DOI : https://dx.doi.org/10.26808/rs.ed.i11v3.01
International Journal of Emerging Trends in Engineering and Development Issue 11, Vol.3 (April- May 2021)
Available online on http://www.rspublication.com/ijeted/ijeted_index.htm ISSN 2249-6149

# PREDICTION OF CARDIOVASCULAR DISEASES USING GENETIC ALGORITHM AND DEEP LEARNING TECHNIQUES

**Kratika Sharma**
Dept. of Information Technology
Chaitanya Bharathi Institute of Technology (A)
Hyderabad, India
Kratikasharma_it@cbit.ac.in

**T. Satya Kiranmai**
Dept. of Information Technology
Chaitanya Bharathi Institute of Technology (A)
Hyderabad, India
tskiranmai_it@cbit.ac.in

## Abstract

The early methods of forecasting the cardiovascular diseases resulted in the reduction of risks by helping make effective decisions about the changes to have occurred in high-risk patients. Cardiovascular diseases are a broad range of diseases that are affecting heart and blood vessels. The health care industry stores lots of medical data, therefore machine learning algorithms can be used to make effective decisions in the prediction of heart diseases. Recent research has delved into uniting these techniques to provide hybrid machine learning algorithms. Medical Diagnosis Systems play a vital role in medical practice and are used by many medical professionals for treatment and diagnosis . In this paper, a medical diagnosis system is presented for predicting the risk of cardiovascular disease. This system is built by combining the relative advantages of genetic algorithm and neural network. Multilayered feed forward neural networks are particularly suited to complex classification problems. The weights of the neural network are determined using genetic algorithm because it finds acceptably good set of weights in less number of iterations. The dataset provided by University of California, Irvine (UCI) machine learning repository is used for training and testing. It consists of 303 instances of heart disease data each having 14 attributes including the class label. First, the dataset is preprocessed in order to make them suitable for training. Genetic based neural network is used for training the system. The final weights of the neural network are stored in the weight base and are used for predicting the risk of cardiovascular disease

***Keywords***— *cardiovascular diseases; genetic algorithm; Multilayered feed forward neural networks; classification problems*

## I. INTRODUCTION

The health care industries collect and store huge amounts of data that contain some concealed information, which is useful for making decisions. Some advanced data mining techniques are used for providing appropriate results and making effective decisions on data. In this dynamic world people want to live a life where they work like a machine in order to earn lot of money and live a contented life therefore in this process they forget to take care of

themselves, because of this there their entire lifestyle change. In this type of lifestyle they have blood pressure, diabetes at a very young age and they don't give enough rest for themselves and eat what is required and they don't even bother about the quality of the food if sick the go for their own medication as a result of all these small negligence it leads to a major threat that is the heart disease Cardiovascular disease technically refers to any disease that affects the cardiovascular system.It is usually used to refer to those related to atherosclerosis. It includes coronary heart disease, rheumatic heart disease, raised blood pressure, cerebrovascular disease, peripheral artery disease, congenital heart disease and heart failure.

## II. LITERATURE SURVEY

### A. *Heart Disease Prediction Using Machine learning and Data Mining Technique*

Researchers have been using several data mining techniques to help health care professionals in the diagnosis of heart disease. However using data mining technique can reduce the number of test that are required. In order to reduce number of deaths from heart diseases there have to be a quick and efficient detection technique. Decision Tree is one of the effective data mining methods used. This research compares different algorithms of Decision Tree classification seeking better performance in heart disease diagnosis using WEKA. The algorithms which are tested is J48 algorithm, Logistic model tree algorithm and Random Forest algorithm. The existing datasets of heart disease patients from Cleveland database of UCI repository is used to test and justify the performance of decision tree algorithms. This datasets consists of 303 instances and 76 attributes. Subsequently, the classification algorithm that has optimal potential will be suggested for use in sizeable data. The goal of this study is to extract hidden patterns by applying data mining techniques, which are noteworthy to heart diseases and to predict the presence of heart disease in patients where this presence is valued from no presence to likely presence.

### B. *Predicting and Diagnosing of Heart Diseases Using Machine Learning*

Prediction and diagnosing of heart disease become a challenging factor faced by doctors and hospitals both in India and abroad. In order to reduce the large scale of deaths from heart diseases, a quick and efficient detection technique is to be discovered. Data mining techniques and machine learning algorithms play a very important role in this area. The researchers accelerating their research works to develop a software with the help machine learning algorithm which can help doctors to take decision regarding both prediction and diagnosing of heart disease. The main objective of this research paper is predicting the heart disease of a patient using machine learning algorithms. Comparative study of the various performance of machine learning algorithms is done through graphical representation of the results.

### C. *Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques*

A popular saying goes that we are living in an "information age". Terabytes of data are produced every day. Data mining is the process which turns a collection of data into knowledge. The health care industry generates a huge amount of data daily. However, most of it is not effectively used. Efficient tools to extract knowledge from these databases for clinical detection of diseases or other purposes are not much prevalent. The aim of this paper is to summarise some of the current research on predicting heart diseases using data mining techniques, analyse the various combinations of mining algorithms used and conclude which technique(s) are effective and efficient. Also, some future directions on prediction systems have been addressed.

