International Journal of Science and Healthcare Research
Vol.9; Issue: 2; April-June 2024
Website: ijshr.com
Research Article *ISSN: 2455-7587*

Heart Failure Predictive Analysis Using Decision Tree Classification

Venkata Subbarao Manne[1]

[1]Research Scholar, Dept of Computer Science and Engineering, Jawaharlal Nehru Technological University, Kakinada, India
[1]Assistant Professor, Dept of Information Technology, Vishnu Institute of Technology, Bhimavaram, India.

DOI: https://doi.org/10.52403/ijshr.20240239

ABSTRACT

With an average age of 28 compared to Western countries, India's young population accounts for half of heart attacks in South Asia, which happen to people under 52. Autopsy reports, which identify the actual cause of death, frequently concentrate on sudden deaths in young adults that have no apparent reason or warning signs. Fat accumulation in the blood vessels of the heart is the cause of abrupt, unexpected natural deaths. The heart stops beating and loses blood as a result of these arteries narrowing or blocking. The body may exhibit subtle symptoms prior to abrupt death, such as shortness of breath, palpitations, tightness in the chest, and chest discomfort. A decision tree classification is a death dataset model that generates labelled classes at leaf nodes and makes judgments at edges to predict class labels for subsequent records. The purpose of the proposed paper study is to predict abrupt natural deaths, which are frequently brought on by smoking, by using regression analysis, a statistical technique that establishes the relationship between independent and dependent variables. The experiment's outcome, which looks at how Artificial Neural Networks (ANN) may be used to forecast heart failure, shows five records out of 50,000 patients from different hospitals. Perceptron's, both single- and multi-layer, were used to gather patient information.

Keywords: Artificial Neural Network, Linear Regression Analysis, Sudden and Unexpected Natural Deaths etc.

1.INTRODUCTION

Sudden deaths (SD), usually referred to as "sudden and unexpected natural deaths," are defined as fatalities without a discernible warning indication[1].The main cause is the accumulation of lipids in the coronary arteries that supply blood to the heart the blood vessels are narrowed or blocked causing the heart to lose blood and not be able to beat sudden death is sudden but it does not happen without warning usually 30 minutes or even a few days before sudden death the body will show some strange signs a crushing pain may be felt in the chest before the onset of sudden death usually in the center or left side of the chest even to the left side of the arm and the left side of the face you may experience palpitations chest tightness and shortness of breath[11]. Symptoms of a heart attack include clear heartbeat, sweating, abdominal pain, headache, nausea, and vomiting, indicating potential danger and activation of sympathetic nerves [2]. The text emphasizes the importance of being vigilant about the causes of sudden death and suggests strategies to prevent it [3]. Staying up late and working overtime during nighttime allows for relaxation and decreased blood supply demand, but it also disrupts sleep and sleep quality. This will make the sympathetic nerves too excited heart oxygen consumption increases [4]. Thus, causing sudden death so you have to get as much rest and sleep as possible [5-6]. Many people have heard of sudden death from exercise but in fact it's not the fault of exercise it's a long time without exercise suddenly high intensity exercise leads to increased oxygen consumption of the heart muscle [7-8]. Thus, sudden death exercise must be gradual do not blindly challenge their own limits. Now a days young people eat takeout almost every day unhealthy diet will cause obesity and high blood fat resulting the accumulation of lipids in blood vessels greatly increase the risk of sudden death [9-10]. So, we must pay attention to a healthy low fat balanced diet. Consume more fish for omega 3 fatty acids, which are beneficial for blood vessels [11-12]. If not, use fish oil with high EPA purity, as higher purity promotes lipid metabolism and reduces sudden death risks. Sudden death is akin to a time bomb hidden within your body, causing you no chance of survival once it occurs. So don't go crazy on the brink of sudden death [13-14].

