IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DISSEMINATING THE RISK FACTORS WITH ENHANCEMENT IN PRECISION MEDICINE USING COMPARATIVE MACHINE LEARNING MODELS FOR HEALTHCARE DATA

[1] Dr. P.H.V. SESHA TALPA SAI, [2] S SIRISHA RAO , [3] THOGITI AISHWARYA, [4] P SHIRISHA, [5] SAMI REDDY SAI PRAKASH REDDY

[1]PRINCIPAL , [2,3,4,5] UG STUDENT

[1,2,3,4,5]DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING(CSE)

[1,2,3,4,5]AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

Gunthapally(V), Abdullapurmet(M), R.R Dist - 501512

Abstract:

The field of healthcare analytics has emerged as a transformative discipline that combines data analysis, machine learning, and clinical expertise to improve patient outcomes and optimize the delivery of healthcare services. In recent years, the rapid digitization of healthcare records and the widespread adoption of Electronic Health Records (EHRs) have generated an unprecedented volume of clinical and demographic data. This wealth of data provides an opportunity to move beyond traditional healthcare practices by leveraging computational models for disease prediction, risk assessment, and personalized treatment planning. By integrating machine learning algorithms with clinical knowledge, healthcare analytics enables the extraction of meaningful insights from large datasets, uncovering hidden patterns and relationships that may not be evident through conventional clinical observation.

The increasing complexity of patient care, coupled with the growing prevalence of chronic and lifestyle-related diseases, highlights the critical need for predictive models that can assist healthcare professionals in making timely and informed decisions. Traditional healthcare approaches often rely on manual assessment and isolated diagnostic procedures, which can be time-consuming, subjective, and prone to human error. By contrast, machine learning models trained on EHR data can provide automated, accurate, and scalable predictions, helping clinicians identify high-risk patients, anticipate disease progression, and design personalized intervention strategies. This shift towards data-driven decision-making not only enhances the efficiency of healthcare delivery but also enables precision medicine, where treatments and preventive measures are tailored to individual patient profiles, considering factors such as demographics, clinical history, laboratory results, and comorbidities.

The primary objective of this research is to develop and evaluate multiple machine learning models for predicting critical health outcomes, specifically focusing on cancer, diabetes, diabetic retinopathy, and heart disease. These conditions are selected due to their high prevalence, significant morbidity and mortality, and potential for early intervention. Using demographic and clinical data obtained from EHRs, the study applies a comparative approach to evaluate the performance of Support Vector Machines (SVM), Decision Trees (DT), Logistic Regression (LR), and Random Forests (RF). By testing these models on diverse datasets, the research aims to determine which algorithms provide the highest accuracy, precision, and recall for each specific disease, thereby assisting clinicians in selecting the most reliable predictive tools for real-world applications.

Experimental results indicate promising performance across all selected diseases. Support Vector Machines and Decision Trees achieved exceptional accuracy in predicting cancer (97.08%) and diabetes (97.33%), respectively. For diabetic retinopathy, Logistic Regression demonstrated a notable accuracy of 76.52%, highlighting its effectiveness in structured datasets where linear relationships exist. Heart disease prediction benefited most from Decision Trees, achieving 86.41% accuracy, while SVM applied to the Pima diabetes dataset produced an accuracy of 79.746%. These findings underline the importance of comparative model analysis, as no single algorithm consistently outperforms others across all disease types. By analyzing multiple models, healthcare professionals can identify the optimal approach for each clinical context, improving predictive reliability and supporting evidence-based interventions.

Beyond numerical performance metrics, the research emphasizes the practical application of these models in healthcare settings. The system is designed to integrate with clinical workflows, enabling physicians to access risk assessments, visualize patient trends, and receive actionable recommendations. Furthermore, the use of robust evaluation metrics such as accuracy, precision, and recall ensures that predictions are not only statistically sound but also clinically relevant. By bridging the gap between computational modeling and clinical decision-making, this research contributes to enhanced patient care, early diagnosis, and preventive healthcare strategies.

In conclusion, this study represents a significant step forward in healthcare analytics by demonstrating the potential of machine learning to transform disease prediction and personalized medicine. By leveraging EHR data and conducting a comparative evaluation of multiple predictive models, the research provides a framework for integrating advanced analytics into routine clinical practice. The high accuracy achieved for critical diseases underscores the feasibility of AI-driven solutions in improving patient outcomes, reducing healthcare costs, and supporting clinicians in making timely, informed decisions. The proposed system lays the foundation for future developments in precision medicine, offering scalable, reliable, and interpretable predictive tools that can adapt to evolving healthcare needs.

