HEART DISEAESE RATE PREDICTION SYSTEM IN E-HEALTH CARE USING MACHINE LEARNING TECHNIQUE

Mrs.E.Padma[1], Tugutla Vishu Vardhan Reddy[2], Yarasani Sai Rithish[3]

[1]Assistant Professor, Dept. of CSE, SCSVMV (Deemed to be University), Kanchipuram, Tamilnadu, India
[2]Student, Dept. of CSE, SCSVMV(Deemed to be University), Kanchipuram ,Tamilnadu, India
[3]Student, Dept. of CSE, SCSVMV(Deemed to be University),Kanchipuram,Tamilnadu, India

---***---

Abstract- *In today's era deaths due to heart disease has become a major issue approximately one person dies per minute due to heart disease. This is considering both male and female category and this ratio may vary according to the region also this ratio is considered for the people of age group. This does not indicate that the people with other age group will not be affected by heart diseases. This problem may start in early age group also and predict the cause and disease is a major challenge nowadays. Here in this paper, we have discussed various algorithms and tools used for prediction of heart diseases.*

Key words: Heart disease classification, disease diagnosis, intelligent system, medical data analytics.

1. INTRODUCTION

The contents of this paper mainly focus on various data mining practices that are valuable in heart disease forecast with the assistance of dissimilar data mining tools that are accessible. If the heart doesn't function properly, this will distress the other parts of the human body such as brain, kidney etc. Heart disease is a kind of disease which effects the functioning of the heart. In today's era heart disease is the primary reason for deaths. WHO-World Health Organization has anticipated that 12 million people die every year because of heart diseases. Some heart diseases are cardiovascular, heart attack, coronary and knock. Knock is a sort of heart disease that occurs due to strengthening, blocking or lessening of blood vessels which drive through the brain or it can also be initiated by high blood pressure. The major challenge that the Healthcare industry faces now-a-days is superiority of facility. Diagnosing the disease correctly & providing effective treatment to patients will define the quality of service. Poor diagnosis causes disastrous consequences that are not accepted. Records or data of medical history is very large, but these are from many dissimilar foundations. The interpretations that are done by physicians are essential components of these data. The data in real world might be noisy, incomplete and inconsistent, so data preprocessing will be required in directive to fill the omitted values in the database. Even if cardiovascular diseases is found as the important source of death in world in ancient years, these have been announced as the most avoidable and manageable diseases. The whole and accurate management of a disease rest on the well-timed judgment of that disease. A correct and methodical tool for recognizing high-risk patients and mining data for timely analysis of heart infection looks a serious want. Different person body can show different symptoms of heart disease which may vary accordingly. Though, they frequently include back pain, jaw pain, neck pain, stomach disorders, and tininess of breath, chest pain, arms and shoulders pains. There are a variety of different heart diseases which includes heart failure and stroke and coronary artery disease. Heart expert's create a good and huge record of patient's database and store them. It also delivers a great prospect for mining a valued knowledge from such sort of datasets. There is huge research going on to determine heart disease risk factors in different patients, different researchers are using various statistical approaches and numerous programs of data mining approaches. Statistical analysis have acknowledged the count of risk factors for heart diseases counting smoking, age, blood pressure, diabetes, total cholesterol, and hypertension, heart disease training in family, obesity and lack of exercise. For prevention and healthcare of patients who are about to have addicted of heart disease it is very important to have awareness of heart diseases.

2. LITERATURE REVIEW

Amin Ul haq et.al., (2020),the author here want to address the problem of features selection by employing preprocessing techniques and standard state of the art four features selection algorithms such as relief, MRMR, LASSO and LLBFS for appropriate subset of features and then applied these features for effective training and testing of the classifiers that identify which feature selection algorithm and classifier gives good results in term of accuracy and computation time.

Salah uddin et.al., (2020)**,** the authors proposed fast conditional mutual information (FCMIM) FS algorithm for feature selection and then these features are input to classifiers for improving prediction accuracy and reducing computation time. The classifiers performances have been compared on features selected by the standard state of the art FS algorithms with the selected features of the proposed FS algorithm and also identify weak features from the dataset which affect the performance of the classifiers. But the only disadvantage is it is difficult to handle this process.

David Springer, Chengyu Liu et.al., (2020),the author explains about the clinical and prognostic significance of non-linear measures of the heart rate variability, applied on the group of patients with coronary heart disease and age matched healthy control group. Three different methods were applied by the author named as Hurst exponent, De trended Fluctuation Analysis and approximate entropy. Hurst exponent of the R-R series was determined by the range rescaled analysis technique. DFA was used to quantify fractal long range correlation properties if heart rate variability.

