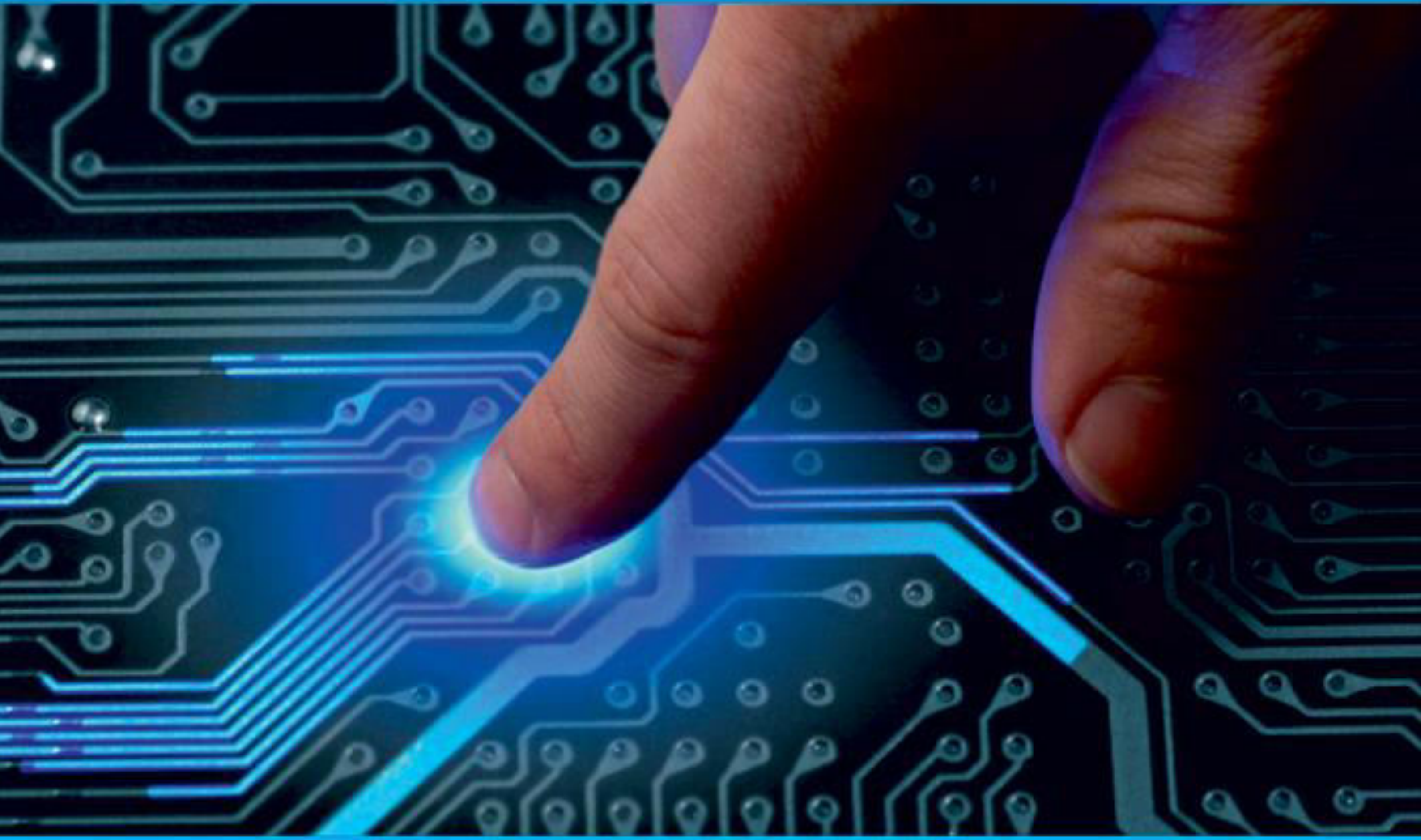
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 12, December 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.625

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Clinical Disease Prediction System Using Machine Learning Techniques

