Grenze International Journal of Engineering and Technology, June Issue

Harnessing Deep Learning for Cardiovascular Risk: A Neural Network Revolution

Lakshmi MJ[1], Ashwini Kodipalli[2], Trupthi Rao[3], Gargi N[4], Rohini B R[5] and Ushasree A[6]
[1,4-5]Dept. of AI&DS, Global Academy of Technology, Bengaluru, Karnataka, India
Email: mjlakshmi66@gmail.com, gargi@gat.ac.in, rohini.br@gat.ac.in
[2]School of Computer Science Engineering, RV University, Bengaluru, Karnataka, India
Email: ashwinik@rvu.edu.in
[3]Dept. of CSE (AI&ML), Dayananda Sagar University, Bengaluru, Karnataka, India
Email: trupthirao-aiml@dsu.edu.in
[6]Dept. of CSE, Gokaraju Lailavathi Engineering College, Hyderabad, India
Email: ushasreemishra@gmail.com

***Abstract*—Cardiac disorders are elevating reason for demise worldwide, claiming millions of lives annually. Early risk assessment and prevention are vital, as CVD is often linked to modifiable lifestyle factors. Leveraging a comprehensive dataset from Kaggle, featuring 12 key attributes, we developed an Artificial Neural Network (ANN) for binary classification to predict CVD risk. Our model, consisting of 6 hidden layers with 512-16 neurons, noted 82% result using the ADAM optimizer and binary cross-entropy loss. indicating a high degree of predictive capability. The ROC curve demonstrated high sensitivity and specificity, while the Loss vs Accuracy graph showed consistent improvements in both learning and validation accuracy. Our ANN model draws precious enlightenment for early CVD risk assessment, enabling potential interventions to reduce mortality rates. This deliberation showcases the inherent of Neural approach in tackling the growing universal epidemic of CVD. The use of neural networks in this context highlights the potential of neural learning approaches in tackling complex healthcare problems.**

***Index Terms*— CVD risk assessment, Artificial Neural Network, ROC curve, Deep Learning.**

I. INTRODUCTION

Cardiovascular disease (CVD) reigns as the leading cause of death worldwide, presenting a paramount health challenge across the globe.WHO underscores that cardiovascular disease (CVD) accounts for 17.9M yearly tolls, with projections anticipating a staggering escalation to 75 million by 2030. Concurrently, the AHA records that a heart attack transpires every 40 seconds within the US. Ethology of many CVD cases is intricately associated with modifiable lifestyle determinants, including tobacco utility, decrease nutrients, physical inactivity, and excessive alcohol intake. Cardio-vascular disease comprises a wide range of challenges impacting blood and vessels, and it can be labelled into four diverse forms-CHD, TIA, PAD and AD. CHD referring to C-Heart disorder includes the blockage and restricts the circulatory flow in Heart. TIA referring to Strokes and Transient Ischemic Attacks causing brief interruptions in circulation and causes high implications on easy blood flux in brain.

Grenze ID: 01.GIJET.12.2.224_21

PAD referring to Peripheral Arterial Disease resulting in ulcers pain and weakness majorly in arm and leg along with hinder blood flow. This affect the heart, including coronary artery disease, which is the blockage or restriction of blood flow to the heart, leading to angina or myocardial infarction Also known as "mini-strokes," TIAs are temporary interruptions in blood flow to the brain, lasting less than 24 hours. TIAs can cause brief episodes of neurological dysfunction, such as weakness, numbness, or difficulty speaking. blood vessels outside of the heart and brain, leading to reduced blood flow to the extremities, such as the legs and arms. PAD can cause symptoms like leg pain, cramping, weakness, and ulcers. AD can affect various arterial beds, including coronary, carotid, and peripheral arteries. Its very important to know the arteries flow and vascular fluids in prior to abort the heart abnormalities' research revolves around to propose a solid inference of features causing abnormalities.

In our study, we leverage a comprehensive tabular dataset from the Kaggle repository, distinguished by 12 essential features. This dataset underpins our Artificial Neural Network (ANN), facilitating precise binary classification to interpret the vulnerable Cardiac Condition The intricate data attributes enhance the predictive capability of our models, offering valuable insights into CVD risk assessment. We came across distinguished ways for building strong and justified scrutiny ranging options like activation types, error computation and deducing permutated optimizers also supported with visuals and standard criteria evaluators. These approaches were supported by visualizations and standard evaluation metrics, enabling us to develop a strong and justified scrutiny model.

