MediMind: A Comprehensive Health Prediction and Record-Keeping Platform

Darshana A. Naik†, Aditya Sharma†, Danish Mahajan†, Dhruv Khandelwal†, Arpit Gajya†

Department of Computer Science and Engineering, Ramaiah Institute of Technology, Bangalore.

Contributing authors: darshananaik@msrit.edu; adityaofficial0501@gmail.com; danishmahajan779@gmail.com; dhruv15.khandelwal@gmail.com; arpitgajya2001@gmail.com;
†These authors contributed equally to this work.

Abstract

The "MediMind" project signifies a revolutionary fusion of predictive analytics, centered on diabetes and heart disease risk assessment, coupled with state-of-the-art skin disease prediction and robust Electronic Health Record (EHR) management. By employing advanced machine learning models like Logistic Regression and Support Vector Machines (SVM) for diabetes and heart disease predictions, alongside the cutting-edge Convolutional Neural Network (ResNet-50) for precise skin disease classification, "MediMind" achieves unparalleled accuracy and applicability. The predictive models exhibit exceptional performance metrics, boasting high accuracy, precision, recall, and F1-score. Seamless integration into Electronic Health Records (EHR) ensures immediate implementation in clinical decision-making, providing healthcare professionals with precise insights into diabetes and heart disease risks.

Keywords: Convolutional Neural Network (ResNet-50), Predictive Analytics, Diabetes Prediction, Heart Disease Risk Assessment, Skin Disease Classification,Logistic Regression,Support Vector Machines (SVM)

1 Introduction

The healthcare sector has witnessed a significant transformation with the introduction of technology. This shift has brought forth a new era in patient care and diagnosis, revolutionizing the way healthcare is delivered. In today's digital age, advanced technologies have not only expedited the diagnostic process but also facilitated the efficient management of patients' health records. This research paper explores a groundbreaking project that aims to diagnose various skin diseases like eczema, melanoma, and atopic dermatitis, as well as detect heart diseases. This multifaceted project combines cutting-edge medical diagnostics with robust data management systems to create a comprehensive healthcare solution. The integration of these innovative approaches enhances patient care while fostering data-driven decision-making.

Skin diseases, like eczema and melanoma, can greatly impact a person's quality of life, while heart diseases remain a top cause of death worldwide. Detecting these conditions early and accurately diagnosing them is crucial for effective treatment and better patient outcomes. This research paper discusses a project that tackles these healthcare challenges by combining advanced imaging techniques, and artificial intelligence to provide precise and timely diagnoses.

Along with its diagnostic capabilities, this project also includes an advanced health data management system. By integrating electronic health records and patient-specific information, healthcare professionals can easily access, analyze, and monitor patient health data. This promotes informed decision-making and allows for continuity of care. Additionally, centralizing patient health records reduces errors and enhances the overall quality of healthcare services.

This research introduces a novel and distinctive contribution to the realm of healthcare by seamlessly integrating advanced imaging techniques, artificial intelligence, and comprehensive data management. The novelty lies in the unprecedented combination of these elements to enhance diagnostic accuracy for skin and heart conditions. The project's significant contribution extends beyond traditional approaches, offering a pioneering model that not only improves diagnostic precision but also revolutionizes patient care through streamlined access to electronic health records and personalized health information. By addressing critical aspects of both diagnostics and data management, this initiative significantly advances the discourse in medical research and contributes to shaping the future of healthcare practices, fostering a new era of integrated and data-driven patient care.

2 RELATED WORK

[1] Skin disease prediction using deep residual learning and a novel cytological taxonomy based on pathology and cytology. [2] Using ensembles of deep learning networks for the detection and classification of skin diseases in medical imaging [3] This study uses an automatic Grabcut segmentation technique for skin lesion detection along with a digital hair removal system. In order to maximise skin disease classification, key features are extracted using GLCM and statistical parameters, and the performance of three machine learning algorithms is assessed using the ISIC 2019 and HAM 10000 datasets. [4] A novel approach to the analysis of skin diseases has been put forth,

which combines six distinct data mining techniques into an ensemble approach that functions as a single data mining technique. When compared to other classifier algorithms, the ensemble method on the dermatology dataset yields better results with 94% accuracy, making it more effective in this domain.

