A COMPARATIVE ANALYSIS OF ALGORITHMS AND SVM BASED IMPLEMENTATION FOR ENHANCED MEDICINE RECOMMENDATION SYSTEM USING MACHINE LEARNING

V Karthick, Associate Professor
Department of CSE Rajalakshmi Engineering College
Chennai 602105,
vkarthick86@gmail.co

Vinisha S , UG Student
Department of CSE
Rajalakshmi Engineering College
Chennai,
210701310@rajalakshmi.edu.in

Yar Zar Min, UG Student
Department of CSE Rajalakshmi Engineering College Chennai,
210701318@rajalakshmi.edu.in

Healthcare is a cornerstone of the medical field, yet medication errors pose significant threats to safety. To address this, we highlight the need for medicine recommendation systems. Our proposed platform helps patients understand symptoms, predict diseases, and receive medication, precaution, and dietary recommendations.We use a supervised learning approach with classifiers such as Support Vector Machines (SVM), Random Forest, Gradient Boosting, K-Nearest Neighbors, and Multinomial Naive Bayes to predict diseases and prescribe medications. Evaluations confirm these classifiers' accuracy. Implementing the SVM model allows patients to input symptoms and receive disease predictions, aiding both patients and healthcare professionals in diagnosis and treatment decisions.This project demonstrates the efficacy of these classifiers, ensuring reliable and precise results.

ABSTRACT

Keywords: Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors, Gradient Boosting, Supervised Machine Learning, Multinomial Naive Bayes, medicine recommendations.

INTRODUCTION

The advent of machine learning, a subset of artificial intelligence, has revolutionized healthcare by providing tools that train computers to emulate human learning and thinking processes. This technology is now widely used to collect and manage patient data, identify healthcare trends, and recommend treatments, medications, and diets. In this project, we present a medical recommendation system utilizing supervised machine learning algorithms, including Support Vector Machines (SVM), Random Forest, Gradient Boosting, K-Nearest Neighbors, and Multinomial Naive Bayes. These algorithms are trained to predict diseases based on symptoms and recommend appropriate medications, precautions, and dietary guidelines. The system, particularly leveraging SVM for its robust classification and regression capabilities, aims to assist both patients and healthcare professionals in making informed decisions, demonstrating significant accuracy and reliability in medical predictions.

LITERATURE REVIEW

S.NO	YEAR	TITLE	AUTHOR'S NAME	DESCRIPTION
1	2021	Computer based disease recommendation system using machine learning	Gupta J P,Singh A	This project involves predicting diseases and recommending medicines, especially for COVID-19, using a symptom-based recommendation system. It employs decision tree, random forest, and Naive Bayes classifiers, analyzing consistent predictions across these algorithms.
2	2023	"A NOVEL INTEGRATED ML WITH NLP FRAMEWORK FOR DRUG RECOMMENDATION SYSTEM."	HARSHAVARDHINI	**This project proposes a sentiment and machine learning-based drug recommendation system that suggests drugs based on patient-provided disease names, incorporating user review sentiments to enhance trust and ensure effective treatment outcomes.**
3	2022	"Diabetes medication recommendation system using patient similarity analytics."	Tan, Wei Ying, et al.	This study presents a diabetes medication recommendation system using patient similarity analytics to provide individualized treatment recommendations, achieving high accuracy in matching actual prescriptions.

