Stacked LSTM and Kernel-PCA-based Ensemble Learning for Cardiac Arrhythmia Classification

Azween Abdullah[1], S. Nithya[2], M. Mary Shanthi Rani[3], S. Vijayalakshmi[4], Balamurugan Balusamy[5]
Faculty of Applied Science and Technology, Pedana University, Kuala Lumpur, Malaysia[1]
The Gandhigram Rural Institute (Deemed to be University), Gandhigram[2, 3]
Christ University, Pune[4]
Shiv Nadar Institution of Eminence[5]
Center for Global Health Research, Saveetha Medical College and Hospital,
Saveetha Institute of Medical and Technical Sciences, India[5]

***Abstract*—Cardiovascular diseases (CVD) are the most prevalent causes of death and disability worldwide. Cardiac arrhythmia is one of the chronic cardiovascular diseases that create panic in human life. Early diagnosis aids physicians in securing life. ECG is a non-stationary physiological signal representing the heart's electrical activity. Automated tools to detect arrhythmia from ECG signals are possible with Machine Learning (ML). The ensemble learning technique combines the power of two or more classifiers to solve a computational intelligence problem. It enhances the performance of the models by fusing two or more models, which extremely increases its strength. The proposed ensemble Machine learning amalgamates the potency of Long Short-Term Memory (LSTM) and ensemble learning, opening up a new direction for research. In this research work, two novel ensemble methods of Extreme Gradient Boosting-LSTM (EXGB-LSTM) are developed, which use LSTM as a base learner and are transformed into an ensemble learner by coalescing with Extreme Gradient Boosting. Kernel Principal Component Analysis (K-PCA) is a significant non-linear dimensionality reduction technique. It can manage high-dimensional datasets with various features by lowering the dimensionality of the data while retaining the most crucial details. It has been applied as a preprocessing step for feature reduction in the dataset, and the performance of EXGB-LSTM is tested with and without K-PCA. Experimental results showed that the first method, fusion of EXG-LSTM, has reached an accuracy of 92.1%, Precision of 90.6%, F1-score of 94%, and Recall of 92.7%. The second proposed method, KPCA with EXGB-LSTM, attained the highest accuracy of 94.3%, with a precision of 92%, F1-score of 98%, and Recall of 94.9% for multi-class cardiac arrhythmia classification.**

Keywords*—Arrhythmia classification; ensemble learning; extreme gradient boosting; kernel PCA; LSTM; machine learning

I. Introduction

Cardiovascular diseases have become a significant public health problem globally, accounting for an average of 30% of all deaths worldwide in 2008. With the rapid increase in the world population, cardiovascular disease prevalence has been on a steady rise, increasing related mortality and morbidity. Recent statistical data indicate that CVD caused 17 million fatalities in 2015 [1].

This rising trend in cardiovascular diseases presents a challenge to both developed and developing countries. Moreover, they cause a major economic burden on nations due to the significant costs associated with treatment and management [2].

Each patient's clinical examinations and medical histories provide the basis of the conventional CVD diagnosis paradigm. These results are evaluated under a set of quantitative medical indicators to categorize the patients based on the taxonomy of medical disorders [3].

One of the most dangerous symptoms of heart disease is car -diac arrhythmia. Early arrhythmia prognosis is becoming more significant in heart failure research since it can support clinical patient treatment. Diagnosing potentially fatal cardiac arrhythmias now requires the competence of cardiologists [4].

Cardiac Arrhythmia is an abnormal phenomenon observed in the electrical activity of the human heart. The irregular heart rhythm indicates numerous complications in the heart chamber. This condition is due to an unhealthy lifestyle and various medical conditions [5]. Bradycardia and tachycardia are the two main categories of arrhythmias. Bradycardia is an arrhythmia in which the heartbeat is less than 60 beats per minute (bpm), while tachycardia is an arrhythmia in which the heartbeat can increase to 100 bpm [6]. A sudden onset of a cardiac problem, such as inadequate heart blood flow, shortness of breath, chest pain, fatigue, or unconsciousness, is how arrhythmia often occurs. Abnormal ECG indicates the symptoms. As a result, it's critical to recognize and address arrhythmia as soon as possible [7]. Cardiologists typically spend a lot of time classifying arrhythmias with great precision. Automatic arrhythmia classifiers based on AI algorithms can assist cardiologists in getting higher precision and spending less time diagnosing patients [8].

The Electrocardiogram (ECG) is a painless method to register the electrical impulse generated by the heart in a human being. It is a time-varying signal that reflects ionic current flow resulting from the contraction and subsequent relaxation of the heart [9]. ECG is composed of the P wave that corresponds to the electrical activity of the atria, the Q, R, and S waves composing the QRS complex that corresponds to depolarization of the ventricles, and the T wave that registers the repolarization of the ventricles [10]. Furthermore, the R-R intervals, the time difference between two R waves, can be obtained from the QRS complex. The different components of the ECG signal have specific shapes in the time and frequency domains, which allow specialists to detect possible anomalies by analyzing the ECG signal in these two domains. Moreover, the ECG signal is stochastic and can include abnormal components such as narrower or wider QRS complexes. The abnormal ECG signal can provide important information about cardiac abnormalities and may aid in diagnosing various heart conditions. However, there is still much to learn about the ECG signal and its uses in clinical settings. In recent years, researchers have made significant progress in enhancing the accuracy and efficacy of ECG signals [11].

The ECG signal is a crucial diagnostic tool for identifying heart diseases and assessing cardiac health, so understanding its properties and analysis techniques is paramount. The abnormal heart rhythms are recorded through ECG to help the physicians identify the abnormal rhythm. Earlier diagnosis of irregular beats can avoid various cardiac complications. Prediction of cardiovascular disease is a tedious process for the cardiologist. Hence, the automated arrhythmia diagnosis will assist clinicians in diagnosing and providing timely patient treatment, averting fatalities [12].

The situation will worsen in areas with inadequate medical professionals and clinical equipment, particularly in emerging nations. This drives the need for a trustworthy, automated, affordable monitoring and diagnosing system. The need for this is growing among healthcare professionals so that using computer-aided diagnosis (CADS) systems can be linked to doing the necessary medical assessments [13].

ML is a powerful tool playing a pivotal role in every field, such as Healthcare, Text classification, Plant disease classification, Robotics, Compression, Banking, etc., [14 - 20].

With advancements in technology and signal processing techniques, the accuracy of ECG signal analysis has significantly improved. Furthermore, using ML algorithms has shown promising results in enhancing accuracy in diagnosing arrhythmia.

