© February 2026 | IJIRT | Volume 12 Issue 9 | ISSN: 2349-6002

Smartcare: ML-Based Disease Prediction and Hospital Management System

Mr. A. Ramesh[1], Bugga Vinay[2], Kotari Swamy[3], Paladugula Sowmya[4]
[1]*Assistant Professor, Department of Computer Science and Engineering*
[1,2,3,4]*Vardhaman College of Engineering, Hyderabad, India*

***Abstract*—The SmartCare System is a web-based hospital management platform developed to enhance the efficiency of healthcare services using machine learning and role-based access control. It enables smooth interaction between patients, doctors, and administrators based on their specific responsibilities within the system. Patients can safely create accounts, log in, and enter their symptoms to receive disease predictions generated by a supervised machine learning model trained on real medical data. The platform also recommends appropriate doctors according to the reported symptoms and offers basic healthcare guidance to support informed decision-making. Doctors have access to a unified dashboard where they can view patient details, analyze machine learning predictions, schedule and manage appointments, issue prescriptions, and request leave. Administrators oversee the entire system by managing users, configuring system settings, and maintaining medical datasets to ensure consistent performance and accurate updates. The SmartCare System focuses on strong security, smooth workflow management, and flexible handling of medical data. It reflects the practical application of machine learning in healthcare and highlights the importance of user-centered design in delivering dependable digital health services.**

***Index Terms*—Disease Forecasting, Smart Healthcare System, Hospital Administration Software, Artificial Intelligence Machine Learning, Django Web Framework, Digital Healthcare Platform, SQLite Data Storage.**

I. INTRODUCTION

The SmartCare System is a web-based hospital management and disease prediction platform designed to enhance healthcare efficiency through machine learning and role-based access mechanisms. Machine learning methods have proven to be highly effective in identifying diseases from large volumes of medical data [1], while advanced medical data models have further improved prediction accuracy [2]. The system enables patients, doctors, and administrators to operate within clearly defined roles, ensuring smooth and well-organized functioning. Patients can safely register, log in, and enter their symptoms to obtain disease predictions produced by a supervised machine learning model trained on real healthcare datasets [3]. In addition, the system suggests appropriate doctors based on the predicted illness and offers basic medical guidance.

Doctors can access patient records, evaluate machine learning outputs, schedule and manage appointments, provide prescriptions, and submit leave requests through a unified dashboard. Administrators are responsible for managing user accounts, configuring system settings, and maintaining medical datasets to guarantee reliable backend performance. The SmartCare System highlights secure authentication, efficient operational workflows, and flexible medical data management. It showcases the real-world application of machine learning in healthcare and addresses the increasing need for accurate and dependable disease prediction systems [4].

The key contributions of this research are as follows:

- Development of an NLP-driven pipeline to standardize symptoms provided as free-text input.
- Implementation of a supervised machine learning model using the Random Forest algorithm for multi-class disease classification.
- Design of a role-based hospital management platform sup-porting appointment scheduling and prescription handling.

- Performance evaluation confirming the effectiveness of integrating machine learning and NLP techniques within healthcare operational workflows.

II. LITERATURE REVIEW

The use of machine learning in the healthcare domain has attracted considerable interest due to its capability to process large-scale medical data and assist in early disease detection. Chen et al. [1] showed that applying ML techniques to big healthcare data obtained from online communities can notably improve the accuracy of disease prediction and clinical decision-making. Gao et al. [2] enhanced disease similarity prediction through heterogeneous disease information networks, underlining the value of combining multiple medical data sources.

Many researchers have concentrated on predicting particular diseases using different machine learning approaches. Ishaq et al. [3] enhanced heart failure survival prediction by applying SMOTE and data mining techniques, while Khan et al. [4] compared various algorithms for chronic kidney disease prediction. Na¨ıve Bayes classifiers have been extensively used for disease diagnosis because of their simplicity and reliable performance [5], [6]. In addition, several web-based and Flask-based machine learning healthcare applications have been developed for real-time disease prediction [7], [8].