*D. Survey of Machine Learning Algorithms for Disease Diagnostic*

In medical imaging, Computer Aided Diagnosis (CAD) is a rapidly growing dynamic area of research. In recent years, significant attempts are made for the enhancement of computer aided diagnosis applications because errors in medical diagnostic systems can result in seriously misleading medical treatments. Machine learning is important in Computer Aided Diagnosis. After using an easy equation, objects such as organs may not be indicated accurately. So, pattern recognition fundamentally involves learning from examples. In the field of bio-medical, pattern recognition and machine learning promise the improved accuracy of perception and diagnosis of disease. They also promote the objectivity of decision-making process. For the analysis of high-dimensional and multimodal bio-medical data, machine learning offers a worthy approach for making classy and automatic algorithms. This survey paper provides the comparative analysis of different machine learning algorithms for diagnosis of different diseases such as heart disease, diabetes disease, liver disease, dengue disease and hepatitis disease. It brings attention towards the suite of machine learning algorithms and tools that are used for the analysis of diseases and decision-making process accordingly.

*E. A Machine Learning Approach for Early Prediction of Breast Cancer*

Nowadays by the rapid digitisation of the data in the Healthcare sector has resulted in the collection of mountains amount of data in various Electronic Health Records (EHR). As the data is the biggest asset in the modern age, whose proper utilisation in the Healthcare sector can lead to the discovery of the dreadful diseases very well in time which in turn will provide high quality of care to patients and at less expenditure. Developing a machine learning models that can help us in prediction the disease can play a vital role in early prediction. These Machine learning methods can be used to classify between healthy people & people with different disease. In the given project the light is been thrown on the same disease by using certain selected machine learning algorithms in WEKA tool and a corresponding evaluation of the selected Machine learning algorithms in terms of accuracy is also performed so as to select the best classifier for the early diagnosis of the said disease with better accuracy results In this paper three different types of models were implemented on the Breast Cancer dataset as Naïve Bayes, Logistic Regression and Random Forest. Out of the three Random Forest lead the top by having accuracy of 98% and sensitivity 99% followed by Logistic Regression with accuracy of 96% and sensitivity 98% and finally with Naive Bayes with accuracy of 91% and sensitivity 94%.

## III. SYSTEM IMPLEMENTATION

Heart disease can be managed effectively with a combination of lifestyle changes, medicine and, in some cases, surgery. With the right treatment, the symptoms of heart disease can be reduced and the functioning of the heart improved. The predicted results can be used to prevent and thus reduce cost for surgical treatment and other expensive.

The overall objective of my work will be to predict accurately with few tests and attributes the presence of heart disease. Attributes considered form the primary basis for tests and give accurate results more or less. Many more input attributes can be taken but our goal is to predict with few attributes and faster efficiency the risk of having heart disease. Decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the data set and databases. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients.

DOI : https://dx.doi.org/10.26808/rs.ed.i11v3.01
International Journal of Emerging Trends in Engineering and Development Issue 11, Vol.3 (April- May 2021)
Available online on http://www.rspublication.com/ijeted/ijeted_index.htm ISSN 2249-6149

### *A. EXISTING SYSTEM*

Numerous works for heart disease prediction is achieved by artificial neural network(ANN) and data mining techniques. Computational biology is often applied in the process of translating biological knowledge into clinical practice, as well as in the understanding of biological phenomena from the clinical data.

## *VARIOUS MACHINE LEARNING ALGORITHMS*

### *A. Logistic Regression*

Logistic regression is one of the most efficient machine learning algorithms and well known for binary classification. The variables are binary dependent variables like true or false, 0s or 1s, pass or fail etc. If there are ordered multiple categories then uses ordinal logistic regression or the variables are having more than one outcomes, then multinomial logistic regression is used. The logistic function as shown, Where e is the numerical constant Euler's number and a is an input we put into the function. The roc curve for the logistic regression.

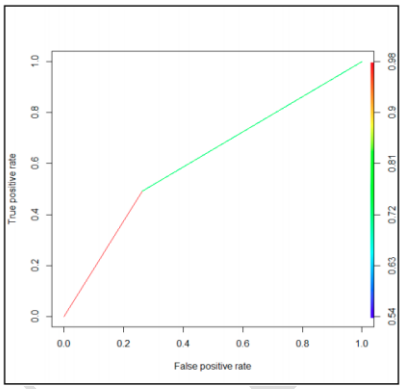


*Fig.1: ROC Curve*

### *B. Naive Bayes*

Navies Bayes classifier shows the probability of each input attribute for the predictable state and provides the probability of event occur. It is one of best classification algorithm in machine learning which uses the Bayesian algorithm. A conditional probability is the likelihood of some conclusion A, given some evidence/observation, B, where a dependence relationship exists between A and B. This probability is denoted as P (A|B) where, P(A) is the probability of event A, P(B) is the probability of event B, P(B|A) is the probability of event B with the condition that event A has taken place.

$$P(A/B) = \frac{(P(B/A)*P(B))}{P(A)}$$

### *C. Random Forest*

Random forest is kind of machine learning method where the weak models are combined to form a dynamic model. It creates decision trees for each attribute. It is a machine learning algorithm used for classification and regression. The random forest tree shows the multiple decision trees that are linked to the system.

DOI : https://dx.doi.org/10.26808/rs.ed.i11v3.01
International Journal of Emerging Trends in Engineering and Development Issue 11, Vol.3 (April- May 2021)
Available online on http://www.rspublication.com/ijeted/ijeted_index.htm ISSN 2249-6149

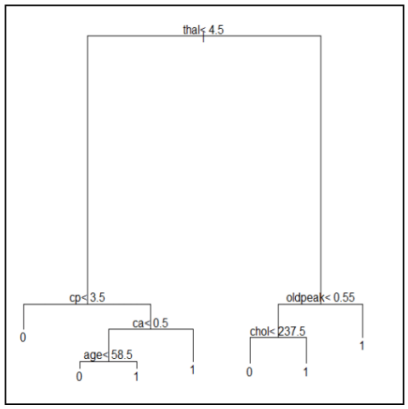


*Fig 2: Random forest tree*

### *D. Gradient Boosting*

In this, gradient boosting technique it provides a variable importance of the attribute that is related to predict the heart disease in this dataset. The following figure shows the variable importance of heart failure prediction with the help of the boosted tree.

