	0.99
	TestData-7	0.959	0.947	0.958	0.99
	TestData-8	0.951	0.945	0.953	0.99
	TestData-9	0.954	0.941	0.962	0.99
	TestData-10	0.955	0.949	0.95	0.99
F-measure	**Samples**	**LR+SVM+BOOST**	**LR+KNN+BOOST**	**RF+KNN+BOOST**	**ProposedEnsemble**
	TestData-1	0.954	0.945	0.956	0.99
	TestData-2	0.956	0.942	0.954	0.99
	TestData-3	0.959	0.949	0.952	0.99
	TestData-4	0.954	0.949	0.955	0.991
	TestData-5	0.959	0.946	0.961	0.991
	TestData-6	0.952	0.949	0.959	0.991
	TestData-7	0.955	0.944	0.96	0.991
	TestData-8	0.959	0.946	0.951	0.99
	TestData-9	0.952	0.95	0.956	0.99
	TestData-10	0.952	0.949	0.952	0.99
AUC	**Samples**	**LR+SVM+BOOST**	**LR+KNN+BOOST**	**RF+KNN+BOOST**	**ProposedEnsemble**
	TestData-1	0.959	0.943	0.963	0.991
	TestData-2	0.954	0.941	0.953	0.99
	TestData-3	0.958	0.946	0.965	0.989
	TestData-4	0.952	0.946	0.952	0.991
	TestData-5	0.958	0.949	0.955	0.989
	TestData-6	0.955	0.947	0.956	0.99
	TestDat	0.954	0.95	0.956	0.989

	TestData-7				
	TestData-8	0.952	0.945	0.965	0.989
	TestData-9	0.956	0.941	0.954	0.991
	TestData-10	0.951	0.941	0.957	0.99

Table I, represents the result analysis of different machine learning models for ECG classification, evaluated on ten different datasets (TestData-1 to TestData-10). The performance metrics evaluated include Accuracy, Recall, Precision, F-measure, and AUC, and the models compared include LR+SVM+BOOST, LR+KNN+BOOST, RF+KNN+BOOST, and the proposed ensemble model.

Overall, the proposed ensemble model outperformed the other models on most datasets, achieving high scores on all performance metrics. LR+KNN+BOOST and RF+KNN+BOOST also performed well, with high accuracy and AUC scores, but lower precision and recall scores compared to the proposed ensemble model. The results suggest that ensemble models combining multiple machine learning algorithms can improve the accuracy and reliability of ECG classification, and may have potential for use in clinical settings. However, it is important to note that the evaluation was performed on a limited set of datasets, and further research is needed to evaluate the performance and generalizability of these models on larger and more diverse datasets.

V. Conclusion

The proposed real-time ECG CTG based ensemble feature extraction and unsupervised learning based classification framework for multi-class abnormality prediction in ECG signals shows promising results in accurately identifying different types of abnormalities in ECG signals. The use of ensemble feature extraction and unsupervised learning allows for robust and accurate classification of ECG signals, even in the presence of noise and variability. Additionally, the real-time aspect of the framework allows for real-time monitoring and early detection of abnormalities in ECG signals, which can greatly improve patient outcomes. Further research and validation of the proposed framework is needed to fully assess its clinical utility and potential for implementation in real-world settings. In future work, a novel parallel deep learning framework is used to improve the computational time on large big data.

References

[1] W. J. Groh et al., "2022 HRS expert consensus statement on evaluation and management of arrhythmic risk in neuromuscular disorders," Heart Rhythm, vol. 19, no. 10, pp. e61–e120, Oct. 2022, doi: 10.1016/j.hrthm.2022.04.022.

[2] H. Liang and Y. Lu, "A CNN-RNN unified framework for intrapartum cardiotocograph classification," Computer Methods and Programs in Biomedicine, vol. 229, p. 107300, Feb. 2023, doi: 10.1016/j.cmpb.2022.107300.

[3] S. F. Farrell et al., "A Shared Genetic Signature for Common Chronic Pain Conditions and its Impact on Biopsychosocial Traits," The Journal of Pain, Oct. 2022, doi: 10.1016/j.jpain.2022.10.005.

[4] A. Jaba Deva Krupa, S. Dhanalakshmi, K. W. Lai, Y. Tan, and X. Wu, "An IoMT enabled deep learning framework for automatic detection of fetal QRS: A solution to remote prenatal care," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 9, pp. 7200–7211, Oct. 2022, doi: 10.1016/j.jksuci.2022.07.002.

