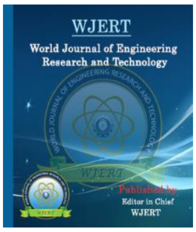
WJERT, 2025, Vol. 11 Issue 12, 286-307. **Original Article** **ISSN 2454-695X**

World Journal of Engineering Research and Technology

www.wjert.org

Impact Factor: 7.029 **Coden USA: WJERA4**



A COMPREHENSIVE ANALYTICAL METHOD FOR PREDICTING DISEASES THROUGH MACHINE LEARNING APPROACHES

Shaikh Abdul Hannan*

Assistant Professor, Faculty of Computing and Information, AlBaha University, AlBaha, Kingdom of Saudi Arabia.

Article Received on 21/10/2025 **Article Revised on 11/11/2025** **Article Published on 01/12/2025**

***Corresponding Author**
Shaikh Abdul Hannan
Assistant Professor, Faculty of Computing and Information, AlBaha University, AlBaha, Kingdom of Saudi Arabia.

https://doi.org/10.5281/zenodo.18171104

How to cite this Article: Shaikh Abdul Hannan*. (2025). A COMPREHENSIVE ANALYTICAL METHOD FOR PREDICTING DISEASES THROUGH MACHINE LEARNING APPROACHES. World Journal of Engineering Research and Technology, 11(12), 286–307.

ABSTRACT

In the area of data mining, supervised machine learning techniques have been a prevalent approach. Recently, a possible use case for these techniques has emerged:disease prediction utilising health data. Early identification of infectious illnesses is thus crucial, and several researchers have created models to identify them early. It requires a lot of conventional clinical studies, which might make predicting the illness more difficult. A thorough analysis of several illness prediction techniques is provided in this work. The present research offers a prediction model that makes use of many information combinations and well-known categorisation techniques. In order to improve the prediction potential of the model, techniques like the Under-Sampling Clustering Over sampling Method (USCOM) address the issue of data imbalance. Adaptive Elephant Herd Optimisation Method (AEHOM) training is used to a multilayeredDeep Convolutional Neural Network (MLDCNN) for the classification job. According to its comprehensive analysis, the proposed Machine Learning-Based Heart Disease Prediction Method (ML-HDPM) has excellent results across many significant assessment variables.The ML-HDPM model performs very well during training, achieving 94.8% precision and 95.5% accuracy. With an accuracy rate of 96.2%, the system's sensitivities (Recall) are very accurate, and its F-score of 91.5% highlights its well-balanced performance. An excellent 89.7% specificity is reported by ML-HDPM, which is significant. The findings demonstrate how ML-HDPM may transform the prediction of heart disease and assist medical practitioners in making precise diagnoses, significantly impacting patient care results.

KEYWORDS: Healthcare Industry, Optimization Method, Training Process, Machine Learning-Based Heart Disease Prediction Method, Precise Diagnoses.

I. INTRODUCTION

According to a McKinsey report, 80% of healthcare expenses in the US are related to treating the many chronic conditions that 50% of Americans have. As living standards increase, chronic disease is becoming more common. Chronic illness treatment costs USD 2.7 trillion on average per year in the US. This sum amounts to 18% of the US GDP annually (J. P. Ellis, 2021). Chronic illnesses pose a serious threat to healthcare in many other nations. As to 2015 Chinese research on diet and chronic diseases, chronic diseases are the leading cause of mortality in China, making up 86.6% of all fatalities. Therefore, risk evaluations are crucial for chronic diseases. The collection of Electronic Health Records (EHR) is becoming simpler as medical data accumulates. Furthermore, advanced heterogeneous vehicular networks were deployed to gather enormous amounts of data on mobile users' health in real time after the first introduction of a bio-inspired highly effective heterogeneous vehicular telematics paradigm. A healthcare institution is using smart clothes to sustainably monitor health. For cost reduction in tree and easy route scenarios for heterogeneous systems, these systems yielded the best results.

To learn from the past and identify valuable patterns in vast, unorganised, and complicated datasets, machine learning algorithms use a range of statistical, probabilistic, and optimisation techniques. Automated text classification, detection of network intrusions, junk email filtering, credit card fraud detection, customer purchasing habits detection, manufacturing process optimisation, and illness modelling are just a few of the many uses for these algorithms. Instead of utilising unsupervised machine learning techniques, the majority of the applications mentioned have used supervised ones. The supervision form involves learning a dataset with a known label in order to create a model for prediction that predicts the results of unlabelled samples.

A doctor needs to be skilled and experienced in order to properly diagnose people who are having a heart attack. Because of this, the duty to use the knowledge and experience of different specialists as well as the clinical screening data gathered in databases to make the analysis process easier is viewed as a useful structure that combines computer-aided patient records and clinical selection tools (J. A. Runstadler, 2021). Additionally, it can improve patient results, decrease treatment errors, improve patient safety, and get rid of pointless disputes. In the medical area, machine learning has already been addressed extensively, especially in relation to disease detection and therapy. Machine learning has the potential to increase diagnostic time and accuracy, according to recent studies. In recent years, machine learning-based AI systems have grown in effectiveness as diagnostic tools.

A variety of disciplines are included in healthcare.It focusses on preventing, identifying, treating, & caring for diseases and conditions that affect the general health and wellbeing of people as well as communities.The range of services is broad and includes a variety of items such medical examinations, interventions, counselling, and public health initiatives.

Algorithms for machine learning are quite good at forecasting the results of a lot of data. The process of transforming enormous amounts of raw data into knowledge that is very useful for predicting and making decisions is called data mining. By producing faster and more accurate diagnoses, machine learning-based technology has the potential to completely transform the healthcare system and provide impoverished people access to high-quality medical treatment worldwide (2020, W. Tang). In addition to lowering the need for pointless diagnostic testing and improving diagnosis accuracy, machine learning may also cut down on the period it takes patients to see their doctors. The prevalence of complicated disorders can be considerably decreased by preventive measures. In an effort to lower mortality and improve clinical decision-making, some clinicians have suggested using machine learning and models of prediction to increase patient identification. Cardiovascular disease risk can be identified using machine learning, which can also help doctors treat patients and give them advise (Albahra, L., 2022).