II. Literature Survey

M.Swathy and K.Saruladha [1] carried a predictive ML and DL investigation on CVD causing lethal death. Their Research revolved around Various CVD causing factors like obese, HBP & consumption of tobacco.it captured the risk in prior with analytics. Supported with solid 92.1% from ML & 91% from DL. Awais Mehmood et.al [2] developed a Convolutional paradigm for Heart Attack by preparing the datasets and using state-of-the-art technique. They gave Cardio Help that generates the chance of facing CVD and their approach recorded 97% accuracy. K.Saikumar & V.Rajesh [3] instigated Radiology dataset for uncovering the cause of CAD risking the heart disease. Their approach emphasized more of Machine and Deep techniques and achieved the productivity of 98.69%. Javed Azmi et.al [4] compiled a study review on CVD detection so that Practitioners get some precautional rings and get the proper aids to prevent heart attack. N. Komal Kumar et.al [5] built a ML model fed with conditions like blocked veins, previous history of attacks, angina. Researcher aimed at aiding CVD by foreseeing things as earlier as possible and the ROC AUC was noted 0.865 with precision of 85%. Yahia Baashar et.al [6] carried a Meta-analysis and compared various supervised machines and neural approach for detecting CVD. At early stages.

The performance achieved from best technique was 91%.Sudarshan Nandy et.al [7] researched on predictive CVD issues to take precative measures to avert the chance of Heart attack and accounted the best score of 95.78%.Aizatul Shafiqah Mohd Faizal et.al [8] reviewed models in two approaches including conventional & AI incorporated biomarkers for CVD risks and assessed multi-modality of Big Data.R. Jane Preetha Princy et.al [9] composed ML model fed with consolidated huge medical datasets from clinics. Operated many techniques and ranked a 73% accuracy by Dtree help in guiding an appropriate treatment. Lubna Riyaz et.al [10] carried their analysis by integrating ML model & ANN neural model with Solid tabular dataset having important Features causing heart diseases and achieved best a performance of 86.91% in neural model. P. Ramprakash et.al [11] developed a method for identifying the principle parameter causing risk leading in heart health. Various clinical and medical data is thoroughly evaluated and using DNN and ANN presence and absence of HD is assessed. Md Ekramul Hossain et.al [12] surveyed type-II diabetes individuals and created a measure for predicting how often these are also target of CVD. The classifiers served a performance range of 79-88 percentage. M.D. Amzad Hossen et.al [13] conducted predictive analysis on CVD illness by forecasting 14 features and operating Supervised ML achieving best score of 92.10%.Anirudda Dutta et.al [14] studied a clinical data to cultivate a model employing SVM and CNN classifiers for assessing prevailing of CHD and models accuracy noted as 79.5%.Pronab Ghosh et.al [15] researched serious illness CVD in Humans by applying Boosting and Classification techniques of ML and noted best accuracy 99%. Khatawakar et al. [16] employes several ML models for detection for heart failure. Rajput et al. [17] used ML classifiers for early detection of Myocardial Infraction. Tn et al. [18] predicted the occurrence of heart attack among young adults using ML algorithms. Shahane et al. [19] used ML models for detection and prediction of diabetes and also, they found its connection with pregnancy. Sagarnal et al. [20] used various DL models for detection of Covid-19 disease.

III. METHODOLOGY

ANN Algorithm:
Input: X, y, architecture, α, T
Initialize weights, biases, activation functions
For each epoch (t = 1 to T):
  For each sample in X:
    Forward pass:
      Compute output of each layer
      Compute predicted output (y)
    Backward pass:
      Compute error
      Compute gradients
      Update weights and biases
  Update weights and biases after each epoch
Output: Trained ANN model

Data Collection: An insightful tabular dataset from the Kaggle repository, tailored for cardiovascular disease (CVD) research, which features 12 key attributes: age, anaemia, creatinine phosphokinase levels, diabetes, ejection fraction, high bp, platelet measures, s-creatinine, serum sodium, sex, smoking status, and time.

Data Preprocessing: To handle the wide spectrum of values and heterogeneous data types effectively, we implemented scaling and normalization, ensuring consistency and enhancing the model's predictive capabilities. This removes outliers if present.

