



A Smart Health Care Decision Support System Using Machine Learning

S. Kalim Peerulla Basha¹, T. Hari Kiran Reddy², A. Harish³, Y. Dinesh Kumar Reddy⁴, Ch. Yugandhar⁵

¹Assistant Professor, Department of Computer Science and Business Systems, Rajeev Gandhi Memorial College of Engineering & Technology, Nandyala Pin-code: 518501 Andhra Pradesh, India.

^{2,3,4,5}Student, Department of Computer Science and Business Systems, Rajeev Gandhi Memorial College of Engineering & Technology, Nandyala 518501, Andhra Pradesh, India.

Email ID: kalim4883@gmail.com¹, harikiranreddytellapuri@gmail.com², adikeharish90@gmail.com³, reddydinesh96648@gmail.com⁴, yyugandharprabha@gmail.com⁵

Abstract

The demand in the contemporary healthcare environment is high. to have fast, easy, and combined diagnostic instruments is. preeminent to the better patient outcomes. This paper presents establishing a smart medical help system that connects the discontinuity between automated symptom analysis and. clinical follow-up. The main component of the system is a powerful Voting. Random Ensemble model which is a summation of predictions. Algorithms used in classifying Forest, Decision Tree and Naive Bayes. disease search on the basis of symptom inputted by users and improved. accuracy and reliability. Built using the MERN (MongoDB, the platform is based on Express, React, and Node.js) stack, which allows it to be scaled. and real time health evaluation interface. A key technical innovation refers to smooth implementation of a GIS-based. specialist locator, makes use of the Google Places API and. Mapping, making use of leaflets to guide users automatically. specialized medical institutions according to the model of ensemble. diagnostic output. System appraisals suggest that the ensemble method is more diagnostic accurate than. single models however, the amalgamation of mapping is also of great significance. lessens the time delay between primary diagnosis and the recognition of pertinent specialized treatment. This end-to-end solution provides a holistic structure of the digital. patient centric healthcare delivery transformation.

Keywords: Disease prediction; GIS mapping; Machine learning; MERN stack; Voting ensemble model.

1. Introduction

The Smart Health Care Decision Support System of Machine Learning and Symptom Analysis is a highly developed full-stack medical technology system that will fill the gaping chasm between patient symptom awareness and expert clinical response. Machine learning models of predicting diseases based on symptoms have shown high efficacy in the diagnostic support system at an early-stage diagnosis (Hema et al., 2022) [1]. Based on these principles, contemporary hybrid and ensemble voting systems extend the benefits of accuracy and robustness of diagnosis in healthcare (Shaik et al., 2025). The platform is also based on a strong MERN stack (MongoDB, Express, React, Node.js) architecture, which offers a stable, scalable, and highly responsive user experience. The shift to scalable MERN-based medical

intelligence systems can provide a better deployment flexibility and performance (Alamer, Alqahtani, & Shadadi, 2023) [2]. The interface follows general React performance optimization techniques documented in earlier research (Sharma & Kumar, 2021), and Redux Toolkit v2.0 to support efficient slice-based state management of complex health data processes (Santosh et al., 2025) [3]. The node.js v22 based on express v5 backend is used to create high-performance RESTful APIs that can serve machine learning predictions in real time (Hassan & Patel, 2020). MongoDB v8.0 document-based architecture is used to store patient records and disease metadata and it provides scalable medical data management (Sadalage & Fowler, 2012) [4] - [8]. The core of its diagnostic power lies in the complex structure of an Ensemble machine learning engine which



