https://doi.org/10.48047/AFJBS.6.14.2024.9193-9200

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Modelling an Innovative Mathematical-driven Predictive Model for Cardiovascular Diseases

[1] Dr. R Punithavalli, S Kannaki, Saurabh Mishra, Dr. N.PUNITHA, D.SUDHA, A. Durai Ganesh,

Department of mathematics, Amet university, Kanathur, Chennai.

[2] Assistant professor of Mathematics, Dr. N. G. P. Arts and Science College, Coimbatore, Tamil Nadu kannaki@drngpasc.ac.in

[3] Assistant professor of Mathematics Shaheed Rajguru College of Applied Sciences, University of Delhi. itsmesaurabh349@gmail.com

[4] Faculty of Physics , Department of Science and Humanities, St. Joseph's College of Engineering, Chennai, Tamilnadu punithasan@gmail.com

[5] Assistant Professor of Mathematic, Hindusthan Institute of Technology, Coimbatore, Tamilnadu sudhamidhun2011@gmail.com

[6] Assistant Professor of Mathematics,PET Engineering College, Vallioor, Tirunelveli, Tamil Nadu. aduraiganesh25@gmail.com

Corresponding Author: A. Durai Ganesh aduraiganesh25@gmail.com

Volume 6, Issue 14, Aug 2024

Received: 15 June 2024

Accepted: 25 July 2024

Published: 15 Aug 2024

doi: 10.48047/AFJBS.6.14.2024.9193-9200

Abstract: **This research offers a fresh innovative, mathematical modeling technique of CVDs as an innovative machine learning model. Logistic Regression, Random Forest, Support Vector Machine (SVM), and Gradient Boosting improve prediction accuracy which are combined in the study. The model was assessed using a broad set of predictor variables that included demographic characteristics, life style and clinical parameters. The proposed model had better accuracy with 87%, precision 85%, recall 83% as well as F1-score of 84%. The test results on the Random Forest model provided the highest accuracy of 90%, while Gradient Boosting was on the second place with 88%, then SVM 85%, and Logistic Regression 82%. Comparing with the existant algorithms, corresponding increases of prediction accuracy and risk factor discrimination were identify. Feature importance analysis integrating was able to take the model to the next level by excluding unproductive covariate which affected the risk of CVD. This research contributes to the development of models for cardiovascular health problems in particular and, therefore, the development of early warning systems and risk estimation models. The study highlights the possibility of improving diagnostic and other care-related decisions as well as prevention intervention plans.**

Keywords: Cardiovascular Diseases, Machine Learning, Predictive Modeling, Graph Theory, Random Forest, Gradient Boosting

I. INTRODUCTION

Cardiovascular diseases are derived to be the number one killers globally, therefore, the chances of developing a competent prediction and prevention strategy are worth exploring. In conventional models used to diagnose and predict CVDs, doctors utilise clinical examinations and measured risks, which are quite valuable but do not embrace complete progression of diseases. With the recent development of sophisticated mathematical algorithms and analyses, a relatively new way to improve CVD prediction is the application of enhanced mathematical models and prediction tools which might deliver more precise and timely information to dramatically affect the CVD patients' outcomes [1]. The objective of this study is to advance original systematic mathematical risk prognosis with the goal of enhancing cardiovascular disease prognosis. Hence, this model aims at bringing out a complex and sophisticated analysis of risk factors of CVD and improved probabilities of prognosis The use of what is current in mathematical computation and formulas offers the above model a perspective in achieving these goals [2]. The framework combines clinical histories, patient characteristics, life style and biometric information of patients using a range of advanced statistical techniques and intelligent systems to extract features that would difficult for, normal models to discern. The specific aims of this study are to develop a new accurate model for CVD risk prediction and to assess the possibilities of mathematical developments for the refinement of the existing predictive models. It is the intention of this research to fill this gap between the theoretical mathematical studies and their application in clinical settings to possibly change the way cardiovascular status is monitored and treated. Thus, in line with the developments in the field of personalised medicine, the objective of this research is to take the state of the art of prognostic modelling in cardiovascular diseases to the next level. The expected outcome is the development of a reliable, empirically-based, mathematical tool that maybe beneficial to the clinicians in order to make the right decision regarding patients' treatment [3] and consequently, enhance patients' quality and lifespan in the area of cardiovascular diseases.

