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

HEALTHCARE PREDICTIVE ANALYTICS FOR DISEASES DIAGNOSIS

[1]Shivam Kumar Yadav (student), [2]Shreyash Narvekar (student), [3]Brijesh Pal (student), [4]Arhama Ansari(student),
[5]Dr. Hasib Shaikh (Assistant Professor)

Department of AIML, Universal College of Engineering, Mumbai, India.

ABSTRACT:

- Healthcare prediction has been a significant factor in saving lives in recent years. In the domain of health care, there is a rapid development of intelligent systems for analyzing complicated data relationships and transforming them into real information for use in the prediction process. Healthcare predictive analytics has become an indispensable tool in modern medicine, offering a proactive approach to disease detection and management. algorithms for effective prediction of various disease occurrences in disease-frequent societies. Therefore, diseases must be accurately predicted and estimated. Hence, reliable and efficient methods for healthcare predictive analysis are essential. The extensive growth in data for monitoring and analyzing the patients outcomes in predicting and diagnosis of chronic diseases lacks in traditional methods and are replaced by technologies to gather the most relevant insights from the medical data by using predictive analytics with very useful tool of machine learning. Healthcare predictive analytics for disease diagnosis has emerged as a transformative approach in modern medicine, facilitating early detection, accurate prognosis, and personalized treatment strategies.

- **Keywords -** ***Predictive Analytics; Machine Learning; Health Care; Prediction Algorithms and Techniques.***

1. INTRODUCTION

Healthcare predictive analytics is a rapidly evolving field that leverages data science and machine learning techniques to analyze historical and real-time healthcare data to make predictions about future events and trends. By identifying potential health risks, optimizing treatment plans, and improving patient outcomes, predictive analytics is transforming healthcare into a more efficient, effective, and patient-centered system. Machine learning algorithms, such as Random Forest Classifier, Logistic Regression, Decision Trees, and Neural Networks, are essential for building predictive models for disease detection. By analyzing patient data and identifying patterns associated with specific diseases, predictive models can assist in early diagnosis, enabling timely intervention and personalized treatment plans. [1] Healthcare professionals play a pivotal role in diagnosing and treating various medical conditions, whether they be diseases, injuries, or mental and physical impairments in individuals. This group encompasses dentists, physicians, surgeons, and their associated colleagues. Additionally, healthcare extends to fields such as nursing, optometry, physical therapy, pharmacy, athletic training, and more. [2] Health Care informatics, a multi-disciplinary field has become synonymous with the technological advancements and data handling challenges. Medical or Health Informatics is a scientific field that deals with the storage, retrieval, and optimal use of medical information, data, and provides knowledge for problem solving and decision making. Technology in Health has an immense development over the years like advances in information gathering, treatments, communications and research.

2. REQUIREMENT SPECIFICATIONS

1.1 Hardware Requirements:

This subsection will provide the minimum requirements that must be fulfilled by the hardware components. The hardware requirements are as follows: -

- **Computer**: You will need a computer with sufficient processing power to run computer vision and machine learning algorithms. A modern CPU with multiple cores and a dedicated GPU (optional but recommended) can significantly enhance the performance.

- Internet Connection
- RAM – minimum 4 gigabytes
- Storage – minimum 100 gigabytes
- Processor – minimum quadcore or hexacore

1.2 Software Requirements:

This subsection will provide the versions of software applications that must be installed. The software requirements are as follows: -

- **Python**: Ensure you have Python installed on your computer. You can use Python 3.x for this project.
- **Database Management System** (DBMS): Flask-SQL Alchemy to store and manage healthcare data.
- **ETL** (Extract, Transform, Load) Tools: Tools like Apache Nifi, libraries to extract and preprocess data from various sources.
- **Data Cleaning and Wrangling** Tools: Software such as Python's pandas, Open for transformation, and feature engineering.
- **Machine Learning** /Deep learning , using algorithm and librarys
- **More** Libraries: Click, Flask, Gunicorn, Importlib-Metadata, Itsdangerous, Jinja2, Joblib, MarkupSafe, Numpy, Python-Dateutil, Pytz, Scipy, Six, Threadpoolct1, Werkzeug, Zipp ,boostrap etc.