3. SCOPE OF THE PROPOSED WORK

The analysis accuracy is reduced when the quality of medical data in incomplete. Moreover, different regions exhibit unique characteristics of certain regional diseases, which may weaken the prediction of disease outbreaks. However, those existing work mostly considered structured data. There is no proper methods to handle semi structured and unstructured. The proposed system will consider both structured and unstructured data. The analysis accuracy is increased by using Machine Learning algorithm.

4. PROPOSED SYSTEM

The proposed system is having a data which is classified if patients have heart disease or not according to the features included. The proposed system can try to use this data to create a model which tries to predict (reading data and data Exploration) if a patient has this disease or not. In the proposed system, logistic regression (classification) algorithm is used and implements Naive Bayes algorithm to getting accuracy result .Finally analysing the result with the help of Comparing Models and Confusion Matrix. From the data having, it should be classified into different structured data based on the features of the patient heart. From the availability of the data, we have to create a model which predicts the patient disease using logistic regression algorithm. Firstly, has to import the datasets. Read the datasets, the data should contain different parameters like age, gender, sex, CP (chest pain), slope, target. The data should be explored so that the information is verified. Create a temporary variable and also build a model for logistic regression. Here, sigmoid function is used which helps in the graphical representation of the classified data. By using logistic regression, naive Bayes the accuracy rate increases.

4.1 NAIVE BAYES ALGORITHM

It is a classification technique based on Bayes Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as 'Naive'. Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

4.2 K-NEAREST NEIGHBOR (KNN) CLASSIFICATION METHOD

K-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. The k-NN algorithm is among the simplest of all machine learning algorithms. The neighbors are taken from a set of objects for which the class (for k-NN classification) or the object property value (for k-NN regression) is known.
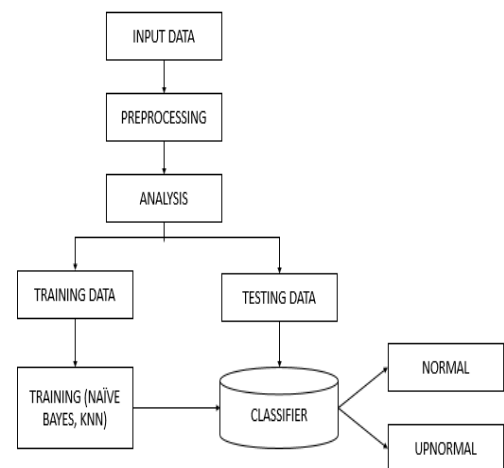
5. IMPLEMENTATION

System Architecture

Figure-1: General Architecture of the proposed system

6. MODULE DESCRIPTION

6.1 User Interface Module

a) Login

This is the first activity that opens the website. User needs to provide a correct contact number and a password, which user enters while registering, in order to login into the app. If information provided by the user matches with the data in the database table then user successfully login into the app else message of login failed is displayed and user need to re-enter correct information. A link to the register activity is also provided for registration of new users.

b) Registration

A new user who wants to access the website needs to register first before login. By clicking on register button in login activity, the register activity gets open. A new user registers by entering full name, password and contact number. A user needs to enter password again in confirm password textbox for confirmation. When user enters the information in all textboxes, on the click of register button, the data is transferred to database and user is directed to login activity again. Registered user then needs to login in order to access the app . Validations are applied on all the textboxes for proper functioning of the app. Like information in each textbox is must that is each textbox, either it is of name, contact, password or confirm password, will not be empty while registering. If any such textbox is empty app will give message of information is must in each textbox. Also data in password and confirm password fields must match for successful registration. Another validation is contact number must be valid one that is of 10 digits. If any such validation is violated then registration will be unsuccessful and then user needs to register again. Message that app will display when one of the field is empty. If all such information is correct user will be directed to login activity for login into the app.

6.2 ADMIN MODULE

In this module, admin can add and view new doctor details, disease details and drug details. And then admin can view feedback provided by various users.

## 6.3 DISEASE ANALYSIS MODULE

In this module, we can analyse the disease and give patient feedbacks if it is life threatening or not and also calculate the how much probability disease will be happening.

6.4 DISEASE PREDICTION MODULE

Patient will specify the symptoms caused due to his illness. System will ask certain question regarding his illness and system predict the disease based on the symptoms specified by the patient and system will also suggest doctors based on the disease.

7. RESULTS
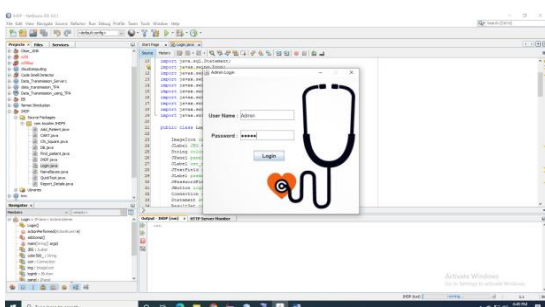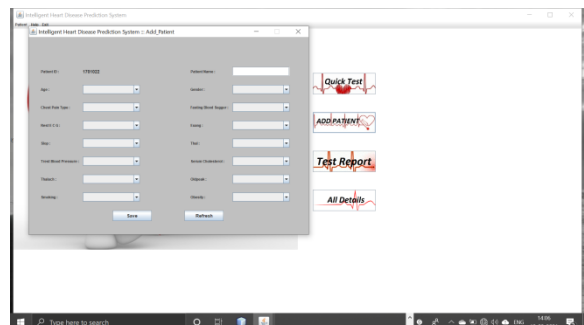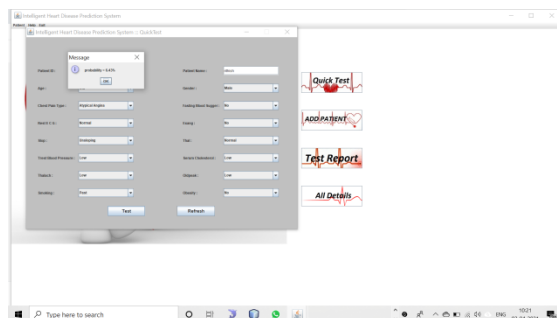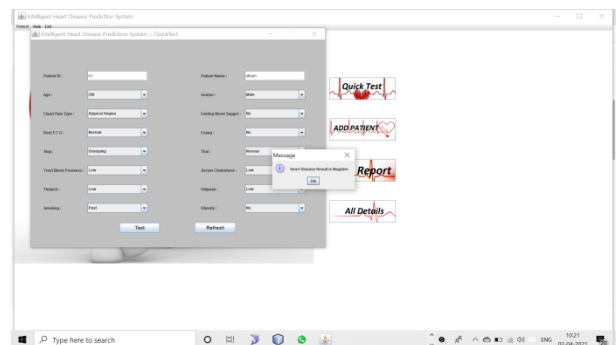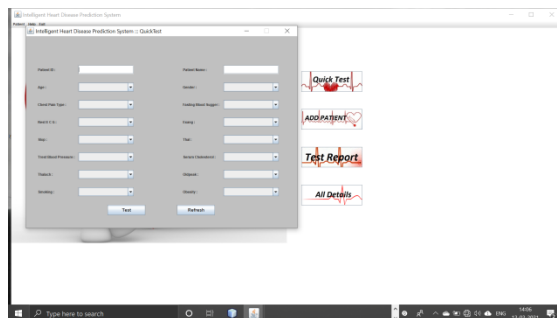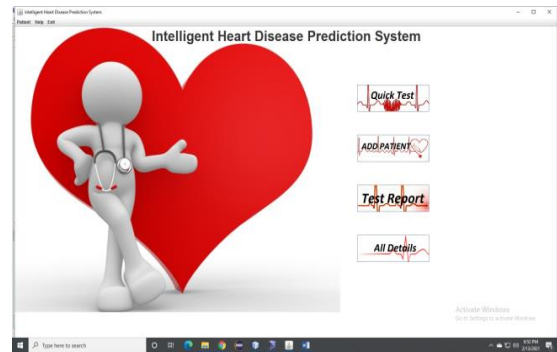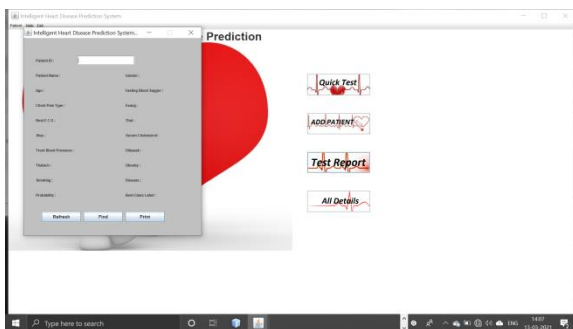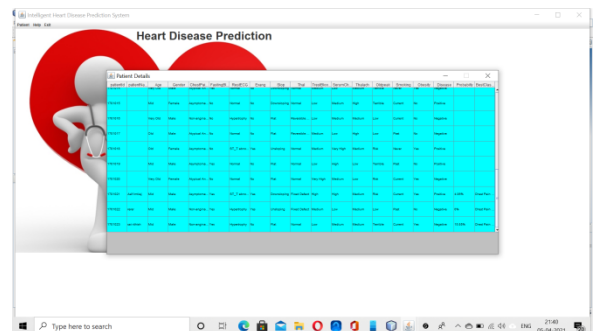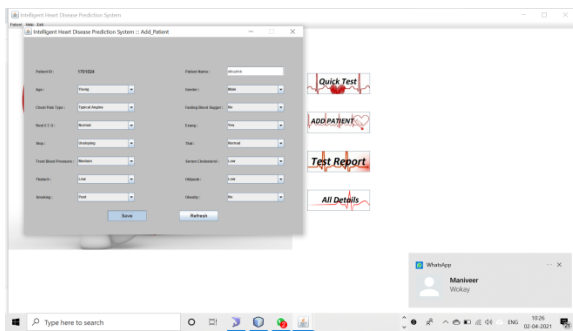
8. CONCLUSION

In this paper, we propose an approach for answering drug queries to support drug prescription. Our focus is on how to obtain and rank answers based on incomplete information and provide personalization. To cope with incomplete and noisy data, we allow both exact and close matches when answering queries. We also present an intuitive approach to display answers to users, which aims to help users to understand the ranked results and possibly refine their queries.
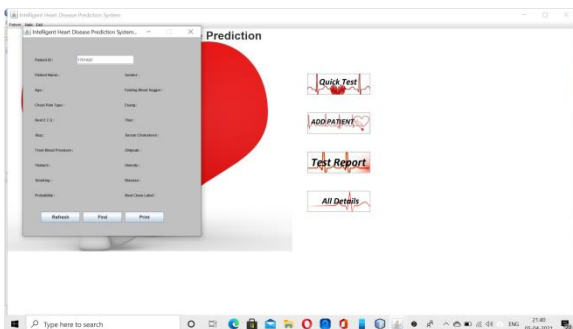
REFERENCES

[1] S. Khemmarat and L. Gao, "Supporting drug prescription via predictive and personalized query system," in PervasiveHealth. IEEE,2015.

[2] C. Knox et al., "Drugbank 3.0: a comprehensive resource for omics research on drugs," Nucleic acids research, vol. 39, no. suppl 1, pp. D1035–D1041, 2011.

[3]M. Kuhn et al., "A side effect resource to capture phenotypic effects of drugs," Molecular systems biology, vol. 6, no. 1, p. 343, 2010.

[4] M. Kanehisa and S. Goto, "Kegg: kyoto encyclopedia of genes andgenomes," Nucleic acids research, vol. 28, no. 1, pp. 27–30, 2000.

[5]T. Fawcett, "An introduction to roc analysis," Pattern recognition letters, vol. 27, no. 8, pp. 861–874, 2006.

[6] K. Sangkuhl et al., "Pharmgkb: understanding the effects of individual genetic variants," Drug Metab. Rev., vol. 40, no. 4, pp. 539–551, 2008.

[7] A. Langer et al., "A text based drug query system for mobile phones,"Int. J. Mob. Commun., vol. 12, no. 4, pp. 411–429, Jul. 2014.

[8]C. Doulaverakis et al., "Panacea, a semantic-enabled drug recommendations discovery framework," J. Biomed. Semant., vol. 5, p. 13, 2014.
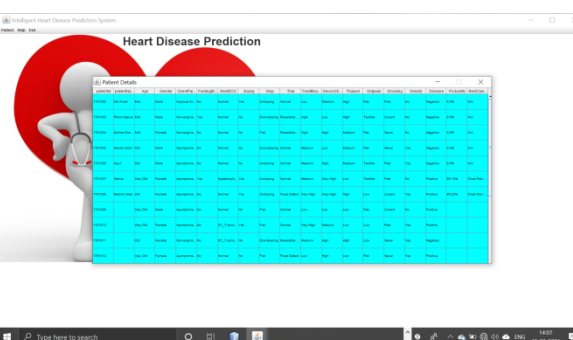
BIOGRAPHIES

Mrs. E. Padma, Assistant Professor in Computer science and Engineering department in SCSVMV (Deemed to be university).

Yarasani sai rithish is pursuing Bachelors of Engineering from SCSVMV (Deemed to be university).

Tugutla Vishnu vardhan reddy is pursuing Bachelors of Engineering from SCSVMV (Deemed to be university).