Classification is one of the most widely used data mining methods. It enables the organisation of different kinds of information, from basic to complex, and divides information into classes. Classification is one kind of supervised training used in data mining. The main goal of categorisation is to make predictions based on the connection between the input and target variables. Numerous classification techniques were used in this work, such as decision trees, random forest (Random Forest), Support Vector Machines (SVM), and (Oğul, R., 2020). According to Melillo and colleagues' research, the CART algorithm had the highest accuracy of any method, coming in at 93.3%. Patients with congestive heart disease and those at lower risk were identified using this technique.

A subfield of artificial intelligence called Machine Learning (ML) uses data to learn and anticipate outcomes. Monitoring, unmonitored, and reinforcement learning are the three primary categories of machine learning algorithms (Ebrahimi M, Razavi A, 2013). To forecast the dependent variable and then attain the required accuracy, supervised learning involves training a model—such as a decision tree, K-nearest, Random Forest, or Logistic Regression neighbours—on provided data and independent variables. Unsupervised learning, on the other hand, uses techniques like the Apriori algorithm & k-means to find pattern in data used for training that are not labelled or classed and then classify them according to the features that were retrieved (Suseendran, G, 2020). Additionally, by using trial and error, the reinforcement learning paradigm teaches machines to learn from previous experiences and make appropriate decisions.

Recent advancements in Machine Learning (ML) approaches for the Internet of Things (IoT) have also been seen. It has been demonstrated that applying by using machine learning algorithms to network traffic data, IoT devices connected to a network may be reliably identified.labelled network traffic data from nine distinct PCs, IoT devices, and mobile phones. They used supervised learning approaches to develop a multi-stage meta classifier (Ekinci, E. & Garip, Z, 2023). The classifier can differentiate between traffic produced by IoT devices and non-IoT devices in the first stage. Every Internet of Things device is linked to a certain class of IoT devices in the second stage.One promising method for deriving precise information from unprocessed sensor data from Internet of Things devices placed in intricate settings is deep learning. Deep learning is also suitable for the cutting-edge computing environment due to its layered structure.

Describe the use of Computer-Aided Decision Support Systems (CADSS) in research and medicine.It has been demonstrated in earlier research that applying data mining techniques to the healthcare sector can forecast diseases more accurately and, in less time, (Afrash, M. R 20220. We suggest using the GA to diagnose cardiac disease. Effective rules for association derived from the GA are used in this method for crossover, tournament selection, and mutation, which leads to the newly suggested fitness function. We employ the popular Cleveland dataset, which is gathered from the UCI machine learning repository, for validation experiments. Later on, we'll see how our findings stand out in comparison to some of the well-known supervised learning methods.Particles Swarm Optimisation (PSO), the most potent evolutionary algorithm, is presented, and some guidelines for heart disease are produced (Harimoorthy, K. & Thangavelu, M., 2021). Overall precision has increased as a result of the rules' haphazard application with encoding techniques. Age, sex, pulse rate, and many other indicators are used to predict heart disease. Neural networks are used in the ML algorithm, which produces more precise and dependable results.

Heart disease is a prevalent and often consequential medical condition that is closely linked to the larger healthcare system. Globally, Heart disease is a leading source of sickness and death, underscoring the importance of healthcare systems in tackling pressing health concerns. Among other conditions affecting the heart and blood vessels, heart disease include heart failure, coronary artery disease, arrhythmias, & valve abnormalities.

Heart disease risk factors include advanced age, predisposition to the disease, tobacco use, physical behaviours, drug misuse, high blood pressure, high cholesterol, sedentary lifestyle, diabetes, being overweight, emotional stress, & poor hygiene habits. Given the severity of cardiovascular disease, a screening protocol must be put in place in order to diagnose it. Numerous diagnostic tests, such as During the screening procedure, medical professionals do tests for blood glucose, cholesterol, blood pressure, electrocardiography (ECG), imaging using ultrasound, cardiac computer tomography (CT), calcium scoring, & stress. The screening process involves a lot of labour-intensive physical labour and human interaction.

1.1 Objectives of the study

- Advanced feature selection using the GA and RFEM ensures that the most relevant characteristics are found, minimises interference, and improves the predictive power of the model.
- By efficiently using deep learning to identify intricate patterns found in complex datasets, an MLDCNN enhances the model's prediction performance when used for classification.
- By facilitating parameter refining, the AEHOM improves the model's performance and encourages efficient convergence during training.

II. LITERATURE REVIEW

Elmogy, M. (2022) Because of the high death rates, colon cancer detection and treatment are seen as social and economic concerns.Nearly 500,000 individuals worldwide get cancer each year, including colon cancer. Numerous screening procedures that may be used to look into polyp pictures and colorectal cancer have been developed since Examining the gland's anatomy by tissue area is the primary method used to determine those in grade of colon cancer. This page offers a comprehensive description of colon cancer diagnosis.Many elements of colon cancer are covered here, including its symptoms, grades, accessible imaging modalities (including histopathological pictures needed for analysis), and typical diagnostic techniques.Additionally, the most popular datasets and measures for evaluating performance are covered. We provide a thorough analysis of the most recent research on colon cancer, divided into two categories: Deep Learning (DL) & Machine Learning (ML). We highlight the main benefits and drawbacks of each technique.These techniques provide strong evidence in support of early cancer detection, which leads to prompt intervention and a lower mortality rate than when symptoms start to show. Additionally, these methods may help stop the advancement of colorectal cancer through the removal of pre-malignant polyps, which can be done with screening tests to help diagnose the condition.Lastly, the current difficulties and potential lines of inquiry that pave the way for further study in this area are discussed.

Jha, V. K. (2024) Diabetes has become a major worldwide health concern, since it is linked to a number of serious consequences, including heart problems, renal disease, and eyesight loss. Using machine learning algorithms in medical services has shown potential in accurately diagnosing and treating diseases, which will reduce the workload for medical workers. Diabetes forecasting is a fast-developing discipline that offers the possibility of patient empowerment and early intervention. A variety of machine learning approaches, such as SVM, Random Forest, Naïve Bayes, and Logistic Regression, are used in our work to create a unique diabetes prediction model.Alongside these fundamental methods, we use ensemble learning to further improve prediction resilience and accuracy. We specifically investigate group techniques like Adaboost, LightGBM, CatBoost, XGBoost, and Bagging. A more accurate and robust final prediction is produced by combining predictions from many base learners. We use a real-world dataset from Kaggle to construct and train our suggested framework in Python.

