



An Efficient and Scalable Heart Disease Diagnosis System with Attribute Impact Based Weights and Genetic Correlation Analysis

Srikanth Meda^{1*}, Raveendra Babu Bhogapathi²

¹ Department of Computer Science, Acharya Nagarjuna University, Guntur 522510, India

² Chalapathi Institute of Engineering & Technology, Lam, Guntur 522034, India

Corresponding Author Email: msk@rvrjce.ac.in

<https://doi.org/10.18280/ria.350105>

ABSTRACT

Received: 17 October 2020

Accepted: 2 January 2021

Keywords:

fuzzy neural networks, cardiovascular disease, attribute impact calculation, genetic correlation analysis algorithm, clustering techniques

Fuzzy neural network (FNN) is playing a vital role in processing of complex data mining applications like medical diagnosis, speech recognition, text processing, image processing etc. Fuzzy neural networks simulate the human brain functionality with fuzzy logic decision making capabilities, to achieve more accuracy in feature selection process of complex data mining applications. Today cardiovascular diseases become a serious global health issue and approximately more than 31% of all global deaths are happening due to cardiovascular diseases reported by WHO. In order to prevent and control the cardiovascular diseases, an efficient and accurate heart disease diagnosis system (HDDS) has to be designed with the state of the art feature based data classifiers. In recent, some research articles introduced HDDS using popular data mining techniques like FNN, but they are suffering from accuracy in allocation of attribute weights and attribute correlation analysis, pattern recognition, forecasting. To address the problems in designing the HDDS, in this paper, Fuzzy Neural Networks has been used with empowered input layer and hidden layers to achieve the high accuracy and performance, while processing the huge set of medical data records. We designed an Attribute Impact calculation procedure to assign the accurate weight values to the attributes and we proposed a Genetic Correlation Analysis algorithm to do correlation analysis which helps in improving the performance.

1. INTRODUCTION

Throughout Cardiovascular Disease (CVD) is the leading cause of mortality in all countries today and approximately 17.7 million of global people are dying out yearly, revealed by WHO in 2016 [1]. coronary artery disease, cardiac arrest, peripheral artery disease, congestive heart failure and ischemic heart disease are the most common types of dangerous heart diseases. Gender, Age, blood pressure (HBP), cholesterol (HDL & LDL), ECG data, diabetes, body mass index value (BMI), heart calcium score (HCS), family history of CVD, smoking, serum creatinine, glycated haemoglobin (HbA1c), stress and mental illness are some of the attributes used in early detection of CVD's. Early detection of CVD risk value from mining the real time medical data records, helps in preventing and control of the heart diseases worldwide.

Today artificial neural networks are widely using in design of different real-time applications like medical diagnosis, speech recognition, text processing, image processing etc. In ANN, supervised learning techniques have been employed with the help of non-linear mathematical models [2] to design the artificial neurons, to behave like the human brain in feature selection, data classification, decision making and forecasting. In recent many studies have been focused on the area of heart disease diagnosis system (HDDS) development using popular machine learning techniques like kNN, fuzzy rules, support vector machine, artificial neural networks and other clustering methods. From the former research articles [3-5] we have noticed that artificial neural networks with fuzzy logic proven

that they are more accurate and reliable than other machine learning techniques. Fuzzy Neural Networks (FNN's) are successfully modeled with genetic algorithms and clinical parameters [6] to forecast the CVD's in prior. Lee [7] and Kahramanli et al. [8] published their articles on how to use the ANN's with Fuzzy logic to select the high impact attributes of dataset from total attributes, to make the processing easy and to improve the performance with more accuracy. Chanda et al. [9] explained that, ANN's are suitable to identify the new patterns from the input data sets of unknown data domain. They also discussed that, the generalization ability of ANN's with fuzzy inference system helps in feature selection and appropriate weigh allocation to the attributes. From our research analysis we observed that the above researches [4-8] were tried to design an efficient HDDS and/or its modules to control the CVD's, but they still remain as unreliable due to the following limitations:

- i) Underutilization of CVD positive datasets while training
- ii) Missing of input attributes correlation mining
- iii) Implementing the regular ANN models for processing and forecasting of CVD's

Earlier HDD systems [5-7] were used the CVD negative or Unconfirmed datasets for training, classification and forecasting. Determining the positive values (CVD occurrence) based on the knowledge of negative features might leads to serious errors as they don't know about the positive features information while classification of data. Allocation of attribute weights using the current manual or predefined static models

will be failed, when they encounter the new set of attributes in future. Improper allocation of static weight values to the input attributes cause to negative impact on result accuracy of ANN's. Henceforth there is a need of designing a dynamic and automatic attribute weight allocation mechanism using the CVD positive and CVD negative data sets.

