iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** II **Month of publication:** February 2023

DOI: https://doi.org/10.22214/ijraset.2023.49331

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

Hyperparameter Optimization for Disease Detection and Analysis

Shital Gajbhiye[1], Omkar Rajendra Bhujbal[2], Nakul Bapusaheb Gite[3], Kshitij Babanrao Khaire[4], Yash Sanjay Kutwal[5]
Department of Information technology, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, India

Keywords: Machine Learning, Self-Reading, Software Development Life Cycle, Unified Modelling Language

I. INTRODUCTION

Health-care information systems tend to capture data in databases for research and analysis in order to assist in making medical decisions. As a result, medical information sys- terms in hospitals and medical institutions become larger and larger and the process of extracting useful information becomes more difficult. Traditional manual data analysis has become inefficient and methods for efficient computer-based analysis are needed. To this aim, many approaches to computerized data analysis have been considered and examined. Data mining represents a significant advance in the type of analytically tools. It has been proven that the benefits of introducing data mining into medical analysis are to increase diagnostic accuracy.

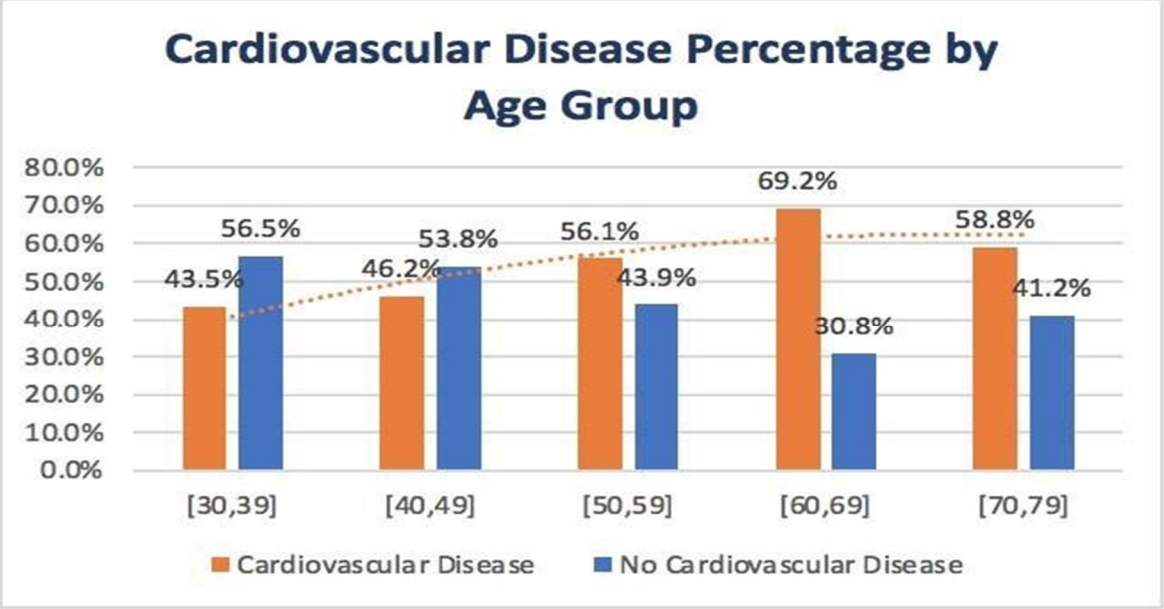


Figure 3

Fig. 1 Cardiovascular Disease Percentage by age group.

Various studies used Machine Learning (ML) algorithms to forecast CVD using clinical datasets. Still, clinical datasets present considerable difficulties owing to class imbalance and their high dimensionality. As a result, employing machine learning without addressing these challenges reduces the efficiency and thus accuracy of the methods. Prior researchers focused on feature selection (FS) and used several ML systems to predict CVD.

Few studies developed support decision systems, including a balancing technique to ad- dress the problem mentioned. developed an HD prediction method, including density-based spatial clustering of applications with noise (DBSCAN), hybrid synthetic minority over-sampling technique-edited nearest neighbor (SMOTE-ENN) to detect and eliminate outliers and balance data distribution. In addition, the classifier forecasts the patient status. Waqar et al. proposed SMOTE-based deep learning to predict heart attacks. The author used SMOTE technique to balance the dataset without feature selection. Recently, Ishaq et al. used SMOTE to balance data distribution and extremely randomized trees (ET) on selected parameters to predict patient survival using RF importance ranking.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

II. AIM

To & develop python based system to detect Heart Disease & Severity Level Clas sification Model Using Machine Learning & Hyperparameter Optimization Methods.