Model Preparation: Designing a neural network entail crafting its architecture: deciding on the layer count, neurons per layer, dropout rates, and choosing activation functions tailored for each layer. In our pursuit of accurately predicting cardiovascular disease through binary classification, the output layer is carefully set to employ either a sigmoid or tanh activation function, crucial for achieving optimal output scaling.

Basic NN constitute 3 layers – input as yellow, hidden as blue and green as output depicted in Figure 1.

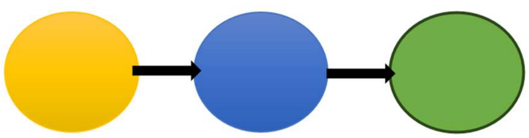


Figure 1. Architecture of the simple NN

Model Training: Examine the dataset both with and without dropouts to measure their influence on model performance. Monitor training accuracy over epochs, adjusting activation functions and batch sizes to fine-tune model training and enhance performance evaluation.

For better understanding dropouts' architecture, we have picturized input layer in yellow, hidden layer in blue output layer in green and black cross mark showing dropout in hidden layer and grey neuron showing dropout in input layer as in Figure 2.

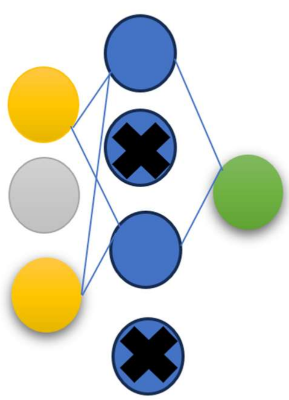


Figure 2. Dropouts

Model Evaluation: Assessed model accuracy and precision, generating detailed reports and visualizing results with ROC curves and error loss graphs for thorough performance analysis.

$$\text{Accuracy}=\frac{\text{TN+}}{\text{TN+TP+FN+F}} \quad (1)$$

$$\text{Precision}=\frac{\text{TP}}{\text{FP+TP}} \quad (2)$$

$$\text{Recall}=\frac{\text{TP}}{\text{FN+TP}} \quad (3)$$

$$\text{F1-Score}=\frac{\text{2TP}}{\text{FN+FP+2TP}} \quad (4)$$

Where, TN=true negative, TP=true positive, FN=false negative, FP=false positive.

IV. RESULTS

Our work leverages deep NN approach for assessing hazard effect of cardiac disfunction. ANN in general comprises three layers input hidden and output layer. Input layer constitutes neurons depending on parameters in dataset that includes 12 features. Our best proposed architecture includes 6 hidden layers comprising neurons ranging from (512–16) respectively. Since this is a binary classification model, the output layer has only 1 neuron. In Table 1 various combination of models is indexed ranging from dropouts, number of hidden layers, number of neurons in each and activation function for each.

TABLE I. SUMMARIZED RESULTS OF COMBINATIONAL MODELS

MODEL NO.	INPUT Dropout/ Hidden Dropout	H- Layer	Optimizer	Epochs	Accuracy
1	N	4	ADAM	200	71%
2	N	6	SGD	200	73%
3	Y	6	SGD-WITH GRADIENT	200	76%
4	Y	4	ADAGRAD	300	75%
5	Y	6	ADAGRAD	300	74%
6	Y	6	RMSPROP	300	69%
7	Y	6	ADADELTA	300	71%
8	Y	4	SGD-WITH MOMENTUM	300	79%
9	Y	6	MINI-BATCH SGD	300	78%
10	Y	6	ADAM	300	82%

From Table I model-10 having 6 hidden layers utilizes ADAM optimizer, Binary_crossentropy for loss and accuracy for metrics it executed for 300 epochs accounting a performance of 82%In Fig 3 and 4 Roc curve is used to graphically analyze the performance of B-classification model, plotting is between FPR and TPR according to changing weights gives information about sensitivity & specificity. Supported by graph for outstanding model 10 and 8 with better score.