Another limitation we noticed that none of the HDDS are tried to utilize the power of attribute correlation and its impact on decision making. In their research article [9] proven that, the impact of correlation mining among multiple attributes while decision making in medical diagnosis. We also strongly believe that, the correlation among attributes of CVD dataset will impact positively on the result accuracy of HDDS. For example, a High BP patient with Inverted ECG values is a dangerous symptom, which alarms the CVD attack at any time. So here the both attributes when they adjoined together, along with their weights we have to provide some additional priority as the combination of these both which are very dangerous. We adopted the genetic algorithm model to find the inter linkage among the attributes by mining the correlations.

Finally, we observed that most of the HDDS are following the generic ANN models for classification, clustering, decision making etc. As well-known proverb says "Different strokes for different folks", we have to define a custom ANN model with cardiovascular specific methodology to deal with CVD datasets. Using the new design models we have to customize the process of ANN's hidden layer development to achieve the more accurate results from HDDS output layer. To achieve the best performance in classification of medical data sets we are determined to use the Fuzzy Neural Networks (FNN) instead of the general ANN's. We deeply observed the performance of FNN's from Kuo et al. [10] article and we inspired from that to insist the FNNs in HDDS design.

In this paper we determined to address the above mentioned limitations of various HDDS modules with respective smart solutions. Our Input layer module receives the input data from CVD(+ve) and CVD(-ve) data repositories for weight calculation and to train the neurons of FNN. Selected input attributes correlation value will be calculated with the help of genetic algorithm based on input data knowledge. While processing the input data at several hidden layers of FNN, attributes will be clustered to improvise their impact on results accuracy. Our proposed custom hidden layer with FNN classifies the data using customized neurons with respective weights allocated by using attribute impact algorithm.

The remainder of this paper is organized as follows: Section-2 describes about the literature review on ANN's in development of HDDS. Section-3 represents the proposed HDDS model with empowered features. Section-4 explains a genetic algorithm for Attribute Impact calculation and Correlation analysis. Section-5 describes the Custom FNN Hidden Layer and section-6 concludes the paper with scheduled future works.

2. LITERATURE REVIEW

This section describes the literature review analysis done by the authors as part research work on HDDS with FNN's.

Artificial Neural Networks (ANN): An Artificial Neural Network is a collection of mathematical neurons to mimic the behavior of a human brain to solve the real time scientific problems using input layer, hidden layers and output layer. While dealing with large amount of datasets in machine

learning, ANN's are more reliable and faster when compared to the other machine learning techniques like kNN, fuzzy rules, support vector machine (SVM) etc. In the year of 1992, Akay [11] published an article on designing of a heart disease diagnosis framework using ANN's. Terrin et al. [12] proposed an architecture to describe about the process of heart disease diagnosis mechanism with the help of input medical data records with respective attributes. In general ANN contains input layer, hidden layers, output layer for processing the real time datasets.

Supervised learning and unsupervised learning both are compatible at training process of ANN's input layer. ANN's are mainly categorized as Radial Basis Function Neural Networks (RBFNN), Kohonen Self Organizing Neural Network (KSONN), Feed Forward Neural Network (FFNN), Recurrent Neural Network (RNN), Modular Neural Network (MNN) and Convolution Neural Network (CNN). Each model of ANN is well suitable at a specific domain data processing, i.e. KSONN are good at medical data classification whereas RBFNN are useful at processing of datasets with no separate training data (self-learning). According to the circumstances we have to choose the compliant model of ANN. Today ANN's are proved their prominence in processing of several real world complex data mining applications like voice to text conversion, speech recognition, face recognition, pattern mining, image processing, medical diagnosis etc.