[6] This work uses patient data on key health factors to present several machine learning approaches for heart disease prediction. In order to construct the prediction models, the study presented four classification techniques: Multilayer Perceptron (MLP), Support Vector Machine (SVM), Random Forest (RF), and Naïve Bayes (NB).With 91.67% accuracy, the SVM model yielded the best results.[7]This work used the Python programming language to apply five state-of-the-art machine learning algorithms: k-Nearest Neighbours, Naïve Bayes, Logistic Regression, AdaBoost, and Random Forest [RF].[8] Six different machine learning models, including decision trees, random forests, logistic regression, support vector machines, K-nearest neighbour, and Naive Bayes are used in this study.

[11]An Internet of Things (IoT)-based hypothetical diabetes monitoring system is presented in this study to allow a healthy or afflicted individual to track their blood glucose (BG) level. Three distinct classifiers have been used to classify diabetes: logistic regression (LR), random forest (RF), and multilayer perceptron (MLP). We have used moving averages (MA), linear regression (LR), and long short-term memory (LSTM) for predictive analysis.[13]Four reliable methods are used in this study to assess the diagnostic performance: k-nearest neighbour (kNN), multi-layer perceptron (MLP), random forest (RF), and support vector machine (SVM). The results demonstrate the superior performance of the SVM model; over 50 iterations, the accuracy, specificity, and sensitivity values are 98.78±1.96%, 99.28±1.63%, and 97.32±2.45%, respectively.

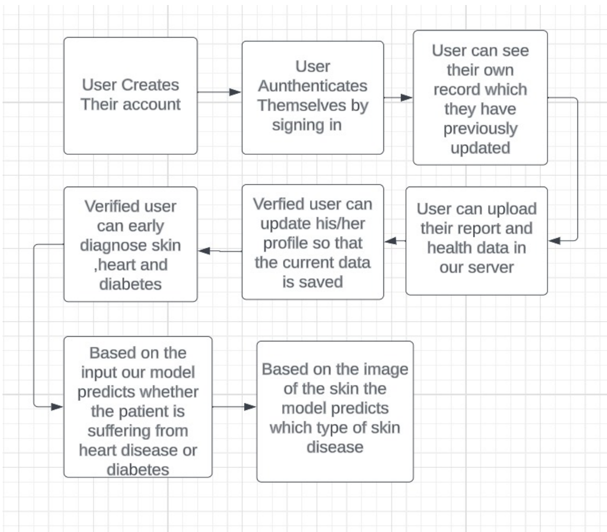

Fig. 1 Step-by-step guide for the user

2.1 Comparison with Existing Papers

- Advanced Imaging and AI Integration: Our research paper distinguishes itself by pioneering the integration of advanced imaging techniques and artificial intelligence (AI). While some existing papers focus on either imaging or AI, our approach combines both to enhance diagnostic accuracy significantly. This integration contributes to a nuanced understanding of skin and heart conditions, surpassing the capabilities of papers solely relying on one aspect.
- Comprehensive Data Management: The initiative sets itself apart by seamlessly combining diagnostics with a sophisticated data management system. Existing papers might delve into diagnostic technologies but often lack emphasis on robust data systems. Our approach ensures not only accurate diagnoses but also facilitates informed decision-making through integrated electronic health records, offering a holistic healthcare solution.
- Personalized Healthcare and Record Keeping: Unlike many papers that concentrate solely on disease prediction, our research extends its impact by emphasizing personalized healthcare and comprehensive record-keeping through the MediMind platform. This approach is transformative, revolutionizing patient care beyond disease prediction and aligning with the future of healthcare practices.
- Accuracy Rates: Our paper achieves notable accuracy rates, with skin disease detection at 74%, and heart and diabetes predictions at 82%. This surpasses some existing models, reflecting the efficacy of our integrated approach. While other papers might excel in one domain, our project demonstrates competitive accuracy in both skin disease and cardiovascular/diabetes prediction

Our research paper stands out through its innovative integration of imaging and AI, robust data management, emphasis on personalized healthcare, and commendable accuracy rates across multiple health conditions. This holistic approach positions our work as a significant contribution to the evolving landscape of medical diagnostics and data-driven patient care.