Schneider, R. (2024) In tropical and sub-tropical places around the globe, dengue fever, a common and quickly spreading arboviral illness, presents significant public health and economic issues. The spatiotemporal differences in dengue prevalence among administrative districts make it difficult to forecast infectious disease outbreaks on a national level. We provide a machine learning ensemble model to anticipate Brazil's Dengue Incidence Rate (DIR) with an emphasis on the population under the age of 19. The model provides state-level DIR estimates one month in advance by combining geographical and temporal data. The qualitative and quantitative effectiveness of the ensemble model across the 27 Brazilian Federal Units is shown by comparative analysis using a dummy model and ablation experiments. Additionally, we demonstrate how this strategy might be applied to Peru, another Latin American nation with unique epidemiological characteristics.

Lalbakhsh, P. (2022) Complexity, uncertainty, and nonlinearity are some traits shared by complex systems. Components usually interact in these phenomena, meaning that one portion of the structure may influence other sections or the other way around. As a result, these phenomena may be applied to the human brain, the global climate of Earth, the spread of diseases, economic institutions, and certain technical systems like power grids and transportation networks. Combining both of these methodologies may result in new hybrid methods with significant performance since both analytical approaches and AI methods have unique qualities when it comes to addressing complicated problems.For this reason, a number of studies have been carried out lately to take use of these combinations in order to forecast the dynamic behaviour and spread of COVID-19. This study aggregates and examines 80 peer-reviewed publications, book chapters, conference proceedings, and preprints published in 2020 that concentrate on using combinations of methods to predict the spread of COVID-19.

Mersha, T. B. (2022) Thanks to high-throughput next-generation sequencing, a large quantity of multi-omics data may now be produced for a variety of uses. The more thorough knowledge of biological systems and molecular pathways of disease development that these data have brought forth has transformed biomedical research. Deep Learning (DL) algorithms have emerged as one of the most promising approaches for analysing multi-omics data in recent years because of their capacity to capture hierarchical and nonlinear characteristics and their predictive accuracy. Multi-omics analysis is increasingly being used in biomedical research to help explain the intricate interactions between molecular levels, but the largest obstacle still lies in integrating and converting multi-omics data into practical functional insights. Enhancing prevention, early diagnosis, and prediction; tracking progression; interpreting patterns and end types; and creating customised therapies are all made possible by multi-omics data.We provide a realistic picture of the benefits, difficulties, and obstacles to applying DL to multi-omics data in this paper, along with a roadmap for multi-omics integrating using DL.

Atashi, A. (2022) A number of clinical, lifestyle, social, and economic variables may contribute to breast cancer, which is regarded as one of the most prevalent diseases in women.Based on data characteristics that are concealed, machine learning may be able to predict breast cancer. UsingThis study used a variety of machine-learning approaches for predicting breast cancer risk based on demographic, laboratory, & mammographic data.The Motamed Cancer Institute (ACECR), located in Tehran, Iran, provided the database used for this analytical study. It had 5,178 separate records, of which 25% were those of breast cancer patients. There were 24 attributes in each record. Each entry in the database contained 24 attributes, and of the 5,178 distinct entries, 25% belonged to people who had breast cancer. This research included Genetic Algorithms (GA), Gradient Boosting Trees (GBT), Random Forests (RF), and Neural Network Algorithms (NNA).First, 20 laboratory and demographic characteristics were used to train the models.To assess how well mammography characteristics predict breast cancer, the models were then trained using all 24 data—demographic, laboratory, and mammographic.

Al-Jumeily, D. (2022) The use of machine learning algorithms to identify (and perhaps forecast) Genetic data on Alzheimer's disease may help predict outcomes more accurately. Comprehensive study on the use of genetic data for the analysis and detection of Alzheimer's disease is still in its early stages. This research examined the scholarly literature on the use of several machine learning approaches to predict Alzheimer's disease based only on genetic data. To identify gaps in the literature, critically assess reporting and algorithmic approaches, and provide the foundation for a wider research program focused on developing novel machine learning-based algorithms for prediction in Alzheimer's disease.Utilising three search engines (PubMed, Web of Science, and Scopus), a systematic evaluation of quantitative research was carried out, including articles published between January 1, 2010, and December 31, 2021."Artificial intelligence," "GWAS," "Alzheimer's disease(s)," and its synonyms were the keywords utilised.There were 24 studies that met the inclusion/exclusion criteria.The examined publications' machine learning approaches varied greatly in their performance (0.59 to 0.98 AUC). The primary results demonstrated that feature selection, hyperparameter search, and validation techniques significantly increase the likelihood of bias in the study.

Azodinia, M. (2023) The purpose of this study article is to forecast Coronary Artery Disease (CAD) based on data collected from 350 participants at a hospital in Armenia. One important aspect that may have a big impact on a patient's survival & life expectancy is CAD. The research takes into account a variety of input factors in order to create accurate predictions, such as the Patient's Age (PA), Sex (S), Family History (FH), Cholesterol level (LOC), Number of Blocked Arteries (NAB), and kind of chest discomfort (TCP). To achieve this crucial objective of CAD prediction, the researchers used three powerful classification algorithms: Random Forest (RF), Support Vector Machine (SVM), & Logistic Regression (LR). Of these, the Random Forest algorithm is notable for its resilience and many benefits, such as its high accuracy, good handling of outliers, capacity to provide insights into feature relevance, and decreased danger of overfitting. This article's study results, which show an accuracy value of 0.95 and a precision value of 0.94, highlight the Random Forest algorithm's remarkable performance. With a life-or-death parameter like CAD, these findings demonstrate the model's capacity for accurate and trustworthy prediction-making. Random Forest performs better than SVM and LR, according to the researchers' comparison study based on statistical factors. The study's findings thus implies that the algorithms should be ranked according to their performance as follows: RF > SVM > LR.

Calhoun, V. (2020) A paradigm change in biomarker discovery has occurred in the neuroimaging community, moving the discipline towards a translational neuroscience age by switching from classic univariate brain mapping methodologies to multivariate prediction models.Predictive modelling, often known as regression-based multivariate models, is a popular and effective method for using neuroimaging characteristics to forecast human behaviour. By focussing on interpreting individual variations in a consistently behavioural phenotype from neuroimaging data, this research provides an intriguing chance to characterise the human brain at the single-subject level. In this review, we provide a summary of current research conducted over the last ten years that look for neuroimaging predictions using machine learning techniques. We begin by reviewing regression-based methods and highlighting the current surge in popularity of connectome-based predictive modelling. Using these approaches, we next methodically discuss current representative research in relation to emotion processing, personality factors, symptom severity, and cognitive performance. Finally, we highlight some of the issues that have surfaced from our assessment of the literature, including the use of deep learning techniques, longitudinal prediction, merging multimodal data, and external validations. We also provide several exciting and difficult future directions.