III. LITERATURE SURVEY

A. An Effective Heart Disease Detection and Severity Level Classification Model Using Machine Learning and Hyperparameter Optimization Methods

Cardiovascular disease (CVD) is the leading cause of death worldwide. A Ma- chine Learning (ML) system can predict CVD in the early stages to mitigate mortality rates based on clinical data. Recently, many research works utilized different machine learning approaches to detect CVD or identify the patient's severity level. Although these works obtained promising results, none focused on employing optimization methods to improve the ML model performance for CVD detection and severity-level classification. This study provides an effective method based on the Synthetic Minority Oversampling Technique (SMOTE) to handle imbalance distribution issue, six different ML classifiers to detect the patient. The proposed model can help doctors determine a patient's current heart disease status. As a result, it is possible to prevent heart disease-related mortality by implementing early therapy. [1].

B. Novel Intelligent Model for Heart Disease Prediction using Dynamic KNN (DKNN) with improved accuracy over SVM

In comparison to Support Vector Machine, the major goal is to forecast the Novel Intelligent model for Heart Disease prediction using Dynamic KNN (SVM). Materials and Procedures: Two machine learning methods, Dynamic KNN (N=92) and Support Vector Machine (N=92), are used to predict heart disease. Dynamic KNN is a simple al- gorithm used for disease prediction. Heart disease dataset is used for disease prediction. For each group 20 samples are taken and it is divided into training and testing dataset. Re- sult and Discussion: Accuracy of Dynamic KNN is 84.44% and Support Vector Machine is 67.21%. There exists an analytical significant difference between Dynamic KNN .

C. Prediction of Heart Disease using Machine Learning Algorithms Application of Sup- port Vector Machine Based on Particle Swarm Optimization in Classification and Prediction of Heart Disease

This heart disease is the number one killer of Chinese residents' health. Early detection of heart disease and timely treatment are of great significance to every heart disease patient. In this article, by mining the physical index data of patients with heart disease, aiming at the problem that the optimal parameters in the traditional support vec- tor machine model are difficult to find.

D. Prediction of Heart Disease using Decision Tree in Comparison with KNN to Im- prove Accuracy

The aim of this survey is to use the irregular forest method to forecast cardiac dis- ease and to achieve better forecasting accuracy by implementing machine learning algo- rithms and comparing their performance to KNN. In this research, two groups—Decision Tree and K-Nearest Neighbor—are examined.

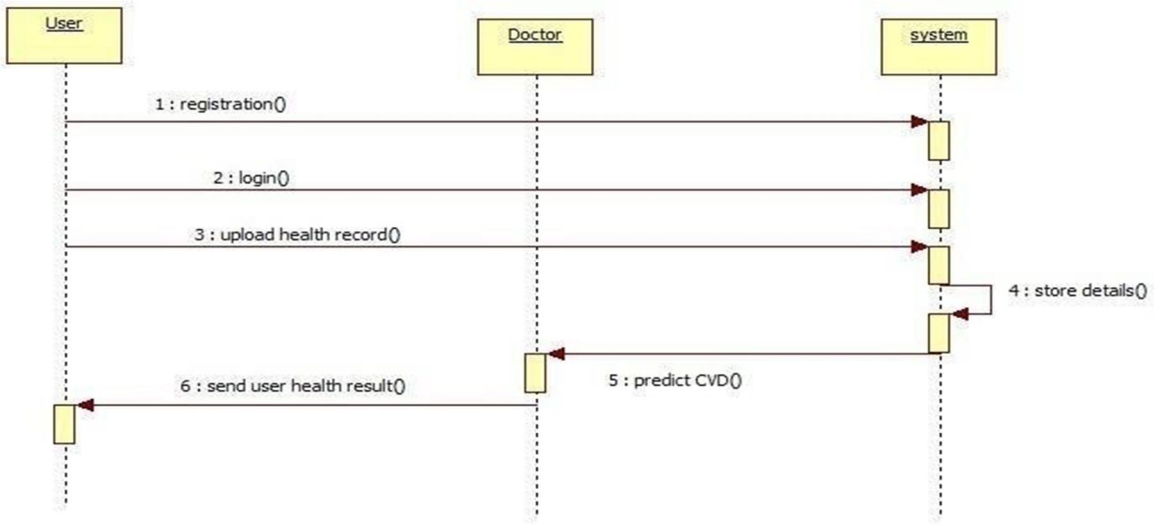
The strategies have been put into practice and tested using a dataset with 1700 records. N=20 iterations on each method are run as part of the programming experiment to identify different grades of model performance with a repeated measures power of 80%. Following testing, the decision tree method and the k-nearest neighbor approach both achieve mean effectiveness of 86.7 percent and 82.5 percent, respectively, for the disease prediction. By using independent samples t-tests, it can be shown that the accuracy of the two algorithms differs statistically significantly (p 0.05). The results of the comparison demonstrate that the Random Forest algorithm performs noticeably better than KNN.

