VER
Vascular & Endovascular Review

Heart Disease Prediction Using Parallel Ensemble Deep Learning Techniques

N. Vellingiri[1]*, S. Ranjitha Kumari[2]

[1]*PhD Research Scholar, Department of computer science, RVS College of Arts and Science, Sulur, Coimbatore.
Email Id: vellingiri.nm@gmail.com
[2]Associate Professor, Department of computer science, RVS College of Arts and Science, Sulur, Coimbatore.
Email Id: ranjithakumari@rvsgroup.com

ABSTRACT

To guarantee proper categorization and allow cardiac specialists to treat patients effectively, it is essential that heart disease be diagnosed and prognoses. Implementation of machine learning in the medical field have expanded as a result of its capacity to recognize patterns in data. The objective of this research is to decrease the amount of deaths from heart disease by creating a model that can accurately forecast these issues. The five primary stages of this study are preprocessing, clustering, extraction of feature, feature selection, and classification. Initially the min-max normalization is used for pre-processing. The K-Means Clustering (KMC) technique is used for clustering with the goal of improving classifier accuracy. This work suggests using a Principal Component Analysis (PCA) to efficiently extract features. After that, the feature selection process is finished using the modified whale optimization algorithm method. To choose the most pertinent and significant characteristics from the cardiovascular illness dataset, it computes the optimal fitness value. Finally, the classification is done using parallel ensemble deep learning model as VGG16, Inceptionv2, ResNet50 and used the Mutation Albatross Optimization Algorithm (MAO) for hyperparameter optimization. To improve health outcomes and lessen the demand on healthcare, these models can be used to create future cardiovascular disease detection systems.

KEYWORDS: Heart disease data, k-means clustering, Principal Component Analysis (PCA), parallel ensemble deep learning model and Mutation Albatross Optimization Algorithm (MAO), VGG16, Inceptionv2, ResNet50.

How to Cite: N. Vellingiri*, S. Ranjitha Kumari, (2025) Heart Disease Prediction Using Parallel Ensemble Deep Learning Techniques, Vascular and Endovascular Review, Vol.8, No.17s, 325-337.

INTRODUCTION

Investigating the hidden patterns in data sets from the medical field is made possible by medical data mining. It is possible to make clinical diagnoses using these patterns. On the other hand, the raw medical data that is now accessible is vast, diverse, and widely dispersed. This data must be gathered in a methodical manner [1]. A healthcare information system may be built from the data. A user-friendly method for detecting hidden patterns is data mining. Heart disease kills 12 million people annually, according to the WHO. In the United States and other wealthy countries, 50% of all fatalities are caused by cardiac diseases. In many underdeveloped nations, it is also the leading cause of deaths [2]. In general, Adult deaths are believed to be mostly caused by it. The phrase "heart diseases" encompasses any conditions that cause problems with the heart. In India and other countries, heart disease accounted for the majority of deaths.

In the United States, every thirty-four seconds, a person died due to heart disease. Heart diseases include cardiomyopathy, or cardiac disease, and coronary heart disease. Heart and blood arterial abnormalities are referred to as "cardiovascular disease" in this context, and the pumping and circulation of blood throughout the body [3]. Several conditions, impairment, and mortality are caused by cardiovascular disease (CVD). Several of the most important and complex tasks in medicine is diagnosing diseases. It is well acknowledged that medical diagnosis is a significant yet challenging undertaking that requires precise and effective execution. It would be really beneficial if this method were automated. Unfortunately, not every doctor is a specialist there are insufficient resource members in certain regions and in every field [4]. Thus, it would probably be very advantageous to integrate all of them into an automated medical diagnosis system. Clinical testing may be accomplished at a lower cost with the application of appropriate computer-based information and/or assistance with decisions resources. For automated systems to be deployed accurately and effectively, a comparison of the many methods available is required [5]. In this study, the variety of descriptive and predictive data mining techniques will be examined techniques recently proposed for heart disease identification.

The healthcare sector generates massive amounts of data. Hidden patterns in this data are found using data mining for clinical diagnosis. Therefore, research conducted over the last several decades has shown Data mining's significance in the healthcare sector [6-8]. Diabetes, high blood pressure, high cholesterol, and a fast heartbeat are some of the symptoms that need to be considered when predicting a cardiac condition. Completing the available medical data often affects the results of heart disease prediction. The medical sector depends heavily on machine learning. Expert systems have emerged for usage in medical settings because of recent developments in artificial intelligence. Additionally, during the last several decades, computational technologies have been developed to enhance physicians' experiences and decision-making skills [9-11]. Predict, identify, and diagnose applying machine learning to various kinds of diseases. Data mining and machine learning techniques have become favour for predicting the probability of developing particular diseases. Data mining is used to forecast diseases in the literature. Despite various studies' efforts to predict disease risk, no accurate results have been obtained. The objective is to accurately estimate heart disease risk. When it comes to forecasting cardiac disease, ensemble approaches have shown to be quite effective. Bagging,

boosting, and random subspace are three effective and popular examples of ensemble learning that are employed in this study to improve classifier performance. This work is motivated by the necessity for such an expert system. In this work, we want to examine the predictive power of several ensemble deep learning algorithms for heart disease constructed prediction models through a variety of techniques, such as VGG 16, Inspection V2, and Resnet 50, to accomplish this objective. Mutation albatross optimization was used to categorize and scale the dataset to enhance the models' convergence.

Section 2 reviews the latest methods for predicting heart disease data, and the following research is structured in the following method. section 3 procedure of the suggested approach. In section 4, an explanation and presentation of the results follows. Section 5 discusses the conclusion and next work.

LITERATURE REVIEW

In recent years, Data mining and machine learning have advanced significantly in the healthcare sector. This section examines some of the most current methods for applying machine learning to identify heart disease data.

Ali et al [12] implement feature fusion with ensemble deep learning for construct an intelligent medical system that forecasts cardiac problems. First, the characteristics that are taken out of sensor data are combined with electronic medical records using the feature fusion approach to provide useful healthcare information. Second, Through the elimination of redundant and unnecessary features and the selection of the most important ones, the information gain strategy lowers computing demand and enhances system performance. The conditional probability technique assigns a certain feature weight to each class, which significantly increases system efficiency. Following that, the predictions of heart disease is trained using the combined deep learning model. Conventional classifiers that use weighting, feature fusion, and feature selection are contrasted with the suggested approach using data from heart disease. The recommended approach has a 98.5% accuracy rate, greater than that of the current techniques. This result suggests that contrast to other existing techniques, our method is more successful in predicting heart disease.

Sonawane et al [13] introduced a multilayer perceptron neural network-based heart disease prediction system. In comparison to previous systems, the neural network in this system has the best accuracy of 98% when predicting when a patient has cardiac disease. It is trained using a back-propagation method and takes 13 clinical characteristics as input. This method is superior to other systems in terms of efficiency, as seen by the accuracy it achieves.

Sabarinathan et al [14] proposed a way to anticipate cardiac diseases and utilizes machine learning algorithms to provide pertinent information. In this study, the decision tree method is used to categorize the characteristics that provide structural information. Age, gender, chest discomfort, and attained heart rate are among the parameters utilized in categorization. The study suggests that cardiac disease may be caused by accurately predicted with 85% accuracy. Here, decision trees are also used for feature selection.

Anooj et al [15] a weighted fuzzy rule-based Clinical Decision Support System (CDSS) was developed to automatically extract information from patient medical records and detect heart disease. The proposed model functions in two stages: first, it generates weighted fuzzy rules, and second, it builds the fuzzy rule-based decision support framework. In the initial stage, weighted fuzzy rules are obtained through a data mining process involving attribute selection and assigning importance values to those attributes. These chosen features and their the final fuzzy-based decision system is then designed using appropriate weighted fuzzy rules. Finally, the suggested system is tested using datasets from the UCI repository, and its precision, specificity, and sensitivity are used to compare its efficacy with that of the neural network-based system.