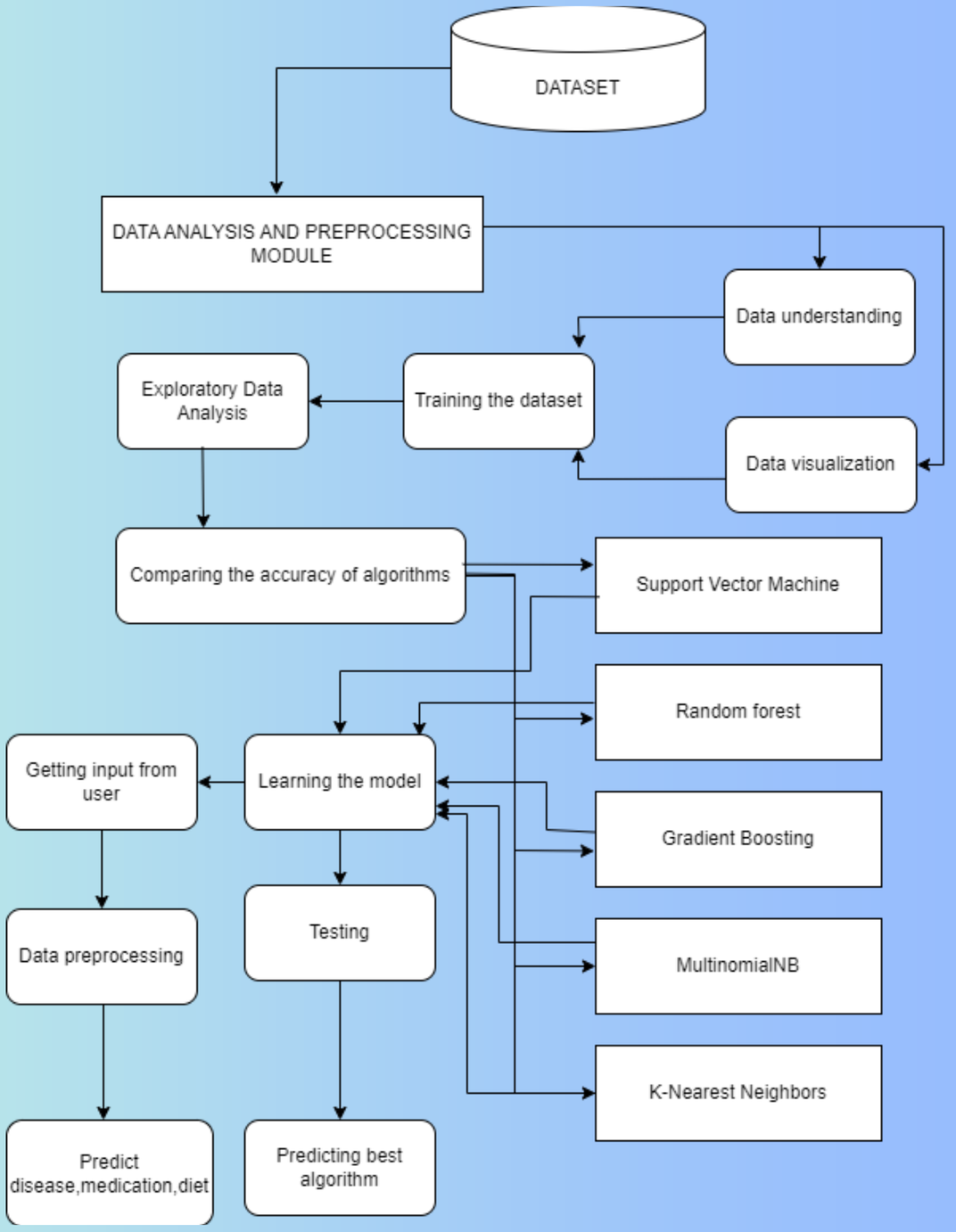
4	2021	"Personalizing medication recommendation with a graph-based approach."	Bhoi, Suman, et al.	This study introduces PREMIER, a personalized medication recommender system leveraging EHR data and drug interaction modeling to enhance prescription accuracy and provide clinically aligned justifications.
5	2024	"REFINE: A Fine-Grained Medication Recommendation System Using Deep Learning and Personalized Drug Interaction Modeling."	Bhoi, Suman, et al.	This study introduces REFINE, a deep learning-based medication recommendation system that enhances personalized treatment by modeling patient data and drug interaction severity, outperforming current techniques.
6	2021	"Machine learning (ML) in medicine: Review, applications, and challenges."	Rahmani, Amir Masoud	This paper provides a comprehensive review and classification of machine learning schemes in healthcare, aiding researchers in understanding current applications, challenges, and future directions.
7	2018	"The use of machine learning algorithms in recommender systems: A systematic review."	Portugal, Ivens, Paulo	This paper systematically reviews the use of machine learning algorithms in recommender systems, identifying trends, challenges, and research opportunities to guide future development.

8	2014	"Survey of faculty perceptions regarding a peer review system."	Gupta J P,Singh A	This study assesses radiologists' views on a mandatory peer review system, finding that many see it as a bureaucratic task rather than a valuable tool for improving medical care.
9	2021	"Artificial intelligence and mapping a new direction in laboratory medicine: a review."	Herman, Daniel S., et al.	This review explores the current and future applications of AI and ML in laboratory medicine, highlighting advancements, challenges, and the need for improved infrastructure and community participation for broader implementation.
10	2018	"MamaBot: a System based on ML and NLP for supporting Women and Families during Pregnancy."	Vaira, Lucia, et al.	This paper explores the use of AI-based chatbots, employing machine learning and natural language processing, to support pregnant women, mothers, and families with young children by providing assistance and instructions.

