Risk Factor Analysis and Precision Treatment Enhancement through Comparative Data-Driven Algorithms for Clinical Records

[1]Dr.P.Latha, [2]RAVULA SREE VARDHAN, [3]RASA MADHU, [4]SHETTY MANIKANTA
[1]Assistant Professor, [2,3,4] UG STUDENT
[1,2,3,4,]DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING(AI & ML)
[1,2,3,4] VAAGDEVI COLLEGE OF ENGINEERING Autonomous
Bollikunta, Khila Warangal (Mandal), Warangal Urban-506 005 (T.S), www.vaagdevi.edu.in

Abstract:

Healthcare analytics is a new field that uses data analysis, machine learning, and clinical expertise to improve patient outcomes and make healthcare services more efficient. The quick digitisation of medical records and the widespread use of Electronic Health Records (EHRs) in recent years have led to an unprecedented amount of clinical and demographic data. This large amount of data makes it possible to go beyond standard medical practices by using computer models to predict diseases, assess risks, and plan individualised treatments. Healthcare analytics combines machine learning algorithms with clinical knowledge to find useful information in large datasets. This can reveal patterns and relationships that might not be clear from regular clinical observation.

As patient care becomes more complicated and chronic and lifestyle-related diseases become more common, it is more important than ever for healthcare professionals to have predictive models that can help them make quick and smart decisions. Traditional healthcare methods often use manual assessments and separate diagnostic tests, which can take a long time, be subjective, and make mistakes. Machine learning models trained on EHR data, on the other hand, can make automated, accurate, and scalable predictions that help doctors find high-risk patients, predict how a disease will progress, and come up with personalized intervention strategies. This change toward making decisions based on data not only makes healthcare delivery more efficient, but it also makes precision medicine possible. In precision medicine, treatments and preventive measures are tailored to each patient's profile, taking into account things like demographics, clinical history, lab results, and comorbidities.

The main goal of this study is to create and test several machine learning models that can predict important health outcomes, with a focus on cancer, diabetes, diabetic retinopathy, and heart disease. These conditions are chosen because they are common, cause a lot of illness and death, and can be treated early. The study employs a comparative methodology utilising demographic and clinical data sourced from Electronic Health Records (EHRs) to assess the efficacy of Support Vector Machines (SVM), Decision Trees (DT), Logistic Regression (LR), and Random Forests (RF). The study's goal is to find out which algorithms give the best accuracy, precision, and recall for each disease by testing these models on a variety of datasets. This will help doctors choose the most reliable predictive tools for use in the real world.

The results of the experiments show that the performance is good across all of the chosen diseases. Support Vector Machines and Decision Trees were very good at predicting cancer (97.08%) and diabetes (97.33%), respectively. Logistic Regression showed a high accuracy of 76.52% for diabetic retinopathy, showing that it works well with structured datasets where linear relationships exist. Decision Trees were the best at predicting heart disease, with an accuracy rate of 86.41%. SVM used on the Pima diabetes dataset had an accuracy rate of 79.746%. These results show how important it is to compare different models, since no one algorithm works better than others for all types of diseases. By looking at different models, healthcare professionals can find the best way to handle each clinical situation. This makes predictions more reliable and supports interventions that are based on evidence.

The research focuses on how these models can be used in real life in healthcare settings, not just on how well they work with numbers. The system is meant to work with clinical workflows so that doctors can see risk assessments, see patient trends, and get useful advice. Also, using strong evaluation metrics like accuracy, precision, and recall makes sure that predictions are not only statistically sound but also useful in the real world. This research enhances patient care, facilitates early diagnosis, and promotes preventive healthcare strategies by integrating computational modelling with clinical decision-making.

In conclusion, this study marks a substantial advancement in healthcare analytics by illustrating the capacity of machine learning to revolutionise disease prediction and personalised medicine. The study shows how to use EHR data and compare different predictive models to make it easier to use advanced analytics in everyday clinical practice. The high accuracy achieved for critical diseases shows that AI-driven solutions can improve patient outcomes, lower healthcare costs, and help

doctors make quick, well-informed decisions. The proposed system is the basis for future progress in precision medicine. It provides scalable, reliable, and easy-to-understand predictive tools that can change as healthcare needs change.