Khazaee et al [16] suggested a novel method for classify two indicators of cardiac arrhythmia and three different categories of ECG beats: normal beats. This system's main components are an optimization module, a classifier module, and a feature extraction module. As the effective feature of the patterns, a suitable set that combines the form and the feature extraction module suggests temporal features. A multi-class support vector machine (SVM)-based classifier is suggested by the classifier module. The optimum subset of characteristics that feed the classifier upstream and the ideal value of the SVM parameters are identified by the optimization module using an optimization method using particle swarms. The results of the simulation demonstrate the excellent recognition accuracy of the suggested method. Only a few characteristics, selected with the aid of a particle swarm optimizer, are needed to attain this level of efficiency.

Chang et al [17] a machine learning algorithm-based AI heart disease diagnostic system was described. This study creates a Python-based healthcare research app because it is more reliable and allows to monitor and construct health monitoring applications. This data processing includes category variables and column conversions. Database gathering, logistic regression, and attribute assessment are the main application development steps. A random forest classification method is developed in order to identify cardiac abnormalities more precisely. This application's accuracy rating of around 83% over training data makes it noteworthy and necessitates data analysis. For research diagnosis, the random forest classifier approach provides more accuracy, is next discussed, along with the trials and results.

Nagavelli et al [18] described heart disease detection machine learning approaches. A weighted Naïve Bayes approach is used for predicting cardiac disease. The second approach functions automatically and applies information theory along with features from the time and frequency domains to analyze and determine ischemic heart disease. This technique chooses two high-performing classifiers for evaluation, such as combining SVM and XGBoost. The third approach utilizes an enhanced SVM model, developed using a dual optimization method, to automatically detect heart failure. A comprehensive Heart Disease

Prediction Model (HDPM) is then incorporated into a CDSS. This model integrates three main components: DBSCAN for detecting and eliminating outliers, a hybrid SMOTE-ENN method to manage class imbalance in the training data, and XGBoost for predicting heart conditions. Overall, machine learning presents promising applications in healthcare for diagnosing, identifying, and forecasting diseases.

Atallah et al [19] developed an ensemble approach using majority voting that may predict if a person may have heart disease. The forecast is based on straightforward, inexpensive medical examinations that may be carried out at any surrounding clinic. Furthermore, the model's improving efficiency is the objective. and trustworthiness of the doctor's diagnosis since it is developed using actual information from both healthy and unhealthy patients. Additional reliable results are obtained than if a single machine learning model were employed since the method employs the overall vote of many models to determine the patient. This method ultimately yielded a 90% accuracy rate using the hard voting ensembles model.

Jevin et al [20] suggested a different method and algorithm to manage the use of Association Rules for mining dispersed medical data sources at various regions (clinics and hospitals). These resources for medical data cannot be transferred to other network regions. Consequently, in order to align the data distribution across the network, the intended global computation has to be segmented into local computations. In each instance of the global computation, the ability to segment computations must be broad enough to accommodate various data distributions and participating nodes. The suggested solution uses agents to represent each dispersed data source. The agent then completes the computation of the global association rule by either traveling to each site and completing local tasks that may be completed there, or by exchanging some minimum summaries with other agents. To protect local data's confidentiality and security, the objective is to complete global tasks with the least amount of communication or movement by participating agents across the network.

Singh et al [21] to solve this problem, a revolutionary diagnostic system was developed, and it performs exceptionally well on training and testing data and shows exceptional intelligence. Algorithms for machine learning (ML) have shown encouraging possibilities for assisting medical practitioners make prompt and precise diagnoses. This study uses supervised machine learning techniques, including PCA, SVM, Random Forest (RF), and Decision Tree (DT). Using a bar plot, we compare their accuracy.

Kanwal et al [22] a model that was provided that applies many machine learning (ML) techniques to provide precise predictions for heart disease (HD). Accurate data for the training model is created through collecting data and pre-processing. Trained machine learning classifiers like Naïve Bayes (NB), logistic regression (LR), SVM, DT, and K Nearest Neighbour (KNN) may be used to identify heart disease. The Relief and LASSO feature selection methods for selecting the most crucial attributes. Various evaluating methods like, sensitivity, accuracy, specificity, MCC, confusion matrix and the precision is utilized to assess the efficacy of the approach. This research used supervised machine learning and feature selection in its comparative analysis. Decision tree gives highest accuracy of 85.21% with all features. On the other hand, with feature selection techniques SVM has an excellent performance.

The identification of medical conditions is seen is a substantial and complex work that must be carried out accurately and effectively. Automating the same would be very advantageous. Clinical evaluations often depend more on the physician's expertise and gut feeling than on the wealth of information hidden in the database. Nevertheless, the level of patient care is lowered by this approach's errors, unexpected biases, and excessive medical costs. The next part introduces the automated heart disease prediction system that uses parallel ensemble deep learning techniques.

PROPOSED METHODOLOGY

This work develops a model that can properly predict cardiac problems in order to lower the number of fatalities caused by these diseases. The five primary stages of this study are preprocessing, clustering, extraction of feature, feature selection, and classification. For the initial stage of pre-processing, the min-max normalization is used. The KMC technique is used for clustering with the goal of improving classifier accuracy. This work suggests using a PCA to efficiently extract features. After that, the feature selection process is finished using the modified whale optimization algorithm method. To choose the most pertinent and significant characteristics from the cardiovascular illness dataset, it computes the optimal fitness value. Finally, the classification is done using parallel ensemble deep learning model as VGG16, Inceptionv2, ResNet50 and applied hyperparameters tuning with the help of MAO.

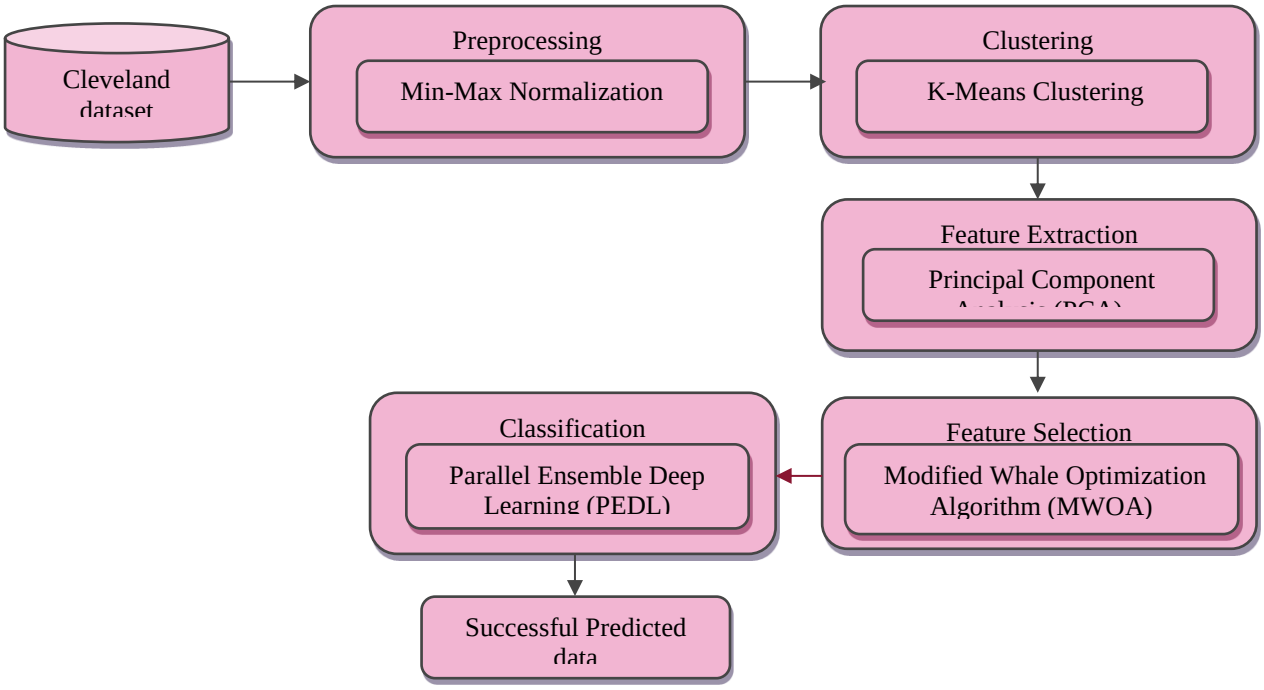

Figure 1. The methodology's suggested procedure

3.1 Data normalizations using min/max normalizations

Normalizing the data is necessary to prevent erroneous findings from being produced by variances in the input data. This study adopts the Min-Max Normalization paradigm, which involves employing a mathematical function to transform numerical values into a new range. The planned study's heart disease dataset is normalized using often used min-max normalizations. The dataset's values are normalized to fall between the minimum and maximum values specified, and each value is then changed using Equation (1)

