

A NOVEL CORONARY HEART STROKE PREDICTION SYSTEM USING MACHINE LEARNING TECHNIQUES

Dr. A. BALAJI¹ CHIGURUPATI BHARGAVI² MAKKENA VASAVI³ VALLALA BHARGAVI⁴ POTHARLANKA KUMARA VENKATA SAI RAGHAVENDRA⁵.

¹Department of Computer Science Engineering, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India.

²Department of CSE & AI, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India.

³Department of CSE & AI, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India.

⁴Department of CSE & AI, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India.

⁵Department of CSE & AI, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India

ABSTRACT: Over the past few decades, cardiovascular diseases have surpassed all other causes of death as the main killers in industrialised, underdeveloped, and developing nations. Early detection of heart conditions and clinical care can lower the death rate. Based on the patient's various cardiac features, we proposed a model for forecasting heart disease and identifying impending heart disease using machine learning techniques. In most cases, input is received through numerical data of various parameters, and output findings are generated in real-time, predicting whether or not the patient has a disease. We'll use a variety of supervised machine learning methods before deciding which one is best for the model. Existing systems rely on classical deep learning models, which are inefficient and imprecise. They aren't as accurate as the proposed model and take a little longer to process.

KEYWORDS: Cardiovascular Disease (CVD), Machine Learning, Deep Learning, WHO (World Health Organization)

1. INTRODUCTION:

Around the world, machine learning is used in many different industries. The healthcare industry is no exception. Machine learning can be quite useful in determining whether or not ailments like loco motor disorders, cardiovascular diseases, and others will exist. If foreseen, such evidence can provide clinicians with insightful knowledge, enabling them to customize their treatment plans and diagnoses. One of the body's major organs is the heart. It forces blood through the blood vessels of the circulatory system. The body's numerous organs are supplied with blood, oxygen, and other materials by the circulatory system, which is essential. The heart is the most important component of the circulatory system. If the heart does not operate properly, it can result in major health problems, including death. The health care sector has vast amounts of medical data; consequently, machine learning algorithms are essential for accurate heart disease prediction. Recent studies have focused on combining these strategies to create hybrid machine learning algorithms. Data pre-processing is used in the research proposal to remove noisy data, fill in blanks when necessary, fill in default values when appropriate, and categorize attributes for prediction and decision-making at multiple levels. To assess the efficacy of the treatment approach, techniques including classification, accuracy, sensitivity, and specificity analysis are performed. An accurate cardiovascular disease prediction model is demonstrated by comparing the levels of accuracy of applying rules to the outcome variables.

2. LITERATURE SURVEY

Theresa Princy.R, et al. (2016) have surveyed different models used for predicting heart disease. The classification techniques used by Theresa were Naive Bayes, Neural network, KNN) LR, Decision tree. The accuracy score was compared with all the models and the comparison was made efficiently [2].

Nagaraj ,M Lutimath, et al.,(2020) research on heart disease prediction using SVM and Naive Bayes classification technique. The way he went by the analysis are Sum of Squared Error, RMSE and Mean Absolute Error, it is established that the SVM emerged as the best classification technique in terms of accuracy.[3].

3. SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

The existing system for cardiovascular stroke prediction relies primarily on classical deep learning models. These models are known to be inefficient and less precise compared to the proposed model. They often require more time for

processing and have lower accuracy. The input data in the existing system typically consists of numerical parameters related to the patient's cardiac features. The output is generated in real-time and provides predictions regarding whether or not the patient has a disease. The system uses traditional machine learning algorithms to make these predictions. However, the inefficiencies and lower accuracy of these models make them less effective in early detection and clinical care for heart conditions, which are critical in reducing the death rate associated with cardiovascular diseases.

LIMITATIONS OF EXISTING SYSTEM

Lower Precision: The existing system may have lower precision and accuracy in predicting heart diseases compared to more advanced machine learning techniques.

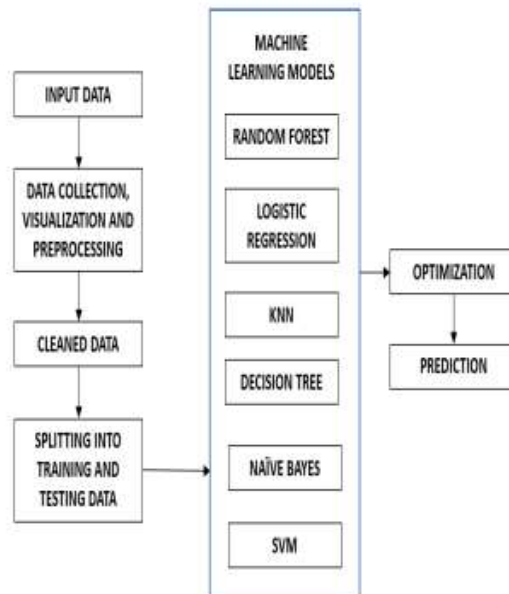
Lack of Scalability: Traditional models may struggle to scale effectively with large datasets or incorporate new data sources and features.