IV. PROPOSED-SYSTEM ARCHITECTURE

Sequential process between user, doctor and system, are shown where user registers himself and login to upload the health records to cloud. The doctor checks the symptoms and notes to system who processes it for prediction of heart disease.

Activity of initialization from user, doctor and system, are shown where user registers himself and login to upload the health records to cloud. The doctor checks the symptoms and notes to system who processes it for prediction of heart disease. Sequential process between user, doctor and system, are shown where user registers himself and login to upload the health records to cloud. The doctor checks the symptoms and notes to system who processes it for prediction of heart disease. Dataset is used to provide training and testing. Data is given as patients health record which is classified depending on dataset values to give result of heart disease.

©IJRASET: All Rights are Reserved



V. ALGORITHM

SMOTE (synthetic minority oversampling technique) is one of the most commonly used over- sampling methods to solve the imbalance problem. It aims to balance class distribution by randomly increasing minority class examples by replicating them. SMOTE synthesises new minority instances between existing minority instances.The SMOTE approach is an oversampling technique that has been frequently utilized in medicine to cope with imbalanced class datasets . SMOTE generates random synthetic data of the mi- nority class from its nearest neighbours using Euclidean distance to increase data instances. Because new samples are created based on original characteristics, they resemble the origi- nal data. SMOTE is widely regarded as one of the most reliable and effective preprocessing methods in machine learning and data analyses.

A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.

VI. CONCLUSION

The proper filters were utilized to improve the validity and rationality of the dataset. The proposed model that consisted of both SMOTE and ML method will ensure the enhancement of prediction accuracy. While the limitation is that it consumes more time during the part of pre-processing. The experimental results show that tree-based models were more effective in achieving higher quality results, and the Hyperparameter optimization method has more impact on improving the model's accuracy. Thus, SMOTE and ET optimized by hyperparameter achieved the highest accuracy for binary and multi-class problems on dataset. The accuracy is compared with other existing methods. The major goal of this study is to improve on previous work by developing a new and unique model-creation method and to make the model relevant and easy to use in a real-world situation.

REFERENCES

[1] A. Abdellatif, H. Abdellatef, J. Kanesan, C. -O. Chow, J. H. Chuah and H. M. Gheni, "An Effective Heart Disease Detection and Severity Level Classification Model Using Machine Learning and Hyperparameter Optimization Methods," in IEEE Access, vol. 10, pp. 79974-79985, 2022, doi: 10.1109/ACCESS.2022.3191669.

[2] K. S. K. Reddy and K. V. Kanimozhi, "Novel Intelligent Model for Heart Disease Pre- diction using Dynamic KNN (DKNN) with improved accuracy over SVM," 2022 Interna- tional Conference on Business Analytics for Technology and Security (ICBATS), 2022, pp. 1-5, doi: 10.1109/ICBATS54253.2022.9758996.