The prevalent developments in artificial neural networks are examined throughout the text. ANNs, which give computer professionals the capacity to perform challenging tasks like pattern identification, planning, and prediction, are a crucial part of machine learning [15-16. Artificial neural networks (ANNs) learn from user experience and repetitive behaviours, much like other machine learning algorithms. They organize text material or visuals related to lung cancer and crisis numbers. Applications for Artificial Neural Networks, or ANN, can be found in chatbots, which are commonly used for text or image classification. The three layers that make up a neural network are typically the input, output, and hidden layers. The input layers are made up of components that convert the input into something that the output layer can use [17]. They are helpful for observing patterns that a human programmer could never extract and train the computer to recognize because of their complexity or sheer number.

2. Proposed Work:

Heart failure is primarily caused by coronary artery disease, fatty deposits, and other risk factors like heart attacks, inherited heart disease, inflammation, high blood pressure, arrhythmia, viral infections, and overactive thyroid. The study revealed a 6.5% annual death rate due to age-specific reasons and pre-existing medical conditions. Lipid buildup in the coronary arteries, which results in heart failure, palpitations, tightness in the chest, and shortness of breath, is frequently the cause of sudden fatalities. The focus of this work is:

- To recognize the Regression Coefficients for Unexpected and Sudden Natural Deaths.
- By using Artificial Neural Network Prediction of Unexpected Death Statistics.
- To predict heart failure by displaying only 5 records from 50,000 patients at various hospitals using single and multi-layer perceptron's.

2.1. Death Data Classification Using Decision Tree Approach: Classification is the task of assigning objects to one of several pre-defined categories. Classification techniques are a systematic approach for building classification models from an input dataset. Many authors discussed several approaches in classification and decision-tree based methods. Decision-tree is fast and easy-to-use approach for rule generation and classification problems. Proposed a general framework based on the decision-tree for machine learning processing persons death data. Many studies have considered positively the induction and analysis of decision trees. In this paper, an input data classification tree for database deaths is introduced. Entire database of the death persons denoted as D which is the root node, divided into two nodes of major objects called Unexpected and Expected nodes representing internal nodes. The object groups being in the Unexpected belong to the internal node. On the other hand, Expected Death which has a group in Expected Death. This uses Decision Tree Classifier to predict future mortality events in heart failure scenarios, aiming to enhance novices' understanding of machine learning techniques. In order to predict class labels for new records, a death dataset model called a decision tree is utilized. The decision tree starts at the root node of death, makes decisions at edges, and outputs the labelled class at leaf nodes like Expected and Unexpected when inner nodes change into Smoking and Natural.

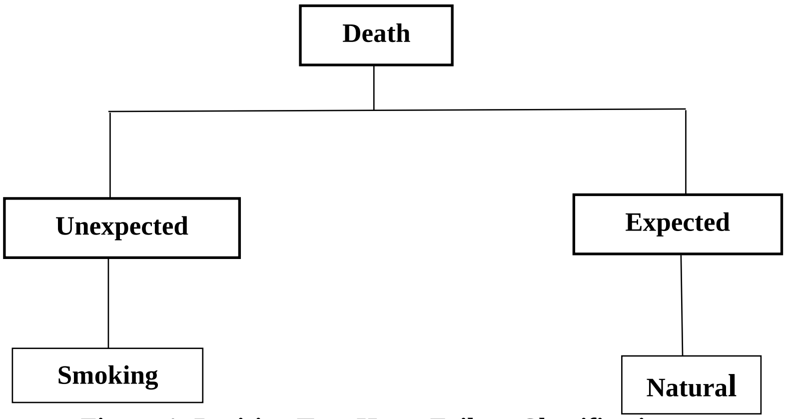

Figure-1: Decision Tree Heart Failure Classification

2.2. Regression Analysis of Unexpected Deaths: Abrupt natural deaths, often caused by smoking, are predicted using regression, a statistical method that estimates the association between independent variables(X) and dependent variables(Y).