Navitha R M, Pooja A

PG Student, Dept. of C.S.E., Vishweshwaraya Technological University, Impact College of Engineering, Sahakarnagar, Bangalore, India

Assistant Professor, Dept. of C.S.E., Vishweshwaraya Technological University, Impact College of Engineering, Sahakarnagar, Bangalore, India

ABSTRACT: In the rapidly evolving landscape of healthcare, the integration of advanced technologies is paramount for improving diagnostic precision and clinical decision-making. This project introduces a novel Clinical Disease Prediction System (CDPS) that harnesses the power of machine learning for disease prediction. Four robust models—Naive Bayes, Decision Tree, Support Vector Machine, and Random Forest—are employed to analyze extensive datasets. The project's methodology involves meticulous steps, including data collection, library integration, data reading, and exploratory data analysis. Subsequently, the dataset is judiciously split into training and testing sets, and the models are fitted and evaluated using metrics such as accuracy. The outcomes affirm the system's proficiency in predictive analytics. Notably, the integration of Gradio and OpenCV enhances the project's usability and accessibility. Gradio facilitates the development of a user-friendly application, while OpenCV introduces text-to-speech functionality, making the system inclusive and versatile. This confluence of machine learning powers, streamlined user interface, and innovative features positions the CDPS as a potent tool for healthcare professionals.

KEYWORDS: Naive Bayes algorithm; Decision Tree algorithm; Support Vector Machine algorithm; and Random Forest algorithm; Gradio interface; OpenCV

I. INTRODUCTION

In an era defined by rapid technological evolution, healthcare stands at the forefront of transformative innovation. The integration of artificial intelligence (AI) and machine learning (ML) into medical practices has emerged as a beacon of progress, offering unprecedented opportunities to enhance diagnostic accuracy and elevate clinical decision-making. Within this context, this paper introduces a groundbreaking Clinical Decision Prediction System (CDPS) designed to leverage the capabilities of machine learning models for predictive disease analysis. As healthcare professionals grapple with an ever-expanding volume of patient data, the need for sophisticated tools that can decipher complex patterns and aid in decision-making has become increasingly apparent.

The motivation behind this paper stems from the imperative to address the challenges inherent in traditional diagnostic approaches. Conventional methods often face limitations in processing vast datasets and discerning subtle patterns indicative of specific diseases. As a response to these challenges, the CDPS integrates four powerful machine learning models—Naive Bayes, Decision Tree, Support Vector Machine, and Random Forest—each meticulously selected for its unique strengths in predictive analytics. The overarching goal is to empower healthcare professionals with a tool that not only augments their diagnostic capabilities but also streamlines the decision-making process, ultimately leading to improved patient outcomes.

The significance of machine learning in healthcare lies in its ability to derive meaningful insights from vast and intricate datasets, enabling a more nuanced understanding of medical conditions. The selected models operate synergistically, with Naive Bayes offering probabilistic insights, Decision Tree providing transparency in decision paths, Support Vector Machine ensuring precise classification, and Random Forest combining the strengths of multiple models for heightened accuracy. This amalgamation of diverse ML approaches equips the CDPS with a comprehensive analytical toolkit, making it well-suited for the intricacies of disease prediction.

