



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Happy Healthing - Daily Health Reading Visualization and Disease Prediction

Prof Kiran Bidua
MIT School of
Engineering, MIT ADT
University, Pune

Mr. Prathmesh Kattukar
MIT School of
Engineering, MIT ADT
University, Pune

Mr. Harshit Bhansali
MIT School of
Engineering, MIT ADT
University, Pune

Ms. Shreya Utekar
MIT School of
Engineering, MIT ADT
University, Pune

Mr. Prash Singhal
MIT School of
Engineering, MIT ADT
University, Pune

Abstract—The Happy healthing is a system that will help to predict disease using Machine Learning. Happy Healthing processes the symptoms provided by the user as input and gives the output as the prediction of the disease. A supervised machine learning algorithm called the Naive Bayes classifier is used in the prediction of the disease. The prediction is calculated by the Naive Bayes algorithm. With an increase in biomedical and healthcare data, accurate analysis of medical data benefits early disease detection and patient care.

Keywords—Natural Language Processing, Artificial Intelligence, Knowledge base, Naive Bayes, Disease prediction.

I. INTRODUCTION

Machine Learning is the domain that uses past data to predict the future. Machine Learning is the understanding of computer systems under which the Machine Learning model learns from data and experience. The machine learning algorithm has two phases: 1) training & 2) testing. Machine learning technology has struggled to predict the disease from a patient's symptoms and from the history of the patient. Healthcare issues can be solved efficiently by using machine learning technology. We are applying complete machine learning concepts to keep track of a patient's health.

ML models allow us to build models to get quickly cleaned and processed data and deliver results faster. By using this system, doctors will make good decisions related to patient diagnoses and, according to that, good treatment will be

given to the patient, which leads to an improvement in patient healthcare services. To introduce machine learning in the medical field, healthcare is the prime example. To improve the accuracy of large datasets, the existing work will be done on unstructured or textual data. The prediction of diseases will be based on linear, KNN, and decision tree algorithms.

II. RELATED WORK

1. Paper Name: Disease phenol type similarity improves the prediction of novel disease associated microRNAs

Author: Duc-Hau Lê Several studies have shown the role of miRNAs (microRNAs) in human disease, and a range of computational methods have been proposed by rating candidate microRNAs according to their relevance to a disease to predict such associations. Network-based approaches are becoming dominant among them as they make effective use of the "disease module" concept in miRNA networks of functional similarity. Of these, the algorithm-based Random Walk with Restart (RWR) method on a functional similarity network miRNA, namely RWRMDA, is state-of-the-art.

The use of this algorithm was motivated by its performance in predicting disease genes since the concept of "disease module" also exists in networks of protein interaction. In addition, for the prediction of disease genes, several other algorithms have also been developed. Nevertheless, they have not been used yet for disease microRNA prediction. In this research, we suggested a method for the prediction of disease-associated miRNAs, namely RWRHMDA. This approach was based on the RWRH algorithm, which was

successfully proposed in a heterogeneous network of genes and disease phenotypes for disease gene prediction. In particular, we used this algorithm to rank candidate miRNAs for disease on a heterogeneous network of phenotypes and miRNAs, which was developed by integrating a functional similarity network of shared target gene based microRNAs and a similarity network of disease phenotypes. We found that RWRHMDA significantly outperformed RWRMDA regardless of parameter settings by comparing the prediction performance of RWRHMDA with that of RWRMDA on a set of 35 disease phenotypes, because it better exploited the "disease module" concept. In addition, eight novel Alzheimer's disease-associated miRNAs were identified using the RWRHMDA system.

2. Paper Name: Defining Disease Phenotypes in Primary Care Electronic Health Records by a Machine Learning Approach: A Case Study in Identifying Rheumatoid Arthritis Shang-Ming Zhou is the author. i) variable selection by comparing relative frequencies of Read codes in the primary care dataset associated with disease cases versus non-disease controls (disease/non-disease based on secondary care diagnosis); ii) predictor/associated variable reduction using the Random Forest method; iii) decision rule induction from decision tree model. The proposed method was then extensively validated on an independent dataset, and compared for performance with two existing deterministic algorithms for RA, which had been developed using expert clinical knowledge.

3. Paper Name: Design and Implementing Heart Disease Prediction Using Naives Bayesian Anjan Nikhil Repaka and Sai Deepak Ravikanti are the authors. Data Mining is a fantastic technique of creation that revolves around discovering and digging out valuable data from vast data collections, which can be further useful in analyzing and drawing up trends for making business-related decisions. Speaking of the medical field, the application of data mining in this field will lead to the discovery and withdrawal of important trends and knowledge that can be useful in clinical diagnosis. The thesis focuses on the diagnosis of heart disease by considering previous evidence and knowledge. In order to predict risk factors for heart disease, SHDP (Smart Heart Disease Prediction) is constructed via Naives Bayesian to achieve this. The rapid development of technology has contributed to a notable increase in the online use of mobile health technology. In a standardized fashion, the necessary data is assembled. The following characteristics are collected from medical profiles to predict the risks of heart disease in a patient, including age, BP, cholesterol, sex, blood sugar, etc. For the Naives Bayesian classification for the prediction of

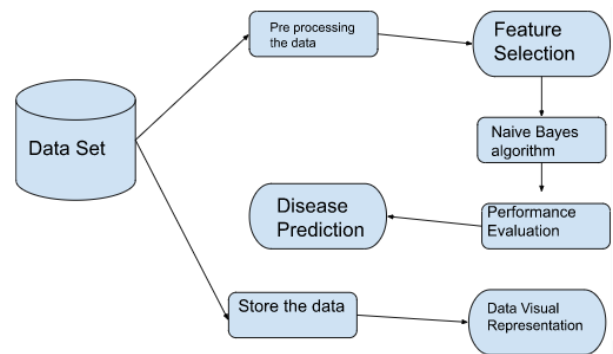
heart disease, the collected attributes serve as an input. The used dataset is divided into two parts: 80% of the dataset is used for training and 20% of the remainder is used for research. The proposed method involves the following phases: selection of datasets, user registration and login (application-based), Bayesian Navy classification, prediction, and stable data transfer using AES (Advanced Encryption Standard). Results are then produced. By making use of data mining approaches that are adopted for heart disease prediction, the study elaborates and introduces several information abstraction techniques. The result shows that the diagnostic framework developed effectively helps predict risk factors for heart disease.

III. PROBLEM STATEMENT AND OBJECTIVE

A. Problem Statement

Currently, the scenario is that if the patient is suffering from any symptoms, then he/she must visit the doctor or hospital to diagnose the disease. But, our main objective is to reduce the time taken by patients to diagnose the disease. Many patients are losing their lives only because of the late diagnosis of their disease. So our main aim is to reduce such deaths. Using Happy Healthing, the user can solve the problem and can get a basic idea of the diseases they are suffering from.

IV. FLOW-CHART



Fig

V. PROPOSED WORK

Our Happy Healthing predicts the majority of chronic diseases. It accepts the structured type of data as input to the machine learning model. Happy Healthing is used by end-users, i.e., patients or any user. In this system, the user will enter all the symptoms from which he or she is suffering. These symptoms are then given to the machine learning model to predict the disease. Algorithms are then applied to which one gives the best accuracy. Then the

system will predict disease on the basis of symptoms. Machine Learning Technology is used by Happy Health. The Nave Bayes algorithm is used to predict the disease by using symptoms.

A. Details of hardware and software

B. Software Requirements

- Python,
- PyCharm
- WINDOWS OS

Hardware Requirement:

- Webcam
- Processor-i3
- Hard disk-5GB
- Memory-2GB RAM

VI. DATASET AND MODEL DESCRIPTION

The dataset used in this system is in a structured format. The dataset which is used contains the disease name with all its symptoms. As our system is based on supervised learning machine algorithms, the dataset has a label of 0 or 1. Then we divide the dataset into a training dataset and a testing dataset. The model is trained by a training dataset. All algorithms were applied to this training dataset and then the machine learning model was trained. Then the testing dataset was provided to the trained model to test the accuracy of the model.

VII. MODULES

The Happy Healthing project is divided into 2 parts.

- **Trend Module:** The first part is trends in daily health for blood pressure and sugar, where there will be a visual representation of the user's reading for the past 5 days. Users can store the details using buttons and a visual representation of the data will be shown for analysis. For analysis Matplotlib and chartjs for graphs are used.
- **The Prediction Module:** The prediction module contains a machine learning model which will take 5 inputs from the user. The user has to select from a drop-down menu option and the data will be passed to the trained model of Naive Bayes. Using the machine learning algorithm, it will predict which disease the user has based on the symptoms.

VIII. LIMITATION

The Happy Healthing System has the following limitations:

- All information on this project is intended for personal educational and medical guidance only; patients should contact their physicians for accurate diagnosis and treatment.
- The patient's medical history has not been taken into account.
- Doesn't allow you to create data mining dimensions.

IX. EVALUATION METHOD

To calculate performance evaluation in the experiment, first, we denote TP, TN, Fp and FN as true positive (the number of results correctly predicted as required), true negative (the number of results not required), false positive (the number of results incorrectly predicted as required), false negative (the number of results incorrectly predicted as not required) respectively. We can obtain four measurements: recall, precision, accuracy, and F1 measures as follows:

$$\frac{\text{TruePositive} + \text{TrueNegative}}{\text{TruePositive} + \text{TrueNegative} + \text{FalsePositive} + \text{FalseNegative}}$$

$$\text{Precision} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}}$$

$$\text{Recall} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}}$$

$$\text{F1-Measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

X. ALGORITHM

NAIVE BAYES:

The Naive Bayes algorithm is a classification algorithm that uses Bayesian techniques and is based on the Bayes theorem in predictive modeling. Some algorithms are more computationally intensive than this one. As a result, it can be used to quickly generate mining models to find relationships between input columns and predictable columns. The Naive Bayes algorithm is primarily used in the creation of classifiers.

Naive Bayes is an easy but amazingly powerful rule for prognostication modeling. The independence assumption that allows decomposing joint likelihood into a product of marginal likelihoods is called 'naive'. This simplified Bayesian classifier is called naive Bayes. The Naive Bayes classifier assumes the presence of a particular feature in a class is unrelated to the presence of any other feature. It is very easy to build and useful for large datasets. Naive Bayes is a supervised learning model. Bayes theorem

provides some way of calculative posterior chance $P(b|a)$ from $P(b)$, $P(a)$ and $P(a|b)$. Look at the equation below:

$$P(b \vee a) = \frac{P(a \vee b)P(b)}{P(a)}$$

Above,

$P(b|a)$ is the posterior chance of class (b,target) given predictor (a, attributes).

$P(b)$ is the a priori probability of class.

$P(a|c)$ is that chance that is the chance of the predictor given class.

$P(a)$ is the a priori probability of a predictor.

In our system, Naïve Bayes decides which symptom is to put in classifier and which is not.

XI. RESULT AND DISCUSSION

The proposed system includes a Disease dataset, which is used to determine whether or not patients have disease based on their characteristics. The total number of records in the dataset is divided into two categories: training and testing data sets. The sigmoid function is commonly used to denote the logistic function; by using it, we can easily represent any real value to the range 1 or 0. The proposed system is applied to this data and attempts to create an accurate model that predicts whether or not the patients have this disease (data exploration and reading data). The accuracy of our model is 82.46%. We can still try to increase the accuracy but by increasing the epoch we can try to achieve more than 90%. SVM and RF are compared with different papers with their accuracy

TABLE I

Model	Train Accuracy	Validation Accuracy	Test Accuracy
Naïve Bayes	86%	88%	85%
SVM	73%	81%	79%
Random Forest	79%	83%	81%

Fig 4: Model Accuracy

XII. CONCLUSION

The main aim of Happy Healthing system is to predict the disease on the basis of the symptoms and allow the user to gain health insights. This system takes the symptoms of the user from which he or she suffers as input and generates final output as a prediction of disease. Average prediction accuracy probability of 100% is obtained. Happy Healthing was successfully implemented using the grails framework. This system gives a user-friendly environment and is easy to use. As the system is based on the web application, the user can use this system from anywhere and at any time. In conclusion, for disease

risk modeling, the accuracy of risk prediction depends on the diversity feature of the data.

XIII. REFERENCES

- [1] D. Kornack and P. Rakic, "Cell Proliferation without Neurogenesis in Adult Primate Neocortex," *Science*, vol. 294, Dec. 2001, pp. 2127-2130, doi:10.1126/science.1065467.
- [2] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.
- [3] R. Nicole, "Title of paper with only first word capitalized," *J. Name Stand. Abbrev.*, in press.
- [4] K. Elissa, "Title of paper if known," unpublished.
- [5] Duc-Hau Le, "Disease phenotype similarity improves the prediction of novel disease-associated microRNAs", "2015 2nd National Foundation for Science and Technology Development Conference on Information and Computer Science"
- [6] Shang-Ming Zhou, "Defining Disease Phenotypes in Primary Care Electronic Health Records by a Machine Learning Approach: A Case Study in Identifying Rheumatoid Arthritis", "May 2, 2016, <https://doi.org/10.1371/journal.pone.0154515>"
- [7] Anjan Nikhil Repaka, "Design and Implementing Heart Disease Prediction Using Naives Bayesian", "Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8".
- [8] Pahulpreet Singh Kohli, "Application of Machine Learning in Disease Prediction", "2018 4th International Conference on Computing Communication and Automation (ICCCA)"
- [9] Deeraj Shetty, "Diabetes Disease Pre-diction Using Data Mining", "2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)