Saleh, A. M. (2024) Agriculture is poised for a revolutionary shift propelled by Machine Learning (ML) in the age of increasing digital data. With a focus on how Machine Learning (ML) is changing agricultural methods, this paper explores the complex relationship between ICT and traditional agriculture. A shining example of innovation in the face of the continuous data deluge that affects data-driven organisations is the combination of precision agriculture and smart farming. Making educated decisions is made easier by machine learning algorithms that analyse historical and current environmental data, soil conditioning, forecast which crops will produce the most, identify illnesses, and optimise irrigation in smart farming. ML-powered drones and driverless cars help precision agriculture by guaranteeing accuracy in crop monitoring, planting, and harvesting. As ML controls fertiliser application, optimisesenergy use, and encourages climate-resilient behaviours, resource efficiency rises.ML is essential for increasing production, reducing environmental impact, and negotiating the challenges of contemporary agriculture, as this thorough evaluation highlights.

III. METHOD AND MATERIALS

Using a multimodal strategy that combines cutting-edge the research intends to provide a comprehensive framework for precisely forecasting cardiac illness using machine learning techniques, selecting features, and reduction of dimensionality methods (Kumar, S. & Yeligeti, S., 2021).

Additionally, the hybrid system includes three essential phases that are crucial to improving the prediction model:data collection, pre-processing, and classification. To maintain model effectiveness and data integrity, a number of the thorough pre-processing phase involves meticulous activity. Among these responsibilities is the meticulous imputation of data that is missing, where the ML-HDPM approach is efficient in accurately estimating database missing values. Additionally, careful feature selection attempts are conducted to determine which are most relevant attributes for predictive modelling. This procedure is made easier by a hybrid approach that combines the strengths of Recursive Feature Elimination Method (RFEM) and Genetic Algorithm (GA) in a synergistic way, making it easier to identify relevant features that are essential for accurate predictions (Hussein, H. I. & Hassan, M. M., 2023).

Furthermore, in order to guarantee consistency in the influence of features, each feature's coefficients are recalculated using a conventional scalar technique, bringing their means to zero and the standard deviations to one. This lessens the likelihood of biases brought on by differences in feature scales. There are two classifications in the database that facilitate classification: class 0 and class 1, which represent the absence and presence of cardiac illness, respectively (Ahsan, M.M.; Siddique, 2021). In particular, 139 cases are classified as class 1, indicating the existence of heart disease, while 164 cases are classified as class 0, indicating the absence of heart disease. The Synthetic Minority Over-Sampling Technique (SMOTE) is used to counteract the innate class imbalance, guaranteeing fair representation of both classes and enhancing the model's ability to forecast outcomes in a variety of situations. The approach achieves a balanced distribution across both kinds by creating fictional examples of under-represented categories. The classification process uses a variety of classifiers, including Along with specific features, these methods include Support Vector Machines (SVM), PCA, Linear Discriminant Analysis (LDA), Naïve Bayes (NB), Decision Trees (DT), & Random Forests (RF). In the end, the classifier predicts whether an individual has heart disease. The procedure used in the recommended method for predicting heart disease is shown in Fig. 1.

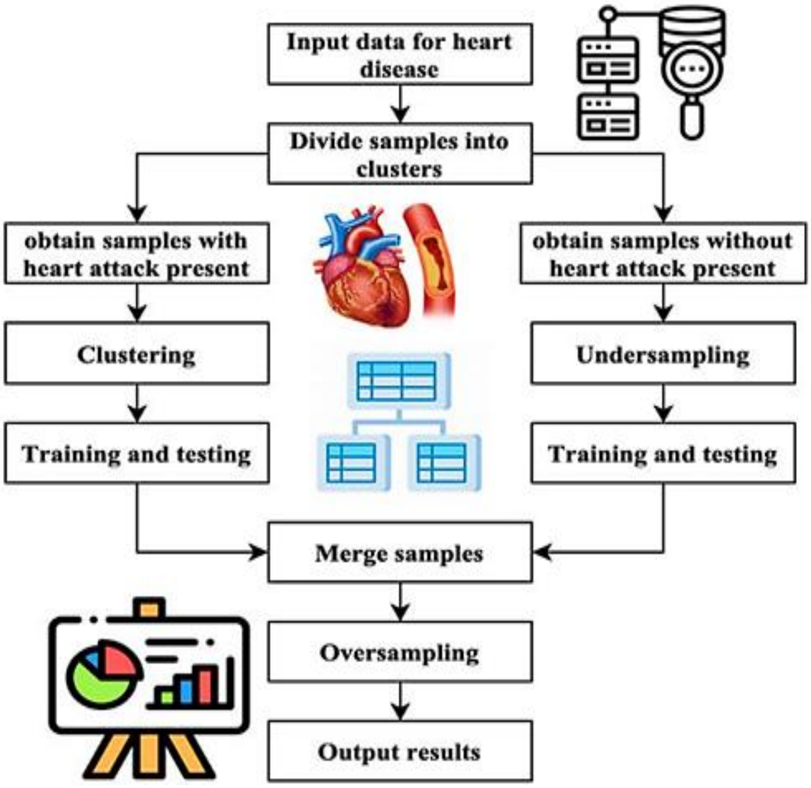


Fig. 1: Procedures for the proposed approach.

The mating pool that contains the genomes chosen by a selection agent results in the following generation (Algethami, A., Fatima, H., 2022). Individuals from mating pools are mixed using a genetic algorithm technique called the crossover agent to create generations of superior offspring. Many kinds of crossover agents exist, including multipoint, two-point, and single-point crossover. Unless a combination is incorporated into the broader population, people in the following generation will likely act similarly to the prior group. To display the diversity, a mutation manager is used, which introduces arbitrary changes to the individuals.Fig. 2 illustrates how the genetic techniques have developed.