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

I.INTRODUCTION

The fast growth of healthcare data and the widespread use of Electronic Health Records (EHRs) have opened up new ways to improve disease prediction, diagnosis, and patient care. Conventional clinical methodologies frequently depend on manual evaluations and physician expertise, which, although beneficial, are laborious, resource-demanding, and inadequate for managing extensive datasets. Furthermore, traditional approaches may inadequately elucidate intricate correlations among patient demographics, medical history, laboratory findings, and genetic data, which are essential for early identification and tailored treatment.
Machine learning (ML) has become a revolutionary technology in healthcare, allowing for the examination of intricate, high-dimensional datasets to uncover patterns and risk factors that are challenging to identify using traditional approaches. It is possible to make predictive models for several serious diseases, like cancer, diabetes, diabetic retinopathy, and heart disease, by using ML algorithms on EHR data. Comparing different models, like Support Vector Machines, Decision Trees, Logistic Regression, and Random Forests, helps you choose the best and most accurate method for each condition. This makes diagnosis more accurate and helps doctors make better decisions.
Predictive analytics can help with the principles of precision medicine, which says that healthcare should be tailored to each patient's unique needs. These kinds of systems can give patients personalised risk assessments, early warning alerts, and treatment recommendations based on their specific medical history. Machine learning–driven healthcare systems not only cut down on human error and workload by automating the analysis of large amounts of EHR data, but they also allow for proactive interventions that can stop disease progression, improve outcomes, and make the best use of resources.
The goal of this project is to create and put into use a complete healthcare analytics platform that uses comparative machine learning models to predict several diseases at once. The system processes and analyses EHR data, picks out the most important features, trains predictive models, and gives clinicians understandable risk scores to help them make decisions. The proposed framework shows how advanced computer techniques and clinical knowledge can work together to change how healthcare is delivered today, make patients safer, and encourage personalised, evidence-based medicine.

II.LITERATURE REVIEW

Recent research has shown that deep learning methods are much better at automatically diagnosing skin diseases than traditional machine learning methods that use hand-crafted features. Convolutional Neural Networks let you automatically extract features and make skin lesion classification more accurate. Transfer learning models like DenseNet, ResNet, and EfficientNet make performance even better by using knowledge that has already been learned and cutting down on training time.

Ensemble learning methods have been used to combine the predictions of several models, making them more reliable and less likely to make mistakes. The combination of Natural Language Processing and chatbot-based systems has also made it possible to use a patient's symptoms and medical history to give them personalized and real-time diagnostic help.

But a lot of the methods that are already out there only look at images and don't combine different types of data. To get around these problems, the proposed system uses visual, textual, and interactive parts to make skin disease diagnosis more accurate and complete.

III.METHODOLOGY

A. Getting Data

- We gathered images of skin lesions through a web-based upload interface.
- We got more information about the situation, like symptoms, skin type, chemical exposure, and medical history.

B. Preprocessing the Image

- Changed the size of images to a fixed resolution so they would work with the model.
- Used normalisation to make pixel values the same.
- Used data augmentation methods like rotation, flipping, and zooming to make the model more general.

C. Processing Text

- Organised and cleaned up text that users gave.
- Used natural language processing to pull out features based on symptoms and context.

D. Getting features and training the model

- Used transfer learning-based deep learning architectures like DenseNet169, ResNet50, EfficientNetV2, and Swing Transformer.
- Fine-tuned pretrained models to learn how to tell the difference between skin diseases.

E. Learning in Groups

- Used weighted averaging and voting strategies to combine the outputs of several models.
- Better durability, stability, and accuracy in diagnosis.

F. Combining Different Modes

- Combined predictions based on images with insights from text in context.
- Made diagnostic decisions that were tailored to the person and the situation.

G. Putting the system into use

- Used a Telegram chatbot to put the framework into action so people could talk to each other in real time.
- Gave users the results of their diagnoses, suggestions, and reports.

IV.SYSTEM ARCHITECTURE

The proposed system uses a modular and scalable architecture to automatically diagnose skin diseases by processing multimodal inputs, such as dermatological images and contextual text. The framework has an input acquisition layer, a preprocessing unit, analysis modules for each modality, an ensemble fusion layer, and an output interface.