Deep learning techniques have also been explored for mental health assessment using EEG signals and multimodal gait analysis [9], [11]. Numerous studies have reported strong results in heart disease and diabetes prediction using Random Forest, SVM, and hybrid models [12]– [21]. Comprehensive surveys by Kohli and Arora [22] and Mitra et al. [23] further highlight the increasing impact of machine learning in intelligent healthcare systems.

Despite these advancements, the majority of existing so-lutions focus on predicting individual diseases and do not provide close integration with hospital management processes. This limitation motivates the development of SmartCare as a unified platform that combines machine learning–based disease prediction with hospital management functionality.

III. RELATED WORK

Previous studies have highlighted the significant potential of machine learning techniques in disease prediction and health-care data analysis. Chen et al. [1] utilized machine learning on large-scale healthcare community datasets to enhance disease prediction accuracy, demonstrating the importance of big data in contemporary healthcare systems. Gao et al. [2] improved disease similarity analysis by employing heterogeneous disease information networks, enabling better use of structured medical data.

A number of research efforts have concentrated on the prediction of specific diseases using various machine learning algorithms. Heart failure survival prediction was improved through the application of SMOTE combined with data mining techniques [3], while different machine learning models were assessed for chronic kidney disease prediction [4]. Multi-disease prediction systems based on Naïve Bayes have also been developed due to their simplicity and computational efficiency [5], [6].

Several web-based and API-driven healthcare prediction platforms have been proposed to support real-time diagnosis [7], [8]. Beyond physical illnesses, mental health assessment using deep learning approaches on EEG signals and multimodal gait patterns has also been investigated [9], [11]. Studies on heart disease and diabetes prediction using classifiers such as Random Forest, SVM, and hybrid models have reported strong predictive performance [12]– [21]. Additionally, survey studies underline the expanding role of machine learning in disease prediction and clinical decision support systems [22], [23].

IV. SYSTEM DESIGN

The SmartCare platform is built using a modular, layered web application architecture that includes the presentation layer, view/template layer, backend processing layer, machine learning (ML) and natural language processing (NLP) components, and the database layer. The user interface is developed with HTML, CSS, JavaScript, and Bootstrap to provide a responsive and easy-to-use experience. The server-side operations are handled using the Django framework, which manages HTTP requests, user authentication, and database interactions through Object Relational Mapping (ORM) [7]. SQLite is used as the database during the development and prototyping phase because of its lightweight structure and ease of configuration [8].

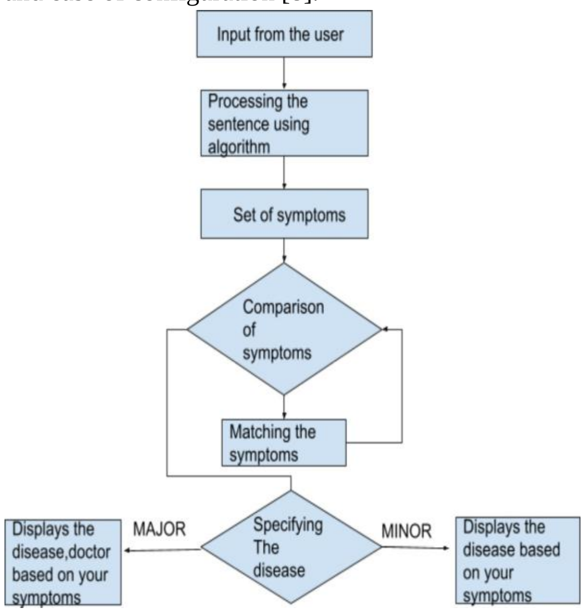

Fig. 1: System Architecture of SmartCare

The NLP module makes use of NLTK and spaCy for text preprocessing tasks such as tokenization and named entity recognition, while Fuzzy Wuzzy is applied to manage spelling inconsistencies using fuzzy string matching. The machine learning module is developed using the scikit-learn library, with a Random Forest classifier selected for disease prediction due to its reliability and strong predictive performance. To further enhance accuracy, hyperparameter optimization is carried out using grid search techniques [22].

