Utilising Neural Networks and Genetic Algorithm for the Prediction of Cardiovascular Illness

Jay Panara
MEng, ECE
University of Waterloo, Canada
jpanara@uwaterloo.ca

Abstract—**Health problems are not unheard of in our societies. The health conditions that are becoming more prevalent each day are impacted by the modern way of life. The number of individuals who are dying from heart disease makes it one of the most concerning illnesses among all others. Hence, this project utilizes self-developed Genetic Algorithm and Neural Network to build an efficient system for predicting these cardiovascular illness in the individuals given the right inputs to the system. Here, the genetic algorithm acts as an optimizer which helps in getting the right weights for the neural network model to get the maximum accuracy in the predicted result. The proposed framework is trained on the dataset obtained from UCI Machine Learning Repository for the Heart Diseases. Additionally, a comparative study between the proposed genetic algorithm neural network-based model with a simple neural network model is performed to understand the efficiency of the proposed framework. The system is further compared and analysed with other state-of-the-art classical machine learning models which were used in [1]. The proposed pipeline has shown to surpass state-of-the-art methods in predicting the cardiovascular illness on various evaluation metrics such as accuracy, precision, recall and computation time.**

Index Terms—**Genetic Algorithm, Neural Networks, Data Preprocessing, Deep Learning, Cardiovascular Diseases, EDA, Machine Learning**

I. INTRODUCTION

The cardiovascular diseases are a set of illnesses that effect the blood vessels and heart. In this disease one or more than one area of blood vessels and heart are impacted. An individual may exhibit asymptomatic condition where there is no feeling about anything or a symptomatic conditions where there is a physical experience of the illness.

There are more than 12 million people worldwide who lose their lives due to cardiovascular illnesses. In most of the developed countries, cardiovascular illness plays a significant factor in deaths of the individuals and is considered the main cause of adult mortality.

High blood sugar, high cholesterol, obesity and hypertension are some of the major risk factors which might lead to cardiovascular illnesses to arise. These diseases sometimes resemble those of other conditions or link to the indications of ageing, making it difficult to detect.

A precise and effective diagnosis of a medical condition is a crucial yet challenging undertaking. In order to assist doctors to diagnose and treat patients more accurately, this system has to be automated. A medical system should address a variety of factors, including how medical information is represented and decision-making, choosing, and adaption of an appropriate model. Medical advancement is always supported by the field of data analysis which improves the level in the expertise of the medical professionals and builds a method for the treatment of a disease. An individual patient's risk level is defined by doctors with the use of a medical diagnosis system.

So in this project, the system that is developed would be beneficial for predicting the cardiovascular illness. The system utilizes the dataset that is obtained from the Heart Disease UCI Machine Learning repository which consists of 4 databases namely, Cleveland, Hungary, Switzerland and Long Beach. The proposed model is build up of neural network which consists of fundamental blocks called neurons. In a layer of a Neural Network model, all the neurons are connected with one another. The weights are applied when the input, in this case Heart Disease Data, is given to the neural network model. These weights regulate the signals between the neurons which explains the impact of input on the output that is produced [2]. Hence, the efficiency of the neural network model is dependent on values of weight assigned to each neuron. So, finding the right weight can be done manually but due to various possibilities, it makes the whole process time consuming to solve. In order to

make the task easier as well as efficient, the help of the optimizers is taken in order to get the right weights. One of the most efficient optimizers is the genetic algorithm which is developed and utilised in this neural network model development. The genetic algorithm is used for optimization and utilises the concepts of natural genetics in its operation. It solves the problems fast with acceptable quality. Thus, an improved model is developed by utilising the efficient optimization abilities of genetic algorithms and the learning capacity of the neural network which is used in forecasting heart diseases. Additionally, a comparative study of the performances of the proposed neural network model with the simple neural network model(without using Genetic Algorithm) is performed. Along with that the efficiency of the model is also evaluated with respect to the outcome in [1]. Various criteria such as accuracy, precision and F1-score are taken into consideration for the evaluation.