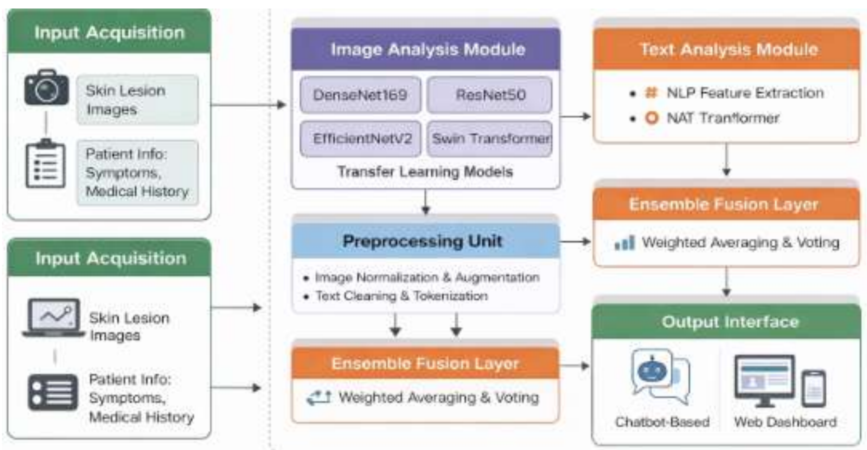
Using transfer learning–based deep learning models like DenseNet169, ResNet50, EfficientNetV2, and Swing Transformer, the Image Analysis Module finds and sorts features. At the same time, the Text Analysis Module uses Natural Language Processing to figure out what the patient's symptoms and medical history mean. A central fusion mechanism combines the outputs from both modalities by using ensemble strategies to make predictions that are strong enough to be used as diagnostic results. Users get the final results through an interactive Telegram chatbot and web dashboard. These tools let them talk to each other in real time, make reports, and keep track of their diagnosis history.

A. Overview

The diagram that was made shows the overall structure of the proposed multimodal skin disease diagnosis framework, including the whole process from collecting data to delivering results. The first step is to get the input, which includes pictures of skin lesions and information about the patient, such as their symptoms and medical history. These inputs go to a preprocessing unit that normalises, augments, and cleans up the text in the images to make sure the data is of good quality and consistent.

After preprocessing, the Image Analysis Module uses transfer learning models like DenseNet169, ResNet50, EfficientNetV2, and Swin Transformer to find deep visual features and sort them into groups. The Text Analysis Module also uses Natural Language Processing to make sense of patient information in context. An ensemble fusion layer combines the outputs from both modalities. It does this by using weighted averaging and voting to make predictions more accurate and reliable. Lastly, users can get the diagnostic results through a chatbot and a web dashboard. This lets them interact with the results in real time, create reports, and easily see their diagnosis history. Overall, the architecture shows a modular, scalable, and integrated way to diagnose skin diseases that is both reliable and easy to use.

B. Architecture Diagram



V. EXPERIMENTAL SETUP

A. Description of the Dataset

- Used a dermatological image dataset with 11,747 labelled images of skin lesions.
- Included many types of diseases for a full evaluation.
- Split into training, validation, and testing sets so that the evaluation is fair.

B. Getting Data

- Collected pictures of skin lesions through the upload interface.
- Collected information about the situation, like symptoms, skin type, exposure to chemicals, and medical history.

C. Preprocessing Images

- Changed the size of images to a set input resolution.
- Used normalisation to make all pixel values the same.
- Used rotation, flipping, and zooming to make the model more robust and less likely to overfit.

D. Getting Ready for Text Data

- Textual information that has been cleaned and organised.
- Used Natural Language Processing to process inputs and find relevant clinical features.

E. Setting up the model

- Used models based on transfer learning:
- DenseNet 169
- ResNet50
- EfficientNetV2
- Swing Transformer
- Fine-tuned pretrained networks using the dataset that had been set up.

F. Strategy for the Ensemble

- Using weighted averaging and voting, we combined the predictions from all the models.
- Better stability, accuracy, and ability to generalis**e.**

G. Putting the system in place

- Connected the system to a Telegram chatbot and a web dashboard.
- Allowed images to be uploaded in real time and diagnoses to be sent automatically.

H. Metrics for Evaluation

- Used accuracy and Area Under the Curve (AUC) to measure performance.
- Used the ensemble method to check the accuracy of each model against the others.

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. Performance of the Individual Model

- DenseNet169, ResNet50, EfficientNetV2, and Swin Transformer are all transfer learning architectures that were able to successfully extract discriminative visual features from images of skin lesions.