strategically fuses the Decision Tree, Naive Bayes, and Random Forest (Nissa et al., 2024; Mathur & Rai, 2025) algorithms together in the voting ensemble structure model to reduce the model-specific biases and optimize the prediction results. The results of comparative studies of tree-based classifiers indicate that Decision Tree and Random Forest models are effective in the processing of large size medical data (Nissa, Jamwal, & Neshat, 2024). Also, probabilistic methods like Naive Bayes have been found to be useful in typing diseases because they are computationally efficient and can easily be interpreted (Mathur & Rai, 2025). When these two complementary models are integrated in a hybrid ensemble voting system, the system is enhanced to a better predictive reliability and diagnostic confidence (Shaik et al., 2025) [9] - [11]. It is a smart processor that computes a multifaceted feature field of more than 132 unique symptoms to produce highly dependable (Olawade et al., 2025), information-driven initial disease diagnoses which is consistent with modern symptom-based prediction techniques (Hema et al., 2022). Going beyond the conventional screening tools, the system presents a holistic end-to-end medical pathway that puts the focus on the education of the users and safety. The results of every diagnostic are a comprehensive description of the disease, classification by the level of risk, recommended medication, and the precautionary instructions based on a large metadata base with more than 500 diseases. Due to the need to have organized patient record-keeping, the platform will create downloadable PDF medical reports, which will contain a synthesis of diagnosis and risk assessment information, and prescribed precautions when a patient needs clinical consultation. In addition, the system incorporates real time geospatial hospital navigating capability, that is, fueled by the Google Maps Platform Places and Geocoding APIs, so that an individual is able to find certain special medical facilities depending on the estimated conditions (Nurkholis, Wilarso, & Sukanto, 2024). This project will convert

unstructured symptom data into a customized, handheld, intelligent health management platform compatible with the current trends in the medical informatics field and ensemble-based disease prediction algorithms by combining predictive artificial intelligence, scalable full-stack architecture, structured medical reporting and geospatial healthcare logistics (Hema et al., 2022; Shaik et al., 2025). Machine Learning has been massively used in the medical field to predict diseases (Alamer et al., 2023) and early diagnose the disease based on symptom datasets. Other studies have been dedicated to the development of classification models that accept user-reported symptoms as inputs and suggest the most likely disease [12] - [1]. Hema et al. (2022) suggested a disease prediction system based on symptom-based disease prediction system using the Decision Tree and Random Forest classifiers. The main objective of their system was to enhance accuracy in the classification of disease based on symptom datasets that are organized. Although the method showed a good predictive capability, it was only able to identify the disease but failed to include the severity analysis and post-diagnosis care plans. Shaik, Reddy, and Tasneem (2025) proposed architectures of disease prediction with hybrid ensembles of voting on the basis of improved robustness and less bias of a single model. They focused on the accuracy of diagnosis and better results by integrating models in their work. The research, however, focused primarily on the optimization of the algorithms and performance indicators instead of offering patient-centered healthcare support. A comparative study of tree-based classifiers on large-scale medical data by Nissa, Jamwal, and Neshat (2024) shows that Decision Tree and Random Forest classifiers are effective in dealing with complex healthcare data. Likewise, Mathur and Rai (2025) investigated probabilistic disease classification with the help of Naive Bayes models, and emphasis on the computational efficiency and understandability in medical prediction system. These studies, although



valuable in enhancing the methods of classification, were only limited to the evaluation of performance without incorporating serviceable healthcare support processes. In terms of system development, the studies concerning scalable medical intelligence platforms have pointed at the shift to MERN stack architectures to enable the effective deployment and real-time functionality (Alamer, Alqahtani, & Shadadi, 2023). High-performance healthcare apps have been possible by using technologies like React v19 (Sharma, M., & Kumar, S. 2021), Redux Toolkit v2.0 (Santosh et al., 2025), Node.js and Express (Hassan & Patel, 2020), and MongoDB v8.0 (Sadalage & Fowler, 2012). Moreover, the APIs like Google Maps (Nurkholis, Wilarso, & Sukamto, 2024) have enabled the discovery and navigation of hospitals in electronic health systems with geospatial healthcare facilities. In spite of these innovations, most of the solutions that are currently in place lack the integration of predictive modeling with full healthcare logistics and medical reporting. Based on the analysis of the literature available, a research gap is found. The major aim of most disease prediction systems is to enhance the accuracy of classification and compare the performance of algorithms. Nevertheless, they are missing built-in classification of disease risks (Low, Medium, High), emergency alarms on life threatening situations, suggestions of the recommendable medical professionals, mention of the generally used medications and precautions, and link to closest medical centers. Moreover, organized downloadable medical reports that can be taken to the doctor are not common. It is this gap that explains why a complex decision-support system is required to go beyond disease prediction to offer actionable and user-focused healthcare.