Linear regression model: $Y = \beta0 + \beta1\ X + \varepsilon$ ---(1)

The linear regression coefficient β_1 associated with a predictor X is the predictable difference in the outcome Y, when comparing two groups that differ by 1 unit in X. β_1 is the expected change in outcome Y per unit change in predictor X, meaning an increase in X by 1 unit is associated with an increase in Y.

The interpretation suggests manipulating **X** leads to a change in **Y**, a causal relationship, and should be avoided unless the data is from an experimental design. The linear regression model was used to study the relationship between smoking and heart rate.

Heart Rate = $\beta_0 + \beta_1$ Smoking + ε --(2)

The Table-1 provides a comprehensive summary of the outcomes of the model.

The text provides an overview of the various factors that contribute to smoking.

	Factors	Standard Error	p-value
Interrupt	**78.7**	**0.8**	**<0.001**
Smoking	**2.9**	**1.3**	**0.03**

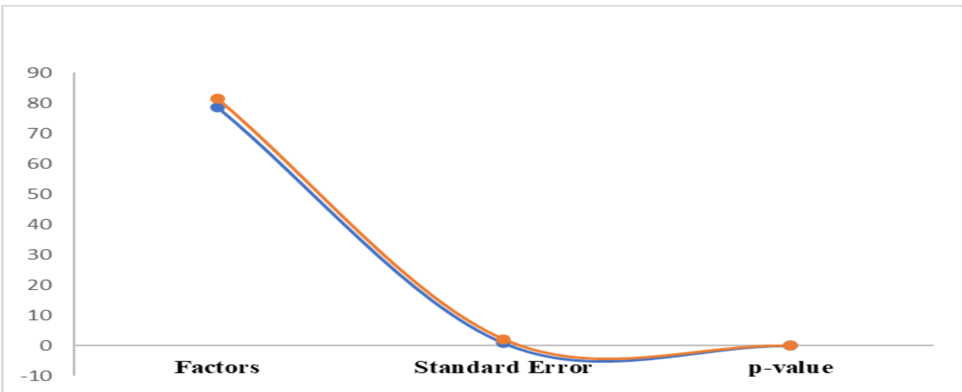

Figure-2: Smoking Factors

The coefficient of smoking is statistically significant ($p < 0.05$) in which implies that within levels of smoking we should expect different average heart rates. But how to interpret the magnitude (Size, Scale or Degree) of this relationship.

1. If there is a binary reaction to smoking (0: non-smoker, 1: smoker): The average heart rate difference between smokers and non-smokers will therefore be $\beta1 = 2.9$. According to the model, smoking raises heart rate by 2.9 beats per minute; however, because of observational data, uncontrolled factors were not taken into account.

Interpreting the standard error: The Standard Error (SE) is a statistical measure used to estimate the uncertainty in the linear regression coefficient. The table provides a useful tool for calculating the p-value and confidence interval for the corresponding coefficient, with SE = 1.3.

The 95% confidence interval can be calculated using the following formula:

95% Confidence Interval = $\beta1 \pm 2 \times SE = 2.9 \pm 2 \times 1.3 = [0.3, 5.5]$.

The range of values that has a 95% probability of include the true value of the parameter being evaluated is known as the 95% confidence interval. There is a 95% confidence interval around 0.3 to 5.5 for the average heart rate difference between smokers and non-smokers. According to the research, the average heart rate of smokers is between 0.3 and 5.6 times higher than that of non-smokers.

Interpreting the interception: For every predictor in the model, the intercept $\beta0$ should be taken as a zero value. Given that smoking = 0, the intercept $\beta0 = 78.7$ shows that the average heart rate among non-smokers is 78.7 beats per minute. For a non-smoker heart rate, the "Interpretation of Linear Regression Intercept" provides comprehensive instructions on how to interpret the intercept in different scenarios.

2. In the event where smoking is a numerical variable (lifetime tobacco use in kilograms), the coefficient $\beta1 = 2.9$ can be understood in this way: The average heart rate difference between two groups of people who differ by one kilogram of lifetime tobacco smoking is 2.9 beats per minute. Alternatively put, we may state this: Among two people with different lifetime tobacco use histories, the one who consumes one kilogram more is likely to have three extra heartbeats per minute.