RESEARCH GAP

- **The current system relies on specific datasets which may not fully represent the wide variety of patient demographics, medical histories, and disease variations encountered in real-world settings. Incorporating more diverse and comprehensive datasets would enhance the system's robustness and applicability across different populations.**
- **The existing model may not be capable of processing real-time data inputs, which is crucial for providing timely and up-to-date recommendations. Future research should focus on integrating real-time data processing capabilities to ensure that recommendations are based on the latest available information.**
- **Seamlessly integrating the recommendation system with existing electronic health records (EHRs) and other healthcare IT systems poses a significant challenge.**

PROPOSED METHODOLOGY



1. DATA COLLECTION

	Disease	Symptom_1	Symptom_2	Symptom_3	Symptom_4
0	Fungal infection	itching	skin_rash	nodal_skin_eruptions	dischromic _patches
1	Fungal infection	skin_rash	nodal_skin_eruptions	dischromic _patches	
2	Fungal infection	itching	nodal_skin_eruptions	dischromic _patches	
3	Fungal infection	itching	skin_rash	dischromic _patches	
4	Fungal infection	itching	skin_rash	nodal_skin_eruptions	
5	Fungal infection	skin_rash	nodal_skin_eruptions	dischromic _patches	
6	Fungal infection	itching	nodal_skin_eruptions	dischromic _patches	
7	Fungal infection	itching	skin_rash	dischromic _patches	
8	Fungal infection	itching	skin_rash	nodal_skin_eruptions	
9	Fungal infection	itching	skin_rash	nodal_skin_eruptions	dischromic _patches
10	Allergy	continuous_sneezing	shivering	chills	watering_from_eyes
11	Allergy	shivering	chills	watering_from_eyes	
12	Allergy	continuous_sneezing	chills	watering_from_eyes	
13	Allergy	continuous_sneezing	shivering	watering_from_eyes	
14	Allergy	continuous_sneezing	shivering	chills	
15	Allergy	shivering	chills	watering_from_eyes	

