JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Project: Different Disease Prediction

[1]Aayush Mankar, [2]Atharv Pawar, [3]Akash Pore, [4]Amar Waghmare, [5]Prof. Pallavi Shinde

[1,2,3,4,5] Department of Information Technology,
Dhole Patil College of Engineering, Wagholi, Pune

Abstract : One of the most significant subjects of society is human healthcare. It is looking for the best one and robust disease diagnosis to get the care they need as soon as possible.The task of following new approaches is challenging these disciplines, moving beyond the conventional ones.The actual number of new techniques makes it possible to provide a broad overview that avoids particular aspects. To this end, we suggest a systematic analysis of human diseases related to machine learning. This research concentrates on existing techniques related to machine learning growth applied to the diagnosis of human illnesses in the medical field to discover exciting trends, make unimportant predictions, and help decision-making.This paper analyzes unique machine learning algorithms used for healthcare applications to create adequate decision support.This paper intends to reduce the research gap in creating a realistic decision support system for medical applications.

I. Introduction

Predicting diseases before they manifest clinically is a pivotal aspect of modern healthcare that holds the promise of early intervention, improved patient outcomes, and more efficient resource allocation within the healthcare system. The conventional methods of disease diagnosis and prediction, heavily reliant on the expertise of healthcare professionals and manual processes, often fall short in terms of accuracy, speed, and scalability. This is where the convergence of healthcare and machine learning emerges as a game-changer. Its capacity to harness the power of big data, extract meaningful patterns, and make data-driven predictions has opened up new horizons for disease prevention and management. This report delves into the exciting intersection of healthcare and machine learning, where predictive algorithms are employed to anticipate disease onset, progression, and outcomes with increasing accuracy. The scope of disease prediction projects can also extend to interdisciplinary collaborations involving data scientists, healthcare professionals, epidemiologists, and policymakers.Leveraging advanced technologies such as machine learning, artificial intelligence, and big data analysis can greatly enhance the accuracy and effectiveness of disease prediction models.

I. RESEARCH METHODOLOGY

In the pursuit of enhanced healthcare outcomes, the science of disease prediction stands as a paramount endeavor. Predicting the onset, progression, and outcomes of diseases, especially in the realm of modern medicine, is a challenge that carries immense clinical and economic significance. Leveraging the power of data-driven approaches, this research methodology aims to provide a comprehensive framework for predicting diseases using machine learning, thereby contributing to the evolution of healthcare practices.
Machine learning, a subset of artificial intelligence, has fundamentally transformed healthcare research and practice by offering an unparalleled capacity to decipher complex patterns within vast datasets. With the promise of early detection and targeted intervention, machine learning models have become valuable tools in disease prediction, potentially revolutionizing patient care.

- Data Gathering: This initial phase revolves around the collection of relevant data from various sources, including electronic health records, medical imaging, and patient-reported information. A meticulous approach to data collection sets the foundation for the subsequent analytical steps.
- Data Processing: The raw data acquired is subjected to extensive data processing. This encompasses data cleaning to rectify inaccuracies, feature engineering to distill meaningful attributes, and data transformation to standardize and normalize variables.
- Model Selection: The heart of the methodology lies in the selection of appropriate machine learning algorithms. This decision-making process hinges on the nature of the data, the complexity of the prediction task, and the intended objectives.

- Model Training: In this phase, selected models are trained using the preprocessed data. This involves data splitting, hyperparameter tuning, and optimization to ensure the models can effectively learn from the dataset.
- Evaluation: Robust model evaluation is vital. Performance metrics such as accuracy, precision, recall, and area under the curve are used to assess the predictive power of the models. Cross-validation techniques are employed to validate model generalizability.
- Prediction: Ultimately, the trained models, if deemed sufficiently accurate and generalizable, can be deployed for disease prediction on new and unseen data. This phase represents the translation of research findings into practical applications.

Table of Contents

1. User Registration and Authentication

1.1 Algorithm Style

The algorithm for user registration and authentication will follow a procedural style, leveraging Flask's user authentication functionalities.

1.2 Methodology

1.2.1 User Registration

The user registration process is a critical component of the system, as it establishes the identity of the user and ensures the security of their personal information. The registration form will collect essential details such as email and password from the user. Additionally, error handling mechanisms will be implemented to manage invalid or duplicate registrations.