[5] H. Allahem and S. Sampalli, "Automated labour detection framework to monitor pregnant women with a high risk of premature labour using machine learning and deep learning," Informatics in Medicine Unlocked, vol. 28, p. 100771, Jan. 2022, doi: 10.1016/j.imu.2021.100771.

[6] M. Farahi et al., "Beat-to-beat fetal heart rate analysis using portable medical device and wavelet transformation technique," Heliyon, vol. 8, no. 12, p. e12655, Dec. 2022, doi: 10.1016/j.heliyon.2022.e12655.

[7] L. Sterckx et al., "Clinical information extraction for preterm birth risk prediction," Journal of Biomedical Informatics, vol. 110, p. 103544, Oct. 2020, doi: 10.1016/j.jbi.2020.103544.

[8] X. Zhang, W. Lu, Y. Pan, H. Wu, R. Wang, and R. Yu, "Empirical study on tangent loss function for classification with deep neural networks," Computers & Electrical Engineering, vol. 90, p. 107000, Mar. 2021, doi: 10.1016/j.compeleceng.2021.107000.

[9] S. Magesh and P. S. Rajakumar, "Ensemble feature extraction-based prediction of fetal arrhythmia using cardiotocographic signals," Measurement: Sensors, vol. 25, p. 100631, Feb. 2023, doi: 10.1016/j.measen.2022.100631.

[10] Y. Lu, X. Zhang, L. Jing, X. Li, and X. Fu, "Estimation of the foetal heart rate baseline based on singular spectrum analysis and empirical mode decomposition," Future Generation Computer Systems, vol. 112, pp. 126–135, Nov. 2020, doi: 10.1016/j.future.2020.05.008.

[11] H. P. van Geijn, A. M. A. Lachmeijer, and F. J. A. Copray, "European multicentre studies in the field of Obstetrics," European Journal of Obstetrics & Gynecology and Reproductive Biology, vol. 50, no. 1, pp. 5–23, Jun. 1993, doi: 10.1016/0028-2243(93)90159-A.

[12] J. Brablik, R. Kahankova, and R. Martinek, "Influence of System Configuration on the Quality of Non-Invasive Fetal Electrocardiography Measurement," IFAC-PapersOnLine, vol. 52, no. 27, pp. 421–426, Jan. 2019, doi: 10.1016/j.ifacol.2019.12.700.

[13] D. J. Jagannath and A. I. Selvakumar, "Issues and research on foetal electrocardiogram signal elicitation," Biomedical Signal Processing and Control, vol. 10, pp. 224–244, Mar. 2014, doi: 10.1016/j.bspc.2013.11.001.

[14] M. G. Frasch et al., "Non-invasive biomarkers of fetal brain development reflecting prenatal stress: An integrative multi-scale multi-species perspective on data collection and analysis," Neuroscience & Biobehavioral Reviews, vol. 117, pp. 165–183, Oct. 2020, doi: 10.1016/j.neubiorev.2018.05.026.

[15] V. P. C. Magboo and Ma. S. A. Magboo, "Prediction of Late Intrauterine Growth Restriction using Machine Learning Models," Procedia Computer Science, vol. 207, pp. 1427–1436, Jan. 2022, doi: 10.1016/j.procs.2022.09.199.

[16] C. Xu, J. Howey, P. Ohorodnyk, M. Roth, H. Zhang, and S. Li, "Segmentation and quantification of infarction without contrast agents via spatiotemporal generative adversarial learning," Medical Image Analysis, vol. 59, p. 101568, Jan. 2020, doi: 10.1016/j.media.2019.101568.

[17] A. M. Oprescu et al., "Towards a data collection methodology for Responsible Artificial Intelligence in health: A prospective and qualitative study in pregnancy," Information Fusion, vol. 83–84, pp. 53–78, Jul. 2022, doi: 10.1016/j.inffus.2022.03.011.

[18] P. Chetlur Adithya, R. Sankar, W. A. Moreno, and S. Hart, "Trends in fetal monitoring through phonocardiography: Challenges and future directions," Biomedical Signal Processing and Control, vol. 33, pp. 289–305, Mar. 2017, doi: 10.1016/j.bspc.2016.11.007.

[19] J. Spilka et al., "Using nonlinear features for fetal heart rate classification," Biomedical Signal Processing and Control, vol. 7, no. 4, pp. 350–357, Jul. 2012, doi: 10.1016/j.bspc.2011.06.008.