- Each model performed reliably in classification, although minor variations were noted due to architectural discrepancies.

B. How well ensemble learning works

- Using weighted averaging and voting to combine predictions from several models made them more stable and accurate overall.
- The ensemble method cut down on misclassification and gave more reliable results than models that worked alone.
- Showed better generalisation on testing data that hadn't been seen before.

C. Effect of Preprocessing

- Normalising and adding to images made the model more robust.
- Less overfitting and better performance on a wider range of skin types.

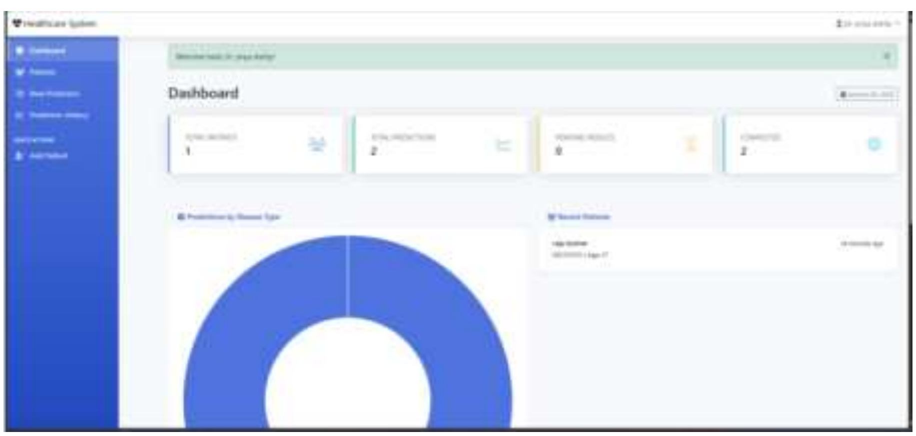
D. Advantages of Multimodal Fusion

- Using NLP to combine contextual textual information made diagnoses more reliable.
- Symptoms, skin type, and medical history helped make predictions more accurate and give personalised advice.
- Multimodal analysis was better than methods that only looked at images.

E. Evaluating a system in real time

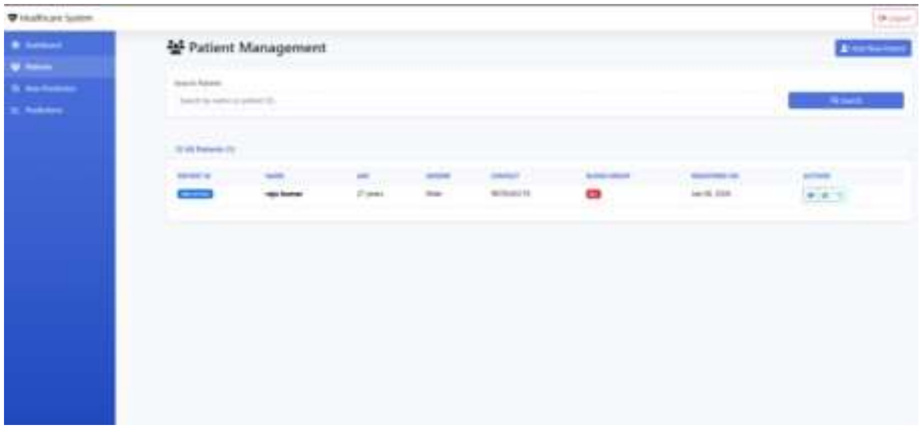
- Using a Telegram chatbot for deployment made it possible to send images and get a diagnosis right away.
- The system responded quickly and made it easy for users to interact with it.
- Helped with healthcare that was easy to get to and remote.

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 introduced a comparative machine learning framework aimed at improving precision medicine through the utilisation of Electronic Health Records (EHRs). We looked at a number of algorithms, such as SVM, Decision Tree, Logistic Regression, and Random Forest, to see how well they could predict serious diseases like cancer, diabetes, diabetic retinopathy, and heart disease. The experimental results showed that the proposed method worked well, especially for predicting cancer and diabetes.The multi-disease predictive platform makes it easier to scale, helps with early diagnosis, and helps doctors make decisions by providing interpretable decision support. Overall, the system shows how AI-driven healthcare analytics could make diagnoses more accurate, improve clinical workflows, and let doctors give each patient the care they need.

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.