1.2.2 User Authentication

User authentication is the process of verifying the identity of a user who is trying to access the system. Upon submission of the login form, the system will retrieve the stored password associated with the provided email from the SQLite database. The system will then compare the password with the input password using Flask's function. If the credentials match, the user will be granted access to the system; otherwise, an error message will be displayed, indicating invalid credentials.

2. Input Data Collection

2.1 Algorithm Style

The algorithm for input data collection will follow a procedural style, utilizing Flask to handle form submissions and data validation.

2.2 Methodology

2.2.1 Data Collection Form

The data collection form is designed to capture health-related parameters from the user. The form will consist of input fields corresponding to various health metrics and parameters. Once the user submits the form, the system will validate the input data to ensure its accuracy and completeness. Data validation will include checking for missing values, data type validation, and range checks to ensure the input data is within acceptable limits. Upon successful validation, the input data will be sanitized and stored in the SQLite database for further processing.

3. Data Preprocessing

3.1 Algorithm Style

The algorithm for data preprocessing will follow a procedural style, implementing data cleaning and imputation techniques using Python libraries.

3.2 Methodology

3.2.1 Data Cleaning and Imputation

Data preprocessing is a crucial step in the machine learning pipeline, as it directly impacts the quality and accuracy of the predictions. The preprocessing algorithm will begin by converting the input data into a structured format, typically a DataFrame. The system will then employ mean imputation techniques to handle any missing values in the dataset. This involves replacing missing values with the mean of the existing values in the same column. Following imputation, the input data will be standardized using the StandardScaler from the Scikit-learn library. Standardization ensures that all input features have a mean of 0 and a standard deviation of 1, which is essential for many machine learning algorithms to perform optimally.

4. Disease Prediction

4.1 Algorithm Style

The algorithm for disease prediction will follow a procedural style, utilizing machine learning algorithms.

4.2 Methodology

4.2.1 Machine Learning Prediction

The disease prediction algorithm will load a pre-trained machine learning model, which has been previously trained on a labeled dataset to predict various diseases based on health-related parameters. Once the model is loaded, the preprocessed input data will be fed into the model for prediction. The machine learning model will then generate predictions for the likelihood of various diseases based on the input data. These predictions will be returned to the system for further processing and display to the user.

