



# Heart Disease Prediction Using CNN Algorithm

VirenViraj Shankar<sup>1</sup> · Varun Kumar<sup>1</sup> · Umesh Devagade<sup>1</sup> · Vinay Karanth<sup>1</sup> · K. Rohitaksha<sup>1</sup>

Published online: 15 May 2020  
© Springer Nature Singapore Pte Ltd 2020

## Abstract

In this paper, we aim to predict accuracy, whether the individual is at risk of a heart disease. This prediction will be done by applying machine learning algorithms on training data that we provide. Once the person enters the information that is requested, the algorithm is applied and the result is generated. Obviously, the accuracy is expected to decrease when the medical data itself are incomplete. We implement the prediction model over real-life hospital data. We propose to use convolutional neural network algorithm as a disease risk prediction algorithm using structured and perhaps even on unstructured patient data. The accuracy obtained using the developed model ranges between 85 and 88%. We have proposed further by applying other machine learning algorithms over the training data to predict the risk of diseases, comparing their accuracies so that we can deduce the most accurate one. Attributes can also be modified in an attempt to improve the accuracy further.

**Keywords** Machine learning · Big data analytics · Deep learning · Medical applications · Convolutional neural network

## Introduction

It is reported that 50% of Americans suffer from at least one chronic disease. Unsurprisingly, this results in 80% of US healthcare fee being spent on chronic disease treatment. With the raise in the living standards, the effect of these diseases also increases. The USA as a whole has spent almost \$2.7 trillion per annum on respective treatments. The USA is not the only country where large sums are spent treating chronic diseases. In China, for example, most people die because of chronic diseases, as reported, this accounts for more than 85% of all deaths in the world's most populated country. Clearly, it is essential that early diagnosis and treatment are essential, not just to save costs, but also to save human life and improve quality of life.

Chen et al. proposed a healthcare system using smart clothing for sustainable health monitoring. Qiu et al. had thoroughly studied the heterogeneous systems and achieved

the best results for cost minimization on tree and simple path cases for heterogeneous systems. Patients' statistical information, test results and disease history are recorded in the EHR, enabling us to identify potential data-centric solutions to reduce the costs of medical case studies.

With the development of big data analytics technology, more attention has been paid to disease prediction from the perspective of big data analysis; various researches have been conducted by selecting the characteristics automatically from a large number of data to improve the accuracy of risk classification, rather than the previously selected characteristics.

To solve these problems, the structured and unstructured data can be combined in healthcare to assess the risk of disease.

### How Model Works?

Figure 1 depicts the various steps carried out during the prediction of heart disease.

1. It starts with the data collection; here in this step, different types of data mainly structured, semi-structured or unstructured can be collected from various sources like hospital, etc.
2. Once the data are collected, the obtained data are first cleaned to remove missing values and to bring under same level of granularity, and then, the cleaned data are classified into training data and test dataset.

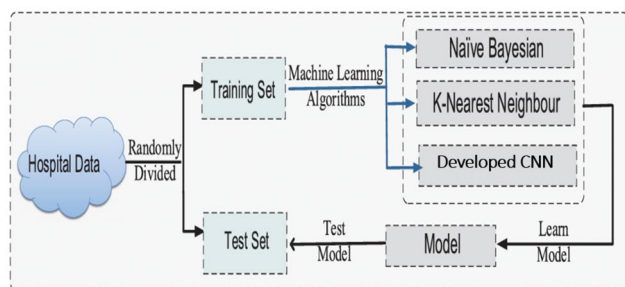
---

This article is part of the topical collection “Advances in Computational Intelligence, Paradigms and Applications” guest edited by Young Lee and S. Meenakshi Sundaram”.

---

✉ Varun Kumar  
varunkumar8156@gmail.com

<sup>1</sup> Department of Computer Science and Engineering,  
JSS Academy of Technical Education, Visveswaraya  
Technological University, Bangalore 56006, India



**Fig. 1** Proposed heart disease prediction model

- After the data segregation, the data are fed into various machine learning algorithms like Naive Bayes, KNN or CNN. This step is mainly done using training data to teach the machine so as to increase its predictive accuracy.
- Once the data have learnt enough, our learned model will be ready for testing.
- The learned model is tested using test data to check its predictive accuracy.
- If the predictive accuracy is up to the desired level, then the model is deployed.