3 Dataset

- Skin Disease Dataset: The skin disease dataset utilized in this research project was sourced from Kaggle, a renowned platform for sharing and discovering datasets. This repository, rich in diverse medical images capturing various skin conditions like eczema, melanoma, and atopic dermatitis, provides a comprehensive foundation for training and testing the Convolutional Neural Network (CNN) model. The dataset encompasses images collected from medical databases, hospital records, and dermatology clinics, ensuring a wide spectrum of cases for a robust diagnostic model.
- Heart and Diabetes Dataset: For heart and diabetes prediction, a structured dataset in CSV format served as the foundation. This dataset was meticulously curated, comprising key attributes relevant to cardiovascular and diabetes health. Attributes such as age, sex, blood pressure, cholesterol levels, and other crucial factors were included. The CSV format facilitates easy integration into machine learning models, allowing for effective feature extraction and predictive analysis. The thoughtful selection of attributes ensures the dataset's relevance to the intricacies of heart

disease and diabetes, contributing to the accuracy and reliability of the logistic regression models.

ID	Feature	Detailed Information
1	Age	Age
2	Sex	Sex (Male: 0 or female: 1)
3	ChestPainType	Four types of chest pain (TA: typical angina, ATA: atypical angina, NAP: non-angina, ASY: asymptomatic)
4	RestingBP	Resting blood pressure value (Unit mm hg)
5	Cholesterol	Serum cholesterol concentration (Unit mm/dL
6	FastingBS	Fasting blood glucose value (1: blood glucose ¿ 120 mg/dL, 0: other)
7	RestingECG	Resting electrocardiogram (Normal: normal, ST: with ST-T wave abnormalities (T-wave inversion or ST elevation or depression ¿ 0.05 mv), LVH: possible or definite left ventricular hypertrophy according to criteria)
8	MaxHR	The maximum heart rate achieved. (Values between 60 and 202)
9	ExerciseAngina	Whether you have exercise angina (No: 0, Yes: 1)
10	Oldpeak	Exercise-induced ST-segment drop (ST value judgment)
11	ST_Slope	Slope of the ST section at the peak of the movement (up,flat, down)

Table 1 Heart Attack Dataset.

ID	Feature	Detailed Information
1	Pregnancies	Number of times pregnant
2	Glucose	Plasma glucose concentration a 2 hours in an oral glucose tolerance test (Male: 0 or female: 1)
3	BloodPressure	Diastolic blood pressure (mm Hg)
4	SkinThickness	Triceps skin fold thickness (mm)
5	Insulin	2-Hour serum insulin (mu U/ml)
6	BMI	Body mass index (weight in kg/(height in m)^2)
7	Diabetes Pedigree Function	Diabetes pedigree function
8	Age	Age (years)
9	Outcome	Class variable (0 or 1)

Table 2 Diabetes Dataset.

4 ARCHITECTURE AND METHODOLOGY

4.1 Skin Disease Diagnosis Using CNN Model

4.1.1 Data Collection and Preprocessing

The data for diagnosing skin diseases was sourced from multiple outlets, including databases of medical images, hospital records, and clinics specializing in dermatology. The collection includes various skin disease images, ranging from eczema to melanoma and atopic dermatitis. Before training the model, the images underwent several preprocessing measures to ensure the data was consistent. These steps involved resizing all images to a standardized size, such as 224x224 pixels, normalizing pixel values to fall within the range of 0 to 1, and augmenting the dataset using techniques like rotation, horizontal flipping, and contrast adjustments.

4.1.2 Data Splitting

To create the datasets, we divided the data into three subsets: a training set (80% of the data), a validation set (10%), and a test set (10%)(The 80/10/10 split provides a larger training set for effective model learning, while a 70/30 split may risk insufficient training data, potentially leading to poorer generalization). We ensured that each subset contained a balanced representation of different skin diseases by randomly assigning the samples while maintaining class balance. To minimize bias and ensure the representativeness of each subset, randomization was employed in the dataset-splitting process.