3. PROBLEM STATEMENT

The growing number of deaths around the world from diseases like breast cancer and other cancer-related diseases shows that healthcare needs to be better and timelier. Lack of medical infrastructure and a low ratio of doctors to patients make this problem worse and make it even more important to find and diagnose these diseases early. Machine learning techniques could be used to do predictive analytics on a lot of data in the healthcare industry. This would help doctors and nurses make better decisions about how to treat and care for their patients. But making a solution that is easy for the general public to use remains a challenge.

4. MACHINE LEARNING'S ROLE IN PREDICTIONS

Machine learning task can be broken into below steps.

1) Collecting data: Whether the data is written on paper, recorded in text files and spreadsheets, or stored in an SQL database, the data need to be gathered in an electronic format suitable for analysis. This data will serve as the learning material an algorithm uses to generate actionable information.

2) Exploring and preparing the data: The quality of any machine learning project is based largely on the quality of data it uses. This stage in the machine learning process tends to require a great deal of human intervention. An often-cited statistic suggests that 80 percent of the effort in machine learning is dedicated to data. Much of this time is spent learning more about the data and its nuances throughout a practice called data exploration.

3) Training a model on the data: The specific machine learning task will inform the selection of an appropriate algorithm, and the algorithm will represent the data in the form of a model.

4) Evaluating model performance: It is very important to estimate how well the algorithm learned from its past experience, since each machine learning model results in a biased solution to the learning problem. The accuracy of the model can be evaluated using a test dataset, depending on the type of model used

5) Improving model performance: It is necessary to utilize the advanced strategies to augment the performance of the model, if better performance is needed. Every now and then, it may be required to change to a different type of model overall.

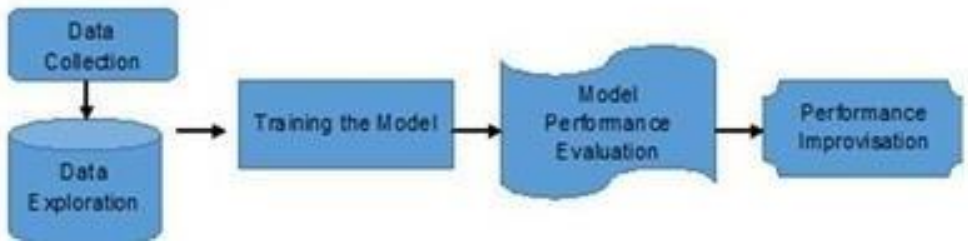


Fig.4.1. Machine Learning Process

5. PROPOSED METHODOLOGY

Predictive analytics in healthcare is a rapidly growing field that utilizes data mining, AI, and machine learning techniques to analyze current and historical healthcare data. The goal is to predict trends and manage the spread of diseases, ultimately improving patient care, reducing costs, and increasing. [3]The proposed methodology for this paper involves the utilization of convolutional neural network and machine learning classification algorithms to predict life-threatening diseases - pneumonia, skin cancer, brain tumors, lung cancer, tuberculosis, and breast cancer. To begin, a vast amount of healthcare data will be collected from various sources such as hospitals, medical centers, and clinics. One common application of predictive analytics in healthcare is disease prediction, where machine learning algorithms are used to identify high-risk individuals or populations based on patient data and past outcomes.[3] Various machine learning classification algorithms such as logistic regression, decision trees, random forests, and support vector machines (SVM) will be explored, and their performance will be evaluated based on metrics such as accuracy, precision, recall, and F1 score.
The working methodology of the system is very simple. The system mainly consist of three modules.