The significance of effective and precise arrhythmia categorization and detection is becoming more evident with the advent of remotely controlled healthcare systems for heart disease patients. Over the past several years, various ML techniques have been built into diagnostic systems to help with the difficult challenge of accurately classifying arrhythmias from recorded ECG data. Selecting the best methods for identifying and categorizing cardiac disease might be challenging. It entails considering context, analyzing data, and the needs of particular patients' details [21].

Effective ML techniques enable implicit programming-free ML. The goal is to create an algorithm that can take a set of patterns and automatically generalize from preliminary data with or without human intervention. Cluster analysis, or clustering, is a component of unsupervised ML techniques [22-26].

Recently, boosting algorithms have been applied to accelerate the performance of ML algorithms. The concrete objective of this research work is to develop ensemble learning for predicting cardiac arrhythmia challenges.

The novelty of the article includes the following significant contributions:

- Development of Ensemble Model for effective Arrhythmia classification.
- Implementation of fusion of Stacked LSTM and XGBoost algorithms in the classification.
- Application of K-PCA for Feature Reduction.
- Identification of 16 types of arrhythmias with enhanced accuracy by hyper-parameter tuning.
- Validation and comparison of the performance of the proposed models with state-of-the-art methods.

The organization of the paper is outlined as follows. Section I presents the introduction. Section II elaborates on the literature survey in related fields and datasets. Section III presents a brief note on the dataset used for experimentation, followed by the methods used for classification. Section IV depicts the experiments and results obtained, followed by a discussion. Finally, the work is concluded in Section V.

II. Literature Review

Numerous methods are currently available for classifying heart arrhythmias. Some techniques use ML techniques, including support vector machines, neural networks, decision trees, and K Means. This section presents a brief survey of ML and DL-based methods for arrhythmia classification.

Ozcift et al. focused on the following two strategies: i) To choose pertinent characteristics from the cardiac arrhythmia dataset, using a correlation-based feature selection algorithm. ii) The effectiveness of the suggested training method is assessed using the performance of selected features with and without simple random sampling using the Random Forest (RF) ML algorithm. This method achieved an accuracy of 90% [27].

Namsrai et al. developed a new ensemble method using feature selection schema to derive several feature subsets from the original dataset. The feature subsets are then used to train classification models, and the classification performance of each feature set is combined with a voting approach. The ensemble classifier reportedly beats the classifiers learned on the original dataset by 7.28 % for Naive Bayes (NB), 2.27 % for the Support Vector Machine (SVM), 17.84 % for the Decision Tree (DT), and 13.43 % for the Bayes Network of F-score [28].

Guvenir et al. presented a Voting Feature Intervals (VFI), a supervised ML algorithm for detecting arrhythmia classification. A majority vote among the class predictions produced by each characteristic individually is the basis for how it operates. The VFI algorithm performs better than other standard algorithms like Naive Bayesian and K-Nearest Neighbour (KNN), according to comparisons [29].

Freund et al. proposed a novel Adaptive Boosting (AdaBoost) algorithm and the ensemble technique based on the Adaptive Boosting algorithm by engrafting the decision tree into a framework. They created the new boosting algorithm by using the multiplicative weight update technique. This boosting approach doesn't need any prior understanding of how well the weak learning algorithm performs. Additionally, they investigated how the novel boosting approach can be applied to problems involving learning functions whose domain, rather than being binary, is an arbitrarily finite set or a bounded segment of the real line [30].

Friedman et al. developed a general gradient descent boosting paradigm for additive expansions based on any fitting criterion. Tools for analyzing Tree Boost models are described, and special enhancements are built for the particular case where the individual additive components are regression trees. Gradient boosting of regression trees generates highly competitive, dependable, and comprehensible regression and classification algorithms that are particularly suitable for mining imperfect data [31].

Khemchandani et al. developed Twin SVM (TSVM) since conventional SVM has computational complexity. The developed twin SVM is faster in computation and has good generalization for binary classification [32].

Dhyani et al. employed SVM and the 3D Discrete Wavelet Transform (DWT) to characterize and analyze ECG signals. SVM is used to classify the ECG using the nine types of heartbeats identified by the different classifiers. In comparison to the complex support vector machine (CSVM) 98.5% and weighted support vector machine (WSVM) 99%, the SVM classifier has a normal precision of 99.02% [33].

Mohanty et al. experimented with the Cubic Support Vector Machine (CSVM) and the C4.5 classifiers for the classification of Ventricular Fibrillation (VF), ventricular tachycardia (VT), and normal sinus rhythm (NSR). Early detection and prevention of ventricular arrhythmias require the ability to predict VT and ventricular fibrillation VF). By combining temporal, spectral, and statistical data, the researchers described a method for identifying and categorizing VT and VF arrhythmias. The results of this study indicate that the suggested approach outperformed cubic SVM, which had an accuracy of 92.23%, with a sensitivity of 90.97%, specificity of 97.86%, and accuracy of 97.02% in the C4.5 classifier [34].

Yadav et al. determined the significance of particular elements pertinent to diagnosing cardiac arrhythmia. They employed a random forest classifier model, measuring the significance of many features like blood pressure, sodium, potassium, calcium, and respiratory rate, among other features. Hyperparameter optimization methods like grid search and genetic algorithms are compared to find the maximum number and depth of trees in the forest. Area Under the Receiver Operator Curve (AUC) score of 0.9787 was the model's highest performance [35].

Kumari et al. researched classifying arrhythmias more accurately and quickly than before. The research uses an SVM classifier using Discrete Wavelet Transform (DWT). DWT was used to extract 190 features from the prepared data. The SVM classifier, which is the best for classification, was given the retrieved features. The testing set was used to assess the results, and a confusion matrix was used to plot the final results. The model's performance accuracy is 95.92 % [36].

Taher et al. tested with a fusion-based model, which is used to choose the highest-ranked features most effectively. The fusion's output is subsequently subjected to an ensemble of classifiers. High-ranked features acquired from several heuristic feature-selection algorithms are fused using a fuzzy-based feature-selection fusion method. The three primary sections of the work are sensing data and preprocessing, feature queuing, selection and extraction from features, and the predictive model. The method performs classification tasks with higher accuracy, F1 measure, recall, and precision. For binary and categorized class modes, respectively, it obtains an accuracy of 98.5% and 98.9% [37].

Ayar et al. suggested a model for binary and multi-class classification, and the model used genetic algorithms for feature selection and the C4.5 algorithm with DT for classification. The average accuracy for binary classification was 86.96%, and for multi-class classification, it was 78.76% [38].

Jadhav et al. proposed an approach based on feature elimination to diagnose arrhythmia in binary classification. With ensemble sizes of 15 and 20, the random subspace ensemble classifier can identify arrhythmias with an accuracy of 91.11% [39].