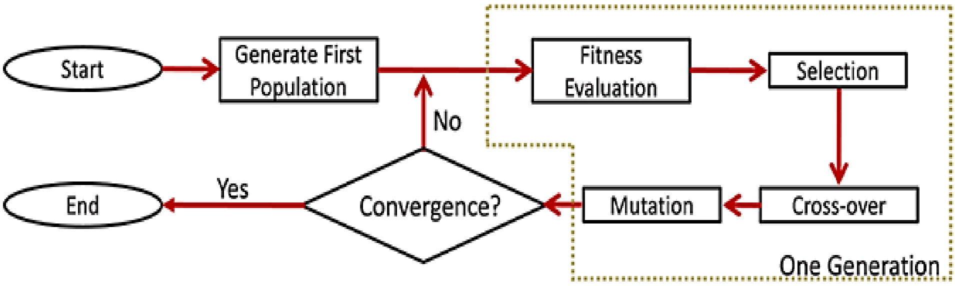


Fig. 2 The Procedure GA.

$V_{max} = \max\{V\}$ ………………………1

$THD = \frac{\sum_{i=0}^{n-1} HC_i}{P_{ff}}$ ………………………2

$H_r = \frac{60}{RR_t}$ ……………………………3

$ZCR = \frac{S_c}{S_c + S_{nc}}$ ………………………4

They used a crossover rate of 5% and a mutation rate of 0.05 (Kaliappan, M. & Lee, M. Y, 2021). The information provided by the user is invalid. The objective function for evaluating the genomes' fitness was the average square error of the genome (Piao, Y. & Lee, J. Y, 2019). A component of the iterative RFEM technique is the recursive removal of unnecessary features. Following training of the classifier using the provided training data, the significance of the features is then computed. The process retrains the model using the remaining subset of features after the least influential elements are eliminated at each iteration. Until the required number of features is achieved, the procedures are repeated. The method receives the number of characteristics to be maintained as an argument. Fig. 3 shows the RFEM method's operation.

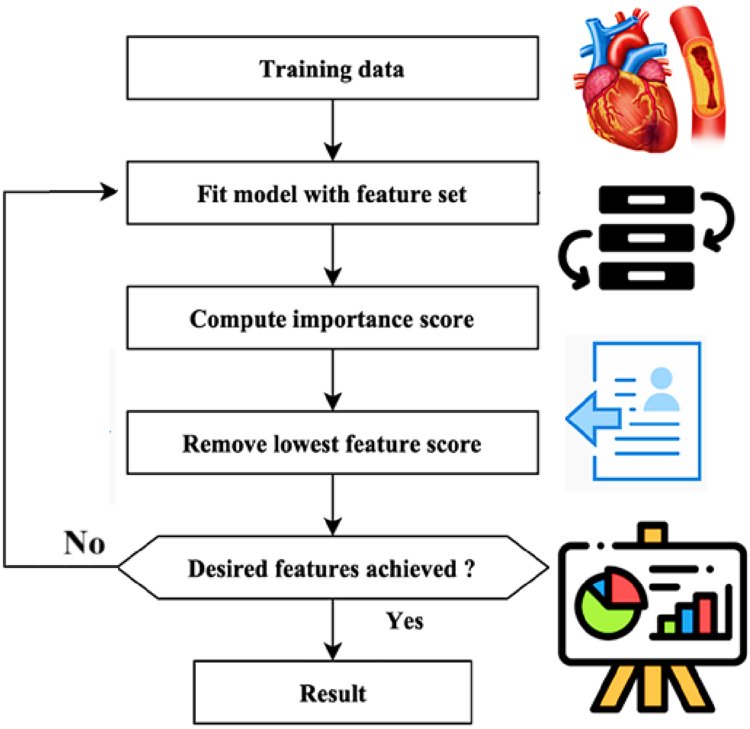


Fig. 3: Process for the RFEM method.

$$Ent = \sum_{x=0}^{m-1}\sum_{y=0}^{n-1}\frac{P_{xy}}{\log(P_{xy})} \text{ ....5}$$

$$Eg = \sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(P_{xy})^2 \text{ ....6}$$

$$SD = \frac{1}{M}\sqrt{\sum_{x=0}^{m-1}(RR_x RR_{x-1} - K)^2} \text{ ....7}$$

Figure 4 displays the system's primary characteristics (Mothe, J. & Ullah, M. Z., 2020). This technique under-samples when sampling with majority labels. In the whole data set, clustering is applied to the subset with minority labels. The underestimating process generates an example set, from which distinct testing, training, and validation sets are generated. This split is done using the 8:1:1 ratio.Integration has been made between the acquired training sets. The sets for testing and validation are complete (Talaat FM, 2024). After oversampling the specimens with minority labels in the combined training set, the SMOTE algorithm subsequently obtains the last stage of data used for training.

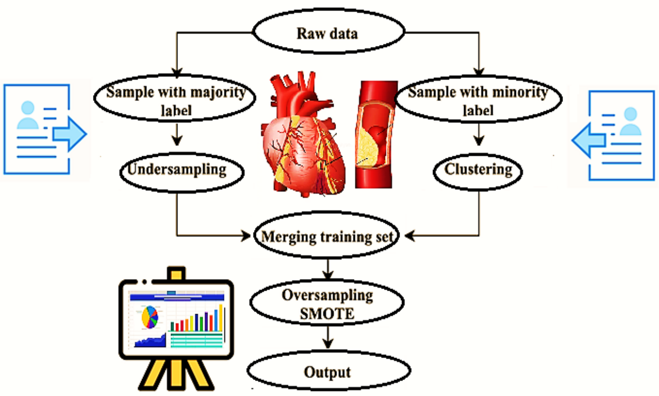


Fig. 4: Over-sampled cluster-based approach.

Figure 5 depicts the UCOM process for predicting cardiac disease. The UCOM algorithm is composed of three distinct parts. The data set of persons without cardiovascular attacks is under-represented since there are considerably more people without heart diseases than there are with them (Min, L. Z. & Ying, C. X., 2019).

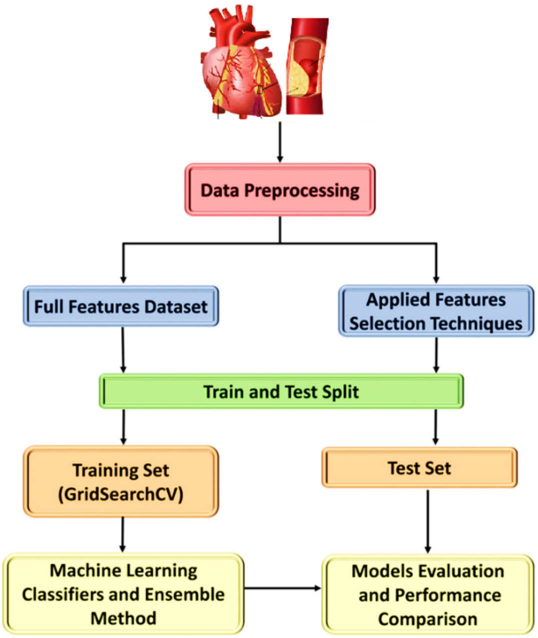


Fig. 5 The UCOM approach to heart disease prediction.