Module 1: Patient module

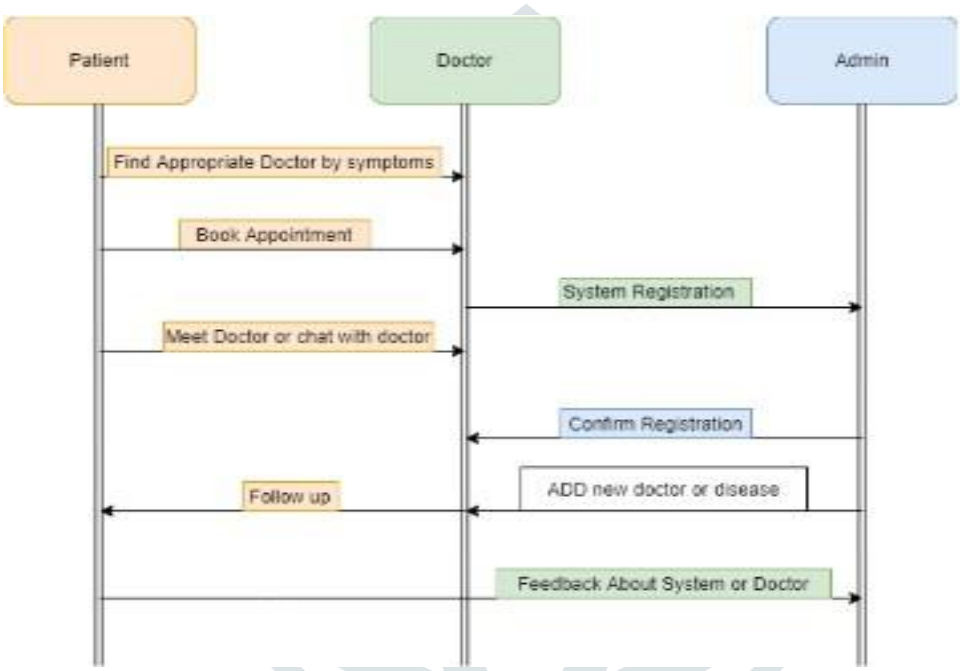


Figure 5.1: sequence diagram of module

The following sequence diagram is the representation of proposed system where we show the interaction between the patient, doctor and Admin. The patient module consists of,

- **Patient Login**: It is authentication with username and password that give access to the system.
- **Registration:** It is a page created for the new user registration who cannot login.
- **Forgot Password**: In this module the user can access his password by enter his username and answering security
- question.
- **Search Doctor**: The user can search the doctor based on Name, Address and Type of diseases treated.
- **Search Diseases**: First the system ask for user Symptom, and when he enters the symptom it asks for a similar symptoms
- based on disease and prediction is done. Based on the disease predicted the doctors available are shown.
- **Feedback**: The user can send Feedback to the System and stored in the database.

6. WORK FLOW CHART

- **Data Collection**: Begin by collecting patient data such as symptoms, medical history, genetic markers, lifestyle choices, And environmental variables.
- **Data Preprocessing**: Clean and preprocess the collected data to ensure it is ready for analysis. This step involves
- handling missing values, encoding categorical variables, and scaling numerical data.
- **Feature Selection**: Implement feature selection techniques like Principal Component Analysis, Local Binary Pattern, Linear
- Discriminant Analysis, and Gabor Filter to identify the most relevant features for disease diagnosis.
- **Model Selection**: Choose appropriate machine learning algorithms like Naive Bayes, Random Forest, Decision Trees, Gradient
- Boosting, Support Vector Machines (SVM), Neural Network, K-Nearest Neighbors, and Logistic Regression for predictive
- analytics1235.

- **Training the Model**: Train the selected machine learning models on the preprocessed data to learn the patterns and relationships
- between the features and the target variable.
- **Testing and Validation**: Evaluate the performance of the trained models using testing data to ensure they can accurately predict
- diseases based on the input symptoms and patient data.
- **Prediction and Diagnosis**: Use the trained models to predict the disease a patient is suffering from based on their symptoms and
- other relevant data.
- **Outcome and Treatment**: Based on the predictions, healthcare professionals can provide personalized treatment plans and
- interventions for patients, improving patient outcomes and overall healthcare quality.