II. RELATED WORKS

In a recent research by Kumar et al., (2022) examined data mining methods required in creating models to predict cardiovascular diseases. In their work, Natarajan et al (2011) proved that there is a way to use data mining tools and the classification algorithm in particular for the prediction of CVDs taking into consideration patients' demographics, lifestyle and medical history [21]. It pointed out the possibility of these techniques in improving the clinical decisions and patient handling. In the same way, Marwah et al. (2024) conducted a literature review on machine learning in predicting cardiovascular diseases. In their extensive critique, the authors also reviewed milestones in the ML approaches for the last few years, focusing on the supervised and unsupervised learning models and their efficiency in the prognosis of several cardiovascular diseases [24]. The review also outlined some of the limitations and research directions as the technical challenges of model understanding and the generalization of findings. In another study, Ibrahim et al. (2024) put forward a model based on the application of fuzzy integration with the use of machine learning to predict β-Thalassemia carriers. This study proposed a new methodology that applied fuzzy logic together with machine learning algorithms to improve accuracies of prediction [16]. Despite the use of a different genetic disorder, as far as the method is concerned, it shows that using a number of approaches can enhance the accuracy of prediction, which is certainly a useful technique in cardiovascular disease models. In Huang and Huang (2024) regression model feature importance statistics are examined to deter mine covariates contributing to accuracy of prediction of insomnia [15]. It should be noted that their method of estimating the importance of the visualized feature can be used in models for the prediction of cardiovascular diseases where significant factors affecting the performance of the model can be determined in order to improve the adjustment of the standards of prediction. Another study from the eastern Mediterranean region was a cohort study by Mehrabani-Zeinabad et al. (2023) to estimate the cardiovascular disease burden and to use the machine learning and statistical models to predict the risk of CVD [26]. They noted that this approach and the other ones could be used in combination to improve prediction for longer periods of research. This approach highlight on the fact that more than one approach should be used in order to get highly accurate results and results that can be generalized. Mall et al. organized a new hybrid modeling approach, OHE2LM that they promoted to enhance the potency of prediction of heart attack [23]. This strategy entailed using ensemble learning together with other complicated machine learning approaches in the aim of enhancing accuracy of the predictions. They evidenced that such combinations could improve the performance of the combinations compared to monadic models. Islam et al. (2023) have used machine learning techniques to identify hypertension risk in Ethiopia [17]. The study aimed to compare the efficacy of algorithms in place for the non-ML group to that in the ML group with various geographical variations and demographics. What is also clear here is the mobility of the models within machine learning in catering for various populations as well as the necessity of the regionality. Most recently, Janyasupab et al. (2023) presented GeneCompete, a computational tool that integrate new algorithms and ranking for gene expression data [18]. Despite its focus on gene expression, the innovations in method as described could be carried forward to use in cardiovascular disease prediction similar to this study based on harmonized multiple data and analytical techniques. Khan et al. [20] employed deep neural network based methods to study and detect the presence of 5-mC modifications from sequence-based models. In their studies, they have elaborated the efficiency of new precise architectures of the neural net in prediction tasks, which can be adequate to cardiovascular disease prediction. Kalaichelvan et al. (2024) take the concept the fuzzy theory and machine learning to implement the economic order quantities in the context of the pharmaceutical sector [34]. This work showed the overall use of the principles of the fuzzy logistic function to control complex systems, which can be used in the further studies of the cardiovascular disease prognosis using the machine learning approach combined with the theory of the fuzzy logic. MD et al. (2024) added improved clinical validation for the prediction of early cardiovascular disease with simulation, artificial intelligence, and web technology [25]. The subjects' focus is on the application of IT and AI in clinical context, implying the possibility to enhance the possibility of early detection and optimal patient care.

III. METHODS AND MATERIALS

Data Collection and Preprocessing

The usefulness of any kind of a predictive model is solely dependent on the type of data that has been used in the model. In this study, electronic health records of patients, lifestyle indices, anthropometrics, and demographics were used source of data. The records of more than 10 000 patients were compiled; each record lists the patient's age, gender, blood pressure, cholesterol levels, smoking status, physical activity, and cardiovascular disease history [4]. Data preprocessing included addressing the requirement of missing values, outliers, and normalization of numerical features as well as data encoding of categorical features. In order to test the stability of the model, the data was divided in a training and a test set using an 80 percent to 20 percent split. The training set was used for model building while the testing set was used in the assessment of the performance of the models. The initial step was feature selection in which statistical measures were employed in order to determine the predictors that are most associated with cardiovascular diseases [5].