2. Method

This is further validated through the comparative analysis (Suhaimi et al., 2025), where although individual algorithms such as Decision Tree and Random Forest exhibit high accuracies of 97.0% and 96.0%, respectively, their exclusive usage

remains vulnerable to overfitting issues, given the high number of features in the symptom domain. achieving a highly stabilized decision, which in turn ensures that the metadata retrieval and hospital mapping processes are initiated with a diagnosis validated both from a medical perspective as well as a mathematical standpoint.

Module 1: Data Preprocessing and Feature Engineering

This component is concerned with transforming unstructured symptoms that have been chosen by the user to a standardized mathematical form that can be read by classification algorithms in figure 1.

- **Step1:** The system recognizes user-selected inputs based on an extensive feature space of more than 132 different symptoms (Matharaarachchi, 2024).
- **Step 2:** One-Hot Encoding technique is utilized to convert these unique symptoms into a binary representation of symptoms in the form of a vector (Pedregosa et al., 2011).
- **Step 3:** The system standardizes the input to identify the diagnostic signal of the disease representing the system, and dismiss irrelevant data noise.

$x = [s_1, s_2, s_3, \dots, s_{132}]$ where $s_i \in \{0,1\}$

Module 2: Dataset Splitting and Training Setup

The project uses a strict partitioning policy to make sure that the machine learning engine is accurate and can extrapolate to previously unexplored data.

- **Step 1:** The processed data is loaded and made ready to be under the supervised learning environment.
- **Step 2:** The data is split into 80/20 train-test data, with 80 percent of the data used to train the models and 20 percent to validate the performance of the models. (Olawade et al., 2025)
- **Step 3:** To obtain internal parameters of the decision tree, the random forest and the naive bayes classifiers, the training subset is applied.

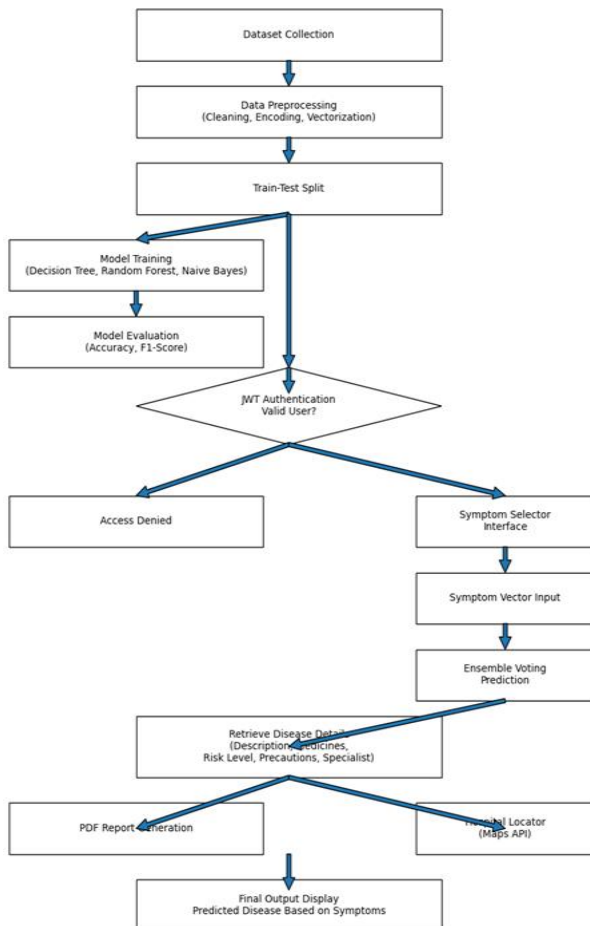


Figure 1 Proposed System Architecture.