Khan et al. (2018) depicted a two-stage cascade structure technique for categorizing cardiac arrhythmia. The first stage used logistic regression (LR), DT, and RF to classify arrhythmias. In contrast, the second stage used Multi-Layer Perceptron’s (MLP) and then LSTM to classify arrhythmias into many classes and binary classes. The accuracy obtained by the first experiment for the LR, DT, and RF was 83.0%, 85.4%, and 87.5%, respectively. The accuracy obtained for the MLP and LSTM experiments was 89.0% and 94.8%, respectively [40].

Itzhak et al. proposed a study on detecting three forms of arrhythmia: Atrial fibrillation, bradycardia, and tachycardia. The authors used random forest, SVM, and logistic regression as classifiers. Their best result was a multiclass random forest classifier that performed with a sensitivity and specificity of 0.92 and 0.86, respectively [41].

Chang et al. found 12 different cardiac rhythm classes using an LSTM model. The LSTM model had an accuracy of 0.982 and an AUC of 0.987 for classifying each of the 12 cardiac rhythms. The F1 score was 0.777, and the precision and recall were 0.692 and 0.625, respectively [42].

Xu et al. developed an innovative ML approach called the Twin Support Vector Machine (TSVM), which seeks to identify two nonparallel planes for each class. The binary classification issue can be solved using traditional TSVM. The multi-class categorization problem can be solved with the Twin-KSVC multi-class classification technique. It combines the benefits of the Support Vector Classification Regression Machine for K-class classification (K-SVCR) with both TSVM and K-SVCR [43].

Mustaqeem et al. selected the wrapper feature selection method in conjunction with SVM-based methods like one-against-one (OAO), one-against-all (OAA), and an error-correction code (ECC) for determining whether an arrhythmia is present or absent. Kappa statistics, accuracy, and Root Mean Square Error (RMSE) measures are used to measure the performance. OAO outperformed all others by scoring 92.07% [44].

Khan et al. tested with Principal Component Analysis (PCA) for feature reduction technique in classifying arrhythmia. The Deep Learning (DL) architecture LSTM is used for classification. An accuracy of 93.5% was attained using PCA and LSTM together [45].

Chen G et al. suggested a wrapper method called Cosine Similarity Measure Support Vector Machines (CSMSVM) by introducing the cosine distance into SVM to reduce features. Traditionally, feature selection methods have extracted features and learned SVM parameters separately or in the attribute space. This may have led to a loss of information about the classification process or increased classification error when the kernel SVM was included. The suggested CSMSVM framework combines feature selection, SVM parameter learning, and removing low-relevance features to achieve better performance [46].

Chandrashekar et al. experimented with several feature selection methods since each underlying algorithm will respond differently to different data types. Using feature selection approaches demonstrates that more information may not be better in ML applications. The feature selection technique for the application is chosen based on the following factors: simplicity, stability, number of reduced features, classification accuracy, storage needs, and computing requirements. It will have advantages like stronger classifier models, improved generalization, and identification of irrelevant variables [47].

Celin et al. used filtering techniques such as low pass, high pass, and Butterworth filter to eliminate the high-frequency noises. The peak detection algorithm is used to locate the peaks, and statistical parameters extract the signal's features. The ML classifiers SVM, Adaboost, ANN, and NB were implemented. The accuracy of the SVM, Adaboost, ANN, and NB classifiers is 87.5%, 93%, 94, and 99.7%, respectively, according to experimental results [48].

Park. J et al. suggested a novel heartbeat classification model that combines an adaptive feature extraction approach and cascade classifiers. The cascade classifiers are constructed from two different random forest classifiers. This can be used with normalized beat morphology features in an environment with limited resources. This model's accuracy was 97.34 %. The random forest classifier was chosen to reduce computational costs and memory even though the k-NN classifier performed better on the initial feature set [49].

Amorim et al. implemented Contourlet, and Shearlet transforms to separate ECG signals into different frequency bands and then extracted features from time-frequency coefficients. They evaluated the performance of KNN, SVM, and Random Forest classifiers on these collected features to categorize seven different forms of beat arrhythmia. With an accuracy of 91.32% and a sensitivity of 90.23%, they used features based on contourlet transforms to get the greatest performance for random forests [50].

Romdhane et al. presented a CNN model to design an algorithm for heartbeat segmentation. This approach is straightforward and efficient because it doesn't need any signal processing that depends on prior knowledge of the signal's morphology or spectrum. The work focused on imbalanced datasets and designed a deep CNN model using a novel focal loss function. The evaluation's findings showed that the targeted loss function increased the overall metrics and the classification accuracy for imbalanced classes. Our suggested strategy achieved 98.41% accuracy, 98.38% F1-score, 98.37% precision, and 98.41% recall [51].

Sharma et al. employed both noisy and denoised (clean) ECG signals. The ECG signals were wavelet decomposed using a dyadic orthogonal filter bank with stop-band energy (SBE) minimization. Characteristics based on fuzzy entropy, Renyi entropy, and fractal dimension were then extracted for a quicker and more precise classification into the five classes. These features were then fed into classifiers, with the KNN classifier attaining maximum accuracy (MAAC) of 98.1%, maximum sensitivity (MASE) of 85.63%, and maximum specificity (MASP) of 98.27% for clean data and MAAC of 98.1%, maximum sensitivity (MASE) of 85.33% for noisy data. This technology can be utilized in intensive care units to help clinicians make fast diagnoses [52].

Chen et al. developed a novel methodology of piecewise linear splines for feature selection and a gradient-boosting algorithm to classify atrial fibrillation and other cardiac dysrhythmias. A piecewise linear spline is used to fit the ECG waveform, and morphological properties associated with the coefficients of the piecewise linear spline are retrieved. The morphological coefficients and heart rate variability features are categorized using XGBoost. The algorithm received an average F1 score on the independent testing set of 81% and an F1 score on a 10-fold cross-validation of 81%. With certain morphological features, the system performs well on multi-label short ECG classification [53].

Assodiky et al. employed LSTM to categorize the types of arrhythmias, using AdaDelta as the adaptive learning rate method and improved performance. It was contrasted with LSTM, which lacked an adaptive learning rate. The best outcome that demonstrated great accuracy was produced by combining LSTM and AdaDelta. The accurate classification rate for train and test data was 98% and 97%, respectively [54].

Although the aforementioned techniques have shown some reasonable accuracy in classifying arrhythmias, more work has to be done, such as increasing accuracy. Most arrhythmia classification techniques currently in use are based on traditional ML techniques, which emphasize careful feature selection and classification less. DL techniques have found widespread usage due to the quick development of tools for working with high-dimensional data and advancements in computing power. Despite several research works reported in the literature, there is still a pressing need for novel strategies to improve the accuracy and effectiveness of classification models for Arrhythmia.

