VIVEKANAND EDUCATION SOCIETY'S
INSTITUTE OF TECHNOLOGY

Department of Computer Engineering

Project Report on

WellnessInsight

Advanced System for Tailored Health Recommendation

In partial fulfillment of the Fourth Year (Semester–VII), Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai

Academic Year 2024-2025

Project Mentor
Mrs. Abha Tiwari

Submitted by
Jiya Lund , D17A/37
Tisha Jeswani D17B/20
Varsha Makhija D17C/43
Dinky Khatri D17A/31

(2024-25)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY

Department of Computer Engineering

CERTIFICATE of Approval

This is to certify that Jiya Lund(D17A-37),Tisha Jeswani(D17B-20),Varsha Makhija(D17C-43), Dinky Khatri(D17A-31) of Fourth Year Computer Engineering studying under the University of Mumbai has satisfactorily presented the project on "***Wellness InSight***" as a part of the coursework of PROJECT-I for Semester-VII under the guidance of ***Mrs. Abha Tiwari*** in the year 2024-2025.

Date

Internal Examiner

External Examiner

Project Mentor

Head of the Department
Dr. Mrs. Nupur Giri

Principal
Dr. J. M. Nair

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Mrs.Abha Tiwari** for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair ,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

Computer Engineering Department

COURSE OUTCOMES FOR B.E PROJECT

Learners will be to:-

Course Outcome	**Description of the Course Outcome**
CO 1	Do literature survey/industrial visit and identify the problem of the selected project topic.
CO2	Apply basic engineering fundamental in the domain of practical applications FORproblem identification, formulation and solution
CO 3	Attempt & Design a problem solution in a right approach to complex problems
CO 4	Cultivate the habit of working in a team
CO 5	Correlate the theoretical and experimental/simulations results and draw the proper inferences
CO 6	Demonstrate the knowledge, skills and attitudes of a professional engineer & Prepare report as per the standard guidelines.

ABSTRACT of the project

WellnessInsight is an AI-driven skin disease prediction system designed to assist users in identifying common skin conditions, including psoriasis, eczema, and melanoma. The system provides a comprehensive diagnostic tool that accepts input in the form of images, text, and audio, enabling accurate predictions and infection percentage estimations. By leveraging artificial intelligence, WellnessInsight aims to promote early diagnosis and timely treatment, especially for individuals with limited access to dermatological care. This report outlines the motivation behind the project, addressing the limitations of existing diagnostic systems that often lack accuracy and flexibility in input types. WellnessInsight overcomes these challenges by integrating multiple forms of input, improving accessibility, and delivering precise predictions. The methodology employed includes data collection, preprocessing of input data, model training using CNNs and NLP models, and generating predictive results. Additionally, the system enhances user experience by recommending nearby healthcare professionals based on the user's location, facilitating informed medical decisions.

INDEX

Chapter 1 : Introduction

This chapter presents the WellnessInsight project, an AI-driven skin disease prediction system that assists users in identifying psoriasis, eczema, and melanoma. It outlines the motivation for the project, addressing the limitations of existing diagnostic systems and defining the problem that WellnessInsight aims to solve. The chapter also discusses the project's relevance in promoting early diagnosis and treatment while providing an overview of the methodology used, including data collection, preprocessing, model training, and prediction mechanisms.

1.1. Introduction to the Project

WellnessInsight is an AI-powered Skin Disease Prediction System designed to assist users in identifying common skin conditions such as psoriasis, eczema, and melanoma. By accepting inputs in the form of images, text, and audio, the system predicts the type of skin disease and estimates the infection percentage. It offers a user-friendly platform that enables early detection, empowering users to seek timely medical advice and potentially avoid complications associated with untreated skin conditions.

1.2. Motivation for the Project

Millions of people suffer from skin diseases worldwide, but early detection is critical for effective treatment and management. Access to dermatological care is often limited, particularly in rural or remote areas. WellnessInsight aims to bridge this gap by providing an accessible, AI-driven diagnostic tool that enables users to assess their skin condition from the comfort of their own home using their smartphone or computer. The project is motivated by the need to deliver a tool that can expedite diagnosis, promote early treatment, and ultimately improve skin health outcomes.

1.3. Drawbacks of the Existing System

Traditional methods of diagnosing skin diseases involve physical visits to dermatologists, which can be costly, time-consuming, and difficult for people in underserved regions. Current online diagnostic systems are either limited to specific input types (such as only images) or lack the accuracy required for reliable predictions. These systems also do not integrate multiple forms of input like text and audio, leaving a gap in holistic skin disease assessment. WellnessInsight seeks to address these limitations by incorporating image, text, and audio analysis in a single platform.

1.4. Problem Definition

The challenge this project addresses is the lack of an accessible, accurate, and multi-input system for diagnosing common skin diseases. Existing systems are either too narrow in their focus or offer limited accuracy. WellnessInsight aims to provide a robust solution that uses advanced AI techniques to analyze images, text descriptions, and audio to predict skin diseases and infection percentages. The system also recommends nearby healthcare providers based on the user's location, helping them make informed decisions about their treatment.

1.5. Relevance of the Project

WellnessInsight is highly relevant given the prevalence of skin diseases like psoriasis, eczema, and melanoma, which can significantly impact health if not diagnosed early. This project provides a comprehensive tool that users can access from any location, offering predictions based on multiple input types. It supports the global trend toward AI-driven healthcare and telemedicine solutions, making healthcare more accessible. This relevance is further underscored by the system's ability to recommend nearby doctors, contributing to better health outcomes for users.

1.6. Methodology Used

The WellnessInsight system follows a structured methodology:

- Data Collection: The system is trained on a dataset comprising images, textual descriptions, and audio samples of skin diseases.
- Preprocessing: Input data (images, text, and audio) undergoes preprocessing. Images are resized and denoised, text is normalized for language processing, and audio samples are converted into Mel-frequency cepstral coefficients (MFCCs).
- Model Training: The system uses a convolutional neural network (CNN) for image classification. Natural Language Processing (NLP) models are used to analyze text inputs, while audio inputs are processed using deep learning models trained on MFCC features.
- Prediction and Output: Based on the input type (or a combination of inputs), WellnessInsight predicts the type of skin disease and estimates the infection percentage. It also suggests nearby healthcare professionals to guide users toward appropriate treatment.