II. LITERATURE REVIEW

There has been a widespread interest in the field of machine learning and how these concepts would be utilised in developing better technologies for health and medicine. This led to research in the domain where a variety of machine learning algorithms and approaches are used. Prafula et al. [3] presented the idea of how machine learning can be implemented in assessing developed smart home sensors and predict the cognitive health of the person.

Genetic algorithm being one of the finest optimizers we could develop and is widely used in the optimization of search tools for challenging issues. It is not just limited to optimization but is widely preferred in the field of machine learning, research and development. Annu Lambora et al. [4] explained about the working of the genetic algorithm, the processes and the basic workflow of the algorithm that go into its development. It also explains the advantages of the Genetic Algorithm and where it is being utilised for practical purposes.

Another example where genetic algorithm proved to be useful is the system that was developed by Latha and Subramanium [5] which was CoActive Neuro-Fuzzy Inference System (CANFIS) and it integrated genetic algorithm in diagnosis of diseases of the heart. They used Neurosolution software for the simulation and utilized the Cleveland dataset for the analysis. The development in the field of genetic algorithm and its application in the domain of machine learning have significantly increased due to its ability to provide highly optimised outputs and requires less information in its development [6].

For the purpose of categorizing data mining tasks, a classifier system based on Neural Network was suggested by Hai et al. [7] Using 13 various datasets from UC-Irvine and another made up repository, they ran their trials. On 5 datasets, they observed that the performance of the neural network based system is comparable to the supervise learning classifier system while performing much better in 6 datasets and on 3 datasets significantly worse.

Liaqat et al. [8] put an equal emphasis on improving the features and removing the issues with overfitting and underfitting that were brought on by the prediction model. They suggested a χ^2 statistical model to filter out extraneous information when the ideally designed Deep Neural Network is sought out via an exhaustive search approach. By contrasting the proposed $\chi^2 - DNN$ model's performance with traditional DNN and ANN models, other cutting-edge machine learning models, and previously published approaches for heart disease prediction, the efficiency of the suggested hybrid model is assessed. In comparison to the earlier described methodologies, the results shown by the proposed model were encouraging.

Yung et al. [9] developed a neuro-genetic system which was used in the stock trading. From 1992 to 2004, the proposed system was tested on 36 businesses listed on the NASDAQ and NYSE. RNN along with genetic algorithm was utilized in the process which performed significantly in the purchase and hold strategy.

Niti et al. [10] developed a neural network based decision support system for the detection of cardiac disease. They used the data obtained from 78 patients for the training of their algorithm which avoids the mistakes done by humans.

Anbarasi et al. [11] developed a method from genetic algorithm for the selection of the feature subset in improving the prediction of heart disease. This method was accurate in predicting the existence of heart disease. The accuracy that was obtained was similar to the one they received previously even after the characteristics were reduced. In this method, they used three approaches for the prediction, namely, Decision Tree, Naive Bayes and Clustering. It was observed that the approach of decision tree outperformed the other two approaches.

Shanthi el al. [12] suggested a decision support system that used artificial neural networks with changing connection weights and genetic algorithm to forecast stroke illness. 150 individuals who showed the symptoms and signs of stroke illness provided the information for their research. The outcome demonstrates that the hybrid

strategy outperforms the conventional artificial neural networks and selection of features utilizing the genetic algorithm.

A prediction algorithm for the coronary artery heart disorder is proposed by Randa et al. [13]. The decision tree methods, such as the Fast Decision Tree and C4.5 pruned tree were employed in the development of the predictive algorithm. The model at the beginning were trained on every feature of the dataset and after that the top features were chosen from the dataset and then utilized in training the model. The strategy of using the best features in the decision tree improved the accuracy percentage from 75.48% to 78.6% and the pruned tree C4.5 method increased the prediction accuracy from 76.3% to 77.5%.

Through the use of artificial neural networks and data mining, Kumaraswamy et al. [14] developed a smart and successful cardiac attack prediction system. For the purpose of forecasting cardiac disease, they also suggested extracting noteworthy patterns. The data pertinent to heart attacks were extracted from the warehouse using K-means clustering. To extract the common patterns they employed the Maximal Frequent Itemset Algorithm (MAFIA) [14] [15].