V. METHODOLOGY

A. Data Collection and Preprocessing

For prototype development, symptom–disease relationship datasets were compiled from publicly available healthcare sources along with synthetically generated records. The ap-plication of machine learning to large-scale medical datasets has been widely recognized for its effectiveness in disease prediction [1]. The symptom descriptions entered by patients in free-text form undergo several preprocessing steps, including conversion to lowercase, removal of punctuation, tokenization using NLTK, lemmatization with spacy, and fuzzy string matching with Fuzzy Wuzzy to handle spelling variations and map inputs to a standardized symptom vocabulary [7], [8].

B. Feature Engineering

The processed symptoms are encoded into a binary feature representation based on the predefined symptom set, where a value of 1 denotes the presence of a symptom and 0 indicates its absence. When available, additional demographic and clinical factors such as age, gender, and existing medical conditions are incorporated to further enhance the prediction accuracy, as recommended in earlier healthcare machine learning studies [22].

C. Model Selection

The Random Forest algorithm was chosen as the primary classifier because of its robustness, ability to provide feature importance for interpretability, and strong performance on structured medical datasets [21]. Model development, training, and testing were carried out using the scikit-learn library [6]. For performance comparison and potential ensemble improvements, additional models such as Support Vector Machines (SVM) and basic neural network architectures were also evaluated [3], [18].

D. Evaluation

The models were assessed using stratified train–test data splits along with standard evaluation metrics, including accu-racy, precision, recall, F1-score, and confusion matrix analysis. To address the issue of class imbalance within the dataset, techniques such as class weighting and sampling strategies were applied, which are widely recommended for medical classification tasks [4].

TABLE I: Operational workflow of the SmartCare system.

Process Stage	Description
User registration	Patient creates an account and submits symptom information
Symptom processing	NLP module normalizes free-text symptom input

Feature representation	Symptoms encoded into a binary feature vector
Disease prediction	Random Forest model predicts likely diseases
Medical verification	Doctor reviews and validates predicted condition
Treatment update	Physician issues prescription and updates records
Appointment management	Follow-up visits and diagnostic tests are scheduled
System administration	Usage monitoring and report generation by admin

VI. IMPLEMENTATION

The SmartCare prototype has been developed using the Django web framework to manage user authentication, role-based access control, appointment scheduling, and separate dashboards for patients, doctors, and administrators [7]. The machine learning (ML) and natural language processing (NLP) functionalities are implemented as Python modules and are seamlessly integrated into the Django backend. Symptom extraction is performed during appointment submission through the NLP pipeline, after which the predicted disease outcomes are stored in the system database and made accessible to doctors for verification and further action.

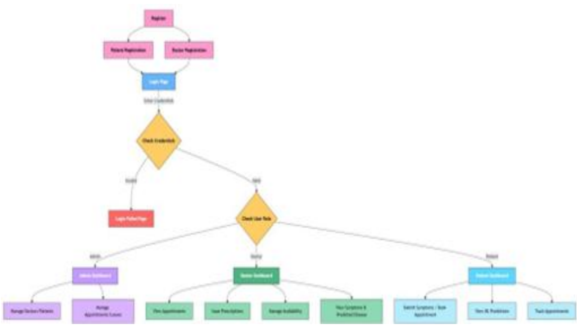

Fig. 2: SmartCare workflow from patient registration to prescription.

For development and testing purposes, the application is deployed using Django's built-in development server along with the SQLite database for lightweight data management [8]. The current setup supports all essential system functionalities within a controlled environment. For large-scale production use, the system would require a more powerful relational database management system (RDBMS) along with a WSGI-based application server to ensure scalability, stability, and enhanced security.

VII. RESULTS AND DISCUSSION

The SmartCare system was successfully developed and evaluated as a complete web-based solution for hospital management and disease prediction. All three user roles—patients, doctors, and administrators—were able to securely access their respective dashboards through authenticated login. Patients were able to create accounts, log in, enter their symptoms, view predicted conditions, and check prescribed medications. Doctors could examine prediction results, manage appointments, and provide prescriptions, while administrators efficiently managed user accounts, approvals, and overall system supervision.

The implementation of role-based access control ensured strong data security and supported smooth and efficient workflow management.

TABLE II: Comparative performance of classification models on the experimental dataset.