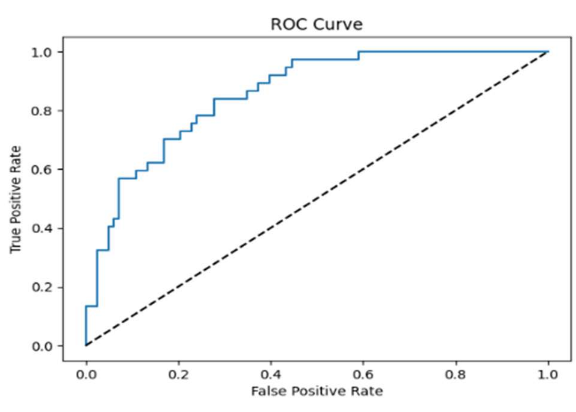


Figure 3. ROC of Model-10

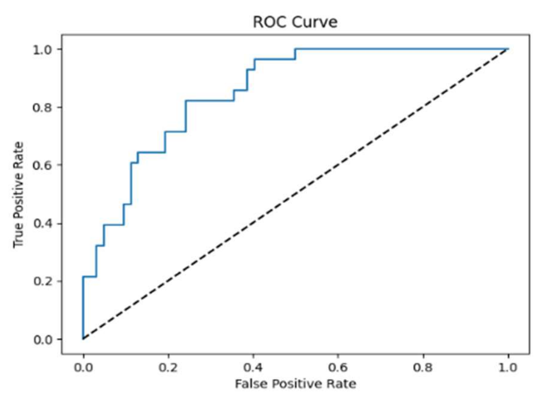
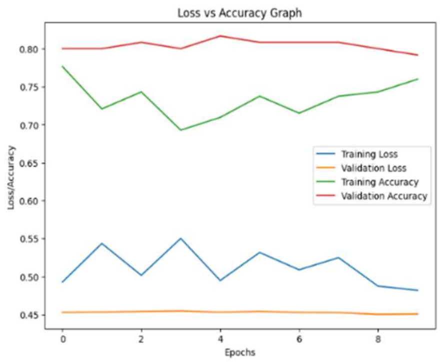


Figure 4. ROC of Model-8

In Figure 5 and 6 Loss vs Accuracy graph is visualized to depict the variation in training & testing accuracy, training and testing loss.



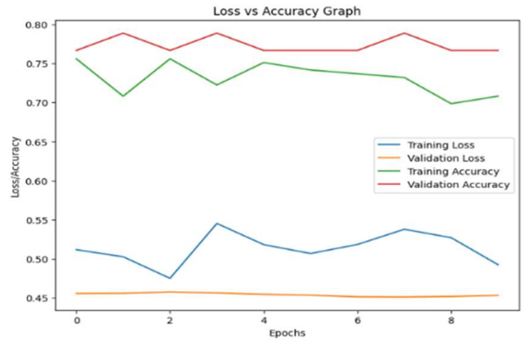
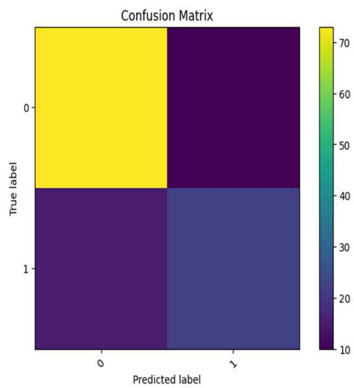


Figure 5. L VS A of model-10

Figure 6. L VS A of model-8

To get more better understanding about actual and computed values confusion metrics is utilized as represented in Figure 7 & 8.



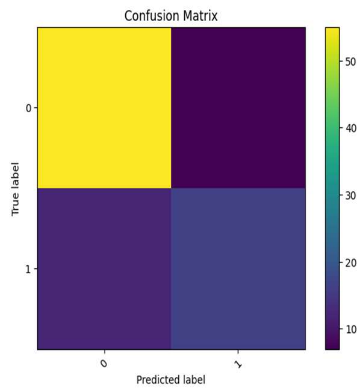
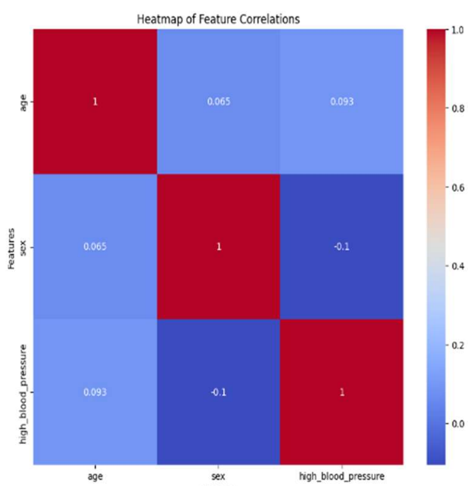


Figure 7. C-metric of model-10

Figure 8. C-metric of model-8

Correlation heatmap is used to analyse the relationship between various parameters present in structured data assembled from Kaggle repository. In Figure 9 correlation between age, sex and high bp can be visualized.