This project used a self-developed genetic algorithm as an optimizer for obtaining the weights of the self-developed neural network model which was then utilized in prediction of the cardiovascular illness which was not done in the above literature. Hence, this research aimed at figuring out whether artificial life technique based optimizer such as the genetic algorithm be utilized in optimizing the weights of the Neural Network model to provide an efficient predictive output that can be used in the real world scenario, which also formed my research question.

III. METHODOLOGY

The goal of this project is to develop a distinctive framework that, via the help of artificial life, would be utilized in helping the healthcare sector for prediction of the dangerous cardiovascular illness. The framework of the proposed pipeline can be seen in Fig. 1. The detail description of each step is explained below:

A. Data:

Dataset used in this experiment is the Heart Disease Dataset available at UCI Machine Learning Repository. The information in the data files are collected from 4 different sources namely, Hungarian Institute of Cardiology, University Hospital-Zurich, University Hospital-

Basel and V.A. Medical Center-Long Beach and Cleveland Clinic Foundation [16]. All the data is combined for the project and then spilt at a ratio of 0.2 to get train and test data. The dataset contains 13 feature classes and 1 label class. The final label class specifies whether the patient has a cardiovascular illness or not.

The purpose of the developed framework is to correctly predict the cardiovascular illness when fed with new data to the model. Both the categories have almost same number of records. The dataset has been enhanced as compared to original dataset by checking for NULL/missing values and removing if any, label encoding, and normalisation.

B. Data Preprocessing:

The data preprocessing section has two parts. The first one is to check whether there are any missing or null values in the data and removing that instance from the dataframe. Followed by outlier detection in which interquartile range process has been used to remove the outliers.

- **Inter Quartile Range (IQR) Process:** In this process, Quartiles are created for the dataset. 10% of the data are considered to be below first quartile point while the second quartile point represents the dataset's median. The process of detecting the outliers is as follows:
  1) The first step is to find the first quartile that is considered as Q1.
  2) The next step is to find the third quartile which is denoted as Q3.
  3) After getting the values for Q1 and Q3, these need to subtracted to get the Inter Quartile Range - $IQR = Q3 - Q1$.
  4) The lower limit is given by, $Q1 - 1.5 * IQR$ and the upper limit is given by, $Q3 + 1.5 * IQR$.
  5) All the values which lie below the lower limit and above the upper limit are considered as outliers.
- **Min-Max Normalisation:** When preparing data for machine learning, the scaling technique known as normalisation is used for the conversion of the numeric column values of the dataset to a scale which is standardized. For every dataset in a model, it is not required. The necessity of this normalization only comes into the picture when, in a model, the ranges of the values in the features differ vastly. Despite the large number of feature normalisation methods available in machine learning, only a few of them are really utilised most frequently [17]. One
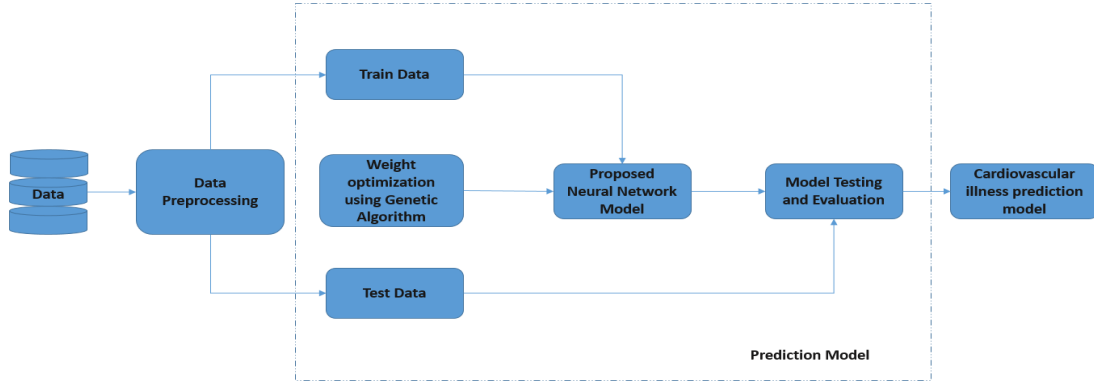
Fig. 1. Methodology