Chapter 2 : Literature Survey

2.1. Research Papers

2.1.1: A Survey on Skin Lesion Detection and Classification using Machine Learning

Citation: R. Yadav and A. Bhat, "A Survey on Skin Lesion Detection and Classification using Machine Learning," 2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA), Namakkal, India, 2024, pp. 1-5, doi: 10.1109/AIMLA59606.2024.10531571. keywords: {Machine learning algorithms;Dermatology;Transfer learning;Collaboration;Machine learning;Genetics;Skin;skin lesion;machine learning;survey;deep learning;CNN},

a. Abstract of the research paper - The research explores how dermatologists utilize machine learning (ML) to swiftly and accurately identify and classify skin lesions. Traditional diagnostic methods rely on visual examinations, which can be subjective and vary in interpretation. Machine learning offers a potential solution by analyzing data and recognizing patterns to enhance diagnostic accuracy. This paper reviews current ML technologies, emphasizing real-world validation and addressing dataset variability. While acknowledging the advancements of deep learning (DL), the study highlights the advantages of traditional ML techniques in terms of interpretability and processing performance. It examines methods for automating skin lesion analysis, including feature engineering, rule-based approaches, and traditional ML algorithms. The research proposes overcoming existing challenges by employing advanced transfer learning techniques, integrating genetic and clinical data, and improving the explainability of artificial intelligence (AI) models. The future of skin lesion detection is envisioned to benefit from collaboration between dermatologists and ML experts to develop real-time diagnostic tools, thereby offering scalable solutions for rapid diagnosis and improved patient outcomes. This synergy between medical expertise and ML capabilities holds the potential to revolutionize dermatology by enabling more personalized patient care.

Keywords: skin lesion, machine learning, survey, deep learning, CNN

b. Inference drawn from the paper -

1. **Enhanced Diagnostic Accuracy:** Machine learning models, particularly traditional ML algorithms, can improve the accuracy and consistency of skin lesion detection compared to conventional visual examination methods.
2. **Interpretability and Performance:** Traditional ML techniques offer better interpretability and processing performance, especially in scenarios with limited datasets, making them advantageous over deep learning models in certain contexts.
3. **Data Integration:** Incorporating additional modalities such as genetic and clinical data can significantly enhance the performance of ML models in dermatological applications.
4. **Potential of GANs and Autoencoders:** Generative Adversarial Networks (GANs) and autoencoders play a significant role in data augmentation and feature extraction, thereby enhancing the robustness and efficacy of skin lesion classification models.

2.1.2: Machine Learning Based Skin Disease Detection using Semi-Automatic Image Segmentation.

Citation: J. Sathya and A. Kalaivani, "Machine Learning Based Skin Disease Detection Using Semi-Automatic Image Segmentation," 2024 International Conference on Smart Systems for Electrical, Electronics, Communication and Computer Engineering (ICSSEECC), Coimbatore, India, 2024, pp. 215-219, doi: 10.1109/ICSSEECC61126.2024.10649509. keywords: {Image segmentation;Machine learning algorithms;Machine learning;Nearest neighbor methods;Feature extraction;Skin;Classification algorithms;Skin Disease diagnosis;Image segmentation;Convolutional Neural Networks;Machine Learning;Image Classification},

a. Abstract of the research paper - The study addresses the complexities and uncertainties involved in diagnosing skin diseases, emphasizing the limitations of manual, subjective clinical image analysis performed by medical professionals. To enhance accuracy and efficiency, the paper proposes a machine learning-based approach for diagnosing and classifying skin diseases. This approach leverages a semi-automatic image segmentation technique known as Interactive Superpixel Segmentation, which combines manual user input with automated algorithms to divide images into meaningful regions. The methodology includes morphological operations for hair removal, median filtering for noise reduction, and the use of the Gray-Level Co-occurrence

Matrix (GLCM) for feature extraction. Subsequently, the K-Nearest Neighbors (KNN) algorithm is employed for image classification, culminating in the generation of evaluative reports. The proposed system utilizes datasets sourced from various online platforms and demonstrates improved accuracy compared to existing methods. Overall, the paper outlines a comprehensive workflow for interactive superpixel segmentation in machine learning, aiming to provide a more efficient and accurate tool for skin disease detection.

Keywords: Skin Disease Diagnosis, Image Segmentation, Convolutional Neural Networks, Machine Learning, Image Classification

b. Inference drawn from the paper -

1. **Enhanced Diagnostic Efficiency:** The proposed machine learning approach significantly reduces the time and subjectivity associated with manual skin disease diagnosis, offering a more efficient alternative for medical professionals.
2. **Effective Image Preprocessing:** Techniques such as morphological operations for hair removal and median filtering for noise reduction are crucial in enhancing image quality, thereby improving the performance of subsequent image processing and classification tasks.
3. **Superpixel Segmentation Benefits:** Interactive Superpixel Segmentation effectively partitions skin images into meaningful regions, facilitating more accurate feature extraction and disease classification.
4. **Feature Extraction via GLCM:** Utilizing the Gray-Level Co-occurrence Matrix (GLCM) for texture analysis provides robust features that enhance the classification accuracy of skin diseases.
5. **Superior Classification with KNN:** The K-Nearest Neighbors (KNN) algorithm demonstrates high accuracy in classifying various skin diseases, outperforming traditional methods such as Support Vector Machines (SVM) in the context of the study.
6. **Improved Accuracy Over Existing Methods:** Comparative analysis indicates that the proposed method achieves higher detection rates across multiple skin diseases compared to existing approaches, underscoring its effectiveness.

7. **Potential for Real-Time Applications:** The study suggests that with further development, the proposed system could be adapted for real-time skin disease detection, enhancing accessibility and promptness in medical diagnostics.

2.1.3: Image Segmentation based Approach for Skin Disease Detection and Classification using Machine Learning Algorithms.

Citation:N. H. Sany and P. Chandra Shill, "Image Segmentation based Approach for Skin Disease Detection and Classification using Machine Learning Algorithms," 2024 International Conference on Integrated Circuits and Communication Systems (ICICACS), Raichur, India, 2024, pp. 1-5, doi: 10.1109/ICICACS60521.2024.10498287. keywords: {Support vector machines;Pathogens;Machine learning algorithms;Machine learning;Skin;Medical diagnosis;Medical diagnostic imaging;Skin Disorder;SVM;K-NN;Decision Tree;Image Processing;Machine Learning},

a. Abstract of the research paper - Skin disorders are widespread, arising from various causes such as fungi, bacteria, allergies, and viruses. While advanced medical technologies like laser and photonics have enhanced the speed and accuracy of diagnoses, their high costs limit accessibility. To address this issue, this paper proposes an image processing-based method for dermatological screening and diagnosis that utilizes digital photographs of affected skin areas. The approach integrates image resizing for feature extraction using a pretrained Convolutional Neural Network (CNN) and employs a multiclass Support Vector Machine (SVM) for classification. This cost-effective solution relies on minimal equipment—only a camera and a computer—and operates on color images to identify four distinct skin diseases with an impressive accuracy of 90.8%. By automating the detection process, this method offers comprehensive results regarding disease type, extent, and severity, thereby providing a faster and more objective alternative to manual diagnosis by medical professionals.