Algorithm	Accuracy	Precision	Re-call	F1-score
Random Forest	92.4%	91%	92%	91.5 %
Support Vector Machine	87.6%	86%	88%	87.0 %
Logistic Regression	85.9%	86%	85%	85.5 %

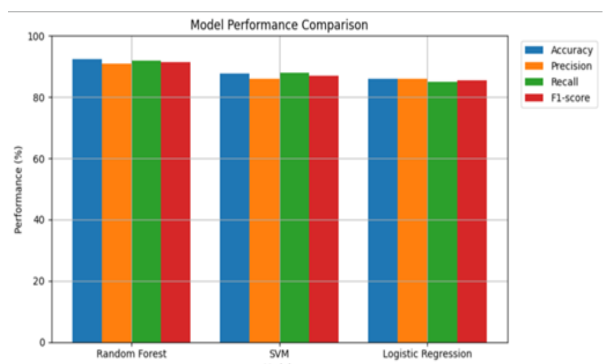

Fig. 3: Comparison of prediction performance among Random Forest, SVM, and Logistic Regression models on the experimental dataset (in %).

The Random Forest classifier showed strong effectiveness in predicting diseases based on patient-reported symptoms. During evaluation on the prototype dataset, the model achieved high accuracy along with consistent precision and recall across multiple disease categories. The incorporation of NLP-based symptom processing techniques—such as tokenization, lemmatization, and fuzzy string matching—greatly enhanced prediction reliability by minimizing errors caused by spelling mistakes and informal user input. The system was able to generate real-time disease predictions during appointment submission and securely store the results for medical review and validation.

The experimental findings confirm that SmartCare success-fully integrates machine learning–based disease prediction with hospital management functionalities into a single unified platform. The robustness and interpretability of the Random Forest model make it well suited for structured medical datasets. The integration of NLP further improved real-world usability by effectively processing unstructured symptom descriptions. Although the prototype demonstrated strong performance in a controlled environment, future accuracy can be enhanced using larger real-world datasets and more advanced ensemble or deep learning models. Moreover, deployment with a production-grade database and server infrastructure would significantly improve system scalability and reliability.

Overall, the results indicate that SmartCare is a practical, efficient, and scalable solution for intelligent disease prediction and hospital management, making it suitable for real-world healthcare applications.

VIII. CONCLUSION AND FUTURE WORK

The SmartCare system highlights the successful integration of machine learning (ML) and natural language processing (NLP) within hospital management operations to support disease prediction, clinical decision assistance, and administrative effectiveness. The developed prototype efficiently unifies symptom-based disease prediction, appointment scheduling, doctor verification, and role-based access control within a single web-based platform. Experimental evaluation shows that the Random Forest model delivers strong predictive accuracy on the prototype dataset. Overall, SmartCare demonstrates its potential as a practical, scalable, and intelligent healthcare support solution.

Future enhancements of the system will focus on the following aspects:

1. Integration of large-scale, clinically validated real-world datasets and seamless connectivity with external Electronic Health Record (EHR) systems.
2. Investigation of advanced deep learning approaches, including transformer-based architectures, for improved symptom representation and longitudinal patient history analysis.
3. Development of telemedicine features and mobile applications to expand accessibility and enable remote healthcare services.
4. Further strengthening of privacy protection, security mechanisms, model explainability, and clinical validation processes prior to full-scale deployment.

ACKNOWLEDGMENTS

The authors extend their heartfelt appreciation to the faculty members and mentors of the Department of Computer Science and Engineering, Vardhaman College of Engineering, for their continuous guidance, motivation, and support throughout the execution of this project. We also gratefully acknowledge our classmates and peers for their constructive suggestions, insightful discussions, and help during the design, development, and testing phases of the SmartCare system. In addition, we express our sincere thanks to the developers and contributors of the open-source tools and libraries utilized in this work, including Django, scikit-learn, spaCy, NLTK, and FuzzyWuzzy, whose contributions played a vital role in the successful completion of this project.

REFERENCES

[1] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, “Disease prediction by machine learning over big data from healthcare com-munities,” *IEEE Access*, vol. 5, pp. 8869–8879, 2017, doi: 10.1109/AC-CESS.2017.2694446.