2. PREPROCESSING

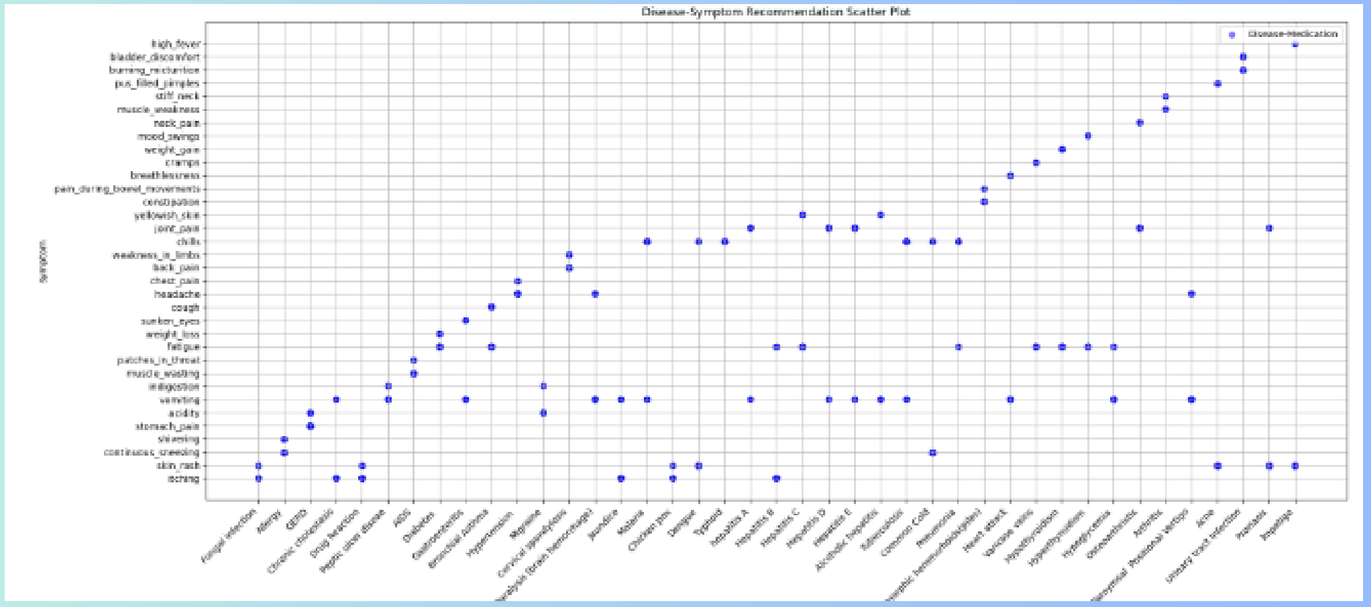
```
dataset['prognosis'].unique()
```

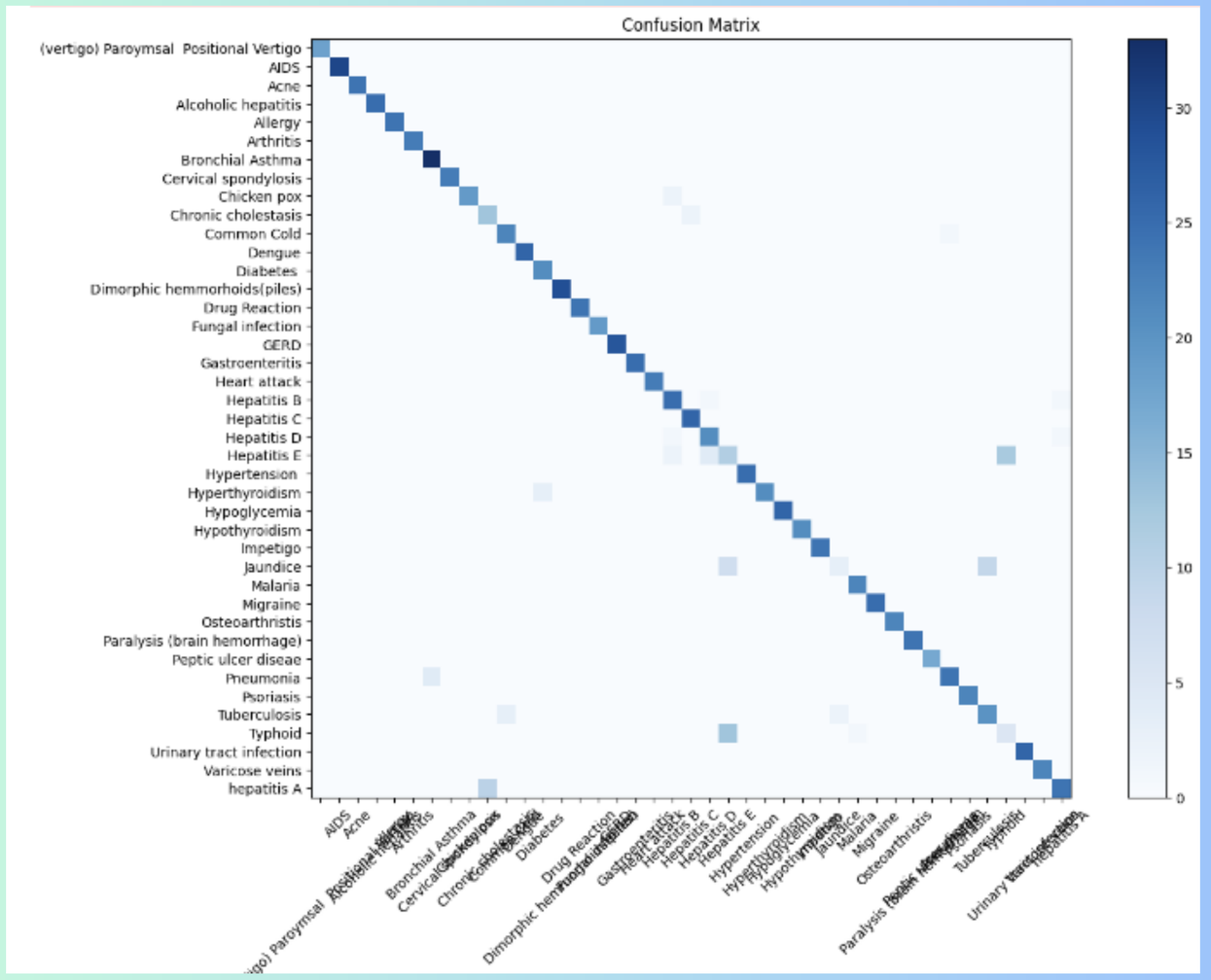
```
array(['Fungal infection', 'Allergy', 'GERD', 'Chronic cholestasis',
       'Drug Reaction', 'Peptic ulcer diseae', 'AIDS', 'Diabetes ',
       'Gastroenteritis', 'Bronchial Asthma', 'Hypertension ', 'Migraine',
       'Cervical spondylosis', 'Paralysis (brain hemorrhage)', 'Jaundice',
       'Malaria', 'Chicken pox', 'Dengue', 'Typhoid', 'hepatitis A',
       'Hepatitis B', 'Hepatitis C', 'Hepatitis D', 'Hepatitis E',
       'Alcoholic hepatitis', 'Tuberculosis', 'Common Cold', 'Pneumonia',
       'Dimorphic hemmorhoids(piles)', 'Heart attack', 'Varicose veins',
       'Hypothyroidism', 'Hyperthyroidism', 'Hypoglycemia',
       'Osteoarthristis', 'Arthritis',
       '(vertigo) Paroymsal  Positional Vertigo', 'Acne',
       'Urinary tract infection', 'Psoriasis', 'Impetigo'], dtype=object)
```

