MediMind: A Comprehensive Health Prediction and Record-Keeping Platform

Darshana A. Naik
Asst. Prof Department of Computer Science
Ramaiah Institute Of Technology
darshananaik@msrit.edu

Aditya Sharma
Department of Computer Science
Ramaiah Institute Of Technology
adityaofficial0501@gmail.com

Danish Mahajan
Department of Computer Science
Ramaiah Institute Of Technology
danishmahajan779@gmail.com

Dhruv Khandelwal
Department of Computer Science
Ramaiah Institute Of Technology
dhruv15.khandelwal@gmail.com

Arpit gajya
Department of Computer Science
Ramaiah Institute Of Technology
arpitgajya2001@gmail.com

***Abstract*—Chronic skin diseases, heart ailments, and diabetes are significant health concerns with substantial implications for patients and healthcare systems if not detected and managed in a timely and accurate manner. This paper introduces a comprehensive approach that integrates modern image processing, machine learning, and deep learning techniques to address these health challenges collectively. By utilizing image processing, advanced algorithms, and deep learning models, we enable the early detection of skin conditions and the prediction of diabetes and heart disease risk. Additionally, our system leverages patient medical records to provide accurate diabetes predictions.**

I. 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.

In this paper, we will explore an integrated healthcare diagnostics and data management system. We'll delve into the technology and methodologies used for diagnosing skin diseases and heart conditions, as well as discuss the architecture and functionality of the data management system. Our research findings aim to contribute to the ongoing conversation in medical diagnostics and data management, while also promoting a better understanding of how technology can enhance patient care and revolutionize the medical profession.

II. RELATED WORK

[1] created an artificial neural network system with a 90% accuracy rate for diagnosing and treating individuals with skin problems. [2] presented a technique for automated dermatological diagnosis. For training and testing purposes, they have utilized several pre-processing methods, including ours, and feed forward-back propagation artificial neural networks. [3] The accuracy percentage of a back propagation neural network (BPNN) prototype used to help dermatologists was 91.2%. [4] Using the support vector machine (SVM) technique, skin diseases like melanoma, basal cell carcinoma (BCC), nevus, and seborrheic keratosis (SK) are categorized.

[5] To recognize skin conditions, a novel approach that combines computer vision and machine learning is proposed. Computer vision is used to extract features from images, whereas machine learning is used to identify skin diseases. The system performed 95% accurately when tested against six different skin conditions. [6]Identified the risk factors of coronary heart disease or atherosclerosis using the inbuilt implementation algorithm and some neural network techniques, and they were only just able to accurately predict whether the test patient has the specified disease or not

[7] The metaclassifier used in the study was logistic regression, and the stacking technique was used to combine the results of the KNN, random forest, and SVM models.

The accuracy rate for the stacked model was 75.1[8] In the study various machine learning algorithms were used for classification tasks, achieving an accuracy rate of 70%. These algorithms included Random Forest, Naive Bayes, Logistic Regression, and KNN. [9] With a Repeated Random approach and Random Forest as the classifier, a high accuracy rate of 89.01% was attained. [10] To predict diabetes, authors proposed a system that makes use of principal component analysis and recursive feature elimination. Deep neural networks and artificial neural networks are used to categorize diabetes. Their accuracy with deep neural networks was 82.67%, and with artificial neural networks, it was 78.62%.

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

III. ARCHITECTURE AND METHODOLOGY

A. Skin Disease Diagnosis Using CNN Model

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.

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%). 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.

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) 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.

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.

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.