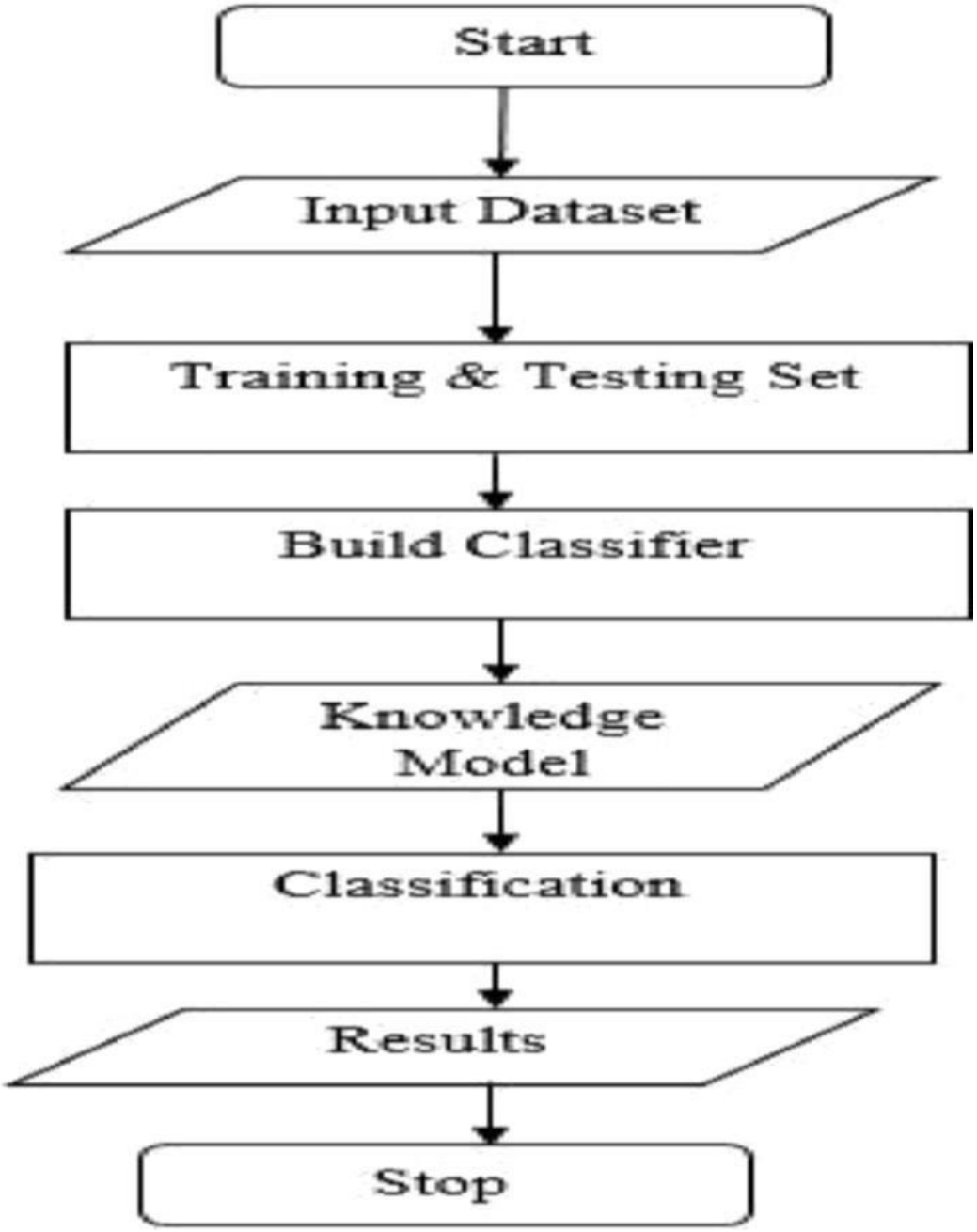


fig.6.1: flowchart of project

7. USE CASE DIAGRAM

Figure 7.1: Use Case diagram of System The use case diagram shows the actors of the system and various tasks performed by them in the system.

3.1. Module : Doctor Module The doctor module is where the doctors are logged in and view the patients consists of many sub modules such as,

- **Doctor Login:** This page is used to login the doctor and give access to the system.
- **My Details:** Here the doctor can update his/her details to the admin and store in system.
- **Notifications:** In this module the doctor gets notified with the symptoms and predicted disease if he can treat the disease.
- **View patient:** In this page the doctor can view the details of patient. Chat module: This is a feature where the doctor and patient can chat after the appointment.
- **View Diseases:** In this module the doctor can view all the details of the diseases and the symptoms.

3.2. Module : Admin Module The admin is the system administrator who has all the access to the modules of the system. The sub modules in the admin module are,

- **Admin Login:** It is used to authenticate the admin to use the system features.
- **Add Doctor:** It can be used to add new doctors to the system.
- **Add Disease:** This can be used to add new diseases based on the doctors request.
- **View Feedback:** This module can be used to see the feedbacks received by the different customers.
- **View Doctors:** This module is used to view the details of the doctor. The other modules are similar to the doctor module.

