IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

MULTIDISEASE DIAGNOSIS WITH DOCTOR RECOMMENDATION SYSTEMUSING DEEP LEARNING ALGORITHM

[1]A.Reeta joldrine , [2]A,Y Hilden Pieo, [3]K.Prasanth, [4]R.Jeeva,[5]P.Arun Bharathi

[1]Assistant Professor, [2,3,4,5]Student [1]Computer Science and Engineering,
[1]Sri Ramakrishna College Of Engineering, Perambalur, Tamil Nadu, India

Abstract : Data mining for healthcare is an interdisciplinary subject of research that has its roots in database statistics and can be used to assess the efficacy of medical treatments. Many existing machine learning models for health care analysis focus on a single ailment at a time. For example, one analysis could be for diabetes, another for cancer, and yet another for cancer disorders. There is no universal approach that can forecast multiple diseases with a single analysis. This project proposes a system that uses Python Flask API to forecast numerous diseases. Diabetes analysis, heart disease analysis, and breast cancer analysis were all used in this investigation. Machine learning methods, Pandas, and the Flask API were used to implement multiple illness analysis. Python pickling is used to save model behaviour, while Python unpickling is used to load the pickle file. The significance of this research is that it analyses diseases and includes all of the parameters that produce the condition, making it possible to detect the disease's maximal impact. Using the KAGGLE dataset, we conduct an exhaustive search of all available feature variables within the data to develop models for cardiovascular, cancer detection, and diabetes detection. Using different time-frames and feature sets for the data (based on laboratory data), machine learning model named as Multi layer perceptron algorithm is implemented to predict the diseases with improved accuracy.

IndexTerms **-Data mining,Healthcare,Python pictling,Python Flask API.**

INTRODUCTION

Data mining is the process of discovering patterns, trends, and insights from large datasets using various statistical and machine learning techniques. The goal of data mining is to extract valuable information from data and use it to make informed business decisions. Data mining involves several steps, including data collection, data cleaning, data integration, data transformation, data reduction, pattern identification, and evaluation. Data mining techniques include clustering, classification, regression, association rule mining, and anomaly detection.

Data mining can be performed on various types of data, including structured, semi-structured, and unstructured data. Structured data is organized and stored in a predefined format, such as databases, spreadsheets, or tables, and can be easily processed by machines. Semi-structured data, on the other hand, has some organizational structure but is not strictly defined, such as XML files, JSON, or HTML documents.

PROBLEMS IN DEEP LEARNING APROACHES

2.1 PREDICTIVE ANALYSIS

Predictive analytics involves using historical data to make predictions about future events or behaviors. Machine learning algorithms can be used for tasks such as predictive modeling and time series forecasting. For example, machine learning algorithms can be used to predict customer churn or forecast sales revenue.

2.2 ANOMALY DETECTION

Anomaly detection involves identifying unusual patterns or events in a dataset that deviate from the norm. Machine learning algorithms can be used for tasks such as fraud detection and intrusion detection. For example, machine learning algorithms can be used to detect credit card fraud or identify network intrusions.

2.3 RECOMMENDER SYSTEM

Recommender systems involve suggesting items or products to users based on their preferences or past behavior. Machine learning algorithms can be used for tasks such as collaborative filtering and content-based filtering. For example, machine learning algorithms can be used to recommend movies or products to users based on their viewing or purchasing history.

2.4 NATURAL LANGUAGE PROCESSING

Natural language processing involves processing and analyzing human language data. Deep learning algorithms can be used for tasks such as sentiment analysis and text classification. For example, deep learning algorithms can be used to classify customer reviews as positive or negative or identify the topic of a news article.

EXISTING SYSTEM

Present days one of the major application areas of machine learning algorithms is medical diagnosis of diseases and treatment. Machine learning algorithms also used to find correlations and associations between different diseases. Nowadays many people are dying because of diabetics and cardiovascular diseases. Prediction and diagnosing of diabetic and heart disease becomes a challenging factor faced by doctors and hospitals both in India and abroad.

In order to reduce number of deaths because of diabetic and heart diseases, we have to predict whether person is at the risk of diabetic and heart disease or not in advance. Data mining techniques and machine learning algorithms play a very important role in this area.

. In this existing system, focused on how data mining techniques can be used to predict diabetic disease in advance such that patient is well treated. An important task of any diagnostic system is the process of attempting to determine and/or identify a possible disease or disorder and the decision reached by this process.

This System mainly focuses on the supervised learning technique called the Random forests for classification of data by changing the values of different hyper parameters in Random Forests Classifier and also implement Navies Bayes algorithm to classify the disease whether it is appeared or not.

PROPOSED SYSTEM

Multiple opportunities for healthcare are created because machine learning models have potential for advanced predictive analytics. In this project, we use supervised machine learning models to predict diabetes, heart disease and cancer disease. Despite the known association between these diseases, we design the models to predict cancer, heart disease and diabetes separately in order to benefit a wider range of patients.

We also consider the prediction of pre-diabetes and undiagnosed diabetes. This project also explores a support vector machine model which combines the results of multiple supervised learning models to increase prediction ability. In this study, we utilize multiple supervised learning models for classification of at-risk patients. In supervised learning, the learning algorithm is provided with training data that contains both the recorded observations and the corresponding labels for the category of the observations.
The algorithm uses this information to build a model that, when given new observations, can predict which output label should be associated with each new observation. Multi-layer perceptron algorithm is used to classify data by separating the classes with a boundary, i.e. a line or multi-dimensional hyper plane.

Optimization ensures that the widest boundary separation of classes is achieved. While MLP often outperforms logistic regression, the computational complexity of the model results in long training durations for model development.

IV. MODULE DESCRIPTION

4.1 Framework Construction

In this module, create the login for admin and user login. Admin can upload the datasets related to heart and diabetic disease. A data set (or dataset, although this spelling is not present in many contemporary dictionaries like Merriam-Webster) is a collection of data. Most commonly a data set corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in question. The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum. ML and AI algorithms for disease detection are algorithmic models that analyze medical data to find signs of diseases before they become severe.The data set may comprise data for one or more members, corresponding to the number of rows. The term data set may also be used more loosely, to refer to the data in a collection of closely related tables, corresponding to a particular experiment or event. In this module, we can upload the datasets related to diabetic, heart and cancer diseases which includes the attributes such as age, gender, height, weight, systolic blood pressure, diastolic blood pressure, cholesterol, glucose, smoke, alcohol, active status, cardio labels.

4.2 Preprocessing

Data pre-processing is an important step in the [data mining] process. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine learning projects. Data-gathering methods are often loosely controlled, resulting in out-of-range values, impossible data combinations, missing values, etc. Analyzing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data is first and foremost before running an analysis. If there is much irrelevant and redundant information present or noisy and unreliable data, then knowledge discovery during the training phase is more difficult. Data preparation and filtering steps can take considerable amount of processing time. In this module, we can eliminate the irrelevant values and also estimate the missing values of data. Finally provide structured datasets.

4.3 Feature Extraction

Feature selection refers to the process of reducing the inputs for processing and analysis, or of finding the most meaningful inputs. A related term, feature engineering (or feature extraction), refers to the process of extracting useful information or features from existing data. Filter feature selection methods apply a statistical measure to assign a scoring to each feature. The features are ranked by the score and either selected to be kept or removed from the dataset. The methods are often uni-variate and consider the feature independently, or with regard to the dependent variable. It can be used to construct the diabetic, heart and cancer diseases. In this module, select the multiple features from uploaded datasets. The MLP when acting as a feature extractor provides a supervized nonlinear mapping of the input space into its hidden layer(s). Feature extraction for exploratory data projection enables high-dimensional data visualization for better data structure understanding and for cluster analysis.The process of transforming raw data into numerical features that can be processed while preserving the information in the original data set. It yields better results than applying machine learning directly to the raw data

4.4 Classification

In this module implement classification algorithm to predict the diabetic and heart diseases. And using machine learning algorithm such as Support vector machine algorithm to predict the diseases. A MLP is a feed forward vector model that maps sets of input data onto a set of appropriate outputs. It consists of multiple vectors of nodes in a directed graph, and each layer is fully connected to the next one. In the MLP algorithm, we plot each data item as a point in n-dimensional space (where n is a number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well. MLPs are simply the coordinates of individual observation. The MLP classifier is a frontier that best segregates the two classes (hyper-plane/ line). User can provide the features and automatically predict the diseases.

4.5 Recommendation

Healthcare industry generates terabytes of data every year. The medical documents maintained are a pool of information regarding patients. The task of extracting useful information or quality healthcare is tricky and important. At present in order to remain healthy, regular body diagnosis is necessary. Today, there are multiple sources available as individual prediction or recommendation system but the need of the hour is to have an integrated model comprising both. Also, it would be more appropriate and convenient if people could get basic diagnosis online 24x7 rather than visiting hospitals & clinics frequently. Thus, reducing cost and saving time. If certain anomalies found in the diagnosis, then recommendation of nearby specialist and hospitals according to user's preference would facilitate in quick and appropriate treatment. Healthcare being a domain evolving continuously and generating a huge amount of data develops a need to use the data for useful knowledge which attracts large organizations to invest heavily in this field. At present in order to remain healthy, regular body diagnosis is necessary. all the needful and adequate information regarding the predicted disease as well as the recommended doctors is provided.

CONCLUSION:

The use of data mining in medical data analysis is an excellent way to consider the existing correlations between variables. We've proven that mining can help us find useful connection even when the traits we're looking for aren't direct indications of the class we're trying to forecast. In our research, we attempted to forecast the likelihood of developing system for predicting diabetic, heart and cancer disease datasets, and we show that the proposed system provide improved accuracy rate in disease prediction. This type of classifier can aid in the early detection of a diabetic patient's and future prediction. Patients can be forewarned to adjust their lifestyle this way. This will result in preventing from multiple diseases there by resulting in low mortality rates as well as reduced cost on health for the state. MLP's have proven to be a classification technique with excellent predictive performance and also been investigated with the help of ROC curve for both training and testing data. Hence this MLP model can be recommended for the classification of the diseases and recommend the doctors based on disease predictions. And also, extend the framework to implement multiple diseases and recommend the diagnosis information like doctor suggestion, prescription and so on. .

REFERENCES

[1] Aamir, Sanam, et al. "Predicting breast cancer leveraging supervised machine learning techniques." Computational and Mathematical Methods in Medicine 2022 (2022).

[2] J. Amin, M. Sharif, M. Yasmin, T. Saba, and M. Raza, ‘‘Use of machine intelligence to conduct analysis of human brain data for detection of abnormalities in its cognitive functions,’’ Multimedia Tools Appl., vol. 79, nos. 15–16, pp. 10955–10973, Apr. 2020.

[3] Z. Du, Y. Yang, J. Zheng, Q. Li, D. Lin, Y. Li, J. Fan, W. Cheng, X.-H. Chen, and Y. Cai, ‘‘Accurate prediction of coronary heart disease for patients with hypertension from electronic health records with big data and machinelearning methods: Model development and performance evaluation,’’ JMIR Med. Informat., vol. 8, no. 7, Jul. 2020, Art. no. e17257.

[4] El-Hasnony, Ibrahim M., et al. "Multi-label active learning-based machine learning model for heart disease prediction." Sensors 22.3 (2022): 1184.

[5] A. Garg, B. Sharma, and R. Khan, ‘‘Heart disease prediction using machine learning techniques,’’ in Proc. IOP Conf. Ser., Mater. Sci. Eng., vol. 1022, 2021, Art. no. 01204

[6] J. A. W. Gold, F. B. Ahmad, J. A. Cisewski, L. M. Rossen, A. J. Montero, K. Benedict, B. R. Jackson, and M. Toda, ‘‘Increased deaths from fungal infections during the coronavirus disease 2019 pandemic—National vital statistics system, United States, January 2020–December 2021,’’ Clin. Infectious Diseases, vol. 76, no. 3, pp. e255–e262, Feb. 2023.

[7] Humayun, Mamoona, et al. "Framework for detecting breast cancer risk presence using deep learning." Electronics 12.2 (2023): 403.

[8] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, ‘‘Heart disease prediction using machine learning algorithms,’’ in Proc. IOP Conf. Ser., Mater. Sci. Eng., vol. 1022, 2021, Art. no. 012072.

[9] Nadakinamani, Rajkumar Gangappa, et al. "Clinical data analysis for prediction of cardiovascular disease using machine learning techniques." Computational intelligence and neuroscience 2022 (2022).

[10] Nasir, Muhammad Umar, et al. "Breast cancer prediction empowered with fine-tuning." Computational Intelligence and Neuroscience 2022 (2022).

[11] Nouman, Aleeza, and Salman Muneer. "A systematic literature review on heart disease prediction using blockchain and machine learning techniques." International Journal of Computational and Innovative Sciences 1.4 (2022): 1-6.

[12] M. Pal, S. Parija, G. Panda, K. Dhama, and R. K. Mohapatra, ‘‘Risk prediction of cardiovascular disease using machine learning classifiers,’’ Open Med., vol. 17, no. 1, pp. 1100–1113, Jun. 2022.

[13] M. Pal and S. Parija, ‘‘Prediction of heart diseases using random forest,’’ J. Phys., Conf. Ser., vol. 1817, no. 1, Mar. 2021, Art.no. 012009.

Rane, Nikita, et al. "Breast cancer classification and prediction using machine learning." International Journal of EngineeringResearch and Technology 9.2 (2020): 576-580.

[14] Rahman, Md Mahbubur. "A web-based heart disease prediction system using machine learning algorithms." NetworkBiology 12.2 (2022): 64.

[15] Sarra, Raniya R., et al. "Enhanced heart disease prediction based on machine learning and $\chi 2$ statistical optimal feature selection model." Designs 6.5 (2022): 87.

[16] Shamrat, FM Javed Mehedi, et al. "An analysis on breast disease prediction using machine learning approaches." International Journal of Scientific & Technology Research 9.02 (2020): 2450-2455.

[18]K. Shah, K. Sharma, and D. Saxena, ‘‘Editorial: Health technology assessment in cardiovascular diseases,’’ Frontiers Cardiovascular Med., vol. 10, Jan. 2023, Art. no. 1108503

[19] A. Srivastava and A. k. Singh, ‘‘Heart disease prediction using machine learning,’’ in Proc. 2nd Int. Conf. Advance Comput. Innov. Technol. Eng. (ICACITE), Apr. 2022, pp. 2633–2635.

[20] Thilagavathi, G., et al. "Heart disease prediction using machine leaning algorithms." 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC). IEEE, 2022.