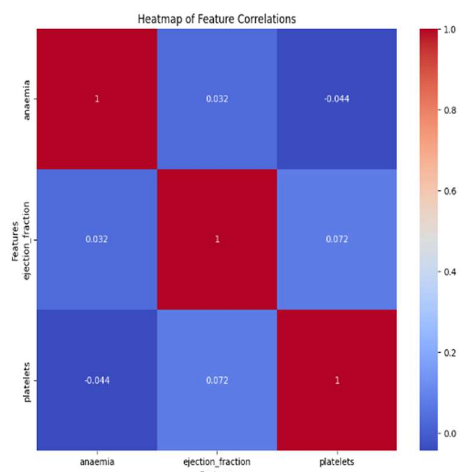


Figure 9. Age vs Sex vs HBP

Figure 9. Anaemia vs platelets vs EF

In Figure 10 correlation between anemia, EF and platelets can be visualized. In Figure 11 correlation between high bp, serum_creatinine creatinine_phosphokinase can be visualized.

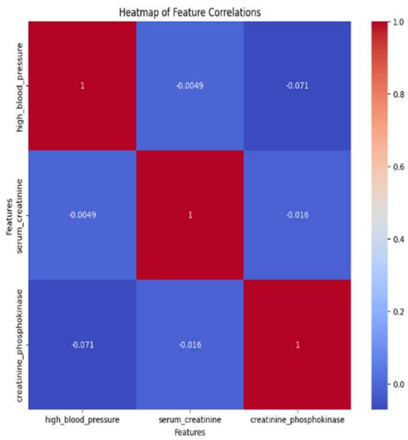


Figure 11. HBP vs SC vs CP

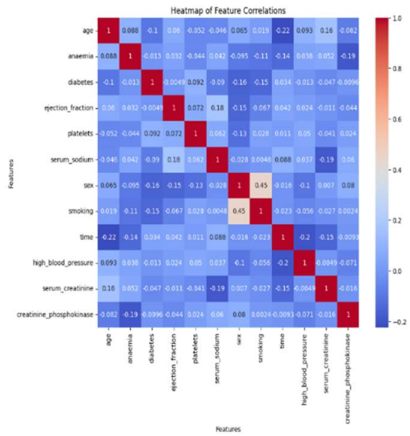


Figure 12. Heatmap for all 12 parameters

Relationship between all 12 features displayed using Heatmap in Figure 12.

VI. Conclusion

In summary, cardiovascular disease (CVD) stands as a significant universe health epidemic, accountable for millions of deaths annually. With its complex spectrum, including coronary heart disease (CHD), Transient Ischemic Attack (TIA), Peripheral Artery Disease (PAD), and Aneurysmal Disease (AD), precise risk assessment becomes crucial. Our study harnessed a vast dataset also an Artificial approach involving neurons to craft a classification method for assessing CVD risk. This innovative model, featuring 6 hidden layers and 12 input features, achieved an impressive 82% accuracy. It offers vital insights into CVD risk, paving the way for early interventions and potentially saving countless lives. By leveraging the power of neural learning and information analytics, we are poised to tackle the escalating CVD threat and improve health outcomes worldwide. Our scrutiny included ROC curve, loss VS accuracy graph, confusion metrics and correlation Heatmap all these gave esteemed analysis on actual and computed values along with some solid overview of ranging TPR and FPR.

References

[1] Swathy, M., & Saruladha, K. (2022). A comparative study of classification and prediction of Cardio-vascular diseases (CVD) using Machine Learning and Deep Learning techniques. ICT Express, 8(1), 109-116.

[2] Mehmood, A., Iqbal, M., Mehmood, Z., Irtaza, A., Nawaz, M., Nazir, T., & Masood, M. (2021). Prediction of heart disease using deep convolutional neural networks. Arabian Journal for Science and Engineering, 46(4), 3409-3422.

[3] Saikumar, K., & Rajesh, V. (2024). A machine intelligence technique for predicting cardiovascular disease (CVD) using Radiology Dataset. International Journal of System Assurance Engineering and Management, 15(1), 135-151.

[4] Azmi, J., Arif, M., Nafis, M. T., Alam, M. A., Tanweer, S., & Wang, G. (2022). A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data. Medical engineering & physics, 105, 103825.