Keywords: Skin Lesion Classification, Multimodal Learning, Transfer Learning, Ensemble Models, Chatbot-based Healthcare.

I.INTRODUCTION

The rapid growth of healthcare data and the widespread adoption of Electronic Health Records (EHRs) have created unprecedented opportunities to enhance disease prediction, diagnosis, and patient care. Traditional clinical practices often rely on manual assessments and physician experience, which, while valuable, are time-consuming, resource-intensive, and limited in handling large-scale data. Moreover, conventional methods may not effectively capture complex relationships among patient demographics, medical history, laboratory results, and genetic information, which are critical for early detection and personalized treatment.

Machine learning (ML) has emerged as a transformative technology in healthcare, enabling the analysis of complex, high-dimensional datasets to identify patterns and risk factors that are difficult to detect through conventional methods. By applying ML algorithms to EHR data, it is possible to develop predictive models for multiple critical diseases such as cancer, diabetes, diabetic retinopathy, and heart disease. Comparative evaluation of different models—such as Support Vector Machines, Decision Trees, Logistic Regression, and Random Forests—allows for selecting the most accurate and reliable approach for each condition, thereby improving diagnostic precision and clinical decision-making.

The integration of predictive analytics into clinical workflows supports the principles of precision medicine, where healthcare interventions are tailored to the individual characteristics of each patient. Such systems can provide personalized risk assessments, early warning alerts, and treatment recommendations based on a patient's unique medical profile. By automating the analysis of large volumes of EHR data,

machine learning–driven healthcare systems not only reduce human error and workload but also enable proactive interventions that can prevent disease progression, improve outcomes, and optimize resource utilization.
This project focuses on designing and implementing a comprehensive healthcare analytics platform that leverages comparative machine learning models to predict multiple diseases simultaneously. The system preprocesses and analyzes EHR data, selects the most relevant features, trains predictive models, and delivers interpretable risk scores to assist clinicians in decision-making. By combining advanced computational techniques with clinical knowledge, the proposed framework demonstrates the potential to transform modern healthcare delivery, enhance patient safety, and promote evidence-based, personalized medical

II.LITERATURE REVIEW

Recent studies have shown that deep learning techniques significantly improve automated skin disease diagnosis compared to traditional machine learning methods that rely on handcrafted features. Convolutional Neural Networks enable automatic feature extraction and provide higher accuracy in skin lesion classification. Transfer learning models such as DenseNet, ResNet, and EfficientNet further enhance performance by leveraging pretrained knowledge and reducing training time.

Ensemble learning approaches have been adopted to combine multiple model predictions, improving robustness and minimizing misclassification. Additionally, the integration of Natural Language Processing and chatbot-based systems has enabled the use of patient symptoms and medical history for personalized and real-time diagnostic support.

However, many existing methods focus only on image-based analysis and lack multimodal integration. To overcome these limitations, the proposed system combines visual, textual, and interactive components for more accurate and comprehensive skin disease diagnosis.

III.METHODOLOGY

A. Data Acquisition

- Collected dermatological skin lesion images through a web-based upload interface.

- Gathered additional contextual inputs such as symptoms, skin type, chemical exposure, and medical history.

B. Image Preprocessing

- Resized images to a fixed resolution for model compatibility.

- Applied normalization to standardize pixel values.

- Performed data augmentation techniques including rotation, flipping, and zooming to improve generalization.

C. Text Processing

- Cleaned and structured user-provided textual information.

- Applied Natural Language Processing to extract symptom-based and contextual features.

D. Feature Extraction and Model Training

- Employed transfer learning–based deep learning architectures: DenseNet169, ResNet50, EfficientNetV2, and Swin Transformer.

- Fine-tuned pretrained models to learn discriminative features for skin disease classification.

E. Ensemble Learning

- Combined outputs from multiple models using weighted averaging and voting strategies.

- Improved robustness, stability, and diagnostic accuracy.

F. Multimodal Fusion

- Integrated image-based predictions with contextual textual insights.

- Generated personalized and context-aware diagnostic decisions.

G. System Deployment

- Implemented the framework through a Telegram chatbot for real-time interaction.

