IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11,S Iss 3, Dec 2022

A Novel MLP Technique on Augmented Dataset for Heart Attack Prediction

[1]Dr.D. Anitha Kumari, [2]Dr. A. Pathanjali Sastri, [3]Dr. A. Radhika, [4] CH.N.V.Jyothirmai

[1]Professor, CSM, TKR College of Engineering and Technology, Hyderabad
[2]Professor, Department of CSE, PSCMR College of Engineering and Technology, Vijayawada, A.P
[3]Associate Professor, Dept of CSE, SRK Institute of Technology, Enikepadu, Vijayawada
[4] Assistant Professor, Department of CSE, PSCMR College Of Engineering and Technology, Vijayawada, A.P

DOI: 10.48047/IJFANS/S1/116

ABSTRACT: Heart Attack is the leading life threatening disease because of the stress life facing by the humans. Traditional approaches are utilizing the 13 attributes related to medical tests like cholestral, threadmill test, and others. But the proposed model integrates the genetic algorithm known as “Particle Swarm Intelligence” algorithm with neural networks to perform the dimensionality reduction and classify the dataset using the customized neural network. The proposed model uses an augmented dataset that contains general information along with medical test reports, which are presented in the form of 55 attributes. The model also has the capability to predict the sub classification like mild, moderate, and severe. Neural Networks helps the automation system to extract the features automatically and implementation of genetic algorithm reduces the features and customization of layers helps the model to find the probability of each option and chooses the one with highest probability.

Keywords: Genetic Algorithms, Neural Networks, Customization, Attribute Selection, Augmentation

INTRODUCTION:

Greek words for "higher level" and "discovery" include "meta" and "heuristic." Typically, this approach finds outstanding answers by avoiding cost complexity. Let's talk about two of the many authors' definitions of these methodologies from various perspectives. In order to swiftly find solutions that are close to ideal, data is organised using learning techniques. A subordinate heuristic is driven by an iterative generating process that integrates cognitively disparate thoughts to achieve this. a master process that iteratively guides and modifies the activities of subpar heuristics to deliver high-quality results in a short amount of time. Iterations may include modifying a single solution—complete or insufficient—or a collection of connected solutions.

The meta-advantages heuristic's allow for the best derivation of search strategies based on the seven approaches. Let's take a closer look at several strategies and their advantages in relation

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,S Iss 3, Dec 2022

to the phases now. In GA, a variety of searching-related problems are resolved, and working with other strategies is relatively simple. The AIS technique, which derives the data through self-adaptation, suggests new locations to search. The SA technique requires less computation time, is simple to implement, and does not require the search for local optimums. Dynamic approaches are the most adaptable, reliable, etc. at graph problems, according to the ACO approach. With its simple implementation and architecture, the PSO approach quickly and efficiently identifies fewer values. ICA technique based on user-implemented

LITERATURE SURVEY:

In[1], Ravish et al say that the cardiovascular system's most crucial component is the heart. It has a weight range of 250–350 grams and is roughly the size of a fist. A technique or model that is effective in foretelling heart attacks is required. In order to effectively evaluate the heart and identify any problems, the authors offer a methodical approach to collecting clinical and ECG data and training an artificial neural network. The group makes predictions using artificial neural networks (ANN). Less than 200 mg/dL of total cholesterol is the ideal range. Another crucial element that needs to be considered in the prognosis of heart disease is age. Diabetes is a genetic condition marked by unusually high blood sugar levels. More clinical trials and recommendations from cardiologists on clinical parameters will assist in creating a working model to reliably predict myocardial infarction since the authors' suggested method did not work for treadmill ECG test data.

In [2], Milli Nabeel et al developed a system for making decisions to diagnose congenital heart disease. The suggested system's fundamental component is a backpropagation neural network. For the purpose of predicting heart disease and blood pressure, the researchers used a neural network. According to a recent research report, scientists have created a novel approach for anticipating the diagnosis of cardiac illness using artificially intelligent methods. A crucial organ of the body is the heart. The effective functioning of the heart is essential to life. Heart disease is typically diagnosed based on physical examinations, symptoms, and physical manifestations in the patient. Based on their training and expertise, doctors can forecast cardiac disease. 13 training and testing attributes from 78 records were used in the research.