III. Materials And Methods

A. Experimented Dataset

The proposed research work has been experimented with the Arrhythmia dataset available in the UCI Irvine ML repository. The dataset contains data on 452 patients in total, and each record has 279 features that describe the traits of an arrhythmia. Age, sex, height, and weight are the first four characteristics that best describe the subject. The remaining 274 features are taken from recordings of typical 12-lead ECGs. The heart muscle depolarizes during each heartbeat, causing small electrical changes on the skin that are detected and amplified by the ECG. A 12-lead ECG captures 12 distinct electrical heart impulses at approximately the same time. Arrhythmia data has an extensive feature size, which may need a feature reduction approach. It has numerous missing values. Some columns with missing values are eliminated, and the experiments are carried out. The goal is to categorize cardiac arrhythmia into one of the three categories and to distinguish between its presence and absence. Class 01 denotes regular ECG classes, Class 02 through Class 15 denotes various arrhythmia classes, and Class 16 denotes the remaining unclassified ones [63].

B. Methods

The proposed work reports on two methodologies EXGB-LSTM and KPCA-EXG-LSTM. The proposed methods classify arrhythmia through a pipeline that includes extreme boosting and stacked LSTM together and implemented in the preprocessed dataset. The flow diagram of the proposed work is shown in Fig.1.

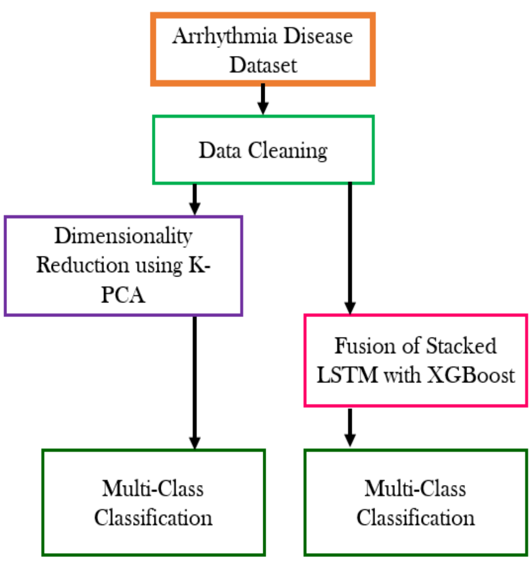

Fig. 1. A flow diagram of EXG-LSTM with K-PCA.

The implementation of the proposed model EXGB-LSTM is shown in the Algorithm 1.

Algorithm 1 EXGB-LSTM
Input: UCI Arrhythmia Machine Learning Dataset *Output: Classification of 16 classes of Arrhythmia*
Begin **Phase I : Data Cleaning** Eliminating colum ns with more number of missing values. **Phase II : Applying Stacked LSTM with XGBoost** Fusion of Stacked LSTM with XGBoost in the Processed dataset. **Phase III: Multi-Class Classification of Arrhythmia** **End**

In the second method, the dataset is dimensionally reduced by K-PCA, followed by ensemble learning using LSTM and extreme boosting and classification. The steps of the proposed K-PCA with the EXGB-LSTM model are outlined below in the Algorithm 2.

Algorithm 2 K-PCA + EXGB-LSTM
Input: UCI Arrhythmia Machine Learning Dataset *Output: Classification of 16 classes of Arrhythmia*
Begin **Phase I : Data Cleaning** Eliminating columns with more missing values. **Phase II : Dimensionality Reduction using Kernel-PCA** **Phase III: Applying Stacked LSTM with XGBoost** Fusion of Stacked LSTM with XGBoost in the Processed dataset. **Phase IV : Multi-Class Classification of Arrhythmia** **End**

C. Kernel Methods

Dimensionality reduction plays a significant role in effectively handling massive dimensional data. The motive of dimensionality reduction may be noise reduction, pre-processing, compression, etc. PCA is a mathematical approach to converting several correlated variables into uncorrelated variables known as Principal components (PC). This PC represents the maximum variance in the dataset. PCA works only for linear structures, and K-PCA has been developed as a non-linear extension of standard PCA. The primary notion of K-PCA is to map the original data into a huge dimensional space through a specific function and then implement the PCA algorithm to it. The linear PCA of the vast dimensional feature space represents a non-linear. Most interesting directions can be found in the original input space by the PCA [64]. It is mainly applied in complex non-linear data such as 3D reconstruction, handwritten digits, bioinformatics face images, natural signals, and geostatistics kriging. Various kernel methods include polynomial, Gaussian, Laplace Radial Basis Function, Sigmoid, and hyperbolic tangent. One of the merits of K-PCA is that original data can be regenerated from its principal components [65] in the following three steps.

- Computing Covariance Matrix

The Covariance Matrix(C) is used to represent the relationship between two random variables and is defined in Eq. (1):

$$C = \frac{1}{n} \sum_{i=1}^{n} \phi x^i \ \phi x_i^T \quad (1)$$

where ϕx^i represents the input feature space, x^T represents the Transpose and x^i , is one of the 'n' multivariate observations. Also, the mean of the data in the feature space is assumed to be zero.

- Decomposition of Eigenvalues

Eigen decomposition of an input data distribution results in the generation of eigenvectors, which form the principal components of the data as given in Eq. (2):

$$\lambda v = Cv \quad (2)$$

where 'λ ' is the eigenvalue and 'v' is the eigenvector which is a linear combination of samples.

- Selection of Eigenvectors and transform data

A function that holds the vectors in the original input space and when the dot product of the vectors in the feature space is returned then it is known as a kernel function as is shown in Eq.(3):

$$K(x_i, x_k) = \phi x_i^T \phi x_k \quad (3)$$

Kernel extracts up to 'n' samples in non-linear PCs without expensive computations [65].

D. LSTM

A conventional Recurrent Neural Network (RNN) is an extension of a feedforward neural network that can handle variable lengths of sequential input, also called an Auto Associative or Feedback Network. The neural network model is structured to preserve previously hidden neurons' output, and it is fed as input to the next cell [66]. LSTM was first proposed in 1997 by Sepp Hochreiter and Jurgen Schmidhuber. A modified version of RNN makes each recurrent unit adaptively capture dependencies of different time scales. It has a forget cell to modulate the flow of information. RNN suffers from complex training time for longer sequences and vanishing gradient problems. LSTM was structured to eliminate gradient problems with three gates namely, an input gate i_t, an output gate o_t, and forget gate. . The design of the LSTM cell is represented in Fig. 2.

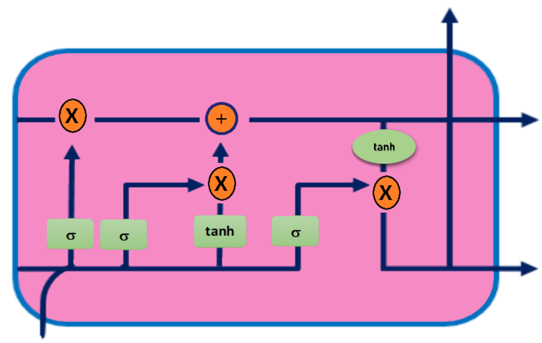

Fig. 2. LSTM cell.

