**Heart Disease Prediction using ML**

**ABSTRACT:**

With big data growth in biomedical and healthcare communities, accurate analysis of medical data benefits early Heart disease detection, patient care, and community services. However, the analysis accuracy is reduced when the quality of medical data is incomplete. Moreover, different regions exhibit unique characteristics of certain regional diseases, which may weaken the prediction of disease outbreaks. In this paper, we streamline machine learning algorithms for effective prediction of Heart disease outbreak in disease-frequent communities. We experiment the modified prediction models over real-life hospital data collected from different parts of county. To overcome the difficulty of incomplete data, we use a latent factor model to reconstruct the missing data. We experiment on a Heart disease based on the symptoms given by the user.It predicts using machine learning algorithms. So,the output is accurate .It uses flask web frame work for GUI. In this we will analyze data using ML algorithms

**INTRODUCTION:**

Cardiovascular disease (CVD) is increasing daily in this modern world. According to the World Health Organization (WHO), an estimated 17 million people die each year from cardiovascular disease, particularly heart attacks and strokes [1]. It is, therefore, necessary to record the most important symptoms and health habits that contribute to CVD. Various tests are performed prior to diagnosis of CVD, including auscultation, ECG, blood pressure, cholesterol and blood sugar. These tests are often long and long when a patient's condition may be critical and he or she must start taking medication immediately, so it becomes important to prioritize the tests [2]. Several health habits contribute to CVD. Therefore, it is also necessary to know which health habits contribute to CVD. Machine learning is now an emerging field due to the increasing amount of data. Machine learning makes it possible to acquire knowledge from a massive amount of data, which is very heavy for man and sometimes impossible [3]. The objective of this paper is to prioritize the diagnostic test and to see some of the health habits that contribute to CVD. Moreover, and above all, the different machine learning algorithms are compared using intelligent optimization algorithms. In this article, manually classified data is used. Manual classification is healthy or unhealthy. Based on a machine learning technique called classification, 70% of the data is supervised or trained and 30% is tested as part of this article. Intelligent optimization algorithms are developed by simulating or revealing certain natural phenomena and are widely used in many research fields because of their versatility [4, 5]. The Particle Swarm Optimization (PSO) algorithm has been successfully applied to heart disease because of its simplicity and generality [6]. However, PSO easily fell into the optimal local solution. In addition, the ACO algorithm was originally introduced for combinatorial optimization. Recently, ACO algorithms have been developed to solve continuous optimization problems. These problems are characterized by the fact that decision variables have continuous domains, unlike discrete problems [7]. Using a single optimization algorithm has the disadvantages of low accuracy and generalizability in solving complex problems. To further explore the application of intelligent optimization in bioinformatics, PSO and ACO are combined in this article, meaning that exploitation and exploration capacity are combined for binary and multi-class heart disease. In this article, the Fast CorrelationBased Feature selection (FCBF) method [8] used to remove redundant and irrelevant features, the results of the PSO optimization are considered the initial values of the ACO, and then the classification model for heart disease is constructed after the parameters are adjusted. In this study, algorithms such as KNearest Neighbour (K-NN), Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF) and Artificial Neural Network (ANN | MLP) are used. It can be concluded that K-Nearest Neighbour and the Random Forest are the best algorithms for the prediction and classification of heart disease dataset.

**Existing system:**

The LIBSVM and the WEKA data mining tool are used to analyze the results of this method. Five data sets (Iris, diabetes disease, breast cancer disease, heart disease and hepatitis) are collected from the Irvine UC machine learning repository for this experiment.

Otoom et al. presented a system for analysis and follow-up. Coronary artery disease is detected and monitored by the proposed system. Cleveland Heart data are taken from the UCI. This dataset consists of 303 cases and 76 attributes/features. 13 features are used out of 76 features. Two tests with three algorithms: Bayes Naive, Support vector machine, and Functional Trees FT are performed for detection purposes. The WEKA tool is used for detection. After testing the Holdout test, the 88.3% accuracy is achieved using the SVM technique.