Interpreting the coefficient of a standardized variable: A measure of a variable having a mean of 0 and a standard deviation of 1 is called a standardized variable, and it is obtained by deducting the mean and dividing the result by the standard deviation. The average heart rate of smokers is **78.7** beats per minute, according to the study's intercept **$\beta_0 = 78.7$**. The expected variation in heart rate between two groups with different lifetime tobacco usage may make the standardized smoking coefficient, **$\beta_1 = 2.9$**, difficult to interpret intuitively. Standardization compares multiple predictors' effects on outcomes, with the largest predictor having the most significant effect. However, it may not yield comparable regression coefficients with different standard deviations or distributions.

3. Classifying smoking as such may lead to information loss if smoking is an ordinal variable (0: non-smoker, 1: light smoker, 2: moderate smoker, and 3: heavy smoker). However, this classification may be useful in scenarios where the predictor and the outcome have a nonlinear relationship. If smoking was divided into several ordered categories, then $\beta1 = 2.9$ would indicate the following: When comparing smokers at one smoking level to the next, it is safe to conclude that they are beating their hearts on average 2.9 beats per minute faster.

4. In the event where there are three levels for the categorical variable of smoking (0: non-smoker, 1: cigarette smoker, and 2: cigar smoker): The unsorted smoking categories denoted by the numbers 0 through 2 are unsuitable for computation. Regression models can incorporate categorical variables with "N" levels once they have been divided into binary variables

with "N-1" levels. The two binary variables, each with a distinct coefficient β, will be produced using the three smoking classifications.

- The first variable, "Smoking cigarettes those who smoke," is coded as follows: "0" if the respondent does not smoke cigarettes, and "1" if they do.
- The second variable, "The cigar smoker," is coded as follows: "0" if the respondent does not smoke cigarettes, and "1" if they do.
- As the reference group, non-smokers will be implicitly (indirectly) coded as if the sum of the codes for cigar and cigarette smokers equals 0, signifying that they are unquestionably non-smokers.

The model becomes: **Heart Rate = $\beta_0 + \beta_1$ Cigarette Smoker + β_2 Cigar Smoker + ε,** where **β_1** will correspond to the difference in the average heart rate between cigarette smokers and non-smokers (the reference group).

β_2 will correspond to the difference in the average heart rate between cigar smokers and non-smokers (the reference group).

The study's findings on the impact of cigarette and cigar smoking on both non-smoking and non-smoking individuals remain unresolved.

2.3. Unexpected Deaths Prediction Analysis Using Artificial Neural Network (ANN): The data on unexpected deaths Text perceives what artificial neural networks view as the prevailing patterns by Zhehao Dai,2024[18]. ANNs, which give computer professionals the capacity to perform challenging tasks like pattern identification, planning, and prediction, are a crucial part of machine learning. Artificial neural networks (ANNs) learn from user experience and repetitive behaviours, much like other machine learning algorithms. They organize text material or visuals related to heart failures (lung cancers) and crisis numbers. Artificial Neural Networks, or ANNs, are widely used for text classification and have applications in the diagnosis of cardiovascular diseases. The three layers that make up a neural network are typically the input, output, and hidden layers. The input layers are made up of components that convert the input into something that the output layer can use. Because of their complexity or sheer quantity, they are useful for seeing patterns that a human programmer could never extract and teach the computer to recognize [16].