The Model of LSTM is working in principle by representing the input sequences as {x_1, x_2, x_3…., x_t}, {y_1, y_2, y_3…., y_t} as an output sequence and {S_1, S_2, S_3…., S_t} as an internal sequence of cell states.

In the LSTM cell, the three gates are computed with Eq..(4-5):

$$i_t = \sigma\,(W_i \,.\, [y_{t-1}\,,\ x_t] + b_i) \quad (4)$$

$$f_t = \sigma\,(W_f \,.\, [y_{t-1},\ x_t] + b_f) \quad (5)$$

$$O_t = \sigma\,(W_o \,.\, [y_{t-1},\ x_t] + b_o) \quad (6)$$

where W_i , W_f , and W_o denote the weight of the corresponding input, and b_i, b_f, and b_o are the respective bias in the cell. The activation function sigmoid is represented as 'σ' in the equation.

The candidate state $\widehat{S_t}$ is computed with the Eq. (7):

$$\widehat{S_t} = \tanh\,(W_s \,.\, [\,y_{t-1}, x_t] + b_s) \quad (7)$$

The current cell state value S_t is calculated from the previous cell state value S_{t-1} using Eq. (8):

$$S_t = f_t \ * S_{t-1} + i_t * \widehat{S_t} \quad (8)$$

The output of the cell y_t is determined by Eq. (9):

$$y_t = o_t \ * \tanh(S_t) \quad (9)$$

The input gate in the LSTM model chose what to store in the current cell. The output gate determines what is to be passed over. The forget gate determines how to forget the current state [65].

E. Stacked LSTM

Multiple hidden layers in stacked LSTM architectures can be used to increase the amount of abstraction gradually. They operate better because they give a more accurate representation of sequence data. In the stacked LSTM architecture, the input from the previous LSTM hidden layer is passed into the current LSTM hidden layer. This has the potential to improve neural network performance significantly. Including enough layers in the model improves the LSTM model's accuracy and dependability. Fewer neurons are needed in each layer, which cuts down on training time. In this study, a stacked LSTM network is suggested with the fusion of XGBoost [67, 68]

F. Extreme Gradient Boosting

The eXtreme Gradient Boost (XGBoost) is a decision tree-based algorithm that transforms weak learners into strong learners. Boosting is an iterative algorithm compared to bagging, a parallel algorithm. The computational procedure for constructing a decision tree in gradient boosting is very time-consuming. XGBoost is an optimized version of the gradient boosting algorithm, which improves the training time. Many models were trained on different subsets of data, and the best-performing model was chosen. It is best suited for non-linearity problems [69, 70]. Combining the models is part of ensemble learning, where different models are combined to optimize the algorithm's performance. Stacking is one of the usual processes in combining models and training an ensemble model on the classification, enabling convenient experimentation and model comparisons. Even though Light Gradient Boosting is faster when compared with the performance it is equivalent [71-73]. The framework of the developed method is presented in Fig. 3.

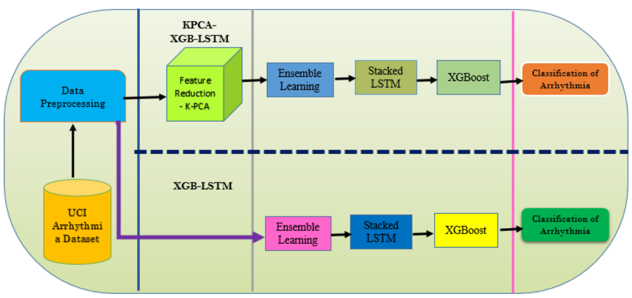

Fig. 3. A framework of EXG-LSTM with K-PCA.

IV. Experimentation and Analysis

A. Experimental Setup

The proposed model uses an Intel Core(R) i5-9300H CPU running at 2.40 GHz, an NVIDIA GeForce® GTX 1050 graphics card, a 64-bit operating system, and 8GB of RAM.

The experiment on ensemble LSTM was carried out with the Arrhythmia dataset. The medical dataset consists of 206 linear valued features and a few categorical features among the 279 features. 452 instances are present in the dataset. It has been distinguished into 16 groups to indicate the existence and non-existence of arrhythmia. Class 1 shows healthy or normal samples. The total number of healthy samples is 245. Classes 2-15 are categorized as different types of arrhythmias. Class 16 indicates the other unclassified arrhythmia. As the dataset may have null values for certain features. Those features having more missing values are removed during the pre-processing stage. The remaining features with fewer missing values are treated by the statistical mean method. The proposed work was tried in two different ways. The dataset was chosen directly after applying data cleaning and data normalization methods to implement LSTM in the first experiment. To improve the performance of LSTM, XGBoost was added, and it showed a better performance. In the second method, the dataset was further subjected to feature reduction by K-PCA sequentially followed by the stacking LSTM with XGBoost. The parameters used for K-PCA is listed in the Table I.

TABLE I. Experimental Parameters

Experimental Parameters for K-PCA	Values
n_components	94.3
Kernel	rbf
Gamma	0.04
Alpha	0.03
fit_inverse_transform	True
eigen_solver	Auto
n_jobs	-1

The dataset is split into 80/20 for training and testing. The results obtained by both proposed methods are shown in Table II.

TABLE II. Performance of Proposed Methods

Classifier	**Accuracy**	**Precision**	**F1-score**	**Recall**	**No. of Features**
EXG-LSTM	92.1	90.6	94.0	92.7	252
KPCA+EXG-LSTM	94.3	92.0	98.0	94.9	135

The proposed model EXG-LSTM reached an accuracy of 92.1%, a precision of 90.6%, a recall of 92.7%, and an F1-score of 94 % with the features 252. The second method performed exceptionally well when applying K-PCA compared with the result of EXG-LSTM by accomplishing an accuracy of 94.3%, a precision of 92 %, a recall of 94.9%, and with an F1-score of 98% by reducing the feature into 135. The proposed methods' results are compared with the state-of-the-art techniques presented in Table II. Using the cardiologist as the gold standard, the attempt is made to lessen this difference by utilizing ML methods.

