American Journal of Applied Sciences and Engineering | *ISSN 2766-7596*
Published by AIR JOURNALS | *https://airjournal.org/ajase*
12011 WestBrae Pkwy, Houston, TX 77031, United States
airjournals@gmail.com; enquiry@airjournal.org

Research article

Predictive Model for the Monitoring and Detection of Heart Disease using Wavelet Based Machine Learning Technique

[1]Nnenna, H. N., [2]Odo, H., [3]Ozoemena, P., & [4]Uzoka, E. C.
[1,2,3]Enugu State University of Science and Technology, Nigeria, [4]Destinet Smart Technologies LTD

Accepted: September 21st, 2022	Published: September 30th, 2022

Citations - APA

Nnenna, H. N., Odo, H., Ozoemena, P., & Uzoka, E. C. (2022). Predictive Model for the Monitoring and Detection of Heart Disease Using Wavelet Based Machine Learning Technique. *American Journal of Applied Sciences and Engineering, 3(5), 1-12. DOI:* https://doi.org/10.5281/zenodo.7153890

ABSTRACT

This research paper on Predictive Model for the monitoring and detection of heart disease using wavelet-based machine learning technique is aimed at enhancing and easing the process of detecting heart diseases efficiently and in an automated manner. This paper adopts the Dynamic Systems Development Model (DSDM) methodology which was originally based on the Rapid Application Development Methodology. This methodology is applied for a fast delivery of the new system within a specified work plan, budget and time. The methodology also features iterative phases such as feasibility and business study, functional and mathematical modelling, implementation and simulations. Artificial Neural Network technique was also integrated with wavelet technique in this study for a clearer productivity and noise reduction during data processing. Results from the simulation of this work were validated using 10-fold validation technique and the Mean Squared Error of 0.0005423 on the average and a regression of 0.9978 on the average performance were achieved.

Keywords: *Heart Disease; Artificial Neural Network; Wavelet Based Machine Learning Technique; Predictive Model*

Vol. 3, No. 5 | 2022 | pp. 1-12 | *DOI: https://doi.org/10.5281/zenodo.7153890*

Introduction

Arrhythmia is type of heart disease which causes improper heart beat (Sidra et al., 2020). This problem is a case where the heart beats very low or too fast, resulting to an irregular pattern. According to (Benjamin, 2007), when this irregular heart beat occurs once in a while, it is normal, but when it continuous, it presents major risk factors such as stroke, brain damage, abnormal blood flow, and even death among other challenges.

The study by Chugh, 2014; Lip, 2012 and Giuseppe (2021) revealed that in the last 20 years, the cases of Arrhythmia had increased by 33%, and has future projection of over 60% by the year 2050. This has remained a matter of concern to the global community and requires urgent attention so as to control and reduced the implications. To address this problem, many studies have proposed the use of various techniques which employed predictive model to monitor and estimate nonlinear heart beat patterns (Sidrah et al., 2020; Purushottam et al., 2016; Mengting et al., 2020; Sharmila and Chellammal, 2018), however despite their success, they also have their limitation which has always been unreliability.

The problem with the conventional heart disease prediction models is that they were developed with Electro Cardiograph (ECG) data which is always attributed with noise. According to Narwaria et al. (2011) noise is a major constituent of ECG signal as interference from the instrument which is instrumentation noise, electrode contact noise, motion artifacts, muscle contraction, base line drift, electrosurgical noise all have the capacity to corrupt the quality of ECG data collected and when used to develop the predictive model like in the conventional systems make them not reliable. Among these noises, the base line wander noise is the most common as it has very low frequency of 0.5 to 0.6Hz and can hide important features of heart disease. This research seeks to address this problem by the development of a signal processing model which ensures that the data used to develop the predictive model are reliable and can be used for intelligent detection of arrhythmia. This when achieved will help control and curtail the global effect of heart disease.

Table 1: Literature Review

Author	Title	work done	Research gap/limitation
Ravish (2014)	Heart Function Monitoring, Prediction and Prevention of Heart Attacks: Using Artificial Neural Networks	Developed a neural network algorithm to be used for monitoring and protection of patients against the issues of heart attack	The performance of the system can be improved by the application of filter
Jayshril (2014)	Prediction of Heart Disease using Multi-layered Perception Neural Network	Performed systematic survey for various machine learning algorithms that can be presented for the diagnosis of heart disease	Noisy arrhythmia data collected for the research
Majid and Omid (2016)	Improving Heart Disease diagnosis by the Evolution of Particle Swarm Optimization Algorithm and Feed Forward Neural Network	Application of an optimization algorithm for monitoring of heart beats while using the neural network for processing the heart data collected during monitoring to decide weather there is heart attack or not	Can be improved on using adaptive filters

Vol. 3, No. 5 | 2022 | pp. 1-12 | *DOI: https://doi.org/10.5281/zenodo.7153890*

Jairam et al., (2016)	Prediction of Heart Abnormally using Particle Swarm Optimization and Radial Neural Network	Developed a hybrid algorithm with ECG dataset and using it to monitor heart issues	Suffers delayed time during training
Alsalamah et al., (2014)	Diagnostic Heart Disease System based on Radial Basis Function Network Classification Approach	Applied the Radial neural network type to develop a diagnostic algorithm for the detection of heart disease	Can be made reliable using adaptive filters

Methodology

This research is devoted to the analysis, implementation, and demonstration of a predictive model for the monitoring and detection of heart disease using wavelet-based machine learning technique. This is a state-of-the-art system proposed for to help facilitate the process of diagnosing patients suffering from arrhythmia which has remained a major problem all over the world today. To this end, the methodology adopted was the Dynamic Systems Development Model (DSDM) methodology, originally based on the Rapid Application Development methodology.

Design Methods

The methods used for the development of the proposed system are data collection, data processing, wavelet transform, training, artificial neural network, prediction.

Data Collection

The ECG data used for the study was sourced from MIT-BIH Arrhythmia database via the physio net repository. The Physio net is a web-based resource that provides both physiologic signals and related open-source software to the biomedical research community. The web site was launched in September 1999 under the body of NIH's National Center for Research Resources (NCRR). It comprises of physio bank component which is the data archive, the physio toolkit (software) and the physio net. The physio bank contained about 35 gigabytes of recorded signals and annotations. The data contained in physio bank is made available via Physio net website or research purpose only. The sample size of the data collected is 96 recorded of heart beat of patients with arrhythmia within the range of 0-128Hz

Data Processing

Heart beat records always contain various form of noise such as muscular noise, frequency noise, base line wander noise among others. The signals when not processed and used to train the artificial intelligence system for time series prediction of heart disease will not be reliable at the end of the day. To address these problems, there is need for signal processing of the noise, so as to provide reliable data for training and hence good result.

Wavelet Transform

The wavelet transform is a mathematical concept developed to convert a signal into another form which either makes certain features of the original signal clearer to study or be identified. This process was used for the transformation of the signal collected from the patient into time-frequency domain. Discrete wavelet transform technique was used due to its ability to simultaneously transform signal simultaneously in both frequency and time domain

Training

After data processing, the next phase of the process is to learn the data using a machine learning algorithm. The learning process is simply a process which enables the algorithm to understand the pattern of heart disease and then store the intelligence as a reference model for time series detection and prediction of heart disease. This was done using the artificial neural network as the selected machine learning algorithm. The neural network was already discussed in detailed in the literature review.

Prediction

This is the process of making time series approximation of the problem sequentially with the reference mode as the base for the decision. In this stage, time series data of the heart beat are collected and then compared it the patterns of the reference model stored during training and then make conclusions based on the results to predict the whether a patient have heart beat.

System Design

System design is the process of developing the elements of a system such as the architecture, modules and components, the different interfaces of those components and the data that goes through them. This is done using structural system design method based on the Level Oriented Design Approach (LORA). In the LORA, there are two general or broad strategies used. The first strategy started with a general definition of a solution to the problem then through a step-by-step process produces a detailed solution (this is called Stepwise Refinement). This is basically dependent on the system requirements and is a top-down process. The other strategy is to start with a basic solution to the problem and through a process of modeling the problem, build up or extend the solution by adding additional features (this is called design by composition). This is independent on the system requirements and is a bottom-up design.

Design of the Wavelet Filter

Wavelet transform was used here as a data processing tool which analyzed the arrhythmia signal which contain noise captured in frequency and time domain. The type of wavelet transform used here is the Discrete Wavelet Transform (DWT) type. In the DWT, the signal is decomposed into two levels such as coarse approximation and detail information. DWT have two sets of functions. Scaling functions are performed by low pass filter and wavelet functions are performed by high pass filter. A signal in time domain is decomposed into different frequency bands by passing it into successive high pass and low pass filters. The original signal is passed through a half band high pass filter which is followed by low pass filter, with the output down sampled. This output is a constituent of various level of wavelet decompositions and the process report until a depth transformation of the data is achieved and then feed for training. The wavelet algorithm is presented below;

Pseudo Code of Wavelet Transform

The following steps are used to provide to apply wavelet transform on the ARRHYTHMIA signal:

1. Start
2. *Load data captured from physio bank*
3. *Identify time and frequency function*
4. *Activate scaling and wavelet function*
5. *Decompose into time domain with high and low frequency bands*
6. *Get the depth transform of the data*
7. *Return*
8. End

Design of the Artificial Neural Network Predictive Model

Artificial Neural Network is an interconnected massive parallel computational models, units or nodes, whose functionality mimic the human behaviour by processing information from the input to the output using the connection strength (weight) obtained by adaptation or learning from a set of training algorithm. The mathematical description of the ANN was presented starting with a single neuron. The neuron is a unit of computation that reads the inputs extracted from the wavelet transform, processes the input and gives the output in processed form. To get the output of the Artificial Neuron from the activation function, we compute the weighted sum of the inputs as;

$$v_k = \sum_{i=1}^{N} w_{ki} x_i \tag{1}$$

Where;
x_i is the neuron's input from the wavelet transform;
w_{ki} is the corresponding weight to the input x_i.
The neuron's output is obtained by sending the weighted sum v_k as the activation function φinput that resolves the output of the specific neuron. $y_k = \varphi(v_k)$. A step function with threshold t can be used to express a simple activation as;

$$\varphi(x) = \begin{cases} 1 \; if x \; \geq t \\ 0 \; if x < t \end{cases} \tag{2}$$

However, bias is most time used instead of a threshold in the network to learn optimal threshold by itself by adding $x_o = 1$ to every neuron in the network. The step activation function for the bias becomes;

$$\varphi(x) = \begin{cases} 1 \text{ if } x \; \geq 0 \\ 0 \; if x < 0 \end{cases} \tag{3}$$

For the learning process to speed up and also adaptive learning capacity, multiple neurons were used as a multi layered network of neurons formed by feeding the output of one neuron to the input of another neuron to form a multi layered feed forward neural network. In the system the layers between the input and output layers are termed hidden layers which are made up of a bunch of neuron nodes with each input feed with the arrhythmia signal for learning.

These neurons were connected by a link that has a weight which represents the connection strength between each interconnected neuron. While the neuron weight is denoted by w_{ij}^l link between unit j in layer l and unit i in layer $l+1$. Also b_i^l represents the bias of the unit i in layer $l+1$. The neural network architecture is therefore presented as a function of these weights and the bias of the neurons as a nonlinear auto regressive model in equation 4;

$$(w, b) = (w^1 b^1, w^2 b^2, w^3 b^3, \dots) \tag{4}$$

Let the activation of unit i in layer l be represented by a_i^l, then the input for the layer labelled as L_1; we have $a_i^1 = x_i$ for the ith input of the whole network. Other layers are given by$a_i^l = f(z_i^l)$, where z_i^l is the total weighted sum of the inputs to unit i in layer l in addition to the bias term. The summation function of the total data from the neurons is presented in the model in equation 5;

$$a_{\mathrm{n}}^2 = f(w_{\mathrm{n}1}^1 x_1 + w_{\mathrm{n}2}^1 x_2 + w_{\mathrm{n}3}^1 x_3 \dots \dots \dots \dots \dots \dots \dots . + b_{\mathrm{n}}^1) \tag{5}$$

Where n is the number of input classes from the arrhythmia dataset.
The equation can be re-written as below;

$$h_{w,b}(x) = a_1^2 = f(w_{1\mathrm{n}}^2 a_{\mathrm{n}}^2 + w_{1\mathrm{n}}^2 a_{\mathrm{n}}^2 + w_{1\mathrm{n}}^2 a_{\mathrm{n}}^2 + b_{\mathrm{n}}^1) \tag{6}$$

Where $h_{w,b}(x)$ is a real number representing the output of the ANN, n is the number of inputs from the dataset. The activation function $f(\mathrm{Zn})$ can be applied to vectors in element-wise as$f([z_1, z_2, z_3, , , , , , , \mathrm{Zn}]) = [f(z_1), f(z_2), f(z_3) \, (\mathrm{Zn}),]$. Therefore equation 6 can be written as;

$$h_{w,b}(x) = a^2 = f(z^2) \tag{7}$$

So, for any given layer l with activationa^l, the activation a^{l+1} of the next layer $l+1$ is obtained as;

$$z^{l+} = w^l a^l + b^l, a^{l+1} = f(z^{l+1}) \tag{8}$$

Also, when an ANN has every neuron in each layer connected to the neurons in the next layer, it is called a fully-connected network. A nonlinear activation function is used in multilayer networks so as to solve nonlinear issues. The activation function used in the model is the sigmoid functions which are like the logistic function as shown in equation 9;

$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{9}$$

Vol. 3, No. 5 | 2022 | pp. 1-12 | DOI: https://doi.org/10.5281/zenodo.7153890

Where z is the activation function, and when z is large, then e^{-z} tends to zero (0), so $\sigma(z) = 1$. Conversely, if z is a small or very large or very large negative number, then e^{-z} tends to one (1), so $\sigma(z) = 0$.

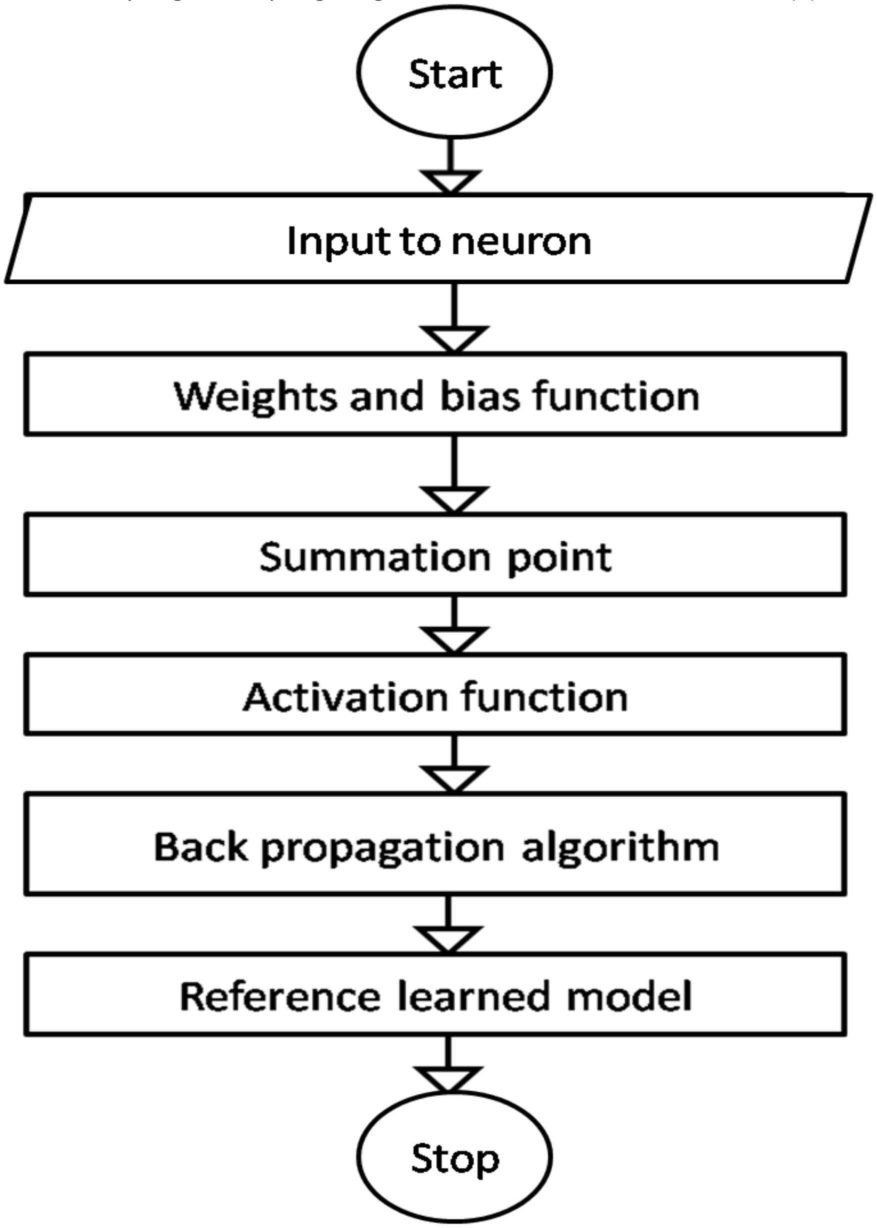

Figure 1: Activity Model of the Neural Network

Implementation

The models were implemented using the mathematical models developed, algorithm and toolbox embedded in Simulink to achieve the new system. The data collected from physio bank was processed with wavelet transform toolbox develop with the algorithm in the pseudocode above and then feed to the neural network neurons modeled in equation 4 for training using the training algorithm in figure 2 to learn the features and implement as a prediction model using neural network toolbox. The neural network architecture configured for the input data is presented in figure 2;

Vol. 3, No. 5 | 2022 | pp. 1-12 | DOI: https://doi.org/10.5281/zenodo.7153890

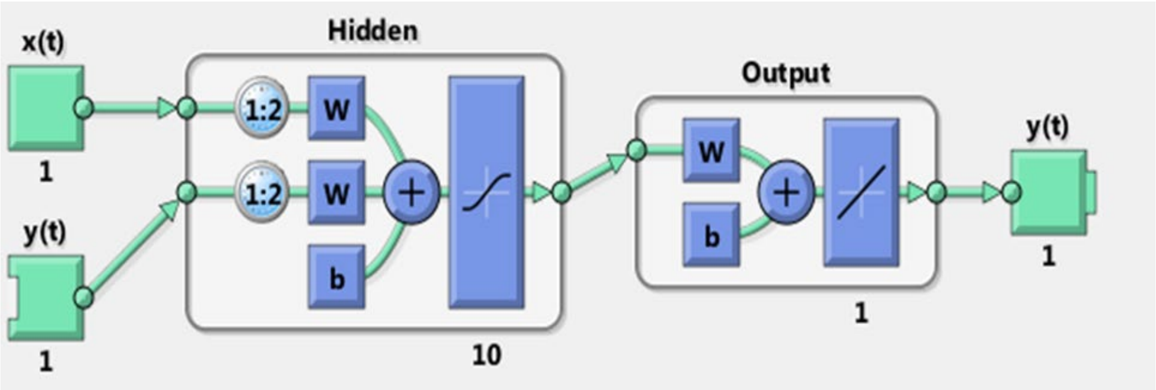

Figure 2: The Neural Network Architecture

The figure 2 presents how the neural network toolbox was used to configure the neural network architecture based on the model in equation 4 with the input of the two classes of the training dataset which represented data of noise and normal arrhythmia rhythm. These data were activated using equation 9 and then train with the back propagation algorithm to learn the data and generate a reference model; The back propagation process was presented as shown in figure 3;

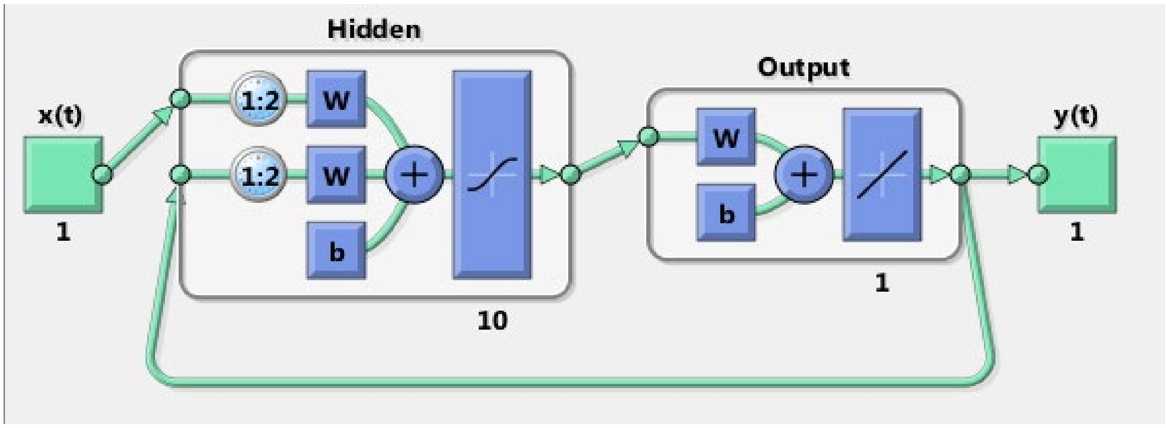

Figure 3: The Back Propagation Process

The figure 3 shows how data from the output are feedback to the input and the weights continuously adjusted until the desired learning result was achieved.
So, through training of different combination of hidden layer neurons and the learning parameter, the optimal number of hidden layer neurons and the learning parameter were obtained as the reference filter model which is the neuro-wavelet filter.