Fuzzy Neural Networks: Artificial Neural Networks (ANNs) recorded very low speed value [5, 13] while processing the huge medical datasets. As ANNs do manipulate each neuron against all possible neurons at each layer, more datasets will be generated for processing. This count will be too high when dealing with the real-time medical data records with dozens of attributes. To overcome the problem of processing the numerous combinations of ANN, several studies [14, 15] proposed the integration of fuzzy logics and genetic systems with ANNs. Using the fuzzy models gained supervised knowledge; we can eliminate the unnecessary neuron combinations before processing them by different ANN hidden layers. Akhil Jabbar et al. [16] proposed NN-FLCS, which is a fuzzy logic control system designed based on ANN's to enhance the learning performance of neural networks with small amount of data records and to classify the data sets with more accuracy by imitating the human brain functionality. Later this technology widely used in share market auto management system design and share trading analysis [3].

Genetic Algorithms with ANNs: GA's are adaptive procedures and follows the Darwin's survival principal. These are the popular search algorithms and proved in processing the high volumes of records with notable performance. Frequently changing genetic operators and their dependent artificial structures helps in calculating the attribute weights and correlations. Selection, crossover and mutation genetic operators help in processing of real-time chromosomes are individuals. GA's are very useful in several data mining operations like pattern mining, data classification, clustering, analysis of large datasets etc.

Heart disease is one of the most dangerous diseases, which scores the highest rate of mortality in the world today [1]. Due to several reasons people are affected by heart disease of different types like coronary artery disease, cardiac arrest, peripheral artery disease etc. Doctors and Medical analysts use to consider several test reports of heart to determine the disease severity and occurrence forecasting. As human being

very few numbers of doctors are with great forecasting capabilities and severity assessment of disease. Sometimes the professional doctors also may predict the wrong value as result, due to the calculation process in disease forecasting is fully complex and tightly coupled with multiple factors (attributes). Persons Gender, Age, Chest Pain level, blood pressure (HBP), cholesterol (HDL & LDL), ECG data, diabetes, body mass index value (BMI), heart calcium score (HCS), family history of CVD, smoking, serum creatinine, glycosylated hemoglobin (HbA1c), stress, mental illness values and disease severity value are some of the attributes plays a vital role in early detection and/or severity determination of heart diseases. To overcome the issues of manual disease forecasting, in 1991 Baxt et al. [6] designed a system with artificial neural networks to find the myocardial infraction using the trained knowledge from clinical parameters. He conducted the analysis on 351 real time patients' data to perform the experiments to forecast the myocardial infraction. In 1992, Akay et al. [11] proposed a genetic algorithm using neural networks for noninvasive diagnosis of cardiovascular problems. Akhil jabbar et al. [16] designed a prediction model using the KNN with Genetic algorithm to diagnosis the heart disease information. The same model with different K values has been applied on total 7 different types of datasets including Heart Disease dataset. Fuzzy rule based clinical decision support system was implemented by Anooj [17] in 2011, which uses the fuzzy logic rules in attribute selection, weights allocation, decision making and forecasting of medical datasets.

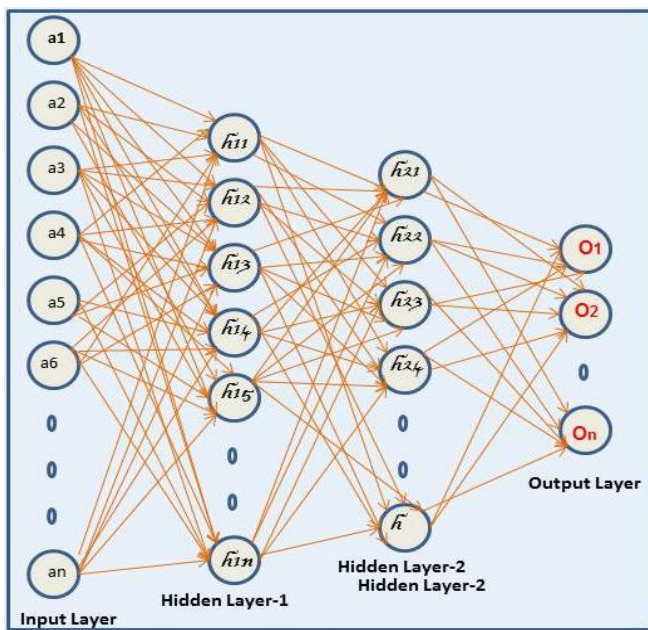


Figure 1. General model of artificial neural networks

Figure 1 describes the general model of ANN's with internal modules input layer, hidden layers and output layer. Each layer input combination is used to be generated as an output, which will be the input to the next consequent layer. As per the requirement number of hidden layers will be formulated and the final one of these hidden layers leads to generate the output which may contain one or more expected output values. The next section explains how we customized the FNN's for designing the HDDS for processing, decision making and forecasting.