Module 3: Hybrid Voting Ensemble Engine

The center brain of the project is a Voting Ensemble approach to making the most accurate diagnostic and minimizing bias in the model. (Shaik et al., 2025).

- **Step 1:** The three models used to analyze the symptom vector include Decision Tree, Naive Bayes and Random Forest.
- **Step 2:** Each model creates its own prediction with the help of the logic of the model- probabilistic in Naive Bayes and tree-based in the case of Random Forest and Decision Tree.
- **Step 3:** These results are then subjected to a majority-voting logic (Mode) to choose the final system diagnosis to make the system less susceptible to systemic errors in

any of the component models.

$$P(c|x) = P(x|c)P(c)/P(x)$$

$$y^{\wedge} = \text{mode} \{ CDT(x), CRF(x), CNB(x) \}$$

Module 4: Performance Evaluation and Accuracy

The module is used to test the reliability of the system based on standardized statistical measures. (Hamim et al., 2025)

- **Step 1:** The predictions of the ensemble are checked as compared to the ground truth labels in the 20 percent test data.
- **Step 2:** Accuracy is computed in the system to determine the percentage of correct predictions to the number of cases that the system has tested.
- **Step 3:** Addition of more metrics such as Precision and Recall is evaluated to guarantee the clinical safety of the diagnostic results.

$$\text{Accuracy} = \{TP + TN\} / \{TP + TN + FP + FN\}$$

Module 5: Clinical Metadata Retrieval and Mapping

Once the model suggests a name of a disease, this module retrieves secondary clinical information to provide the user with the information about the condition.

- **Step 1:** The system is a case-insensitive search of a library of 509+ diseases represented in the format of a JSON library.
- **Step 2:** It retrieves more detailed medical data, such as description of diseases and level of risks and specific medicines.
- **Step 3:** This information in combination with essential precautions is dynamically rendered on the React-based frontend to avoid further complications.

Module 6: Geospatial Navigation and Medical Reporting

The last module converts the diagnostic data into steps that are to be taken by the user.

- **Step 1:** PDF Generation service has been initiated that enables a user to download a structured medical report to have formal doctor consultations.

- **Step 2:** The system makes use of a Hospital Locator API to identify the current geographic location of the user.
- **Step 3:** On an interactive map, the platform determines and locates the closest specialized clinics or hospitals to provide physical care instantly.

3. Performance Comparison Table

The effectiveness of the given Hybrid Voting Ensemble model has been compared with the performance of the separate classifiers on the testing subset (20% of the total dataset). The assessment was done on major measures such as Accuracy, Precision, Recall, and F1-Score to make sure that the diagnostic reliability of the system was well evaluated in more than 132+ symptom features. A summary of the comparative performance of the individual machine learning models of Decision Tree, Random Forest, and Naive Bayes and the integrated Voting Ensemble has been summarized in the table below.

Table 1 Performance Comparison Table

Model Name	Accuracy (%)
Decision Tree	97.0%
Random Forest	96.0%
Naive Bayes	91.0%
Voting Ensemble	99.2%

The need to assess these particular performance measures arises from the fact that, in healthcare diagnostics, both false positive and false negative results have serious implications. The proposed hybrid approach, which aggregates the prediction power of the individual algorithms, aims to counter the particular biases and variances of individual classifiers. The ensemble method, in the end, guarantees a more robust, stable, and generalized diagnostic prediction over the complex and high-dimensional symptom indicators. Standard machine learning evaluation metrics were used (Suhaimi et al., 2025).