8. WORKING OF SYSTEM

The working principle is quite simple where all the people are required to login before using the system. The system uses some intelligent data mining techniques for predicting the diseases based on the symptoms. The doctor gets notified whenever the user searches for the disease. Admin has access to all the other modules in the system. The diagrammatic representation of system architecture is given below,

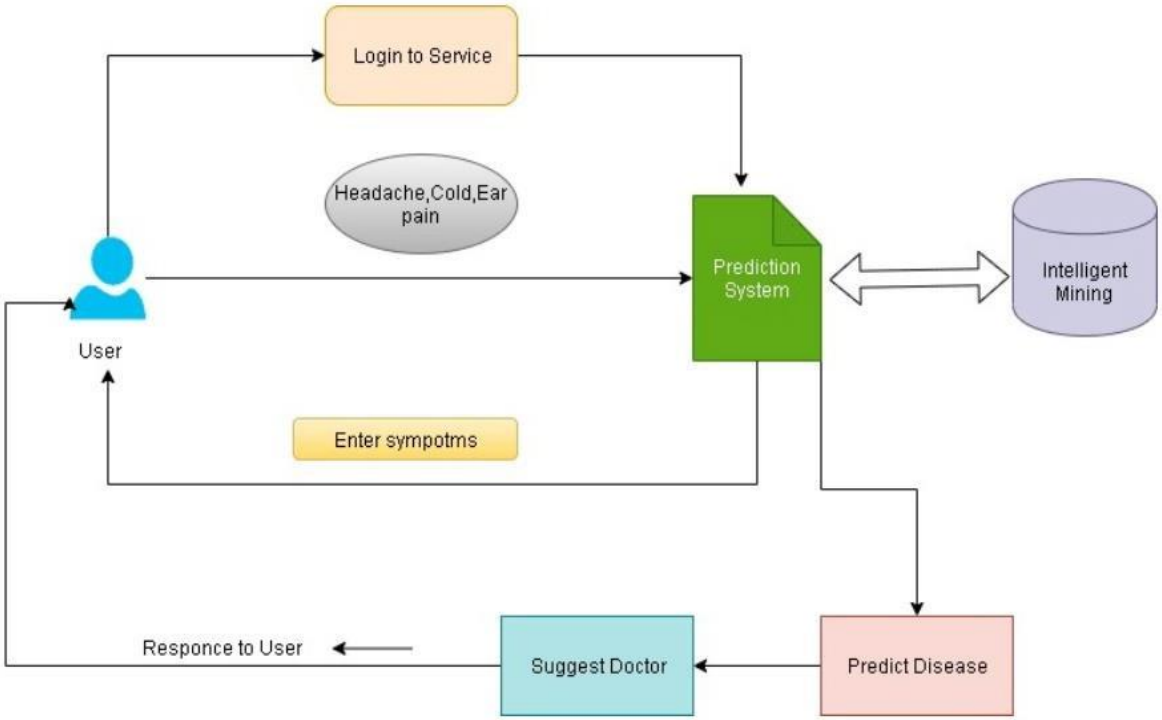
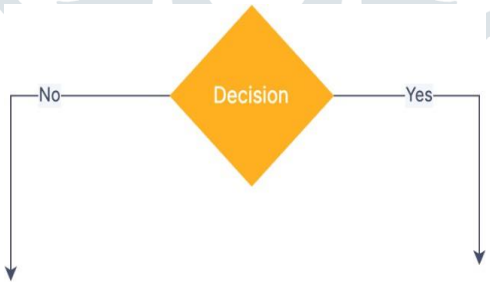


Figure 8.1: System architecture

- **Decision tree**

A Decision tree uses series of decisions that gives many outcomes for combination of decisions. It can be used as decision making tool in many fields. It is similar to binary tree in the form of appearance where there is a root node, leaf node and branches. Root nodes are the starting point of the tree and branches are ending. These decision tree algorithm is combined along with one of the classification techniques to obtain good accuracy and results.