3. OUR EMPOWERED HEART DISEASE DIAGNOSIS SYSTEM(HDDS)

In this section, we discuss about the empowered HDDS model with proposed genetic attribute correlation analysis for efficient classification of data and custom hidden layer development for reliable and accurate forecasting.

Proposed HDDS Architecture: Initially we represent the overview of the custom architecture diagram of proposed HDDS with prominent modules and components as shown in Figure 2. A dataset is a collection of data records; each record represents a patient's health information in the form of attributes. The diagram shown CVD+ve and CVD-ve datasets have been prepared from the people who are suffering from heart diseases and who are under diagnosis process respectively. Both datasets are available to the "Dataset preprocessing" module with components like Cleaning, Integration, imputation etc. At this level, noisy data removing, data de-duplication, data transforming, data reduction and missing value imputation [18] operations are performed on raw medical datasets to make data reliable and process able.

After this, the pre-processed data will be sent to Un-supervised learning module for training the system for future classifications and forecasting. In this module first the attributes are analyzed with all possible values and the correlation among the attributes is mined to determine their impact on disease forecasting. Based on their impact identified from correlation mining weights will be allocated using the new allocation process. Finally these weighted attributes will be connected as input to the FNN input layer. Instead of processing the data with General ANN hidden layer model, we are designing a new custom FNN hidden layer model for grouping the attributes to different levels based on the attribute set impact on disease confirmation. Correlations related knowledge of previous module is used in grouping the attributes to certain levels to set the high priority to more impacting attribute group and to set the low priority to less impacting attribute group. This custom model with attribute grouping process protects the diagnosis system results from false positive values effect. The detailed description of proposed custom HDDS model explores the design of each module.

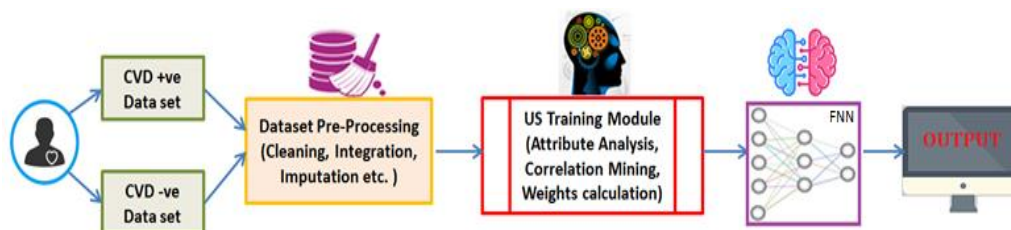


Figure 2. An architecture diagram of proposed HDDS model

Dataset PreProcessing: This is the first process module of proposed heart disease diagnosis model, which cleanse the data, integrates and imputes [18] with missing values. CVD+ve and CVD-ve are the data sets selected as input for this module for training and classification respectively. CVD+ve dataset is a collection health records of heart disease patients who are currently in safe zone with medication, under treatment with serious illness, recently diagnosed and died due to cardiac problem. To record the high accuracy from results, we are preparing the dataset with the data collected from different stages of patients suffering from heart disease. This dataset is used to train the proposed diagnosis model to get the data classification knowledge. This CVD+ve dataset with 250 cardiac patient records have been collected from local government cardiac center, with total 16 attributes by following the patient health record privacy rules.

CVD-ve dataset is the general people health record set, which contains the three different groups of people are patients, healthier and the people with more chances of heart disease occurrence (at max risk to disease). As we know that “prevention is better than cure”, our diagnosis system has to forecast the people who are at the risk of heart disease in future. Using the CVD+ve knowledge we can easily train the model using supervised machine learning data classification techniques and classify the heart disease (+ve) patient records.