In [3], Chaitrali Dangare and Sulabha Apte reported that an effective and dependable Decision Support System on Heart Disease employing Neural networks was developed to improve diagnosis accuracy and shorten diagnosis time. A system that accurately predicts heart illness using artificial intelligence known as a neural network has been developed by researchers. The outcome of the experiment demonstrates that the system diagnoses heart disease with almost 100% accuracy using neural networks. Heart disease prediction from a variety of characteristics or indications is a complex problem that might result in erroneous assumptions and unforeseen outcomes. The study comprised 94 instances in all. The approach incorporates 13 medical indicators, including sex, blood pressure, and cholesterol.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,S Iss 3, Dec 2022

In [4], Srabanti Arora et al stated that by utilizing an artificial neural network, heart illness can be predicted with about 100% accuracy and in a shorter amount of time. A novel method of heart disease prediction has been proposed by researchers. The support vector machine, out of the four classification algorithms, was found to be the most effective method for predicting heart disease. Many researchers are creating various hybrid methodologies by merging various data mining techniques to aid doctors and other medical professionals in predicting heart disease. Machine learning methods used to identify various diseases include Naive Bayes, artificial neural networks, support vector machines, decision trees, etc. The group suggests creating a Web-based software or tool in the future that uses decision tree approaches to make it simple for individuals who live in remote areas to identify the condition. There are other methods that can be used, such as association mining.

In [5], P. Ramprakash et al revealed that a variety of conditions that affect the heart are referred to as heart disease. Heart disease risk factors include age, smoking, heredity, smoking, sex, hypertension, a poor diet, drinking alcohol, and not exercising. In order to forecast the risk of coronary heart disease in individuals with past cases of mental health problems, a machine learning model was developed. Manual diagnosis of heart diseases is challenging, hence standard techniques are employed. According to the World Health Organization, heart conditions are thought to have caused 24% of deaths in India. The ability to detect patients with heart disease will be helped by this research. The sklearn library is used to split the data into training and testing portions. 242 examples or instances are chosen from a total of 303 samples and used to create the model. The remaining information is utilized to evaluate how well the generated model has been used.

In [6], Hamza Turabieh et al report that heart disease is the leading cause of death. The effectiveness of the heart, which pumps blood to every region of the body, is essential to human survival. The Cleveland database in the UCI repository for machine learning provided the data set. and put forth a fresh algorithm to forecast cardiac disease. Because a back-propagation neural network (BPNN) can map intricate non-linear connections between input and output variables, researchers have found that it is preferable to other modeling techniques for unstructured issues. To forecast cardiac disease, the authors put forth a number of hybrid models. The suggested model can support 84.24% and 86.8%, respectively, for the Pima Indians diabetes data set as well as the Cleveland heart disease data set, which are both good accuracy values. In the study, 3 standard healthcare information sets were used. The researchers contend that as compared to a standard back-propagation ANN, the ANN-GWO model took almost twice as long to discover the best model. The group thinks further research is required to determine the best architecture for neural networks.

Table 1: Merits & Demerits of Existing Approaches

S. No	Author	Method	Merits	Demerits
1.	Ravish	ANN	Fault tolerance and scalability	More computational power is needed, likely to overfit

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,S Iss 3, Dec 2022

2.	Al-Milli	decision support system	back propagation algorithm, 13 medical attributes are employed	Sensitive to noise, high dependency
3.	Dangare Chaitrali	data mining and artificial neural network(HDPS)	multilayer perceptron neural network along with backpropagation algorithm is used	Many other models outperformed this model due to overfitting
4.	Maji	artificial neural network	Less time, verified with 10-fold validation	low performance and not efficient
5.	Ramprakash	Machine Learning	Variety range of parameters are taken	Should include the parameter family history,
6.	Turabieh	back-propagation neural network	Good accuracy, can map complex non-linear relations	The results can be more optimized

PROPOSED METHODOLOGY: While feature selection finds the appropriate set of features for a specific target variable, structure learning uncovers the relationships among all the parameters, generally by showing these links as a graph. When using a feature selection technique, it is fundamentally assumed using a feature selection technique that the information may contain certain features that can be removed with little to no data loss since the features are duplicated or irrelevant. Since a relevant trait may become redundant in the presence of another relevant feature that really is closely related, the terms "irrelevant" and "redundant" are not interchangeable. It is best to separate the methods for feature extraction and feature selection. Feature extraction creates novel features from the functionality of the original features, whereas feature selection only returns a subset of the features. In areas with many there are many of features but just a small number of samples or data points, feature selection approaches are frequently used. The accuracy of learning is increased, computation time is decreased, and a better comprehension of the learning method or data is made possible by removing irrelevant data. Most of the time, not every character in the data can be used to actually build a machine learning model. Repetitive variables decrease a model's ability to generalize and may also lower a classifier's order to achieve accuracy. A complicated model is also created when additional factors are added to an existing one. The selection of features are done using the PSO algorithm as discussed below