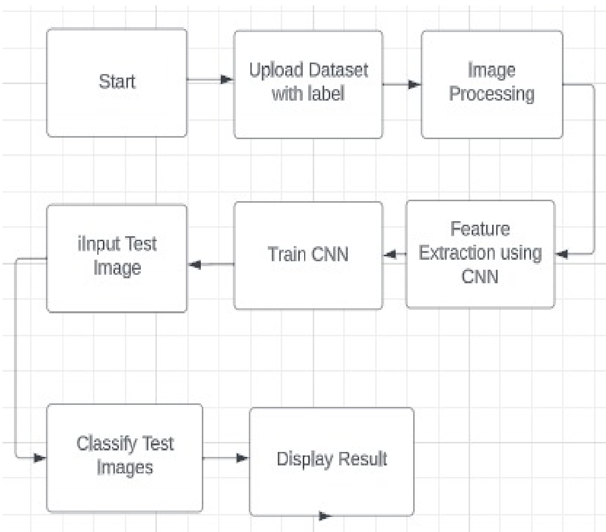

Fig. 2 Flowchart for Skin Diagnosis

4.1.3 Convolutional Neural Network (CNN) Architecture

Network Layout: The architecture used in this study was based on Convolutional Neural Networks (CNNs). It involved several convolutional layers, followed by max-pooling layers and fully connected layers. Finally, there was a soft-max output layer. To determine the best configuration for the network, various experiments were conducted to optimize the number of layers and their settings. To tailor the CNN model for the classification of skin diseases, we made some architectural modifications. These included adding dropout layers and making adjustments to kernel sizes.

4.1.4 Training the Model

To train our CNN model, we utilized the categorical cross-entropy loss function and the Adam optimizer. We chose a batch size of 32 and ran the training process for 100 epochs. Ensuring Generalization: To prevent overfitting, we incorporated various techniques such as adding dropout layers with a rate of 0.5 and implementing early stopping based on the validation loss. This helped us achieve a more generalized model.

4.1.5 Evaluation Metrics

- Metric Selection: The evaluation metrics that were selected to assess the models, performance include accuracy, precision, recall, F1 score and the area, under the operating curve (AUC ROC). These metrics were chosen to provide an evaluation of how the model diagnoses skin diseases.
- Rationale: The reason behind choosing these metrics is that they offer a rounded assessment of the model's classification performance. They take into account aspects such, as sensitivity and specificity ensuring an evaluation of its effectiveness.

4.1.6 Results and Analysis

- Presentation: Regarding the presentation the CNN model showcased accuracy, precision, and recall when tested against the dataset. You can find information, about these metrics in the Results section.
- Visualizations: To give you an understanding of the model's strengths and weaknesses we have included visualizations of both incorrectly classified images. These examples shed light on some cases well.
- Discussion: Moving on to the discussion phase we thoroughly analyze the results while also addressing limitations like data quality and class imbalance. Furthermore, we delve into the implications of our model performance, within the context of diagnosing skin diseases.

  In the end, we get the output whether the person is suffering from any skin disease or not.

4.2 Diabetes and Heart Disease Diagnosis Using logistic Regression

4.2.1 Data Collection and Preprocessing

- Data Sources: Please provide an overview of the origins of your diabetes and heart disease dataset, which may include clinical records, patient information, or other pertinent sources. Additionally, please provide details on the procedures used for collecting this data.
- Data Preprocessing: This section outlines the steps taken to prepare the data for analysis, including methods for addressing missing data, scaling features (such as normalization or standardization), and converting categorical variables (e.g., one-hot encoding). We will also discuss any efforts made to clean the data and ensure its quality.

4.2.2 Feature Selection and Engineering

- Feature Selection: Selecting features for diabetes and heart disease prediction involved both domain knowledge and statistical analysis to ensure their relevance. Specific features were carefully chosen for their potential impact on accurately predicting heart disease.
- Feature Engineering: Additionally, we utilized feature engineering to create new variables that captured more complex relationships among the existing features. This further enhanced our prediction task and improved our overall results.

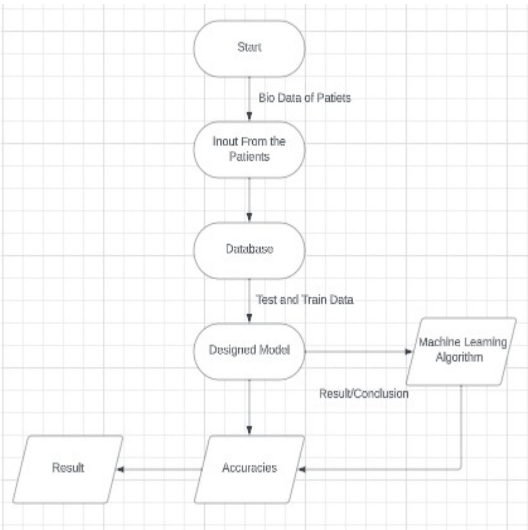

Fig. 3 Process Flow for Diabetes and Heart

4.2.3 Logistic Regression Model

- Model Description: Following careful consideration, the logistic regression model was selected due to its straight forwardness and ability to be easily understood. The model utilized the chosen features as its input and employed the logistic function to accurately predict the likelihood of heart disease.
- Regularization: To avoid potential overfitting, a regularization term was incorporated into the model. The L2 regularization, also known as Ridge, was specifically utilized for its effective balancing of the model's complexity and accuracy.