$A_{fx} = C_x\{\sum_{x=0}^{m-1} F_x W_x$ ………..........……8

$C_x = \exp(-(Fx)^2)$ ……..................…..…9

$P_c, E_n = \frac{1}{VE_n}\sum_{y=0}^{VE_n} b_{E_{n,y,l}}$ ……..…………10

The presented method integrates a comprehensive approach to heart disease prediction, encompassing feature selection, gathering data, machine learning approaches, and (Alizadehsani, R., 2013). Important aspects are found after the patient data has been obtained and analysed.

IV. EXPERIMENTAL RESULTS

Comprehensive testing is used in the simulation section to evaluate the suggested approach's efficacy. Accuracy, sensitivity, specificity, and precision are among the metrics that are calculated to assess the prediction efficacy of the approach. A comparison with current models confirms that the suggested method is effective.

In order to enhance performance within the limitations of the research, the modified bee technique—a variant of the traditional bee method for optimization—was specifically created via repetitive parameter exploration and optimisation.The MATLAB 9.2 model-development program was used for these testing on a machine setup with an i5 CPU and 3 GB of RAM.

Fig. 6 displays the accuracy analysis of the Machine Learning Hybrid Deep Predictive Model (ML-HDPM) approach to heart disease prediction. Testing accuracy (86.8%) and training accuracy (99.8%) are two areas where ML-HDPM performs noticeably better than competing algorithms.Several innovative techniques are included into the model, which is responsible for this noteworthy development.

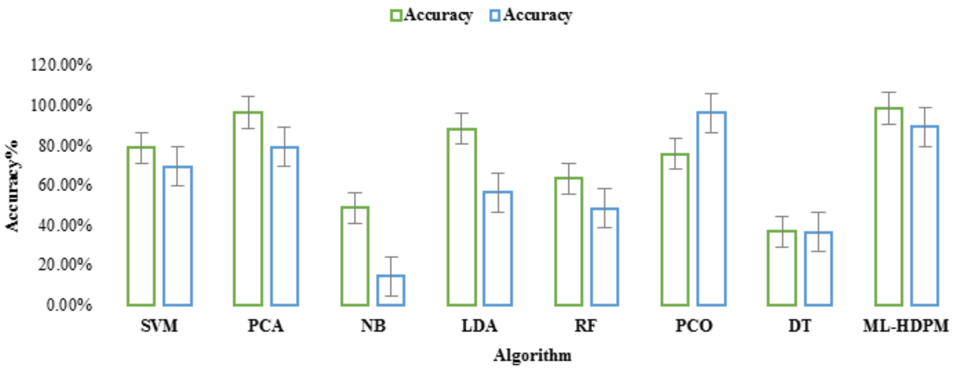


Fig. 6: Evaluation of the Predictive Accuracy of Heart Disease.

The efficiency of the Machine Learning Hybrid Deep Predictive Model (ML-HDPM) technique in the accuracy analysis of heart disease prediction is shown in Figure 7.Remarkably, ML-HDPM performs better than other algorithms in terms of accuracy rates in both the testing (79.4%) and training (98.3%) phases. The holistic approach of ML-HDPM, which integrates many crucial approaches, is responsible for this accomplishment.

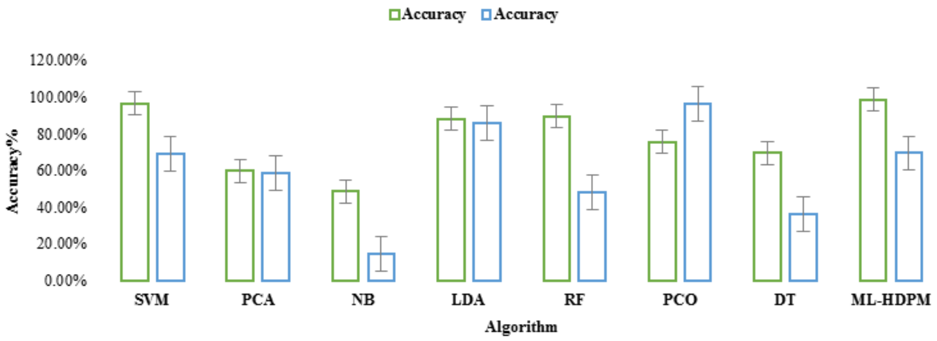


Fig. 7: Reliable assessment of heart disease diagnosis.

A comparison of the False Positive Rate (FPR) for several heart disease prediction methods is shown in Figure 8. Comparing the Machine Learning Hybrid Deep Prediction Model (ML-HDPM) method to other methods, the FPR values are much lower in both the training (18.9%) and testing (17.8%) phases. This accomplishment is the result of the ML-HDPM framework's beneficial use of feature selection, data balancing, and improved deep learning elements.

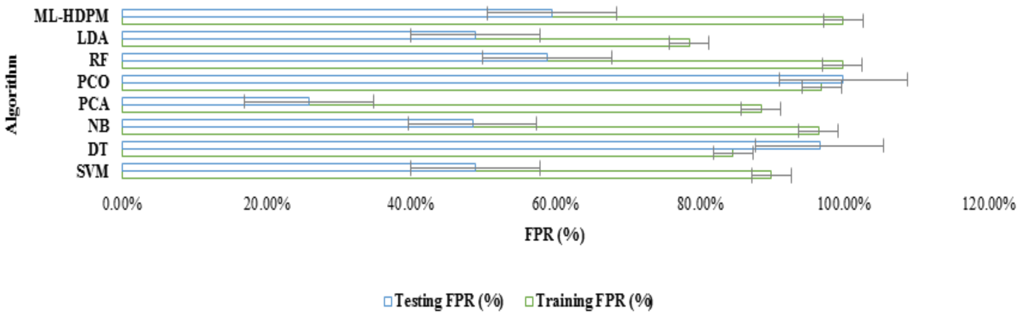


Fig. 8: Examination of the false-positive rate in the prediction of heart disease.