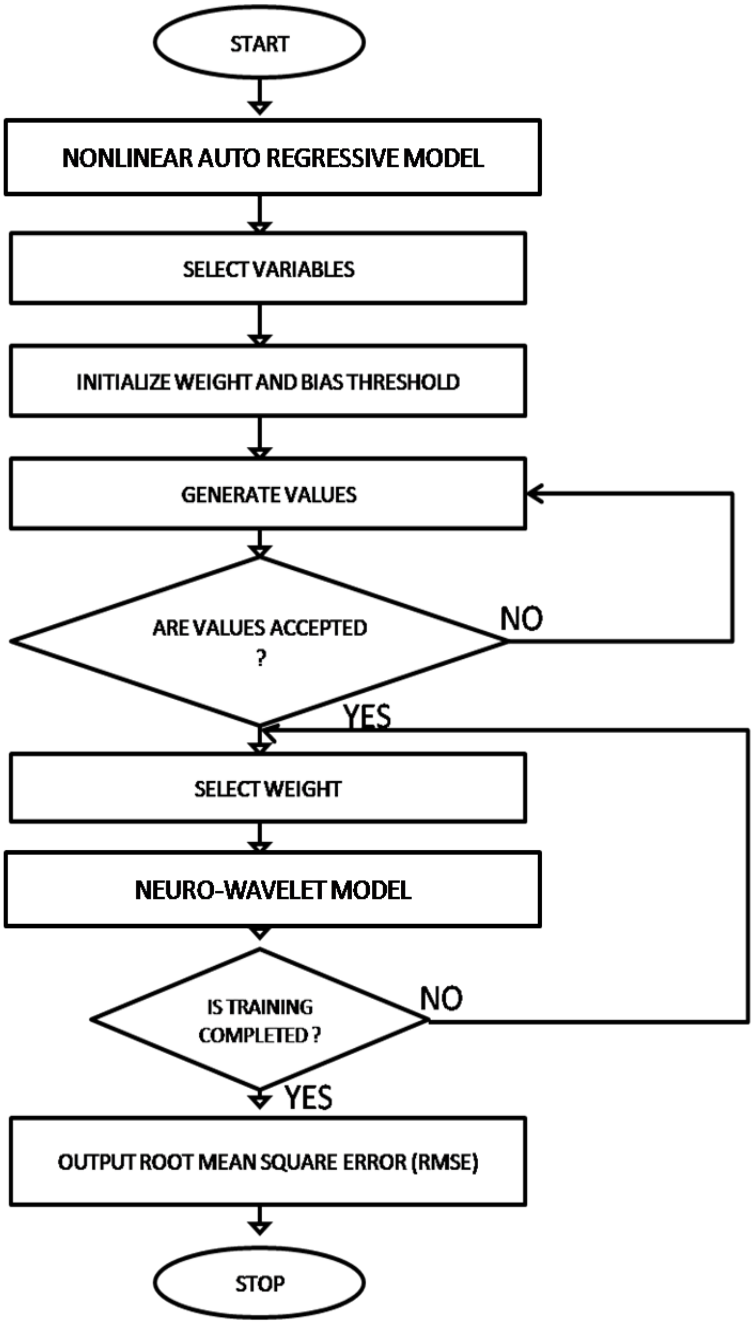

Figure 4: the training algorithm (Back propagation)

Figure 4 presented how the neural network algorithm is integrated with the neuro-wavelet model for an efficient system performance in improving the prediction of heart disease.

Results

The training was done after the neural network architecture was configured and the training algorithm selected with the necessary input parameters based on the setting in table 2;

Vol. 3, No. 5 | 2022 | pp. 1-12 | *DOI: https://doi.org/10.5281/zenodo.7153890*

Table 2: Neural Network Parameters

Parameters	Values
Maximum number of epoch to train	1000
Hidden layer	10
Epoch between display	100
Maximum time to train in sec	Infinity
Maximum validation failure	6
Scale factor for tolerance	0.001
Scale factor for step size	0.1
Initial step size	0.01
Cost horizon	7
Control horizon	2
Input number	2
Gradient	0.840

However, before the training process proceeds the training sets were divided into three subsets as shown in figure 5;

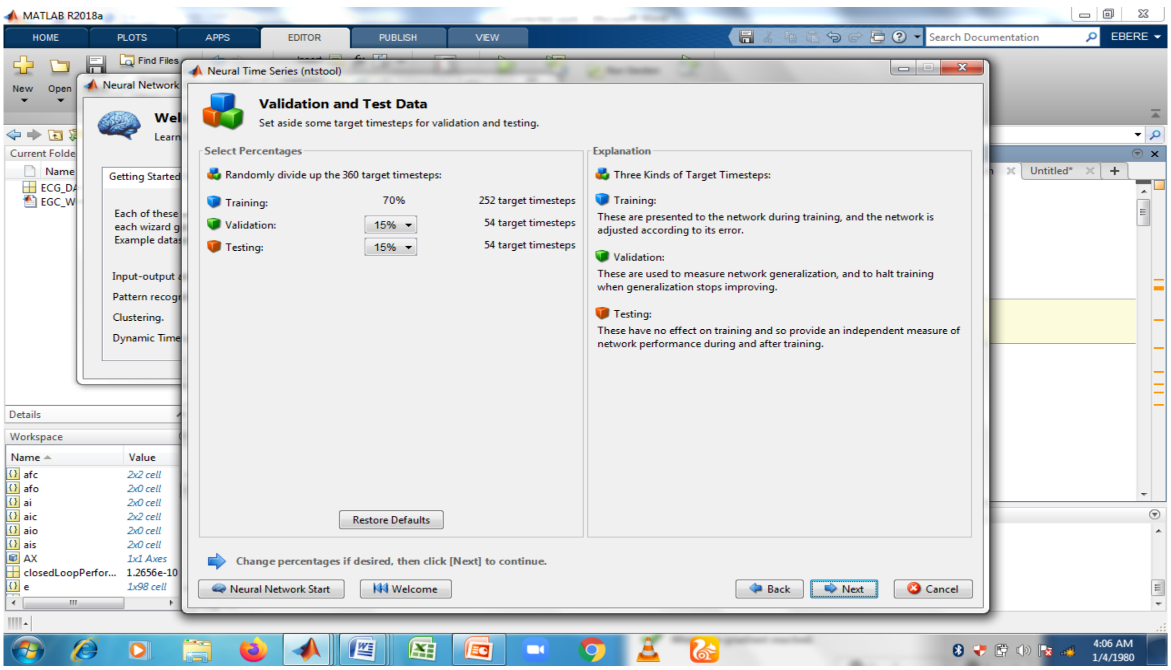

Figure 5: Training, Test and Validation Sets

The figure shows how the data were divided into three ratios of 70:15:15 for training, test and validation sets respectively. The training set was used for the normal training and learning of the neural network, then simultaneously the testing and validation process to ensure that the prediction performance is very précised.

