JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Heart Disease Prediction System Using Machine Learning Techniques.

[1]Assoc. Prof. Sukhwinder K. Bhatia,
[2]Anand Vishwas Jadhav, [3]Adesh Pradip Jadhav, [4]Adarsh Hari Lohar

Associate Professor, Dept. of Electronics & Telecommounication Engineering, JSPM's Imeperial College Of Engineering & Research Wagholi, Maharashtra, India
Student, Dept. of Electronics & Telecommounication Engineering, JSPM's Imeperial College Of Engineering & Research Wagholi, Maharashtra, India

Abstract : With the rapid advancement of technology and data, the healthcare sector is currently one of the most important study areas. Managing the vast amount of patient data is challenging. Big Data Analytics makes handling this data simpler. There are numerous methods used all around the world to cure various ailments. A new method that aids in disease detection and prediction is machine learning. In this study, machine learning is used to predict disease based on symptoms. On the presented dataset, machine learning algorithms like Naive Bayes, Decision Tree, and Random Forest are used to forecast the disease. Using the use of the Python programming language, it is implemented.

I. INTRODUCTION

Currently, when someone has a specific illness, they must make an expensive and time-consuming doctor visit. Also, it could be challenging for the user if they are far from a hospital or doctor because the illness cannot be identified. So, it might be simpler for the patient and the process if the aforementioned procedure could be carried out utilising an automated programme that can save time and money. There are other systems that use data mining techniques to examine the patient's risk status and predict heart-related diseases. Web-based software called heart Disease Predictor makes heart disease predictions for users based on their reported symptoms. Data sets for the heart disease prediction system were gathered from several websites that deal with health. The user will be able to determine the likelihood of the condition based on the listed symptoms with the use of the heart disease predictor. People are always interested in learning new things as internet usage increases daily. When an issue happens, people always try to turn to the internet for assistance. Compared to hospitals and doctors, more people have access to the internet. When an issue happens, people always try to turn to the internet for assistance. Compared to hospitals and doctors, people have access to the internet. When someone has a certain ailment, they do not immediately have a choice.

Motivation:

The methods used to forecast heart disease currently only use a tiny dataset. Our system's goal is to operate on a bigger dataset in order to boost system efficiency as a whole. Our technology is simple to use and produces results rapidly. The heart disease prediction feature relies on natural language processing, which enables users to enter their health-related problems. Three classification algorithms, namely Nave Bayes, Decision Tree, and Random Forest, are employed at various levels of evaluations in a comparative study and analysis to support this work. Although these machine learning methods are widely utilised, predicting cardiac disease is a crucial task requiring the highest level of accuracy. Thus, a variety of levels and assessment strategy types are used to evaluate the three algorithms.

II. LITERATURE SURVEY

1.Design And Implementing Heart Disease Prediction Using Naives Bayesian-

Data mining is an excellent approach that focuses on exploring and extracting important information from large collections of data. This information may then be useful in studying and identifying trends to help with business-related decision-making. In the medical area, the application of data mining can lead to the discovery and extraction of interesting patterns and information that can be useful in making clinical diagnoses. The research focuses on diagnosing cardiac illness by taking into account prior data and knowledge. In order to accomplish this, Navies Bayesian modelling is used to construct SHDP (Smart Heart Disease Prediction), which predicts risk factors for heart disease.

2.Application of Machine Learning in Disease Prediction-

Machine learning is increasingly being used in the diagnosis of medical conditions. This can be attributed mostly to advancements in illness categorization and recognition systems, which are able to provide information that helps medical professionals identify fatal diseases early on and so considerably improve patient survival rates. Using three distinct illness databases (Heart, Breast cancer, and Diabetes) available in the UCI repository for disease prediction, we employ various classification methods in this study, each with its own advantages. Each dataset's features were chosen using backward modelling and the p-value test. The study's findings support the notion that machine learning might be used to detect diseases early.

3.Disease phenotype similarity improves the prediction of novel disease- associated microRNAs

Many computational methods have been developed to anticipate such relationships by ranking prospective microRNAs according to their relevance to an illness, which has been demonstrated in numerous studies to play roles in miRNAs (microRNAs) on human disease. Network-based approaches are taking the lead among them since they effectively take advantage of the "disease module" notion in miRNA functional similarity networks. The most recent of which is RWRMDA, a Random Walk with Restart (RWR) algorithm-based technique on a miRNA functional similarity network. Because the "dis- ease module" idea also applies to protein interaction networks, the use of this method was motivated by its efficacy in predicting illness genes.