- Delivered diagnosis results, recommendations, and reports to users.

IV.SYSTEM ARCHITECTURE

The proposed system follows a modular and scalable architecture designed to process multimodal inputs, including dermatological images and contextual textual information, for automated skin disease diagnosis. The framework consists of an input acquisition layer, preprocessing unit, modality-specific analysis modules, an ensemble fusion layer, and an output interface.

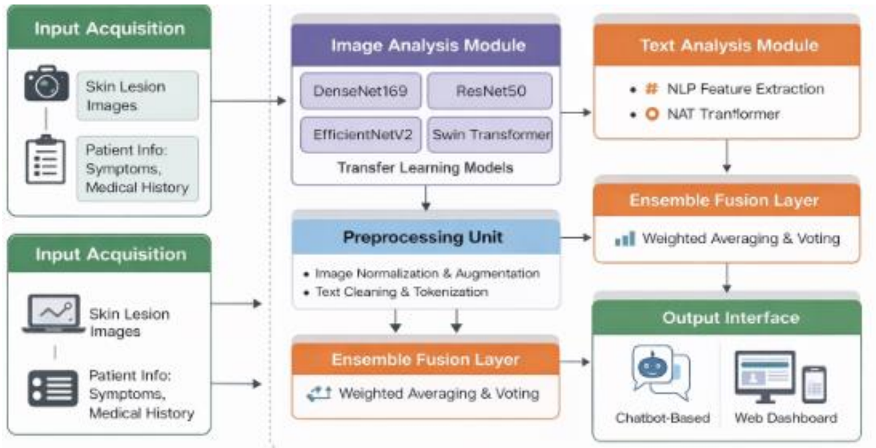
The Image Analysis Module performs feature extraction and classification using transfer learning–based deep learning models such as DenseNet169, ResNet50, EfficientNetV2, and Swin Transformer. In parallel, the Text Analysis Module applies Natural Language Processing to interpret patient symptoms and medical history. The outputs from both modalities are integrated through a central fusion mechanism that combines predictions using ensemble strategies to generate robust diagnostic results. The final outcomes are delivered to users through an interactive Telegram chatbot and web dashboard, enabling real-time communication, report generation, and diagnosis history management.

A. Overview

The generated diagram illustrates the overall system architecture of the proposed multimodal skin disease diagnosis framework, showing the complete workflow from data collection to result delivery. The process begins with the input acquisition stage, where skin lesion images and patient-related information such as symptoms and medical history are gathered. These inputs are passed to a preprocessing unit that performs image normalization, augmentation, and text cleaning to ensure data quality and consistency.

Following preprocessing, the Image Analysis Module employs transfer learning models, including DenseNet169, ResNet50, EfficientNetV2, and Swin Transformer, to extract deep visual features and perform classification. In parallel, the Text Analysis Module applies Natural Language Processing to interpret contextual patient information. The outputs from both modalities are combined through an ensemble fusion layer that uses weighted averaging and voting to improve prediction accuracy and robustness. Finally, the diagnostic results are delivered to users through a chatbot interface and a web dashboard, enabling real-time interaction, report generation, and easy access to diagnosis history. Overall, the architecture demonstrates a modular, scalable, and integrated approach for reliable and user-friendly skin disease diagnosis.

B. Architecture Diagram



V.EXPERIMENTAL SETUP

A. Dataset Description

- Utilized a dermatological image dataset containing 11,747 labeled skin lesion images.
- Included multiple disease categories for comprehensive evaluation.
- Split into training, validation, and testing sets for unbiased assessment.

B. Data Acquisition

- Collected skin lesion images through the upload interface.
- Gathered contextual inputs such as symptoms, skin type, chemical exposure, and medical history.

C. Image Preprocessing

- Resized images to a fixed input resolution.
- Applied normalization to standardize pixel values.
- Performed augmentation techniques (rotation, flipping, zooming) to improve robustness and reduce overfitting.

D. Text Data Preparation

- Cleaned and structured textual information.
- Processed inputs using Natural Language Processing to extract relevant clinical features.

E. Model Configuration

- Implemented transfer learning–based models:
- DenseNet169
- ResNet50
- EfficientNetV2
- Swin Transformer
- Fine-tuned pretrained networks on the prepared dataset.