Figure 9 shows the True Positive Rate (TPR) comparison of several algorithms for heart disease prediction in order to illustrate the enhanced performance of the machine learning Hybrid Deep Predictive Model (ML-HDPM) method. With TPR values of 99.7% and 87.8%, respectively, ML-HDPM achieves notably higher TPR values throughout the training and testing phases.

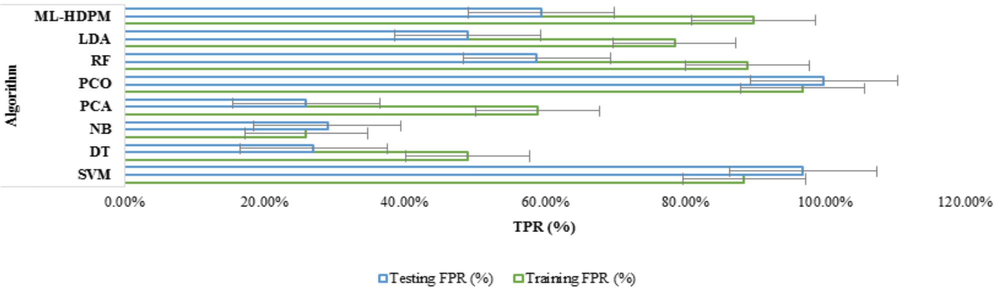


Fig. 9: Examination of true positive rates for the prediction of heart disease.

When compared to other algorithms, the ML-HDPM technique achieves superior F-scores, achieving higher testing (87.9%) and training (99.8%) scores. The results of the F score for each model are shown in Fig. 10. The methodology's completeness, which includes data balance, feature selection, and deep learning optimisation techniques, is credited with this accomplishment.

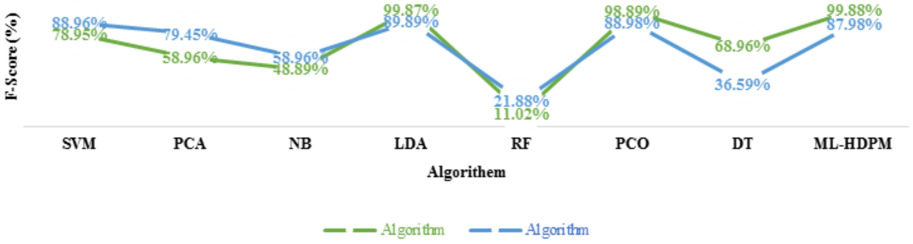


Fig. 10: F-Score analysis for heart disease prediction.

According to the study's simulation results, the Machine Learning Hybrid Deep Predictive Model (ML-HDPM) technique is highly effective in properly predicting heart disease. ML-HDPM beats other algorithms, showcasing its remarkable performance in both sets with a testing accuracy of 95.6% and a training accuracy of 96.36%. The accuracy scores of 69.5% for testing and 96.5% for training, which emphasise the model's efficacy across various datasets, further corroborate this edge (Fig. 11).

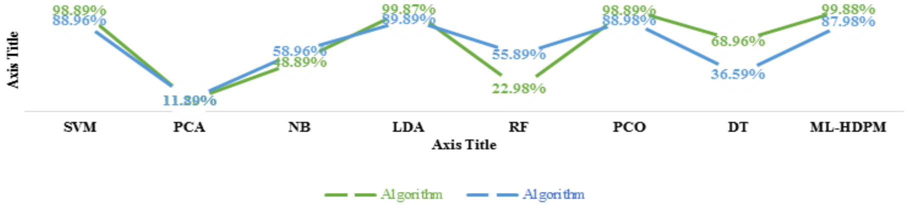


Fig. 11: Evaluation of the prediction specificity for hearing impairment.

V. CONCLUSION

In complicated situations, infectious diseases may make individuals die, and they affect people everywhere. There have been several contributions using a range of methods to detect and classify infectious diseases. ML algorithms may also aid in the early diagnosis of infectious diseases.

Furthermore, the application of Machine Learning (ML) models to support medical diagnostics, screening, and medical forecasting is expanding due to its exceptional ability to identify and categorise patients. The amount of complex and erroneous health data that clinical practitioners currently deal with makes it difficult for them to make well-informed judgements.Although the speed of decision-making in cardiac problems has the potential to reduce complications and improve the individual's well-being, machine learning algorithms have proven essential in the prediction, diagnosis, and treatment of several illnesses due to their high accuracy.

Physical patient care has grown hazardous after the emergence of COVID-19. The healthcare system must be maintained in the meantime. Even if telemedicine and online consultations are becoming more and more popular, it is still important to consider an alternative strategy that can highlight the importance of real medical facilities. Using home robots for medical treatment rather than hospitalisation is recommended by a number of recent research.

The privacy of patient data is a growing problem in many countries. A number of nations have also voiced legal concerns about the ethics of AI and ML when used on real medical data. Therefore, rather of depending only on data gathering and processing, future study may try to create synthetic data.

This research sorted raw data using machine learning techniques to provide a novel and creative heart disease diagnosis. In medicine, predicting cardiac disease is a challenging but essential endeavour. However, the fatality rate might be considerably decreased if the disease is discovered early and preventative measures are implemented immediately.

The research provides a comprehensive approach termed ML-HDPM that attempts to enhance the prediction of cardiovascular disease by addressing the drawbacks of traditional diagnostic methods. This approach makes use of the MLDCNN model improved with

AEHOM and incorporates a number of techniques, including feature selection, data balancing, and deep learning.

FUTURE WORKS

Further research and methodological advancement are also necessary, according to the report, particularly in regards to problems with feature selection, data quality, and optimisation techniques.

Further research attempts must concentrate on expanding the study's scope to include a greater range of cardiovascular illnesses and examining the practical use of the findings in real-world healthcare settings. The study highlights the potential of advanced machine learning methods to transform the diagnosis and treatment of cardiovascular disease.

VI. REFERENCES

1. T. Boyles, A. Stadelman, J. P. Ellis et al., “The diagnosis of tuberculous meningitis in adults and adolescents: protocol for a systematic review and individual patient data metaanalysis to inform a multivariable prediction model,” Wellcome Open Research, 2021; 4.
2. L. K. Borkenhagen, M. W. Allen, and J. A. Runstadler, “Influenza virus genotype to phenotype predictions through machine learning: a systematic review,” Emerging Microbes and Infections, 2021; 10(1): 1896–1907.
3. E. E. Kpokiri, G. Marley, W. Tang et al., “Diagnostic infectious diseases testing outside clinics: a global systematic review and meta-analysis,” in Open forum infectious diseases, Oxford University Press, US., 2020.
4. N. K. Tran, S. Albahra, L. May et al., “Evolving applications of artificial intelligence and machine learning in infectious diseases testing,” Clinical Chemistry, 2022; 68: 125–133.