$$v' = \frac{v - \min_A}{\max_A - \min_A}(\text{new\_max}_A - \text{new\_min}_A) + \text{new\_min}_A \quad (1)$$

Where,
A –Data Attributes,
Min (A), Max (A) –stands for A's min. and max. absolute values
v'- New values of inputs
v - Old values of inputs
New_ max (A), new_ min (A) imply max and min value ranges (boundaries of required ranges).

3.2 K-means Clustering

The most effective centroid is identified by iterating over the centroids using the K-means clustering approach. Presumably, it recognizes the quantity of clusters contain. This technique is also known as the flat clustering method. The number of clusters the algorithm finds in the data is represented by the letter "K" in K-means [23]. By grouping the data points together, this method reduces their total squared distances from the centroid. Remember that less variety in a cluster means more identical data points.

The steps mentioned below will assist us in comprehending the operation of the K-Means clustering approach-

- *Step 1:* The number of clusters, K, that this method must produce must first be specified.
- *Step 2:* Next, create clusters out of K data points that were selected at random. Set the data into groups proportionate to the number of accessible data points.
- *Step 3:* Finding the cluster centroids is the next stage.
- *Step 4:* The following steps are iterated through to assign data points to clusters that do not fluctuate to get the optimal centroid.
- 4.1 First, the centroids and data points would be compared to get the total squared distances.
- 4.2 The centroid, or cluster nearest to the others, must now receive each data point.
- 4.3 Average all of the cluster data points to get the centroids for the clusters.

The problem is solved using K-means using expected maximum. Allocate data points use expectation to find the nearest cluster, then Maximum to find each cluster's centroid. In the E-step, data points are assigned to the nearest cluster. M-steps determine cluster centroid. The mathematical solution is broken down below; to skip it.
The function that the objective is

$$J = \sum_{i=1}^{m}\sum_{k=1}^{K} w_{ik}\|x^i - \mu_k\|^2 \quad (2)$$

where $w_{ik} = 1$ for data point xi if it belongs to cluster k; otherwise, $w_{ik} = 0$. Also, μ_k is the centroid of x^i's cluster.

3.3 Dimensionality Reduction using PCA

On normalizing data their dimensionalities are reduced. PCA is being used in this effort to reduce dimensionality. Principal component analysis reduces the dimension and generates additional features by combining the original features in a linear fashion. Instances of the dataset are mapped via PCA from d-dimensional spaces to k-dimensional subspaces where k<d. The Principal Components (PC) are the collection of k newly created dimensions. Each PC aims to maximize variance while avoiding variation that has already been taken into account by all of its previous components [24]. Therefore, each subsequent component covers less variation than the first, which covers the most variation. Computation of principal components:

$$PC_i = a_1X_1 + a_2X_2 + \cdots + a_dX_d \quad (3)$$

Where PC_i – Principal Component 'i'; X_j – original feature 'j'; a_j – numerical coefficient for X_j.

3.4 Implementing a modified whale optimization method for feature selection

The WOA technique was inspired by humpback whale hunting strategies and the creation of an equation for such strategies. The whales hunt by circling and consuming their prey, which is mostly small fish, using a bubble-net technique. Whales emerge and start to surface, creating an enormous ring of bubbles deep under the fish. The bubbles serve as a trap, forcing the fish to the surface. Fish that are rising to the surface are being pursued by whales [25]. In theory, hunting consists of three stages: circling, exploitations, and explorations, where the grade of fish consumed is determined by the explorations. Phase of encirclement: The fish are found and surrounded by whales. The starting point for the optimal site is originally arbitrary and indeterminate. In reaction to the random initiations, other agents adjust their positions, and these updated positions develop into the most effective positions for achieving targets.

In animal swarms, individuals may greatly improve their personal attributes by learning from their neighbours. To generate whales' adaptable neighbourhoods and strengthen ties between groups, the adaptive social learning approach creates neighbourhoods by analysing social ranking, social effects, and social networks. To improve population variety while maintaining computation accuracy, a novel strategy that is based on neighbourhood updates is developed.

The position of the whale and its surroundings may be shown using equations (4) and (5).

$$\vec{H} = |\vec{C}x\vec{X} * (t) - \vec{X}(t)| \quad (4)$$

$$\vec{X}(t+1) = \vec{X}(t) - \vec{A} \times \vec{H} \quad (5)$$

Equations (6) and (7) provide vector coefficients denoted by $\vec{A}$ and $\vec{C}$, t denotes the current iteration, positional vectors are provided by phrase $\vec{X}$, and arbitrary replies started randomly are denoted by $\vec{X} *$.

$$\vec{A} = 2\vec{a} \times \vec{r} - \vec{a} \quad (6)$$

$$\vec{C} = 2x\vec{r} \quad (7)$$

A random integer between 0 and 1 is represented by the linear reduction of $\vec{a}$'s elements from 2 to 0 at each iteration.

Humpback whales employ bubble-net hunting strategies, which include circling their target. To choose the best course of action, whales surround their prey, which includes fish, and then adjust their placements. Equation (8) and equation (9) show the main mathematical component of the WOA.

$$X(t+1) = X^*(t) - A.|C.X^*(t) - X(t)| \; if \; p < 0.5 \quad (8)$$

$$X(t+1)|C.X^*(t) - X(t)|.e^{bl} \cos(2\pi t) + X^*(t) \; if \; p \geq 0.5 \quad (9)$$

In this case, X stands for whale position vectors, t for time or iteration indices, and X* for the ideal solutions that were produced; A=2a. (r-a); C=2.r; r denotes random vectors with values between 0 and 1; b denotes constant values dependent on certain pathways; in this work, its value is 1; As iterations continue, parameter vectors reduce linearly from 2 to 0; The random number, l, falls between -1 and 1. Given that the probabilities in Eqs. (10) and (11) are 50% and 50%, whales have an equal probability of randomly selecting either route during the optimization stage. To update the whales' locations, p, a random variable between 0 and 1, is used to alternate between (8) and (9). If vector A is in the bubble-net stage, its random value is [-1, 1], but during the searching stage, it may be more or less than 1. Equation (10) is an example of the searching technique.

$$X(t+1) = X_{rand} - A.|C.X_{rand} - X(t)| \quad (10)$$

This random search algorithm, which gives priority to the search when |A| is greater than one, causes the WOA method to do a worldwide search.