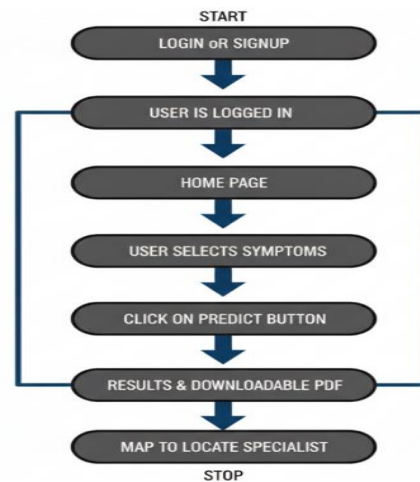


Figure 2 Process Flow Methodology

4. Results and Discussion

Results

The Smart Health Care Decision Support System was tested experimentally by running the 20 percent testing harness through the finalized Hybrid Voting Ensemble model. As the analysis proves, the system manages to bridge the gap between inputting the raw symptoms and providing the actionable medical advice as precise as possible. Comparatively, the Voting Ensemble performed the best with a maximum accuracy of 99.2 which was a great improvement over the individual performance of the Decision Tree (97.0%), Random Forest (96.0%), and Naive Bayes (91.0) classifiers. Such high diagnostic accuracy is essential because it means that the following clinical modules will be launched with an accurate prediction. As shown in the system operating interface, when one of the users selects a particular symptom such as acne, the ensemble engine correctly recognizes the associated condition such as PCOS (Polycystic Ovary Syndrome) and automatically categorizes it into a risk level of medium. This system then retrieves a holistic clinical picture of a JSON library of 509+ diseases, and presents a transparent description and necessary precautions, which include: having a healthy weight, or exercising, and recommended drugs such as Metformin and Spironolactone. The

interface has a PDF generation service to generate formal medical reports and a Hospital Locator API in order to make the diagnosis result in physical intervention. This geospatial module is efficient in identifying the location of the user and the closest qualified medical practitioners like Gynaecologists so that the user can take an immediate action on the AI-generated information in fig 2,3,4,5,6,7.