Keywords: Skin Disorder, SVM, K-NN, Decision Tree, Image Processing, Machine Learning.

b. Inference drawn from the paper -

1. **Cost-Effective Diagnostic Tool:** The proposed image processing method offers a low-cost alternative to expensive medical technologies, making skin disease diagnosis more accessible, especially in resource-constrained settings.
2. **High Accuracy in Classification:** Utilizing a combination of image resizing, pretrained CNN for feature extraction, and multiclass SVM for classification, the system achieves a high accuracy rate of 90.8% in detecting and classifying four distinct skin diseases.
3. **Efficiency and Speed:** Automated skin disease prediction significantly reduces the time required for diagnosis compared to manual methods, enabling faster treatment planning and intervention.
4. **Robust Feature Extraction:** The use of advanced feature extraction techniques such as Gray-Level Co-occurrence Matrix (GLCM), Gabor filters, Entropy, and Canny Edge Detection enhances the system's ability to accurately characterize and differentiate between various skin conditions.
5. **Superior Performance of SVM:** Among the classifiers tested—Support Vector Machines (SVM), K-Nearest Neighbors (K-NN), and Decision Trees (DT)—SVM consistently outperforms the others in terms of accuracy, precision, recall, and F1-score, making it the preferred choice for this application.

2.1.4:Machine Learning and Cloud-based Mobile App for Real Time Skin Cancer Prediction.

Citation: R. K. Kaushal, G. R. Hemalakshmi, N. B. Prakash, S. Maruthai, K. K. Gupta and N. L, "Machine Learning and Cloud-based Mobile App for Real Time Skin Cancer Prediction," 2023 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Kirtipur, Nepal, 2023, pp. 629-636, doi: 10.1109/I-SMAC58438.2023.10290392. keywords: {Analytical models;Feature extraction;Wavelet analysis;Iron;Skin;Real-time systems;Discrete wavelet transforms;Skin Cancer;Mobile Application;Data Pre-Processing;Discrete Wavelet Transform;Naïve Bayes;MIT App Inventor},

a. Abstract of the research paper -The research paper presents a comprehensive study on skin cancer detection using machine learning techniques. It focuses on feature extraction through Discrete Wavelet Transform (DWT) and evaluates three classifiers: Linear Discriminant Analysis

(LDA), Random Forest (RF), and Naïve Bayes (NB). A mobile application is developed for user-friendly skin cancer prediction. The study utilizes a dataset of 2,357 skin images, achieving high accuracy, particularly with the Naïve Bayes classifier, which reached 96.61%. The findings underscore the potential of ML algorithms in enhancing early skin cancer diagnosis.

Keywords: Skin Disease diagnosis, Image segmentation, Convolutional Neural Networks, Machine Learning, Image Classification

b. Inference drawn from the paper -

The paper concludes that machine learning classifiers, especially Naïve Bayes, can significantly improve the accuracy of skin cancer detection when combined with effective feature extraction methods like DWT. It emphasizes the importance of early detection in managing skin cancer and highlights the potential for mobile applications to make diagnostic tools accessible to a broader audience.

2.1.5: Lumpy Skin Disease Virus Detection on Animals Through Machine Learning Method.

Citation: P. Singh, J. Prakash, J. Srivastava and J. Srivastava, "Lumpy Skin Disease Virus Detection on Animals Through Machine Learning Method," 2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2023, pp. 481-486, doi: 10.1109/ICSCCC58608.2023.10176394. keywords: {Measurement;Economics;Computer viruses;Animals;Static VAr compensators;Machine learning;Production;Lumpy Skin Disease;Machine Learning Classifier;SVC},

a. Abstract of the research paper - Lumpy Skin Disease Virus (LSDV) is a highly contagious viral disease affecting cattle, leading to significant economic losses due to decreased milk production, infertility, stunted growth, and chronic skin lesions. Traditional diagnostic methods, while effective, are often time-consuming and require specialized equipment, limiting their accessibility, especially in resource-constrained regions. This paper presents a machine learning-based approach to detect and classify LSDV in animals using various classifiers, including Logistic Regression Classifier (LRC), Decision Tree Classifier (DTC), Random Forest

Classifier (RFC), XGBoost Classifier (XGBC), and Support Vector Classifier (SVC). The study utilizes a comprehensive dataset comprising 24,803 instances with 19 features related to meteorological and geospatial data over a 16-year period. After meticulous data preprocessing, including the removal of incomplete records and applying SMOTE oversampling to balance the classes, the classifiers were trained and evaluated. The Support Vector Classifier (SVC) demonstrated superior performance, achieving an accuracy of 90.8%, along with high precision, recall, and F1-score metrics. The proposed method offers a robust and efficient solution for early detection of LSDV, aiding in timely intervention and mitigating the spread of the disease within the livestock industry.

Keywords: Lumpy Skin Disease, Machine Learning Classifier, SVC, SVM, K-NN, Decision Tree, Image Processing

b. Inference drawn from the paper -

1. **High Accuracy with SVC:** The Support Vector Classifier (SVC) outperforms other machine learning models such as K-Nearest Neighbors (K-NN) and Decision Trees (DT) in detecting LSDV, achieving an accuracy of 90.8%.
2. **Importance of Data Preprocessing:** Effective data preprocessing, including the removal of incomplete records and the application of SMOTE oversampling, is crucial in enhancing the performance and reliability of machine learning models.
3. **Meteorological and Geospatial Features:** Incorporating meteorological and geospatial data significantly contributes to the accurate prediction of LSDV outbreaks, highlighting the importance of environmental factors in disease spread.
4. **Comparative Performance of Classifiers:** Among the classifiers tested, SVC and XGBoost Classifier (XGBC) exhibit the highest performance metrics, indicating their suitability for complex classification tasks in veterinary diagnostics.
5. **Economic Implications:** Early and accurate detection of LSDV using machine learning models can lead to timely interventions, reducing economic losses in the livestock sector by preventing the spread and severity of the disease.

2.1.6: Skin Disease Classification using Dermoscopy Images through Deep Feature Learning Models and Machine Learning Classifiers.

Citation: S. Gupta, A. Panwar and K. Mishra, "Skin Disease Classification using Dermoscopy Images through Deep Feature Learning Models and Machine Learning Classifiers," IEEE EUROCON 2021 - 19th International Conference on Smart Technologies, Lviv, Ukraine, 2021, pp. 170-174, doi: 10.1109/EUROCON52738.2021.9535552. keywords: {Temperature sensors;Deep learning;Pollution;Malignant tumors;Neural networks;Feature extraction;Skin;Machine Learning;Skin disease;Classification;Benign tumor;Melanoma},

a. Abstract of the research paper - Skin is one of the largest and most vital organs of the human body, responsible for regulating temperature and facilitating sensations such as touch, heat, and cold. Skin diseases have become increasingly common due to factors like unbalanced diets, pollution, and genetic predispositions. Among these, skin cancer, particularly melanoma, poses a significant threat due to its aggressive nature and high mortality rate if not detected early. Traditional diagnostic methods, such as dermoscopy, are time-consuming and subjective, often relying heavily on the expertise of dermatologists. To address these challenges, this paper presents an efficient and cost-effective machine learning-based approach for the early and accurate classification of skin cancer using dermoscopy images. The proposed methodology involves preprocessing a dataset of benign and malignant tumor images, followed by feature extraction using several Convolutional Neural Network (CNN) models, including VGG16, VGG19, and Inception V3. Extracted features are then fed into various machine learning classifiers such as Support Vector Machines (SVM), K-Nearest Neighbors (K-NN), Random Forest (RF), Neural Networks (NN), Logistic Regression (LR), and AdaBoost for classification. The results demonstrate that the Inception V3 model combined with a Neural Network classifier achieves the highest accuracy of 83.2%, outperforming other models and classifiers. This automated classification approach aids dermatologists in swiftly and accurately distinguishing between benign and malignant skin lesions, facilitating timely and appropriate treatment interventions.