of it is Min-Max Normalization. Scaling is another name for this method. The dataset's attribute values are shifted and re-scaled with the aid of the Min-Max scaling algorithm, which results in the values being present from anything between 0 and 1. The formula for the Min-Max Normalization is given as follows:

$$X_n = \frac{X - X_{minimum}}{X_{maximum} - X_{minimum}}$$

where X_n is the value of normalization, $X_{maximum}$ is the maximum value of a feature and $X_{minimum}$ is the minimum value of a feature.

C. Weight optimization using Genetic Algorithm

An adaptive heuristic search algorithm known as the Genetic Algorithm was developed from the Theory of Evolution in Nature by Charles Darwin. In the field of machine learning, it is employed to address the optimization issues. It is a crucial algorithm since it speeds up the process of solving difficult issues that would otherwise take a lengthy time. The different phases [18] of the self-developed genetic algorithm for the proposed model is shown in Fig. 2 and detailed description of it is as followed:

- **Initialization:** The initialization phase of the genetic algorithm is to first generate the population. In the project, the first generation is the neural network model generated with random weights. This process marks the initialization stage of the process.
- **Fitness Assignment:** The next step is the fitness assignment. It refers to individual's capacity to compete with others and the individuals are rated according to their fitness scores. The greater the fitness score, greater are the chances of being selected.
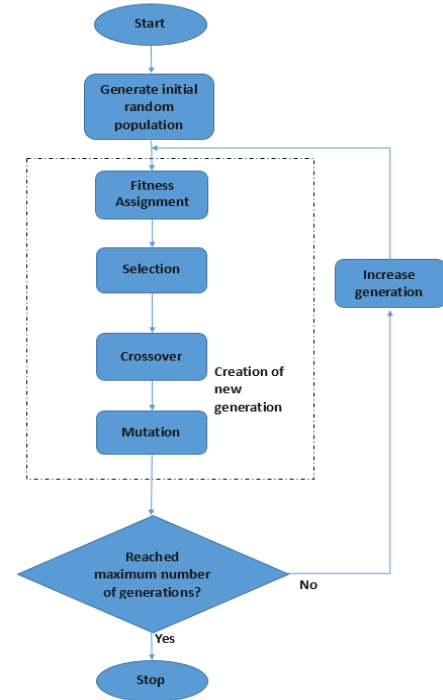

Fig. 2. Structure of Genetic Algorithm

In this case, all the neural network models from the pool is tested for its accuracy and the neural network pool is then sorted with respect to its fitness or accuracy obtained.

- **Selection:** In this phase, the individual with the highest fitness is selected. Here, the top two neural network with the highest fitness are selected for the next stage.
- **Crossover:** The crossover is the crucial part of the process. In this phase, a random crossover point is chosen and after that point there is a switch in the

genetic information of the two parents to create a new offspring. In this project, a random point is selected and from this point the weights of the top two neural networks are swapped, resulting in the formation of new weights which would be given to the neural network model for the further process.

- **Mutation:** To preserve the population variety, the mutation function plays a key role in introducing randomness in the genes of the offspring. In this project, the mutation multiplies the weights with random numbers from 3 to 9. It helps to resolve the problem of premature convergence and improves the diversity.

- **Termination:** A halting criterion is used for the termination of the algorithm. It checks if the threshold fitness is achieved and if it has, then algorithm terminates otherwise the process is repeated until the maximum fitness is achieved.

D. Neural Network Model

The fundamental building block of deep learning, neural networks are renowned for simulating the behaviour of the human brain while tackling challenging data-driven issues. The neurons that make up the neural network architecture replicate the behaviour of the brain. A Neural Network's main function is to convert input into an actionable output. To generate the required output, the input data is processed by many layers of neurons layered together. A neural network typically has an input layer, an output layer, and one or more hidden layers. A layer is made up of many neurons piled in a row, while a multi-layer neural network is made up of several layers that are stacked on top of one another.