TABLE III. Performance of the Proposed Methods with State-of-the-Art Methods

Authors	**Classifier**	**Accuracy**	**Class**
Zuo et al. [55] (2008)	KDF-WKNN	70.66	2
Kohli et al. [56] (2007).	OAK	73.40	2
Niazi et al. [57] (2015).	KNN with IFSFS	73.80	2
Prathibhamol et al. [58] (2017).	P-CA-CRA	80.00	16
Pandey et al. [59] (2019).	RNN	83.10	2
Durga [60] (2021).	DT	84.13	-
Jadhav et al. [39] (2014).	MLP	86.67	2
Mitra et al. [61] (2013).	CFS+LM	87.71	2
Raut et al. [62] (2008).	MLPNN	90.00	7
Mustaqeem et al. [44] (2018).	OAO	92.07	16
Khan et al. [45] (2021).	PCA + LSTM	93.50	16
Proposed Method 1	**EXG-LSTM**	**92.10**	**16**
Proposed Method 2	**KPCA+EXG-LSTM**	**94.30**	**16**

Table III clearly reveals the superior performance of our proposed methods. It is worth noting that our proposed methods are capable of classifying 16 classes of arrhythmia with the highest accuracy of 94.30 when compared with state-of-the-art methods as depicted in Fig. 4.

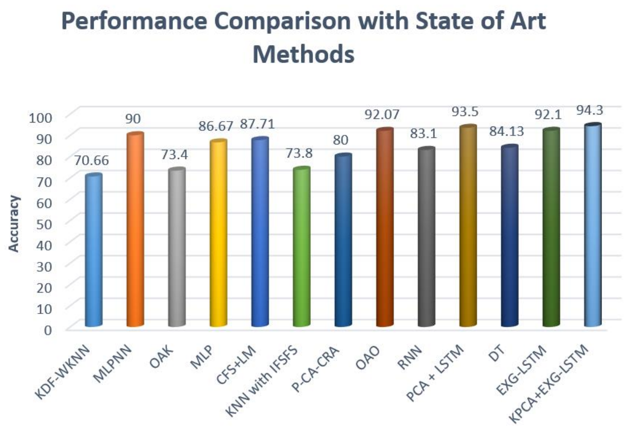

Fig. 4. Proposed models with state-of-the-art methods.

Arrhythmias are potentially fatal. Hence, it is imperative that prediction and analysis be employed in the medical profession with the greatest possible accuracy. The proposed approach highlights these issues by selecting a few notable features using a feature selection mechanism, which enhances classification efficiency.

V. CONCLUSION

A novel hybrid DL pipeline for multi-class Arrhythmia classification has been proposed. This approach involves selecting maximally distinct features using the feature selection technique and helps improve the performance of the classification of arrhythmia data. The work implied Kernel-PCA and a Stacked LSTM DL technique with the fusion of XGBoost for classification to determine whether an arrhythmia was present or absent. This research work aims to upgrade the performance of LSTM through the novel pipeline with XGBoost. It was good in its execution and recorded a high accuracy of 92.1%. Concurrently, feature reduction was done, and the performance of the ensemble of LSTM is monitored. Both the proposed models performed well. Among them, EXG-LSTM with K-PCA has reached the highest accuracy of 94.3%. This research works on stacking LSTM shows that the strength of LSTM was boosted by ensembling, and showed better results in classifying the classes of arrhythmia efficiently. Arrhythmias are deadly disorders. Hence prediction and analysis must be exceedingly accurate before being applied in the medical field. The suggested method draws attention to these problems by choosing several standout features with an improved feature selection mechanism, which enhances classification efficiency with reduced number of features of 135. The future direction of the research would be to construct an interpretable ML classifier to augment the classifier's accuracy that will serve as an effective handy diagnostic tool for physicians.

REFERENCES

[1] M. Nishizaki, "Life-threatening arrhythmias leading to syncope in patients with vasospastic angina", Journal of Arrhythmia, 2017, 33(6), 553-561.

[2] R. J. Koene, W.O. Adkisson, & D. G. Benditt, "Syncope and the risk of sudden cardiac death: Evaluation, management, and prevention" Journal of arrhythmia, 2017, 33(6), 533-544.

[3] M. M. Al Rahhal, Y. Bazi and H. A. Hichri. "Deep learning approach for active classification of electrocardiogram signals," Information Sciences, 2016, vol. 345, pp. 340–354.

[4] A. Krizhevsky, I. Sutskever, & G.E. Hinton, "Imagenet classification with deep convolutional neural networks", Communications of the ACM, 2017, 60(6), 84-90.

[5] H. Sugumar, S. Prabhu, A. Voskoboinik, & P.M. Kistler, "Arrhythmia induced cardiomyopathy. Journal of arrhythmia, 2018, 34(4), 376-383.

[6] A. Mustaqeem, S. M. Anwar and M. Majid. "Multiclass classification of cardiac arrhythmia using improved feature selection and SVM invariants," Computational and Mathematical Methods in Medicine, vol. 2018, pp. 1–11.

[7] E. Alickovic and A. Subasi. "Medical decision support system for diagnosis of heart arrhythmia using DWT and random forests classifier," Journal of Medical Systems, 2016, vol. 40, no. 108, pp. 1–12.

[8] P. Rajpurkar, A.Y. Hannun, M. Haghpanahi, C. Bourn, & A.Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks". arXiv preprint arXiv:1707.01836.

[9] S. Chen, A. Li, & J.M. Roveda, Artificial Intelligence in Cardiac Arrhythmia Classification. learning, 120, 268-275.

[10] S. U. Kumar and H. H. Inbarani, "Neighborhood rough set-based ECG signal classification for diagnosis of cardiac diseases," Soft Computing, 2017, vol. 21, no. 16, pp. 4721–4733.

[11] Y. Ozbay and B. Karlik "A recognition of ECG arrhythmias using artificial neural networks," in Proc. IEEE Engineering in Medicine and Biology Society, Istanbul, Turkey, pp. 1680–1683.

[12] Z. Ebrahimi, M. Loni, M. Daneshtalab, & A. Gharehbaghi, "A review on deep learning methods for ECG arrhythmia classification". Expert Systems with Applications: 2020, X, 7, 100033.

[13] C. G. Nayak, G. Seshikala and U. Desai, "Identification of arrhythmia classes using machine-learning techniques", International Journal of Biology and Biomedicine, 2016, vol. 1, pp. 48–53.