DESCRIPTION OF SYMPTOMS

nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain	acidity	ulcers_on_tongue	...	blackh
1	0	0	0	0	0	0	0	...	
1	0	0	0	0	0	0	0	...	
1	0	0	0	0	0	0	0	...	
0	0	0	0	0	0	0	0	...	
1	0	0	0	0	0	0	0	...	

DESCRIPTION-SYMPTOMS SCATTER PLOT





CONFUSION MATRIX FOR VARIOUS DISEASE

ALGORITHM USED

SUPPORT VECTOR MACHINE

It is a supervised machine learning algorithm that classified the data by finding optimal line or hyperplane ny maximizing the distance between each class in N dimensional space.It is used for both classification and regression problems.It shows 87% accuracy here with the confusion matrix of [[85 8][18 89]].

RANDOM FOREST

Random forests also called random decision forests refers to an ensemble learning method for classification, regression which is operated by constructing a multitude of decision trees at training time.It shows 90% accuracy here with the confusion matrix of [[88 5][15 92]].

GRADIENT BOOSTING

Gradient Boosting is a boosting algorithm that combines several weak learners into strong learners, in which each new model is trained to minimize the loss function such as mean squared error or cross-entropy of the previous model using gradient descent

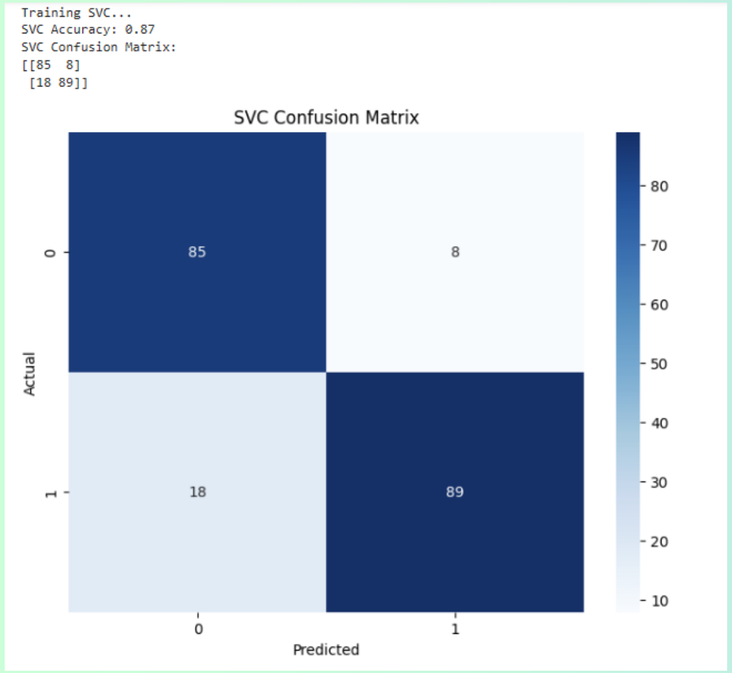
K-NEAREST NEIGHBORS

The K-Nearest Neighbors (KNN) algorithm is a popular machine learning technique used for classification and
regression tasks. It relies on the idea that similar data points tend to have similar labels or values. During the training phase, the KNN algorithm stores the entire training dataset as a reference.

