

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, India
vkarthick86@gmail.com

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

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

ABSTRACT -Healthcare plays an important role in our life it stands as a [1] centerstone of medical field ,nowadays people are seeking their medical related information in online .Presently various types of medical information are available across various websites which possesses challenges for users to get their information for their well being.Even medication errors which represent threat to human safety.By overiewing these challenges we underscore the essentiality of deploying a [2]medicine recommendation systems in health care in order to help patient to give accurate medications and precautions based on the symptoms. It provides a user with a platform designed to assist the patient with their symptoms and predict the disease based on the symptoms and gives the overview of the medication to be followed ,precautions and diets .It provides a powerful tool to assist users in understanding and managing the health.With the availability of medical information from the dataset accurate result pattern can be recognized.It also provides the tailored recommendation of top 5 medicines,prescriptions details based on input ensuring reliable and precise result.In Order to get this result we employ a supervised learning approach where the labeled data is used in order to predict the disease with prescription.We consider a multifaceted set of classifier which involves Support vector

machine,Randomforest,GradientBoosting,K-Nearest Neighbors and Multinomial Naive Bayes to predict how efficiently the model works.This project result demonstrated the efficacy of SVM classifiers in accurately predicting the result . Finally, we implement the selected SVM model in a practical setting, allowing patients to input their symptoms, and the system predicts the most likely disease. This system serves as a valuable tool for healthcare professionals in aiding diagnosis and treatment decisions.

Keywords - *Support Vector Machine (SVM),Random Forest,,K-Nearest Neighbors,Gradient Boosting,Supervised machine learning,MultinomialNB.*

I. INTRODUCTION

The advent of Machine learning comes under the category of AI.Machine learning act as a tool that are used in healthcare and medical professionals which involves training the computer to emulate how people learn and think.This can be used in health care to collect and manage patient data ,identify the trend in health care and recommend medicine ,treatment,diet etc.Nowadays hospitals and healthcare related fields began to use machine learning to improve the

resolution and reduce the risk in medical field which has led to several new career and job opportunities. In this project we present a Medical recommendation system that harness trained data set ,this is implemented using the supervised machine learning algorithm , Support vector machine algorithm is used here to predict the disease with the given symptoms and provide with the disease , their description,diet to be followed and precautions.Support vector machine is a machine learning algorithm which is used for classification and regression tasks that are used to analyze and find the pattern of the data used for classification and regression analysis.It is a powerful algorithm that are commonly used to predict the result from the dataset.In regression task SVM constructs a hyperplane and predict the target and minimize the error.

Jay Prakash Gupta [3] proposed a project that involves predicting the disease and medicine especially for COVID affected people which was first found in China and became a pandemic. The medicines were not available for them to cure the disease .In this situation the doctors and specialists were manually recommending the medicine based on the symptoms. This causes huge loss of life because of lack of medicine availability. So they implemented a disease recommendation system based on symptom and also they came with an idea of developing the medicine based on the symptoms of disease.He implemented using three mining algorithm which includes decision tree classifier,Random forest classifier and Naive bayes classifier and assigned different prediction level based on the result obtained from more than one classifier and he analyzed that some disease predicted by all the three classifier are same.

Kizar Abbas[4] proposed a blockchain based drug supply chain management and recommendation system using deep learning models which includes natural language processing,N gram for predicting the number of words in a sequence,LightGMB for sentiment analysis.

The work by Ivens Portugal [5] demonstrates a

recommendation system to provide users with product and service recommendation.He made this study to identify the use of machine learning algorithm in recommendation system to identify research information.His characterization adopted new machine learning algorithms in recommendation system and used big data technologies

Suvendu Kumar Nayak [6] in his research developed a disease prediction and drug recommendation prototype by using multiple approaches for machine learning algorithms and predicted the disease based on the symptoms and drug recommendation provided.

[7]C.Shilpa proposed a online recommender system that are being used in hospitals,medical professionals and drugs.This medical suggestion can be used during pandemic floods and cyclone hit.She used the decision tree to predict the accurate result ,In time of emergency this system provides safe medication

Mathur [8] research shows a content based recommender system for medicine using machine learning algorithms .This research is mainly focused on machine learning models as well as examining to determine whether the content based recommender system approach is utilized to build drug recommendation models.

Adegun[9] made a research on CAD Based machine learning project for reducing Human-Factor-Related error in medical image analysis.He used machine learning techniques such as deep learning methods.This provides a user friendly website for detecting the error in medical image.