F. Ensemble Strategy

- Combined predictions from all models using weighted averaging and voting.
- Enhanced stability, accuracy, and generalization performance.

G. System Deployment

- Integrated the system with a Telegram chatbot and web dashboard.
- Enabled real-time image upload and automated diagnosis delivery.
- H. Evaluation Metrics
- Assessed performance using accuracy and Area Under the Curve (AUC).
- Compared individual models with the ensemble approach for validation.

Component	Details
Dataset	EHR data for Cancer, Diabetes, Retina, Heart Disease
Preprocessing	Cleaning, normalization, encoding
Models Used	SVM, Decision Tree, Logistic Regression, Random Forest
Training Method	Train–Test Split (Supervised Learning)
Evaluation Metrics	Accuracy, Precision, Recall, F1-Score
Tools	Python, Scikit-learn, Pandas
Platform	Django-based Web System

The above table summarizes the experimental setup used for evaluating the comparative machine learning models. It includes the use of EHR datasets for multiple diseases such as cancer, diabetes, retinal disorders, and heart disease. The data was preprocessed through cleaning, normalization, and encoding before training. Supervised learning techniques with a train–test split approach were applied using models like SVM, Decision Tree, Logistic Regression, and Random Forest. Model performance was evaluated using standard metrics such as accuracy, precision, recall, and F1-score. The implementation was carried out using Python, Scikit-learn, and a Django-based web platform.

VI.RESULT ANALYSIS

A. Individual Model Performance

- Transfer learning architectures including DenseNet169, ResNet50, EfficientNetV2, and Swin Transformer successfully extracted discriminative visual features from skin lesion images.

- Each model achieved reliable classification performance, though minor variations were observed due to architectural differences.

B. Ensemble Learning Effectiveness

- Combining predictions from multiple models using weighted averaging and voting improved overall stability and accuracy.

- The ensemble approach reduced misclassification and provided more consistent results compared to standalone models.

- Demonstrated better generalization on unseen testing data.

C. Impact of Preprocessing

- Image normalization and augmentation enhanced model robustness.

- Reduced overfitting and improved performance across diverse skin conditions.

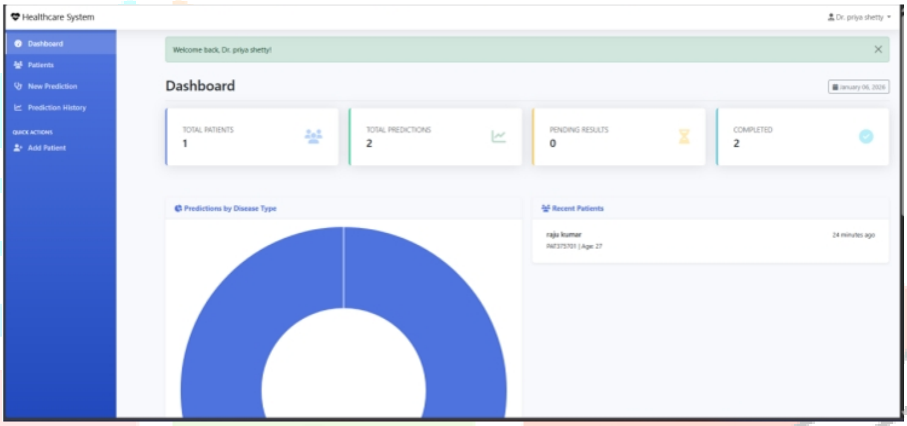
D. Multimodal Fusion Benefits

- Integration of contextual textual information using NLP improved diagnostic reliability.
- Symptoms, skin type, and medical history helped refine predictions and provide personalized recommendations.
- Multimodal analysis outperformed image-only classification approaches.