For the purpose of this study, multiple different Neural Network models were designed and trained on the available dataset for the task of cardiovascular illness prediction. After various experimental trials, an optimised neural network model was found whose configurations are described in Table I.

TABLE I
NN MODEL CONFIGURATION

Hidden Layers	5 hidden layers (All dense layers)
Number of Neurons	Each of the dense layers have 512, 256, 128, 64, 10 neurons and the last one with 1 output Neuron.
Activation Function	Used ReLU and Sigmoid.

The GA-NN model architecture can be observed in Fig. 3.

```
Model: "neural__network__model_40"

Layer (type)                Output Shape            Param #
=================================================================
dense_246 (Dense)           (None, 512)             7168

dense_247 (Dense)           (None, 256)             131328

dense_248 (Dense)           (None, 128)             32896

dense_249 (Dense)           (None, 64)              8256

dense_250 (Dense)           (None, 10)              650

dense_251 (Dense)           (None, 1)               11

=================================================================
Total params: 180,309
Trainable params: 180,309
Non-trainable params: 0
```

Fig. 3. GA-NN Model Architecture

E. Overall working of the proposed model

Python is the programming language that was used in the development of the model and the dataset that was used for the proposed model would be the UCI Heart Disease Dataset consisting of various instances and attributes. This dataset required preprocessing and by preprocessing it means cleansing, integrating, transforming and reducing the data. It is a crucial step since it fills the missing numbers, reduces noise, spots outliers and rectifies any mistakes in the dataset. This processed data would be later on split into train and test data. Initially for the first generation of the neural networks, the weights are given randomly. These pool of neural networks are then assigned the fitness value and are sorted according to their fitness scores. From these sorted neural network models, the top two are selected for the cross-breeding where the weights of these neural networks are crossovered. After the crossover of the weights, the offspring or the crossovered weights are given to the mutation function for adding mutation in it. This mutated weights are given to the neural network model that was initially developed. The model is checked for its fitness and if it does not clear the fitness threshold then the obtained weights and the neural network model is utilized to repeat the whole process again till the optimized new weights with the right fitness is obtained. Once the result crosses the fitness threshold, the final result of the neural network model with the best weights is achieved.

IV. EXPERIMENTAL RESULTS

The various evaluation metrics used for the experiments are given below:

A. Evaluation Metrics

- **Accuracy:** It is a metric that determines how the model performed on the data for which it's predictions were to be tested. The accuracy is the ratio of correctly classified examples according to the labels provided divided by the total examples or the total predictions made by the model.

$$Accuracy = \frac{\text{Number of correctly classified examples}}{\text{Total number of examples}}$$

- **Precision:** For quantifying the correct number of positive predictions, precision metric is used. This helps us in calculation of the accuracy of minority classes. Precision being the ratio of correctly predicted outputs over total number of outputs that belong to that class or category. For binary classification task it is defined as,

$$Precision = \frac{TP}{TP + FP}$$

TP = True Positive FP = False Positive
Similarly, it above formula of precision can be extended for multi-class classification as,

$$Precision = \frac{TP_1 + TP_2}{(TP_1 + TP_2) + (FP_1 + FP_2)}$$

TP_1 = True Positive for class 1, TP_2 = True Positive for class 2, FP_1 = False Positive for class 1, FP_2 = False Positive for class 2.

- **Recall:** Recall serves as gauge of how well the model detects True Positives. It is calculated by taking the total number of True Positives and dividing it with the sum of total number of True Positives and False Negatives. It takes values between 0.0 and 1.0.

$$Recall = \frac{TP}{TP + FN}$$

In case of multi-class classification,

$$Recall = \frac{TP_1 + TP_2}{(TP_1 + TP_2) + (FN_1 + FN_2)}$$

TP_1 = True Positive for class 1, TP_2 = True Positive for class 2, FN_1 = False Negative for class 1, FN_2 = False Negative for class 2.