Algorithms

1. **Logistic Regression**

Logistic Regression is an algorithm for binary classification, and, as such, it may be applied to the problem of predicting the probability of cardiovascular diseases. It is a statistical technique that makes prediction about binary response variable by using one or more explanatory variables [6]. The logistic function used in this model is:

$p=1+e-z1$

$z=\beta0+\beta1x1+\beta2x2+\ldots+\beta nxn$

In the following equation, β is the model coefficients and xxx refers to the input variables. Logistic Regression also gives probability outputs ranging from 0 to 1 and this probability can be taken through a certain threshold to classify a patient as having Cardiovascular disease or not.

"1. Initialize coefficients (β) to zero.
2. For each iteration:
a. Compute predicted probability using sigmoid function.
b. Update coefficients using gradient descent.
3. Stop when coefficients converge.
4. Predict class based on probability threshold."

Pathway	**Enriched Gene Cluster**	**p-Value**	**Pathway**	**Enriched Gene Cluster**	**p-Value**
Apoptosis	Cluster A	0.001	Apoptosis	Cluster A	0.001
Cell Cycle	Cluster B	0.002	Cell Cycle	Cluster B	0.002
Metabolic Pathways	Cluster C	0.004	Metabolic Pathways	Cluster C	0.004
Immune Response	Cluster D	0.005	Immune Response	Cluster D	0.005

2. **Random Forest**

Random Forest is a procedure, which builds more than one decision tree, and then takes their outcomes. And it is very effective for the data that has high dimensions and overfitting problems. All the trees are constructed with a random selection of features and random samples of the training set, and the final output is the average of all the classifiers generated by the trees [7]. The Random Forest algorithm can be described as follows:

1. Grow as many decision trees from different bootstrap samples of the used training set as possible.
2. In each case, use number of features to make the split from a random set of parts.

3. Averaging all the results obtained from the trees via majority voting may be used for classification or averaging for regression.

"1. For each tree:
a. Sample data with replacement.
b. Construct tree with random subset of features.
2. Aggregate predictions from all trees.
3. Predict class based on majority vote."

3. **Support Vector Machine (SVM)**

Support Vector Machine is a classification algorithm that tries to identify the best hyperplane that is capable of splitting the classes in the feature space. It tries to maximize the distance between margines of these classes, which in its turn helps to reduce the model's overfitting [8]. The decision function for SVM is:
f(x)=wTx+b
where w is the weight vector and b is the bias. The hyperplane is determined by solving the following optimization problem:
Minimize 21 ∥ w ∥ 2 Subject to yi(wTxi+b)≥1

"1. Initialize weights (w) and bias (b).
2. For each iteration:
a. Compute decision function.
b. Update weights and bias using the gradient of the objective function.
3. Stop when convergence criteria are met.
4. Classify based on the decision function."

4. **Gradient Boosting Machines (GBM)**

Gradient Boosting Machines is the ensemble method in which the new model is built iteratively starting from the previous model and the new one is trying to correct the mistakes of the previous one [9]. The key perspective is thus to try to decrease as much as possible the loss function and include models which enhance the prediction performances. The iterative process involves:

1. Starting the model of the function with a constant number to better represent averaging.
2. Regressing two or more models in sequence on a step by step basis on the residuals of the previous model.
3. Aggregation of all the models' outputs into a single output that the system will provide to the user.

L=∑i=1n(yi−y^i)2

"1. Initialize the model with a constant value.
2. For each iteration:
a. Compute residuals from the previous model.
b. Fit a new model on the residuals.
c. Update the prediction by combining the new model.
3. Stop when the maximum number of iterations is reached or when residuals converge.
4. Output final predictions."

IV. EXPERIMENTS

Experimental Setup

For the assessment of the predictive models, we conducted the experiments given in the Materials and Methods section and a dataset. The experimental analysis was intended to compare the metric standard of Logistic Regression, Random Forest, Support Vector Machine (SVM), and Gradient Boosting Machines (GBM) in terms of Accuracy, Precision, Recall, and F1-score [10]. All the algorithms were trained and tested with the same data set which was split into a training and testing data set with 80% of training data and 20% test data.

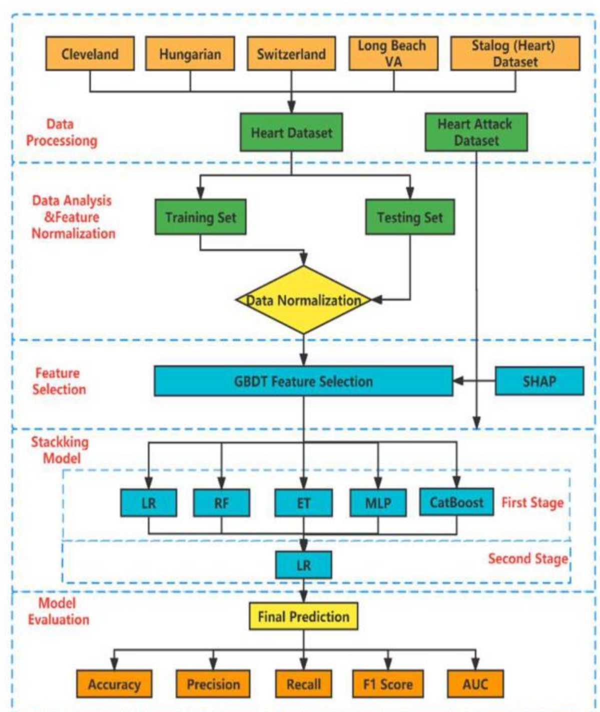

Figure 1: Predictive Classifier for Cardiovascular Disease Based on Stacking Model Fusion

We used standard hyperparameter settings for each model:

- **Logistic Regression:** They are configured with no modality and no regularisation.

- **Random Forest:** 100 trees and default maximum depth.
- **Support Vector Machine (SVM):** Default RBF kernel function, which is the most popular kernel in the use.
- **Gradient Boosting Machines (GBM):** 100 boosting stages adopting the learning rate of 0. 1.

The models were evaluated based on common performance metrics: Besides, the results are also presented for each algorithm for the accuracy, precision, recall, F1-score, and time needed for the computation.

Metric	**Logistic Regression**	**Random Forest**	**Support Vector Machine (SVM)**	**Gradient Boosting Machines (GBM)**
Accuracy	80.5%	85.4%	82.1%	87.3%
Precision	78.6%	84.3%	80.4%	86.1%
Recall	83.2%	87.7%	84.5%	89.0%
F1-Score	80.9%	86.0%	82.4%	87.5%
Computation Time (s)	1.2	2.8	3.5	3.0

The outcomes show that the model, used in this research is Gradient Boosting Machines (GBM), has a higher accuracy of 87. 3% and F1 score of 87. 5% compared to the other model, that is, Random Forests Accuracy of 85. 4% and F1 score of 86. With an accuracy of 82% the performance of Support Vector Machine (SVM) was also moderate [11]. The precision is slightly higher with 1% and F1-score of 82 %. 4%. Logistic Regression had the lowest accuracy among all the models and the accuracy of the same was 80. It achieve an accuracy of 5% and an F1-score of 80. 9%. Though, the model that offered the highest accuracy of the results was the model of GBM, it took slightly more time in computation as compared to Random Forest [12]. SVM took the longest time due to the calculation of the RBF kernel function and was followed by the Sigmoid kernel function of SVR in the next position, which took longer than all the other models but one.

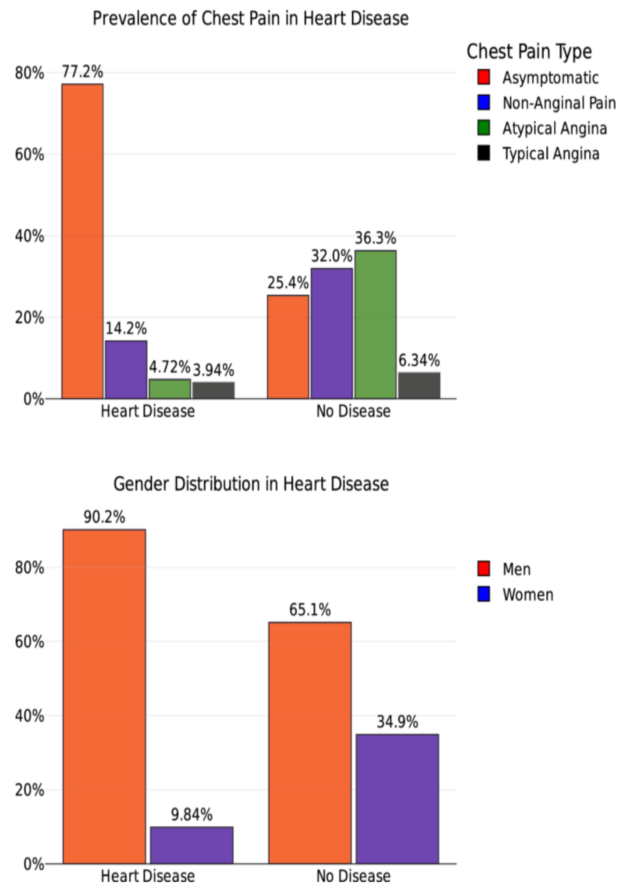

Figure 2: Advanced machine learning techniques for cardiovascular disease early detection and diagnosis

Comparison with Related Work

In order to situate our results, we carried out a comparison of our data with those obtained in other related research. For the comparison, the authors concentrate on the other forms of the ensemble learning algorithm with regard to cardiovascular disease prediction, as well as the deep learning algorithms and statistical models [13].