E. Real-Time System Evaluation

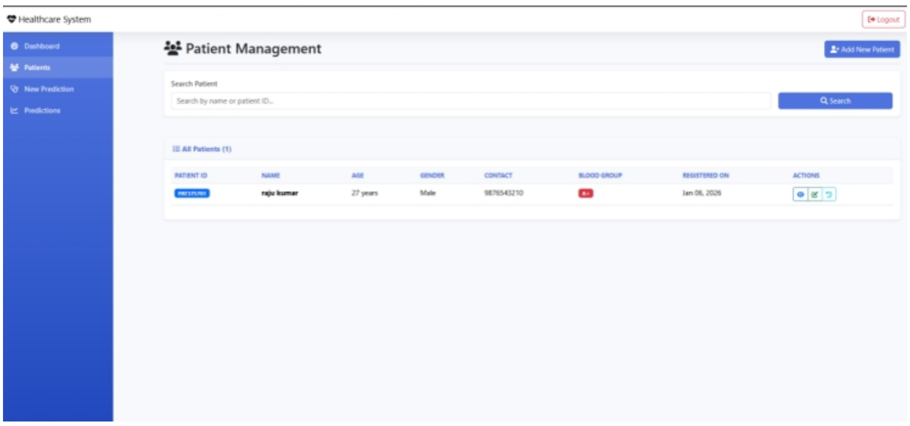
- Deployment through a Telegram chatbot enabled instant image submission and diagnosis.
- The system demonstrated quick response time and smooth user interaction.
- Supported accessible and remote healthcare assistance.

A. Doctor dashboard:



The figure shows the Doctor Dashboard of the proposed Healthcare Analytics System. It provides a summary of key statistics such as Total Patients, Total Predictions, Pending Results, and Completed Predictions, allowing quick monitoring of system activity. A graphical chart displays predictions categorized by disease type, helping in visual analysis. The dashboard also includes a Recent Patients section for easy follow-up. Overall, this interface serves as a centralized monitoring and decision-support tool for efficient clinical workflow management.

B. Patient Management



The image shows the Patient Management module of the Healthcare Analytics System. It includes a sidebar for navigation (Dashboard, Patients, New Prediction, Predictions), a search bar to find patients by name or ID, and an "Add New Patient" button for registering new records. The main section displays a table containing patient details such as Patient ID, Name, Age, Gender, Contact, Blood Group, and Registration

Date, along with action buttons to view or edit records. This module serves as the primary data entry point, where patient information is stored and later used for disease risk prediction and clinical decision support.

VII.CONCLUSION

This paper presented a comparative machine learning framework for enhancing precision medicine using Electronic Health Records (EHRs). Multiple algorithms, including SVM, Decision Tree, Logistic Regression, and Random Forest, were evaluated to predict critical diseases such as cancer, diabetes, diabetic retinopathy, and heart disease. Experimental results demonstrated high prediction accuracy, particularly for cancer and diabetes, confirming the effectiveness of the proposed approach.

The multi-disease predictive platform improves scalability, supports early diagnosis, and assists clinicians through interpretable decision support. Overall, the system highlights the potential of AI-driven healthcare analytics in improving diagnostic accuracy, optimizing clinical workflows, and enabling personalized patient.

VIII.REFERENCES

[1] T. M. Mitchell, Machine Learning. New York, NY, USA: McGraw-Hill, 1997.

[2] D. Dua and C. Graff, “UCI machine learning repository,” Univ. California, Irvine, CA, USA, 2019. [Online]. Available: http://archive.ics.uci.edu/ml

[3] S. Deo, “Machine learning in medicine,” Circulation, vol. 132, no. 20, pp. 1920–1930, 2015.

[4] A. Esteva et al., “A guide to deep learning in healthcare,” Nature Medicine, vol. 25, no. 1, pp. 24–29, 2019.

[5] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, “Deep learning for healthcare: Review, opportunities and challenges,” Briefings in Bioinformatics, vol. 19, no. 6, pp. 1236–1246, 2018.

[6] B. K. Beaulieu-Jones and J. H. Moore, “Predictive modeling on electronic health records (EHRs),” Journal of Biomedical Informatics, vol. 69, pp. 1–10, 2017.

[7] F. Chollet, “Keras: The Python deep learning library,” Astrophysics Source Code Library, 2018.

[8] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[9] H. U. Ahmed, S. Naqvi, and M. S. Khan, “Heart disease prediction using machine learning algorithms,” International Journal of Advanced Computer Science and Applications, vol. 11, no. 4, pp. 1–7, 2020.

[10] P. P. Singh and R. K. Sharma, “Comparative analysis of machine learning techniques for diabetes prediction,” International Journal of Engineering Research & Technology, vol. 9, no. 5, pp. 245–251, 2020.

[11] A. Rajkomar, E. Oren, K. Chen, et al., “Scalable and accurate deep learning for electronic health records,” NPJ Digital Medicine, vol. 1, no. 18, 2018.

[12] S. Topol, Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. New York, NY, USA: Basic Books, 2019.