3.5 Classification using Parallel Ensemble Deep Learning

Deep learning and ensemble learning are two methods in machine learning that perform more than conventional algorithms. The former describes techniques that combine many base models into a single framework to generate a more powerful model that performs better than the basis models. An ensemble method's effectiveness depends on many factors, such as the training and combination of the baseline models [26]. There are standard methods for creating an ensemble model that have been effectively used in a variety of domains in the literature. In contrast, deep learning-based methods improve machine learning forecasts in several fields. Adjusting appropriate hyper-parameters is the most significant challenge in deep learning, which makes it a laborious and time-consuming process. This is true even though there are many different deep learning architectures, and they can handle complicated issues and extract features automatically. A lot of recent research attempts to use ensemble learning to approach deep learning to get around this challenge. Additionally, it provides a thorough explanation of the several aspects or elements that affect the effectiveness of ensemble approaches. Furthermore, a number of research projects that used ensemble learning across a variety of fields are presented and appropriately classified. It calculates an optimal fitness value from the cardiovascular diseases dataset to decide which aspects are most crucial and essential. Finally, the classification is done using parallel ensemble deep learning model as VGG16, Inceptionv2, ResNet50 and applied hyperparameters tuning with the help of Mutation Albatross Optimization Algorithm (MAO).

3.5.1. VGG 16

The VGG architecture, commonly known as VGGNet, is a conventional neural network (CNN). To improve the efficiency of the model and deepen these CNNs, VGG was created. A common multi-layer deep CNN architecture is the Visual Geometry Group, or VGG [27]. When there are 16 or 19 convolutional layers in a VGG-16 or 15, it is not considered "deep". According to Figure 2, VGG architecture creates advances object recognition models. On several tasks and datasets, VGGNet beats ImageNet.

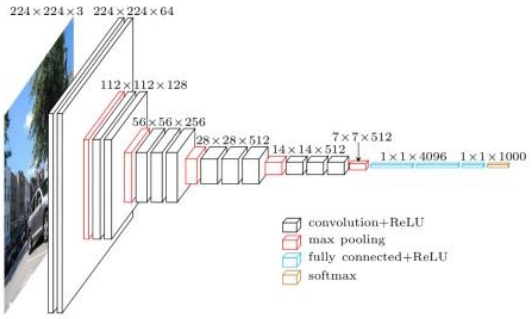

Figure 2. Architecture of VGGnet

a) VGG Architecture

VGGNets are built on the basic characteristics of convolutional neural networks (CNNs). Using small multilayer filters, the VGG network is developed. There are three fully connected layers and thirteen convolutional layers in the VGG-16. Let's rapidly go over the VGG design:

Input: The VGGNet receives the 224 × 224 image. The model's creators cut the 224x224 region in the middle of each image to ensure that ImageNet's input size remains constant.

Convolutional Layers: To identify left/right and up/down, the smallest receptive field (3x3) is used by VGG convolutional layers. The input is linearly altered using 1x1 convolution filters. ReLU units, substantial AlexNet improvements that reduce training time, follow. ReLU produces positive inputs and 0 otherwise. Setting the combined stride at 1 pixel maintains convolutional spatial resolution.

Hidden Layers: Each VGG hidden layer uses ReLU. Training time and memory consumption are increased by Local Response Normalization (LRN), hence VGG rarely uses it. Additionally, performance is not improved by this method.

Fully-Connected Layers: Layers in VGGNets connect in three ways. 4096 channels compose the first two levels, while 1000 channels per class compose the third layer.

b) VGG 16

VGG16 is a 16-layer convolutional neural network model based on VGGNet. The top-5 test accuracy for the VGG16 model in ImageNet is around 92.7%. Among the datasets is ImageNet. In contrast to AlexNet, it makes major advances by gradually replacing many 3x3 kernel-sized filters in addition to the massive kernel-sized filters. VGG16 was trained for weeks using Nvidia Titan Black GPUs. Images may be classified into 1000 categories using the 16-layer VGGNet-16, including keyboard, mouse, pencil, and animal. The 224-by-224 image input is used by the model.

VGGnet is a 16-layer deep neural network. With more than 138 million parameters, VGG16 is an enormous network. It's a large network even nowadays. However, as figure 4 illustrates, the network's attractiveness stems from the simplicity of the VGGNet16 design. Its architecture alone suggests that it is rather homogeneous. A pooling layer reduces the width and height after a few convolution layers. Take a look at how many filters are used: there are roughly 64 available, would gradually double to around 128 and 256. Use filters with a value of 512 for last layers.

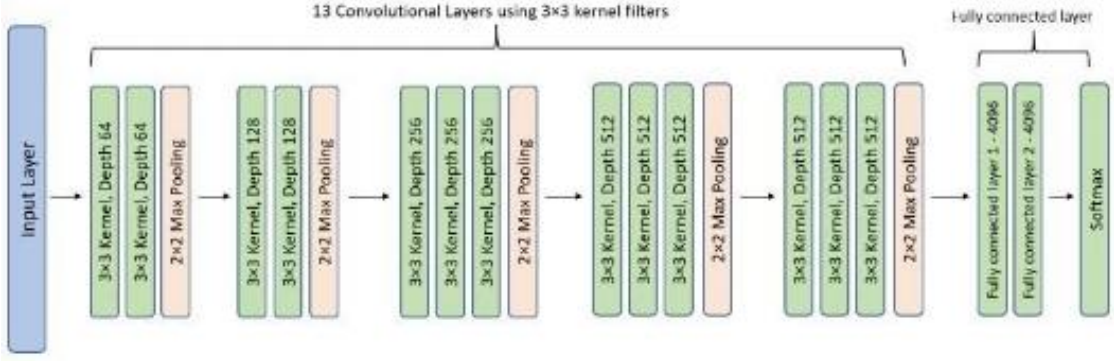

Figure 3. Architecture of VGG16

The above discussion emphasized multiple limitations of VGG16 when used for feature extraction in image datasets. VGG16 was initially created to process large-scale data, capable of identifying 1,000 classes from a dataset containing over a million images. Consequently, utilizing it on smaller datasets with limited training samples often results in severe under fitting. Identifying whether a disease exists or not from images is often challenging because the visual patterns are highly similar, making it hard for models to extract discriminative features. As a result, using deeper layers for training frequently leads to misclassification. This issue can be minimized by applying regularization within the VGG16 layers. In this study, a Regularization-based Adaptive factor (RA) is proposed as a solution to the under fitting problem.

3.5.2. Inspection v2

Inception v2 represents the improved version of the Inception CNN architecture, where batch normalization is a key component. Because batch normalization enhances training performance, this version eliminates techniques like dropout and local response normalization. Unlike Inception v1, which occasionally applies larger filters such as 5×5 convolutions that significantly reduce input dimensions, Inception v2 avoids this issue. The use of such larger filters in the earlier model can lead to a drop in classification accuracy [28]. This happens because a sharp reduction in input dimensions can cause the network to lose important information. Additionally, using larger filters such as 5×5 increases computational load compared to employing multiple smaller 3×3 filters. In terms of factorization, a 3×3 convolution can be replaced with a sequence of asymmetric filters first a 1×3 filter and then a 3×1 filter. This approach preserves the same receptive field as a standard 3×3 kernel but reduces the computational cost by about one-third. This factorization technique is ineffective in the initial layers where the input resolution is large; it becomes beneficial only when the input size is within the range of approximately 12 to 20 pixels. Inception V1 also introduced an auxiliary classifier, which helps the network converge more efficiently. This auxiliary branch is intended to lessen the impact of the vanishing gradient problem by directing useful gradients toward earlier layers, thereby reducing overall loss during training.