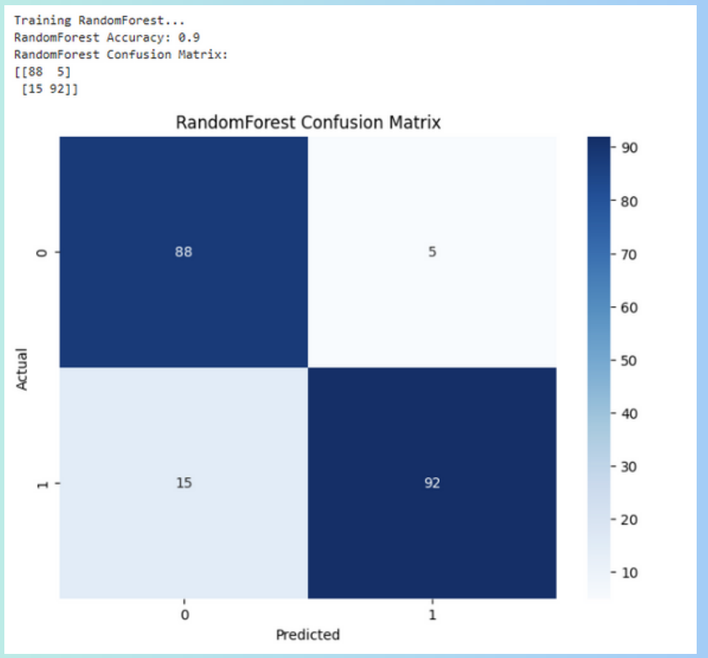
MULTINOMIALNB

Naive Bayes classifier for multinomial models. The multinomial Naive Bayes classifier is suitable for classification with discrete features .It shows 91% accuracy here with the confusion matrix of [[88 5][38 69]].