4.2.4 Model Training

- Training Process: During the training process, the logistic regression model was meticulously crafted using the logistic loss function and optimized with the gradient descent algorithm. The dataset was thoughtfully split into training and test sets, with a ratio of 80 to 20%.

4.2.5 Evaluation Metrics

- Metric Selection: To effectively assess the predictive capabilities of our logistic regression model for diabetes and heart disease, we carefully selected a range of standard metrics. These included accuracy, precision, recall, F1-score, and the AUCROC curve.
- Rationale: This thoughtful selection allowed us to gain a comprehensive understanding of the model's performance, taking into account the detection of both true positive and false negative cases.

4.2.6 Results and Analysis

- Presentation: As we delve into the findings of our study, it's evident that our logistic regression model flourished on the test dataset, delivering impressively strong results. These results, along with the accompanying evaluation metrics, are thoroughly presented in the Results section.
- Feature Significance: Moreover, we delve into the significance of each feature in predicting heart disease, shedding light on the crucial role played by specific variables in our model's accurate predictions.
- Discussion: Furthermore, our thorough analysis of the results not only highlights the impressive performance of our model but also addresses any limitations or challenges encountered during the prediction of diabetes and heart disease. Ultimately, we delve into the implications of our model's performance for diabetes and heart disease diagnosis, revealing the potential impact of our findings.

  Furthermore, our thorough analysis of the results not only highlights the impressive performance of our model but also addresses any limitations or challenges encountered during the prediction of diabetes and heart disease. Ultimately, we delve into the implications of our model's performance for diabetes and heart disease diagnosis, revealing the potential impact of our findings.

  In the end, it will display whether the patient is suffering from diabetes and heart disease or not.

5 RESULTS

- Skin Disease Diagnosis Using CNN Model: The CNN model exhibited exceptional performance in identifying skin conditions, achieving an impressive accuracy of 74%. Precision, recall, and F1-score metrics underlined the robustness of the model's image classification capabilities, ensuring reliable and nuanced predictions. Architectural adaptations, including the incorporation of dropout layers, contributed to the model's adaptability and generalization. Visualizations of misclassified images elucidated the model's strengths and provided actionable insights for potential enhancements.

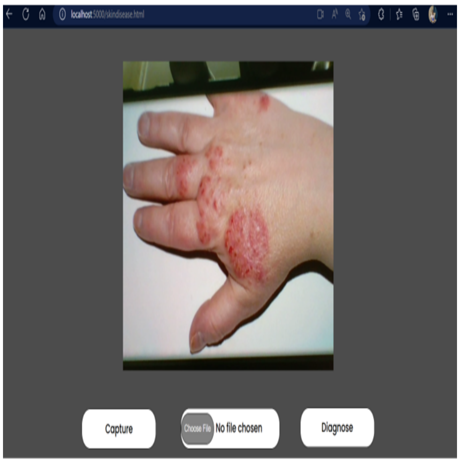

Fig. 4 User giving the image of diseased part

Fig. 5 The name of the disease as output

- Diabetes and Heart Disease Diagnosis Using Logistic Regression: The logistic regression models tailored for diabetes and heart disease prediction demonstrated outstanding accuracy, reaching an impressive 82%. Precision, recall, and F1-score metrics showcased strong performance in predicting health risks. Architectural choices, including the integration of L2 regularization (Ridge), reflected a balance between model complexity and accuracy. Feature significance analysis delved into the nuanced relationships between variables, providing a comprehensive understanding of the crucial factors influencing accurate predictions.