A. Baldominos, A. Puello, H. Oğul, R. Colomo-palacios, and S. Member, “Predicting infections using computational intelligence – a systematic review,” IEEE Access, 2020; 8: 31083–31102.

5. Ahmad LG, Eshlaghy A, Poorebrahimi A, Ebrahimi M, Razavi A. Using three machine learning techniques for predicting breast cancer recurrence. J Health Med Inform, 2013; 4(124): 3.
6. Thiyagaraj, M. & Suseendran, G., editors. Enhanced prediction of heart disease using particle swarm optimization and rough sets with transductive support vector machines

classifier. In Data Management, Analytics and Innovation: Proceedings of ICDMAI 2019, 2; 2020: Springer.

7. Ay, Ş., Ekinci, E. & Garip, Z. A comparative analysis of meta-heuristic optimization algorithms for feature selection on ML-based classification of heart-related diseases. J. Supercomput, 2023; 1–30.
8. Afrash, M. R. et al. Machine learning-based clinical decision support system for automatic diagnosis of COVID-19 based on clinical data. J. Epidemiol. Biostat, 2022; 8(1): 77–89.
9. Harimoorthy, K. & Thangavelu, M. Multi-disease prediction model using improved SVM-radial bias technique in the healthcare monitoring system. J. Ambient. Intell. Humaniz. Comput, 2021; 12: 3715–3723.
10. Tharwat, M., Sakr, N. A., El-Sappagh, S., Soliman, H., Kwak, K. S., &Elmogy, M. Colon cancer diagnosis based on machine learning and deep learning: modalities and analysis techniques. Sensors, 2022; 22(23): 9250.
11. Modak, S. K. S., & Jha, V. K. Diabetes prediction model using machine learning techniques. Multimedia Tools and Applications, 2024; 83(13): 38523-38549.
12. Sebastianelli, A., Spiller, D., Carmo, R., Wheeler, J., Nowakowski, A., Jacobson, L. V., ... & Schneider, R. A reproducible ensemble machine learning approach to forecast dengue outbreaks. Scientific reports, 2024; 14(1): 3807.
13. Jamshidi, M., Roshani, S., Daneshfar, F., Lalbakhsh, A., Roshani, S., Parandin, F., ... &Lalbakhsh, P. Hybrid deep learning techniques for predicting complex phenomena: A review on COVID-19. AI, 2022; 3(2): 416-433.
14. Kang, M., Ko, E., & Mersha, T. B. A roadmap for multi-omics data integration using deep learning. Briefings in Bioinformatics, 2022; 23(1): bbab454.
15. Rabiei, R., Ayyoubzadeh, S. M., Sohrabei, S., Esmaeili, M., & Atashi, A. Prediction of breast cancer using machine learning approaches. Journal of biomedical physics & engineering, 2022; 12(3): 297.
16. Alatrany, A. S., Hussain, A. J., Mustafina, J., & Al-Jumeily, D. Machine learning approaches and applications in genome wide association study for Alzheimer's disease: A systematic review. IEEE Access, 2022; 10: 62831-62847.
17. Ghorbani, H., Krasnikova, A., Ghorbani, P., Ghorbani, S., Hovhannisyan, H. S., Minasyan, A., ... &Azodinia, M. Improving the estimation of coronary artery disease by classification machine learning algorithm. In 2023 IEEE 6th International Conference and

Workshop Óbuda on Electrical and Power Engineering (CANDO-EPE), October. 2023; (000159-000166): IEEE.

18. Sui, J., Jiang, R., Bustillo, J., & Calhoun, V. Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: methods and promises. Biological psychiatry, 2020; 88(11): 818-828.
19. Mohyuddin, G., Khan, M. A., Haseeb, A., Mahpara, S., Waseem, M., & Saleh, A. M. Evaluation of machine learning approaches for precision farming in smart agriculture system: a comprehensive review. IEEE access, 2024; 12: 60155-60184.
20. Ashish, L., Kumar, S. &Yeligeti, S. WITHDRAWN: Ischemic heart disease detection using a support vector machine and extreme gradient boosting method. Mater. Today Proc, 2021. https://doi.org/10.1016/j.matpr.2021.01.715.
21. Hassan, D., Hussein, H. I. & Hassan, M. M. Heart disease prediction based on pre-trained deep neural networks combined with principal component analysis. Biomed. Signal Process. Control, 2023; 79: 104019.
22. Ahsan, M.M.; Siddique, Z. Machine Learning-Based Heart Disease Diagnosis: A Systematic Literature Review. arXiv 2021; arXiv: 2112.06459.
23. Ahmad, G. N., Ullah, S., Algethami, A., Fatima, H. & Akhter, S. M. H. Comparative study of optimum medical diagnosis of human heart disease using machine learning technique with and without sequential feature selection. IEEE Access, 2022; 10: 23808–23828.
24. Jackins, V., Vimal, S., Kaliappan, M. & Lee, M. Y. AI-based smart prediction of clinical disease using random forest classifier and naive Bayes. J. Supercomput, 2021; 77: 5198–5219.
25. Piao, M., Piao, Y. & Lee, J. Y. Symmetrical uncertainty-based feature subset generation and ensemble learning for electricity customer classification. Symmetry, 2019; 11(4): 498.
26. Déjean, S., Ionescu, R. T., Mothe, J. & Ullah, M. Z., editors. Forward and backward feature selection for query performance prediction. In Proceedings of the 35th Annual ACM Symposium on Applied Computing, 2020.
27. Talaat FM (2024) Revolutionizing cardiovascular health: integrating deep learning techniques for predictive analysis of personal key indicators in heart disease.
28. Reddy, N. S. C., Nee, S. S., Min, L. Z. & Ying, C. X. Classification and feature selection approaches by machine learning techniques: heart disease prediction. Int. J. Innov. Comput, 2019; 9(1): 9–46.

29. Alizadehsani, R. et al. A data mining approach for diagnosis of coronary artery disease. Comput. Methods Programs Biomed, 2013; 111(1): 52–61.