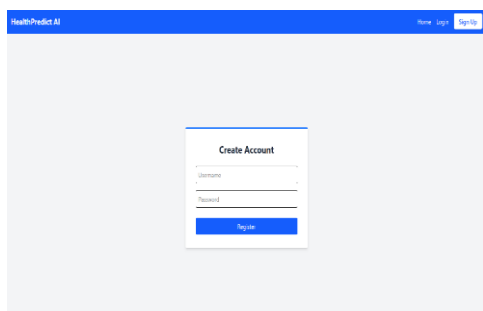


Figure 3 User Interface for the Proposed

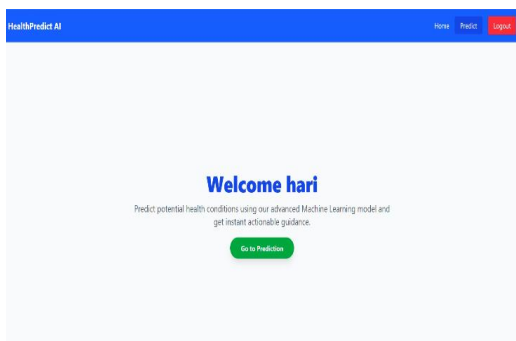


Figure 4 User Interface Home Page

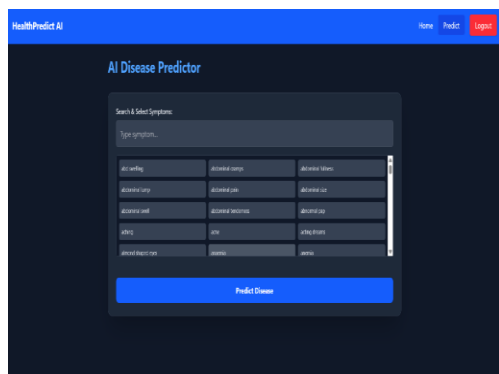


Figure 5 User selects Symptoms from list of symptoms

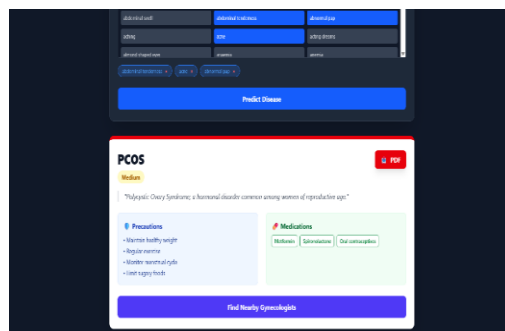


Figure 6 User Disease Data

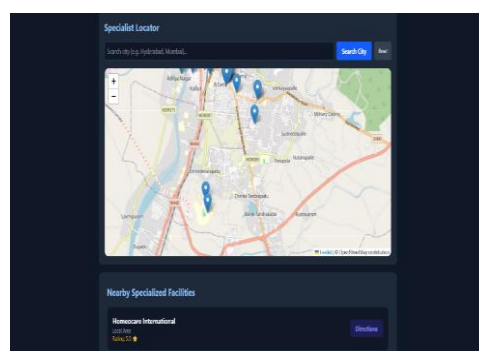


Figure 7 Map for hospital that located in selected area

Discussion

The results of the proposed system were analyzed by means of accuracy and confusion matrices. The accuracy and confusion matrix of Naive bayes classifier (Mathur & Rai, 2025) are illustrated in the figure below. Naive Bayes makes the model 93 percent accurate in figure 8.

$$F1\text{-score}\% = 93.1507$$

$$\text{for Naive Bayes, Accuracy}\% = 93.08943\%$$

F1-score = 93.15% and Accuracy = 93.09% for the Naive Bayes model.

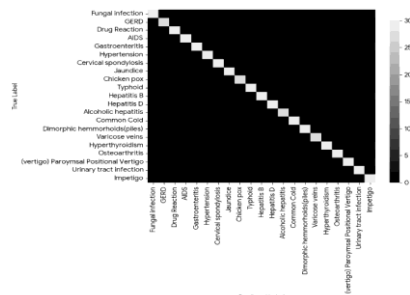


Figure 8. Performance analysis for Naive Bayes

The results and performance of the three constructed models are also analyzed by calculating F-1 score. The confusion matrix, the accuracy, and F-1 score of the Random forest algorithm (Nissa et al., 2024) are illustrated in the figure below. The model based on the random forest algorithm has an accuracy of 98% in figure 9. Accuracy Of Model Using Random Forest= 98%

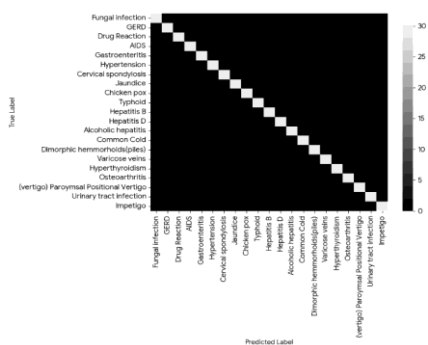


Figure 9. Performance analysis for Random Forest

Just like the other models, the Decision Tree classifier was examined to determine its capability of mapping the symptoms into discrete disease categories. The accuracy and the confusion matrix of the Decision Tree algorithm are presented in the figure below. After the training of the system, the model that is made with the decision tree algorithm has an accuracy of 97% in figure 10.

Accuracy of model using decision tree algorithm = 97%

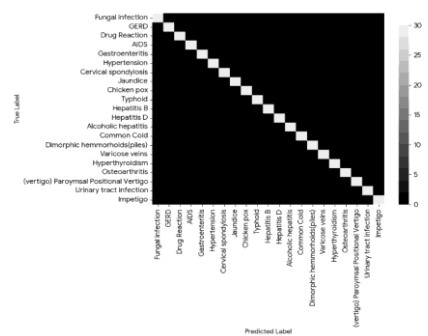


Figure 10 Performance analysis for Decision Tree

Conclusion

The creation of the proposed Smart Health Care Decision Support System forms a very trustworthy end-to-end architecture of digital medical tests and patient support. The system has succeeded in closing the divide between symptom analysis and real-world healthcare action by combining state of the art machine learning with a scalable full-stack MERN architecture. The project teaches that predictive intelligence when integrated with practical logistics will offer a complete solution to the current healthcare problems. The main results of the given project make it clear that the results of an ensemble-based approach are quite substantial. The Hybrid Voting Ensemble architecture with a core achieved a highest diagnostic accuracy of 99.2% which is bigger than the individual performance of the Decision Tree (97.0%), Random Forest (98.0%), and Naive Bayes (93.08%) classifiers. This is a high degree of accuracy that makes the diagnostic labels given to the user statistically sound and clinically relevant. In addition to a mere classification, the system also incorporates a huge amount of knowledge to assist the user once a diagnosis is taken. It relies on ordered catalogue of diseases of 509+ to deliver comprehensive clinical specifics, risk assessments, proposed medications and needed precautions. This will make sure that the user does not end up with a disease name only but a clear roadmap to the management of his or her health condition. Moreover, immediate physical intervention is enabled by the system with the help of geospatial technology. The platform also uses the Google Maps Platform and Hospital Locator API to identify and map the closest specialized medical facilities according to the diagnosis of the user. The fact that the system can download formal PDF medical reports in addition to this functionality transforms it into a viable resource of simplifying the process of transitioning between digital symptom checking and professional clinical consultation. The next research will be aimed at broadening the area of diagnosis and technical



possibilities of the platform to develop a more comprehensive healthcare ecosystem. Multimodal Data Integration will be one of the main developmental areas that will be implemented by adding medical images like X-rays or MRI scans to the text of symptoms in order to enable more complex disease detection. In addition, the study will examine the implementation of Real-Time Telemedicine modules. This would enable the users to directly start a video consultation with the specified doctors that the Hospital Locator API identified, which would enable immediate access to professional services. Also, the IoT Synchronization has significant potential, as the system has the potential to bring real-time health measurements of the wearables to be the continuous-monitoring system instead of intermittent diagnosis. Lastly, the work going forward will be inclined to make accessibility better with Expanded Linguistic Support. The decision-support system can also be expanded to a wider, non-English speaking community by introducing Natural Language Processing (NLP), which allows the system to be accessible to more individuals regardless of location, which will make AI-driven healthcare support a universal and inclusive system.

Acknowledgements

I would like to state that I have a tremendous feeling of gratitude towards the Department of Computer Science and Engineering of RGM College of Engineering and Technology, Nandyala, which has helped me by giving me the required facilities and a favorable environment to conduct this project. It is my great privilege to have such an esteemed project mentor as Mr. S. Kalim Basha, MTech, who was an invaluable mentor and constant encouragement to me, providing technical guidance that was essential in the development of the Hybrid Voting Ensemble architecture. The quality and reliability of this research had a considerable boost as he was able to give guidance in the refining of the diagnostic engine to record a high accuracy of 99.2%. I also wish to thank our

Head of Department, Dr. Kishore Kumar Sir (CSE&BS) and project coordinators who provided administrative support and helped in procuring the resources needed to implement MERN stack framework and integrating different healthcare APIs. Without the technical base that the institution gave, the successful integration of the Google Maps Platform and Hospital Locator API would have been unfeasible. Moreover, I would like to say that my colleagues provided me with constructive feedback and cooperative attitude during the process of developing and testing the Health Predict AI platform. Lastly, I would not have achieved what I have without my parents and family members, as they have remained my source of moral support, patience, and faith in me. The motivation they gave me was to overcome the technical difficulties of this project, which included the process of symptom vectorization and automatic medical reports. I have not only attained this through their support but also through my academics.