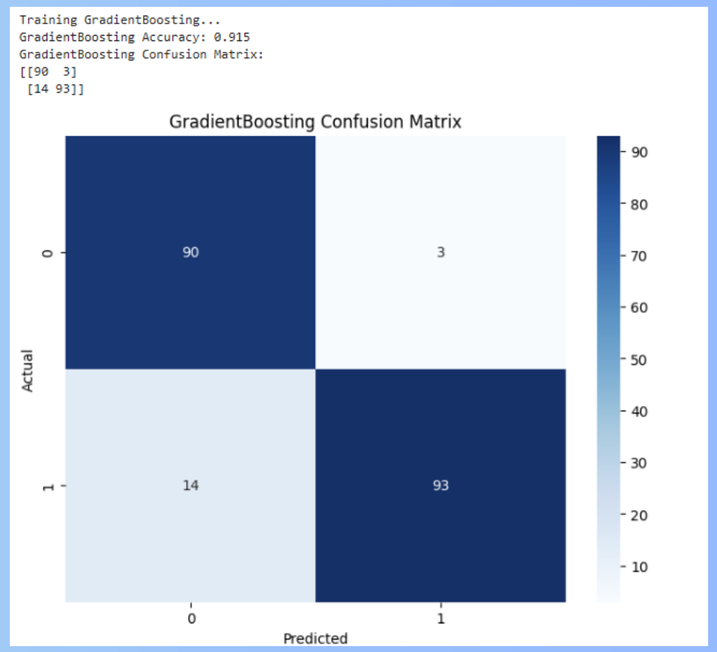
DISCUSSION AND RESULT



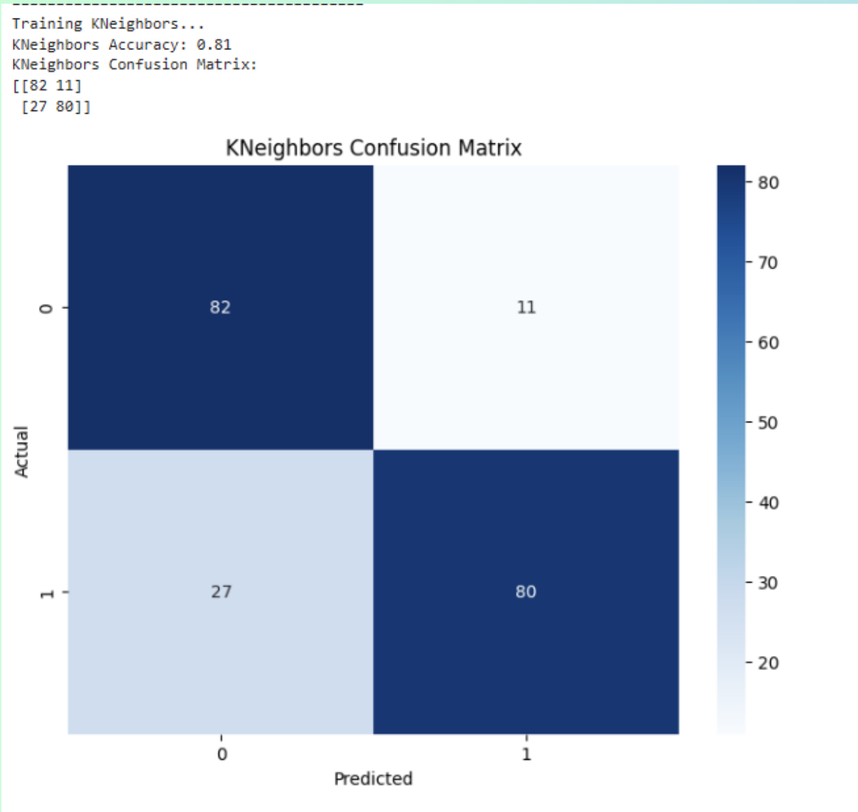
Accuracy and confusion matrix of SVM



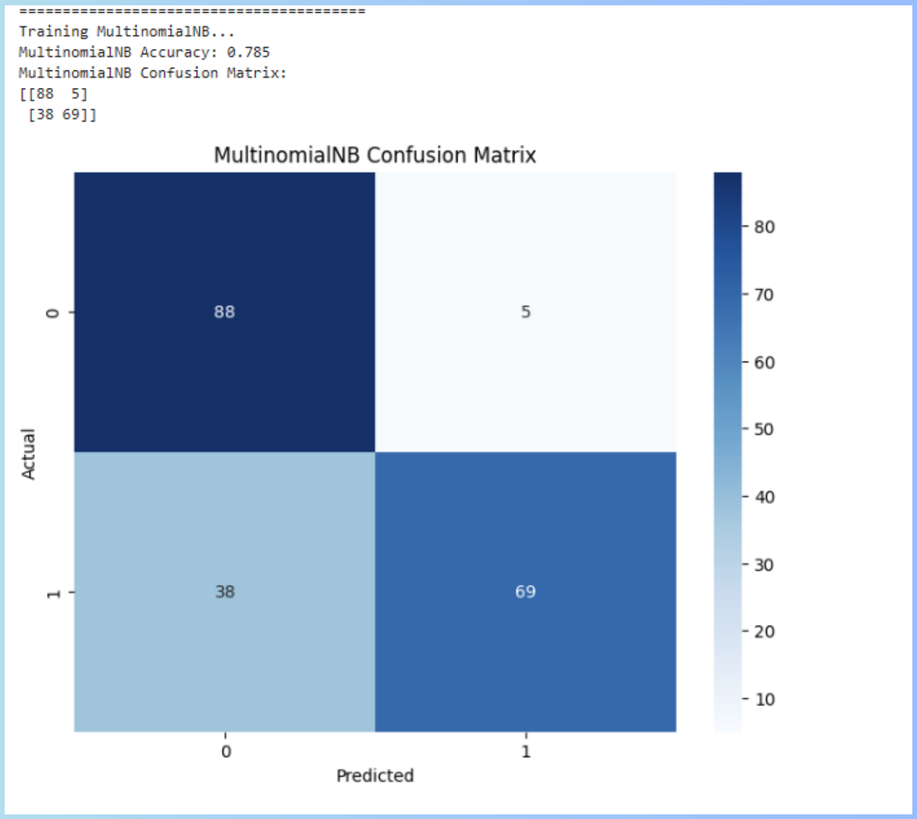
Accuracy and confusion matrix of Random forest