Formula to find the position vector and velocity vector

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,S Iss 3, Dec 2022

$$\mathbf{v}_i^{t+1} = \underbrace{\mathbf{v}_i^t}_{inertia} + \underbrace{c_1 U_1^t (pb_i^t - p_i^t)}_{personal\ influence} + \underbrace{c_2 U_2^t (gb^t - p_i^t)}_{social\ influence}$$

- Inertia: Makes the particle move in the same direction and with the same velocity
- Personal Influence: Improves the individual, Makes the particle return to a previous position, better than the current, Conservative
- Intensification: explores the previous solutions, finds the best solution of a given region
- Diversification: searches new solutions, finds the regions with potentially the best solution. Position vector can be calculated using the equation (1)

$$X_i = V_i + X_i \text{-} \quad (1)$$

Let us see the mathematical application of PSO using the data given below:

0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	1	0	0

Let the objective function be $f(x) = -x^2 + 5x + 20$ Since all are at initial stage the position and

$$P_{best,i}^{t+1} = \begin{cases} P_{best,i}^t & \text{if } f_i^{t+1} > P_{best,i}^t \\ x_i^{t+1} & \text{if } f_i^{t+1} \leq P_{best,i}^t \end{cases}$$

Initializations:

T = max iterations =10

Range = (0,1)

Population size = 5

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,S Iss 3, Dec 2022

Dimension of the problem = 2 Constants: c1= 2, c2 = 1 r1, r2 are randomly generated numbers Iteration 0:

1.x1=0,x2= 0,x3=0, x4=0,x5=0 Let the initial position be x,i=1,2,3,4,5

2.Objective function: f^0_1=20 ,f^0_2=20 ,f^0_2=20,f^0_3=20,f^0_4=20

3. Initial velocities is zero v01=v02=v03=v04=v05=0

4.Pbest => p0best+1=0 ,p1best+2 =0 , ……p5best+5=0

5.Gbest=min(p_{best})=0

Table 2: F(x) & Velocity Computation

x1	x2	x3	x4	x5
0	0	0	0	0
0	0	0	0	0
0	0	1	0	0
0	1	0	0	0
0	0	0	0	0

Transformation

x1	x2	x3	x4	x5
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

Iteration 1:

$v_{11}=wv_{11}+c_1r_1(p_{11}^b-x_{11})+c_2r_2(p_g - x_{11})$

=0+2*(0.3)(0-0)+1*(0.2)(0-0)

v_{11}=0

$x_{11}=v_{11}+x_{11}$ 0+0=0 **∈(0,1)**

$v_{12}=wv_{12}+c_1r_1(p_{12}^b-x_{12})+c_2r_2(p_g - x_{12})$

=0+2*(0.3)(0-0)+1*(0.2)(0-0)

v_{12}=0

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,S Iss 3, Dec 2022

$x_{12}=v_{12}+x_{12}$ 0+0=0 $\in$**(0,1)**

…

$v_{33}=wv_{33}+c_1r_1(p_{33}^{b}-x_{33})+c_2r_2(p_g- x_{33})$

=0+2*(0.3)(0-0)+1*(0.2)(0-1)

=-0.6

$x_{33}=v_{33}+x_{33}$ = -0.6

When it comes to automation of medical analysis of any large amounts of data is a key responsibility for any operator. Using MLP, hidden patterns in data are strengthened in order to extract the best results for users based on patient behaviour. To identify the patterns, the author used approaches from machine learning, deep learning, and data mining. The main problem with pattern recognition is the resource usage for relationships and supported symptoms. Based on the back and front propagation methodologies, MLP is employed. Two techniques—sigmoid and tanh—are employed for backpropagation. Based on the fusion of two approaches, adagrade and adadelta, the pitfall approach is examined. Front propagation is often simple, and values are determined based on parameters.

RESULTS & DISCUSSION:

	Age	Weight	Length	Sex	BMI	DM	HTN	Current Smoker	EX-Smoker	FH	Obesity	CRF	CVA	Airway disease	Thyroid Disease	CHF	DLP	BP	PR	Edema	Weak Peripheral Pulse	Lung rales	Systolic Murmur
0	53	90	175	1	29.387755	0	1	1	0	0	1	0	0	0	0	0	1	110	80	0	0	0	0
1	67	70	157	0	28.398718	0	1	0	0	0	1	0	0	0	0	0	0	140	80	1	0	0	0
2	54	54	164	1	20.077335	0	0	1	0	0	0	0	0	0	0	0	0	100	100	0	0	0	0
3	66	67	158	0	26.838648	0	1	0	0	0	1	0	0	0	0	0	0	100	80	0	0	0	0
4	50	87	153	0	37.165193	0	1	0	0	0	1	0	0	0	0	0	0	110	80	0	0	0	1
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
298	58	84	168	1	29.761905	0	0	0	0	0	1	0	0	0	0	0	0	100	76	0	0	0	0
299	55	64	152	0	27.700831	0	0	0	0	0	1	0	0	0	0	0	0	100	60	0	0	0	1
300	48	77	160	0	30.078125	0	1	0	0	1	1	0	0	0	0	0	0	130	70	0	0	0	0
301	57	90	159	0	35.599858	1	0	0	0	0	1	0	0	0	0	0	0	100	60	0	0	0	0
302	56	85	170	0	29.411765	0	1	1	0	0	1	0	0	0	0	0	0	120	80	0	0	0	0