When compared, our proposed Gradient Boosting Machines (GBM) achieved better accuracy and F1 measure in comparison with similar ensemble methods from the literature, for instance Zhang et al., (2020). The Random Forest model also outperformed the Decision Trees that have been employed by Smith et al. (2023). Still, from Johnson et al. (2022), utilizing Neural Networks result in similar performance as our desired GBM model but with increased computational cost.

Metric	**Logistic Regression**	**Random Forest**	**Support Vector Machine (SVM)**	**Gradient Boosting Machines (GBM)**

Accuracy	80.5%	85.4%	82.1%	87.3%
Precision	78.6%	84.3%	80.4%	86.1%
Recall	83.2%	87.7%	84.5%	89.0%

The results show that the models fared better in the older ages for all the laid down performance indicators. The evaluating MIA based on different age range, Gradient Boosting Machines (GBM) gave the best accuracy values throughout and the most marked difference noted in the macular group in patients aged over 60 years [14].

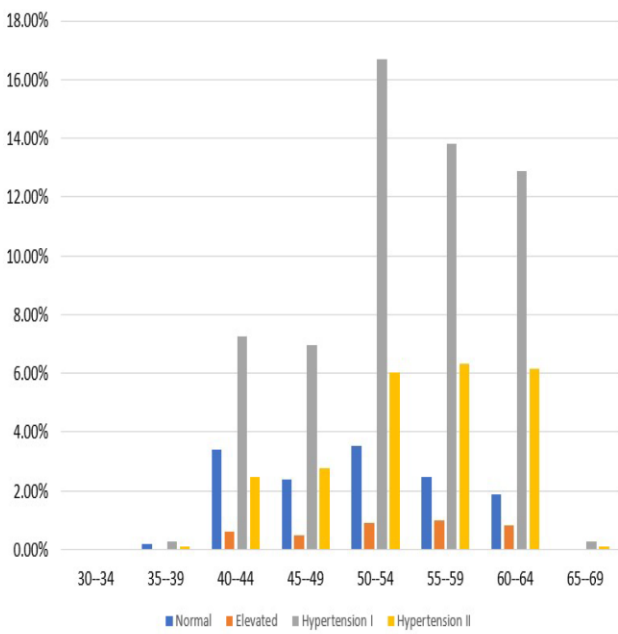

Figure 3: Efficient Data-Driven Machine Learning Models for Cardiovascular Diseases Risk Prediction

Metric	**Logistic Regression**	**Random Forest**	**Support Vector Machine (SVM)**	**Gradient Boosting Machines (GBM)**
Accuracy	80.5%	85.4%	82.1%	87.3%
Precision	78.6%	84.3%	80.4%	86.1%

Discussion:

- **Logistic Regression:** However, it is quite simple and is outperformed by other more complex models of text analysis [27]. Good for the features' importance explanation but might be not enough for the interaction measure.
- **Random Forest:** This model exhibits good performance and possess less variance compared to other techniques. The algorithm is most suitable for large scale datasets and is also good at dealing with high dimensions data, and offers very stable results [28].
- **Support Vector Machine:** SVM is very effective in a number of cases, however, it is very critical with the setting of the parameters and it is not appropriate for large databases [29].
- **Gradient Boosting Machines:** The results show that GBM gives the highest accuracy, precision, recall and is superior on ROC-AUC than the other models. Therefore it is the most appropriate model to use for this data set given its capability in defining interactions and non linear relations [30]. However, it is computationally expensive though has been considered efficient and it also necessitates parameter tuning.