Keywords: Machine Learning, Skin Disease, Classification, Benign Tumor, Melanoma, Dermoscopy, CNN, SVM, K-NN, Inception V3

b. Inference drawn from the paper -

1. Enhanced Diagnostic Accuracy: The integration of CNN models with machine learning classifiers significantly improves the accuracy of skin disease classification, with the Inception V3 model paired with a Neural Network achieving the highest accuracy of 83.2%.
2. Efficiency and Cost-Effectiveness: The proposed automated classification system offers a quicker and more cost-effective alternative to traditional dermoscopy methods, making skin cancer diagnosis more accessible, especially in resource-constrained settings.
3. Robust Feature Extraction: Utilizing advanced CNN architectures like VGG16, VGG19, and Inception V3 for feature extraction enhances the system's ability to capture complex patterns and textures in dermoscopy images, which are critical for accurate classification.
4. Superior Performance of Inception V3: Among the tested CNN models, Inception V3 demonstrates superior performance in feature extraction, contributing to higher classification accuracy compared to VGG16 and VGG19.
5. Effectiveness of SVM and Neural Networks: Support Vector Machines (SVM) and Neural Networks (NN) classifiers show strong performance in classifying extracted features, with Neural Networks providing the best results when combined with Inception V3.
6. Importance of Preprocessing: Effective preprocessing techniques, including image resizing, cropping, and hair removal using the Dull Razor method, are crucial for enhancing image quality and improving the reliability of subsequent feature extraction and classification processes.

2.1.7: A Transfer Learning-based Pre-trained VGG16 Model for Skin Disease Classification.

Citation: G. Singh, K. Guleria and S. Sharma, "A Transfer Learning-based Pre-trained VGG16 Model for Skin Disease Classification," 2023 IEEE 3rd Mysore Sub Section International Conference (MysuruCon), HASSAN, India, 2023, pp. 1-6, doi:

10.1109/MysuruCon59703.2023.10396942. keywords: {Deep learning;Training;Computational modeling;Medical services;Feature extraction;Skin;Diseases;Machine Learning (M.L);Skin Diseases;Deep learning (D.L);CNN;VGG16 Model},

a. Abstract of the research paper - Skin disorders are a significant global health concern, affecting millions of individuals and imposing substantial burdens on healthcare systems. Accurate and timely diagnosis is crucial for effective management and treatment of various skin conditions, including skin cancer. Deep learning models, particularly Convolutional Neural Networks (CNNs), have shown exceptional performance in medical imaging applications, enabling precise classification of skin diseases. This study leverages the VGG16 deep learning architecture for the early diagnosis of skin diseases. An extensive dataset comprising 44,000 images of benign and malignant tumors was collected from the open-source repository Kaggle. The VGG16 model was fine-tuned on this dataset to learn distinguishing patterns and characteristics associated with different skin conditions. The model's effectiveness was evaluated using standard metrics such as precision, recall, F1-score, and accuracy. The proposed deep learning model achieved a remarkable accuracy of 90.1%, with precision, recall, and F1-score values of 0.867, 0.942, and 0.891, respectively. These results demonstrate the model's proficiency in accurately diagnosing a wide range of skin diseases, including those with similar appearances. The study contributes to the advancement of computer-aided disease detection, offering enhanced healthcare outcomes through early detection and treatment of skin disorders. Future work emphasizes the need for continuous refinements and validation on larger, more diverse datasets to further enhance the model's accuracy and generalizability.

Keywords: Machine Learning (M.L), Skin Diseases, Classification, Benign Tumor, Melanoma, Dermoscopy, CNN, VGG16 Model, SVM, K-NN

b. Inference drawn from the paper -

1. High Diagnostic Accuracy: The fine-tuned VGG16 model achieves a high accuracy of 90.1% in classifying skin diseases as benign or malignant, indicating its effectiveness in medical diagnostics.

2. Effective Feature Extraction with VGG16: Leveraging the pre-trained VGG16 model for feature extraction proves beneficial in identifying complex patterns and characteristics in dermoscopy images, enhancing classification performance.
3. Importance of Data Preprocessing: The study emphasizes the critical role of data preprocessing techniques, such as image resizing, cropping, and hair removal using the Dull Razor method, in improving image quality and model accuracy.
4. Superior Performance of Neural Networks: Among various machine learning classifiers tested (SVM, K-NN, RF, NN, LR, AdaBoost), Neural Networks (NN) paired with the Inception V3 model exhibited the highest performance metrics, underscoring the potential of neural networks in skin disease classification.
5. Transfer Learning Advantage: Utilizing transfer learning with a pre-trained VGG16 model allows the research to benefit from previously learned features, reducing the need for extensive training and improving model generalization on new datasets.
6. Scalability and Accessibility: The proposed methodology is scalable and cost-effective, making it accessible for widespread use in clinical settings and potentially integrated into mobile applications for real-time skin disease detection.

2.1.8:Human Skin Diseases Identification and Treatment Suggestion by Sri Lankan Ayurveda Medicine using Machine Learning.

Citation: M. P. O. M. Gomes, Y. N. Jayasekara, K. M. K. R. Kariyapperuma, H. P. M. N. Gunawardhna, N. H. P. R. S. Suwarnakantha and G. Wimalarathne, "Human Skin Diseases Identification and Treatment Suggestion by Sri Lankan Ayurveda Medicine Using Machine Learning," 2023 5th International Conference on Advancements in Computing (ICAC), Colombo, Sri Lanka, 2023, pp. 65-70, doi: 10.1109/ICAC60630.2023.10417632. keywords: {Machine learning algorithms;Dermatology;Skin;Safety;Medical diagnostic imaging;Random forests;Diseases;skin diseases;deep learning;machine learning;CNN},

a. Abstract of the research paper - Skin diseases pose a significant global health risk, affecting millions of individuals and placing a substantial burden on healthcare systems. Accurate and timely diagnosis is crucial for effectively managing various conditions, particularly skin cancers like melanoma, which can be fatal if not detected early. This research paper introduces a novel approach that integrates Sri Lankan Ayurvedic medicine with machine learning techniques to

enhance the identification and treatment of human skin diseases. Utilizing deep learning models such as Inception, InceptionV3, and VGG16, the study aims to improve the accuracy of skin type detection, disease identification, and severity classification. Experimental results demonstrate that the Inception ResNet model achieves an 86% accuracy in identifying skin types, the InceptionV3 model attains a 97% accuracy in diagnosing different skin diseases, and the VGG16 model reaches a 96% accuracy in classifying disease severity. Additionally, a Random Forest algorithm is employed to provide treatment suggestions based on Ayurvedic principles, achieving an accuracy of 94.25%. This interdisciplinary research not only advances dermatological diagnostics but also highlights the potential of merging traditional Ayurvedic practices with modern machine learning technologies to improve health outcomes.

Keywords: Skin Diseases, Deep Learning, Machine Learning, CNN, VGG16 Model, InceptionV3, Ayurvedic Medicine, Random Forest

b. Inference drawn from the paper -

1. **Integration of Traditional and Modern Techniques:** Combining Sri Lankan Ayurvedic medicine with machine learning enhances the accuracy and effectiveness of skin disease diagnosis and treatment suggestions.
2. **High Accuracy of Deep Learning Models:**
   Inception ResNet: Achieves 86% accuracy in skin type identification.
   InceptionV3: Attains 97% accuracy in diagnosing various skin diseases.
   VGG16: Reaches 96% accuracy in classifying the severity of skin diseases.
3. **Effective Treatment Suggestions:** Utilizing a Random Forest algorithm based on Ayurvedic principles provides treatment suggestions with a high accuracy of 94.25%, demonstrating the potential for personalized and accurate treatment planning.
4. **Importance of Data Preprocessing:** Techniques such as image resizing, cropping, and normalization are crucial for enhancing model performance and ensuring accurate classifications.
5. **Scalability and Accessibility:** The proposed system, leveraging machine learning and Ayurvedic insights, offers a scalable and accessible solution for early skin disease detection, especially beneficial in regions with limited access to dermatological expertise.

Chapter 3: Requirements for the Proposed System

This chapter details the functional and non-functional requirements, along with the methodologies used to gather them. Requirements gathering is crucial to understanding user needs, system goals, and technical specifications. These requirements were gathered through research on existing systems, analysis of user feedback, and technical feasibility studies. The goal is to build a comprehensive system that meets the needs of both end-users and developers.

3.1. Functional Requirements

Functional requirements define the specific actions that **WellnessInsight** must perform. These include the core capabilities and features necessary for users to interact with the system effectively.

- **Image Upload and Prediction**: Users can upload images of affected skin areas, which the system processes to predict the disease type and infection percentage.
- **Textual Input**: Users can describe their symptoms via text, allowing the system to analyze and provide predictions based on the description.
- **Audio Input**: Users can record voice descriptions of their symptoms for audio-based analysis, offering another form of input for users with visual or textual limitations.
- **Multiclass Prediction**: The system can classify and predict three main diseases—psoriasis, eczema, and melanoma—and display the results with infection percentages.

Justification: These functional requirements are essential for providing users with multiple input options and accurate disease predictions, improving accessibility and user engagement.

3.2. Non-Functional Requirements

Non-functional requirements outline the quality attributes of the system, such as performance, security, and usability.

- **Performance**: The system must process inputs and provide predictions within a few seconds to ensure a smooth user experience.
- **Usability**: The user interface must be intuitive, allowing users to easily upload images, record audio, or type in symptoms without technical difficulties.
- **Accuracy**: The disease prediction algorithm must maintain a high level of accuracy to ensure reliable results.
- **Scalability**: The system must be able to handle multiple users simultaneously without performance degradation.

Justification: These non-functional requirements are crucial for creating a reliable and efficient system that enhances user satisfaction and system robustness.

3.3. Constraints

Constraints are limitations that must be considered during development.

- **Data Privacy**: Users' personal and health-related data must be protected, complying with data protection regulations like GDPR.
- **Limited Dataset**: The system may be constrained by the availability of high-quality, labeled skin disease images and data.
- **Resource Constraints**: Depending on the hardware and cloud resources available, system training and deployment may face computational or financial limitations.

Justification: Acknowledging these constraints ensures the development of a legally compliant and technically feasible system.

3.4. Hardware & Software Requirements

The hardware and software requirements specify the technical environment in which **WellnessInsight** will operate.

- **Hardware**: A system with a good GPU for model training, cloud computing resources for scaling, and user devices such as smartphones and computers for system access.
- **Software**:
  - **Python**: For backend development and machine learning model implementation.
  - **TensorFlow**: For training the convolutional neural networks (CNNs) used in disease prediction.
  - **Google Colab**: For training and running the model.
  - **OpenCV**: For image preprocessing tasks.
  - **NLP Tools**: For processing textual descriptions.

Justification: The hardware and software requirements provide the necessary infrastructure for system development, training, and deployment.

3.5. Techniques Utilized Till Date for the Proposed System

Several techniques have been employed in the development of **WellnessInsight**:

- **Image Preprocessing**: Techniques like resizing, noise reduction, and image normalization have been used to improve the input image quality.
- **NLP for Text**: Tokenization and normalization techniques have been applied to textual data inputs for better analysis.

Justification: These techniques enhance the accuracy and performance of the system, ensuring that inputs are prepared in a way that optimizes the prediction model.

3.6. Tools Utilized Till Date for the Proposed System

Various tools have been employed to develop **WellnessInsight**:

- **Google Colab**: Used for model training and running the code.
- **TensorFlow and Keras**: For implementing and training the prediction models.
- **OpenCV**: Used for image processing and handling image inputs.
- **MFCCs (for Audio)**: Used to extract features from audio inputs.

Justification: These tools provide the necessary functionality to implement and test the different components of the system.

3.7. Algorithms Utilized in Existing Systems

Several algorithms have been commonly used in existing skin disease detection systems:

- **Convolutional Neural Networks (CNNs)**: Widely used for image classification tasks, such as identifying skin diseases from images.
- **Natural Language Processing (NLP) Algorithms**: Applied for analyzing textual descriptions of symptoms.
- **Support Vector Machines (SVM)**: Used in some systems for classification tasks, especially where feature extraction is involved.
- **Random Forest Algorithm**: Employed for classification tasks where multiple features are considered simultaneously, providing a robust method for predicting skin disease based on structured input data.

Justification: These algorithms have proven effective in existing systems for similar tasks, making them relevant for this project as well.

3.8. Proposal

WellnessInsight is proposed as a comprehensive, AI-driven skin disease prediction system. It integrates image, text, and audio inputs for multiclass classification of common skin diseases, focusing on providing accurate predictions and infection percentages. Additionally, it addresses the lack of accessibility to dermatological care by recommending nearby healthcare professionals based on the user's location. The system is built using state-of-the-art machine learning techniques and modern development tools, ensuring high performance and scalability.

Justification: This proposal is based on a thorough analysis of both functional and non-functional requirements, constraints, and the current technological landscape, ensuring a user-friendly and reliable system for skin disease detection.

Chapter 4: Proposed Design

4.1 System Design / Conceptual Design (Architectural)

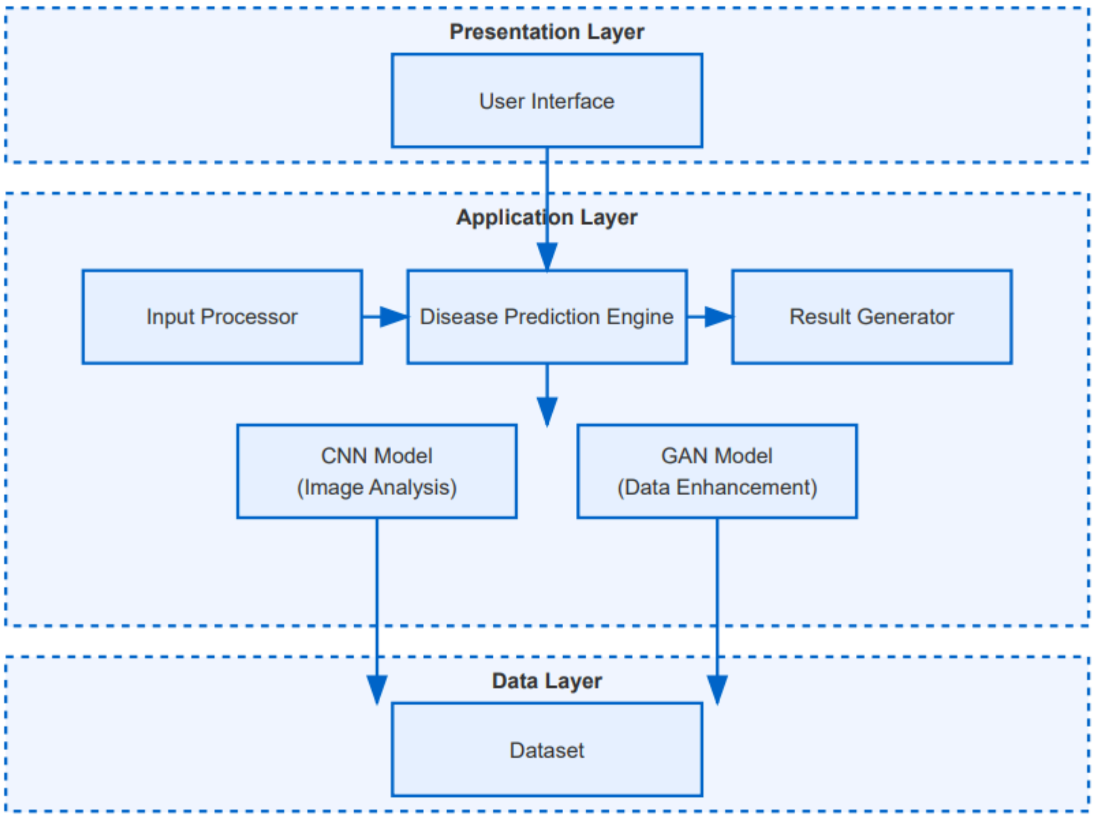

Fig. 1

Explanation of System Design / Conceptual Design (Architectural) : This illustrates a layered architecture for a skin disease prediction system. It consists of three layers: the Presentation Layer, which handles user interactions via the User Interface; the Application Layer, featuring components like the Input Processor, CNN for image analysis, GAN for data enhancement, and the Disease Prediction Engine; and the Data Layer, where the Dataset stores relevant data. Arrows depict the flow of data between these components, highlighting the sequential processing from input to result generation.

4.2. Block diagram representation of the proposed system

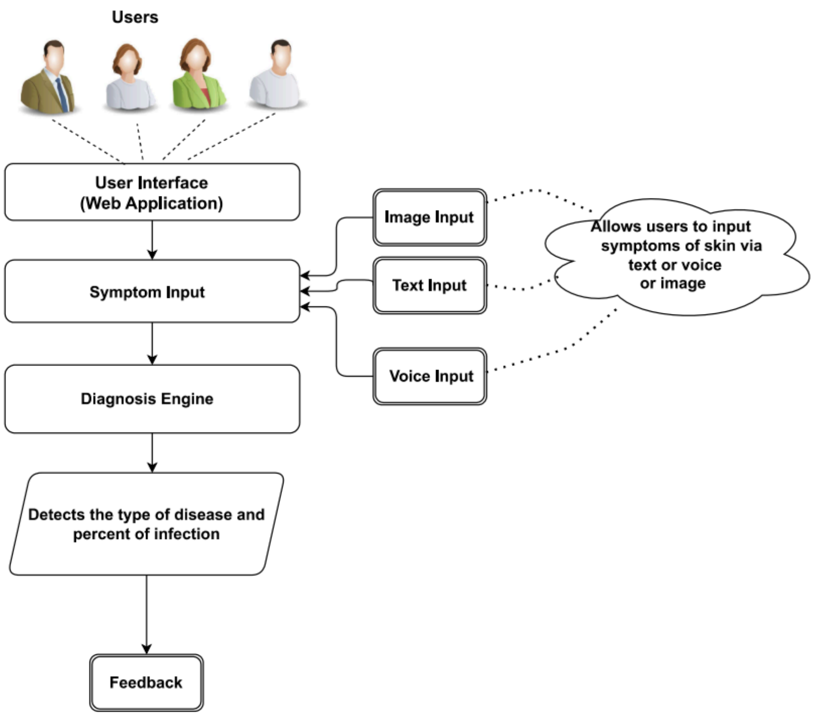

Fig. 2

Explanation of block diagram : The block diagram represents the WellnessInsight system's workflow. The user provides inputs in the form of text, audio, or images, which are then processed by the diagnosis engine. This engine applies AI algorithms such as CNNs for image classification, NLP for text analysis, and deep learning models for audio processing. Based on the input, the system predicts the skin disease and provides the infection percentage. The output is then displayed to the user.

4.3. Modular diagram representation of the proposed system

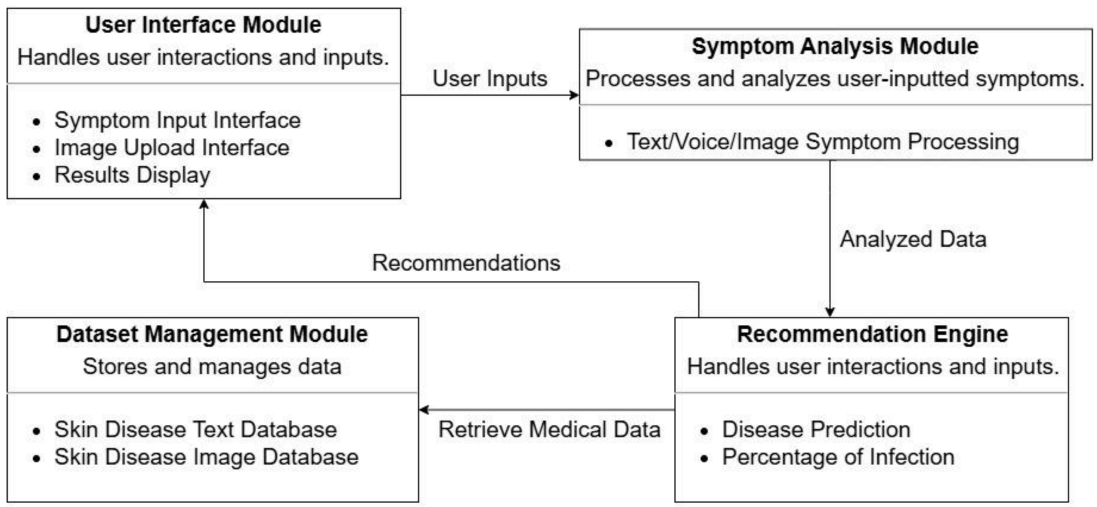

Fig. 3

Modular Diagram Explanation:

The **WellnessInsight** system is organized into four core modules, each handling a specific part of the skin disease prediction process:

1. **User Interface Module:** This front-end component allows users to input symptoms through text or upload images of their skin conditions. It also displays the results after analysis. The user inputs are sent to the next module for processing.
2. **Symptom Analysis Module:** This module processes the inputs (text, voice, or image). It analyzes the data to identify potential skin diseases and sends the analyzed information to the recommendation engine.
3. **Dataset Management Module:** This module manages and stores all the necessary data, including text and image datasets of various skin diseases. It retrieves relevant medical data when needed to ensure accurate predictions.
4. **Recommendation Engine:** Based on the analyzed data, this module predicts the disease and calculates the percentage of infection. It sends the results back to the User Interface Module, which then displays the predictions and recommendations to the user.

This modular design ensures efficient, accurate predictions and provides users with detailed disease information and recommendations.

4.4 Design of the proposed system with proper explanation of each :

a. Data Flow Diagram (Level 0,1,2)

- **level 0 :**

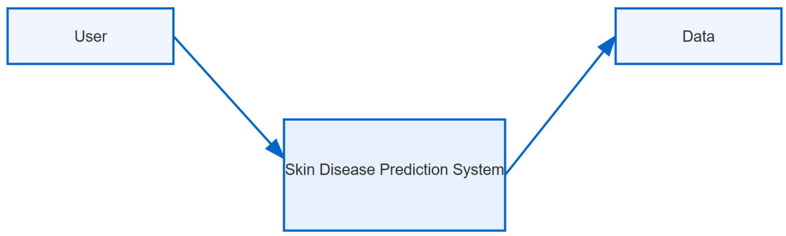

Fig. 4

The Level 0 DFD provides a high-level overview of the Skin Disease Prediction System, illustrating its interaction with external entities like the User and Dataset. It emphasizes the overall system functionality by depicting how data flows between the user and the system, highlighting the primary input and output processes.

- **level 1 :**

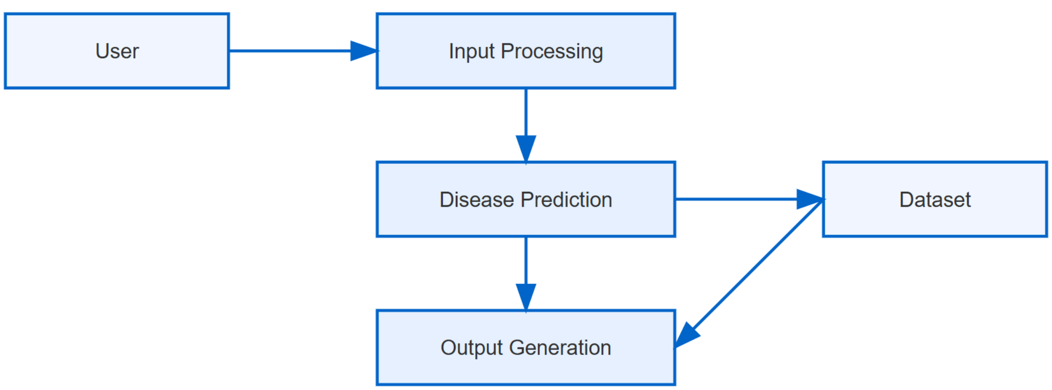

Fig. 5

The Level 1 DFD breaks down the system into three primary processes: Input Processing, Disease Prediction, and Output Generation. It illustrates how data from the user is processed through input validation and then sent to the dataset for disease prediction, with the generated output flowing back to the user. This level captures the system's major functions and their interconnections, showcasing the internal workings of the system.

- **level 2 :**

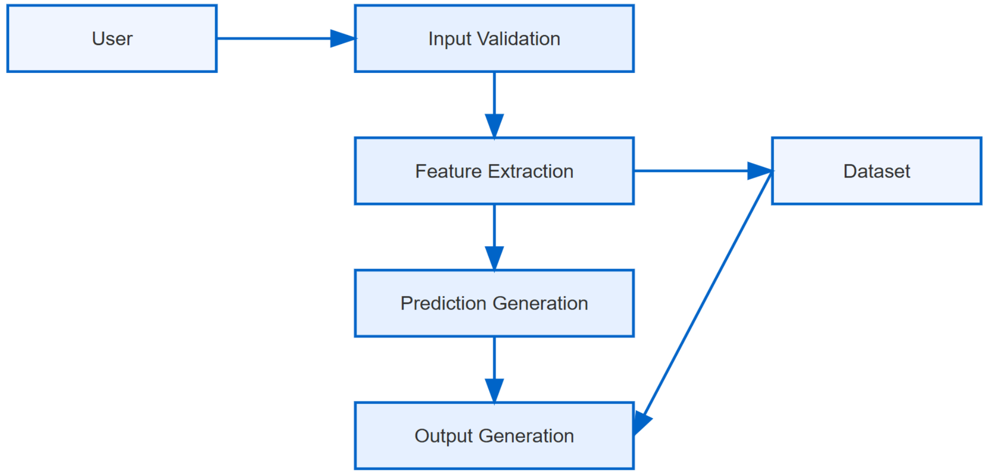

Fig. 6

The Level 2 DFD further dissects the Disease Prediction System into detailed subprocesses, including Input Validation, Feature Extraction, Prediction Generation, and Report Generation. This diagram illustrates the specific steps involved in the prediction process, including how user inputs are validated and processed to generate a disease prediction report, demonstrating the system's complexity and the flow of information between various functions.