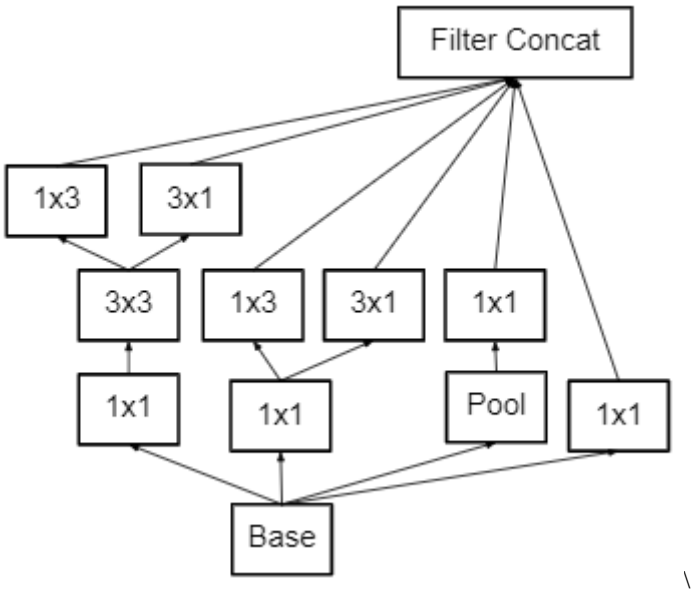

Figure.4. Inspection V2 architecture

In the inception V2 design (illustrated in Figure 5), the 5×5 convolution block is substituted with two consecutive 3×3 convolutions. This modification not only reduces the computation cost but also speeds up processing, as a single 5×5 filter requires roughly 2.78 times more computation than a 3×3 filter. Therefore, replacing it with two 3×3 layers enhances the overall efficiency of the model. Additionally, this version applies factorization by breaking n×n convolutions into sequential 1×n and n×1 operations. As was previously mentioned, a 3x3 convolution may be decomposed into a 1x3 and subsequently a 3x1 convolution, which has a computational cost that is 33% lower than that of a 3x3 convolution. Instead of making the module deeper, the representational bottleneck problem was resolved by increasing the feature banks. This would prevent the information loss that occurs as it gets deeper.

3.5.3. Resnet 50

Increasing depth is difficult because multiplications may make the gradient transmitted back through the network unstable, resulting in unreasonably large or small values. A frequent issue that arises is gradients dispersion. The usage of Xaiver initialization, normalization of batches and switching to ReLU as the activation function are some of the techniques that have been found to prevent gradient dissipation. The dispersion of the gradient has been successfully addressed [29]. Degradation, or a decline in network performance as depth grows, there exists an additional issue about network deepening. The efficiency of the model is dependent on the network's depth, as shown by experience. In concept, deeper models should provide better results since as the number of network layers increases, the network may be able to extract more intricate feature patterns. The experiment indicated profound network degradation. As network depth increases, accuracy saturates or decreases. Additionally, training set accuracy is decreasing. Confirm this is not overfitting. The training set should be precise since overfitting is predicted. If this is addressed, network depth increases by a number of orders. ResNet's residual network fixes this. In the end, y = F(x)+x. The two mapping options ResNet offered were identity mapping, and residual mapping, the area outside the "curved curve" in Figure 6. The residual mapping is the "difference", whereas the identity mapping is x or y - x, resulting in F(x). For classification tasks, ResNet-50 used convolution to the input first, then four residual blocks and complete connection. The ResNet-50 network structure, shown in Figure 7, consists of 50 Conv2D processes.

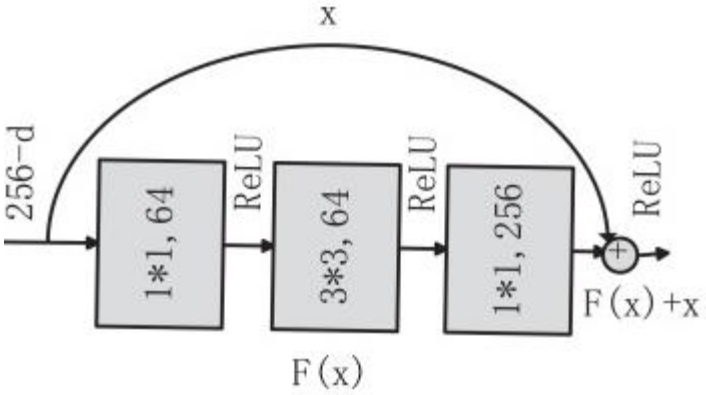

Fig. 5. Residual block.

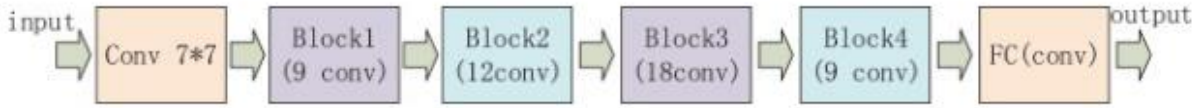

Fig. 6. ResNet-50.

At CNN's final stage, the fully connected (FC) layer summarizes the preceding levels. Considering the local amplification, local feature extraction, and feature engineering processes of the preceding convolution and pooling, the subsequent FC layer might be seen as feature weighting. The FC layer's structure enables efficient collection of the complex convolutional layer-generated nonlinear combinations of features. The FC layer will learn an artificial nonlinear function. This is the fundamental learning strategy. After formatting for the multiple-layer perceptron, the feed forward neural network receives the image after it has been flattened into column vectors. After then, the flattened data is used in every training session. This approach allows the model to categorize the picture using Softmax and other classification algorithms while differentiating between its primary characteristics and certain low-level features.

3.5.4. Parameter optimization Mutation Albatross Optimization Algorithm (HMAOA)

Long-duration flights are made possible by the intriguing flying technique known as "dynamic soaring," which uses ambient energy. The energy needed to fly for that long is provided by the wind near the surface [30]. In mountainous or above-sea regions, the phrase "wind shear/gradient" describes the notable change in wind speed with height. Energy is captured from the geographical distribution of wind speed by flying regularly over the wind gradient zone. Among other soaring species, this has been seen in albatrosses and eagles. In actuality, soaring birds are able to cover long distances almost without flapping their wings because the dynamic soaring method drastically lowers energy consumption. See figure 7 for an example of the four distinct flying phases that make up the dynamic soaring manoeuvre: Windward ascension, high altitude turns, leeward descent, and low altitude turn are the first four options. For potential energy, the bird exchanges kinetic energy. as it climbs higher and enters the headwind to perform this move. The bird can maintain height increases for a certain amount of time because to the lift that the wind gives it. After that, the bird turns abruptly and descends with a tailwind. It keeps going downward, exchanging kinetic energy for potential energy, which is then converted into velocity increases. The bird is the low-altitude turn because to its enhanced velocity, or momentum and starts a new dynamic soaring cycle at a low altitude point. Energy-neutral, or near-neutral, manoeuvres are ideal for the dynamic soaring cycle.

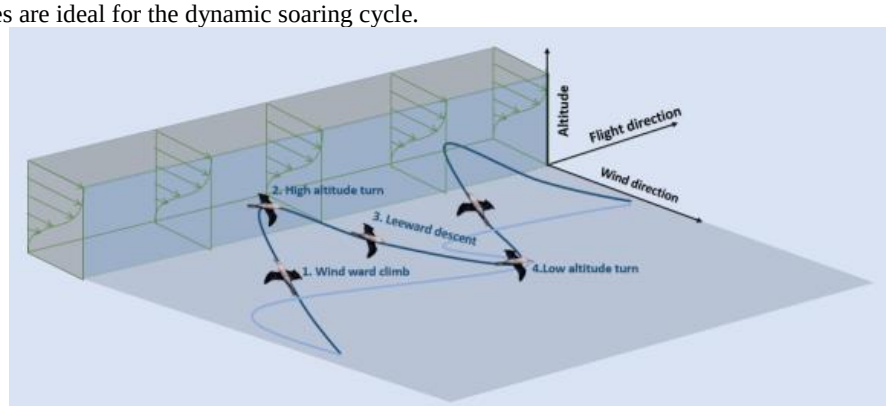

Figure 7. Albatrosses conduct dynamic soaring maneuvers

The dynamic soaring phenomena makes it abundantly evident that albatrosses may traverse hundreds of kilometers with little energy expenditure. This implies that they must have an energy-sourcing mechanism in place. This may be done by selecting a flight path that allows for the gathering of wind energy, as shown experimentally. The fact that they locate such way on their own, in real time, is equally noteworthy. Assuming there is enough wind, a natural way to accomplish this type of manoeuvre is to alter or disrupt possible actions (controls) on a regular basis, such rolling and/or pitching. After that, the bird senses its energy level and decides whether to adjust its course to consume less energy by using the related changes in wind, velocity, and

height. [31]. Keep in mind that the albatross's nostrils serve as an airspeed sensor. The procedure is then carried out independently until the course or trajectory is recognized and tracked, after which the actions are modified appropriately. The periodic perturbations are easy to repeat and memorize. In extremum searching, bird pitching and rolling disruption modulates. Finding that the objective function does not need a mathematical representation for extremum seeking systems is like experiencing wind, velocity, and height. Finally, the demodulation and parameter update stage is comparable to modifying the action based on the energy assessment data. As a result, Extremum seeking systems may provide a complete and organic modelling and dynamic soaring's control system foundation.