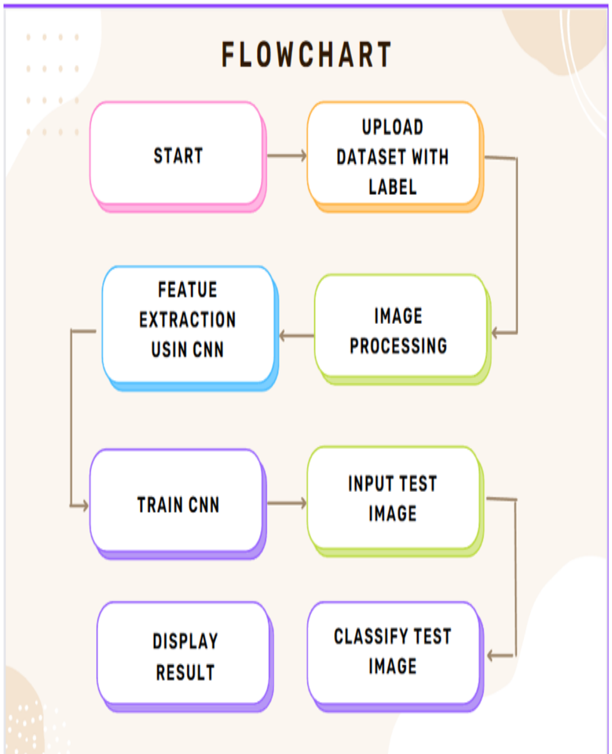

Fig. 2. Flowchart for Skin Diagnosis

B. *Heart Disease Diagnosis Using logistic regression*

1) *Data Collection and Preprocessing:*

- Data Sources: Please provide an overview of the origins of your 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.

2) *Feature Selection and Engineering:*

- Feature Selection: Selecting features for 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.

3) *Logistic Regression Model:*

- Model Description: Following careful consideration, the logistic regression model was selected due to its straightforwardness 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) *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%.

5) *Evaluation Metrics:*

- Metric Selection: To effectively assess the predictive capabilities of our logistic regression model for 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.

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 heart disease. Ultimately, we delve into the implications of our model's performance for 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 heart disease. Ultimately, we delve into the implications of our model's performance for heart disease diagnosis, revealing the potential impact of our findings.
  In the end, it will display whether the patient is suffering from heart disease or not.

The system demonstrated an ability to predict the specified diseases with an accuracy ranging from 70% to 74%. The model's loss visibly decreased sharply initially and then gradually tapered off towards the conclusion. Similarly, the accuracy showed a sudden increase followed by a stabilization towards the conclusion. After thorough testing with various epoch values, the decision was made to set the epoch at 1200.

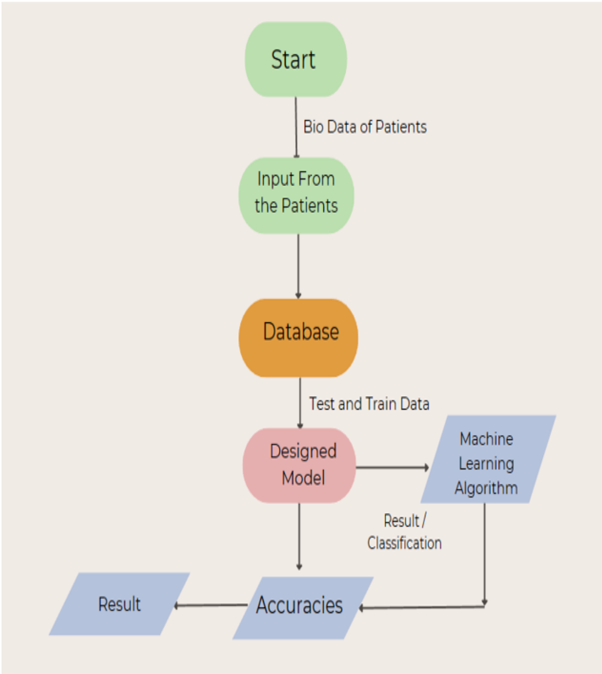

Fig. 3. Process Flow for Diabetes and Heart

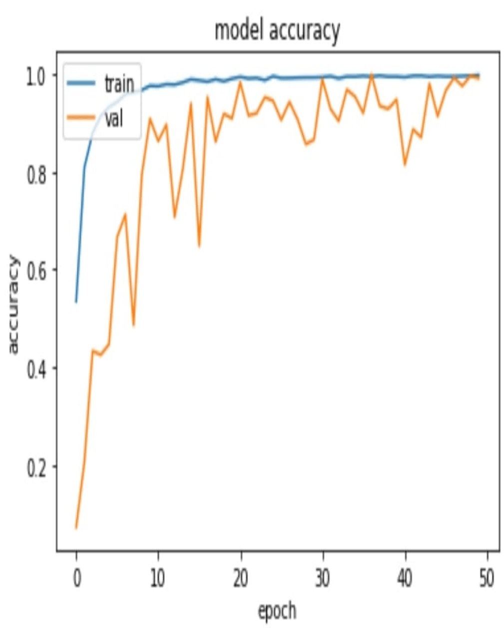

Fig. 4. Accuracy chart, demonstrating accuracy percent in training

IV. RESULTS

MediMind provides a comprehensive platform for record-keeping, health prediction, and personalized healthcare management. The platform uses advanced algorithms and machine learning to analyze patient data and predict potential health issues. By storing health records in a secure database, healthcare organizations can ensure that sensitive patient information is protected. The platform can improve patient care by providing healthcare providers with comprehensive patient health records and making it easier for them to manage patient records efficiently

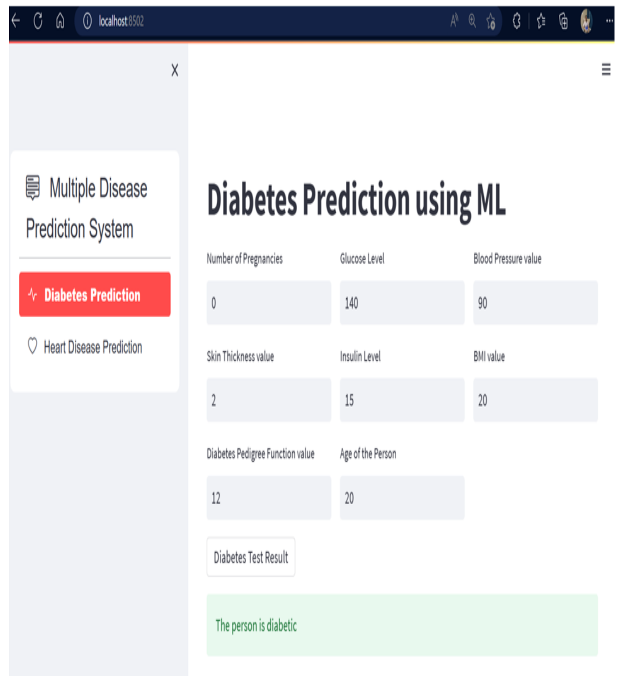

Fig. 5. The User Interface for Diabetes prediction

The proposed research work on MediMind is significant for several reasons:

- Personalized healthcare: MediMind has the potential to revolutionize healthcare by providing individuals with personalized and data-driven healthcare. This can lead to improved health outcomes, reduced healthcare costs, and a better quality of life for patients.
- Early detection and prevention: MediMind can facilitate early disease detection and prevention by utilizing cutting-edge algorithms and machine learning to anticipate possible health problems. Better patient outcomes and less strain on the healthcare system may come from this.
- Secure data management: MediMind provides a secure platform for storing and managing health records, which is critical in today's world where cybersecurity threats

are a major concern. This can help prevent data breaches and protect sensitive patient information.

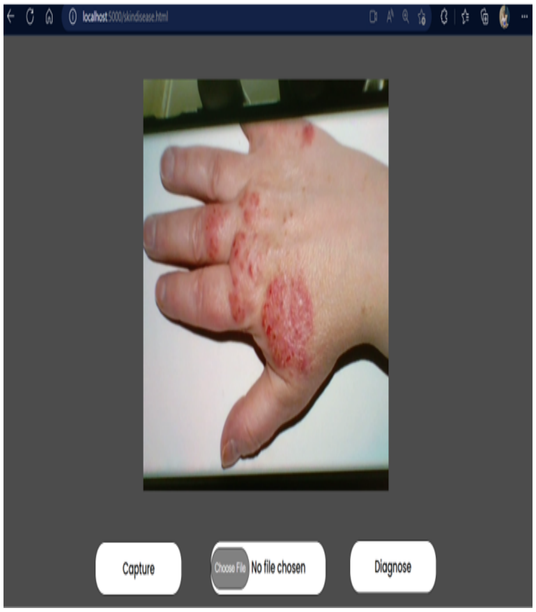

Fig. 6. User giving the image of diseased part

Fig. 7. The name of the disease as output

V. 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

VI. Conclusion

In conclusion, the development of MediMind, a comprehensive health prediction and record-keeping platform, has the potential to significantly improve healthcare outcomes by leveraging machine learning algorithms to predict and prevent diseases, while also providing a centralized platform for managing patient health records.