In Classification, information is classified under various marks as indicated by certain parameters given in info and then the names are anticipated for the information. Naïve Bayes is a simple but it is very powerful prediction algorithm. It is simple technique that is used for the classification, model and it can assign class labels to problem instances. It is trained by using the supervised learning. The most useful part is that it requires less number of training data.

$$p(C_k \mid x_1, \ldots, x_n)$$

$$p(C_k \mid \mathbf{x}) = \frac{p(C_k)\, p(\mathbf{x} \mid C_k)}{p(\mathbf{x})}$$

Probability is calculated by using,

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

The decision tree algorithm is used to take the decisions based on the questions asked and the answers obtained from the input. Later these algorithm can be used to fetch the desired result.

9. RESULTS AND DISCUSSION

In the proposed system, we anticipate significant improvements and enhancements compared to the existing system. Users will benefit from a more interactive and feature-rich web application. The Agent, with its advanced capabilities, is expected to provide more accurate and comprehensive answers to user queries.

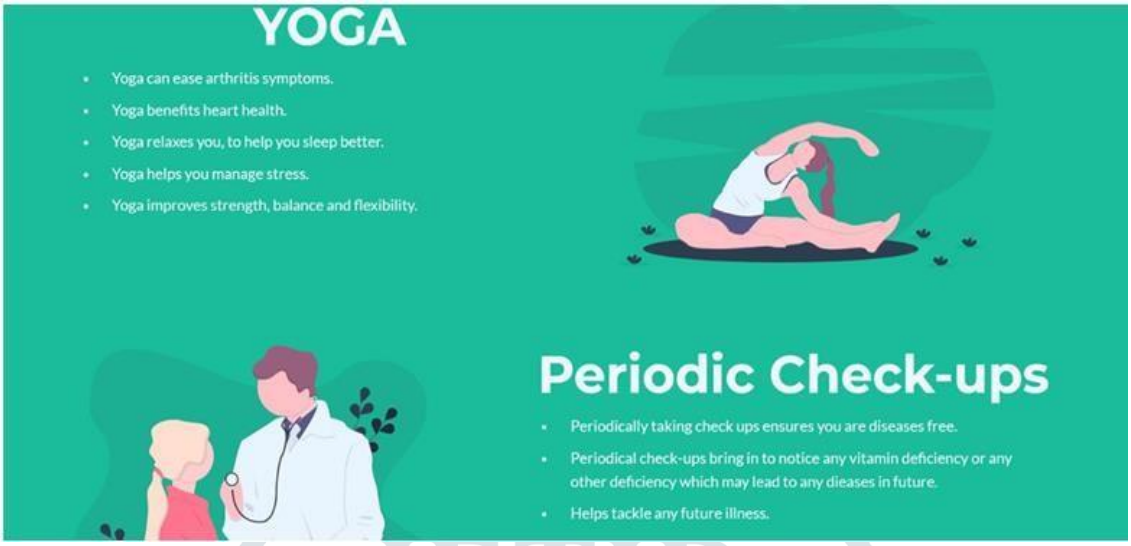
Figure 9.1

Homepage: Find Your Disease

Figure 9.2

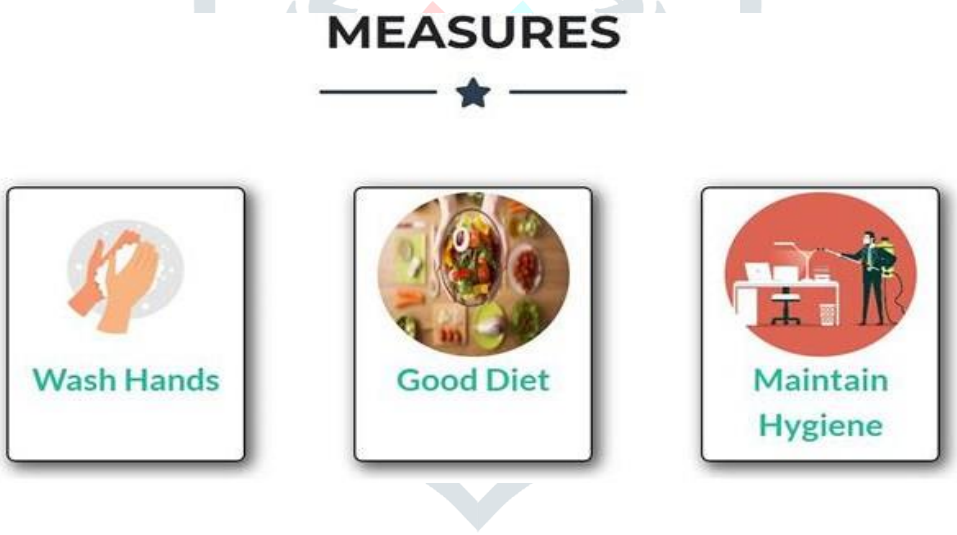
Prevent: An advice and benefits of Workout / Exercise