3.2 PROPOSED SYSTEM

In the proposed system, the machine learning model is trained in such a way that the system could predict if the person has the chance of getting cardiovascular disease. It predicts and alerts them about the impending danger. The method used for prediction uses Logistic regression, SVM, Naïve Bayes, Machine Learning algorithms which are widely used in classification-based problem. Based on the characteristics of the patient's heart, the data we have should be divided into several structured data sets.

After data is collected, they are pre-processed to remove all the null values in the dataset. The Null values in the dataset will have significant effect on the training of the ML model. The dataset will have a lot of data which needs to be normalized before they are trained. Data transformation is the process of converting data into a format that is more conducive to data mining.

4. SYSTEM ARCHITECTURE



In the proposed system, the machine learning model is trained in such a way that the system could predict if the person has the chance of getting cardio vascular disease. It predicts and alerts them about the impending danger. The method used for prediction uses Logistic regression, SVM, Naïve Bayes, Machine Learning algorithms which are widely used in classification based problem. Based on the characteristics of the patient's heart, the data we have should be divided



7. CONCLUSION

An increasing death rate due to heart disease is a cause of concern for every citizen. Also, the increase in population that decreases the efficiency of hospitals in giving timely treatment. Therefore, this calls out for an immediate solution. It used machine learning models such as RF, NB, DT etc. It is possible to detect when a patient has a heart-related problem. To reduce the over-fitting of the models, we created synthetic data. We analysed the dataset completely, cleaned the data, and generated a new feature, BMI, in order to improve our model's performance. The Gradient Boosting Classifier performed best in terms of train score 80.29% in training accuracy score, and the ML algorithm works best in terms of the test score, i.e. 72.68%.

Future Enhancement: In future, we can implement multiple feature selection technique that extracts optimal feature to develop models and design an application using real-time data from hospitals.

REFERENCES

- [1] Fahd SalehAlotaibi, "Implementation of Machine Learning Model to Predict Heart Failure Disease" International Journal of Advanced Computer Science and Applications(IJACSA), 10(6), 2019. <http://dx.doi.org/10.14569/IJACSA.2019.0100637>
- [2] J. Thomas and R. T. Princy, "Human heart disease prediction system using datamining techniques," 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT), 2016, pp. 1-5, doi: 10.1109/ICCPCT.2016.7530265.
- [3] J. Thomas and R. T. Princy, "Human heart disease prediction system using datamining techniques," 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT), 2016, pp. 1-5, doi: 10.1109/ICCPCT.2016.7530265.
- [4] Rajdhan, ApurbAgarwal, AviSai, Milan Ghuli, Poonam. (2020). Heart Disease Prediction using Machine Learning. International Journal of Engineering Research and. V9.10.17577/IJERTV9IS040614.
- [5] Suthaharan, S. (2016). Support vector machine. In Machine learning models and algorithms for big data classification (pp. 207-235). Springer, Boston, MA.
- [6] Jiang, L., Cai, Z., Wang, D., Jiang, S. (2007, August). Survey of improving k- nearest-neighbor for classification. In Fourth international conference on fuzzy systems and knowledge discovery (FSKD 2007) (Vol. 1, pp. 679-683). IEEE.
- [7] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30, 3146-3154.
- [8] Selent, D. (2010). Advanced encryption standard. Rivier Academic Journal, 6(2), 1-14.
- [9] Yegnanarayana, B. (2009). Artificial neural networks. PHI Learning Pvt. Ltd.
- [10] Amin-Naji, M., Aghagolzadeh, A., Ezoji, M. (2019). CNNs hard voting for multi-focus image fusion. Journal of Ambient Intelligence and Humanized Computing, 1-21.

[11] Nuttall, F. Q. (2015). Body mass index: obesity, BMI, and health: a critical review. *Nutrition today*, 50(3), 117.

[12] Zheng, A., Casari, A. (2018). *Feature engineering for machine learning: principles and techniques for data scientists.* " O'Reilly Media, Inc."

[13] Cai, J., Luo, J., Wang, S., Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing*, 300, 70-79.