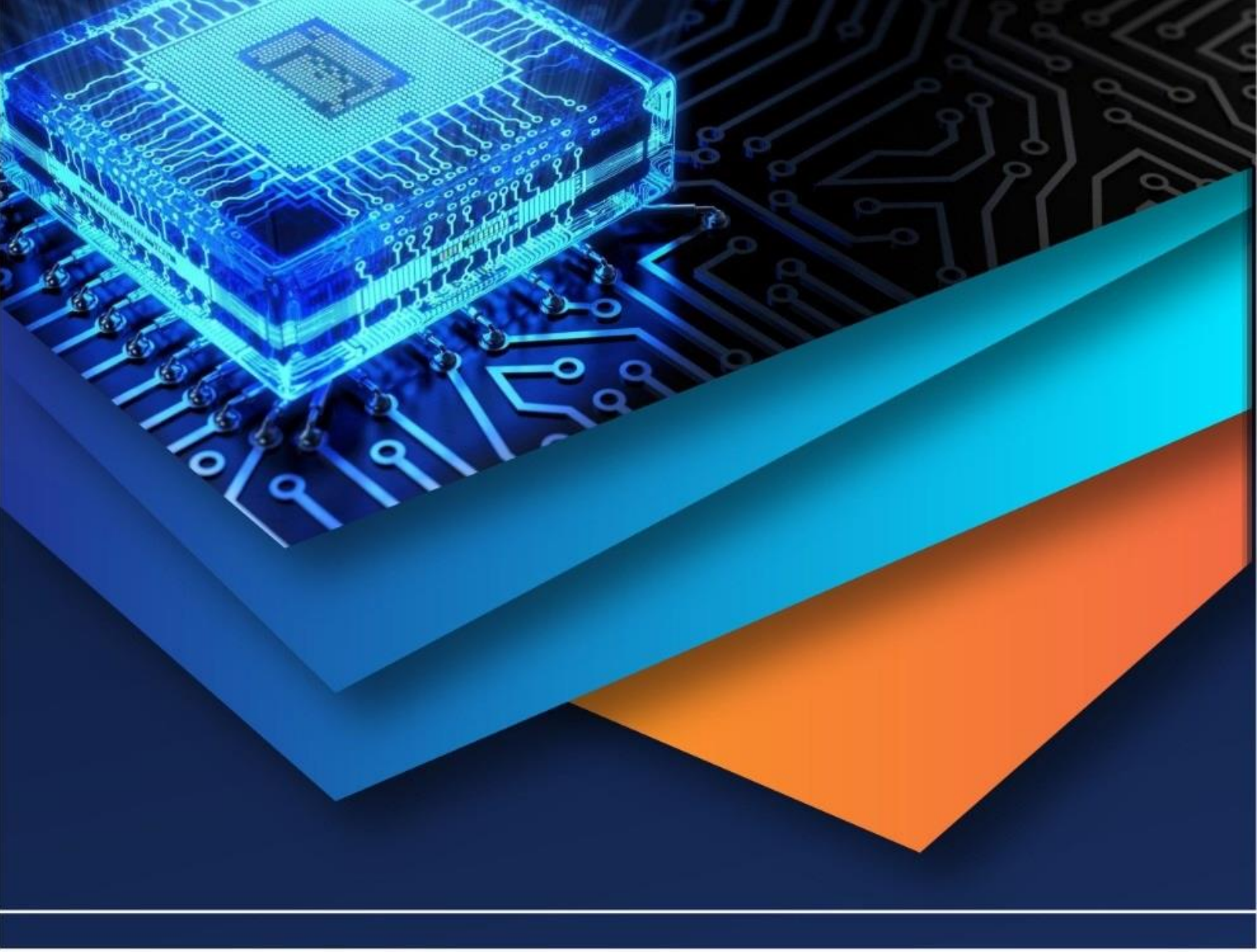
[3] T. Xue and Z. Jieru, "Application of Support Vector Machine Based on Particle Swarm Optimization in Classification and Prediction of Heart Disease," 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), 2022, pp. 857-860, doi: 10.1109/ICSP54964.2022.9778616.

[4] G. S. Reddy Thummala and R. Baskar, "Prediction of Heart Disease using Decision Tree in Comparison with KNN to Improve Accuracy," 2022 International Conference on In- novative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), 2022, pp. 1-5, doi: 10.1109/ICSES55317.2022.9914044.

[5] https://towardsdatascience.com/predicting-presence-of-heart-diseases-using-machine-learning-36f00f3edb2c

[6] https://iopscience.iop.org/article/10.1088/1757-899X/1022/1/012072/meta Heart Disease Prediction using Machine Learning

[7] https://www.youtube.com/watch?v=EUzexQGkH8k



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)