4.Efficient Heart Disease Prediction System using Decision Tree-

In the modern way of life, cardiovascular disease (CVD) is a major cause of morbidity and mortality. Cardiovascular disease identification is a crucial but difficult task that requires careful attention to detail, efficiency, and the right automation. Every person is not equally skilled, including doctors. As not all doctors are equally experienced in every subspecialty, there are numerous locations where we struggle to find skilled and specialised medical professionals. An automated approach for diagnosing illnesses would improve medical care and lower costs. In this study, we developed a system that can quickly identify the guidelines for estimating patients' risk levels based on the health-related information provided.

III. SYSTEM ANALYSIS

System Architecture:

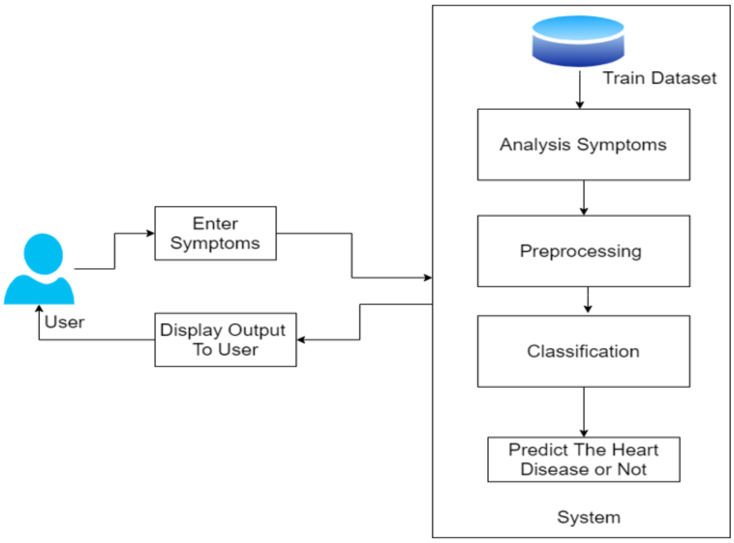

Fig 1. System Architecture

IV. METHODOLOGY

The working of the system begin with the collection of dataset and selecting the important attributes. Then the required data is preprocessed into the required format. The data is then divided into two parts training and testing data. The algorithms are applied and the model is trained using the training data. The accuracy of the system is obtained by testing the system using the testing data. This system is implemented using the following modules.

1.) Collection of Dataset
2.) Selection of attributes
3.) Data Pre-Processing
4.) Balancing of Data
5.) Disease Prediction

SVM:

SVM modeling is a promising classification approach for predicting medication adherence in HF patients. This predictive model helps stratify the patients so that evidence-based decisions can be made and patients managed appropriately. Further, this approach should be further explored in other complex diseases using other common variables. Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

Desicion Tree:
A decision tree is a graph that uses a branching method to illustrate every possible outcome of a decision. Decision trees can be drawn by hand or created with a graphics program or specialized software. Informally, decision trees are useful for focusing discussion when a group must make a decision. Programmatically, they can be used to assign monetary/time or other values to possible outcomes so that decisions can be automated

Random Forest:
Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

Naïve bayes:

- Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.
- It is mainly used in text classification that includes a high-dimensional training dataset.
- Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
- It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.
- Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.

ANN:
Artificial Neural networks (ANN) or neural networks are computational algorithms. It intended to simulate the behavior of biological systems composed of "neurons". ANNs are computational models inspired by an animal's central nervous systems. It is capable of machine learning as well as pattern recognition.

Anaconda:
For scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), Anaconda is a free and open-source version of the Python and R programming languages that attempts to simplify package management and deployment. Data-science packages are included in the distribution that are compatible with Windows, Linux, and macOS. It is created and maintained by Anaconda, Inc., a company that was established in 2012 by Peter Wang and Travis Oliphant. It is referred to as Anaconda Distribution or Anaconda Individual Edition as an Anaconda, Inc. product, but the firm also produces Anaconda Team Edition and Anaconda Enterprise Edition, all of which are paid options. The package management system conda is in charge of managing package versions in Anaconda.

This package manager turned out to be valuable on its own and for uses other than Python, so it was split off as a distinct open-source package. Miniconda is a boot-strapped, scaled-down variation of Anaconda that only contains conda, Python, the programmes they depend on, and a select few more packages.

More than 250 packages are pre-installed in the anaconda distribution, while more than 7,500 additional open-source packages, the conda package manager, and the virtual environment manager may all be downloaded from PyPI. Moreover, Anaconda Navigator, a graphical user interface, is provided as an alternative to the command line interface (CLI).