Alghamdi [10] researches about predicting diabetes mellitus using SMOIT and ensemble machine learning approach. He used 5 different models for predicting the diabetes such as decision tree,Naive bayes,Logistic regression,Logistic model tree and random forest.This study shows the potential of ensembling and SMOTE approaches for predicting diabetes.

This research suggest that [17] recommendation system suggest the right item, it suggest items to the user based on the users rating suggested by the recommendation system .It provides a recommendation for books ,movies,paper etc.This paper is about suggesting the medicine to the users according to the disease using “recommenderlab”.It gives a good medicine according to the user rating available in database.

Medical fields have been undergoing many automation during the past decades .Even though the automation increases, the application faces[18] lots of problems in efficiency and accuracy in healthcare.In order to address this issue this study show developed a health care system where the individuals can first visit the doctor at the initial state and after diagnosis recommendation is done by the health care system.

In this work we are trying to implement a Medicine recommendation system where the disease is predicted based on the input that are provided by the used i.e symptoms , based on the symptoms and disease the top 5 medications to cure the particular disease is predicted and their precautions,diet to be followed and workouts and predicted.We are implementing this by using Support machine vector model and predicted based on the symptoms.Also the dataset are trained by using 5 different algorithm which includes Support vector machine,Randomforest,Gradient Boosting,K-Nearest Neighbors and Multinomial Naive inorder to get the accurate result Each model is trained using the same dataset inorder to compare the accuracy of the result provided by the model.

I. MATERIALS AND METHODS

This project contains a training dataset for symptoms which contains a list of symptoms in a column with a prognosis column representing the disease based on symptoms. And also about dataset which contains the description of the disease,diet to be followed,top 5 medicines to be included for the particular disease,their precautions,disease with their symptoms and workouts.These are trained using supervised machine learning algorithm which includes

Support vector machine,Randomforest,Gradient Boosting,K-Nearest Neighbors and Multinomial Naive Bayes to predict how efficiently it works.And the final output is implemented using Support Machine Vector.

Hardware requirements of the project include

- Operating System (Windows/Linux)
- RAM (4 GB Minimum)
- Secondary Storage (256 GB Minimum)

Software requirements include

- Internet browser (Chrome/Edge/Mozilla Firefox)
- Internet connection
- Jupyter Notebook
- Anaconda Navigator

II. EXISTING SYSTEM

The recent study about the project [11] made a prediction of disease .His system accurately predict the disease based on the symptoms given as input .The algorithm used here includes classification algorithm such as Naive bayes ,Random forest,Logistic regression and KNN and utilized to anticipate the disease.Even this system can provide whether the person contains the particular disease or not with the output true or false.[12] Some research is about to present a overview of collecting the social media data pharmacovigilance purpose and also classify the interaction between drug,disease and ADRs.[13]IoMT-Assisted patient diet recommendation system through machine learning model is focused on to implement both deep learning and machine learning models such as multilayer perceptron,Recurrent Neural Network,Long-Short term memory and Gated Recurrent Units. [14] Some model is mainly aimed to focus on predicting five possible disease based on the given dataset through the use of classification and supervised learning algorithms.This research mainly aim to provide patients with accurate result.[15]The recent paper shows that the analysis and survey of various diseases not only predict the disease based on the symptoms but also aim to predict the vulnerability

of the disease. Comparing the usefulness [16] of traditional statistical methods and machine learning algorithms is done in medical fields and aims to incorporate increasing AI companies from diagnosis to treatment of disease. Some researchers [19] found that selecting the recommendation system for the health care is a difficult task. Since it is time consuming due to the vast amount of algorithms. This paper examined the usage of different algorithm in health care system and also assist new researchers with the help of previous research. [20] This study shows recommendation system for health application of mosquito borne disease using hybrid approach.

III. PROPOSED SYSTEM

Our proposed solution aims to provide a adaptable method or platform for users where users can get the details of their health condition based on the input provided by them and also the dataset is trained and analyzed for finding the disease based on symptoms also provide along with the description of the disease, precautions to be followed, top 5 medications and diet to be followed. The system uses 5 different algorithms in order to find the accuracy and confusion matrix. This system uses numpy, pandas, make-classification libraries to accurately predict the output. The algorithms used here are Support vector machine, Random forest, Gradient Boosting, K-Nearest Neighbors and Multinomial Naive Bayes. The accuracy are compared and finally implemented using Support Machine Vector. It proves to be a reliable solution with high user desirability and at the same time ensures feasibility. Also our interface allows users to input their symptoms creates a seamless user experience along with prescription details.