[5] Kumar, N. K., Sindhu, G. S., Prashanthi, D. K., & Sulthana, A. S. (2020, March). Analysis and prediction of cardio vascular disease using machine learning classifiers. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 15-21). IEEE.

[6] Baashar, Y., Alkawsi, G., Alhussian, H., Capretz, L. F., Alwadain, A., Alkahtani, A. A., & Almomani, M. (2022). Effectiveness of artificial intelligence models for cardiovascular disease prediction: Network meta-analysis. Computational intelligence and neuroscience, 2022(1), 5849995.

[7] Nandy, S., Adhikari, M., Balasubramanian, V., Menon, V. G., Li, X., & Zakarya, M. (2023). An intelligent heart disease prediction system based on swarm-artificial neural network. Neural Computing and Applications, 35(20), 14723-14737.

[8] Faizal, A. S. M., Thevarajah, T. M., Khor, S. M., & Chang, S. W. (2021). A review of risk prediction models in cardiovascular disease: conventional approach vs. artificial intelligent approach. Computer methods and programs in biomedicine, 207, 106190.

[9] Princy, R. J. P., Parthasarathy, S., Jose, P. S. H., Lakshminarayanan, A. R., & Jeganathan, S. (2020, May). Prediction of cardiac disease using supervised machine learning algorithms. In 2020 4th international conference on intelligent computing and control systems (ICICCS) (pp. 570-575). IEEE.

[10] Riyaz, L., Butt, M. A., Zaman, M., & Ayob, O. (2022). Heart disease prediction using machine learning techniques: a quantitative review. In International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 3 (pp. 81-94). Springer Singapore.
[11] Ramprakash, P., Sarumathi, R., Mowriya, R., & Nithyavishnupriya, S. (2020, February). Heart disease prediction using deep neural network. In 2020 international conference on inventive computation technologies (ICICT) (pp. 666-670). IEEE.
[12] Hossain, M. E., Uddin, S., & Khan, A. (2021). Network analytics and machine learning for predictive risk modelling of cardiovascular disease in patients with type 2 diabetes. Expert Systems with Applications, 164, 113918.
[13] Hossen, M. A., Tazin, T., Khan, S., Alam, E., Sojib, H. A., Monirujjaman Khan, M., & Alsufyani, A. (2021). Supervised Machine Learning-Based Cardiovascular Disease Analysis and Prediction. Mathematical Problems in Engineering, 2021(1), 1792201.
[14] Dutta, A., Batabyal, T., Basu, M., & Acton, S. T. (2020). An efficient convolutional neural network for coronary heart disease prediction. Expert Systems with Applications, 159, 113408.
[15] Ghosh, P., Azam, S., Jonkman, M., Karim, A., Shamrat, F. J. M., Ignatious, E., ... & De Boer, F. (2021). Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. IEEE Access, 9, 19304- 19326.
[16] Khatawakar, D. R., Akhil, M. S., Thanushree, V. V., Kodipalli, A., Rao, T., & Rohini, B. R. (2023, June). Analysis and prediction of heart failure using machine learning algorithms. In 2023 International Conference on Computational Intelligence for Information, Security and Communication Applications (CIISCA) (pp. 255-260). IEEE.
[17] Rajput, K. S. C., Kumar, M., Kodipalli, A., BR, R., & Rao, T. (2023, June). Prediction of Myocardial Infraction Using Machine Learning Algorithms. In 2023 International Conference on Computational Intelligence for Information, Security and Communication Applications (CIISCA) (pp. 326-332). IEEE.
[18] Tn, K., Meghana, S., Kodipalli, A., Rao, T., & Kamal, S. (2023, March). Prediction of early heart attack possibility using machine learning. In 2023 2nd International Conference for Innovation in Technology (INOCON) (pp. 1-5). IEEE.
[19] Shahane, S., Siri, T. C., Bhoomika, R., Kodipalli, A., Rao, T., & Rohini, B. R. (2023, June). Prediction of Diabetes using Machine Learning Algorithms and Correlation with Pregnancy. In 2023 3rd International Conference on Intelligent Technologies (CONIT) (pp. 1-7). IEEE.
[20] Sagarnal, C., Devamane, S. B., Hosamani, R., & Rao, T. (2021). Deep Learning Approaches for COVID-19 Diagnosis. In IDDM (pp. 116-126).