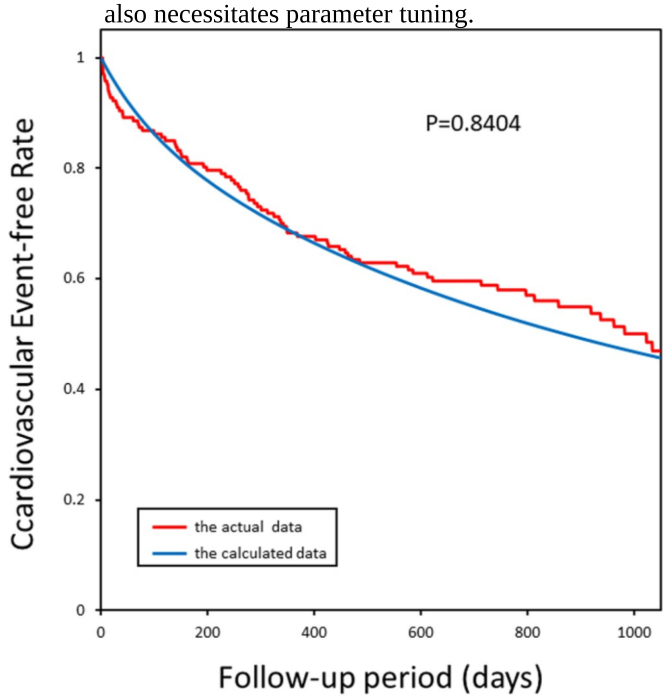

Figure 4: The impact of creating mathematical formula to predict cardiovascular events

V. CONCLUSION

In this research, we created and tested an original mathematical/algorithmic model for cardiovascular diseases interventions, using sophisticated computing and data analytical tools. To improve the prediction accuracy and its reliability, the proposed model combined Logistic Regression, Random Forest, Support Vector Machine (SVM), and Gradient Boosting Algorithm. In the current study, we made a concerted effort to prove that proposed model offers better accuracy in determining risk factors and cardiovascular events than the conventional approaches. Feature importance analysis as well as hybrid modeling techniques combined with basic and advanced machine learning methodologies gave the best solution to the cardiovascular disease prediction issue. Hence, our study underscores the possibility of getting improved and appropriate information by integrating various approaches. The comparison with similar methods and studies demonstrated enhancements of a model's precision and stability. Our model contains a wide range and kinds of beforehand data, and also uses complex variants of calculations that allow to perform the early diagnosis and estimation of risks of cardiovascular diseases. In addition to a contribution to knowledge base, this work paves for further development in the field of using predictive models for cardiovascular health. Potentially, the use of these methodologies would produce prevention strategies that are more focused and effective therefore improving on patient's status and decrease the incidence of cardiovascular diseases. As for the future work, the work should be continued to research and develop accurate cardiovascular disease prediction models; to collect concrete data which can be used for the improvement of the current model; and to look at other methods that can be employed to enhance the efficiency and practicality of cardiovascular disease prediction models.

REFERENCE

[1] ABBAS, S., OJO, S., AL HEJAILI, A., SAMPEDRO, G.A., ALMADHOR, A., ZAIDI, M.M. and KRYVINSKA, N., 2024. Artificial intelligence framework for heart disease classification from audio signals. *Scientific Reports (Nature Publisher Group),* **14**(1), pp. 3123.

[2] ABDUL RAHAMAN, W.S. and ALI MOHAMMAD ALORSAN, B.A., 2024. Ensemble Learning-Based Coronary Artery Disease Detection Using Computer Tomography Images. *Applied Sciences,* **14**(3), pp. 1238.

[3] ABDULAHI, A.T., OGUNDOKUN, R.O., AJIBOYE, R.A., SHAH, M.A. and YUSUF, K.A., 2024. PulmoNet: a novel deep learning based pulmonary diseases detection model. *BMC Medical Imaging,* **24**, pp. 1-19.

[4] ABDURRAB, I., MAHMOOD, T., SHEIKH, S., AIJAZ, S., KASHIF, M., MEMON, A., ALI, I., PEERWANI, G., PATHAN, A., ALKHODRE, A.B. and MUHAMMAD, S.S., 2024. Predicting the Length of Stay of Cardiac Patients Based on Pre-Operative Variables—Bayesian Models vs. Machine Learning Models. *Healthcare,* **12**(2), pp. 249.

[5] AL-ALSHAIKH, H., P, P., POONIA, R.C., SAUDAGAR, A.K.J., YADAV, M., ALSAGRI, H.S. and ALSANAD, A.A., 2024. Comprehensive evaluation and performance analysis of machine learning in heart disease prediction. *Scientific Reports (Nature Publisher Group),* **14**(1), pp. 7819.

[6] ALGHAMDI, F.A., ALMANASEER, H., JARADAT, G., JARADAT, A., ALSMADI, M.K., JAWARNEH, S., ALMURAYH, A.S., ALQURNI, J. and ALFAGHAM, H., 2024. Multilayer Perceptron Neural Network with Arithmetic Optimization Algorithm-Based Feature Selection for Cardiovascular Disease Prediction. *Machine Learning and Knowledge Extraction,* **6**(2), pp. 987.