IV. METHODOLOGY

Our system uses SVM (Support Machine Vector), Random Forest, Gradient Boosting, K-Nearest Neighbors and Multinomial Naive Bayes algorithm for training the model with the dataset samples.

Comparison is done between the algorithms to find the accuracy, precision and confusion matrix. And finally our system Medicine recommendation system is built using Support Vector Machine. The following steps are included to describe the flow of training the samples and recognizing the disease. Dataset collection of top 5 medicines with diseases, description of the disease, precautions, a set of symptoms is given to the particular disease is compared with the input symptoms using Support vector machine in order to predict the accurate result.



Figure 1: Block diagram of proposed work

A. Data collection:

Before applying the machine learning algorithms, exploratory data analysis (EDA) also called as first data preprocessing is applied to clean the and remove all the noises and outliers. Machine Learning algorithms can now perform better on the processed data. The diseases along with symptoms

their medication,diet and precautions are present in the dataset given below:

	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	

Figure 2 . Dataset sample

B.Data Cleaning:

Data cleaning is an important step involved in machine learning which involves identifying and removing irreverent data from the data set.It involves handling all the missing values in the dataset.These are the diseases that are included in the dataset:

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

array(['Fungal infection', 'Allergy', 'GERD', 'Chronic cholestasis',
      'Drug Reaction', 'Peptic ulcer disease', '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',
      'Osteoarthritis', 'Arthritis',
      '(vertigo) Paroymsal Positional Vertigo', 'Acne',
      'Urinary tract infection', 'Psoriasis', 'Impetigo'], dtype=object)
```

Figure 3.Diseases

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain	acidity	ulcers_on_tongue	...	blackheads	scurring	skin_p
0	1	1	1	0	0	0	0	0	0	0	...	0	0	0
1	0	1	1	0	0	0	0	0	0	0	...	0	0	0
2	1	0	1	0	0	0	0	0	0	0	...	0	0	0
3	1	1	0	0	0	0	0	0	0	0	...	0	0	0
4	1	1	1	0	0	0	0	0	0	0	...	0	0	0

5 rows x 133 columns

Figure 4. Description of symptoms

C.Training the dataset

SVM classifier library is imported for further execution.After importing all the dataset the dataset is trained using SVM classifier .The input is given by the user based on the input the

disease,medication,precaution,diet to be followed is given as output.

D.Plots of the dataset:

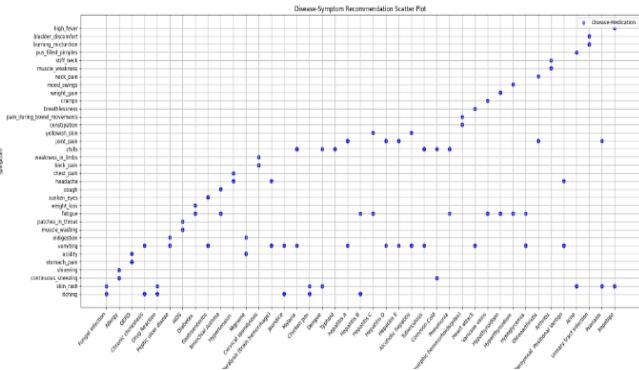


Figure 5.Disease-Symptoms scatter plot

The above figure shows the disease symptom scatter plot for various diseases.

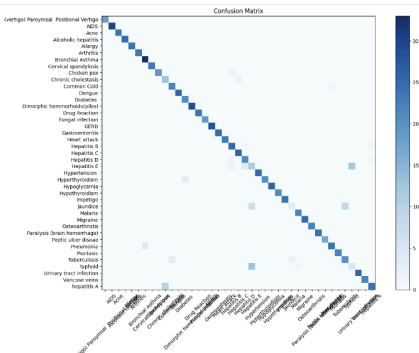


Figure 6.Confusion matrix

E.Comparing the Accuracy:

Comparison among the algorithms is done to make sure which algorithm works best for this project.The libraries used here includes: sk-learn which include make-classification,train_test_split,SVM,Random forest,Gradientboosting,MultinomialNB,accuracy_score,confusion matrix,numpy,matplotlib,seaborn.

SVM shows accuracy with 0.87 and confusion matrix: [[85 8] [18 89]]

Random forest shows accuracy with 0.90 and confusion matrix:

```
[[88 5]
 [15 92]]
```

GradientBoosting shows accuracy with 0.915 and confusion matrix:

```
[[90 3]
 [14 93]]
```

KNeighbors shows accuracy with 0.81 and confusion matrix: [[82 11] [27 80]]

MultinomialNB shows accuracy with 0.785 and confusion matrix: [[88 5] [38 69]]

F. SVM Classifier:

It is a supervised machine learning algorithm that classified the data by finding optimal line or hyperplane by 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]].

G. Random forest:

Random forests or decision forests refers to an ensemble learning method for classification, regression and other tasks which is operated by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. It shows 90% accuracy here with the confusion matrix of [[88 5] [15 92]].

H. GradientBoosting:

Gradient Boosting is a algorithm for boosting that combines several weak 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. It shows 91.5% accuracy here with the confusion matrix of [[903] [14 93]].

I KNeighbors:

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. It shows 91% accuracy here with the confusion matrix of [[82 11] [27 80]].

J. 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]].

V. RESULTS

Thus our project works effectively in predicting the disease, medication, precautions and diet recommendation of the input symptoms. And comparison is made between the algorithm each of which has different accuracy level with confusion matrix.

```
Enter your symptoms..... itching
-----predicted disease-----
Fungal infection
-----description-----
Fungal infection is a common skin condition caused by fungi.
-----precautions-----
1 : Bath twice
2 : use detol or neem in bathing water
3 : keep infected area dry
4 : use clean cloths
-----medications-----
5 : ['Antifungal Cream', 'Fluconazole', 'Terbinafine', 'Clotrimazole', 'Ketoconazole']
-----diet-----
6 : Avoid sugary foods
7 : Consume probiotics
8 : Increase intake of garlic
9 : Include yogurt in diet
10 : Limit processed foods
11 : Stay hydrated
12 : Consume green tea
13 : Eat foods rich in zinc
14 : Include turmeric in diet
15 : Eat fruits and vegetables
-----diet-----
16 : ['Antifungal Diet', 'Probiotics', 'Garlic', 'Coconut oil', 'Turmeric']
```

Figure 7. Result obtained by our project
Accuracy comparisons of different algorithms:

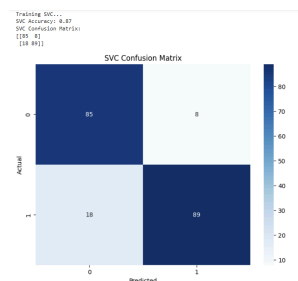


Figure 8 .Accuracy and confusion matrix of SVM Classifier

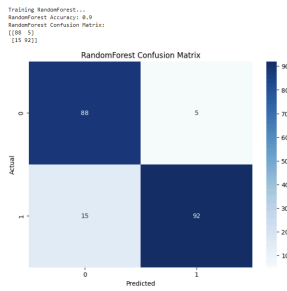


Figure 9 .Accuracy and confusion matrix of Random Forest Classifier

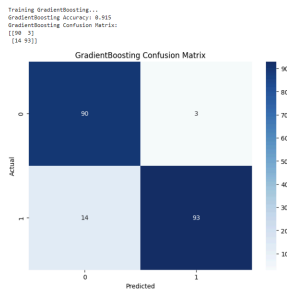


Figure 10 Accuracy and confusion matrix of GradientBoosting

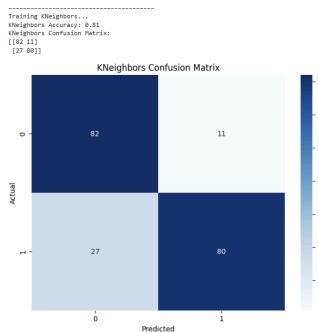


Figure 11 .Accuracy and confusion matrix of KNeighbors

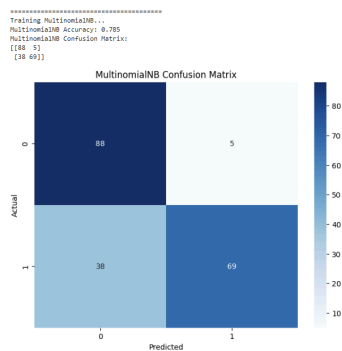


Figure 12 Accuracy and confusion matrix of MultinomialNB

VI. DISCUSSION

The model was trained using the dataset sample which contains 130 sample symptoms with 42 disease samples with description and top 5 medication samples. By comparing the accuracy of 5 different algorithms which includes Support machine vector, RandomForestclassifier, GradientBoosting, KNeighbors and MultinomialNB we infer that all these algorithms shows excellent result but GradientBoosting shows the accuracy of 91% which is greatest among all.

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

Table 1.Training Accuracy

The previous work that we considered was done by using mining algorithms such as Naive Bayes classifier, Random Forest and Decision tree Here each classifier Predicted different result and also aim to predict only disease based on symptoms .But our proposed work uses 5 different algorithms with different accuracy level and confusion matrix and finally recommendation system is designed using SVM classifier.

VII. CONCLUSION

Thus our project in developing a medicine recommendation system using SVM classifier efficiently helps people in identifying the disease with the set of input symptoms and predict the medications that can be used to cure the disease, precaution, diet to be followed etc. And also comparison among the

