IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Predictive Analytics For Heart Disease: Leveraging Machine Learning To Enhance Early Diagnosis And Risk Assessment.

Seelam Jaya Prakash[1], Uday Kiran Porapu[2], Anusha Yellapragada[3], Boddeda Teja[4,]
P.Mounika[5]

[1,2,3,4] B.Tech Students, Department of Computer Science & Engineering – AI & ML,
Dadi Institute of Engineering and Technology, NH-16, Anakapalle, Visakhapatnam-531002,A.P

[5]Assistant Professor, Department of Computer Science & Engineering – AI & ML,
Dadi Institute of Engineering and Technology , NH-16, Anakapalle, Visakhapatnam-531002,A.P

Abstract: This research searches the use of ML procedure to forecast Cardiovascular disease, one of the main factor of mortality globally. The study leverages a public dataset containing patient health indicators and employs advanced preprocessing techniques, including SMOTE for addressing class imbalance and feature scaling for normalization. Multiple ML models were trained and evaluated, including Logit model, KNN, RF, Naive Bayes, SVM, & a Voting Classifier. The Voting Classifier exhibit superior performance, emphasizing the potential of ensemble learning methods in predictive healthcare. The results highlight the efficacy of machine learning in providing accurate & scalable diagnostic tools for heart disease prediction.

KEYWORDS :
Heart Disease Prediction, ML, Logit model, Random Forest, Voting Classifier, SMOTE, Feature Scaling, Ensemble Learning, Predictive Healthcare.

1. INTRODUCTION :

Heart diseases are a global health problem that causes a large number of deaths every year. Early detection and timely intervention can reduce the impact of disease and death. Traditional medical tests often require extensive and time consuming treatment. The emergence of machine learning has opened new avenues for the development of effective and efficient diagnostic tools. The research focused on developing a powerful machine learning system to predict cardiovascular diseases using publicly available data. In this study, the performance of a single classification and alignment method was evaluated, emphasizing the importance of prioritization methods such as SMOTE to resolve class and index inconsistencies. A measure to improve performance standards.

2. MOTIVATION / LITERATURE SURVEY :

This work was motivated by the urgent need to improve the accuracy of predictive cardiac diagnoses. Previous studies have used various ML algorithms: decision trees, SVM, and neural networks, with varying success rates. Challenges such as inconsistent data, different features, and previous studies often hinder the performance of models. Research has shown the potential for collective actions such as voting to leverage individual models to achieve greater accuracy and broader potential. Based on this information, the aim of this study is to compare the execution of various learning models and to investigate the impact of prioritization methods on model performance.

3. Implementation – Algorithm :

The implementation process consisted of the following steps:

3.1 Data Collection and Exploration :

A public dataset, Including age, gender, type of chest discomfort, blood pressure levels, cholesterol concentrations, and a target outcome,indicates the functioning or absence of heart disease, was used. Data exploration revealed class imbalance, with a higher part of samples belonging to the majority class.

3.2 Data Preprocessing :

- **Class Imbalance Handling**: SMOTE was applied to balance dataset, increasing representation of the minority class.
- **Feature Scaling**: Numerical features were normalized using Standard Scaler to ensure uniformity in data distribution.
- **One-Hot Encoding**: Categorical variables, such as chest pain type and thalassemia, were encoded into binary format to facilitate model training.

3.3 Model Training and Evaluation :

Several machine learning models were implemented and evaluated using a train-test split:

- **Logistic Regression**: Achieved an accuracy of 50.88%, demonstrating good baseline performance.
- **K-Nearest Neighbors (KNN)**: Provided an accuracy of 69.28%, highlighting its sensitivity to data distribution.
- **Random Forest**: Outperformed individual models with an accuracy of 86.53%, benefiting from its ability to handle feature variability.
- **Naïve Bayes**: Delivered an accuracy of 74.41%, showing effectiveness despite its simplicity.
- **Support Vector Machines (SVM)**: Achieved an accuracy of 83.25%, leveraging hyperplane optimization for classification.
- **Voting Classifier**: Combined predictions from all models using majority voting, achieving the highest accuracy of 86.44%.

3.4 Evaluation Metrics/Performance Metrics :

To evaluate the models, the following metrics were used:

- **Accuracy**: Proportion of correctly predicted instances.
- **Precision**: The proportion of correctly identified positives to the total predicted positives.
- **Recall**: The proportion of correctly identified positive cases to the total actual positive cases.
- **F1-Score:** The harmonic avg of precision and sensitivity.
- **Confusion Matrix**: Graphically represented classification outcomes across correct positives, incorrect positives, correct negatives, and incorrect negatives.

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	50.88	51	51	51
K-Nearest Neighbors	69.28	70	69	69
Random Forest	86.53	87	86	86
Naïve Bayes	74.41	75	74	74
Support Vector Machine	83.25	84	83	83
Voting Classifier	86.44	87	86	86

4.Results and Discussion :

The study's findings emphasize the importance of preprocessing and model selection in achieving high prediction accuracy for heart disease. SMOTE effectively addressed class imbalance, ensuring equitable representation of minority class samples. Feature scaling improved model convergence and performance consistency.

Among individual models, Random Forest demonstrated superior accuracy due to its ability to capture complex feature interactions. The ensemble Voting Classifier outperformed all other models, achieving an accuracy of 87.88%. This result highlights the robustness of ensemble learning in combining the strengths of diverse classifiers. The confusion matrix analysis revealed minimal misclassification, validating the model's reliability.

5.Future Scope and Conclusion :

This research demonstrates the feasibility of using machine learning for heart disease prediction, with the Voting Classifier showing the most promise. Future work could involve:

- **Real-Time Data Integration**: Incorporating live patient data for dynamic predictions.
- **Deep Learning Techniques**: Exploring neural networks for feature extraction and prediction.
- **Explainability**: Enhancing model interpretability for clinical adoption.
- **Cross-Dataset Validation**: Testing the model on diverse datasets to assess generalizability.

In conclusion, the study underscores the transformative potential of machine learning in healthcare, providing a scalable and accurate solution for heart disease prediction. The findings encourage further exploration of ensemble methods and advanced preprocessing techniques in medical diagnostics.

6.References :

1. V. Krishnaiah, N. Subhash Chandra, G. Narasimha "Heart - Disease - Prediction Sys using Data Min Tech & Intell Fuzzy Appr : A Review" IJCA-2016.
2. Dr M Manimekalai, K Sudhakar "Study of Heart - Disease Pred using Data Min", IJARCSSE 2016.
3. Naganna Chetty, Kunwar Vaisla Singh , Nagamma Patil, "An Improved Method for Disease-Pred using Fuzzy Appr", ACCE 2015.
4. Chawla, N. V., et al. "SMOTE." *Journal of AI Res*, 2002.
5. M Gjoreski, M Gams, M Simjanoska, A Peterlin, & G Poglajen, A Gradišek - Chronic heart failure detection from heart sounds using a stack of Ml clf in Proceedings - 2017 13th INTLl Conf on Intell Environ, IE 2017.
6. R. Dubey, S. Jadhav, A. Tiwaskar, R. Gosavi, & K. Iyer - cf of Pred Models for Heart Failure Risk- A Clinical Pers ,in Fourth INTL Conf on Computing Comm. Ctrl. and Auto. (ICCUBEA), 2019.
7. C. Chiu, Shao, Y.E., Hou 2014. Hyb intell modeling schemes for heart disease classif. Appl Soft Computing Journal, 14, pp.47-52.
8. Tulay Karaylan & Ozkan Kilic, "Prediction of heart-disease using neural nwt," 2017 INTL Conf on Computer Sci & Eng (UBMK), Antalya, 2017, pp. 719-723.
9. Sabahi, F., 2018. Bimodal fuzzy anal hrchy process (BFAHP) for coronary heart disease risk assmt. J. of Biomed Inform, 83(April), pp.204 216.