- **F1-Score:** The harmonic mean of recall and precision is known as the F1-Score [19]. In order to get high precision and high recall which is generally desired, we use F - Score calculations given by,

$$F1 - Score = \frac{2 * Precision * Recall}{(Precision + Recall)}$$

- **Computational Time:** It is the time taken for the machine learning model to be trained.
- **Number of Misclassifications:** Number of Misclassifications is a measure that determines Misclassification rate which is a machine learning measure that indicates the proportion of data that a classification model erroneously forecasted. It is determined by:

$$\text{Misclassification Rate} = \frac{\text{False Pos. + False Neg.}}{\text{Total Prediction}}$$

Henceforth, number of misclassifications are the number of inaccurate predictions made by the classifier. It usually is used along with other evaluation parameters, such as accuracy, to analyse model performance since sometimes a model of high accuracy might still have high number of misclassifications in a particular class. This would suggest that the model is not performing well for records belonging to that particular class.

B. Comparative study of simple NN Model with the GA-NN Model

The results of the Genetic Algorithm based Neural Network model is compared with the simple Neural Network Model on the above evaluation metrics.

TABLE II
SIMPLE NN MODEL VS GA-NN MODEL

Evaluation Parameters	Simple NN Model	Genetic Algorithm based NN Model
Testing Accuracy	92.20%	97.56%
Precision	94.06%	98.00%
Recall	90.48%	95.28%
F1-Score	92.23%	97.58%
Computational Time	10.98 sec	5.22 sec
Misclassification Rate	16/205 = 0.078	5/205 = 0.024

From TABLE II it is observed that the testing accuracy of the Genetic Algorithm based NN model is 97.56% that is more than the testing accuracy of the simple NN model which is 92.20%. The precision of the simple NN model is 94.06% whereas the precision of GA-NN model is 98.00% which is almost 4% higher than the simple NN model. The Recall value of the GA-NN model is higher by 5%, and values to be 95.28%, when compared with the Recall score of the simple NN-model which is 90.48%. The GA-NN model also performed well with respect to the F1-score and has the value of 97.58% whereas the value of simple NN model is 92.23%. The computational time of the GA NN model is 5.22 sec and for the simple NN model is 10.98 sec. This shows

that the computational time of the Genetic Algorithm based NN model is almost half compared to the Simple NN model. The misclassification rates of NN model and GA NN model are 0.078 and 0.024 respectively. The misclassification rate is approximately reduced by more than half when GA NN model was used.
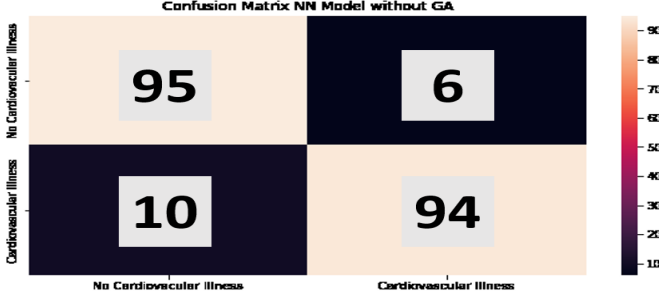

Fig. 4. Confusion Matrix of NN model without GA

Fig. 4 shows the heatmap of the confusion matrix of NN model without GA representing misclassifications in predicitng the cardiovascular illness. According to the confusion matrix, the category of 'No Cardiovascular illness' has 10 misclassified records whereas the category of 'Cardiovascular illness' is misclassified 6 times.
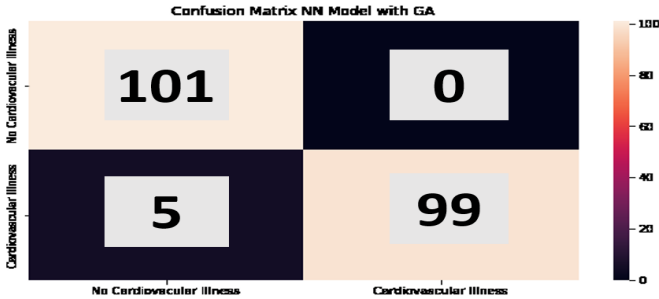

Fig. 5. Confusion Matrix of NN model with GA

Fig. 5 shows the heatmap of the confusion matrix of NN model with GA representing misclassifications in predicting the cardiovascular illness. According to the confusion matrix, the category of the 'No Cardiovascular illness' is misclassified 5 times whereas there are no misclassifications for the category of 'Cardiovascular illness'. This depicts greater predictive abilities of the model when Genetic Algorithm is employed in the working of Neural Network Model.