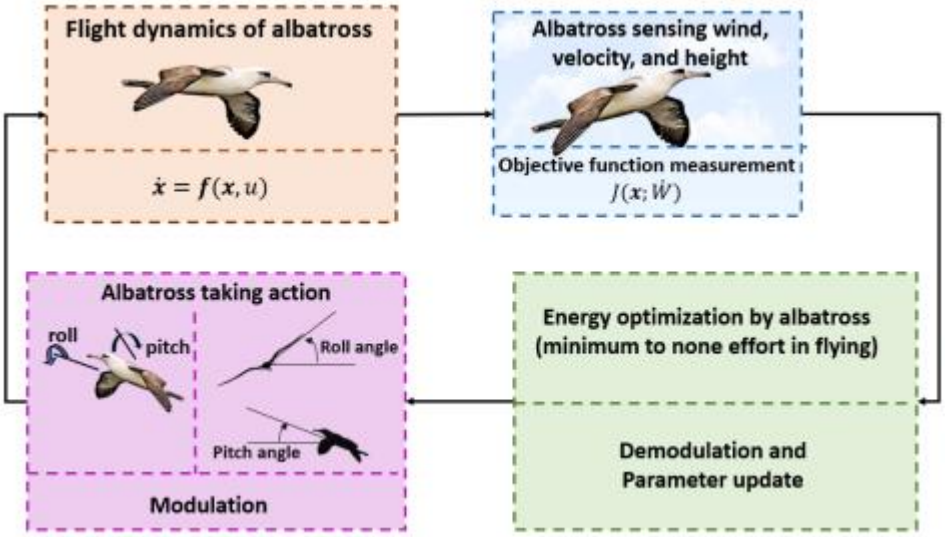

Figure 8. Extremum searching and dynamic soaring systems are parallel

The dynamic soaring system typically uses the bird/mimicking-object to control the movements u(t) = [CL, φ], which, subsequently stand for rolling and pitching movements. To perform the move, CL might be constant or free (changing). A single-input single-output (SISO) system for the dynamic soaring solution with constant CL may be constructed by using the system output is the desired function J, while the free input is the rolling φ. As a result, we may take use of the advanced and established techniques available in the area of extremum searching for extremum seeking controllers with a single input and a single output. Furthermore, we believe the first results from this issue should be on a single-input and single-output basis before adding more variables since the use of extremum searching is novel and unexplored in connection with the dynamically increasing challenge. To propose it as a special paradigm for dynamic soaring modelling and control, we will now modify that structure. First, the flight dynamics model of albatrosses and imitating objects may be maintained. In addition to figure 8. Rolling movement φ is the only input. The target function J is defined as g(x, W˙), from the wind shear/gradient state variables and the flight dynamic system. Maximize or minimize the objective function J with θ = φ. The roller action/control input maximizes or decreases the system's objective function. This allows the albatross/mimicking object system to be addressed by the extremum seeking system toward the extremum (maximum/minimum) of its intended function by using the best rolling action/control. In terms of form, the high pass filter is s/(s + ωh) while the low pass filter is wl/(s + ωl). A sin(ωt) is the modulating signal, and b sin(ωt – φ phase) is the demodulating signal. The frequencies given by ω, ωh, and ωl stand for the high pass filter, low pass filter, and input/modulating signal, in that combination. The φ phase also indicates the phase difference between the modulating and demodulating signals.

3.5.5. Mutation Albatross Optimization Algorithm (MAOA)

An iterative search for disease data is conducted after initialization, and if the best diagnosis is found, the procedure is terminated. The approach defines the objective function in the first phase of execution using equation (12) and its solution space. The MAOA algorithm produces an initial spider population depending on the number of SN for the best selection of disease data after obtaining the parameters of the proposed model. The spiders' SN information, which includes the total amount of spiders, is maintained in a fixed-size memory remains constant during the MAOA simulation. The web and fitness values of the spiders (samples) are determined using the final weight from equation (12) and recorded after their placements are randomly created in the disease selection space. After the installation procedure is complete, the algorithm begins the iteration phase, which uses the constructed spider positions to perform in the CH selection search space.

Estimate the joining weight via the equation (8),

$$We_i = we_1\left(\frac{R_e}{En_{initial}}\right) * we_2\left(\frac{Deg_i}{Max(Deg_i)}\right) * we_3\left(\frac{En_{TX(k,d)}}{En_{tx}}\right) * we_4\left(\frac{dis_{mini}}{dis_{maxi}}\right) * we_5\left(\frac{Mobility_i}{max(Mobility)}\right) \quad (11)$$

we_1, we_2, we_3, we_4 and we_5 =1. Thus,we_1, we_2, we_3, we_4 where a similar strategic approach is used to weight the indications. The node is selected as a CH and given the greatest weight for that specific circular region. The optimum cluster heads, node degree, transmission power, residual energy, and the MAOA algorithm have been selected.

a) Mutation Rate (MR) for initial population

For x_i, compute the opposing solution and its fitness function fit_{xi} the AOA technique is used in this phase to enhance the original population X's solution. The two populations, *X* and *Xi*, are then combined using the suggested technique to create a new population. This new population's best of N solutions are selected using both populations' fitness function values, which account

for the initial population X's upgrade.

The radius r_m, also known as the mating radius, determines if they are next to each other, males and females in charge of the community are mating:

$$r_m = \frac{\sum_{j=1}^{n}(x_j^{high} - x_j^{low})}{2n} \quad (12)$$

In the case when there is only one female and several males in the neighbourhood, to replace the parents with the least fitness function, to determine which parents will have offspring, a roulette wheel system is used. This describes the stages of the proposed mutation rate-based AOA algorithm part and depend on enhancing the AO algorithm's performance through the use of the mutation-based learning methodology. Finding the optimal option for the cluster head selection is the objective of this development.

b) Updated stage

To generate a new population, from the union of two populations, the best N solutions are selected; x_{best}, the population's best suggestions is chosen. The selection method provides the best solution that captures the disease's data subset, which is repeated until the stopping requirements are reached.

RESULTS AND DISCUSSION

This section provides a detailed presentation of the suggested model's experimental results. The suggested idea is implemented using Mat Lab. The new EDL-MWOA model is compared with the existing HFCNN, CNN and Swarm-ANN models for the Cleveland database of the UCI data repository in terms of precision, recall, accuracy, and f-measure. At https://archive.ics.uci.edu/ml/machine-learning-databases/heartdisease/, can access the Cleveland database. The term "multivariate" describes this kind of dataset, which are numerical data analyses that provide or incorporate wide ranges of mathematical or statistical variables. Fourteen characteristics make up this composite: exercise-induced angina, resting blood pressure, serum cholesterol, fasting blood sugar, high heart rate, thalassemia, main vessel counts, and comparison of rest slopes of peak exercise ST segments to previous peak ST depressions caused by activities. Cleveland's database contains 920 incidences. Proposed model performance measures are compared using these metrics.

