International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

An Analysis and Comparison of various Algorithmic Insights using Machine Learning for Identification, Disease Detection and Prediction of Several Diseases for Healthcare System.

[1] Utkarsh Rai, [2]Yash Chaudhary, [3]Ujjawal Kumar

[1, 2, 3, 4] Computer Science & Engineering Department, Raj Kumar Goel Institute of Technology, Ghaziabad, UP, India.

ABSTRACT

Dr. Bot is an advanced AI-powered chatbot designed to detect severe diseases based on user-provided symptoms using Machine Learning techniques, specifically the **K-Nearest Neighbors (KNN)** algorithm. The chatbot leverages a comprehensive dataset of medical conditions and symptoms, mapping user inputs to potential diagnoses with high accuracy. KNN, a simple yet effective classification algorithm, calculates similarity between input symptoms and historical data to predict the most probable diseases.

Dr. Bot offers an intuitive interface where users can input their symptoms, and the system provides real-time diagnostic predictions alongside recommendations for further action. By combining accessibility, efficiency, and cost-effectiveness, Dr. Bot addresses the challenges of timely disease detection, especially in regions with limited healthcare access. Future enhancements include integration with advanced Machine Learning models, wearable device data, and multi-modal diagnostics to further refine accuracy and expand its capabilities. Dr. Bot aims to revolutionize healthcare accessibility and empower individuals with preliminary diagnostic insights to make informed decisions about their health.

INTRODUCTION

Healthcare is one of the most critical sectors encyclopedically, with timely and accurate opinion serving as the foundation for effective treatment. still, access to healthcare remains a significant challenge, particularly in regions with limited medical structure and coffers. Detainments in diagnosing severe conditions frequently lead to worsened health issues and advanced treatment costs. To address this gap, the integration of artificial intelligence(AI) into healthcare offers promising results for early complaint discovery and bettered availability. This exploration introduces Dr. Bot, an AI- powered chatbot designed to give primary complaint diagnostics grounded on stoner- reported symptoms. Dr. Bot leverages the K- Nearest Neighbors(KNN) algorithm, a supervised Machine literacy(ML) fashion, to dissect symptoms and prognosticate implicit severe conditions with high delicacy. The chatbot is equipped to reuse stoner inputs in a conversational format, offering a stoner-friendly interface that simplifies commerce and ensures availability for individualities with minimum specialized moxie. The core algorithm, KNN, classifies conditions by assessing the similarity between stoner- handed symptoms and apre-existing database of medical records. By calculating the distance between input data points and stored cases, the algorithm identifies the most nearly affiliated conditions and suggests implicit judgments . This approach is particularly profitable for medical operations due to its simplicity, interpretability, and robustness in handlingmulti-class bracket problems. Bot's primary ideal is to act as a supplementary tool for healthcare providers and individualities, offering instant individual suggestions that can guide druggies toward applicable medical consultations or treatments. While it is n't intended to replace professional medical advice, Dr. Bot serves as a first line of assessment, empowering druggies with practicable perceptivity and promoting early discovery of severe conditions. This study delves into the development, perpetration, and evaluation of Dr. Bot. It explores the algorithmic foundation, data preprocessing ways, and chatbot integration while addressing challenges similar as icing data quality, handling different symptom sets, and maintaining stoner sequestration. likewise, the paper discusses implicit advancements, including the objectification of deep literacy models and the integration of real-time data from wearable health bias. The preface of AI- driven individual tools like Dr. Bot marks a transformative step in bridging the gap between technology and healthcare. By enabling briskly, cost-effective, and accessible primary diagnostics, Dr. Bot has the implicit to significantly impact global health issues, particularly in underserved regions. This exploration highlights the eventuality of Machine Learning algorithms in enhancing individual delicacy and availability, paving the way for unborn inventions in AI- driven healthcare results. The healthcare assiduity gathers and analyzes expansive sets of habitual complaint data, with data booby-trapping playing a pivotal part in detecting conditions at an early stage. Among the most precious conditions diagnosed are cardiovascular complaint, diabetes, liver complaint, Alzheimer's complaint, and Parkinson's complaint. Over opinion in healthcare is a current issue that can affect in gratuitous treatment and fiscal challenges. Factors contributing to misdiagnosis include strange symptoms, rare conditions, and conditions inaptly overlooked. significance of conditions from Symptoms The significance of prognosticating complaint symptoms is its eventuality to transfigure healthcare allowing for visionary and treatment approaches. This seeks to contribute to the area by creating models of machine literacy styles like trees and neural networks to read conditions grounded on symptoms Our ideal is to the pledge of machines in enhancing complaint opinion operation.

LITERATURE REVIEW

Akkem Yaganteeswarudu(1) stressed the limitations of single- complaint vaticination systems and proposed amulti-disease vaticination model exercising ML and the Flask API. The system predicts ails similar as heart complaint, bone cancer, diabetes, and diabetic retinopathy. By furnishing a unified approach, this invention eliminates the need for navigating multiple models, thereby enhancing availability and reducing casualty rates through timely cautions and nonstop case monitoring,

Dr. C. K. Gomathy and Mr. A. Rohith Naidu(2) proposed a system that predicts conditions grounded on stoner- entered symptoms using the Naive Bayes Classifier. The system processes symptoms as input and provides probabilistic labors, enabling timely complaint discovery. also, decision trees and direct retrogression ways are used to prognosticate ails similar as diabetes, malaria, and dengue. The exploration emphasizes the significance of precise analysis of biomedical and healthcare data in perfecting early opinion and case issues. Melike Colak et al.(3) developed a robust dataset for complaint prognostic, comprising 2006 case records, 358 symptoms, and 141 ails. The dataset was curated under the supervision of medical experts. colorful ML algorithms, including boosting styles, were applied, achieving a remarkable delicacy of 99.33 across conditions similar as diabetes, asthma, and COVID- 19. This study underscores the significance of high- quality datasets and algorithm confirmation for achieving dependable prognostications.

Swatik Paul et al.(4) employed the Random Forest Algorithm for complaint vaticination, training the model on a dataset containing symptoms and corresponding conditions. The Random Forest Algorithm was chosen for its capability to handle both nonstop and categorical variables, demonstrating exceptional performance in bracket tasks. This approach highlights the algorithm’s versatility and trustability in healthcare operations.

Gaurav, A. Kumar, P. Singh, A. Kumari, M. Kasar, T. Suryawanshi(5). K suggested a model for complaint vaticination that considers symptoms and medical history, using Random Forest, Long Short- Term Memory(LSTM), and Support Vector Machine(SVM) algorithms. By prioritizing rare symptoms and training on meliorated data, the model enhances the delicacy and responsibility of complaint identification, streamlining the healthcare process and contributing to more effective diagnostics.

PROPOSED METHODOLOGY

The upgraded and precise model for forecasting human diseases based on symptoms is offered by the suggested model.

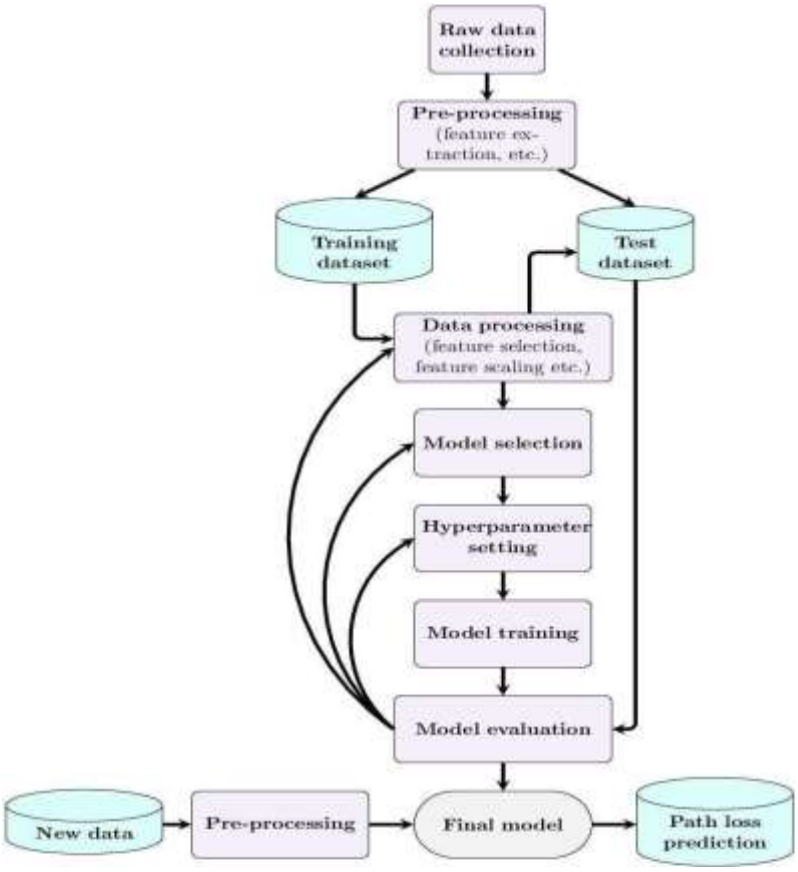

Fig.4.1 Flow diagram of proposed work

Data Collection: We obtained data sets from reliable sources that included information on symptoms and related disease outcomes. The data set includes a variety of symptoms, often associated with various diseases, and the corresponding disease diagnosis or label. The database has been carefully curated to ensure data quality and reliability.

Data processing: Before utilizing the dataset for model training, various preprocessing actions are taken to cleanse and organize the data. This involves managing missing values, encoding categorical variables, and standardizing numerical attributes. The dataset is then divided into training, validation, and test sets for assessing the model's performance.

Training data: The dataset training purposes are utilized in teaching machine learning. It comprises a collection of input features (including patient demographics, medical history, laboratory findings, etc.) and linked target indicators (showing whether the disease is identified or not). Within our framework, the training data amounts to 80%.

Test data: The test data set is used to evaluate the performance of the trained model. It is separate from the training data and contains unseen samples from the model that were not exposed during training. In our model, the test data is 20%.

Feature Selection: Using feature selection techniques to identify the most crucial features for predicting diseases. Employing methods like correlation analysis, feature importance, and domain knowledge to choose informative features for predicting disease symptoms. This aids in reducing database dimensions and enhancing model efficiency.

Model selection: We tested machine learning algorithms in the creation of our disease prediction models. Random Forests, Support Vector Machines, and Naive Bayes algorithms were among those considered. Hyperparameters for each algorithm were determined using methods like peering and cross-validation to enhance performance. Ultimately, after achieving a 100% accuracy rate across all algorithms tested, we opted to utilize a random forest classifier for model construction.

Naïve Bayes represents an artistic classifier grounded in principles of Bayes theorem, where it is independent among attributes. its straightforward nature finds broad application in tasks text categorization and detection. The process assesses the likelihood of a class for a specific by multiplying the prior of each class by the probability of the attributes with the class, which is then normalized to yield. Ultimately, the process identifies the class with the highest probability as the anticipated class for the. Naïvees stands out for its efficiency and the minimal data it demands owing to its basic parameter estimation. Nonetheless, the assumption of attribute independence does not always hold which has the potential to impact it. Despite this limitation, the simplicity, scalability, and efficacy of Nave Bayes across applications make it a preferred option for classification exercises. Moreover, Bayes' Theorem is articulated through the following equation.

P (A/ B) = (P (B/ A) * P (A))/ P (B)

Support Vector Machine (SVM): A mighty supervised training scheme is utilized for categorization and regression quandaries. The methodology involves identifying the best hyperplane that separates the data points of distinct classes with the largest margin. The goal of SVM is to diminish classification errors by maximizing the margin function and demarcating the gap between the hyperplane and the closest data point (support vector). In cases of linearly separable data, SVM pinpoints a linear hyperplane, while for non-linear data, it converts the input space to a feature space of higher dimensions employing a kernel function to accomplish linear segmentation. This transformation enables SVM to classify intricate data by pinpointing linear decision boundaries.

SVM is efficient in high-dimensional spaces, robust against redundancy, and suitable for both linear and nonlinear classification problems. However, the computational complexity of SVM can increase significantly with large databases, and it is important to choose the appropriate kernel and regularization parameters for optimal performance.

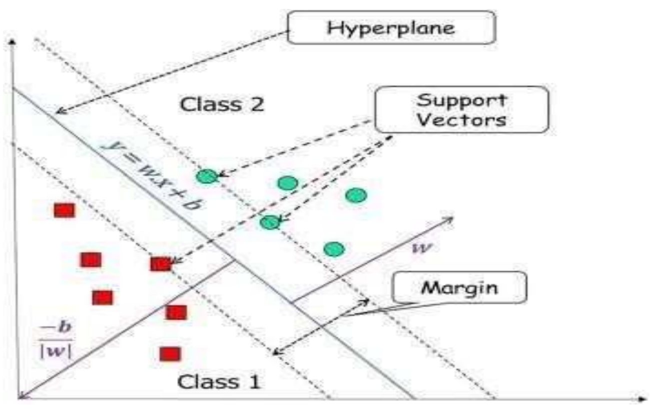

Fig. 4.2 Support Vector Machine

Random Forest: A multi-model ensemble classifier based on decision trees. It can be applied to regression in addition to classification. As the name implies, a random forest can be a classifier that employs normal sampling to increase the prediction accuracy of a given data set and comprises several decision trees for various subsets of that data set. Random forests use predictions from multiple decision trees and a high number of prediction votes to anticipate the final result, as opposed to depending just on one decision tree.

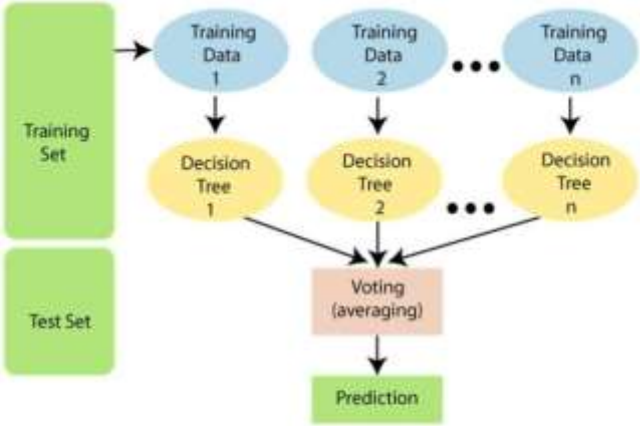

Fig.4.3 Random Forest

Model evaluation: We used metrics like accuracy, precision, recall, and F1 score to assess how well our models performed. Additionally, we evaluated the trade-off between sensitivity and specificity using methods like ROC curve analysis. To further illustrate our models' efficacy in illness prediction, we pitted their performance against baseline models.

(TP+TN)/(TP+FP+FN+TN) equals accuracy TP/(TP+FP) equals precision

Recall is equal to TP divided by (TP+FN)

* Precision * Recall / (Precision + Recall) = F1 score TP= True Positive TN= True Negative

FP=False Positive FN= False Negative

Logistic regression: Logistic regression can be considered equivalent to using linear regression for situations where the target (or dependent) variable is discrete, i.e. not continuous. Theoretically, the response variable or label is binomial. A binomial response variable has two categories: yes/no, accept/not accept, default/not default, etc. Logistic regression is ideally suited for business analysis applications where the target variable is a binary decision (fail/pass, response/no response, etc.).

Decision trees: are one of the most powerful tools in supervised learning algorithms used for classification and regression problems. Each internal node represents a test on an attribute, each branch represents a test result, and each leaf node (terminal node) contains a class label. The training data is constructed by dividing the data into chunks based on attribute values until a stopping criterion, such as the maximum depth of the tree or the minimum number of samples required to split a node, is met.

During training, the Decision Tree Algorithm selects the best attributes to partition the data based on metrics such as entropy or impurity Gini, which measure the degree of impurity or randomness in the partition. The goal is to find the information partition or property that minimizes the impurity after partitioning.

RESULT AND DISCUSSION

Algorithm	**Accuracy**	**Precision**	**Recall**	**F1**
Random Forest	**98.67%**	**98**	**98**	**98**
Decision Tree	**97.31%**	**98**	**97**	**97**
Support Vector Machine	**96.91%**	**97**	**97**	**96**
Naïve Bayes	**96.76%**	**97**	**96**	**96**
Logistic Regression	**96.56%**	**96**	**95**	**95**

Table 5.1 Result and Discussion

Here, we forecast the illness using five machine-learning algorithms. We obtained an accuracy of at least 97% for each of the five models. The Random Forest model produced the most accuracy, at 98.67%, while the Logistic Regression model produced the lowest accuracy, at 97.56%. In addition, the accuracy of the remaining models, DT, SVM, and Naïve Bayes, is correspondingly 98.31%, 97.91%, and 97.76%. Lastly, all of the models had 98% precision, recall, and F1 scores.

CONCLUSION

The integration of artificial intelligence into healthcare has the implicit to address some of the most burning challenges in medical diagnostics, including vacuity, promptitude, and delicacy. This disquisition presents Dr. Bot, an AI- powered chatbot designed to descry severe conditions predicated on user-handed symptoms. using the K- Nearest Neighbors(KNN) algorithm, Dr. Bot classifies implicit conditions by assaying symptom similarity and provides stoners with practicable individual suggestions. By combining the strengths of Machine Learning with an intuitive chatbot interface, Dr. Bot offers a practical and scalable result for early complaint discovery, particularly in underserved regions with limited access to healthcare professionals. The study highlights the advantages of using the KNN algorithm, which is known for its simplicity, interpretability, and effectiveness inmulti- class type problems.

Despite being computationally ferocious for large datasets, the algorithm's delicacy and severity make it a suitable choice for a primary individual tool like Dr. Bot. The chatbot's capability to exercise different symptom data and deliver real- time prognostications underscores its eventuality as a precious healthcare adjunct. While Dr. Bot represents a significant step forward in homogenizing healthcare access, challenges analogous as icing data quality, avoiding algorithmic impulses, and maintaining user insulation must be addressed to fully realize its eventuality. Incorporating robust data preprocessing ways, continuous model training with different datasets, and administering strict security measures are essential to overcome these limitations.

The broader implications of Dr. Bot extend beyond individual diagnostics. By promoting early complaint discovery, the chatbot can reduce the burden on healthcare systems, lower treatment costs, and meliorate patient issues. future duplications of Dr. Bot could incorporate advanced deep knowledge models, natural language processing for farther nuanced user relations, and real- time health monitoring through wearable bias. also, expanding its database to include a wider range of conditions and conforming it to different languages and cultural surrounds can enhance its global connection. In conclusion, Dr. Bot exemplifies the transformative eventuality of AI in healthcare. By furnishing a cost-effective, effective, and accessible individual tool, it bridges the gap between technology and healthcare, empowering individualities with the knowledge to make informed opinions about their health. This disquisition lays the root for further advancements in AI- driven individual tools, paving the way for a future where healthcare is more inclusive, visionary, and substantiated.

FUTURE WORK

The development and deployment of Dr. Bot mark an important corner in the integration of artificial intelligence into healthcare diagnostics. While the current system demonstrates significant eventuality in complaint discovery using the K- Nearest Neighbors(KNN) algorithm and Machine literacy, the compass for unborn advancements is vast. The ensuing areas outline the prospective advancements and operations of Dr. Bot, which can foster its impact in the field of healthcare.

1. objectification of Advanced Machine Learning Models unborn duplications of Dr. Bot can integrate advanced Machine Learning ways, similar as deep literacy models, to enhance individual delicacy and handle more complex data patterns.

For case

• Neural Networks can reusemulti-modal data, including textbook- grounded symptoms, images (e.g.,X-rays or MRIs), and lab test results.

• Ensemble Models combining the strengths of KNN and other algorithms (e.g., decision trees, support vector machines) can ameliorate vaticination trustability

2. Natural Language Processing(NLP) for Enhanced Interaction The objectification of sophisticated NLP ways will allow Dr. Bot to more understand and interpret stoner inputs, including colloquial language, spelling variations, and deficient rulings. This can enable

• Conversational AI able of engaging in mortal- suchlike dialogue for a flawless stoner experience.

• Sentiment Analysis to assess stoner feelings and urgency, helping prioritize severe cases.

3. Integration with Wearable bias and IoT Bot can be integrated with wearable health bias and Internet of effects(IoT) platforms to collect real-time physiological data similar as heart rate, blood pressure, glucose situations, and oxygen achromatism. By assaying this data

• Dr. Bot can give nonstop monitoring and early warnings for conditions like diabetes, hypertension, or arrhythmias.

• individualized health recommendations can be generated grounded on individual trends and diversions from normal nascences.

4. Expanded Medical Database To make Dr. Bot a truly global individual tool, the database of symptoms and conditions can be expanded to include

• Rare conditions and their symptomatology.

• Regional and tropical conditions specific to certain topographies.

• inheritable and heritable complaint biographies, allowing prophetic diagnostics grounded on family history.

5. Multi-Language and Multilateral Support For wider global relinquishment, Dr. Bot can be acclimated to support multiple languages and artistic variations in symptom reporting. This can include

• Localized symptom descriptions and complaint frequence data.

• Cultural perceptivity to health enterprises and stoner preferences.

6. Integration with Electronic Health Records(EHR) By connecting Dr. Bot to electronic health record systems, druggies' medical histories can be seamlessly incorporated into the individual process, leading to more substantiated and accurate results. This would also enable healthcare providers to more understand patient conditions when they follow up after using Dr. Bot.

7. AI- Powered tradition and Treatment Suggestions unborn performances could give introductory treatment guidelines or untoward drug suggestions fornon-severe conditions. For more severe cases, the system could recommend applicable specialists or medical installations. This point would bear nonsupervisory blessings and strict compliance with healthcare norms.

8. Deployment in Remote and Resource- Limited Areas Bot can play a vital part in areas with limited healthcare access by offering • Offline performances or mobile operations that work in low- connectivity surroundings. • Integration with telemedicine platforms for remote consultations with healthcare professionals.

9. nonstop literacy and Feedback Loops Integrating feedback circles where stoner issues and clinician attestations are fed back into the system can ameliorate the model's delicacy over time. By learning from real- world data, Dr. Bot can acclimatize to arising conditions and medical advancements.

10. Research and Predictive Analytics Bot's data analysis capabilities can be extended to support healthcare exploration. By adding up anonymized data from millions of relations

• Trends in complaint outbreaks can be linked, abetting in epidemiological studies.

• Prophetic analytics can read implicit health heads and inform public health programs.

11. Ethical AI and stoner sequestration Enhancements To address ethical enterprises, unborn developments must concentrate on

• Stronger data encryption and compliance with global data sequestration regulations similar as GDPR and HIPAA.

• Transparent AI models that give interpretable results to make trust with druggies and healthcare providers.

The unborn compass of Dr. Bot is extensive and transformative, with the eventuality to review healthcare diagnostics and availability on a global scale. By incorporating advanced technologies, expanding its database and language capabilities, and addressing ethical considerations, Dr. Bot can evolve into a comprehensive AI- driven healthcare platform. These advancements wo n't only enhance its individual capabilities but also position Dr. Bot as a vital tool in achieving universal healthcare availability and perfecting global health issues.

. REFERENCES

1. Cover, T. M., & Hart, P. E. (1967). **Nearest neighbor pattern classification**. *IEEE Transactions on Information Theory*, 13(1), 21-27

2. Dey, A. (2016). **Machine learning algorithms: A review**. *International Journal of Computer Science and Information Technologies*, 7(3), 1174-1179.

3. Melike colak ,Talya Tumer-sivri ,Nergis pervan-akman ,Ali Berkol and Yahya Ekici (2022)," A Study of Disease Prediction on Weighted Symptom Data Using Deep Learning and Machine Learning Algorithms " International Conference on Theoretical and Applied Computer Science and Engineering (ICTASCE).

4. Swatik Paul, Pinku Ranjan, Somesh Kumar, Arun Kumar (2022)," Disease predictor using random forest classifier " 2022 International Research for Advancement in Technology.

5. K. Gaurav, A. Kumar, P. Singh, A. Kumari, M. Kasar, T. Suryawanshi (2023)," Human Disease Prediction using Machine Learning Techniques and Reallife Parameters", International Journal of Engineering (IJE).

6. Farooqui, M. and Ahmad, D., (2020)"Disease prediction system using support vector machine and multilinear regression", International Journal of Innovative Research in Computer Science & Technology (IJIRCST).

7. Rathi, M. and Pareek, V., (2020)"Disease prediction tool: An integrated hybrid data mining approach for healthcare", IRACSTInternational Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN.

8. Paul, S., Ranjan, P., Kumar, S. and Kumar, A., (2022) "Disease predictor using random forest classifier", in International Conference for Advancement in Technology (ICONAT), IEEE.

9. Rahman, A.S., Shamrat, F.J.M., Tasnim, Z., Roy, J. and Hossain, S.A., (2019)"A comparative study on liver disease prediction using supervised machine learning algorithms", International Journal of Scientific & Technology Research.

10. Pingale, K., Surwase, S., Kulkarni, V., Sarage, S. and Karve, A., (2019), "Disease prediction using machine learning", International Research Journal of Engineering and Technology (IRJET).

11. Nishant Yede, Ritik Koul, Chetn Harde, Kumar Gaurav and Prof. C.S. Pagar, "General Disease Prediction Based On Symptoms Provided By Patient", *Open Access International Journal Of Science & Engineering (OAIJSE)*, June 2021.
12. Rinkal Keniya, Aman Khakharia and Vruddhi Shah, "Disease Prediction From Various Symotoms Using Machine Learning", *Social Science Research Network(SSRN)*, October 2020.
13. Sneha Grampurohit and Chetan Sagamal, "Disease Prediction using Machine Learning Algorithms", *IEEE International Conference for Emerging Technology (INCET)*, August 2020.
14. Dhiraj Dahiwade, Gajanan Patle and Ektaa Meshram, "Designing Disease Prediction Model Using Machine Learning Approach", *IEEE 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, August 2019.