Figure 9.3



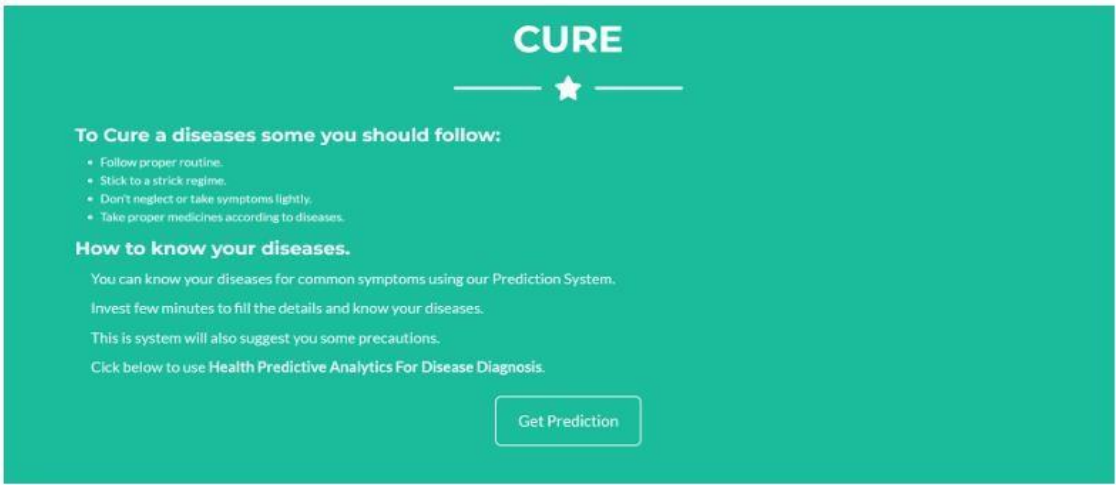
Yoga and Periodic Checkups

Figure 9.4



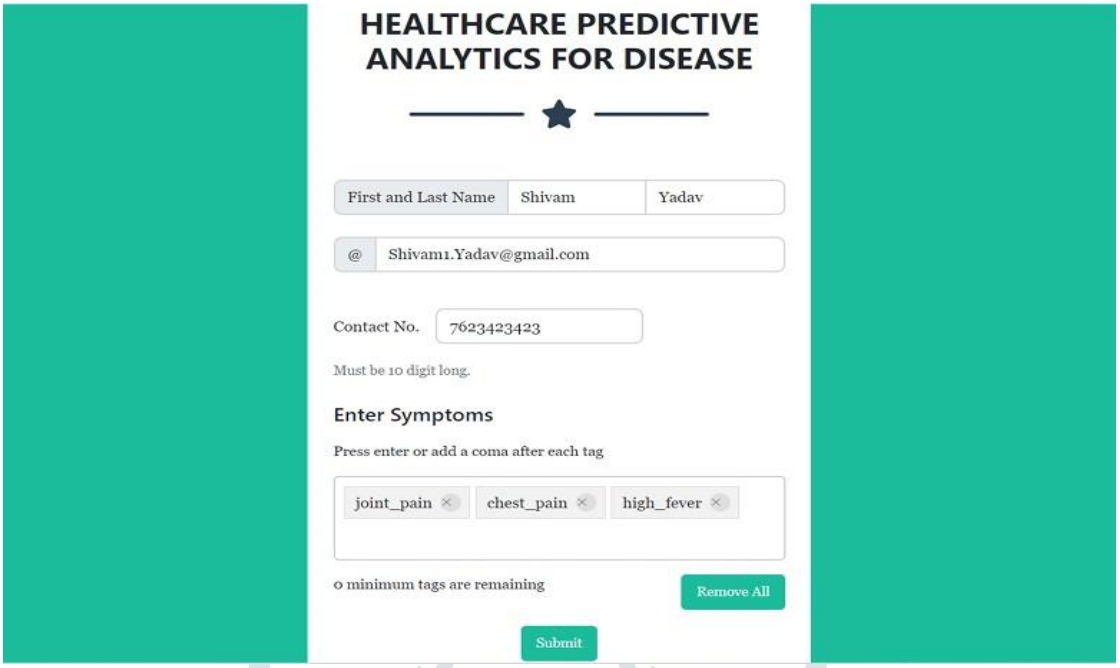
Measures: Wash hand before eating food, keep healthy diet and maintain hygiene