We selected the real time Korean dataset KNHANES-VI [15] for training and classification. Each health record of CVD+ve and CVD-ve datasets contains the attributes are Gender, Age, Chest Pain level, blood pressure (HBP), cholesterol (HDL & LDL), ECG data, diabetes, body mass index value (BMI), heart calcium score (HCS), family history of CVD, smoking, serum creatinine, glycated hemoglobin (HbA1c), stress, mental illness values and disease severity value(ds). The disease severity attribute of medical records is having four different stages are mild, moderate and severe. Mild means the person is affected by heart disease very recently and he/she is in safe zone, Moderate means the person already had heart disease since a period of time and now he is under medication to control the disease, severe means the person is under medical treatment in hospital and fighting with disease seriously. This disease severity (ds) value is used as threshold in decision making while allocating the weight to attributes and even in mining the correlations also. By referring different heart disease dataset attributes from the open forums databases [16], we selected these 16 attributes to determine and forecast heart disease. Before utilization of these datasets, they have undergone with many preprocessing techniques like Cleaning, Integration and imputation to make data compatible for FNN processing.

4. UNSUPERVISED CORRELATION ANALYSIS

In this module as described in Figure 2, unsupervised learning process is implemented among the input attributes of data records to find the attribute correlation to calculate the attribute weights. We have chosen the KSONN (Kohonen Self Organizing Neural Network) maps for attribute weighting and correlation analysis, as this is more compatible and widely used neural network type in the field of medical analysis [17]. We designed an unsupervised KSONN custom neural network algorithm for calculating the weights and to identify the correlation value among the attributes for grouping the using a threshold value as shown in algorithm-1. The most common

KSONN algorithm classifies the attributes using the random weight values in two ways are:

$$G(k, a_1) = \begin{cases} 1 & \frac{d}{x} d(a_1, w) \leq \lambda \\ 0 & \frac{d}{x} d(a_1, w) > \lambda \end{cases} \quad (1)$$

or

$$G(k, a_1) = \exp\left(-\frac{d^2(a_1, w)}{2\lambda^2}\right) \quad (2)$$

From the above equations we noticed that the group’s function G calculating the value of a_1 from dataset k . In first method to determine the result value is either 0 or 1, the function d is evaluated with attribute a_1 with a random weight value. The outcome of this value is compared against the threshold λ to determine the result value. In second case, the exponent function is used to evaluate the attribute a_1 value among 0 to 1 (with accurate floating points) instead of either 0 or 1 as mentioned in first equation. The first equation is used to take the direct decisions like true or false, whereas the second equation is used to return the real floating point value with fractional part without ceiling or floor methods to round it. In our approach we are following the second equation model with some customizations as described in the algorithm-1 to return the real floating value as result.

Algorithm-1

Genetic Attribute Impact calculation and Correlation analysis algorithm

Input: CVD+ve dataset P with records $R_1, R_2 \dots R_n \in P$ and attributes $a_1, a_2, \dots a_n \in \forall R_x$

Output: Training Knowledge,

Each attribute a_x with weight $a_{x,w}$ and

attributes correlation set as $G = \{ga_1, ga_2, \dots ga_n\}$

Process Begin:

Set $u =$ size of P , $v =$ col size of R ;

Set $i=j=0$;

Weighted Hash Map $WM =$ new $Map(v)$;

define a matrix var $M1_{(u \times v)}$

// transforming the input data to matrix $M1$

for each record $(R_1 \rightarrow R_n) \in P$ do

for each attrib $a_1 \rightarrow a_n \in R$ do

$M1[i][j] = R_{i+1}(a_{i+1})$;

$i++$;

end

$j++$;

end

reset $i=0, j=0$;

// calculate weights of each column (attribute) of input matrix

// $ds -$ Disease severity

for each column(vector v) $\rightarrow M1$ do

$w(a_{i+1}) =$ setPriority($a_{i+1}, M1(v_i), M1(v_{(ds)})$);

$WM(i,j) = WM(a_{i+1}, \delta)$ // $w(a_{i+1}) : \Leftrightarrow \delta$

$i++, j++$;

end for


```

Echo WM ; // attributes with weights
// correlation analysis using WM with each attribute
weights
reset i=1, j=1;
for each elem of weighted map WM from i → v do
    displayOnGraph(G1, ai, WM(i) );
end
reset i=1, j=1;
Define Distance Matrix M2(v x v);
for each elem of Gi → v do
    for each elem of Gj → v do
        M2(i-1 x j-1) = calculate distance (Gi, Gj);
        i++;
    end
    j++;
end
G = get_correlation_Sets(M2);
return G; // vector with correlation sets
Process End;