**Disadvantages:**

The common objective of all these techniques is to classify hearth disease using hybrid classification techniques. However, they used only one classification and optimization technique.

**Proposed System:**

We Propose machine learning algorithms for effective prediction of Heart disease outbreak in disease-frequent communities. We take input as collection of heart disease dataset from kaggle website and then train and test using random forest algorithm and then predict based on user input of respective patient.

**Advantages:**

* The proposed approach presented a systematic way to achieve the desired results by taking into account different technical optimizations with different machine learning algorithms.
* The proposed methods are compared to supervised algorithms based on existing approximate sets and classification accuracy measurements are used to evaluate the performance of the proposed approaches

**LITERATURE SURVEY**

In the above study we will see different data mining techniques that were used to classify the heart diseases. In year 2000, research conducted by ShusakuTsumoto [5] says that as we human beings are unable to arrange data if it is huge in size we should use the data mining techniques that are available for finding different patterns from the available huge database and can be used again for clinical research and perform various operations on it.

and perform various operations on it. Y. Alp Aslandogan, et. al. (2004), worked on three different classifiers called K-nearest Neighbour (KNN), Decision Tree, Naïve Bayesian and used Dempsters’ rule for this three viewpoint to appear as one concluding decision. This classification based on the combined idea show increased accuracy [6].

Franck Le Duff (2004), worked on creating Decision tree quickly with clinical data of the physician or service. He suggested few data mining techniques which can help cardiologists in the predication survival of patients. The main drawback of the system was that the user needs to have knowledge of the techniques and we should collect sufficient data for creating an suitable model [8].

Kiyong Noh, et. al. (2006) made use of a classification technique for removal of multi-parametric structures by accessing HRV and ECG signals. Kiyong used the FPgrowth algorithm as the foundation of this technique that is associative. A rule consistency degree was gained which allows a robust press on trimming designs in the method of producing designs[10]

HeonGyu Lee, et. al. (2007), operated for the operation systems of Arithmetical and cataloguing for the addition chief of the multi-parametric feature through direct and nonlinear features of Heart Rate Variability (HRV). The dissimilar classifiers existing are cataloguing grounded on Decision Tree (C4.5), Multiple Association Rules (CMAR) and Bayesian classifiers, and Support Vector Machine (SVM) that are investigated for the valuation of the linear and nonlinear features of the HRV tables [11].

**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

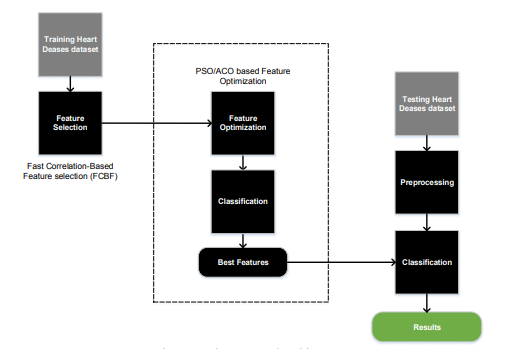
* System : Pentium Dual Core.
* Hard Disk : 120 GB.
* Monitor : 15’’ LED
* Input Devices : Keyboard, Mouse
* Ram : 1GB.

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows 7.
* Coding Language : python
* Tool : anaconda, flask framework
* Database : MYSQL

**SYSTEM DESIGN**

**ARCHITECTURE :**

****

**Conclusion:**

The amount of Heart diseases can exceed the control line and reach to maximum point. Heart disease are complicated and each and every year lots of people are dying with this disease By using this all systems one of the major drawbacks of these works is mainly focus only to the application of classify techniques and algorithms for heart disease prediction, by all these studying various data cleaning and mining techniques that prepare and build a dataset appropriate for data mining. So that I can use this Machine Learning in that logistic regression algorithms by predicting if patient has heart disease or not. Any nonmedical employee can use this software and predict the heart disease and reduce the time complexity of the doctors.