II. RELATED WORK

In [1] Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Delivery: This paper provides a comprehensive survey of the application of machine learning in healthcare, focusing on the evolving landscape of smart healthcare systems. It explores various machine learning models, including Naive Bayes, Random Forest, Decision Tree, and SVM, for diagnosis and disease prediction. In [2] A Survey of Machine Learning in Big Data Analytics for Healthcare: This survey paper discusses the integration of machine learning in big data analytics for healthcare. It reviews the use of machine learning models such as Naive Bayes, Random Forest, Decision Tree, and SVM in healthcare data analysis and prediction. In [3] Machine Learning for Healthcare: This review paper provides an in-depth analysis of machine learning techniques used in healthcare applications. It covers the role of Naive Bayes, Random Forest, Decision Tree, and SVM models in healthcare diagnosis and disease prediction. In [4] Applications of Machine Learning in Healthcare: This paper surveys the various applications of machine learning in healthcare, with a particular emphasis on diagnosis and prediction. It explores the strengths and limitations of Naive Bayes, Random Forest, Decision Tree, and SVM models in the context of smart healthcare systems. In [5] Machine Learning and Data Mining Techniques in Personalized Healthcare: A Review: This review paper focuses on the personalization of healthcare through machine learning and data mining techniques. It discusses the use of Naive Bayes, Random Forest, Decision Tree, and SVM models for tailoring healthcare solutions to individual patients. In [6] A Comprehensive Survey of Machine Learning for Healthcare: This survey paper provides a broad overview of machine learning applications in healthcare. It delves into the utilization of machine learning models, including Naive Bayes, Random Forest, Decision Tree, and SVM, in healthcare diagnosis and prediction. In [7] Machine Learning in Healthcare Informatics: This paper reviews the application of machine learning in healthcare informatics. It investigates the role of machine learning models in symptom diagnosis and disease prediction and discusses their implications for improving healthcare delivery.

III. PROPOSED ALGORITHM

1. Machine Learning Models: Utilizing four different machine learning models - Naive Bayes, Random Forest, Decision Tree, and Support Vector Machine (SVM) - to diagnose symptoms and predict diseases. This multi-model approach aims to enhance the accuracy and robustness of healthcare decision-making. `
2. Gradio for the User Interface: Integrating Gradio as the user interface framework. makes it easier to create interactive interfaces for machine learning models, allowing healthcare professionals or users to input symptoms and receive diagnostic and predictive results in a user-friendly manner.
3. Voice Assistant Powered by OpenAI: Incorporating OpenAI's voice assistant technology to enable voice interactions with the system. This can enhance user accessibility and convenience, allowing users to interact with the healthcare system using natural language commands and queries.

Your proposed method appears to be comprehensive, incorporating both machine learning models for data analysis and prediction and user-friendly interfaces for easy interaction. The integration of voice assistant technology can further improve user engagement and accessibility. It's essential to develop and test this system while considering privacy and security measures to protect patient data. Additionally, ongoing training and updates of the machine learning models would be crucial to ensure the system remains accurate and up-to-date in its diagnostic and predictive capabilities.

IV. PSEUDO CODE

Importing all the necessary modules

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

Reading the data

```
 df_train = pd.read_csv('Training.csv')
df_train.head()
df_train.shape
df_train.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4920 entries, 0 to 4919
Columns: 133 entries, itching to prognosis
dtypes: int64(132), object(1)
memory usage: 5.0+ MB
df_train.describe() # statistical analysis of the data
df_train.isnull().sum().sum() #checking the null values
df_train.dtypes.unique()
df_train.columns
# Looking how much percent each disease having
df_train['prognosis'].value_counts() # as we can see that all the classes are balanced
df_train['prognosis'].value_counts(normalize = True).plot.bar()
plt.subplots_adjust(left = 0.9, right = 2 , top = 2, bottom = 1)
```

Machine Learning Models

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.naive_bayes import MultinomialNB
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
# Splitting the dataset
print ("Naive Bayes")
nb=MultinomialNB()
clf_nb=nb.fit(x_train,y_train)
print ("Acurracy: ", clf_nb.score(x_test,y_test))
y=df_train['prognosis'] #target
x=df_train.drop(['prognosis'],axis=1) #symptoms
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=.33,random_state=42)
Naive Bayes
Acurracy:  1.0

print ("DecisionTree")
dt = DecisionTreeClassifier(min_samples_split=20)
clf_dt=dt.fit(x_train,y_train)
print ("Acurracy: ", clf_dt.score(x_test,y_test))
DecisionTree
Acurracy:  0.9772167487684729

print ("Support Vector Machine")
svm = SVC(kernel='linear', C=1, gamma=0.1)
clf_svm=svm.fit(x_train,y_train)
print ("Acurracy: ", clf_svm.score(x_test,y_test))
Support Vector Machine
Acurracy:  1.0

print ("Random Forest")
rf=RandomForestClassifier(n_estimators=50,n_jobs=5,random_state=33,criterion="entropy")
clf_rf=rf.fit(x_train,y_train)
```

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