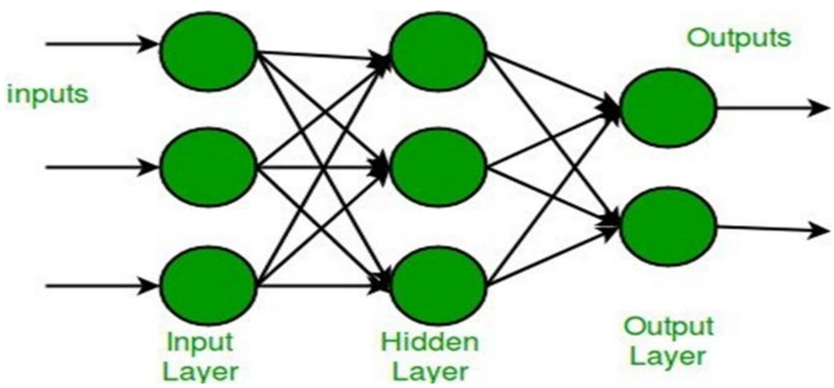

Figure-3: Layers of Artificial Neural Networks

3.Experimental Result: The experimental result explores the use of Artificial Neural Networks (ANN) in predicting heart failure is to display the 5 records out of 50,000 patients at various hospitals collected by leveraging the control of single and multi-layer perceptron's.

Table-2: Patients Information

Age	Anemia	Creatinine Phosphokinase	Diabetes	Ejection Fraction	High_blood_pressure
25	0	582	0	20	1
24	0	7861	0	38	0
25	0	146	0	20	0
26	1	111	0	20	0
27	1	160	1	20	0

The model is trained using x_train and y_train input features and labels, with 200 epochs and batch size 6 for efficiency. Validation data is used to monitor performance on unseen data. Early stopping callbacks monitor validation loss. Verbosity level 2 provides detailed output. Training history stores training and validation metrics over epochs. Dropout (0.5) is a regularization technique in neural networks that randomly sets 50% of input units to zero during training, preventing overfitting and improving model generalization. The final dense layer in a neural network, often used for binary classification, produces a probability value between 0 and 1, with sigmoid activation squeezing the output to the range [0, 1]. The input layer uses all independent features, with 12 units of activation function provided, and is referred to as the ‘relu'. The model uses an Adam optimizer with a 0.001 learning rate for training, and a binary cross entropy loss function for binary classification tasks, with accuracy as a measure.

array ([[1.19294523e+00, -8.71104775e-01,1.65728387e-04, ..., 7.35688190e-01, 6.87681906e-01, -1.62950241e+00],
[-4.91279276e-01, -8.71104775e-01,7.51463953e+00, ..., 7.35688190e-01, -6.87681906e-01, -1.60369074e+00],
[3.50832977e-01, -8.71104775e-01,-4.49938761e-01, ..., 7.35688190e-01, 1.45416070e+00, -1.59078490e+00],
[-1.33339153e+00, -8.71104775e-01,1.52597865e+00,...,-1.35927151e+00, -6.87681906e-01, 1.90669738e+00],
[-1.33339153e+00, -8.71104775e-01,1.89039811e+00,..., 7.35688190e-01, 1.45416070e+00, 1.93250906e+00],
[-9.12335403e-01, -8.71104775e-01,-3.98321274e-01,..., 7.35688190e-01, 1.45416070e+00, 1.99703825e+00]])

Early stopping measures validation loss during training, waits for 10 epochs, and restores best weights from the lowest validation loss to prevent overfitting and retain the best-performing model. The early stopping parameters include monitoring the val_loss, providing 10 minutes of patience, restoring the best weights, and having an EPOCHS of 69.

Epoch 1/200
38/38-3s-loss:0.9452-accuracy:0.5312-val_loss: 0.7542 - val_accu racy: 0.4933 - 3s/epoch -66ms/step
Epoch 2/200
38/38-0s-loss:0.8759-accuracy:0.5938-val_loss:0.7225-val_accuracy:0.4800-256ms/epoch -7ms/step
Epoch 3/200
38/38-0s-loss:0.7662-accuracy:0.5670-val_loss:0.7000-val_accuracy:0.5333-164ms/epoch-4ms/step
Epoch 4/200

38/38- 0s -loss:0.7616-accuracy: 0.6161-val_loss:0.6864-val_accuracy:0.5867-191ms/epoch-5ms/step
Epoch 5/200