[14] M. Mary Shanthi Rani, P. Chitra, S. Lakshmanan, M. Kalpana Devi, R. Sangeetha, & S. Nithya, "DeepCompNet: A Novel Neural Net Model Compression Architecture", Computational Intelligence and Neuroscience, 2022, https://doi.org/10.1155/2022/2213273

[15] N. Karthikeyan, M. S. Rani, "ECG Classification Using Machine Learning Classifiers with Optimal Feature Selection Methods", In Evolutionary Computing and Mobile Sustainable Networks, 2022, pp. 277-289. Springer, Singapore. https://doi.org/10.1007/978-981-16-9605-3_19

[16] R. Sangeetha, M. Mary Shanthi Rani, R. Joseph R, "Optimized Deep Neural Network for Tomato Leaf Diseases Identification", In International Advanced Computing Conference,2021, pp. 562-576, Springer, Cham. https://doi.org/10.1007/978-3-030-95502-1_42

[17] S. Nithya, & M.S. Rani, "Deep Learning Model for Arrhythmia Classification with 2D Convolutional Neural Network", In Innovations in Information and Communication Technologies, 2022, pp.1-11, Springer, Singapore. https://doi.org/10.1007/978-981-19-3796-5_1

[18] S. Nithya, & M.S.Rani, "Stacked Variational Autoencoder in the Classification of Cardiac Arrhythmia using ECG Signals with 2D-ECG Images", In International Conference on Intelligent Innovations in Engineering and Technology, 2022, pp. 222-226). IEEE Xplore. https://doi: 10.1109/ICIIET55458.2022.9967575.

[19] S. Gupta, R.Sangeeta, R.S. Mishra, G. Singal, T. Badal, & D.Garg, "Corridor segmentation for automatic robot navigation in indoor environment using edge devices", Computer Networks, 2020, 178, 107374. https://doi.org/10.1016/j.comnet.2020.107374

[20] S. Nithya, & M.S. Rani, "XACML: Explainable Arrhythmia Classification Model Using Machine Learning", In International

Advanced Computing Conference, 2022, pp. 219-231, Cham: Springer Nature Switzerland.

[21] C. H. Hsing, "A hybrid intelligent model of analyzing clinical breast cancer data using clustering techniques with feature selection," Applied Soft Computing, 2014, vol. 20, pp. 4–14.

[22] E. R. Hruschka and N. F. Ebecken, "Extracting rules from multilayer perceptron's in classification problems: A clustering-based approach," Neurocomputing, 2006, vol. 70, no. 3, pp. 384–397.

[23] M. Nilashi, "A soft computing approach for diabetes disease classification," Health Informatics Journal, 2016, vol. 24, no. 4, pp. 379–393.

[24] M. Nilashi, O. Ibrahim and A. Ahani, "Accuracy improvement for predicting Parkinson's disease progression," Scientific Reports, 2016, vol. 6, no. 1, pp. 34–181.

[25] M. Nilashi, O. Ibrahim, H. Ahmadi and L. Shahmoradi. "A knowledge-based system for breast cancer classification using the fuzzy logic method," Telematics and Informatics, 2017, vol. 4, no. 34, pp. 133–144.

[26] K. Polat. "Classification of Parkinson's disease using a feature weighting method on the basis of fuzzy C-means clustering," International Journal of Systems Science, 2012, vol. 4, no. 4, pp. 597–609.

[27] A. Ozçift, A. Random forests ensemble classifier trained with data resampling strategy to improve cardiac arrhythmia diagnosis. Computers in biology and medicine, 2011, 41(5), 265-271.

[28] E. Namsrai, T. Munkhdalai, M. Li, J.H. Shin, O.E. Namsrai, K.H.A. & Ryu, "A feature selection-based ensemble method for arrhythmia classification". Journal of Information Processing Systems, 2013, 9(1), 31-40.

[29] H. A. Guvenir, B. Acar, G. Demiroz, & A. Cekin, "A supervised machine learning algorithm for arrhythmia analysis", In Computers in Cardiology, 1997, (pp. 433-436). IEEE.

[30] Y. Freund, R E. Schapire, "A decision-theoretic generalization of online and an application to boosting", Journal of computer and system sciences, 55(1), 119-139. (1997). https://doi.org/10.1006/jcss.1997.1504

[31] J H. Friedman, "Greedy function approximation: a gradient boosting machine", Annals of Statistics, 2001, 1189-1232.

[32] R. Khemchandani, S. Chandra S, "Twin support vector machines for pattern classification", IEEE Transactions on pattern analysis and machine intelligence, 29(5), 905-910, 2007. https://doi.org/10.1109/TPAMI.2007.1068

[33] S. Dhyani, A. Kumar, & S. Choudhury, "Analysis of ECG-based arrhythmia detection system using machine learning". MethodsX, 2023, 10, 102195.

[34] M. Mohanty, S. Sahoo, P. Biswal, & S. Sabut, "Efficient classification of ventricular arrhythmias using feature selection and C4. 5 classifier", Biomedical Signal Processing and Control, 2018, 44, 200-208.

[35] S. S. Yadav, & S. M. Jadhav, Detection of common risk factors for diagnosis of cardiac arrhythmia using machine learning algorithm. Expert systems with applications, 2021, 163, 113807.

[36] C. U. Kumari, A. S. D. Murthy, B. L. Prasanna, M. P. P. Reddy, & A. K. Panigrahy, "An automated detection of heart arrhythmias using machine learning technique: SVM", Materials Today: Proceedings, 2021, 45, 1393-1398.

[37] F. Taher, H. Alshammari, L. Osman, M. Elhoseny, A.Shehab, & E. Elayat, "Cardiac Arrhythmia Disease Classifier Model Based on a Fuzzy Fusion Approach". Computers Materials & Continua, 2023, 75(2), 4485.

[38] M. Ayar, M., & S. Sabamoniri, "An ECG-based feature selection and heartbeat classification model using a hybrid heuristic algorithm", Informatics in Medicine Unlocked, 2018, 13, 167-175.

[39] S. Jadhav, S. Nalbalwar, & A. Ghatol, "Feature elimination based random subspace ensembles learning for ECG arrhythmia diagnosis", Soft Computing, 2014, 18, 579-587.

[40] M. A. Khan, M.R. Karim, & Y. Kim, "A two-stage big data analytics framework with real world applications using spark machine learning and long short-term memory network", Symmetry, 2018, 10(10), 485.

[41] S. B. Itzhak, S. S. Ricon, S. Biton, J. A. Behar, & J. A. Sobel, Effect of temporal resolution on the detection of cardiac arrhythmias using HRV features and machine learning. Physiological Measurement,2022, 43(4), 045002.

[42] K. C. Chang, P. H. Hsieh, M. Y. Wu, Y. C. Wang, J. Y. Chen, F. J Tsai,., ... & T.C. Huang, Usefulness of machine learning-based detection and classification of cardiac arrhythmias with 12-lead electrocardiograms. Canadian Journal of Cardiology, 2021, 37(1), 94-104.

[43] Y. Xu, R. Guo, L. Wang, "A twin multi-class classification support vector machine", Cognitive Computation, 2013, 5(4), 580-588. https://doi.org/10.1007/s12559-012-9179-7

[44] A. Mustaqeem, S.M. Anwar, M. Majid, "Multiclass classification of cardiac arrhythmia using improved feature selection and SVM invariants", Computational and mathematical methods in medicine, 2018. https://doi.org/10.1155/2018/7310496

[45] M. A.Khan, Y. Kim, "Cardiac arrhythmia disease classification using LSTM deep learning approach", CMC-COMPUTERS MATERIALS & CONTINUA, 2021, 67(1), 427-443.