algorithms are made by finding the accuracy. The result shows that Gradient Boosting has the highest accuracy percentage of 91%. SVM classifier shows the accuracy of 87%. From this we infer that Gradient Boosting is best for with the accuracy of 91%. Therefore based on the comparative analysis among these algorithms in this research we gained a predictive knowledge of evaluating metrics such as accuracy confusion matrix, precision and we gained the insights of weakness and strength of all these algorithm. Finally, implementing this medicine recommendation system represents a significant milestone for building the gap between patients and the real time medical application. Additionally this implementation of medicine recommendation system enhanced the engagement among individuals and adherence to treatment plans.

VIII. REFERENCES

- [1] Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol.* 2017;2. doi:10.1136/svn-2017-000101
- [2] Reehan PM, Nikhilesh N, Revanth R, Chandra VV, Suvanam SB. Disease prediction and medicine recommendation system. *AIP Conf Proc.* 2024;2742: 020089.
- [3] 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.
- [4] 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.
- [5] “The Use of Machine Learning Algorithms in Recommender Systems: A Systematic Review.” 2018. *Expert Systems with Applications* 97 (May): 205–27.
- [6] 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.
- [7] 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
- [8] Mathur U. A Content Based Recommender System for Medicine using Machine Learning Algorithm. masters, Dublin, National College of Ireland. 2022. Available: <https://norma.ncirl.ie/6223/1/utkarshmathur.pdf>
- [9] Adegun AA, Ogundokun RO, Adebisi MO, Asani EO. 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. pp. 164–172.
- [10] Alghamdi M, Al-Mallah M, Keteyian S, Brawner C, Ehrman J, Sakr S. Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford Exercise Testing (FIT) project. *PLoS One.* 2017;12: e0179805.
- [11] Website. Available: https://www.researchgate.net/profile/Anand-Kumar-220/publication/353016433_Disease_Prediction_and_Doctor_Recommendation_System_using_Machine_Learning_Approaches/links/60e45592299b1ea9ee5f3ab/Disease-Prediction-and-Doctor-Recommendation-System-using-Machine-Learning-Approaches.pdf
- [12] Machine learning-based recommendation system for disease-drug material and adverse drug reaction: Comparative review. *Materials Today: Proceedings.* 2022;51: 304–313.

[13] Iwendi C, Khan S, Anajemba JH, Bashir AK, Noor F. Realizing an Efficient IoMT-Assisted Patient Diet Recommendation System Through Machine Learning Model. [cited 24 May 2024]. Available:<https://ieeexplore.ieee.org/abstract/document/8964364/>

[14] Gada D. Disease Prediction System using Machine Learning. 2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA. IEEE; 2022. doi:10.1109/iccubea54992.2022.10011002

[15] Galphat Y, Dayaramani C, Raghani D, Kithani L, Kriplani Y. Disease Prediction System using Machine Learning. [cited 24 May 2024]. Available:<https://ieeexplore.ieee.org/abstract/document/10127575>

[16] Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of Conventional Statistical Methods with Machine Learning in Medicine: Diagnosis, Drug Development, and Treatment. *Medicina*. 2020;56: 455.

[17] Abhishek A, Bindal AK, Yadav D. Medicine recommender system: A machine learning approach. *AIP Conf Proc*. 2022;2576: 050014.

[18] Rustam F, Imtiaz Z, Mehmood A, Rupapara V, Choi GS, Din S, et al. Automated disease diagnosis and precaution recommender system using supervised machine learning. *Multimed Tools Appl*. 2022;81: 31929–31952.

[19] Website. Available: <https://www.tandfonline.com/doi/abs/10.1080/1463922X.2022.2061078>

[20] Thomas B, John AK. Machine Learning Techniques for Recommender Systems – A Comparative Case Analysis. *IOP Conf Ser: Mater Sci Eng*. 2021;1085: 012011.