```
print ("Accuracy: ", clf_rf.score(x_test,y_test))
Random Forest
Acurracy:  1.0
df_test = pd.read_csv('Testing.csv')
df_test.head()
```

V. SIMULATION RESULTS

The dataset is a balanced dataset i.e. there are exactly 140 samples for each disease, and no further balancing is required. We can notice that our target column i.e. prognosis column is of object data type, this format is not suitable to train a machine learning model. So, we will be using a label encoder to convert the prognosis column to the numerical data type. Label Encoder converts the labels into numerical form by assigning a unique index to the labels. If the total number of labels is n, then the numbers assigned to each label will be between 0 to n-1in Fig.1.

In Fig.2.we can see that the models are performing very well on the unseen data. Now we will be training the models on the whole train data present in the dataset that we downloaded and then test our combined model on test data present in the dataset.

Creating a function that can take symptoms as input and generate predictions for disease in Fig.3.

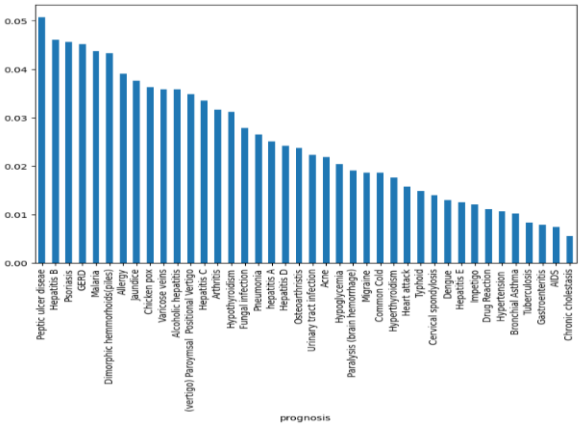
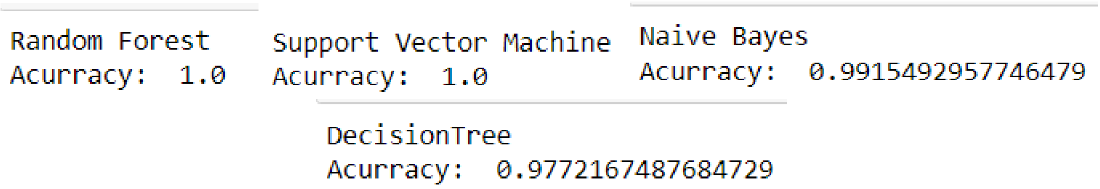


Fig.1: Accuracy of different models



```
Random Forest          Support Vector Machine     Naive Bayes
Acurracy:  1.0         Acurracy:  1.0             Acurracy:  0.9915492957746479