V. EVALUATION OF THE PROPOSED MODEL WITH THE LITERATURE

This section evaluates the Genetic Algorithm based Neural Network model with the outputs and results obtained in work done in [1].

Metric \ Model	MLP	SVM	RF	NB
Accuracy	81.67%	88.33%	86.67%	86.67%
Precision	92.31%	95.65%	95.45%	91.67%
Recall	54.55%	78.57%	75%	78.57%
F1 Score	68.67%	86.27%	83.99%	84.61%

Fig. 6. Reference Paper Models [1]

In [1], four machine learning models were utilized in developing the predictive model. These four models comprised of Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), Random Forest (RF) and Naive Bayes (NB). All the four models are compared based on the evaluation metrics of Accuracy, Precision, Recall and F1-score. It is observed that SVM got the accuracy of 88.33% which is the highest among the models and the Multi-Layer Perceptron performed the worst with 81.67%. When these values are compared with the self developed neural network in this project, the self developed neural network received the accuracy of 92.20% which is higher than the results in [1] and the proposed GA based NN model outperformed all of these models with 97.56% accuracy. The precision score of the SVM is the highest with a value of 95.65% and the lowest is of Naive Bayes, with a value of 91.67%. This is still lower when compared to the proposed GA based NN model of this project. It has the precision score of 98.00%. The Recall score in [1] is very low as seen in Fig. 6, with a highest score of 78.57% obtained in SVM and NB. The proposed neural network model and the proposed GA-NN model in this project easily outshine it with values of 90.48% and 95.28% respectively. The F1-score in Fig. 6 shows that SVM has the highest score of 86.27% and Multi-Layer Perceptron having the lowest score of 68.67%. The proposed model of Neural Network has the F1-Score of 92.23% and the Genetic Algorithm based Neural Network has the highest F1-score of 97.58% when compared to all the models.

From the comparison and evaluation of the models in [1] and the proposed Genetic Algorithm based Neural Network model, it is observed that the GA-NN model outperformed the state-of-the-art-models with respect to all the evaluation metrics.

VI. CONCLUSION

In this research, the UCI heart disease dataset has been utilized. At the beginning of the project code various relations and dependency have been demonstrated

between the variables and columns in the dataset. The dataset was processed by removing the outliers and null values from the data. This analysis and processing gave an important understanding of the dataset so that efficient working code can be developed out of it. The research focused on developing a genetic algorithm which used the artificial life techniques. These techniques include fitness assignment, selection, crossover and mutation. The Genetic Algorithm here acted as an excellent optimizer in selecting the right weights for the self-developed neural networks. The Genetic Algorithm is executed in loop until the weights which are selected is more than the fitness threshold. These weights are then given to the neural network that is developed for the prediction of the cardiovascular illness. The neural network that is used is a 5 hidden layer network with relu and sigmoid activation function. This Genetic Algorithm based Neural Network is compared with a simple Neural Network for the evaluation. Various evaluation parameters such as testing accuracy, precision, recall, F1-score and misclassification rate are used for the evaluation. The scores that were obtained for the Genetic Algorithm based Neural Network model was then compared with scores obtained for the simple Neural Network model and it was observed that the GA based NN model excelled in all the evaluation parameters. This GA based NN model was then compared with the work done by Boukhatem et al. [1]. There were various models, such as, SVM, NB, MLP and RF that were utilized in prediction of the heart disease in [1]. It was observed that the accuracy, precision, recall and F1-score which were obtained in [1] were lower compared to the proposed GA-NN model. Hence, GA based NN model outshined in every aspect of the evaluation.

This project successfully figured out that machine learning models combined with feature selector and optimizer, which in this case is the Neural network model utilising artificial life technique i.e. Genetic Algorithm, provide an efficient predictive output that can be used in the real world scenario. Hence, successfully answering to the research question which was mentioned in the section of Literature Review.