Figure 9.5



Cure : How to know your diseases and to cure a diseases some you should follow

Figure 9.6

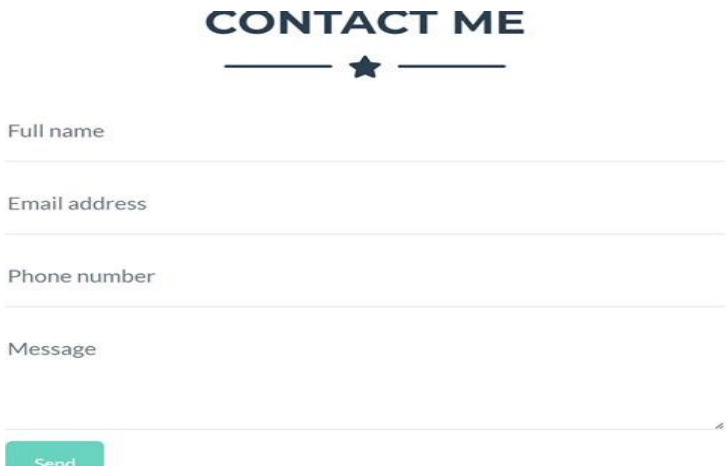


Get Prediction (Form) : Enter your details and Symptoms your suffering

Figure 9.7

Recommendation: Disease you are suffering, Medicines and Precautions

Figure 9.8



Contact: You can send us an message using contact form

Figure 9.9

About Us

6. CONCLUSION

Predictive analytics in healthcare sometimes referred to as just "predictive analytics healthcare" is a process of analyzing historical healthcare data to identify patterns and trends that may be predictive of future events. Predictive analytics in healthcare can be used to predict the likelihood of particular health conditions, clinical decisions, trends, and even spread of disease.[3] In conclusion, this paper aimed to develop machine learning models for predicting diseases - pneumonia, skin
cancer, brain tumors, lung cancer, tuberculosis, and breast cancer. The models were developed using two popular convolutional neural network architectures - Inception v3 and VGG-16 - and were evaluated using accuracy, precision, recall, and F1 score metrics.[2]
that Machine Learning has given medical providers new tools to work with, novel ways to practice medicine. It also confirms that machine learning tools and techniques are decisive in health care province and exclusively used in the diagnosis and predictions of various types of disease. In conclusion, the use of convolutional neural network and machine learning models in early prediction of diseases has the potential to revolutionize the healthcare industry and improve patient outcomes. This paper provides a solid foundation for future research in the development of accurate and reliable predictive models, which can aid in early detection and treatment of diseases. The diagnosis of the disease is necessary as it can only be treated with proper medications but not cured. This paper presents a review of predicting the chronic disease with different feature selection method and classification algorithm using various dataset

7. REFERENCE

[1] Prof. Dr.Khaled Alnowziser "Improving Healthcare Prediction of Diabetic Patients Using KNN Imputed Features and Tri-Ensemble Model" January 2024.
[2] Dr. B. Nithya, Dr. V. Ilango "Predictive Analytics in Health Care Using Machine Learning Tools and Techniques". May-2017.
[3] Dr. Aziz Makandar1, Miss. Nayan Jadhav2 "Classifying and Predictive Analytics for Disease Detection". May-2023.
[4] Mohd Javaid, Abid Haleem, Ravi Pratap Singh, Rajiv Suman, Shanay Rab "Significance of machine learning in healthcare: Features, pillars and applications". June 2022.
[5] Sunday Adeola Ajagbe, Matthew O. Adigun "Deep learning techniques for detection and prediction of pandemic diseases". March 2023.
[6] I. Preethi , Dr.K. Dharmarajan "Diagnosis of chronic disease in a predictive model using machine learning algorithm". December 2023.