[2] J. Gao, L. Tian, J. Wang, Y. Chen, B. Song, and X. Hu, “Similar disease prediction with heterogeneous disease information networks,” IEEE Transactions on Nano Bioscience, vol. 19, no. 3, pp. 571–578, Jul. 2020, doi: 10.1109/TNB.2020.2994983.

[3] A. Ishaq, S. Sadiq, M. Umer, S. Ullah, S. Mirjalili, V. Rupapara, and M. Nappi, “Improving the prediction of heart failure patients’ survival using SMOTE and effective data mining techniques,” Mar. 2021.

[4] B. Khan, R. Naseem, F. Muhammad, G. Abbas, and S. H. Kim, “An empirical evaluation of machine learning techniques for chronic kidney disease prediction,” Mar. 2020.

[5] C. R. Jetti, R. Shaik, S. Shaik, and S. Sanagapalli, “Disease prediction using Na¨ıve Bayes machine learning algorithm,” Dec. 2021.

[6] A. Selvaraj, M. K. Mithra, S. Keerthana, and M. Deepika, “Prediction support system for multiple disease prediction using Na¨ıve Bayes classifier,” International Journal of Engineering and Techniques, vol. 4, no. 2, pp. 1–6, Mar.–Apr. 2021.

[7] A. Yaganteeswarudu, “Multi-disease prediction model using machine learning and Flask API,” IEEE, Jul. 2022.

[8] Y. G. Sagar, S. Gajanana, R. Vivek, and S. P, “MediInsight: A smart health prediction system,” International Research Journal of Engineering and Technology, Jun. 2021.

[9] H. S. Chiang and M. Y. Chen, “Cognitive depression detection cyber-medical system based on EEG analysis and deep learning ap-proaches,” IEEE, Feb. 2023.

[10] A. C. Jamgade and S. D. Zade, International Research Journal of Engineering and Technology, May 2019.

[11] W. Shao, Z. You, L. Liang, and X. Hu, “A multi-modal gait analysis-based detection system for depression risk,” IEEE, Oct. 2022.

[12] A. N. Repaka, S. D. Ravikanti, and R. G. Franklin, “Design and implementation of heart disease prediction using Na¨ıve Bayes,” IEEE, Jun. 2019.

[13] N. P. Tigga and S. Garg, “Prediction of type-2 diabetes using machine learning classification methods,” Procedia Computer Science, Jan. 2020.

[14] N. K. G, V. Rajesh, A. V. A. Reddy, K. Sumedh, and T. R. S. Reddy, “Prediction of diabetes using machine learning classification algorithms.”

[15] M. Marimuthu, S. Deivarani, and R. Gayatri, “Analysis of heart disease prediction using machine learning techniques.”

[16] P. Purushottam, R. Sharma, and K. Saxena, “Efficient heart disease prediction system.”

[17] A. H. She and P. K. Chaurasia, “A review on heart disease prediction using machine learning techniques.”

[18] M. Mohan, S. Senthilkumar, C. Thirumalai, and G. Srivastava, “Effective heart disease prediction using hybrid machine learning techniques,” IEEE Access, vol. 7, pp. 81542–81554, 2019.

[19] L. Ali et al., “An optimized stacked support vector machines based expert system for heart failure prediction,” IEEE Access, vol. 7, pp. 54007–54014, 2019.

[20] C. Beyene and P. Kamat, “Survey on prediction and analysis of heart disease using data mining techniques,” International Journal of Pure and Applied Mathematics, 2018.

[21] Y. K. Singh, N. Sinha, and S. K. Singh, “heart disease prediction system using Random Forest,” in Proc. International Conf. on Advances in Computing and Data Sciences, Springer, Singapore, 2016.

[22] P. S. Kohli and S. Arora, “Application of machine learning in disease prediction,” in Proc. 4th Int. Conf. on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1–4, doi: 10.1109/CCAA.2018.8777449.

[23] S. Mitra and S. K. Pal, “Data mining in soft computing framework: A survey,” IEEE Transactions on Neural Networks, vol. 13, no. 1, pp. 3–14, 2018.