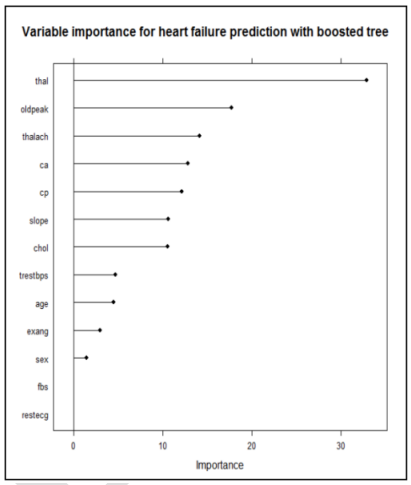


*Fig 3: Variable importance graph*

### *E. Accuracy Module*

The module predicts the accuracy by using machine learning algorithms. In this, each algorithm provides different accuracy rate for taken attributes which is the cause of the cardiovascular disease. You can calculate the accuracy of your model with:

$$\text{True Negative Rate, Specificity} = \frac{P}{P+Q}$$

$$\text{False Positive Rate, 1- Specificity} = \frac{Q}{Q+P}$$

sum to 1

$$\text{True Positive Rate, Sensitivity} = \frac{R}{R+S}$$

$$\text{False Negative Rate} = \frac{S}{R+S}$$

sum to 1

DOI : https://dx.doi.org/10.26808/rs.ed.i11v3.01
International Journal of Emerging Trends in Engineering and Development Issue 11, Vol.3 (April- May 2021)
Available online on http://www.rspublication.com/ijeted/ijeted_index.htm ISSN 2249-6149

From Confusion Matrix Specificity and Sensitivity can be derived as illustrated below:
Where, P=Number of true negatives, Q=Number of false positives and R=Number of true positives, S=Number of false negatives. Sensitivity is the approach that identify the people those with the cardiovascular disease (true positive rate) and specificity is the approach that identify the people those without the cardiovascular disease (true negative rate).

### *B. PROPOSED SYSTEM*

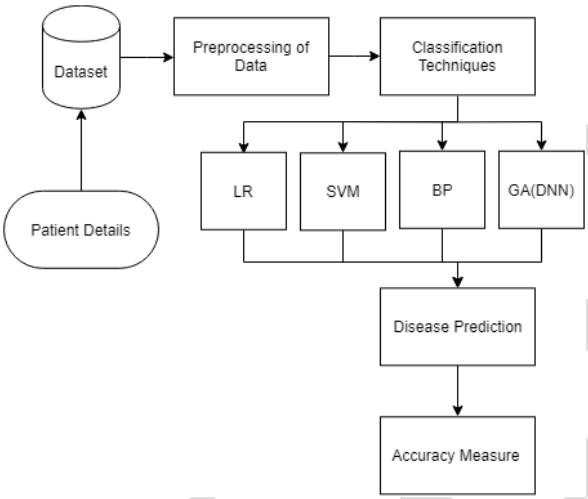


*Fig 4 Flow of the proposed system*

The above figure shows functioning of the system is described step by step
Step-1. The dataset contains the details of the patients.
Step-2. After identifying the data from the available resources, they are further selected for processing which includes data cleaning, removal of noise i.e. missing data
Step-3 Classification algorithms such as back propagation , svm , Logistic Regression and genetic algorithm with multi layer feed forward network are used here.
Step-5 It also finds the accuracy of the algorithms and compares the accuracy among all the algorithms.

The proposed work is implemented with three main modules that is machine learning models such as svm and logistic regression and further extended to implement genetic algorithm with fitness function as deep learning model and the results of these models are compared for the better accuracy along with cleveland dataset. All the models above are implemented using the python libraries such as scikit-learn ,tensorflow etc. The pre-processing is done by replacing the null values and also in appropriate values using python code.

*Cleveland Dataset*
This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML reaserchers to this date. The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4. Experiments with the Cleveland database have concentrated on simply attempting to distinguish presence (values 1,2,3,4) from absence (value 0).

DOI : https://dx.doi.org/10.26808/rs.ed.i11v3.01
International Journal of Emerging Trends in Engineering and Development Issue 11, Vol.3 (April- May 2021)
Available online on http://www.rspublication.com/ijeted/ijeted_index.htm ISSN 2249-6149

This dataset contains information concerning heart disease diagnosis. The data was collected from the Cleveland Clinic Foundation, and it is available at the UCI machine learning Repository. Six instances containing missing values.

Format:

A data frame with 297 observations on the following 14 variables.