Results

The figure 5 presents the performance of the system after testing using mean square error analyzer;

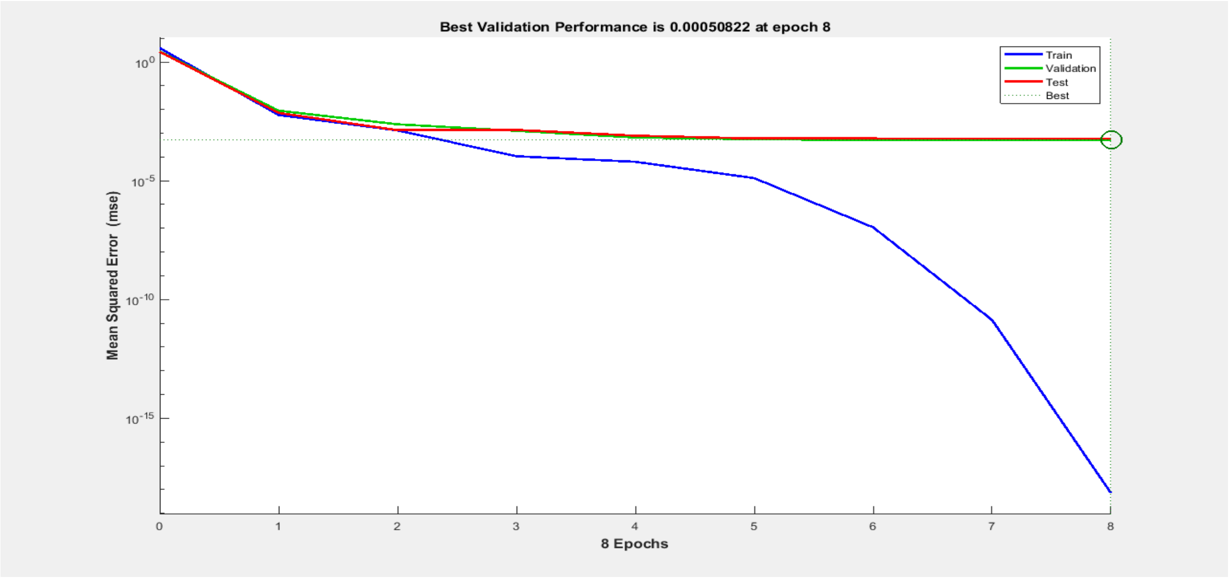

Figure 6: Mean Squared Error (MSE) System Performance

Figure 6 presents how the new system has dropped in its MSE rate across different test iterations. It depicts how the neural network algorithm and adaptive filters learns from the previous errors and uses the back propagation algorithm to improve the training parameters to achieve an optimal result in performance. From the result presented in figure 6, the error rate of 0.000506822 was achieved which is approximately 0 errors. This means that the system successfully operated without any error. The figure 7 will present the regression result of the system.

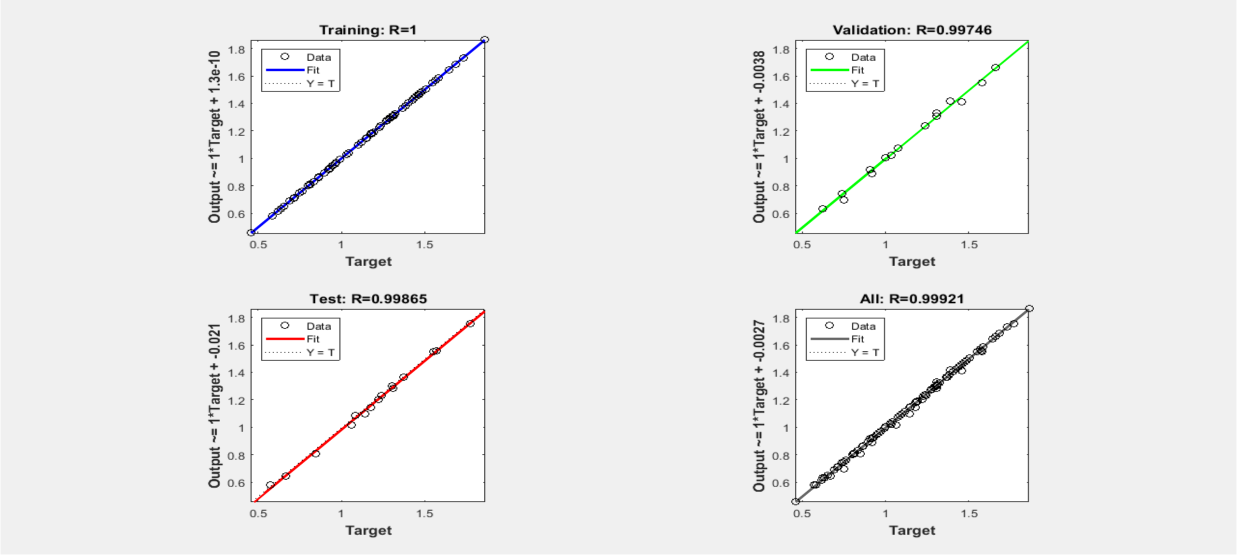

Figure 7: Regression Results of the System

The aim of the regression is to enhance the predictive performance of the system by optimizing the regression value to be equal to 1. From figure 7, it can be seen that the system achieved a regressive value of 0.99921 which is approximately equal to 1. This means that the system presented in this paper achieves an optimal heart disease predicting performance with almost 0 errors in performance.

