International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 12 | December - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Machine Learning Implementation on Medical Domain to Identify Disease Insights

Mary Ratna Manjari, K.Aravind, D.Aravind, D.Akhilesh Goud, Y.Arun, Y.Aryan
Dr RamaRao, Dr Thayabba Khatoon
Artificial Intelligence & MachineLearning
Department Of Computer Science And Engineering
Malla Reddy University, Hyderabad,Telangana,India

ABSTRACT

Machine learning has become as a part of our lives and we are living with the technology. We need to understand what is happening with the health of the person to be precise we need to analyze our own health. In this scenario we are implementing machine learning methodology in health care information for the problem statement of personalizing the medical information which is a private information we need to make it safe while using and implementing some sort of algorithms. In this paper we are discussing about understanding the human disease patterns and using which random forest and other machine learning models work and predict the actual procedure a person has to follow to get a good health and avoid the different health loss activities we are doing regularly. In this random forest is the most accurate algorithm worked with this concept and we need to analyze the other reasons for understanding which kind of information is most useful for performing machine learning. Machine learning cannot be implemented for all type of issues in the real time. But we can maintain a better break through of machine learning implementation on medical issues as mentioned in this article. We are performing a better algorithm to understand the human problems related to health care and we are proposing with sample implementations and explanation with relevant results. We tried to implement IT algorithm which gives the trust on the algorithm based on the truth maintenance system.

I. INTRODUCTION

The project "identifying disease insights using machine learning" aims to explore how machine learning techniques can be used to analyze different types of data and extract useful insights for disease diagnosis. The types and sources

data that can be used for disease diagnosis, such as electronic health records, laboratory results, medical images, and genomic data. The methods and tools for preprocessing, cleaning, and transforming the data to make it suitable for machine learning analysis, The machine learning algorithms and models that can be used for disease diagnosis, such as classification, regression, clustering, and deep learning. The evaluation and validation of the machine learning models, such as accuracy, precision, recall, and confusion matrix.

The ethical and social implications of using machine learning for disease diagnosis, such as privacy, security, fairness, and accountability. The project will use Python as the programming language and various libraries and frameworks, such as scikit-learn, TensorFlow, Keras, and PyTorch, for machine learning implementation.
The project "identifying disease insights using machine learning" is intended for anyone who is interested in learning how machine learning can be applied to healthcare problems and how to use machine learning tools and techniques for disease diagnosis. The project will provide a comprehensive and hands-on introduction to the topic and will help the learners to develop the skills and knowledge needed to use machine learning for disease diagnosis.

II. LITERATURE REVIEW

ML has many applications in various fields, including healthcare. One of the challenges in healthcare is to identify and diagnose diseases using various sources of information, such as symptoms, medical tests, images, and genomic data. Identifying disease insights using ML can help improve the accuracy, efficiency, and effectiveness of disease diagnosis and treatment. Several studies have explored the use of ML Techniques for disease diagnosis in different domains. Ahsan et al1 conducted a comprehensive review of the literature on machine-learning-based disease diagnosis (MLBDD) and summarized the most recent trends and approaches in terms of algorithm, disease type, data type, application, and evaluation metrics. They also highlighted the key results and future opportunities in the MLBDD area. Alanazi 2 proposed a system that uses convolutional neural networks (CNN) and K-nearest neighbor (KNN) algorithms to identify and predict chronic diseases such as diabetes, heart disease, and kidney disease based on the patient's symptoms and living habits. They compared their system with other algorithms such as Naïve Bayes, decision tree, and logistic regression and showed that their system achieved better performance. Kumar et al3 presented a systematic literature review on the use of artificial intelligence techniques to diagnose various diseases such as Alzheimer, cancer, diabetes, chronic heart disease, tuberculosis, stroke and cerebrovascular,

hypertension, skin, and liver disease. They analyzed the advantages and disadvantages of different techniques and discussed the ethical and social implications of using AI for disease diagnosis. Singh et al4 provided a literature review on ML techniques used in the agricultural sector, mainly focusing on the classification and detection of plant diseases. They described the different types of data and methods used for plant disease diagnosis and suggested some future research directions. In this kind of methodology, we have the ancient
books where we can get the related information regarding different tree root or the leaves information which are used for the specific kind of disease treatment implementations. For an instant in ancient days we used to use cumin seeds for the digestion of the person if the person is suffering with constipation. In the save way the medication now a days we are using the same kind of chemical composition of the cumin seeds for the digestion. In this scenario first we need to the understand the two systems which are existing in the current world and we need to understand the implementation of the third way which is an unknown to lot more people

III. PROBLEM STATEMENT

Disease diagnosis is a vital and complex process in healthcare that requires accurate and timely analysis of various sources of information, such as symptoms, medical tests, images, and genomic data. However, the current methods of disease diagnosis are often manual, subjective, and error-prone, which can lead to misdiagnosis, delayed treatment, and increased costs. Moreover, the rapid growth and diversity of data in healthcare pose new challenges and opportunities for disease diagnosis, such as data quality, integration, and interpretation. The project "identifying disease insights is using machine learning" aims to bridge the gap between the current and the desired situation of disease diagnosis by using machine learning (ML) techniques to analyze different types of data and extract useful insights for disease diagnosis. ML is a branch of artificial intelligence (AI) that enables computers to learn from data and make predictions or decisions. ML has many applications in various fields, including healthcare.

The project hopes to achieve the following goals and objectives:

To improve the accuracy, efficiency, and effectiveness of disease diagnosis by using ML models that can handle large and diverse datasets, learn from patterns and relationships, and provide reliable and consistent results.

To enhance the understanding and explanation of disease diagnosis by using ML models that can provide interpretable and transparent insights, such as the features, factors, and causes of diseases, and the evidence and rationale for the predictions or decisions.

To empower the users and stakeholders of disease diagnosis by using ML models that can support and augment their decision-making, such as the patients, doctors, nurses, researchers, and policymakers.

The project will also explore the ethical and social implications of using ML for disease diagnosis, such as privacy, security, fairness, and accountability, and propose solutions and recommendations to address them.

The project "identifying disease insights using machine learning" will benefit the healthcare domain and society in general by providing a comprehensive and hands-on introduction to the topic and by developing the skills and knowledge needed to use ML for disease diagnosis. The project will also contribute to the advancement of science and technology by creating and sharing new knowledge and innovations in the field of ML and healthcare. The project will ultimately improve the quality and accessibility of healthcare services and outcomes, and enhance the health and well-being of individuals and communities.

IV. METHODOLOGY

The project will follow a standard machine learning workflow, which consists of the following steps: data collection, data preprocessing, data analysis, model evaluation, and model deployment. The project will use publicly available datasets from various sources, such as Kaggle, UCI Machine Learning Repository, and COVID-19 Open Research Dataset, that contain information about different diseases, such as chronic diseases, cancer, COVID-19, and plant diseases. The project will select the datasets that are relevant, reliable, and representative for the problem domain and the research objectives. The project will apply various techniques to clean, transform, and prepare the data for machine learning analysis.

DATA PREPROCESSING TECHNIQUES

Data cleaning: The project will remove or handle any missing, noisy, or inconsistent data, such as outliers, duplicates, or errors, that can affect the quality and performance of the machine learning models.

Data transformation: The project will convert the data into a suitable format and scale for machine learning analysis, such as encoding categorical variables, normalizing numerical variables, or applying feature engineering techniques to create new or modify existing features.

Data reduction: The project will reduce the dimensionality and complexity of the data by selecting or extracting the most relevant and informative features for machine learning analysis, such as applying feature selection methods, such as filter, wrapper, or embedded methods, or feature extraction

methods, such as principal component analysis (PCA) or linear discriminant analysis (LDA).

Data analysis: The project will use various machine learning techniques to analyze the data and extract useful insights for disease diagnosis.

Model evaluation: The project will evaluate and validate the machine learning models using various metrics and techniques, such as accuracy, precision, recall, f1-score, confusion matrix, cross-validation, or bootstrap, to measure the performance and reliability of the models. The project will also use explainable and interpretable machine learning techniques, such as feature importance, partial dependence plots, or SHAP values, to understand and explain the predictions or decisions of the models and to provide meaningful and transparent insights for disease diagnosis.

Model deployment: The project will deploy and integrate the machine learning models into a web or mobile application that can provide user-friendly and interactive interfaces for disease diagnosis. The project will use different tools and platforms, such as Flask, Django, or Streamlit, to develop and host the application. The project will also monitor and update the machine learning models to ensure their accuracy and efficiency over time.

ML ALGORITHMS

Machine learning algorithms are methods that can learn from data and make predictions or decisions. They can be used to detect common diseases like cold, cough, etc. based on the symptoms or other features of the patients.

Decision tree: A decision tree is a graphical representation of a series of rules or questions that can classify the data into different categories. For example, a decision tree can ask questions like “Do you have a fever?”, “Do you have a sore throat?”, or “Do you have a runny nose?” to determine if the patient has a cold, a flu, or an allergy.

Support vector machine: A support vector machine is a mathematical model that can find the best boundary or line that separates the data into different classes. For example, a support vector machine can find the best line that separates the patients who have a cold from those who do not, based on their symptoms or other features

K-nearest neighbor: A k-nearest neighbor is a simple algorithm that can classify the data based on the similarity or distance to the nearest neighbors. For example, a k-nearest neighbor can classify a patient as having a cold or not, based on the symptoms or other features of the k most similar or closest patients in the data

Naïve Bayes: A naïve Bayes is a probabilistic model that can calculate the likelihood probability of the data belonging to different classes, based on some assumptions or prior knowledge. For example, a naïve Bayes can calculate the probability of a patient having a cold or not, based on the probability of having a cold given the symptoms or other features of the patient

V. ARCHITECTURE

These are the various sources of data that can be used for disease diagnosis, such as electronic health records, laboratory results, medical images, and genomic data. Data preprocessing This is the component that applies various techniques to clean, transform, and prepare the data for machine learning analysis. The data preprocessing techniques include data cleaning, data transformation, and data reduction. Machine learning models these are the models that use various machine learning techniques to analyze the data and extract useful insights for disease diagnosis. The machine learning models include supervised learning, unsupervised learning, and deep learning models.

It shows the steps involved in training and validating a model, and making predictions. The flowchart starts with the dataset, which is split into train and validation data. The train data is preprocessed and used to train different classifiers, such as SVM, Naive Bayes, and Random Forest. The validation data is used to make predictions using the trained classifiers and to evaluate their performance. The flowchart also shows the inputs and outputs of each step, such as features, labels, accuracy, and confusion matrix. The flowchart is a visual representation of how machine learning can be used to solve a classification problem.

DISEASE CLASSIFICATION

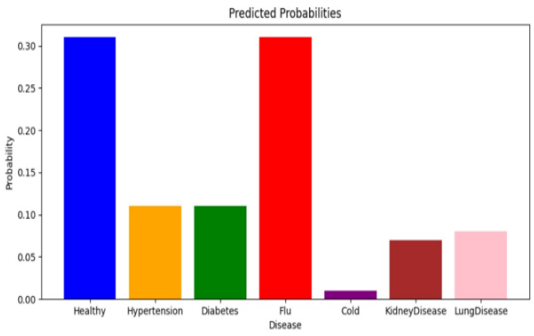
Disease classification is the process of assigning a disease to a specific category or group based on its characteristics, such as symptoms, causes, or effects. Disease classification can help in diagnosis, treatment, prevention, and research of diseases. There are different systems and methods of disease classification, such as the International Classification of Diseases (ICD), the Diagnostic and Statistical Manual of Mental Disorders (DSM), or the machine learning approach.

DISEASE DETECTION

Machine learning can be used to detect diseases based on symptoms using various algorithms and models, such as decision tree, support vector machine, k-nearest neighbor, naive Bayes, or logistic regression. These algorithms and models can analyze the symptoms or other features of the patients and assign them to different classes or categories of diseases, such as cold, flu

RESULT

The software is open in a window with a blue header and white body. The window has a table with columns for age, gender, weight, blood pressure, sugar level, and cholesterol level. The table has rows for different predictions for different combinations of the above parameters. There is a pop-up window with the predicted disease for the selected row in the table. The background is a dark grey color with other windows open.



VI. CONCLUSION

The project "identifying disease insights using machine learning" is a comprehensive and hands-on introduction to the topic of using machine learning techniques to analyze different types of data and extract useful insights for disease diagnosis. The project covers the following topics:

The types and sources of data that can be used for disease diagnosis, such as electronic health records, laboratory results, medical images, and genomic data.

The methods and tools for preprocessing, cleaning, and transforming the data to make it suitable for machine learning analysis.

The machine learning algorithms and models that can be used for disease diagnosis, such as classification, regression, clustering, and deep learning.

The evaluation and validation of the machine learning models, such as accuracy, precision, recall, and confusion matrix.

The ethical and social implications of using machine learning for disease diagnosis, such as privacy, security, fairness, and accountability.

FUTURE WORK

The future work of the project "identifying disease insights using machine learning" could include the following aspects:

To use more diverse and comprehensive datasets from different sources and regions, to increase the generalizability and applicability of the machine learning models for disease diagnosis.

To explore more advanced and novel machine learning techniques, such as reinforcement learning, transfer learning, or federated learning, to improve the performance and efficiency of the machine learning models for disease diagnosis.

To incorporate more domain knowledge and expert feedback into the machine learning models, to enhance the explainability and interpretability of the machine learning models for disease diagnosis.

VII. REFERENCE

"Radiogenomics for Precision Medicine With a Big Data Analytics
Perspective" Andreas S. Panayides et al, Volume 23, No:5, September 2019

"Intelligent Analysis of Medical Big Data Based on Deep Learning" by HANQING SUN et al, Volume 7, 2019, special section on deep learning algorithms for internet of medical things

"Harnessing the power of machine learning in dementia informatics research: Issues, opportunities and Challenges", Gavin Tsang et al., Volume 3, 2020, IEEE Access

From webpage "Disease Prediction Using Machine Learning "– GeeksforGeeks

From webpage "Identification and Prediction of Chronic Diseases Using Machine Learning Approach"

From webpage "Identifying disease genes using machine learning and gene functional similarities"