```

Algorithm1 receives the CVD+ve dataset P as input, which is a collection of records from R1 to Rn and any record R_c ∈ P contains a set of attributes like a1 to an. Total 16 attributes are selected from the dataset KNHANES-VI to analyze the heart disease information, with the following attributes are age(Age), gender(GN), Chest Pain level (CPL), blood pressure(HBP), cholesterol(HDL), ECG data (ECG), diabetes (DB), body mass index value (BMI), heart calcium score (HCS), family history of CVD(H_CVD), smoking(SM), serum creatinine(SC), glycated hemoglobin(HbA1c), stress(Str), mental illness (MIL) values and disease severity value(DS). The output of the algorithm is an attribute vector with all weighted (δ) attributes, a vector G with identified correlation set like gal to gan and the trained knowledge (TK) for future classifications and forecasting.

Now the algorithm main process begins with transforming the pre-processed CVD+ve dataset to a matrix M1(u x v) for mathematical processing the input data. After transforming each vector (column) of M1 contains an attribute set input data like all patients age values vector, all patients ECG value vector etc. Each vector with a specific attribute of all patients data has given to the setPriority function with attribute name, attribute vector M1 (v_i) and disease severity vector(M1(v(ds))). This setPriority function calculates the weights at each attribute level using the all values from that attribute vector. To calculate the given attribute weight v_i δ using the attribute vector, the values of attribute vector are grouped with the help of disease severity (ds) value. As disease severity is having three possible values, we group each attribute vector values also to three groups like mild, mode rate and severe. At each group level the upper bounds (ub) and lower bounds (lb) is calculated as bounds mean average (σ) and the actual group values mean average is calculated as data values mean average(μ) value.

$$\sigma_1 = \frac{\text{lower bound} + \text{upper bound}}{2} \quad (3)$$

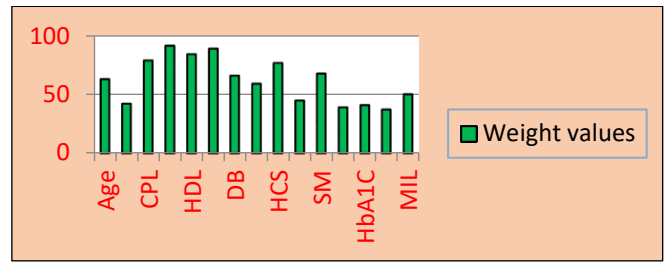
$$\mu = \sum_{i=1}^n \left[\frac{(k + x_i)}{n} \right] \quad (4)$$

Finally these three groups' relevant ub and lb value mean

average intercepted with actual data value mean average values to achieve high accuracy. The whole process is implemented in setPriority() function and finally the priority of the attribute will be returned by this function using the below equation:

$$w_i \delta = \frac{1}{n} + \sum_{i=1}^n (\sigma_i + \mu_i) \quad (5)$$

The below Graph 1 is representing the weight values, which are obtained using the attribute impact calculation, at each attribute level after processing the CVD+ve dataset. Among them Chest Pain Level (CPL) and ECG are recorded as high weighted attributes as they impact on on heart diseases at high rate. Serium Creatinine (sc) and Stress (str) are having low weight value as they have very less impact on heart diseases.



Graph 1. Attribute weight values calculated using attribute impact algorithm

5. CUSTOM FNN HIDDEN LAYER

As described above, with the help of CVD+ve knowledge we can easily train the classification model using KSONN self-learning maps, whereas classification of the complex medical data for disease forecasting we are customizing the FNN hidden layer to achieve more accuracy and to improve the performance and reliability. The trained knowledge from CVD+ve dataset is enough for separating the heart disease suffering patients from the given CVD-ve dataset. Then the leftover records are belongs to the people who are healthier and the people who are at risk to disease in future. The former classification techniques [4, 8] were identified the healthier list and announced the left over people records as at risk to disease. Their process given the people at risk result value accuracy is very low, as the classification was not done with comprehensive analysis.