REFERENCES

[1] C. Boukhatem, H. Y. Youssef, and A. B. Nassif, "Heart disease prediction using machine learning," in *2022 Advances in Science and Engineering Technology International Conferences (ASET)*, 2022, pp. 1–6.

[2] A. Wiki, "Weights and Biases," https://machine-learning.paperspace.com/wiki/weights-and-biases, [Online; accessed 4-November-2022].

[3] M. S.-E. Prafulla N Dawadi, Diane J Cook, "Automated Cognitive Health Assessment Using Smart Home Monitoring of Complex Tasks," https://pubmed.ncbi.nlm.nih.gov/25530925/, 2013, [Online; accessed 4-November-2022].

[4] A. Lambora, K. Gupta, and K. Chopra, "Genetic algorithm- a literature review," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMIT-Con)*, 2019, pp. 380–384.

[5] L. Parthiban and R. Subramanian, "Intelligent heart disease prediction system using canfis and genetic algorithm," *Intl J Biol Life Sci*, vol. 3, 01 2007.

[6] S. R, "Genetic Algorithms and its use-cases in Machine Learning ," https://pubmed.ncbi.nlm.nih.gov/25530925/, 2021, [Online; accessed 4-November-2022].

[7] H. H. Dam, H. A. Abbass, C. Lokan, and X. Yao, "Neural-based learning classifier systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 1, pp. 26–39, 2008.

[8] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, "An automated diagnostic system for heart disease prediction based on χ^2 statistical model and optimally configured deep neural network," *IEEE Access*, vol. 7, pp. 34 938–34 945, 2019.

[9] Y.-K. Kwon and B.-R. Moon, "A hybrid neurogenetic approach for stock forecasting," *IEEE Transactions on Neural Networks*, vol. 18, no. 3, pp. 851–864, 2007.

[10] N. R. Niti Guru, Anil Dahiya, "Decision Support System for Heart Disease Diagnosis Using Neural Network," https://www.delhibusinessreview.org/v_8n1/v8n1i_word.pdf, 2007, [Online; accessed 5-November-2022].

[11] A. Masilamani, ANUPRIYA, and N. C. S. N. Iyenger, "Enhanced prediction of heart disease with feature subset selection using genetic algorithm," *International Journal of Engineering Science and Technology*, vol. 2, 10 2010.

[12] D. Shanthi, G. Sahoo, and N. Saravanan, "Evolving connection weights of artificial neural networks using genetic algorithm with application to the prediction of stroke disease," *International Journal of Soft Computing*, vol. 4, pp. 95–102, 01 2009.

[13] R. El-Bialy, M. A. Salamay, O. H. Karam, and M. E. Khalifa, "Feature analysis of coronary artery heart disease data sets," *Procedia Computer Science*, vol. 65, pp. 459–468, 2015, international Conference on Communications, management, and Information technology (ICCMIT'2015). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050915029622

[14] B. Shantakumar and Y. Kumaraswamy, "Intelligent and effective heart attack prediction system using data mining and artificial neural network," *European Journal of Scientific Research*, vol. 31, pp. 642–656, 01 2009.

[15] S. B. Patil and Y. S. Kumaraswamy, "Extraction of significant patterns from heart disease warehouses for heart attack prediction," 2009.

[16] U. M. L. Repository, "Heart Disease Data Set," https://archive.ics.uci.edu/ml/datasets/heart+Disease, 1988, [Online; accessed 2-November-2022].

[17] Javapoint, "Normalization in Machine Learning," https://www.javatpoint.com/normalization-in-machine-learning.

[18] J. Point, "Genetic Algorithm in Machine Learning," https://www.javatpoint.com/genetic-algorithm-in-machine-learning, 2021, [Online; accessed 18-November-2022].

[19] Z. LT, "Essential Things You Need to Know About F1-Score," https://towardsdatascience.com/essential-things-you-need-to-know-about-f1-score-dbd973bf1a3, 2021, [Online; accessed 8-November-2022].