The accuracy is the ratio of accurately collected optimistic discoveries to all predicted positive annotations.

$$\text{Precision} = TP/TP+FP \quad (13)$$

The ratio of accurately recognized positive findings compared to every finding within the actual group is referred to as sensitivity.

$$\text{Recall} = TP/TP+FN \quad (14)$$

F1 is the accuracy and recall weighted average. Negative and false positive results are acceptable.

$$\text{F1 score} = 2*(\text{Recall} * \text{Precision})/ (\text{Recall} + \text{Precision}) \quad (15)$$

Positives and negatives' accuracy is determined using the following formula:

$$\text{Accuracy} = (TP+FP)/(TP+TN+FP+FN) \quad (16)$$

The terms TP, FP, TN, and FN stand for True Positive, False Positive, True Negative, and False Negative, respectively.

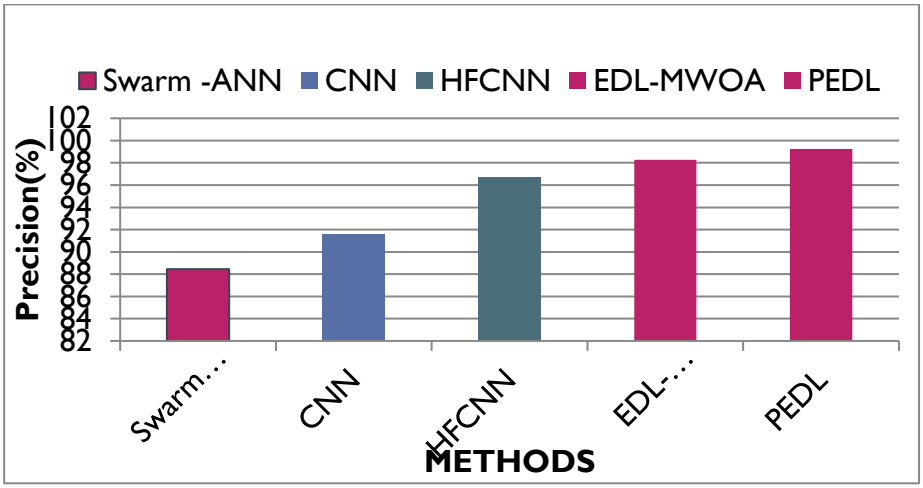

Figure 9. precision Comparison between the Suggested and Current Approaches

According to the findings, if the classifier model receives sufficient data, the recommended model may correctly predict heart disease. By employing improved feature extraction approaches, the suggested model performs better. These outcomes also show that the suggested model is capable of outperforming other models. Figure 9 shows that the metrics for the Swarm-ANN and CNN approaches are 88.45% and 91.57%, respectively, the HF-CNN technique has 96.7%, EDL-MWOA method has 98.26%,

and the proposed PEDL method has an accuracy rate of 99.24%.

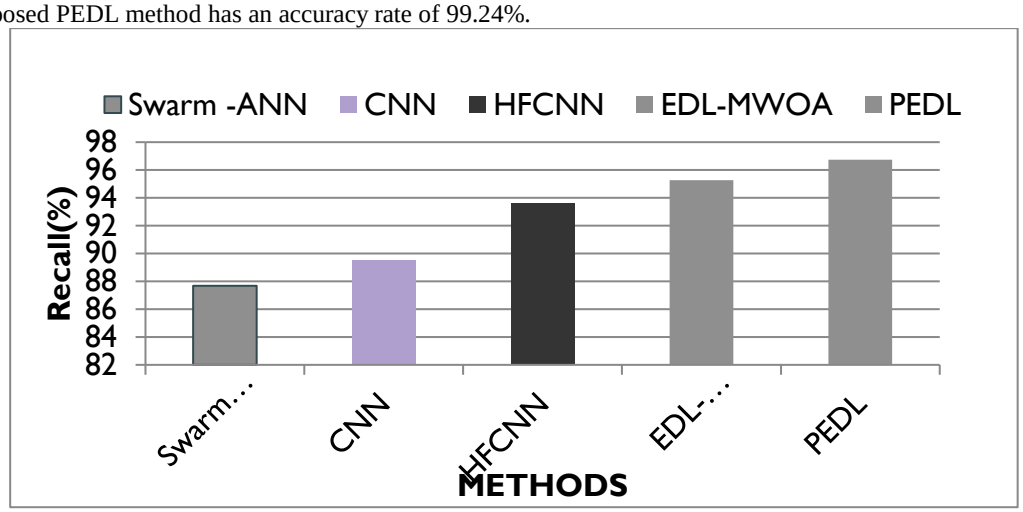

Figure 10. Recall comparison between the proposed and current approaches.

The proposed PEDL performs better in recall than earlier methods, as shown in Figure 10. With a recall of 96.74% 95.27%, the proposed approach is more effective than the Swarm-ANN method (89.54%), CNN method (87.68%), and HFCNN methodology (93.57%). Recall increases rapidly in the beginning of training, and using the recommended MWOA based feature selection, it is obvious that this reduces the distance between the points, stabilizing the output.

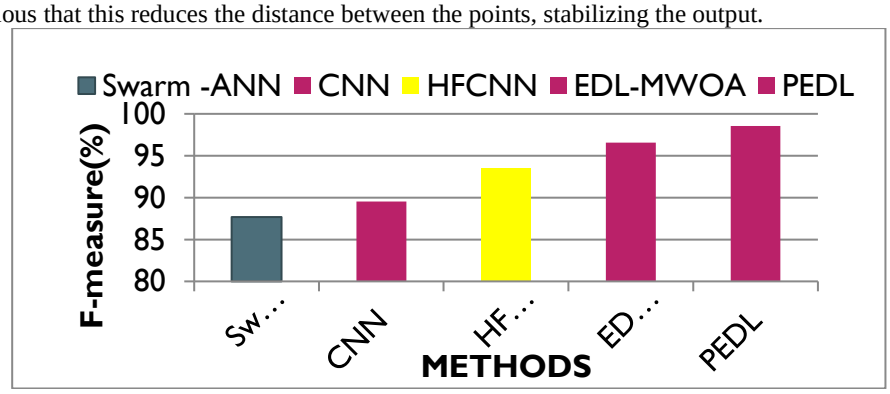

Figure 11. F-value review between the suggested and current approaches

The suggested PEDL performs more effectively than figure 11 shows previous approaches in terms of F-measure. By combining the MWOA technique with the feature selection strategy, the suggested system's performance was noticeably enhanced. After selecting the characteristics, used both general and particular feature selection methods to assess their relevance.

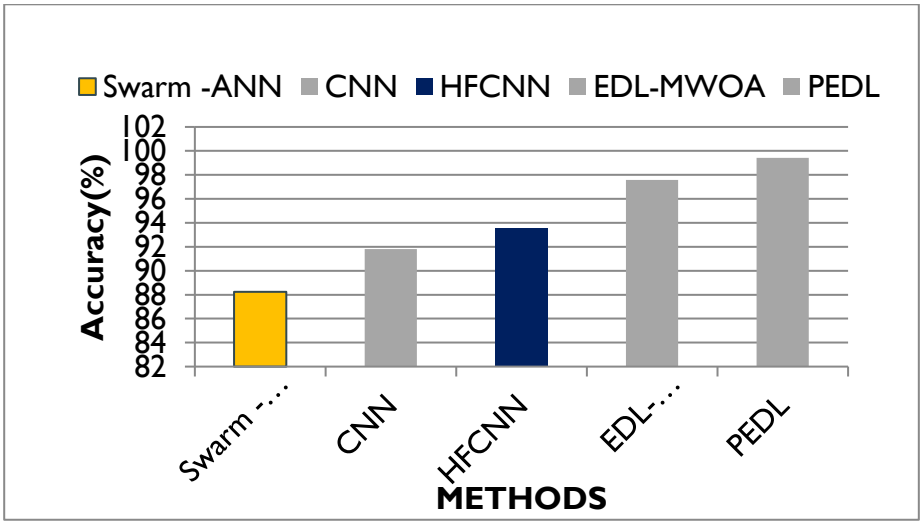

Figure 12. Evaluation of the Suggested and Current Methods' Accuracy

The suggested PEDL model has the best accuracy (99.84%), as seen in Figure 12. This classifier has better accuracy than others. Based to the results, if the classifier model is given enough information, the suggested model may be able to predict heart disease with consistency. These findings also show that the model that was suggested outperforms existing models in data categorization with more characteristics. All baseline models' accuracy improved significantly after utilizing the suggested method for feature

extraction and selection.