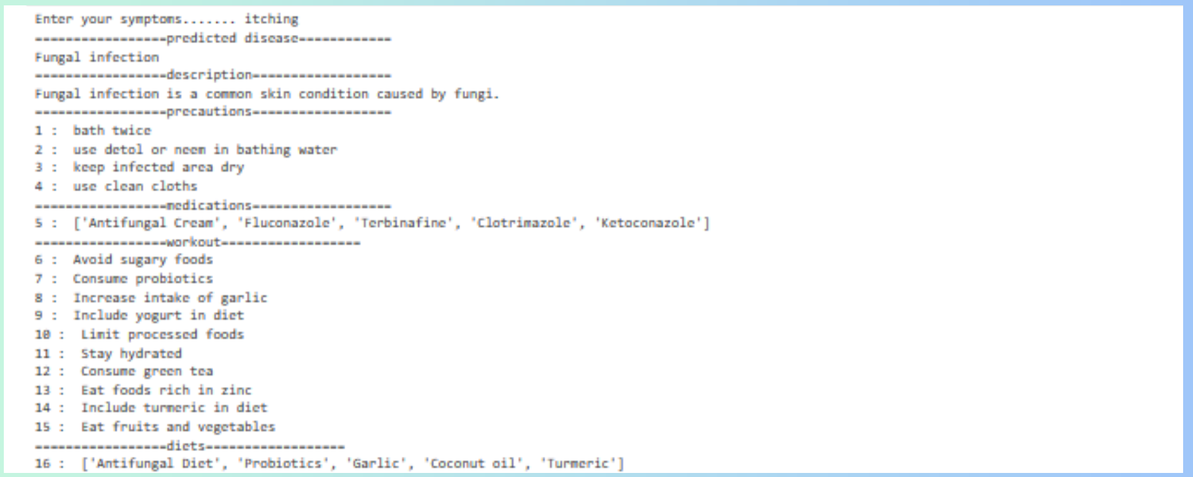
Accuracy and confusion matrix of Gradient boosting



Accuracy and confusion matrix of KNeighbors



Accuracy and confusion matrix of MultinomialNB



OUTPUT SCREENSHOT OF INPUT SYMPTOMS

Algorithm used	Accuracy
Support Vector Machine	87%
Random Forest Classifier	90%
Gradient Boosting	91.5%
KNeighbors	81%
MultinomialNB	78.5%

Comparing the accuracy of different algorithms

CONCLUSION AND FUTURE WORK

This project presents a comprehensive medicine recommendation system leveraging supervised machine learning algorithms to predict diseases based on patient symptoms and provide tailored medication recommendations, along with necessary precautions and dietary advice. By employing algorithms such as Support Vector Machine, Random Forest, Gradient Boosting, K-Nearest Neighbors, and Multinomial Naive Bayes, the system ensures high accuracy in disease prediction and medication recommendations. The effectiveness of these classifiers in accurately predicting results demonstrates the system's potential to aid healthcare professionals in diagnosing and treating patients more efficiently.

Future enhancements could include integrating more diverse datasets to improve the robustness of the predictions, incorporating real-time data processing for up-to-date recommendations, and expanding the system's capabilities to include a broader range of diseases and treatment options. Additionally, developing a user-friendly interface and ensuring seamless integration with existing healthcare systems will enhance the system's usability and adoption in clinical settings. By continually refining the algorithms and incorporating user feedback, this medicine recommendation system aims to become an invaluable tool in the healthcare industry, ultimately improving patient outcomes and reducing medication errors.

REFERENCES

[1] Gupta JP, Singh A, Kumar RK. A computer-based disease prediction and medicine recommendation system using machine learning approach. Int J Adv Res Eng Technol (IJARET). 2021;12: 673–683.

[2] Abbas K, Afaq M, Ahmed Khan T, Song W-C. A Blockchain and Machine Learning-Based Drug Supply Chain Management and Recommendation System for Smart Pharmaceutical Industry. Electronics. 2020;9: 852.

[3] Portugal, Ivens, Paulo Alencar, and Donald Cowan. "The use of machine learning algorithms in recommender systems: A systematic review." Expert Systems with Applications 97 (2018): 205-227.

[4] Nayak SK, Garanayak M, Swain SK, Panda SK, Godavarthi D. An intelligent disease prediction and drug recommendation prototype by using multiple approaches of machine learning algorithms. IEEE Access. 2023; 1–1.

[5] Silpa C, Sravani B, Vinay D, Mounika C, Poorvitha K. Drug recommendation system in medical emergencies using machine learning. 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA). IEEE; 2023. doi:10.1109/icidca56705.2023.10099607

[6] Mathur, Utkarsh. A Content Based Recommender System for Medicine using Machine Learning Algorithm. Diss. Dublin, National College of Ireland, 2022.

[7] Adegun, Adekanmi Adeyinka, et al. "CAD-based machine learning project for reducing human-factor-related errors in medical image analysis." Handbook of research on the role of human factors in IT project management. IGI Global, 2020. 164-172

[8] Galphat, Yugchhaya, et al. "Disease Prediction System using Machine Learning." 2023 2nd Edition of IEEE Delhi Section Flagship Conference (DELCON). IEEE, 2023.

[9] Iwendi, Celestine, et al. "Realizing an efficient IoMT-assisted patient diet recommendation system through machine learning model." IEEE access 8 (2020): 28462-28474.