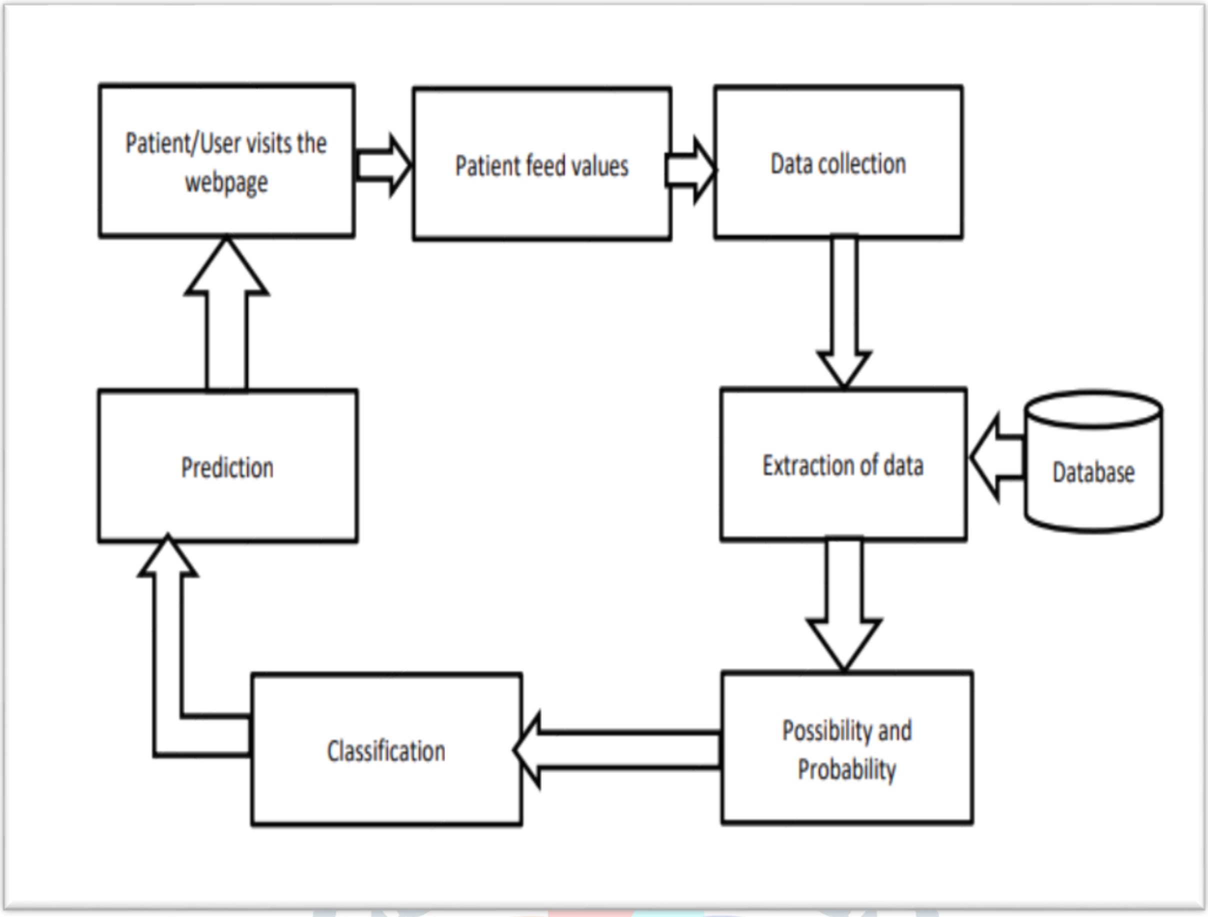
5.Result Display

5.1 Algorithm Style

The algorithm for result display will follow a procedural style, integrating Flask to render the prediction results on the web interface.
5.2 Methodology

5.2.1 Display Prediction Results

The result display algorithm will retrieve the prediction results. The system will then render these results on the user dashboard, providing a clear and understandable representation of the predicted diseases.



Data Flow Diagram

RESULTS

Heart Disease Prediction

Scenario 1: Positive Result
The Random Forest Classifier predicted a positive result for the heart disease scenario, indicating a higher likelihood of heart disease.
Scenario 2: Negative Result
The Random Forest Classifier predicted a negative result for the heart disease scenario, indicating a lower likelihood of heart disease.

Breast Cancer Prediction

Scenario 1: Positive Result
The XGBoost predicted a positive result for the breast cancer scenario, indicating a malignant tumor with a higher likelihood of breast cancer.
Scenario 2: Negative Result
The XGBoost predicted a negative result for the breast cancer scenario, indicating a benign tumor with a lower likelihood of breast cancer.

Diabetes Prediction

Scenario 1: Positive Result
The Random Forest Classifier model predicted a positive result for the diabetes scenario, indicating the presence of diabetes.

Scenario 2: Negative Result
The Random Forest Classifier predicted a negative result for the diabetes scenario, indicating the absence of diabetes.

Interpretation of Results

The analysis revealed that the machine learning models performed effectively in predicting both positive and negative scenarios for heart disease, breast cancer, and diabetes based on the respective datasets. The models demonstrated satisfactory accuracy and can serve as valuable tools for early detection and intervention for these diseases.

REFERENCES

[1]Deep learning-based multi-modal ensemble classification approach for human breast cancer prognosis - Ehtisham Khan Jadoon, Fiaz Gul Khan, Sajid Shah, Ahmad Khan, and Muhammed Elaffendi
[2]Detection of breast cancer through clinical data using supervised and unsupervised feature selection techniques - Amin Ul Haq, Jian Ping Li, Abdus Saboor, Jalaluddin Khan, Samad Wali, Sultan Ahmad, Ghufran Ahmad Khan, and Wang Zhou
[3]Prediction of diabetes empowered with fused machine learning - Usama Ahmed, Shabib Aftab, Muhammad Farhan Khan
[4]Effective feature engineering technique for heart disease prediction with machine learning - Azam Mehmood Qadri, Ali Raza, Kashif Munir, and Mubarak S. Almutairi
[5]Similar disease prediction with heterogeneous disease information networks - Jianliang Gao, Ling Tian, Jianxin Wang, Yibo Chen