**FUTURE WORK**

Today’s, world most of the data is computerized, the data is distributed and it is not utilizing properly. By Analyzing the available data we can also use for unknown patterns. The primary motive of this research is the prediction of heart diseases with high rate of accuracy. For predicting the heart disease we can use logistic regression algorithm, naviebayes, sklearn in machine learning. The future scope of the paper is the prediction of heart diseases by using advanced techniques and algorithms in less time complexity

**REFENRENCES:**

* K. Gai, M. Qiu, Z. Xiong, and M. Liu, “Privacy-preserving multi-channel communication in Edge-of-Things,” Future Generation Computer Systems, 85, 2018, pp. 190-200.
* K. Gai, M. Qiu, and X. Sun, “A survey on FinTech,” *Journal of Network and Computer Applications*, 2017, pp. 1-12.
* Abbas, K. Bilal, L. Zhang, and S. U. Khan, “A cloud based health insurance plan recommendation system: A user cen-tered approach, “*Future Generation Computer Systems*, vols. 43-44, pp. 99-109, 2015.
* N. Khan, ML M. Kiah, S. A. Madani, M. Ali, and S. Sham-shirband, “Incremental proxy re-encryption scheme for mo-bile cloud computing environment,”*The Journal of Supercom-puting*,Vol. 68, No. 2, 2014, pp. 624-651.
* R. Wu, G.-J. Ahn, and H. Hu, “Secure sharing of electronic health records in clouds,” In *8th IEEE International Conference on Collaborative Computing: Networking, Applications and Work-*
* A. Abbas andS. U. Khan, “A Review on the State-of-the-Art Privacy Preserving Approaches in E-Health Clouds,” *IEEE Journal of Biomedical and Health Informatics,*vol. 18, no. 4, pp. 1431-1441, 2014.
* [7] M. H. Au, T. H. Yuen, J. K. Liu, W. Susilo, X. Huang, Y. Xiang, and Z. L. Jiang, “A general framework for secure sharing of personal health records in cloud system,” Journal of Comput-er and System Sciences, vol. 90, pp, 46-62, 2017.
* [8] J. Li, “Electronic personal health records and the question of privacy,” *Computers*, 2013, DOI: 10.1109/MC.2013.225.
* [9] D. C. Kaelber, A. K. Jha, D. Johnston, B. Middleton, and D. W. Bates, “A research agenda for personal health records (PHRs),” *Journal of the American Medical Informatics Association*, vol.15, no. 6, 2008, pp. 729-736.
* [10] S. Yu, C. Wang, K. Ren, and W. Lou, “Achieving secure, scala-ble and fine-grained data access control in cloud computing,” in *Proceedings of the IEEE INFOCOM*, March 2010, pp. 1-9.
* [11] S. Kamara and K. Lauter, “Cryptographic cloud storage,” *Financial Cryptography and Data Security*, vol. 6054, pp. 136–149, 2010.
* T. S. Chen, C. H. Liu, T. L. Chen, C. S. Chen, J. G. Bau, and T.C. Lin, “Secure Dynamic access control scheme of PHR in cloud computing,” *Journal of Medical Systems*, vol. 36, no. 6, pp. 4005–4020, 2012.
* [13] K. Gai, M. Qiu, “Blend arithmetic operations on tensor-based fully homomorphic encryption over real numbers,” *IEEE Transactions on Industrial Informatics*, 2017, DOI: 10.1109/TII.2017.2780885..
* [14] M. Li, S. Yu, Y. Zheng, K. Ren, andW. Lou, “Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption,” *IEEE Transactions on Parallel and Distributed Systems*, 2013, vol. 24, no. 1, pp. 131–143.
* [15] “Health Insurance Portability and Accountability,” http://www.hhs.gov/ocr/privacy/hipaa/administrative/privacyrule/, accessed on October 20, 2014.
* [16] Z. Xiao and Y. Xiao, “Security and privacy in cloud compu-ting,” *IEEE Communications Surveys and Tutorials*, vol. 15, no. 2, pp. 1–17, Jul. 2012.