Through the use of advanced deep learning algorithms, such as neural networks and logistic regression models, MediMind can accurately predict and diagnose diseases, leading to earlier interventions and improved patient outcomes.

References

[1] L. G. Kabari and F. S. Bakpo, Member, IEEE, "Diagnosing Skin Diseases Using an Artificial Neural Network", c2009 IEEE

[2] M. Shamsul Arifin, M. Golam Kibria, Adnan Firoze, M. Ashraful Amin, Hong Yan, "Dermatological Disease Diagnosis Using Colour-skin Images", Proceedings of the 2012 International Conference on Machine Learning and Cybernetics, Xian, 15-17 July, 2012

[3] Shuzlina Abdul-Rahman, Ahmad Khairil Norhan, Marina Yusoff, Azlinah Mohamed, Sofianita Mutalib, "Dermatology Diagnosis with Feature Selection Methods and Artificial Neural Network", 2012 IEEE EMBS International Conference on Biomedical Engineering and Sciences — Langkawi — 17th - 19th December 2012

[4] Suganya R. (2016) "An Automated Computer Aided Diagnosis of Skin Lesions Detection and Classification for Dermoscopy Images." International Conference on Recent Trends in Information Technology.

[5] Kumar, V., Kumar, S., & Saboo, V. (2016) "Dermatological Disease Detection Using Image Processing and Machine Learning." IEEE

[6] Jabbar M A, Deekshatulu B L & Chandra P (2013, March). Heart disease prediction using lazy associative classification. In 2013 International Mutli-Conference on Automation, Computing,Communication, Control and Compressed Sensing (iMac4s) (pp. 40- 6). IEEE

[7] Shorewala, V. Early detection of coronary heart disease using ensemble techniques. Inform. Med. Unlocked 2021, 26, 100655

[8] Maiga, J.; Hungilo, G.G.; Pranowo. Comparison of Machine Learning Models in Prediction of Cardiovascular Disease Using Health Record Data. In Proceedings of the 2019 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), Jakarta, Indonesia, 24–25 October 2019; pp. 45–48.

[9] Ouf, S.; ElSeddawy, A.I.B. A proposed paradigm for intelligent heart disease prediction system using data mining techniques. J. Southwest Jiaotong Univ. 2021, 56, 220–240

[10] . Vijayashree and J. Jayashree, " An Expert System for the Diagnosis of Diabetic Patients using Deep Neural Networks and Recursive Feature Elimination," International Journal of Civil Engineering and Technology, vol. 8, pp. 633-641, Dec. 2017.

[11] L. B. Goncalves and M. M. Bernardes, "Inverted Hierarchical Neuro-Fuzzy BSP System: A Novel NeuroFuzzy Model for Pattern Classification and Rule Extraction in Databases," in IEEE Transactions on Systems, Man, and Cybernetics, vol. 36, no. 2, pp. 236- 248, Mar. 2006.

[12] L. Han, S. Luo, H. Wang, L. Pan, X. Ma and T. Zhang, "An Intelligible Risk Stratification Model Based on Pairwise and Size Constrained Kmeans," in IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 5, pp. 1288-1296, Sept. 2017.

[13] Xu, Weifeng, Jianxin Zhang, Qiang Zhang, and Xiaopeng Wei. "Risk prediction of type II diabetes based on random forest model." In Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2017 Third International Conference on, pp. 382-386. IEEE, 2017.

[14] Song, Yunsheng, Jiye Liang, Jing Lu, and Xingwang Zhao. "An efficient instance selection algorithm for k nearest neighbor regression." Neurocomputing 251 (2017): 26-34.

[15] Komi, Messan, Jun Li, YongxinZhai, and Xianguo Zhang. "Application of data mining methods in diabetes prediction." In Image, Vision and Computing (ICIVC), 2017 2nd International Conference on, pp. 1006-1010. IEEE, 2017