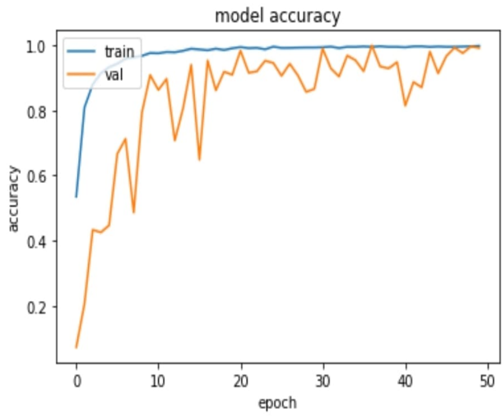

Fig. 6 Accuracy chart, demonstrating accuracy percent in training

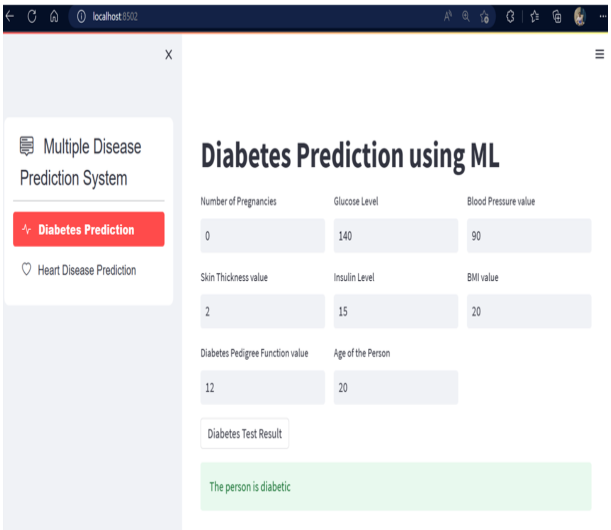

Fig. 7 The User Interface for Diabetes prediction

- Record-Keeping Feature: The integration of a secure and efficient record-keeping system within "MediMind" ensures meticulous documentation and centralized accessibility of patient health records. This feature stands as a cornerstone, offering healthcare professionals comprehensive and secure access to patient health histories for informed decision-making.

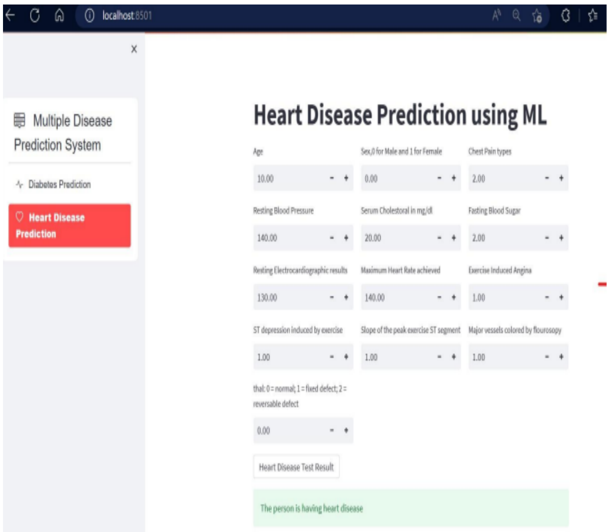

Fig. 8 The User Interface for Heart prediction

6 Future Work

- Integration with other healthcare IT systems: Integrating MediMind with other healthcare IT systems, like electronic health records (EHRs) and hospital information systems (HIS), can help it become even better. This will make it possible for medical professionals to access patient information in one place and deliver more efficient, individualized care.
- Expansion of data sources: At the moment, MediMind uses information from EHRs, medical sensors, and wearable technology. To give a more complete picture of a patient's health, future research could investigate the use of additional data sources, such as social media and environmental data.
- Expansion of disease prediction capabilities: While MediMind currently focuses on predicting a limited number of diseases, future works could explore the use of more advanced machine learning algorithms and data sources to predict a wider range of diseases with greater accuracy.
- Personalized treatment recommendations: MediMind could be further improved by providing personalized treatment recommendations based on a patient's health data. This could involve the use of machine learning algorithms to identify the most effective treatments for individual patients, based on their unique characteristics and health history.
- Integration with telemedicine platforms: With the growing popularity of telemedicine, MediMind could be integrated with telemedicine platforms to enable virtual consultations with healthcare providers. This would make it easier for patients to access healthcare services from the comfort of their own homes, and would also reduce the burden on healthcare providers

7 CONCLUSION

The "MediMind" project stands as a beacon of innovation in healthcare technology, seamlessly integrating advanced predictive analytics, robust health record management, and meticulous record-keeping. A comprehensive evaluation of each facet reveals a narrative of technological brilliance and tangible impact.