Our proposed custom hidden layer classifies the data using customized neurons with respective weights allocated by using genetic attribute impact algorithm. In order to reduce the impact of low weight values and to increase the impact of high weight values on determining the results, we are customizing the input neurons as N number of sets based on correlations among attributes. For this correlation calculation we are using the machine learning K-means [19] clustering algorithm and the other novel machine learning techniques [20]. Initially each record from CVD-ve dataset will be taken for forecasting and all attributes of that record will be weighted with previously calculated knowledge. At this time each attribute value with weight value (δ) given as input to K Means technique for clustering the neighbor attributes(q) using the centroids (p) as represented below:

$$c_1 = \sum_{p=1}^n a \sum_{q=1}^n \delta \|a_p^q \cdot a_p^\delta\| \quad (6)$$

After this clustering process, each cluster will be evaluated as part neural network hidden layer to forecast the people records who are at risk to cardiovascular disease. As described by Prasad and Aruna [21], we implemented the hidden layers and forecasting using our proposed correlation based group (attribute set) data. The experimental results with CVD+ve and CVD-ve datasets (Figure 3) as training data and real time classifiable data, proven that our proposed approach is having the notable accuracy in forecasting the medical data, scalability and reliability [22, 23].

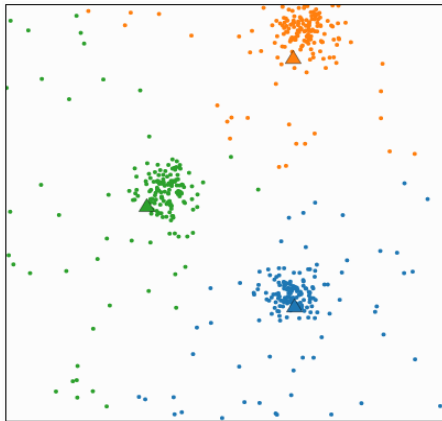


Figure 3. CVD-ve data set process results healthier (green), at risk (blue), CVD patients (orange)

6. CONCLUSIONS

Today artificial neural networks are widely using in medical diagnosis as they are well suitable in complex data processing. As per our research analysis ANN's are suffering from accuracy and scalability limitations while processing the medical data due to Underutilization of CVD positive datasets, Missing of input attributes correlation mining and implementing the regular ANN models for processing and forecasting of CVD's. In this paper, we proposed FNNs to design an accurate and scalable Heart Disease Diagnosis System (HDDS) with attribute correlation mining and CVD specific custom hidden layer development, to forecast people about to risk with heart diseases. To obtain the more accurate results we applied the attribute impact algorithm on CVD+ve dataset to calculated weights and to perform correlations among them. Simulations conducted on CVD-ve dataset proven that our process achieved high accuracy and scalability due to the utilization of FNNs with genetic correlation algorithm and custom hidden layer development with KMeans clustering techniques. In future, we are planning to extend this research work to a well-defined framework to verify the integrity of different modules and the interconnectivity among them.

REFERENCES

[1] Samantaray, L., Hembram, S., Panda, R. (2020). A new Harris Hawks-Cuckoo search optimizer for multilevel

thresholding of thermogram images. *Revue d'Intelligence Artificielle*, 34(5): 541-551. <https://doi.org/10.18280/ria.340503>

[2] Tommandru, S., Sandanam, D. (2020). An automated framework for patient identification and verification using deep learning. *Revue d'Intelligence Artificielle*, 34(6): 709-719. <https://doi.org/10.18280/ria.340605>

[3] Jabbar, M.A., Deekshatulu, B.L., Chandra, P. (2013). Classification of heart disease using artificial neural network and feature subset selection. *Global Journal of Computer Science and Technology Neural & Artificial Intelligence*, 13(3): 4-8. <https://computerresearch.org/index.php/computer/article/view/367>.

[4] Liu, J., Tang, Z.H., Zeng, F., Li, Z., Zhou, L. (2013). Artificial neural network models for prediction of cardiovascular autonomic dysfunction in general Chinese population. *BMC Medical Informatics and Decision Making*, 13(1): 1-7. <http://www.biomedcentral.com/1472-6947/13/80>

[5] http://www.who.int/cardiovascular_diseases/global-hearts/Global_hearts_initiative/en/, accessed on 17 Mar. 2019.