Figure 1: Transformed Data

Figure 1 represents the data transformation step, which converts the categorical data into numerical as part of pre-processing to make the process fast.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,S Iss 3, Dec 2022

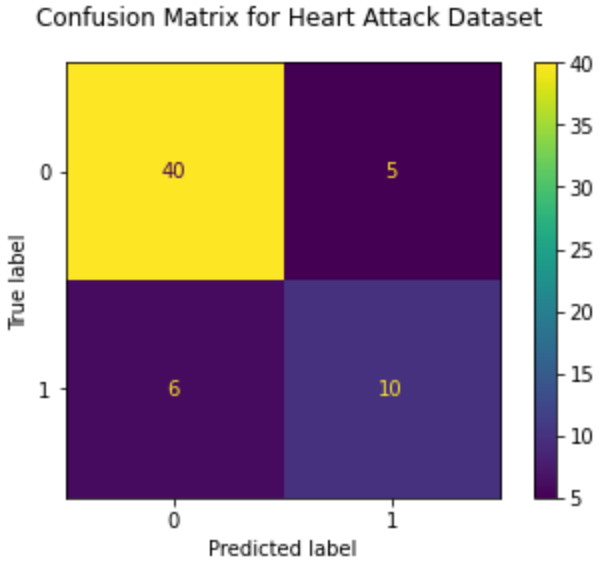

Figure 2: Confusion Matrix

Figure 2 presents confusion matrix generated on the test data in which most of the positive and negative records are correctly classified where less number of type-1 and type-2 are occurred.

	precision	recall	f1-score	support
0	0.87	0.89	0.88	45
1	0.67	0.62	0.65	16
accuracy			0.82	61
macro avg	0.77	0.76	0.76	61
weighted avg	0.82	0.82	0.82	61

Figure 3: Classification Report

Figure 3 projected classification reports with can show case all the metrics based on normal and heart attack cases. All the metrics in report are equal i.e. 82% because the dataset is a balanced dataset.

CONCLUSION: The proposed methodology used PSO algorithm for reducing the number of attributes then it has implemented a 5-layer neural network which helps us to identify whether a person is suffering from heart attack or not along with that it also performs the sub classification, which determines the stage of attack. In this regard, the system is more efficient than the traditional approaches.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11,S Iss 3, Dec 2022

REFERENCES:

1. Ravish, D. K., Shanthi, K. J., Shenoy, N. R., & Nisargh, S. (2014). *Heart function monitoring, prediction and prevention of Heart Attacks: Using Artificial Neural Networks. 2014 International Conference on Contemporary Computing and Informatics (IC3I).* doi:10.1109/ic3i.2014.7019580

2. Al-Milli, Nabeel. "Backpropagation neural network for prediction of heart disease." *Journal of theoretical and applied information Technology* 56.1 (2013): 131-135.

3. Dangare, Chaitrali and Apte, Sulabha, A Data Mining Approach for Prediction of Heart Disease Using Neural Networks (November 14, 2012). International Journal of Computer Engineering and Technology (IJCET), Volume 3, Issue 3, October-December 2012, Available at SSRN: https://ssrn.com/abstract=2175569

4. Maji, S., Arora, S. (2019). Decision Tree Algorithms for Prediction of Heart Disease. In: Fong, S., Akashe, S., Mahalle, P. (eds) Information and Communication Technology for Competitive Strategies. Lecture Notes in Networks and Systems, vol 40. Springer, Singapore. https://doi.org/10.1007/978-981-13-0586-3_45

5. Ramprakash, P., Sarumathi, R., Mowriya, R., & Nithyavishnupriya, S. (2020). *Heart Disease Prediction Using Deep Neural Network. 2020 International Conference on Inventive Computation Technologies (ICICT).* doi:10.1109/icict48043.2020.9112443

6. Turabieh, H. (2016) A Hybrid ANN-GWO Algorithm for Prediction of Heart Disease. *American Journal of Operations Research*, **6**, 136-146. doi: 10.4236/ajor.2016.62016.