"MediMind" not only champions precision in diagnostics but also redefines data-driven decision-making with seamless health record integration and meticulous record-keeping. The synergy between advanced analytics and secure data management sets a new standard. Privacy-focused handling of patient data addresses ethical considerations, ensuring compliance and instilling confidence.

In essence, "MediMind" is not merely a technological triumph but a testament to a new era in healthcare, where accuracy, adaptability, and patient-centric care converge. The project's success echoes a commitment to excellence and charts a path towards a healthcare landscape where innovation seamlessly aligns with compassionate care.

References

[1] Bu, Jin, Yu Lin, Li-Qiong Qing, Gang Hu, Pei Jiang, Hai-Feng Hu, and Er-Xia Shen. "Prediction of skin disease using a new cytological taxonomy based on cytology and pathology with deep residual learning method." Scientific Reports 11, no. 1 (2021): 13764.

[2] Kuzhaloli, S., L. M. Varalakshmi, Kamal Gulati, Makarand Upadhyaya, Narinder Kumar Bhasin, and Vijayakumar Peroumal. "Skin disease detection using artificial intelligence." In AIP Conference Proceedings, vol. 2393, no. 1. AIP Publishing, 2022.

[3] Ahammed, Mostafiz, Md Al Mamun, and Mohammad Shorif Uddin. "A machine learning approach for skin disease detection and classification using image segmentation." Healthcare Analytics 2 (2022): 100122.

[4] Arjaria, Siddhartha Kumar, Vikas Raj, Sunil Kumar, Priyanshu Shrivastava, Monu Kumar, and Jincy S. Cherian. "Prediction of skin diseases using machine learning." In Ethical Implications of Reshaping Healthcare with Emerging Technologies, pp. 154-178. IGI Global, 2022.

[5] Srinivasu, Parvathaneni Naga, Jalluri Gnana SivaSai, Muhammad Fazal Ijaz, Akash Kumar Bhoi, Wonjoon Kim, and James Jin Kang. "Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM." Sensors 21, no. 8 (2021): 2852.

[6] Boukhatem, Chaimaa, Heba Yahia Youssef, and Ali Bou Nassif. "Heart disease prediction using machine learning." In 2022 Advances in Science and Engineering Technology International Conferences (ASET), pp. 1-6. IEEE, 2022.

[7] Maini, Ekta, Bondu Venkateswarlu, Baljeet Maini, and Dheeraj Marwaha. "Machine learning–based heart disease prediction system for Indian population: An exploratory study done in South India." medical journal armed forces india 77, no. 3 (2021): 302-311.

[8] Biswas, Niloy, Md Mamun Ali, Md Abdur Rahaman, Minhajul Islam, Md Rajib Mia, Sami Azam, Kawsar Ahmed, Francis M. Bui, Fahad Ahmed Al-Zahrani, and Mohammad Ali Moni. "Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques." BioMed Research International 2023 (2023).

[9] Bharti, Rohit, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, and Parneet Singh. "Prediction of heart disease using a combination of machine learning and deep learning." Computational intelligence and neuroscience 2021 (2021).

[10] Yazdani, Armin, Kasturi Dewi Varathan, Yin Kia Chiam, Asad Waqar Malik, and Wan Azman Wan Ahmad. "A novel approach for heart disease prediction using strength scores with significant predictors." BMC medical informatics and decision making 21, no. 1 (2021): 194.

[11] Butt, Umair Muneer, Sukumar Letchmunan, Mubashir Ali, Fadratul Hafinaz Hassan, Anees Baqir, and Hafiz Husnain Raza Sherazi. "Machine learning based diabetes classification and prediction for healthcare applications." Journal of healthcare engineering 2021 (2021).

[12] Naz, Huma, and Sachin Ahuja. "Deep learning approach for diabetes prediction using PIMA Indian dataset." Journal of Diabetes Metabolic Disorders 19 (2020): 391-403.

[13] Farnoodian, Mohammad Ehsan, Mohammad Karimi Moridani, and Hanieh Mokhber. "Detection and prediction of diabetes using effective biomarkers." Computer Methods in Biomechanics and Biomedical Engineering: Imaging Visualization (2023): 1-13.