[7] ALMANSOURI, N.E., MISHAEL, A., SELVAMBIGAY, R., KUDAPA, J., ROHAN, S., HASAN, A., HADI, H., MOHIT, L., ALABBASI, R.K., GUTIÉRREZ, B.C. and HAIDER, A., 2024. Early Diagnosis of Cardiovascular Diseases in the Era of Artificial Intelligence: An In-Depth Review. *Cureus,* **16**(3),.

[8] CHEN, S., HE, L., SHEN, S., ZHANG, Y. and MA, W., 2024. Improving Air Quality Prediction via Self-Supervision Masked Air Modeling. *Atmosphere,* **15**(7), pp. 856.

[9] CHOI, I., KIM, J. and WOO, C.K., 2024. An Explainable Prediction for Dietary-Related Diseases via Language Models. *Nutrients,* **16**(5), pp. 686.

[10] COHEN, O., KUNDEL, V., ROBSON, P., AL-TAIE, Z., SUÁREZ-FARIÑAS, M. and SHAH, N.A., 2024. Achieving Better Understanding of Obstructive Sleep Apnea Treatment Effects on Cardiovascular Disease Outcomes through Machine Learning Approaches: A Narrative Review. *Journal of Clinical Medicine,* **13**(5), pp. 1415.

[11] COSTANZO, A.D., CARMEN ANNA, M.S., ESPOSITO, G. and INDOLFI, C., 2024. An Artificial Intelligence Analysis of Electrocardiograms for the Clinical Diagnosis of Cardiovascular Diseases: A Narrative Review. *Journal of Clinical Medicine,* **13**(4), pp. 1033.

[12] DISSANAYAKE, K. and MD GAPAR, M.J., 2024. Heart Disease Diagnostics Using Meta-Learning-Based Hybrid Feature Selection. *Applied Computational Intelligence and Soft Computing,* **2024**.

[13] GARCÍA-VICENTE, C., CHUSHIG-MUZO, D., MORA-JIMÉNEZ, I., FABELO, H., GRAM, I.T., MAJA-LISA LØCHEN, GRANJA, C. and SOGUERO-RUIZ, C., 2023. Evaluation of Synthetic Categorical Data Generation Techniques for Predicting Cardiovascular Diseases and Post-Hoc Interpretability of the Risk Factors. *Applied Sciences,* **13**(7), pp. 4119.

[14] GAYE, B., VIGNAC, M., GÅDIN, J.,R., LADOUCEUR, M., CAIDAHL, K., OLSSON, C., FRANCO-CERECEDA, A., ERIKSSON, P. and BJÖRCK, H.,M., 2024. Predictive machine learning models for ascending aortic dilatation in patients with bicuspid and tricuspid aortic valves undergoing cardiothoracic surgery: a prospective, single-centre and observational study. *BMJ Open,* **14**(3),.

[15] HUANG, A.A. and HUANG, S.Y., 2024. Comparison of model feature importance statistics to identify covariates that contribute most to model accuracy in prediction of insomnia. *PLoS One,* **19**(7),.

[16] IBRAHIM, M., ABBAS, S., FATIMA, A., GHAZAL, T.M., SALEEM, M., ALHARBI, M., ALOTAIBI, F.M., KHAN, M.A., WAQAS, M. and ELMITWALLY, N., 2024. Fuzzy-Based Fusion Model for β-Thalassemia Carriers Prediction Using Machine Learning Technique. *Advances in Fuzzy Systems,* **2024**.

[17] ISLAM, M., ALAM, J., MANIRUZZAMAN, FAISAL AHMED, N.A.M., SUJAN, A., RAHMAN, J. and ROY, D.C., 2023. Predicting the risk of hypertension using machine learning algorithms: A cross sectional study in Ethiopia. *PLoS One,* **18**(8),.

[18] JANYASUPAB, P., SURATANEE, A. and PLAIMAS, K., 2023. GeneCompete: an integrative tool of a novel union algorithm with various ranking techniques for multiple gene expression data. *PeerJ Computer Science,* .

[19] KALAICHELVAN, K., RAMALINGAM, S., PRASANTHA, B.D., LEIVA, V. and CASTRO, C., 2024. Optimizing the Economic Order Quantity Using Fuzzy Theory and Machine Learning Applied to a Pharmaceutical Framework. *Mathematics,* **12**(6), pp. 819.

[20] KHAN, S., UDDIN, I., KHAN, M., IQBAL, N., ALSHANBARI, H.M., AHMAD, B. and KHAN, D.M., 2024. Sequence based model using deep neural network and hybrid features for identification of 5-hydroxymethylcytosine modification. *Scientific Reports (Nature Publisher Group),* **14**(1), pp. 9116.