Spyder:
Spyder is a Python-based integrated development environment (IDE) that is open-source and cross-platform. The scientific Python stack's most well-known packages, such as NumPy, SciPy, Matplotlib, pandas, IPython, SymPy, and Cython, as well as other open-source programmes, are all integrated with Spyder. It is distributed with the MIT licence.

Since 2012, a group of scientific Python developers and the community have been maintaining and constantly improving Spy- der, which was first designed and built by Pierre Raybaut in 2009.

In addition to supporting interactive tools for data inspection and embedding Python-specific code quality assurance and introspection tools like Pyflakes, Pylint, and Rope, Spyder is expandable through first- and third-party plugins. It is cross-platform through Anaconda and is accessible on Windows, macOS, and several popular Linux distributions, including Arch Linux, Debian, Fedora, Gentoo, openSUSE, and Ubuntu.

The Python bindings for PyQt or PySide can be used with Spyder, which uses Qt for its graphical user interface. The ability to utilise any back-end is provided via QtPy, a thin abstraction layer created by the Spyder project and later adopted by numerous other programmes.

V. RESULT AND IMPLEMENTATION

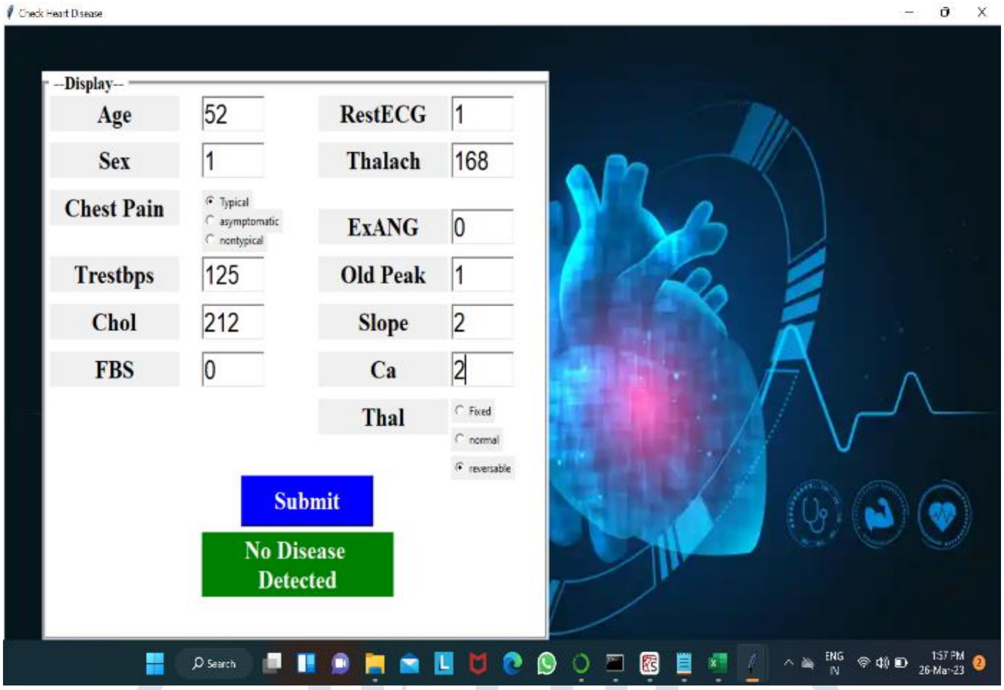

Fig 2 No Disease Detected

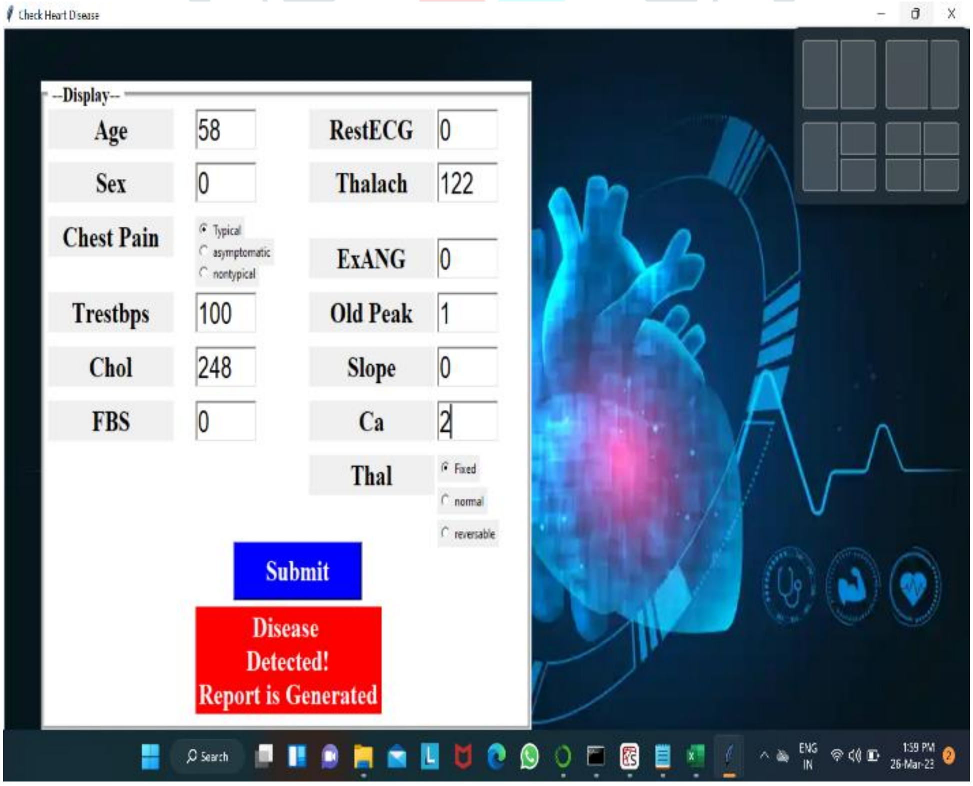

Fig 3. Disease Detected

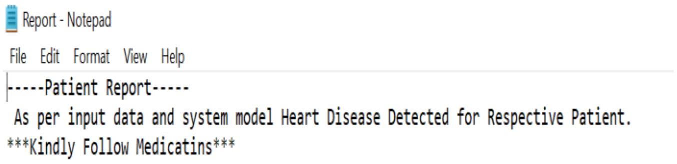

Fig 3. Disease Detected Report

VI. CONCLUSION

This project aims to predict the heart disease on the basis of the symptoms. The project is designed in such a way that the system takes symptoms from the user as input and produces output i.e. predict heart disease. Early detection of cardiac disease can help high-risk patients make decisions about lifestyle modifications that will lessen problems, which can be a significant advancement in the field of medicine. Each year, more people are diagnosed with cardiac illnesses. This calls for an early diagnosis and course of action. The medical community as well as patients may benefit greatly from this use of appropriate technology support. SVM, Decision Tree, Random Forest, Naive Bayes, Logistic Regression, Adaptive Boosting, and Extreme Gradient Boosting are just a few of the seven machine learning algorithms employed in this study to evaluate performance. The characteristics that are anticipated to cause heart disease in people are Average prediction accuracy probability of 55obtained. heart Disease Predictor was successfully implemented using grails framework. It is essential to create a system that can forecast heart diseases precisely and effectively given the rise in fatalities caused by heart diseases. The goal of the study was to identify the most effective ML algorithm for heart disease identification. The UCI machine learning repository dataset is used in this work to examine the accuracy scores of the Decision Tree, Logistic Regression, Random Forest, and Naive Bayes algorithms for predicting heart disease.

VII. REFERENCES

1. D. Tian, J. Zhou, Y. Wang, Y. Lu, H. Xia, and Z. Yi, “A dynamic and self- adaptive network selection method for multimode communications in hetero- geneous vehicular telematics,” IEEE Transactions on Intelligent Transporta- tion Systems, vol. 16, no. 6, pp. 3033–3049, 2015.

2. M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, C. Youn, “Wearable 2.0: En- able Human-Cloud Integration in Next Generation Healthcare System,” IEEE Communications, Vol. 55, No. 1, pp. 54–61, Jan. 2017

3. M. Chen, Y. Ma, J. Song, C. Lai, B. Hu, ”Smart Clothing: Connecting Human with Clouds and Big Data for Sustainable Health Monitoring,” ACM/Springer Mobile Networks and Applications, Vol. 21, No. 5, pp. 825C845, 2016

4. J. Wang, M. Qiu, and B. Guo, “Enabling real-time information service on telehealth system over cloud-based big data platform,” Journal of Systems Ar- chitecture, vol. 72, pp. 69–79, 2017

5. Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, “Healthcps: Healthcare cyber-physical system assisted by cloud and big data,” IEEE Sys- tems Journal, 2015.

6. K. Lin, J. Luo, L. Hu, M. S. Hossain, and A. Ghoneim, “Localization based on social big data analysis in the vehicular networks,” IEEE Transactions on Industrial Informatics, 2016.
7. K. Lin, M. Chen, J. Deng, M. M. Hassan, and G. Fortino, “Enhanced finger- printing and trajectory prediction for iot localization in smart buildings,” IEEE Transactions on Automation Science and Engineering, vol. 13, no. 3, pp. 1294–1307, 2016.

8. S. Bandyopadhyay, J. Wolfson, D. M. Vock, G. Vazquez-Benitez, G. Ado- mavicius, M. Elidrisi, P. E. Johnson, and P. J. O’Connor, “Data mining for censored time-to-event data: a bayesian network model for predicting cardio- vascular risk from electronic health record data,” Data Mining and Knowledge Discovery, vol. 29, no. 4, pp. 1033–1069, 2015.

9. B. Qian, X. Wang, N. Cao, H. Li, and Y.-G. Jiang, "A relative similarity based method for interactive patient risk prediction," Data Mining and Knowledge Discovery, vol. 29, no. 4, pp. 1070–1093, 2015.

10. J. Wan, S. Tang, D. Li, S. Wang, C. Liu, H. Abbas and A. Vasilakos, "A Manu- facturing Big Data Solution for Active Preventive Maintenance", IEEE Trans- actions on Industrial Informatics, DOI: 10.1109/TII.2017.2670505, 2017.

11. N. Nori, H. Kashima, K. Yamashita, H. Ikai, and Y. Imanaka, "Simultane- ous modeling of multiple diseases for mortality prediction in acute hospital care," in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015, pp. 855–864.

12. S. Basu Roy, A. Teredesai, K. Zolfaghar, R. Liu, D. Hazel, S. Newman, and A. Marinez, "Dynamic hierarchical classification for patient risk-ofreadmission," in Proceedings of the 21th ACM SIGKDD international conference on knowl- edge discovery and data mining. ACM, 2015, pp. 1691–1700.

13. Y.-D. Zhang, X.-Q. Chen, T.-M. Zhan, Z.-Q. Jiao, Y. Sun, Z.-M. Chen, Y. Yao, L.-T. Fang, Y.-D. Lv, and S.-H. Wang, "Fractal dimension estima- tion for developing pathological brain detection system based on minkowski- bouligand method," IEEE Access, vol. 4, pp. 5937–5947, 2016.

14. S.-M. Chu, W.-T. Shih, Y.-H. Yang, P.-C. Chen, and Y.-H. Chu, "Use of tradi- tional chinese medicine in patients with hyperlipidemia: A population-based study in taiwan," Journal of ethnopharmacology, vol. 168, pp. 129–135, 2015.

15. S.-H. Wang, T.-M. Zhan, Y. Chen, Y. Zhang, M. Yang, H.-M. Lu, H.- N. Wang, B. Liu, and P. Phillips, "Multiple sclerosis detection based on biorthogonal wavelet transform, rbf kernel principal component analysis, and logistic re- gression," IEEE Access, vol. 4, pp. 7567–7576, 2016

16. Y.-D. Zhang, X.-Q. Chen, T.-M. Zhan, Z.-Q. Jiao, Y. Sun, Z.-M. Chen, Y. Yao, L.-T. Fang, Y.-D. Lv, and S.-H. Wang, "Fractal dimension estimation for developing pathological brain detection system based on minkowski-bouligand method," IEEE Access, vol. 4, pp. 5937–5947, 2016.

17. S.-M. Chu, W.-T. Shih, Y.-H. Yang, P.-C. Chen, and Y.-H. Chu, "Use of traditional chinese medicine in patients with hyperlipidemia: A population-based study in taiwan," Journal of ethnopharmacology, vol. 168, pp. 129–135, 2015.

18. S.-H. Wang, T.-M. Zhan, Y. Chen, Y. Zhang, M. Yang, H.-M. Lu, H.- N. Wang, B. Liu, and P. Phillips, "Multiple sclerosis detection based on biorthogonal wavelet transform, rbf kernel principal component analysis, and logistic regression," IEEE Access, vol. 4, pp. 7567–7576, 2016.

19. Y.-D. Zhang, Z.-J. Yang, H.-M. Lu, X.-X. Zhou, P. Phillips, Q.-M. Liu, and S.-H. Wang, "Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation," IEEE Access, vol. 4, pp. 8375–8385, 2016.