[6] Baxt, W.G. (1991). Use of an artificial neural network for the diagnosis of myocardial infarction. *Annals of Internal Medicine*, 115(11): 843-848. <https://doi.org/10.7326/0003-4819-115-11-843>

[7] Lee, C.C. (1990). Fuzzy logic in control systems: Fuzzy logic controller. I. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(2): 404-418. <https://doi.org/10.1109/21.52551>

[8] Kahramanli, H., Allahverdi, N. (2008). Design of a hybrid system for the diabetes and heart diseases. *Expert Systems with Applications*, 35(1-2): 82-89. <https://doi.org/10.1016/j.eswa.2007.06.004>

[9] Chadha, R., Mayank, S., Vardhan, A., Pradhan, T. (2016). Application of data mining techniques on heart disease prediction: a survey. In *Emerging Research in Computing, Information, Communication and Applications*, pp. 413-426. <https://doi.org/10.1007/978-81-322-2553-938>

[10] Kuo, R.J., Chen, C.H., Hwang, Y.C. (2001). An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network. *Fuzzy Sets and Systems*, 118(1): 21-45. [https://doi.org/10.1016/S0165-0114\(98\)00399-6](https://doi.org/10.1016/S0165-0114(98)00399-6)

[11] Akay, M. (1992). Noninvasive diagnosis of coronary artery disease using a neural network algorithm. *Biological Cybernetics*, 67(4): 361-367. <https://doi.org/10.1007/BF02414891>

[12] Terrin, N., Schmid, C.H., Griffith, J.L., D'Agostino Sr, R.B., Selker, H.P. (2003). External validity of predictive models: A comparison of logistic regression, classification trees, and neural networks. *Journal of Clinical Epidemiology*, 56(8): 721-729. [https://doi.org/10.1016/S0895-4356\(03\)00120-3](https://doi.org/10.1016/S0895-4356(03)00120-3)

[13] Jagielska, I., Matthews, C., Whitfort, T. (1999). An investigation into the application of neural networks, fuzzy logic, genetic algorithms, and rough sets to automated knowledge acquisition for classification problems. *Neurocomputing*, 24(1-3): 37-54. [https://doi.org/10.1016/S0925-2312\(98\)00090-3](https://doi.org/10.1016/S0925-2312(98)00090-3)

[14] Rajeswari, K., Vaithyanathan, V., Neelakantan, T.R.

- (2012). Feature selection in ischemic heart disease identification using feed forward neural networks. *Procedia Engineering*, 41: 1818-1823. <https://doi.org/10.1016/j.proeng.2012.08.109>
- [15] Cook, S., Ladich, E., Nakazawa, G., Eshtehardi, P., Neidhart, M., Vogel, R., Windecker, S. (2009). Correlation of intravascular ultrasound findings with histopathological analysis of thrombus aspirates in patients with very late drug-eluting stent thrombosis. *Circulation*, 120(5): 391-399. <https://doi.org/10.1161/circulationaha.109.854398>
- [16] Akhil Jabbar, M., Deekshatulua, B.L., Chandra, P. (2013). Classification of heart disease using K- nearest neighbor and genetic algorithm. *Procedia Technology*, 1: 85-94. <https://doi.org/10.1016/j.protcy.2013.12.340>
- [17] Anooj, P.K. (2012). Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. *Journal of King Saud University-Computer and Information Sciences*, 24(1): 27-40. <https://doi.org/10.1016/j.jksuci.2011.09.002>
- [18] Korea Center for Disease Control and Prevention, The Six Korea National Health & Nutrition Examination Survey 2013 (KNHANES VI), January 2017, <http://knhanes.cdc.go.kr/>, accessed on 17 Mar. 2019.
- [19] Mumtaz, K., Duraiswamy, K. (2010). A novel density based improved k-means clustering algorithm–Dbkmeans. *International Journal on Computer Science and Engineering*, 2(2): 213-218.
- [20] <https://archive.ics.uci.edu/ml/datasets/heart+Disease>, accessed on 17 Mar. 2019.
- [21] Prasad, R., Aruna, C. (2016). Scalable and accurate missing value imputation with least-missing-column-values-impute-first and K-NN clustering strategies. *Advanced Science Letters*, 22(10): 2876-2880. <https://doi.org/10.1166/asl.2016.7097>
- [22] Narayana, M.S., Babu, N., Prasad, B.V.V.S., Kumar, B.S. (2011). Clustering categorical data--study of mining tools for data labeling. *International Journal of Advanced Research in Computer Science*, 2(4).
- [23] Hartati, S. (2010). A kohonen artificial neural network as a DSS model for predicting CAD. In 2010 International Conference on Distributed Frameworks for Multimedia Applications, pp. 1-5.