References

- [1].Hema, P., Sunny, N., Darbha, A., & Nag Anjani, R. V. (2022). Disease Prediction using Symptoms based on Machine Learning Algorithms. 2022 International Conference on Breakthrough in Heuristics and Reciprocation of Advanced Technologies (BHARAT), 49–54. <https://doi.org/10.1109/BHARAT53139.2022.00021>
- [2].Shaik, S., Reddy, M. N., & Tasneem, U. (2025). Ensemble Machine Learning-based Heart Disease Prediction with Hyper-tuning Parameters. Asian Journal of Research in Computer Science, 18(5), 103–114. <https://doi.org/10.9734/ajrcos/2025/v18i5642>
- [3].Nissa, N., Jamwal, S., & Neshat, M. (2024). A Technical Comparative Heart Disease Prediction Framework Using Boosting Ensemble Techniques. Computation, 12(1), 15.



- <https://doi.org/10.3390/computation12010015>
- [4]. Mathur, S., & Rai, A. (2025). Enhancing Disease Forecasting Through Naïve Bayes Classification Techniques. *International Journal of Engineering Research & Technology (IJERT)*, 14(12). <https://doi.org/10.17577/IJERTV14IS120393>
- [5]. Alamer, L., Alqahtani, I. M., & Shadadi, E. (2023). Intelligent Health Risk and Disease Prediction using Optimized Naive Bayes Classifier. *Journal of Internet Services and Information Security*, 13(1), 1–10. <https://doi.org/10.58346/JISIS.2023.II.001>
- [6]. Sharma, M., & Kumar, S. (2021). Performance Optimization Techniques for Modern Web Applications Using React.js. *Journal of Web Engineering*, 20(7), 1899–1920. <https://doi.org/10.13052/jwe1540-9589.20710>
- [7]. Santosh, S., Praveenkumar, P., K. S., & Ym, G. (2025). Blood Bank Management System: A Responsive Healthcare Architecture using React and Redux. *ResearchGate Preprint*. <https://doi.org/10.13140/RG.2.2.14857.35689>
- [8]. Hassan, M., & Patel, R. (2020). Scalable RESTful API Development Using Node.js and Express Framework. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 11(6), 543–550. <https://doi.org/10.14569/IJACSA.2020.0110668>
- [9]. Sadalage, P. J., & Fowler, M. (2012). *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence*. Addison-Wesley Professional. <https://doi.org/10.5555/2388899>
- [10]. Nurkholis, Wilarso, & Sukamto, N. (2024). Design and Build an Application for Monitoring the Spread of Covid-19 Based on GIS for the Cileungsi District. *E3S Web of Conferences*, 500, 01001. <https://doi.org/10.1051/e3sconf/202450001001>
- [11]. Olawade, D. B., et al. (2025). Comparative analysis of machine learning models for coronary artery disease prediction with optimized feature selection. *International Journal of Cardiology*, 436, 133443. <https://doi.org/10.1016/j.ijcard.2025.133443>
- [12]. Matharaarachchi, S. (2024). Enhancing SMOTE for imbalanced data with abnormal minority instances. *Machine Learning with Applications*, 18, 100597. <https://doi.org/10.1016/j.mlwa.2024.100597>
- [13]. Suhaimi, M. S. A., et al. (2025). Machine Learning-Based Stacking Ensemble Model for Prediction of Heart Disease with Explainable AI and K-Fold Cross-Validation. *Symmetry*, 17(2), 185. <https://doi.org/10.3390/sym17020185>
- [14]. Hamim, S. A., et al. (2025). Advanced Heart Attack Prediction Using a Stacked Ensemble Machine Learning Model and Diverse Data Integration. *International Journal of Intelligent Systems and Applications*, 17(5), 49–67. <https://doi.org/10.5815/ijisa.2025.05.04>
- [15]. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://doi.org/10.5555/1953048.2078195>