System Validation

This study applies a 10-fold validation technique to ensure the reliability of the accuracy being recorded for this work. The results achieved from testing this model is presented in table 3;

Table 3: System Validation

Epochs	**Mean Squared Error (MSE)**	**Regression**
1	0.000346	1.105
2	0.000964	0.973
3	0.000767	1.008
4	0.000674	1.205
5	0.00057	0.929
6	0.00087	0.891
7	0.000278	0.793
8	0.000307	0.991
9	0.000294	0.958
10	0.000353	1.125
Average	0.0005423	0.9978

From the values in table 3, it can be seen that the system performance accuracy was gradually becoming optimized as the system performed more iterations of training in the neural network model. Thereby achieving an average optimal value of 0.0005423 on the MSE rates (approximately 0 which is the optimal value for MSE) and 0.9978 on the regression/prediction accuracy which is approximately equal to 1 (the optimal value for regression).

Conclusion

This research has successfully developed a predictive model for the monitoring and detection of heart disease using wavelet-based machine learning technique. The researcher developed a neural network-based security algorithm for the early detection of heart diseases. Dynamic Systems Development Model (DSDM) methodologywas applied for the effective development of the system and implemented with Simulink. The result after testing and validation recorded MSE of 0.0005423 and R= 0.9978. The implication of the results showed that the new algorithm developed was able to correctly monitor, detect and prevent heart diseases especially in the case of arrhythmia. An enhanced data processing and computation in the hidden layers, then right choice of activation used, training algorithm adopted and also the quality of data used to train the neural network and achieve the early heart disease detection algorithm. These features of the new algorithm make it to stand-out from the others with better performance.

Vol. 3, No. 5 | 2022 | pp. 1-12 | *DOI: https://doi.org/10.5281/zenodo.7153890*

References

Alsalamah, M., Amin, S., Halloran, J., (2014). Diagnosis of heart disease by using a radial basis function network classification technique on patients' medical records. IEEE MTT-S International Microwave Workshop Series on RF and Wireless Technologies for Biomedical and Healthcare Applications (IMWS-Bio2014), London. 12(3), 1-4

Benjamin, M. (2006). Introduction to Heart Disease. Available at http://www.mentalhelp.net/poc/view_doc.php?type=doc&id=4496 (Accessed 10 July 2020).

Chugh S.S. (2014). Worldwide epidemiology of a trial fibrillation: A global burden of disease 2010 study. Circulation, PMID 24345399. 9(3), 20-27

Giuseppe Lippi (2021). Global epidemiology of Arrhythmia: An increasing epidemic and public health challenges. *National Library of Medicine International Journal of Strokes* PIMD: 31990238. 8(2), 23-28.

Jairam P. Kelwadeand Suresh Salankar (2016). Prediction of heart abnormalities using Particle swarm optimization in Radial basis Function neural network. *IEEE*:793-797. 5(2), 12-18.

Jayshril S. Sonawane (2014). Prediction of Heart Disease Using Multilayer Perceptron Neural Network. *IEEE*, 5(1), 2-7.

Lip G. (2012). The global burden of atrial fibrillation and stroke: A systematic review of the epidemiology of a trial fibrillation in regions outside North America and Europe. *Chest, PMID*: 22459778, 3 (1), 32-37.

Majid and Omid (2016). Improving the Heart Disease Diagnosis by Evolutionary Algorithm of PSO and Feed Forward Neural Network, *IEEE*, 4 (2), 48-53

Mengting Shen, Luqiao Zhang, Xue Luo and Jing Xu (2020). Arrhythmia Detection Algorithm Based on Manual Extraction Features and Automatic Extraction Features. 7th Annual International Conference on Geo-Spatial Knowledge and Intelligence. IOP Conf. Series: *Earth and Environmental Science* 4 (1), 52-58.

Narwaria, R.P. &Verma, S. & Singhal, P.K. (2011). Removal of Baseline wander and powerline Interference from ECG signal - A survey approach. *International Journal of Electronics Engineering,* 3 (1), 107–111.

Purushottam F., Kanak, S., & Richa, S. (2016). Efficient Heart Disease Prediction System, 8 (3), 962-969.

Ravish D. (2014). Heart Function Monitoring, Prediction and Prevention of Heart Attacks: Using Artificial Neural 2S

Sharmila R, S. Chellammal, T. (2018). A conceptual method to enhance the prediction of heart diseases using the data techniques*, International Journal of Computer Science and Engineering*, 7(3), 56-64.

Sidrah Liaqat, Kia Dashtipour, Adnan Zahid, Khaled Assaleh, Kamran Arshad and Naeem Ramzan (2020). Detection of Arrhythmia Using a Machine Learning Approach. *Information*, 11(2), 540 - 549