38/38-0s-loss:0.6958-accuracy: 0.6384-val_loss:0.6790-val_accuracy:0.6400-194ms/epoch-5ms/step
Epoch 6/200
38/38-0s-loss:0.6796-accuracy:0.6741-val_loss:0.6719-val_accuracy:0.6667-153ms/epoch-4ms/step
Epoch 7/200

The model is trained using X_train and y_train input features and labels, with 200 epochs and batch size 6 for efficiency. Validation data is used to monitor performance on unseen data. Early stopping callbacks monitor validation loss. Verbosity level 2 provides detailed output. Training history stores training and validation metrics over epochs. The text provides information on early stopping parameters, including monitoring, patience, restoration of best weights, and total number of trained epochs based on validation loss history.

The Loss is 0.5872 and Test Accuracy: 69.33%.

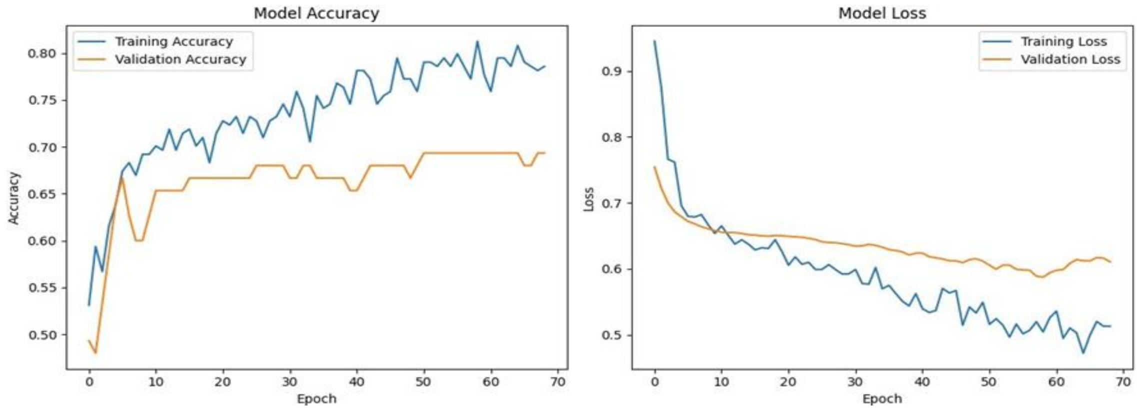

Figure-4: Training Data Accuracy and Loss of Suden deaths

4.CONCLUSION

Heart failure is primarily caused by coronary artery disease, fatty deposits, inflammation, high blood pressure, arrhythmia, viral infections, hyperactive thyroid, and heart attacks. Age-specific causes and pre-existing medical disorders account for 6.5% of annual deaths. Artificial neural networks are being used to predict unexpected and sudden natural deaths.

Declaration by Authors

Ethical Approval: Not Required

Acknowledgement: None

Source of Funding: None

Conflict of Interest: The authors declare no conflict of interest.