## Dataset and Model

### Hospital Data

The medical datasets that we have considered in this study are hospital data and the same are stored in our database [1]. This includes 13 attributes being processed. The structured data being considered include laboratory data as well as basic patient details such as the patient's age, gender and cholesterol level. All this datum is required and relevant to diagnose heart disease in patients. As future scope, unstructured data can also be considered, some of which are listed in Table 1.

### Disease Risk Prediction

Here, in this model, we have mainly focused on the prediction of heart disease. The aim of the designed model is to predict the likelihood of the individual being a current or future heart disease patient. Here, the model requires the individual to input values which correlate with various

relevant attribute values of the patient  $= (x_1; x_2; \dots; x_n)$ . This will contain general, laboratory and medical information, all of which will be fed to the algorithm to generate results with a higher accuracy than the other algorithms that were considered.

### Data Preprocessing

As expected, missing data will lead to reduced accuracy of predictions. This loss of data may be due to human error, etc. Thus, to retain accuracy, we need to fill it in. Before data are applied to the model, missing attributes are filled and unnecessary attributes, if any, are weeded out. This is taken care in preprocessing process, and dataset is divided randomly into training and test data to generate accuracy value to evaluate the design.

## Literature Survey

Poornima Singh et al. [2] proposed heart disease prediction system based on neural network. The proposed method considered 15 attributes for prediction. The training model used multilayer perceptron neural network with back propagation. The dataset considered was the structured data, and the model was able to provide 100% accuracy.

Gomathi et al. [3] used Naive Bayes and decision tree data mining techniques for predicting different types of diseases. They mainly concentrated on prediction of heart diseases, diabetics and breast cancer. The results were derived from the confusion metrics.

Miranda et al. [4] proposed Naive Bayes classifier approach for the prediction of the cardiovascular diseases. The authors have considered few important risk factors for deciding the heart disease. The proposed idea has led to 85% of accuracy, sensitivity and specificity.

## Methods

### Convolution Layer

The first layer is the convolution layer. The layer retains the relationship between pixels of images. It does this by learning its properties. This operation is performed by dividing the input image into smaller boxes of pixels. Basically,

**Table 1** Structured and unstructured datasets

Data category	Item	Description
Structured data	Demographics of the patient	Patient's gender, age, height, weight, etc.
	Lining habits	Whether the patient smokes, has a genetic history, etc.
	Examination items and results	Including 682 items, such as blood, etc.
	Diseases	Patient's disease, such as cerebral infarction
Unstructured data	Patient's readme illness	Patient's readme illness and medical history
	Doctor's records	Doctor's interrogation records

it is a mathematical operation which takes into account two things: image matrix and filter/kernel.

## Padding

In many occasions, the filter applied does not fit the input image properly. This leaves us with two options:

1. One workaround would be adding padding (zeroes in the case of zero padding) for a better fit.
2. Using valid padding—weeding out portions of the image where filter did not fit appropriately and using the rest of it only (Fig. 2).

## ReLU

The rectified linear unit is the most popular activation function employed for deep learning [5]. ReLu functions by returning the input value if it is positive and failing that, 0. This can be represented as:

$$\text{ReLU}(a) = \text{Max}(0, a)$$

Graphically, this can be represented in Fig. 3.

Activation functions serve two primary purposes:

1. Help a model account for interaction effects. What is an interactive effect? It is when one variable A affects a prediction differently depending on the value of B. For example, if my model wanted to know whether a certain body weight indicated an increased risk of diabetes, it would have to know an individual's height. Some body

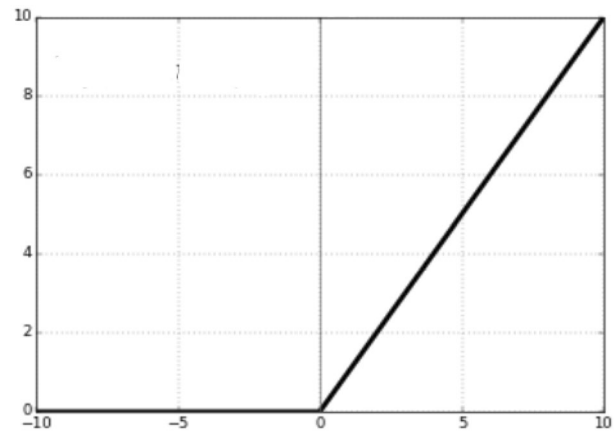


Fig. 3 Graphical representation of ReLu function

weights indicate elevated risks for short people, while indicating good health for tall people. So, the effect of body weight on diabetes risk depends on height, and we would say that weight and height have an interaction effect.

2. Help a model account for nonlinear effects. This just means that if I graph a variable on the horizontal axis, and my predictions on the vertical axis, it is not a straight line. Or said another way, the effect of increasing the predictor by one is different at different values of that predictor.

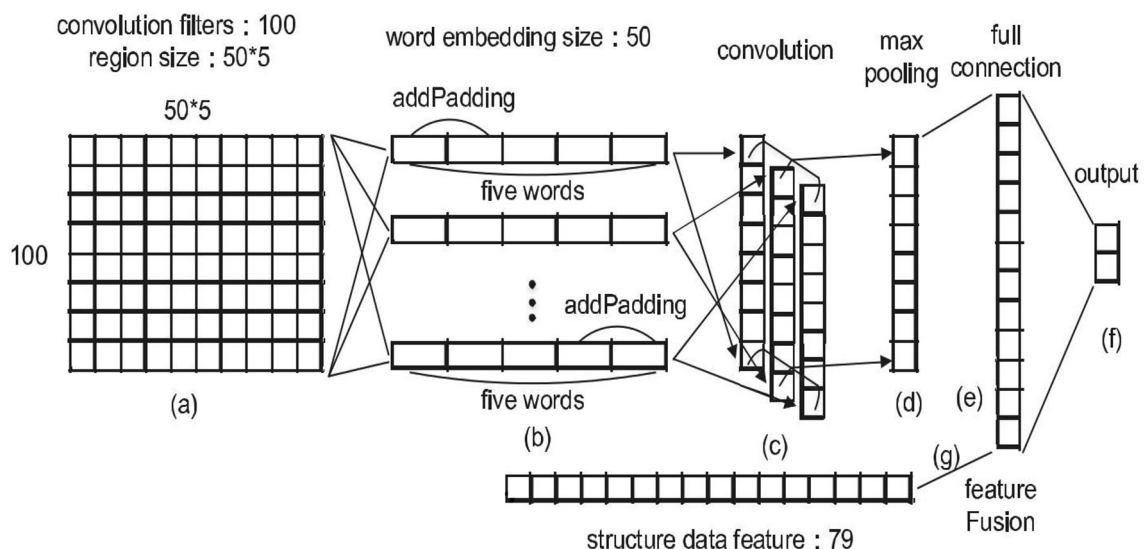


Fig. 2 CNN at work

## Pooling

The pooling section is used to drop unnecessary parameters as needed. Spatial pooling reduces the dimensionality of input data but retains its important attributes. This can be of different types:

- **Max Pooling:** It takes the largest element from the feature map.
- **Average Pooling:** The largest element could be considered here as well.
- **Sum Pooling:** Sum of every element in the feature map is calculated and used.

## Fully Connected Layer

Next comes the fully connected or FC layer, which takes the matrix once it is transformed into vectors. As illustrated in Fig. 2, the feature map matrix is converted into vector form ( $v_1, v_2, v_3, \dots$ ) and is fed into a neural network. The FC layer takes these vectored features and integrates them to form a model.

Finally, another activation function is applied for classification. The activation function used in CNN is usually sigmoid or softmax function

## Sigmoid Function

It is a mathematical function that has an “S”-shaped curve or sigmoid curve [6]. It is mainly used for binary classifications. It is defined by the formula and shown graphically in Fig. 4.

The main advantage of this activation function stems because its derivative is easy to determine. Based on the convention, we can expect the output value in the range of  $-1$  to  $1$ .

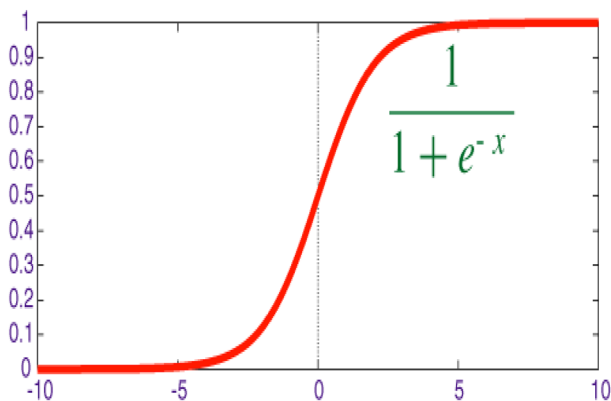


Fig. 4 Sigmoid function

## Softmax Function

Softmax function calculates the probability distribution of the event over ‘n’ different events. Below are the few properties of softmax function [5].

- Calculated probabilities will range from 0 to 1.
- The sum of all these computed probabilities will be equal to 1.

## Adam Optimization Algorithm

Adam optimization is an alternative to the normally used stochastic gradient descent to update neural network weights iteratively, based on training data fed [7].

Adam varies from classical stochastic gradient descent as stochastic gradient descent uses just one learning rate alpha which is applied for all weight updates, and the same is not modified during the training process. In the case of Adam, a learning rate is maintained for each network weight and separately adapted during the learning process.

Adam incorporates the advantages of two extensions of stochastic gradient descent such as:

- Adaptive Gradient Algorithm:** a learning rate is maintained for each attribute value, improving gradient descent performance.
- Root Mean Square Propagation:** here also, learning rate is also maintained a learning rate for every attribute, but here, it is based on the average of recent magnitudes of the gradients for the weight, e.g., how quickly it is changing. This means the algorithm does well on online and ever-changing applications flooded with noise.

### Propose Algorithm

Input

- o Alpha is the learning rate.
- o  $\beta_1$  and  $\beta_2$  are hyper parameters having default values -  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ .
- o epoch = max no of iterations.
- o Initialize V, S, v and s to zero. Where V and v are weighted average of past gradients and S and s are weighted average of the squares of the past gradients before bias correction.
- o W is the input provided and b is the bias that gets multiplied with the weights at each neurons.

Output

1. Update V and v like momentum.

$$V = \beta_1 \times V + (1 - \beta_1) \times dW$$

$$v = \beta_1 \times v + (1 - \beta_1) \times db.$$

2. Update S and s like Rmsprop.

$$S = \beta_2 \times S + (1 - \beta_2) \times dW^2$$

$$s = \beta_2 \times s + (1 - \beta_2) \times db^2.$$

3. After bias correction.

$$V_{corrected} = V / (1 - \beta_1 t)$$

$$v_{corrected} = v / (1 - \beta_1 t)$$

$$S_{corrected} = S / (1 - \beta_2 t)$$

$$s_{corrected} = s / (1 - \beta_2 t)$$

4. Update parameters W and b.

$$W = W - \text{learning rate} \times (V_{corrected} / \sqrt{S_{corrected} + \epsilon}).$$

$$b = b - \text{learning rate} \times (v_{corrected} / \sqrt{s_{corrected} + \epsilon}).$$

5. Repeat step 4 until no further correction is needed.

t-> no of iterations

Epoch specifies no of iterations

Basically, the formula for the algorithms is

$$\text{Output} = X * W + b$$

where X=input parameters, i.e., the input parameters considered for prediction of heart diseases example age, sex, chest pain, etc.

W is the weights that are changed in each iteration so that the accuracy of the algorithm can be improved.

b is the bias that is added to each node in the neural network. Learning rate is unique for each weight. So for

updating the weights and the bias, we use Adam's optimizer, i.e., the proposed algorithm is used for updating the weights and the bias used in the neural network. The updated values are multiplied with the given input parameters and the classifier function like ReLu and sigmoid are used for the classification. The weights and bias are updated iteratively until we get the sufficient accuracy for the model.

Classification

Sigmoid (output)= class label 0 or class label 1

ReLu (output)= class label 0 or class label 1

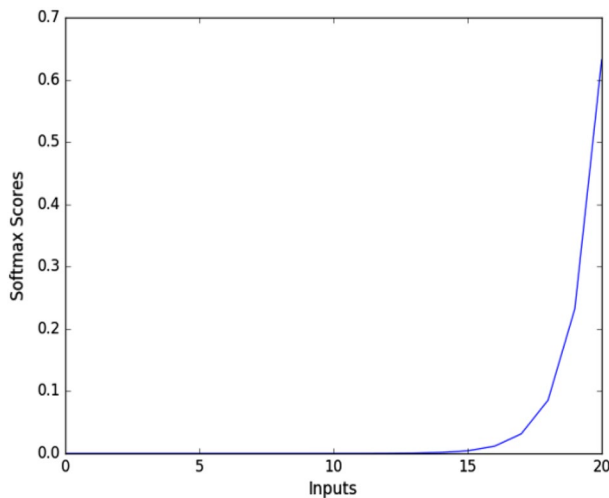


Fig. 5 Softmax function

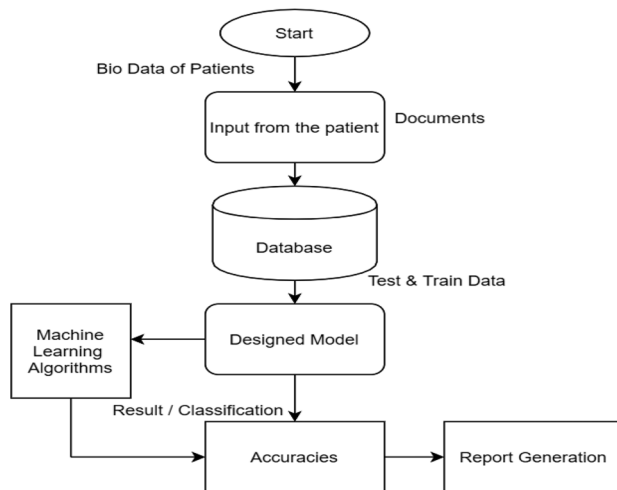


Fig. 6 Flowchart of the prediction model

For the different nodes, we have used different activation functions, i.e., for the hidden layers, we have used ReLu as activation function and for the final output layer, we have used sigmoid activation function for the classification of the class label (Fig. 5).

## Flowchart

Figure 6 gives the view of how the flow of data occurs from one phase to other, like data collection, data cleaning, data splitting into training and test data, then training the model using training data and finally predicting the accuracy using test data and then report generation

## Results

Figure 7 indicates the details about the various attributes that are responsible for heart disease. Apart from these attributes, there are various other attributes as well, but those do not affect the disease as much as these attributes do

Figure 8 shows basically two graphs for comparison. There are two class labels: One indicates about the person who has heart disease indicated by the value 1, i.e., if the green scatters on the graph, we can say as class 1. The other class label indicates about the person who does not have heart disease indicated by the value 0, i.e., if the red scatters on the graph, we can say as class 0. The x-axis in the graph indicates the number of values that are being plotted from the test data chosen from the whole dataset. The y-axis indicates the class labels of the datasets used to plot the graph. The left part of the graph shows us the plotting of the class labels of actual test data used for testing of the trained model. The right part of the graph shows us the plotting of predicted class labels done by the developed model/algorithm on the same data used for the testing of the developed model.

Figure 9 shows the comparison for different algorithms which can be used to predict heart, namely CNN, Naive Bayes, KNN, etc. It is observed that the model accuracy is highest for the model which is designed using CNN and then comes Naive Bayes algorithm which has little less accuracy than that of CNN and then comes KNN algorithm with least accuracy.

Further CNN algorithm can be used with both structured and unstructured data. Images can also be used with CNN algorithm to predict different diseases. The reason is that we

SQLyog Community Edition- MySQL GUI - [ - root@localhost]

File Edit Favorites DB Table Objects Tools Window Help

root@localhost  
information\_s  
cardiac  
cardiac  
Views  
Stored Pro  
Functions  
mysql  
test

Query

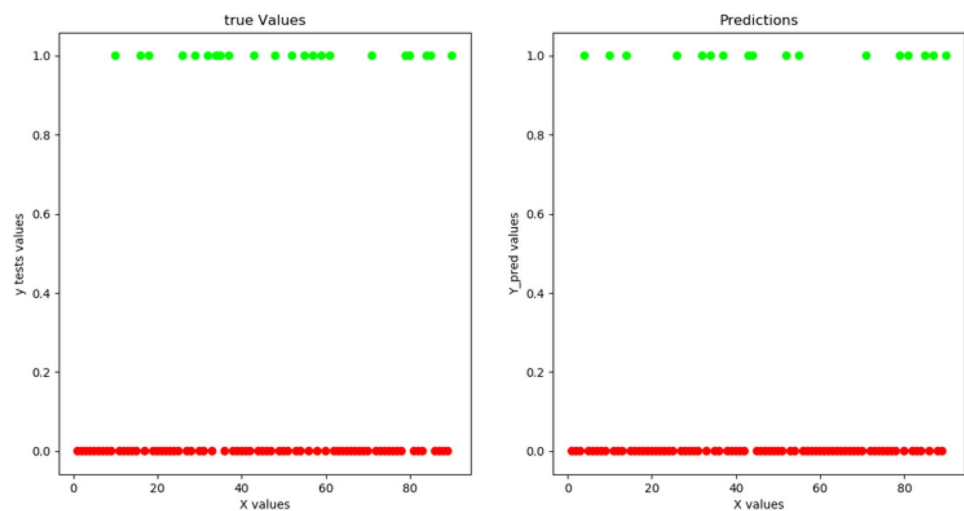
1 Result 2 Messages 3 Table Data 4 Objects 5 History

Show All or Limit 0 50 Refresh

age	sex	pai	bp	chol	fbs	ecg	maxhr	eiang	eist	slope	vessels	thal	pana	res
63	1	1	145	233	1	2	150	0	2.3	3	0	6	umesh	The Person has Mild heart Disease
60	1	4	130	206	0	2	132	1	2.4	2	2	7	7	The Person has Severe heart Disease
63	1	1	145	233	1	2	150	0	2.3	3	0	6	umesh	The Person does not have heart Disea
60	1	4	130	206	0	2	132	1	2.4	2	2	7	vinay	The Person has Severe heart Disease
*	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)

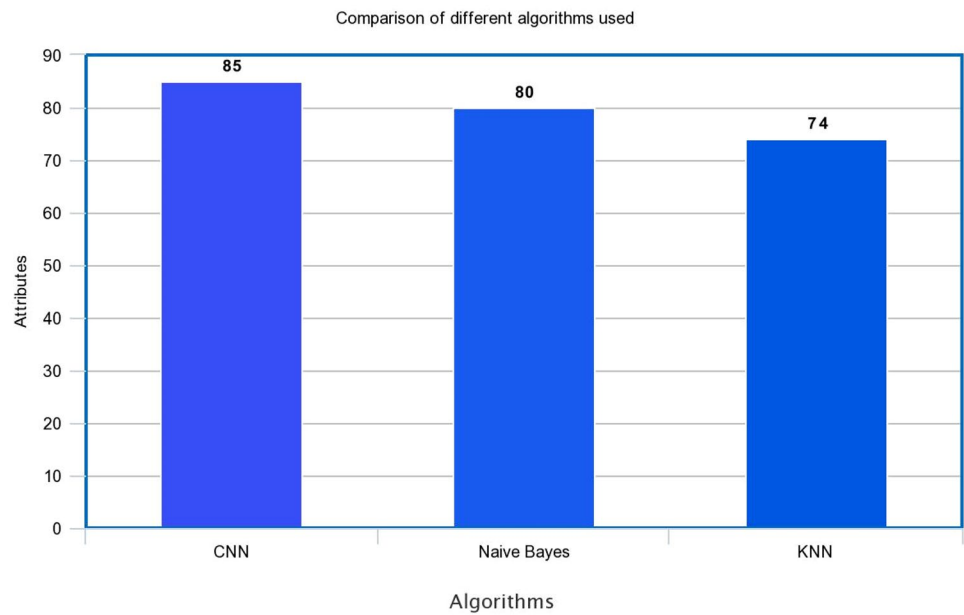
**Fig. 7** Database containing the attributes for predicting heart disease

**Fig. 8** Class labels of test data from the dataset with the actual class labels





**Fig. 9** Graphs of different algorithms



have used a very good optimization algorithm to boost the accuracy power of CNN algorithm.

## Conclusion

Convolutional Neural Network algorithm is a means for early heart disease risk determination using structured data. The accuracy obtained using our model goes up to 85–88%. For future endeavors, we propose to extend our algorithm to incorporate unstructured data as well. As of now, all attributes and laboratory tests considered have been approved by medical doctors.

## References

1. Sayad AT, Halkarnikar PP. Diagnosis of heart disease using neural network approach. *Int J Adv Sci Eng Technol*. 2014;2:88–92.
2. Singh P, Singh S, Pandi-Jain GS. Effective heart disease prediction system using data mining techniques. *Int J Nanomed*. 2018;13(T-NANO 2014 Abstracts):121–4.
3. Gomathi K, ShanmugaPriyaa D. Multi disease prediction using data mining techniques. *Int J Syst Softw Eng*. 2016;4(2):12–4.
4. Miranda E, Irwansyah E, Amelga AY, Kom S, Maribondang MM, Kom S, Salim M, Kom S. Detection of cardiovascular disease risk's level for adults using Naive Bayes classifier. *Healthc Inf Res*. 2016;22(3):196–205.
5. Agarap AF. Deep learning using rectified linear units (relu) (2018). arXiv preprint [arXiv:1803.08375](https://arxiv.org/abs/1803.08375).
6. Yin X, Goudriaan JAN, Lantinga EA, Vos JAN, Spiertz HJ. A flexible sigmoid function of determinate growth. *Ann Bot* 2003;91(3):361–71.
7. Kingma, D.P., Ba, J., Adam: A method for stochastic optimization (2014). arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.