                 DecisionTree
                 Acurracy:  0.9772167487684729
```

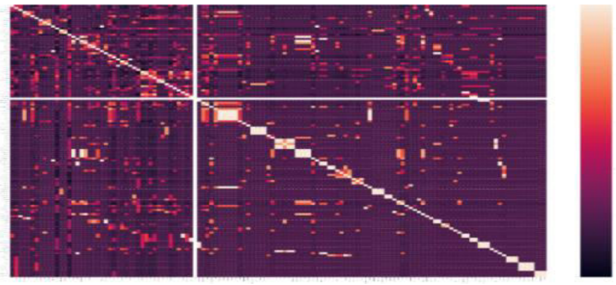
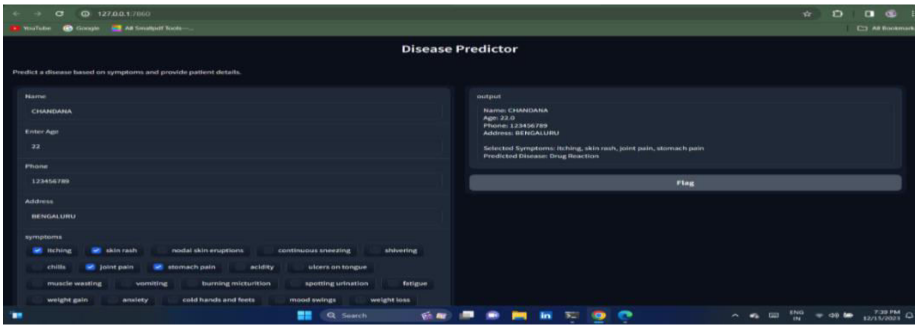


Fig.2: Confusion Matrix of the proposed model



Disease Predictor

Predict a disease based on symptoms and provide patient details.

Name
CHANDANA

Enter Age
22

Phone
123456789

Address
BENGALURU

symptoms
☑ itching ☑ skin rash □ nodal skin eruptions □ continuous sneezing □ shivering
□ chills ☑ joint pain ☑ stomach pain □ acidity □ ulcers on tongue
□ muscle wasting □ vomiting □ burning micturition □ spotting urination □ fatigue
□ weight gain □ anxiety □ cold hands and feets □ mood swings □ weight loss

output
Name: CHANDANA
Age: 22.0
Phone: 123456789
Address: BENGALURU

Selected Symptoms: itching, skin rash, joint pain, stomach pain
Predicted Disease: Drug Reaction

7:39 PM
12/15/2023

Fig.3: Disease Predicted basis of symptoms

VI. CONCLUSION AND FUTURE WORK

This paper not only demonstrates the effectiveness of machine learning in healthcare but also underscores the importance of user-centric design in practical applications. The CDPS holds promise as a valuable contribution to the healthcare technology landscape, offering a reliable and accessible platform for informed clinical decision-making. The system design and analysis of the CDPS paper reflect a holistic approach that integrates advanced machine learning, user-centric design with Gradio, and innovative features with OpenCV. The resulting system not only showcases analytical prowess but also prioritizes user accessibility and inclusivity, marking a paradigm shift in the intersection of healthcare and technology. The outcomes of the CDPS paper underscore its effectiveness in predictive analytics, user interface development, and the integration of innovative features. The project not only showcases technological advancements in healthcare but also holds promise for practical and impactful applications in clinical decision support.

REFERENCES

1. Belard, Arnaud, Timothy Buchman, Jonathan Forsberg, Benjamin K. Potter, Christopher J. Dente, Allan Kirk, and Eric Elster. "Precision diagnosis: a view of the clinical decision support systems (CDSS) landscape through the lens of critical care." Journal of clinical monitoring and computing 31 (2017): 261-271.
2. Zikos, Dimitrios, and Nailya DeLellis. "CDSS-RM: a clinical decision support system reference model." BMC medical research methodology 18, no. 1 (2018): 1-14.
3. Davenport, T. H., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. Future Healthcare Journal, 6(2), 94-98.
4. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. The New England Journal of Medicine, 380(14), 1347-1358.
5. Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. The New England Journal of Medicine, 375(13), 1216-1219.
6. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24-29.
7. Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22), 2402-2410.
8. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453.
9. Goldhahn, J., Rampton, V. B., & Spinas, G. A. (2018). Could artificial intelligence make doctors obsolete? BMJ, 363, k4563.
10. Razzak, M. I., Naz, S., & Zaib, A. (2018). Deep learning for medical image processing: Overview, challenges and the future. In Classification in BioApps (pp. 323-350). Springer.
11. Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., Kitai, T., & Artificial Intelligence in Cardiology: How to Use Artificial Intelligence to Predict Individual Patients With Cardiac Diseases. European Journal of Preventive Cardiology, 27(5), 520-522.
12. Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain? arXiv preprint arXiv:1712.09923.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING