Predicting heart failure using deep neural network

1st Minh Tuan Le
International University, Vietnam National University Ho Chi Minh City,
Ho Chi Minh City, Viet Nam
lmtuan@hcmiu.edu.vn

2nd Minh Thanh Vo
International University, Vietnam National University Ho Chi Minh City,
Ho Chi Minh City, Viet Nam
vmthanh@hcmiu.edu.vn

3rd Linh Mai
International University, Vietnam National University Ho Chi Minh City,
Ho Chi Minh City, Viet Nam
mlinh@hcmiu.edu.vn

4th Son V.T Dao
International University, Vietnam National University Ho Chi Minh City,
Ho Chi Minh City, Viet Nam
dvtson@hcmiu.edu.vn

***Abstract*— In the current health system, the diagnosis of heart failure is a difficult task and plays an important role in the early and effective treatment of patients. It is also based on the available diagnose data, from which a medical professional can make the best diagnosis for a patient. This process is very complicated, so with the development of machine learning, medical professionals will be supported to be able to make predictions of early heart failure with high accuracy. In this research, we want to predict heart failure using multilayer perceptron neural network (MLP). The prediction of heart failure dataset with the highest accuracy of 88% is better than the other research.**

Keywords—heart failure, machine learning, multilayer perceptron

I. INTRODUCTION

The term "heart failure" refers to a condition in which the heart's contraction is not being as effective as it should be. Heart failure itself is not a disease but has a variety of causes. In the majority of cases, heart failure is a decrease in the ability of the heart to contract, leading to a decrease in blood flow to the organs of the body, and water stagnation in the organs, so it is called heart failure. congestion (blood stasis). The best way to prevent heart failure is to control heart failure risk factors and conditions, such as coronary artery disease, high blood pressure, high cholesterol, diabetes, or obesity [1]

Edema in heart failure is a result of water stasis in the body. Because when heart failure, the contractile force of the heart muscle relaxes, blood to the organs in the body is not as complete as normal. The body tries to find a way to compensate for the deficiency by secreting substances that make the heart contract harder and faster. The immediate effect is that the heart squeezes better, but in the long run it will reduce the function of the heart.[2]

When the blood flow to the kidneys is insufficient, the organ stores water and salt in the body, which, if normally, must be eliminated through the urine. At first, water and salt are retained to increase blood circulation in the body, supplying the kidney with the lack of blood. This is the body's compensation mechanism; it works for a period of time. If not detected and treated in time, this compensation mechanism will not only lose its effectiveness, but also worsen heart failure.

Excess fluid, just accumulating slowly up. When the fluid accumulates much, it will seep through the blood vessel wall, causing water retention in many organs. Stagnant fluid in areas like the legs causes edema. Simultaneously, fluid accumulates in the pleural space, making it difficult to breathe more heavily.[3]

In the medical field, the prognosis of heart failure can be seen as the decision-making process in predicting diseases. Diagnosing the disease with this decision-making process will be less expensive, simple, effective, and higher accuracy. Today, the number of people who get sick and die from heart failure is increasing. Therefore, it is necessary to detect the disease early and accurately to be able to provide the appropriate treatment process to save the lives of many patients. However, due to the complex process and various disease symptoms, it is very difficult to diagnose patients with heart failure, causing delays in treatment. Therefore, it is necessary to develop a predictive system for heart failure to help medical professionals diagnose heart failure early and accurately. Machine learning algorithms such as: PCA, SVM have been successfully applied to help medical expert diagnose various diseases such as heart failure, diabetes. ANN has also been applied by researchers in the medical field [4-6]. In this research, we are applying the multilayer perceptron for predicting heart failure.

The structure of this paper is as follow. The literature of this research is reviewed in the section II. Section III show the methodology and the experimental result in Section IV. Conclusion and future work are described in the last section.

II. LITERATURE REVIEW

Heart Failure Clinical Records Data set has been used in this project, which is records heart failure patient from the Faisalabad Institute of Cardiology and the Allied Hospital in Faisalabad (Punjab, Pakistan). The dataset is available on UCI repository. The target of this binary classification has two categorial value: 1-yes (patient is sick) and 0-no (patient is healthy). The attribute for predicting is “DEATH_EVENT” which is contain two categorical values and is considered as a binary classification problem. Table 1 list the number of instance number of attribute and features of the dataset

Table 1. Dataset description

Number of Instances	Number of Attributes	Features
299	13	age; anaemia; reatinine_phosphokinase; diabetes; ejection_fraction; high_blood_pressure; platelets; serum_creatinine; serum_sodium; sex; smoking; time

The Perceptron is a classic algorithm used to classify two concepts; it is like a linear line dividing point. The perceptron takes inputs and related weights (representing the importance of the input), aggregating them together to produce output, for use in classification. Perceptron's have been known for quite some time with implementations since 1950. A multilayer perceptron (MLP) is an implement of multiple layers of perceptron that are interconnected, forming a simple neural feedforward controller. A Feedforward network is a network of one or more layers of neurons, in which the signal wires travel only one way from the input through the layers, to the output. This MLP is useful in non-linear functions that a single perceptron cannot perform.[7-9]

III. Research Methodology

This system includes two steps: data pre-processing which involved normalization and outlier detection, then followed by decision tree, multilayer perceptron
The outliers in this research is using IQR method of outlier detection and then, decision tree and multilayer perceptron have been used for classifying the heart failure patient. The detail of this method is described in this section.

A. Data Preprocessing

Normalization: before training data, we have to normalize this dataset, since the gradient descent will be effective with normalized (scaled) values. We may get the values in the different scales if we would not normalize the data. In order to adjust weights, our model would take more time to train on this data. However, if we normalize our data by using normalization techniques, we will have numbers in the same scale which will make our model train much faster and gradient descent will be effective in this case.

IQR Method of Outlier Detection: outlier detection is used for pre-processing data IQR (short for "interquartile range") is the middle spread, also known as the quartile range of the dataset. This concept is used in statistical analysis to help draw conclusions about a set of numbers. IQR is used for the range of variation because it excludes most outliers of data.

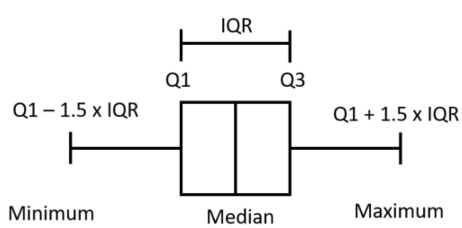

Figure. 1. A Box-Whisker plot

In Figure 1, minimum, maximum is the minimum and maximum value in the dataset. The Median is also called second quartile of the data. Q1 is the first quartile of the data, it means that 25% of the data is lies between minimum and Q1. And Q3 is the third quartile of the data, it says that 75% of the data lies between minimum and Q3. The equation below is the Inter-Quartile Range or IQR, which is the difference between Q3 and Q1

$$IQR = Q3 - Q1 \tag{1}$$

To detect the outliers using this technique, we have to define a new range, which is called a decision range, and any data point lies outside this range is considered as outlier. This range is shown as given below:

$$Low\ Bound: (Q1\ - 1.5 * IQR) \tag{2}$$

$$Uper\ Bound: (Q3 + 1.5 * IQR) \tag{3}$$

Any data less than the *Lower Bound or* more than the *Upper Bound* is considered as an outlier. Our data follows Gaussian Distribution Figure 2

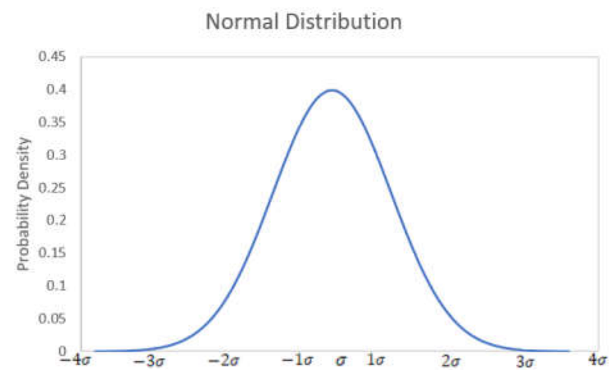

Figure. 2. Gaussian Distribution.

About 68,26% of the data lies within one standard deviation (σ) of the mean (μ) . About 95.44% data lies within (2 σ) of the mean (μ). 99.72% of data lies within (3σ) of the mean (μ). And the rest of the data lies outside (3σ) of the mean (μ). Q1 (first quartiles) and Q3 (third quartiles) lies at -0.675σ and $+0.675\sigma$ from the mean (μ).

B. Parameter Matrix

The dataset contains 299 instances and 12 attributes. Each of these attributes are physiological measurements. The patients in this this dataset includes 194 men and 105 women and the range of their ages between 40 and 95 years old. Features, measurement, and range are listed in Table II.

Table II. Features, measurement, meaning, and range of the dataset

Feature	Explanation	Measurement	Range
Age	Age of the patient	years	[40, ….,95]
Anaemia	Decrease of red blood cells or hemoglobin	Boolean	0,1
High blood pressure	If the patient has hypertension	Boolean	0,1
creatinine phosphokinase (CPK)	Level of the CPK enzyme in the blood	mcg/L	[23, ...,7861]
Diabetes	If the patient has diabetes	Boolean	0, 1
Ejection fraction	Percentage of blood leaving the heart at	%	[14, ...,80]

Feature	Explanation	Measurement	Range
	each contraction		
Sex	Platelets in the blood	binary	0, 1
Platelets	woman or man	kiloplatelets/ mL	[25.01, ...,850.00]
Serum creatinine	Level of serum creatinine in the blood	mg/dL	[0.50, …9.40]
Serum sodium	Level of serum sodium in the blood	mEq/L	[114, ...,148]
Smoking	If the patient smokes or not	Boolean	0, 1
Time	Follow-up period	days	[4, ...,285]
[target] death event	If the patient deceased during the follow-up period	Boolean	0, 1

C. Decision Tree algorithm

Decision Tree works for both categorial and continuous input and output variables. In this method, we have to split the population or sample into two or more homogeneous sets based on most significant splitter / differentiator in input variables. The following steps show the general rules of building a tree:

- Pick the best attributes
- The bet attribute is one which best splits or separates the data
- Ask the relevant question
- Follow the answer path
- Go to step 1 until you arrive to the answer

Firstly, for building tree algorithm to predict the correct output. We have to analyze our data, features and categorical (dummy values), which exist in data. Secondly, to find the best features for our tree, we should work with entropies and discriminative powers with the following formulas:

$$Entropy: H(X) = \sum_{i=1}^{n} P(x_i) * log\ P(x_i) \qquad (4)$$

$$Discriminative\ power = entropy(parent) - (weight\ average) * entropy \quad (5)$$

In our data, an attribute like an age, in order to split it, the middle is found to separate it in two parts. This approach also was used for other attributes. After that, best potential has been found to splits to decide which part will be left or right node of our tree. While building our tree with nodes, we determine best splits and based on this column and value, our data is separated into two parts (left child and right child of node in a tree).

In recursive part, in each level of tree, repeating above explained approaches to build our tree. When there is no question to ask, current node is set as a leaf node.

D. Multilayer perceptron

The multilayer perceptron neural network is indicated that made up of multilayer layer. For some complex problem, which are not linearly separable. Therefore, in order to solve this problem, one or more layers are added in single layer perceptron, so it is known as multilayer perceptron. The MLP network is also known as feed-forward neural network having one or more hidden layer as can be seen in Figure 3.

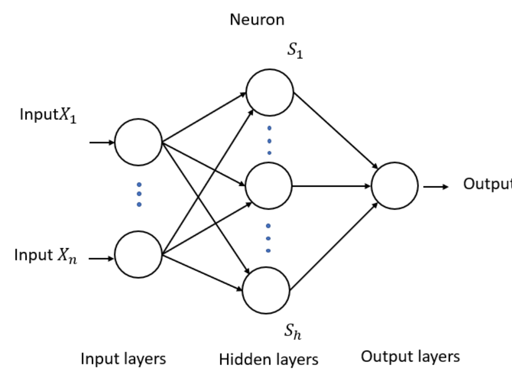

Figure. 3. A multilayer perceptron

In the figure above, this neural network has an input layer with 12 neurons (represent 12 features), one hidden layer with 5 neurons for each hidden layer, and an output layer with 1 neuron (binary classification 0 and 1).

- Input layer: call input variable (x1, …, xp), also called the visible layer
- Hidden layer: the layer of node lies between the input and output layer.
- Output layer: this layer produces the output variables.

Our input shape will be the matrix with the shape of (batch size, number of attributes - 1), because of one hidden layer we will have two weights matrices, one weight will be between input nodes and hidden layers node and another weight will be for interconnection between hidden layer and output layer. Thus first weight will be a matrix with the shape of (number of features - 1 ,number of neuron in hidden) and another weight will be a matrix with the form of (number of neuron in hidden, number of classes) and output layer will be a matrix with the form of (batch size, number of classes).

Mini-batch training technique will be applied to our dataset. Batch size as 4 are supposed to used. So our train matrix will be in shape of (batch size, number of neuron in hidden).We will load the data instance from our dataset as a number of batch size and we will start to train on that batch, moreover we will do forward and backpropagation on that mini-batch. Moreover, after finishing that mini-batch training we will take another mini batch from our dataset and we will go in this order to train and update our parameters once we see all the instances we will finish one epoch of train.

Backpropagation is the algorithm for training some neural networks, such as for training the multilayer perceptron neural network. In this algorithm, weights and biases has been adjusted repeatedly for minimizing the actual output vector of the network and output vector

IV. Exprimental Result

A. Performance Evaluation

A true positive is an outcome where the model correctly predicts the positive class.

A true negative is an outcome where the model correctly predicts the negative class.

A false positive is an outcome where the model incorrectly predicts the positive class.

A false negative is an outcome where the incorrectly predicts the negative class.

Precision: This measure gives the accuracy in cases classified as positive.

$$Precision = \frac{\text{TP}}{TP + FP} \tag{6}$$

Recall: This measure gives the accuracy on positive cases classification.

$$Recall = \frac{\text{TP}}{TP + FN} \tag{7}$$

Specificity: It gives the accuracy on the non-fraud cases classification.

$$Specificity = \frac{\text{TP}}{FN + TN} \tag{8}$$

False Positive Rate: False positive rate is known as false alarm rate. In fraud detection, the legal application is classified as fraud.

$$FPR = \frac{\text{FP}}{FP + TN} \tag{9}$$

Accuracy: The number of correct predictions of positive application divided the total number of predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

B. Result

Decision Tree: with depth 4, the accuracy achieves 86.57%. It can be caused because of having lot of continuous features which requires split data several times.

Multilayer perceptron: the evaluation of this model based different metrics such as accuracy, recall, precision, and f1 score. This model will be sensitive. Because in this type of example we should avoid from False Negative. False Negative means when patient has disease and we say that you are healthy, so we are predicting that he is not sick but in reality, he is sick. There is another situation like False Positive. Predicting that he is sick but in reality, he is not sick. So, compare these two situations, the first situation is much more dangerous since if patient is sick and we say that you are healthy, he is in danger. So, it should avoid from False Negative

Table III. Performance with different number of epochs

# of epochs	Accuracy	Recall	Precision	F1 score
10	0.7941	0.9411	0.8135	0.8727
50	0.8382	0.9166	0.8627	0.8888
100	0.7941	0.9	0.8333	0.8653
150	0.8529	0.9	0.9	0.9
200	0.8676	0.9019	0.9166	0.88

The performance with different number of neurons in Table IV. The highest accuracy of 0.8823 for 10 neurons in hidden layer

Table IV. The performance with different of Neurons

# of neural	Accuracy	Recall	Precision	F1 score
5	0.8088	0.7941	0.7941	0.7647
10	0.8823	0.8775	0.9038	0.8431
15	0.8653	0.8431	0.8392	0.8431
20	0.8737	0.86	0.8703	0.8431

Test error changes averagely after each 10 epochs. If it is increasing, we should stop because we are going to overfit in this case, so this stopping is called early stopping. For example, in Figure 4, the situation is that it should stop to keep weight and prevent increasing of our test error.

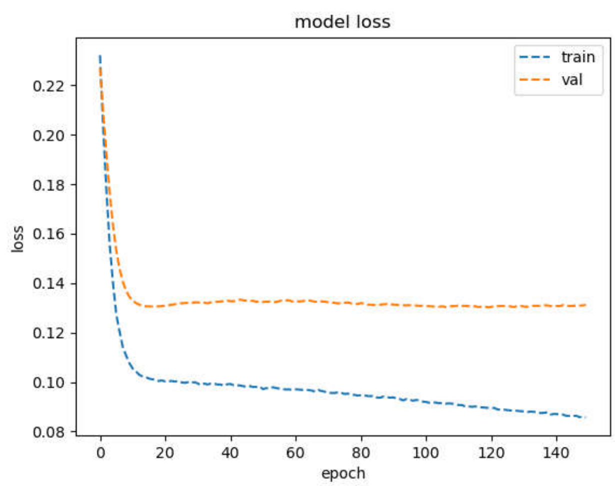

Figure. 4. Model train and validation loss

V. Conclusion and Future Works

This paper represents the application of neural networks in predicting heart failure. We apply the IQR method of outlier detection to the heart failure dataset for detecting and removing outliers. We also apply the multi-layer perceptron techniques have been applied to achieve a higher accuracy.

Comparing with the results shown in Section II, our proposed result achieves an accuracy 88%. In the future, we will propose the technique consists of feature selection to find the best features in the heart failure dataset and classification, which using multilayer perceptron neural network to get higher accuracy.

References

[1] E. E. Tripoliti, T. G. Papadopoulos, G. S. Karanasiou, K. K. Naka, and D. I. Fotiadis, "Heart Failure: Diagnosis, Severity Estimation and Prediction of Adverse Events Through Machine Learning Techniques," *Comput. Struct. Biotechnol. J.*, vol. 15, pp. 26–47, Jan. 2017, doi: 10.1016/j.csbj.2016.11.001.

[2] L. Hussain *et al.*, "Detecting Congestive Heart Failure by Extracting Multimodal Features and Employing Machine Learning Techniques," *BioMed Research International*, Feb. 18, 2020. https://www.hindawi.com/journals/bmri/2020/4281243/ (accessed Aug. 10, 2020).

[3] "Survival analysis of heart failure patients: A case study." https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0181001 (accessed Aug. 10, 2020).

[4] C. Dangare and S. Apte, "A Data Mining Approach for Prediction of Heart Disease Using Neural Networks," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2175569, Nov. 2012. Accessed: Aug. 10, 2020. [Online]. Available: https://papers.ssrn.com/abstract=2175569.

[5] S. Smiley, "Diagnostic for Heart Disease with Machine Learning," *Medium*, Jan. 12, 2020. https://towardsdatascience.com/diagnostic-for-heart-disease-with-machine-learning-81b064a3c1dd (accessed Aug. 10, 2020).

[6] S. Angraal *et al.*, "Machine Learning Prediction of Mortality and Hospitalization in Heart Failure With Preserved Ejection Fraction," *JACC Heart Fail.*, vol. 8, no. 1, pp. 12–21, Jan. 2020, doi: 10.1016/j.jchf.2019.06.013.

[7] M. Jahangir, H. Afzal, M. Ahmed, K. Khurshid, and R. Nawaz, "An expert system for diabetes prediction using auto tuned multi-layer perceptron," in *2017 Intelligent Systems Conference (IntelliSys)*, Sep. 2017, pp. 722–728, doi: 10.1109/IntelliSys.2017.8324209.

[8] M. Jahangir *et al.*, "Auto-MeDiSine: an auto-tunable medical decision support engine using an automated class outlier detection method and AutoMLP," *Neural Comput. Appl.*, vol. 32, no. 7, pp. 2621–2633, Apr. 2020, doi: 10.1007/s00521-019-04137-5.

[9] J. S. Sonawane and D. R. Patil, "Prediction of heart disease using multilayer perceptron neural network," in *International Conference on Information Communication and Embedded Systems (ICICES2014)*, Feb. 2014, pp. 1–6, doi: 10.1109/ICICES.2014.7033860.

[10] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, p. 16, Feb. 2020, doi: 10.1186/s12911-020-1023-5.