CONCLUSION

Three data mining classification approaches are used to develop a heart disease prediction system. From a database containing previous heart disease instances, the algorithm retrieves concealed data. The five primary stages of this study are preprocessing, clustering, extraction of feature, feature selection, and classification. Based on data exploration and business information, five mining objectives are established. The trained models are used to assess the objectives. Every model has unique accuracy-related strengths, accessibility to extensive information, and simplicity of model interpretation, and they could all respond to complex inquiries. Only parallel ensemble deep learning can determine the influence and correlation between the medical characteristics and heart disease. The suggested EDL-MWOA model has the best accuracy (99.84%). This classifier has better accuracy than others. The obtained findings demonstrate that the suggested model may reliably predict cardiac disease if the classifier model is given enough information. These findings also show that the model that was suggested outperforms existing models in data categorization with more characteristics. Additionally, this prediction system may include other data mining methods, such as suggestion Rules and Time Series. It is also possible to utilize continuous data rather of just categorical data.

REFERENCES

1. Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. In *IOP conference series: materials science and engineering* (Vol. 1022, No. 1, p. 012072). IOP Publishing.
2. Shah, D., Patel, S., & Bharti, S. K. (2020). Heart disease prediction using machine learning techniques. *SN Computer Science*, *1*(6), 345.
3. Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*, *7*, 81542-81554.
4. Marimuthu, M., Abinaya, M., Hariesh, K. S., Madhankumar, K., & Pavithra, V. (2018). A review on heart disease prediction using machine learning and data analytics approach. *International Journal of Computer Applications*, *181*(18), 20-25.
5. Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective heart disease prediction using machine learning techniques. *Algorithms*, *16*(2), 88.
6. Hossain, M. I., Maruf, M. H., Khan, M. A. R., Prity, F. S., Fatema, S., Ejaz, M. S., & Khan, M. A. S. (2023). Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison. *Iran Journal of Computer Science*, *6*(4), 397-417.
7. Ayon, S. I., Islam, M. M., & Hossain, M. R. (2022). Coronary artery heart disease prediction: a comparative study of computational intelligence techniques. *IETE Journal of Research*, *68*(4), 2488-2507.
8. Assegie, T. A., Rangarajan, P. K., Kumar, N. K., & Vigneswari, D. (2022). An empirical study on machine learning algorithms for heart disease prediction. *Int J Artif Intell ISSN*, *2252*(8938), 1067.
9. Sharma, V., Yadav, S., & Gupta, M. (2020, December). Heart disease prediction using machine learning techniques. In *2020 2nd international conference on advances in computing, communication control and networking (ICACCCN)* (pp. 177-181). IEEE.
10. Raja, M. S., Anurag, M., Reddy, C. P., & Sirisala, N. R. (2021, January). Machine learning based heart disease prediction system. In *2021 International conference on computer communication and informatics (ICCCI)* (pp. 1-5). IEEE.
11. Ansari, G. A., Bhat, S. S., Ansari, M. D., Ahmad, S., Nazeer, J., & Eljialy, A. E. M. (2023). Performance evaluation of machine learning techniques (MLT) for heart disease prediction. *Computational and Mathematical Methods in Medicine*, *2023*(1), 8191261.
12. Ali, F., El-Sappagh, S., Islam, S. R., Kwak, D., Ali, A., Imran, M., & Kwak, K. S. (2020). A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. *Information Fusion*, *63*, 208-222.
13. Sonawane, J. S., & Patil, D. R. (2014, February). Prediction of heart disease using multilayer perceptron neural network. In *International conference on information communication and embedded systems (ICICES2014)* (pp. 1-6). IEEE.
14. Sabarinathan, V., & Sugumaran, V. (2014). Diagnosis of heart disease using decision tree. *International Journal of Research in Computer Applications & Information Technology*, *2*(6), 74-79.
15. Anooj, P. K. (2012). Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. *Journal of King Saud University-Computer and Information Sciences*, *24*(1), 27-40.
16. Khazaee, A. (2013). Heart beat classification using particle swarm optimization. *International Journal of Intelligent Systems and Applications*, *5*(6), 25.
17. Chang, V., Bhavani, V. R., Xu, A. Q., & Hossain, M. A. (2022). An artificial intelligence model for heart disease detection using machine learning algorithms. *Healthcare Analytics*, *2*, 100016.
18. Nagavelli, U., Samanta, D., & Chakraborty, P. (2022). Machine learning technology-based heart disease detection models. *Journal of Healthcare Engineering*, *2022*(1), 7351061.
19. Atallah, R., & Al-Mousa, A. (2019, October). Heart disease detection using machine learning majority voting ensemble method. In *2019 2nd international conference on new trends in computing sciences (ictcs)* (pp. 1-6). IEEE.
20. Jevin, M. J., Jayant, H., Sanjay, R., Hemasai, V., & Venkatasrinivas, P. V. (2023). Heart disease identification method using machine learning classification in e-healthcare. *Heart Disease*, *10*(3).
21. Singh, A., Mahapatra, H., Biswal, A. K., Mahapatra, M., Singh, D., & Samantaray, M. (2024). Heart disease detection using machine learning models. *Procedia Computer Science*, *235*, 937-947.
22. Kanwal, A., Ahmad, K. T., & Aslam, N. (2022). Detection of heart disease using supervised machine learning. *VFAST Transactions on Software Engineering*, *10*(3), 58-70.

23. Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE access*, *8*, 80716-80727.
24. Richardson, M. (2009). Principal component analysis. *URL: http://people. maths. ox. ac. uk/richardsonm/SignalProcPCA. pdf (last access: 3.5. 2013). Aleš Hladnik Dr., Ass. Prof., Chair of Information and Graphic Arts Technology, Faculty of Natural Sciences and Engineering, University of Ljubljana, Slovenia ales. hladnik@ ntf. uni-lj. si*, *6*, 16.
25. Mirjalili, S., & Lewis, A. (2016). The whale optimization algorithm. *Advances in engineering software*, *95*, 51-67.
26. Vo, H. V., Du, H. P., & Nguyen, H. N. (2024). APELID: Enhancing real-time intrusion detection with augmented WGAN and parallel ensemble learning. *Computers & Security*, *136*, 103567.
27. Yang, H., Ni, J., Gao, J., Han, Z., & Luan, T. (2021). A novel method for peanut variety identification and classification by Improved VGG16. *Scientific Reports*, *11*(1), 15756.
28. Iqbal, H., Chawla, H., Varma, A., Brouns, T., Badar, A., Arani, E., & Zonooz, B. (2022). AI-driven road maintenance inspection v2: reducing data dependency & quantifying road damage. *arXiv preprint arXiv:2210.03570*.
29. Koonce, B. (2021). ResNet 50. In *Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization* (pp. 63-72). Berkeley, CA: Apress.
30. Krishnan, K., Subramaniasivam, A., Ravichandran, K., & Subramanyam, N. (2021, March). Albatross optimization algorithm: a novel nature inspired search algorithm. In *International conference on emerging trends and technologies on intelligent systems* (pp. 203-216). Singapore: Springer Singapore.
31. Richardson, P. L., Wakefield, E. D., & Phillips, R. A. (2018). Flight speed and performance of the wandering albatross with respect to wind. *Movement ecology*, *6*(1), 3.