- V1 age(continuous)
- V2 sex
- V3 cp, chest pain type:1,2,3,4
- V4 trestbps: resting blood pressure(continuous)
- V5 cholesterol(continuous)
- V6 fps: fatsing blood sugar>120? yes=1, no =0
- V7 restecg: resting electrocardiographic results, 0,1, 2
- V8 thalach: maximum heart rate achieved(continuous)
- V9 exang: exercise induced angina (1 = yes; 0 = no)
- V10 oldpeak = ST depression induced by exercise relative to rest (continuous)
- V11 slope: the slope of the peak exercise ST segmentV12 ca: number of major vessels (0-3) colored by flourosopy
- V13 thal: 3 = normal; 6 = fixed defect; 7 = reversible defect
- V14 diagnosis of heart disease: 1: < 50 2: > 50

For heart disease prediction system we implement the below algorithms and compare the results:

- Back propagation algorithm
- SVM Technique
- Logistic Regression
- Genetic algorithm along with deep neural network

*1. Back propagation*

The back-propagation algorithm can be employed effectively to train neural networks; it is widely recognised for applications to layered feed-forward networks, or multi-layer perceptrons The back-propagation learning algorithm can be divided into two phases: propagation and weight update .

Phase 1: Propagation

1. Forward propagation of a training pattern's input through the neural network in order to generate the propagation's output activations.
2. Back propagation of the propagation's output activations through the neural network using the training pattern's target in order to generate the deltas of all output and hidden neurons.

Phase 2: Weight update For each weight-synapse:

1. Multiply its output delta and input activation to get the gradient of the weight.
2. Bring the weight in the opposite direction of the gradient by subtracting a ratio of it from the weight. Repeat the phase 1 and 2 until the performance of the network is good enough. The results illustrated the peculiar strength of each of the methodologies in comprehending the objectives of the specified mining objectives.

This phase is broken down into 6 parts:

1. Initialize Network.
2. Forward Propagate.

3. Back Propagate Error.
4. Train Network.
5. Predict.
6. Dataset Case Study.

These steps will provide the foundation that you need to implement the backpropagation algorithm

*Initialise the network*

It accepts three parameters, the number of inputs, the number of neurons to have in the hidden layer and the number of outputs.

You can see that for the hidden layer we create **n_hidden** neurons and each neuron in the hidden layer has **n_inputs + 1** weights, one for each input column in a dataset and an additional one for the bias.

You can also see that the output layer that connects to the hidden layer has **n_outputs**neurons, each with **n_hidden + 1** weights. This means that each neuron in the output layer connects to (has a weight for) each neuron in the hidden layer.

*Forward Propagate*

We can calculate an output from a neural network by propagating an input signal through each layer until the output layer outputs its values.We call this forward-propagation.

It is the technique we will need to generate predictions during training that will need to be corrected, and it is the method we will need after the network is trained to make predictions on new data.

We can break forward propagation down into three parts:

1.Neuron Activation.
2.Neuron Transfer.
3.Forward Propagation

*1.Neuron Activation:*

The first step is to calculate the activation of one neuron given an input.

The input could be a row from our training dataset, as in the case of the hidden layer. It may also be the outputs from each neuron in the hidden layer, in the case of the output layer.

Neuron activation is calculated as the weighted sum of the inputs. Much like linear regression

activation = sum(weight_i * input_i) + bias

*2. Neuron Transfer:*

Once a neuron is activated, we need to transfer the activation to see what the neuron output actually is.

Different transfer functions can be used. It is traditional to use the sigmoid activation function, but you can also use the tanh (hyperbolic tangent) function to transfer outputs. More recently, the rectifier transfer function has been popular with large deep learning networks.

The sigmoid activation function looks like an S shape, it's also called the logistic function. It can take any input value and produce a number between 0 and 1 on an S-curve. It is also a function of which we can easily calculate the derivative (slope) that we will need later when backpropagating error.

*3.Forward propagate:*

Forward propagating an input is straightforward.We work through each layer of our network calculating the outputs for each neuron. All of the outputs from one layer become inputs to the neurons on the next layer.

DOI : https://dx.doi.org/10.26808/rs.ed.i11v3.01
International Journal of Emerging Trends in Engineering and Development Issue 11, Vol.3 (April- May 2021)
Available online on http://www.rspublication.com/ijeted/ijeted_index.htm ISSN 2249-6149

*Back Propagate Error*

The backpropagation algorithm is named for the way in which weights are trained.
Error is calculated between the expected outputs and the outputs forward propagated from the network. These errors are then propagated backward through the network from the output layer to the hidden layer, assigning blame for the error and updating weights as they go.
The math for backpropagating error is rooted in calculus, but we will remain high level in this section and focus on what is calculated and how rather than why the calculations take this particular form.
This part is broken down into two sections.
1.Transfer Derivative.
2.Error Backpropagation.

*Train Network*

The network is trained using stochastic gradient descent.
This involves multiple iterations of exposing a training dataset to the network and for each row of data forward propagating the inputs, backpropagating the error and updating the network weights.
This part is broken down into two sections:
1.Update Weights.
2.Train Network.

*Prediction*

A function named **predict()** that implements this procedure. It returns the index in the network output that has the largest probability. It assumes that class values have been converted to integers starting at 0.

*2. SVM*

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data ,the algorithm outputs an optimal hyperplane which categorizes new examples. In two dimentional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side.
The implementation is based on libsvm. The fit time complexity is more than quadratic with the number of samples which makes it hard to scale to dataset with more than a couple of 10000 samples.
The multiclass support is handled according to a one-vs-one scheme.
For a dataset consisting of features set and labels set, an SVM classifier builds a model to predict classes for new examples. It assigns new example/data points to one of the classes. If there are only 2 classes then it can be called as a Binary SVM Classifier.
There are 2 kinds of SVM classifiers:

1. Linear SVM Classifier
2. Non-Linear SVM Classifier

*SVM Linear Classifier:*

In the linear classifier model, we assumed that training examples plotted in space. These data points are expected to be separated by an apparent gap. It predicts a straight hyperplane dividing 2 classes. The primary focus while drawing the hyperplane is on maximizing the distance from hyperplane to the nearest data point of either class. The drawn hyperplane called as a maximum-margin hyperplane.In Linear Classifier, A data point considered as a p-dimensional vector(list of p-numbers) and we separate points using (p-1) dimensional

DOI : https://dx.doi.org/10.26808/rs.ed.i11v3.01
International Journal of Emerging Trends in Engineering and Development Issue 11, Vol.3 (April- May 2021)
Available online on http://www.rspublication.com/ijeted/ijeted_index.htm ISSN 2249-6149

hyperplane. There can be many hyperplanes separating data in a linear order, but the best hyperplane is considered to be the one which maximizes the margin i.e., the distance between hyperplane and closest data point of either class.
The Maximum-margin hyperplane is determined by the data points that lie nearest to it. Since we have to maximize the distance between hyperplane and the data points. These data points which influences our hyperplane are known as support vectors.

*SVM Non-Linear Classifier:*
In the real world, our dataset is generally dispersed up to some extent. To solve this problem separation of data into different classes on the basis of a straight linear hyperplane can't be considered a good choice. For this Vapnik suggested creating Non-Linear Classifiers by applying the kernel trick to maximum-margin hyperplanes. In Non-Linear SVM Classification, data points plotted in a higher dimensional space.Vapnik proposed Non-Linear Classifiers in 1992. It often happens that our data points are not linearly separable in a p-dimensional(finite) space. To solve this, it was proposed to map p-dimensional space into a much higher dimensional space. We can draw customized/non-linear hyperplanes .
This function helps to build a high dimensional feature space. There are many kernels that have been developed. Some standard kernels are:

1. Polynomial (homogeneous) Kernel**:** The polynomial kernel function can be represented by the above expression. Where $k(x_i, x_j)$ is a kernel function, $x_i$ & $x_j$ are vectors of feature space and d is the degree of polynomial function.
2. Polynomial(non-homogeneous)Kernel: In the non-homogeneous kernel, a constant term is also added. The constant term "c" is also known as a free parameter. It influences the combination of features. x & y are vectors of feature space.
3. RadialBasisFunctionKernel**:**
   It is also known as RBF kernel. It is one of the most popular kernels. For distance metric squared euclidean distance is used here. It is used to draw completely non-linear hyperplanes.

$$K(x,x') = \exp\left(\frac{-\|x-x'\|^2}{2\sigma^2}\right)$$

where x & x' are vectors of feature space. is a free parameter. Selection of parameters is a critical choice. Using a typical value of the parameter can lead to overfitting our data.

*3. Logistic Regression*
Simple logistic regression is useful for finding relationship between two continuous variables. One is predictor or independent variable and other is response or dependent variable. It looks for statistical relationship but not deterministic relationship. Relationship between two variables is said to be deterministic if one variable can be accurately expressed by the other. For example, using temperature in degree Celsius it is possible to accurately predict Fahrenheit. Statistical relationship is not accurate in determining relationship between two variables. For example, relationship between height and weight.
This class implements logistic regression using lib linear, newton-cg, sag of lbfgs optimizer. The newton-cg, sag and lbfgs solvers support only L2 regularization with primal formulation. The liblinear solver supports both L1 and L2 regularization, with a dual formulation only for the L2 penalty.

DOI : https://dx.doi.org/10.26808/rs.ed.i11v3.01
International Journal of Emerging Trends in Engineering and Development Issue 11, Vol.3 (April- May 2021)
Available online on http://www.rspublication.com/ijeted/ijeted_index.htm ISSN 2249-6149

Randomness and unpredictability are the two main components of a regression model.
Prediction = Deterministic + Statistic
Deterministic part is covered by the predictor variable in the model. Stochastic part reveals the fact that the expected and observed value is unpredictable. There will always be some information that are missed to cover. This information can be obtained from the residual information.
Let's explain the concept of residue through an example. Consider, we have a dataset which predicts sales of juice when given a temperature of place. Value predicted from regression equation will always have some difference with the actual value. Sales will not match exactly with the true output value. This difference is called as residue.
Residual plot helps in analyzing the model using the values of residues. It is plotted between predicted values and residue. Their values are standardised. The distance of the point from 0 specifies how bad the prediction was for that value. If the value is positive, then the prediction is low. If the value is negative, then the prediction is high. 0 value indicates prefect prediction. Detecting residual pattern can improve the model.
Non-random pattern of the residual plot indicates that the model is,

1. Missing a variable which has significant contribution to the model target
2. Missing to capture non-linearity (using polynomial term)
3. No interaction between terms in model
4. Characteristics of a residue
5. Residuals do not exhibit any pattern
6. Adjacent residuals should not be same as they indicate that there is some information missed by system.

*Metrics for model evaluation*
*R-Squared value*

- This value ranges from 0 to 1. Value '1' indicates predictor perfectly accounts for all the variation in Y. Value '0' indicates that predictor 'x' accounts for no variation in 'y'.