[46] G. Chen, & J.Chen, "A novel wrapper method for feature selection and its applications", Neurocomputing, 2015, 159, 219-226.

[47] G. Chandrashekar, & F. Sahin, "A survey on feature selection methods", Computers & Electrical Engineering, 2014, 40(1), 16-28.

[48] S. Celin, & K. Vasanth, "ECG signal classification using various machine learning techniques", Journal of medical systems, 2018, 42(12), 241.

[49] J. Park, M.Kang, J.Gao, Y.Kim, & K.Kang, "Cascade classification with adaptive feature extraction for arrhythmia detection", Journal of medical systems, 2017,41,1-12.

[50] P. Amorim, T. Moraes, D. Fazanaro, J. Silva, & H. Pedrini, "Shearlet and contourlet transforms for analysis of electrocardiogram signals", Computer Methods and Programs in Biomedicine, 2018, 161, 125-132.

[51] T. F Romdhane, & M.A. Pr, "Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss", Computers in Biology and Medicine, 2020, 123, 103866.

[52] M.Sharma, R.S. Tan, & U.R. Acharya, "Automated heartbeat classification and detection of arrhythmia using optimal orthogonal wavelet filters", Informatics in Medicine Unlocked, 2019, 16, 100221.

[53] Y. Chen, X. Wang, Y. Jung, V. Abedi, R. Zand, M. Bikak, & M. Adibuzzaman, "Classification of short single-lead electrocardiograms (ECGs) for atrial fibrillation detection using piecewise linear spline and XGBoost", Physiological measurement, 2018, 39(10), 104006.

[54] H. Assodiky, I. Syarif, & T. Badriyah, " Arrhythmia classification using long short-term memory with adaptive learning rate", EMITTER International Journal of Engineering Technology, 20158, 6(1), 75-91.

[55] W. M. Zuo, W. G. Lu, K. Q. Wang, H. Zhang, "Diagnosis of cardiac arrhythmia using kernel difference weighted KNN classifier". In Computers in Cardiology (pp. 253-256). IEEE, 2008. https://doi.org/10.1109/CIC.2008.4749025

[56] N. Kohli, N.K. Verma, A. Roy A, " SVM-based methods for arrhythmia classification in ECG", In International conference on computer and communication technology (ICCCT) (pp. 486-490). IEEE, 2010. https://doi.org/0.1109/ICCCT.2010.5640480

[57] K. A. K. Niazi, S. A. Khan, A. Shaukat, M. Akhtar M, "Identifying the best feature subset for cardiac arrhythmia classification", In 2015 Science and Information Conference (SAI) (pp. 494-499). IEEE, 2015. https://doi.org/10.1109/SAI.2015.7237188

[58] Cp P, A. Suresh, G. Suresh, "Prediction of cardiac arrhythmia type using clustering and regression approach (P-CA-CRA)", In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 51-54). IEEE, 2017. https://doi.org/10.1109/ICACCI.2017.8125815

[59] S. K. Pandey, R.R. Janghel, "ECG arrhythmia classification using artificial neural networks", In Proceedings of 2nd International Conference on Communication, Computing and Networking (pp. 645-652). Springer, Singapore. 2019. https://doi.org/10.1007/978-981-13-1217-5_63.

[60] S. Durga, E. Daniel, S.D. Kanmani J.M. Philip, "Cardiac arrhythmia classification using sequential feature selection and decision tree

classifier method", International Journal of Innovative Computing and Applications, 2021, 12(4), 175-182.

[61] M. Mitra, R.K. Samanta, "Cardiac arrhythmia classification using neural networks with selected features", Procedia Technology, 2013, 10, 76-84. https://doi.org/10.1016/j.protcy.2013.12.33.

[62] R. D.Raut, S.V. Dudul, "Arrhythmias classification with MLP neural network and statistical analysis", In 2008 First International Conference on Emerging Trends in Engineering and Technology (pp. 553-558). IEEE, 2008. https://doi.org/10.1109/ICETET.2008.260

[63] D. Dua, C. Graff, " UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine", CA: University of California, School of Information and Computer Science. (2019).

[64] M. Kallas, C. Francis, L. Kanaan, D. Merheb, P. Honeine, H. Amoud, Multi-class SVM classification combined with kernel PCA feature extraction of ECG signals. In 2012 19th International Conference on Telecommunications (ICT) (pp. 1-5). IEEE, 2012. https://doi.org/10.1109/ICTEL.2012.6221261

[65] S. Yang, H. Shen, "Heartbeat classification using discrete wavelet transform and kernel principal component analysis", In IEEE 2013 Tencon-Spring (pp. 34-38). IEEE. 2013. https://doi.org/ 10.1109/TENCONSpring.2013.6584412

[66] N. Merrill, C. Olson, "Unsupervised Ensemble-Kernel Principal Component Analysis for Hyperspectral Anomaly Detection", In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 112-113.

[67] Y. Dong, Y. Zhang, F. Liu, & X. Cheng, "Reservoir Production Prediction Model Based on a Stacked LSTM Network and Transfer Learning. ACS omega, 2021, 6(50), 34700-34711.

[68] A. Sahar, & D. Han, "An LSTM-based indoor positioning method using Wi-Fi signals", In Proceedings of the 2nd International Conference on Vision, Image and Signal Processing ,2018, pp. 1-5.

[69] T. I. Poznyak I. Chairez Oria, A.S. Poznyak, "Background on dynamic neural networks", Ozonation and Biodegradation in Environmental Engineering, 2019,57-74.

[70] X. Yang, Z. Chen Z, "A Hybrid Short-Term Load Forecasting Model Based on CatBoost and LSTM", In 2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP) (pp. 328-332). IEEE, 2021. https://doi.org/ 10.1109/ICSP51882.2021.9408768

[71] T. Chen, C. Guestrin, "xgboost: A scalable tree boosting system", In Proceedings of the 22nd ACM sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785-794. 2016. https://doi.org/10.1145/2939672.2939785

[72] R. Lorbieski, S.M. Nassar, "Impact of an Extra Layer on the Stacking Algorithm for Classification Problems", J. Comput. Sci., 14(5), 613-622. 2018. DOI: 10.3844/jcssp.2018.613.622

[73] S. Sang, F. Qu, P. Nie, "Ensembles of Gradient Boosting Recurrent Neural Network for Time Series Data Prediction", IEEE Access. 2021. https://doi.org/ 0.1109/ACCESS.2021.3082519