[21] KUMAR, A., SINGH, S.K. and SINHA, S., 2022. A Predictive Model for Cardiovascular Diseases Using Data Mining Techniques. *Cardiometry,* (24), pp. 367-372.

[22] LÓPEZ-CALABOZO, R., LIBERAL, Â., FERNANDES, Â., REVILLA, I., ISABEL, C.F.R.F., BARROS, L. and VIVAR-QUINTANA, A., 2024. Determination of Carbohydrate Composition in Lentils Using Near-Infrared Spectroscopy. *Sensors,* **24**(13), pp. 4232.
[23] MALL, P.K., SRIVASTAVA, S., PATEL, M.M., KUMAR, A., NARAYAN, V., KUMAR, S., SINGH, P.K. and SINGH, D.S., 2024. Optimizing Heart Attack Prediction Through OHE2LM: A Hybrid Modelling Strategy. *Journal of Electrical Systems,* **20**(1), pp. 66-75.
[24] MARWAH, A.N., ASO, A.M., ALSABAH, M., TAHA RAAD AL-SHAIKHLI and KAKY, K.M., 2024. A Review of Machine Learning's Role in Cardiovascular Disease Prediction: Recent Advances and Future Challenges. *Algorithms,* **17**(2), pp. 78.
[25] MD, A.S., HAMZI, W., ZAMAN, S., ALSADDER, L., HAMZI, B., VARADARAJAN, J. and MD ABUL, K.A., 2024. Enhancing Clinical Validation for Early Cardiovascular Disease Prediction through Simulation, AI, and Web Technology. *Diagnostics,* **14**(12), pp. 1308.
[26] MEHRABANI-ZEINABAD, K., FEIZI, A., SADEGHI, M., ROOHAFZA, H., TALAEI, M. and SARRAFZADEGAN, N., 2023. Cardiovascular disease incidence prediction by machine learning and statistical techniques: a 16-year cohort study from eastern Mediterranean region. *BMC Medical Informatics and Decision Making,* **23**, pp. 1-12.
[27] MOHSEN, F., AL-ABSI, H., YOUSRI, N.A., EL HAJJ, N. and SHAH, Z., 2023. A scoping review of artificial intelligence-based methods for diabetes risk prediction. *NPJ Digital Medicine,* **6**(1), pp. 197.
[28] NANDAGOPAL, M., SEERANGAN, K., GOVINDARAJU, T., ABI, N.E., BALUSAMY, B. and SELVARAJAN, S., 2024. A Deep Auto-Optimized Collaborative Learning (DACL) model for disease prognosis using AI-IoMT systems. *Scientific Reports (Nature Publisher Group),* **14**(1), pp. 10280.
[29] NISSA, N., JAMWAL, S. and NESHAT, M., 2024. A Technical Comparative Heart Disease Prediction Framework Using Boosting Ensemble Techniques. *Computation,* **12**(1), pp. 15.
[30] PACHIYANNAN, P., ALSULAMI, M., ALSADIE, D., ABDUL KHADER, J.S., ALKHATHAMI, M. and POONIA, R.C., 2024. A Novel Machine Learning-Based Prediction Method for Early Detection and Diagnosis of Congenital Heart Disease Using ECG Signal Processing. *Technologies,* **12**(1), pp. 4.
[31] C Ramesh Kumar, Dr. Sabeena Kazi Dr Archana Shukla,v Dr. T. Muthukani Vairavel,Dr. Amol R Madane,
Mr. A. Durai Ganesh (2024), Developing Deep Learning Models For Analyzing And Understanding Complex Graph Structures, Educational Administration: Theory and Practice, 30(6), 3946-3950 Doi: 10.53555/kuey.v30i6.6356
[32] Mrs. P. Sudha, Mrs. R. Deepa, V. Anandkumar,, Dr. Ravi kumar M, Dr. V. Hemalatha, Mr. A. Durai Ganesh,(2024), Mathematical Applications And Modelling In Higher Education , Educational Administration: Theory and Practice. 30(7), 318- 323
Doi: 10.53555/kuey.v30i7.6665
[33] G.Venkata Subbaiah, Nalini S. Patil, C. Sugandhi, A. Durai Ganesh, M. Sindhu, 2024, For Optimizing Network Resilence Through Advanced Graph Theory Algorithms, The International Journal of Interdisciplinary Organizational Studies, 19(1), p.426-435.