1. Regression sum of squares (SSR)
This gives information about how far estimated regression line is from the horizontal 'no relationship' line (average of actual output).
Error $=\sum_{i=1}^{n}(\text{Predicted\_output} - \text{average\_of\_actual\_output})^2$
2. Sum of Squared error (SSE)
How much the target value varies around the regression line (predicted value).
Error $=\sum_{i=1}^{n}(\text{Actual\_output} - \text{predicted\_output})^2$
Value of R2 may end up being negative if the regression line is made to pass through a point forcefully. This will lead to forcefully making regression line to pass through the origin (no intercept) giving an error higher than the error produced by the horizontal line. This will happen if the data is far away from the origin.
This is related to value of 'r-squared' which can be observed from the notation itself. It ranges from -1 to 1.

$$r = (+/-)\ \text{sqrt}(r^2)$$

If the value of b1 is negative, then 'r' is negative whereas if the value of 'b1' is positive then, 'r' is positive. It is unitless.

*4. Genetic algorithm along with deep neural network*

*1. Deep Neural Network*

Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans: learn by example. Deep learning is a key technology behind driverless

cars, enabling them to recognize a stop sign, or to distinguish a pedestrian from a lamppost. It is the key to voice control in consumer devices like phones, tablets, TVs, and hands-free speakers. Deep learning is getting lots of attention lately and for good reason. It's achieving results that were not possible before.

In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Models are trained by using a large set of labeled data and neural network architectures that contain many layers.

In a word, accuracy. Deep learning achieves recognition accuracy at higher levels than ever before. This helps consumer electronics meet user expectations, and it is crucial for safety-critical applications like driverless cars. Recent advances in deep learning have improved to the point where deep learning outperforms humans in some tasks like classifying objects in images.

While deep learning was first theorized in the 1980s, there are two main reasons it has only recently become useful:

1. Deep learning requires large amounts of labeled data. For example, driverless car development requires millions of images and thousands of hours of video.
2. Deep learning requires substantial computing power. High-performance GPUs have a parallel architecture that is efficient for deep learning. When combined with clusters or cloud computing, this enables development teams to reduce training time for a deep learning network from deep neural networks.
3. The term "deep" usually refers to the number of hidden layers in the neural network. Traditional neural networks only contain 2-3 hidden layers, while deep networks can have as many as 150.
4. Deep learning models are trained by using large sets of labeled data and neural network architectures that learn features directly from the data without the need for manual feature extraction.

Deep learning is a specialized form of machine learning. A machine learning workflow starts with relevant features being manually extracted from images. The features are then used to create a model that categorizes the objects in the image. With a deep learning workflow, relevant features are automatically extracted from images. In addition, deep learning performs "end-to-end learning" – where a network is given raw data and a task to perform, such as classification, and it learns how to do this automatically.Another key difference is deep learning algorithms scale with data, whereas shallow learning converges. Shallow learning refers to machine learning methods that plateau at a certain level of performance when you add more examples and training data to the network.

A key advantage of deep learning networks is that they often continue to improve as the size of your data increases.

The Keras library provides wrapper classes to allow you to use neural network models developed with Keras in scikit-learn.

There is a KerasClassifier class in Keras that can be used as an Estimator in scikit-learn, the base type of model in the library. The KerasClassifier takes the name of a function as an argument. This function must return the constructed neural network model, ready for training. This network has 22 inputs neurons ,16 hidden units and 2 outputs neurons. We have used "softmax" and "relu" activation function in this network.

### *2. Genetic Algorithm*

Genetic Algorithms(GAs) are adaptive heuristic search algorithms that belong to the larger part of evolutionary algorithms. Genetic algorithms are based on the ideas of natural selection and genetics. These are intelligent exploitation of random search provided with historical data

to direct the search into the region of better performance in solution space. They are commonly used to generate high-quality solutions for optimization problems and search problems.

Genetic algorithms simulate the process of natural selection which means those species who can adapt to changes in their environment are able to survive and reproduce and go to next generation. In simple words, they simulate "survival of the fittest" among individual of consecutive generation for solving a problem. Each generation consist of a population of individuals and each individual represents a point in search space and possible solution. Each individual is represented as a string of character/integer/float/bits. This string is analogous to the Chromosome.

In a genetic algorithm, a population of candidate solutions (called individuals, creatures, or phenotypes) to an optimization problem is evolved toward better solutions. Each candidate solution has a set of properties (its chromosomes or genotype) which can be mutated and altered; traditionally, solutions are represented in binary as strings of 0s and 1s, but other encodings are also possible.

The evolution usually starts from a population of randomly generated individuals, and is an iterative process, with the population in each iteration called a *generation*. In each generation, the fitness of every individual in the population is evaluated; the fitness is usually the value of the objective function in the optimization problem being solved. The more fit individuals are stochastically selected from the current population, and each individual's genome is modified (recombined and possibly randomly mutated) to form a new generation. The new generation of candidate solutions is then used in the next iteration of the algorithm. Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population.

A typical genetic algorithm requires:

1. a genetic representation of the solution domain,
2. a fitness function to evaluate the solution domain.

A standard representation of each candidate solution is as an array of bits.Arrays of other types and structures can be used in essentially the same way. The main property that makes these genetic representations convenient is that their parts are easily aligned due to their fixed size, which facilitates simple crossover operations. Variable length representations may also be used, but crossover implementation is more complex in this case. Tree-like representations are explored in genetic programming and graph-form representations are explored in evolutionary programming; a mix of both linear chromosomes and trees is explored in gene expression programming.

Once the genetic representation and the fitness function are defined, a GA proceeds to initialize a population of solutions and then to improve it through repetitive application of the mutation, crossover, inversion and selection operators.