5.REFERENCES

1. Allison G. Yow, Venkat Rajasurya, Sandeep Sharma: Sudden Cardiac Death, STAT Pearls, National Library of Medicines, August 8, 2022.
2. A De Groote, M. Wantier, G. Cheron, M. Estenne, and M. Paiva: Chest wall motion during tidal breathing, Journal of Applied Physiology, vol. 83, no. 5, pp. 1531–1537, 1997.
3. E. Aubert, L. Welkenhuysen, J. Montald et al.: Laser method for recording displacement of the heart and chest wall, Journal of Biomedical Engineering, vol. 6, no. 2, pp. 134–140, 1984.
4. Shetty, and C. Naik, Different data mining approaches for predicting heart disease, IIRJET., Vol. 5, no. 9, pp. 217–281, 2016.
5. S. Arthy and G. Murugeswari: A survey on heart disease prediction using data mining techniques, Int. J. Eng. Res. Computer Science Engineering (IJERCSE), vol. 5, no. 4. pp. 280–287, 2018.
6. E.M. Strand, W.T. Jone: A neural network for tracking the prevailing heart rate of the electrocardiogram, DOI: 10.1109/CBMSYS.1990.109420, IEEE Xplore: 06 August 2002.
7. Jyotismita Talukdar EMAIL logo and Thipendra P. Singh: Early prediction of cardiovascular disease using artificial neural network, Journal of Behavioural Robotics, https://doi.org/10.1515/pjbr-2022-0107,Published by De Gruyter Open Access February 17, 2023.
8. J. Tu and J. Lin: Respiration harmonics cancellation for accurate heart rate measurement in non-contact vital sign detection, in Proceedings of the 2013 IEEE MTT-S International Microwave Symposium Digest (MTT), Denver, CO, USA, June 2013.
9. J. Talukdar, and B. Dewangan: Analysis of cardiovascular diseases using artificial neural network, In: International Conference on Parallel, Distributed and Grid Computing (PDGC), IEEE, 2018, pp. 132–137, 10.1109/PDGC.2018.8745900.

10. K. Vembandasamy, R. Sasipriya, and E. Deepa: Heart diseases detection using Naive Bayes algorithm, IJISET-International Journal Innovative Science, Eng. & Technol., vol. 2, no. 9. pp. 441–444, 2015.
11. Madeline Knott, Hope Gillette: What to know about sudden death syndrome, Medical News Today, November 7, 2023.
12. R. M. Arthur, J. R. Cox and F. M. Nolle: Digital analysis of the electroencephalogram the blood pressure wave and the electro. Cardiogram, *Proceedings of IEEE*, vol. 60, pp. 1137-1164, October 1972.
13. Sakaki M, Yoo HJ, Nga L, et al: Heart rate variability is associated with amygdala functional connectivity with MPFC across younger and older adults, Neuroimage, 139:44–52,2016.
14. S. Sharma and M. Parmar: Heart diseases prediction using deep learning neural network model, Int. J. Innovative Technol. Explor. Eng., vol. 9, no. 3. pp. 2244–2248, 2020.
15. Thanh Han-Trong, Hoang Nguyen Viet: An Efficient Heart Rate Measurement System Using Medical Radar and LSTM Neural Network, Journal of Electrical and Computer Engineering, Volume 2022, Article ID 4696163, https://doi.org/10.1155/2022/4696163, 2022.
16. Yi-Chun Chen, Chun-Chieh Hsiao, Wen-Dian Zheng, Ren-Guey Lee, Robert Lin: Artificial neural networks-based classification of emotions using wristband heart rate monitor data, Medicine (Baltimore), PMID:31415420 PMCID: PMC6831309, DOI: 10.1097/MD.0000000000016863, 2019Aug; 98(33):e16863,2019.
17. Y. Iwata, H. T. Thanh, G. Sun, and K. Ishibashi: High accuracy heartbeat detection from CW-Doppler radar using singular value decomposition and matched filter, Sensors, vol. 21, no. 11, p. 3588, 2021.
18. Dai Z, Ko T, Fujita K, Nomura S, Uemura Y, Onoue K, Hamano M, Katoh M, Yamada S, Katagiri M, Zhang B, Hatsuse S, Yamada T, Inoue S, Kubota M, Sawami K, Heryed T, Ito M, Amiya E, Hatano M, Takeda N, Morita H, Yamanishi Y, Saito Y, Komuro I. Myocardial DNA Damage Predicts Heart Failure Outcome in Various Underlying Diseases. JACC Heart Fail. 2024 Apr; 12(4):648-661. doi: 10.1016/j.jchf.2023.09.027.

How to cite this article: Venkata Subbarao Manne. Heart failure predictive analysis using decision tree classification. *International Journal of Science & Healthcare Research.* 2024; 9(2): 289-296. DOI: *https://doi.org/10.52403/ijshr.20240239*

******