Genetic algorithms are based on an analogy with genetic structure and behavior of chromosome of the population. Following is the foundation of GAs based on this analogy –

- Individual in population compete for resources and mate
- Those individuals who are successful (fittest) then mate to create more offspring than others
- Genes from "fittest" parent propagate throughout the generation, that is sometimes parents create offspring which is better than either parent.
- Thus each successive generation is more suited for their environment.

*Search space*

The population of individuals are maintained within search space. Each individual represent a solution in search space for given problem. Each individual is coded as a finite length vector

DOI : https://dx.doi.org/10.26808/rs.ed.i11v3.01
International Journal of Emerging Trends in Engineering and Development Issue 11, Vol.3 (April- May 2021)
Available online on http://www.rspublication.com/ijeted/ijeted_index.htm ISSN 2249-6149

(analogous to chromosome) of components. These variable components are analogous to Genes. Thus a chromosome (individual) is composed of several genes (variable components).

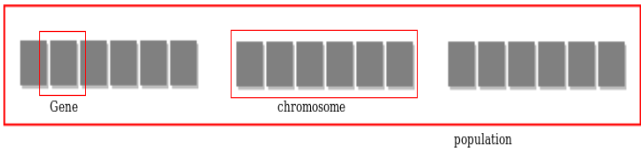


*Fig 5: Screenshot of gene, chromosome, population representation*

*Fitness Score*

A Fitness Score is given to each individual which shows the ability of an individual to "compete". The individual having optimal fitness score (or near optimal) are sought.

The GAs maintains the population of n individuals (chromosome/solutions) along with their fitness scores .The individuals having better fitness scores are given more chance to reproduce than others. The individuals with better fitness scores are selected who mate and produce **better offspring** by combining chromosomes of parents. The population size is static so the room has to be created for new arrivals. So, some individuals die and get replaced by new arrivals eventually creating new generation when all the mating opportunity of the old population is exhausted. It is hoped that over successive generations better solutions will arrive while least fit die.

Each new generation has on average more "better genes" than the individual (solution) of previous generations. Thus each new generation have better "partial solutions" than previous generations. Once the offspring produced having no significant difference than offspring produced by previous populations, the population is converged. The algorithm is said to be converged to a set of solutions for the problem.

Operators of Genetic Algorithms:

For each new solution to be produced, a pair of "parent" solutions is selected for breeding from the pool selected previously. By producing a "child" solution using the above methods of crossover and mutation, a new solution is created which typically shares many of the characteristics of its "parents". New parents are selected for each new child, and the process continues until a new population of solutions of appropriate size is generated. Although reproduction methods that are based on the use of two parents are more "biology inspired", some research[4][5] suggests that more than two "parents" generate higher quality chromosomes.

These processes ultimately result in the next generation population of chromosomes that is different from the initial generation. Generally the average fitness will have increased by this procedure for the population, since only the best organisms from the first generation are selected for breeding, along with a small proportion of less fit solutions. These less fit solutions ensure genetic diversity within the genetic pool of the parents and therefore ensure the genetic diversity of the subsequent generation of children.

Opinion is divided over the importance of crossover versus mutation. There are many references in Fogel (2006) that support the importance of mutation-based search.

Although crossover and mutation are known as the main genetic operators, it is possible to use other operators such as regrouping, colonization-extinction, or migration in genetic algorithms.

It is worth tuning parameters such as the mutation probability, crossover probability and population size to find reasonable settings for the problem class being worked on. A very small mutation rate may lead to genetic drift (which is non-ergodic in nature). A recombination rate that is too high may lead to premature convergence of the genetic

DOI : https://dx.doi.org/10.26808/rs.ed.i11v3.01
International Journal of Emerging Trends in Engineering and Development Issue 11, Vol.3 (April- May 2021)
Available online on http://www.rspublication.com/ijeted/ijeted_index.htm ISSN 2249-6149

algorithm. A mutation rate that is too high may lead to loss of good solutions, unless elitist selection is employed

*Selection Operator:*

The idea is to give preference to the individuals with good fitness scores and allow them to pass there genes to the successive generations.

*Crossover Operator:*

This represents mating between individuals. Two individuals are selected using selection operator and crossover sites are chosen randomly. Then the genes at these crossover sites are exchanged thus creating a completely new individual (offspring). For example –

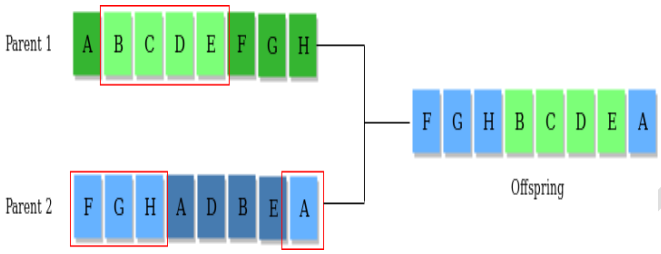


*Fig 6: crossover representation*

*Mutation Operator:*

The key idea is to insert random genes in offspring to maintain the diversity in population to avoid the premature convergence. For example –

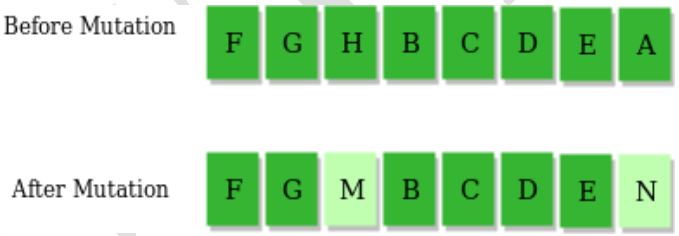


*Fig 7: mutation representation*

## IV. RESULTS

In this section ,we have presented the accuracy results of various methods we have implemented like:

### *A. Backpropagation*

For back propagation the accuracy obtained is 82.0.

```
Scores: [73.33333333333333, 85.0, 83.33333333333334, 86.66666666666667, 81.66666666666667]
Mean Accuracy: 82.000%
```

*Fig 8:Screenshot of backpropagation accuracy*

### *B. SVM*

For svm the accuracy obtained is 81.5.

```
SVM: 81.57894736842105 %
```

*Fig 9: Screenshot of SVM accuracy*

DOI : https://dx.doi.org/10.26808/rs.ed.i11v3.01
International Journal of Emerging Trends in Engineering and Development Issue 11, Vol.3 (April- May 2021)
Available online on http://www.rspublication.com/ijeted/ijeted_index.htm ISSN 2249-6149

### *C. Logistic Regression*

For logistic regression the accuracy obtained is 82.89.

```
Logistic Regression: 82.89473684210526 %
```

*Fig 10:Screenshot of logistic regression accuracy*

### *D. Genetic Algorithm*

For genetic algorithm the accuracy obtained is 94.34.

```
[False  True  True  True  True  True  True False  True False False  Tr
ue
  True]
after genetic algorithm
genetic4test.py:58: UserWarning: Update your `Dense` call to the Keras
 2 API: `Dense(9, input_dim=9, activation="relu", kernel_initializer="
uniform")`
  model.add(Dense(layers[0], input_dim=ip, init='uniform', activation=
activ))


Test Score:  94.34662865180718
(base) ksagars-MacBook-Pro:genetic Dr K Sagars
```

*Fig 11: Screenshot of genetic algorithm accuracy*

*Comparison of tables*

*TABLE I. BACKPROPOGATION*

| Dense Layers | Accuracy | Time taken |
|---|---|---|
| 1 | 81.46553 | 0.67s |
| 2 | 74.76878 | 0.89s |
| 3 | 79.35577 | 0.78s |
| 4 | 82.67846 | 0.67s |
| 5 | 82.64764 | 0.69s |

*TABLE II. GENETIC ALGORITHM*

| Dense layers | Accuracy | Time Taken |
|---|---|---|
| 1 | 74.33444 | 654s |
| 2 | 74.56775 | 711s |
| 3 | 92.7746 | 625s |
| 4 | 94.5666 | 728s |
| 5 | 83.67655 | 720s |

## *V. CURRENT LIMITATIONS AND FUTURESCOPE*

The dataset on Heart Disease is taken and analyzed to predict the severity of the disease. This system is developed for the analysis of medical data. The data in the dataset is preprocessed to make it suitable for classification. The dataset consists of 76 attributes on the whole .We have not considered all of them as some of them were irrelevant. The dataset was not too large and was convenient to run the algorithm on limited number of layers in neural network . Further the execution time of the system may be reduced . Also different datasets are not considered to compare the algorithm accuracy.

For future work there are many interesting aspects. We can enhance the system by using Genetic Algorithms with Principal Component Analysis(PCA) to reduce the dimension of the dataset and is used to predict the cardiovascular diseases risk. To get a clear depiction of

the condition Digital Image Processing can be included .. We can enable images to analyse and predict the risk. Also we can increase the data to several Mb and Gb with current technologies. Not just the current dataset we have used ,we can implement on several other datasets and see if there is efficiency maintained by the algorithm. All the attributes ca be taken into consideration to predict the risk of cardiovascular disease Other advance techniques like GANs can also be performed.

## VI. CONCLUSION

There are several fatal diseases among which heart related are the most prominent. Medical practitioners gather information and conduct a variety of surveys on coronary diseases to study the symptoms of heart patients and disease progression. The proposed heart disease prediction system has been designed as a deep neural network .The Cleveland dataset is used. A genetic based neural network approach is used to predict the severity of the disease. The neural network has been used as a fitness function in Genetic Algorithm. The weights for the neural network are determined using genetic algorithm Based on the best feature set generated in genetic algorithm the accuracy has been calculated.

## REFERENCES

[1] Jaymin Patel, Prof.Tejal Upadhyay, Dr.Samir Patel “Heart disease prediction using Machine learning and Data Mining Technique" Volume 7.Number1 Sept 2015- March 2016.

[2] Igor Kononenko ”Machine learning for medical diagnosis: history, state of art& perspective" Elsevier - Artificial intelligence in Medicine, Volume23, Aug 2001

[3] Sanjay Kumar Sen” Predicting and Diagnosing of Heart Disease Using Machine Learning Algorithms”- International Journal of Engineering And Computer Science ISSN:2319-7242Volume6Issue 6 June 2017

[4] G.Parthiban, S.K.Srivasta "Applying Machine learning methods in Diagnosing Heart disease for Diabetic Patients" International Journal of Applied Information Systems (IJAIS) – ISSN: 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 3– No.7, August 2012

[5] Sana Bharti, Shailendra Narayan Singh" Analytical study of heart disease comparing with different algorithms”: Computing, Communication & Automation(ICCCA),2015InternationalConference.

[6] Thenmozhi.K and Deepika.P, Heart Disease Prediction using classification with different decision tree techniques. International Journal of Engineering Research & General Science, Vol 2(6), pp 6-11, Oct2014.

[7] Animesh Hazra, Subrata Kumar Mandal, Amit Gupta,Arkomita Mukherjee and Asmita Mukherjee” Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review”- Advances in Computational Sciences and Technology ISSN 0973-6107, Volume10, Number7(2017).