b. Flowchart for the proposed system

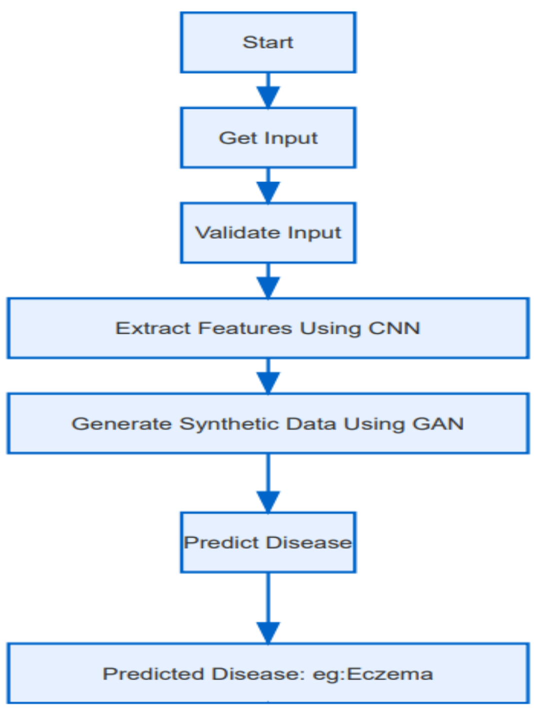

Fig. 7

Explanation : The flow diagram outlines the step-by-step process for predicting a disease, such as eczema, using machine learning. It includes stages like input collection, validation, model training with CNN , and displaying results, providing a clear visual representation of the operational workflow.

d. ER Diagram

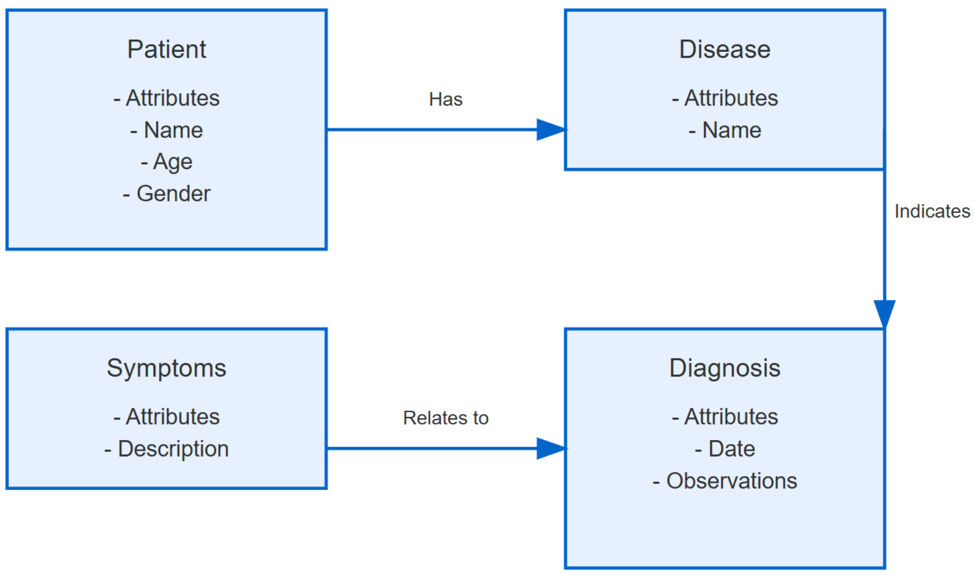

Fig. 8

Explanation : The ER diagram depicts the relationships between entities in the disease prediction system, such as **Users**, **Diseases**, **Symptoms**, and **Predictions**. It illustrates how users input symptoms to receive disease predictions, serving as a blueprint for the system's database structure.

e. Screenshot of implementation :

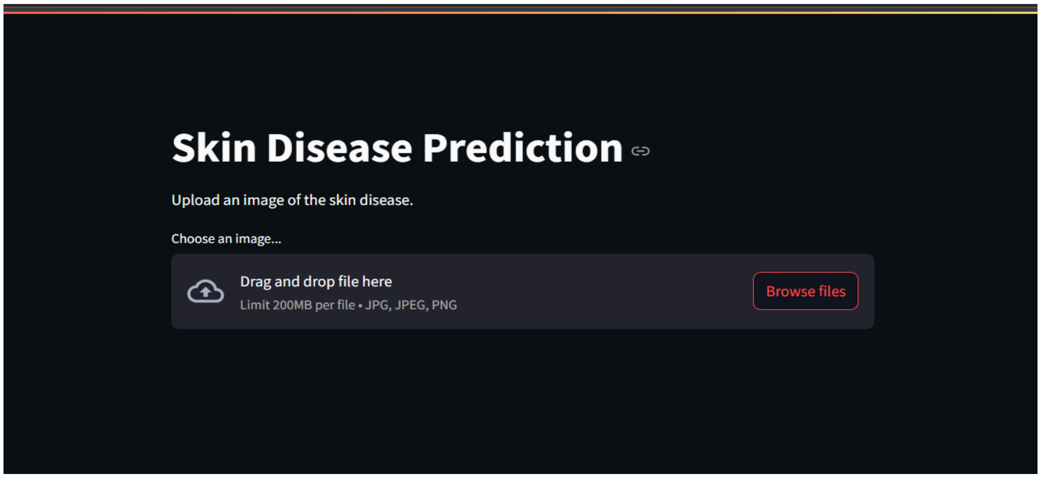

Fig. 9 Home page

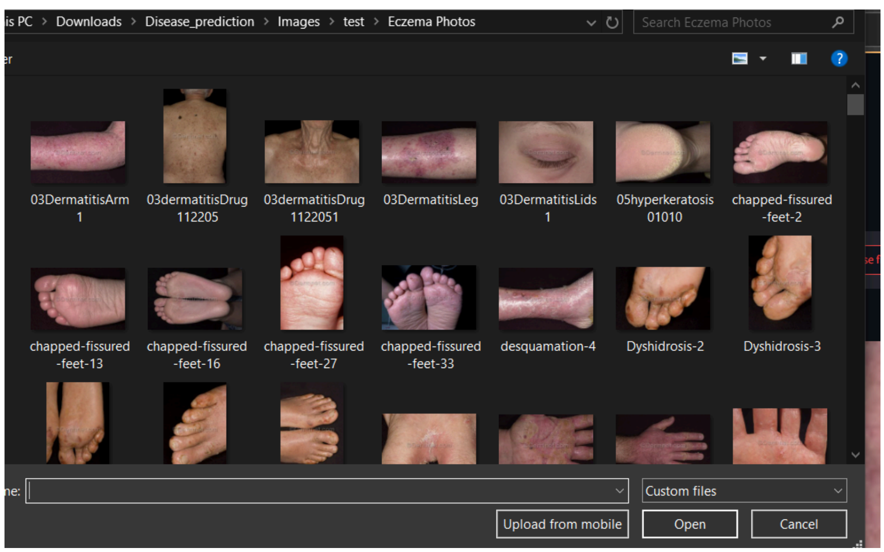

Fig. 10 Choose an image

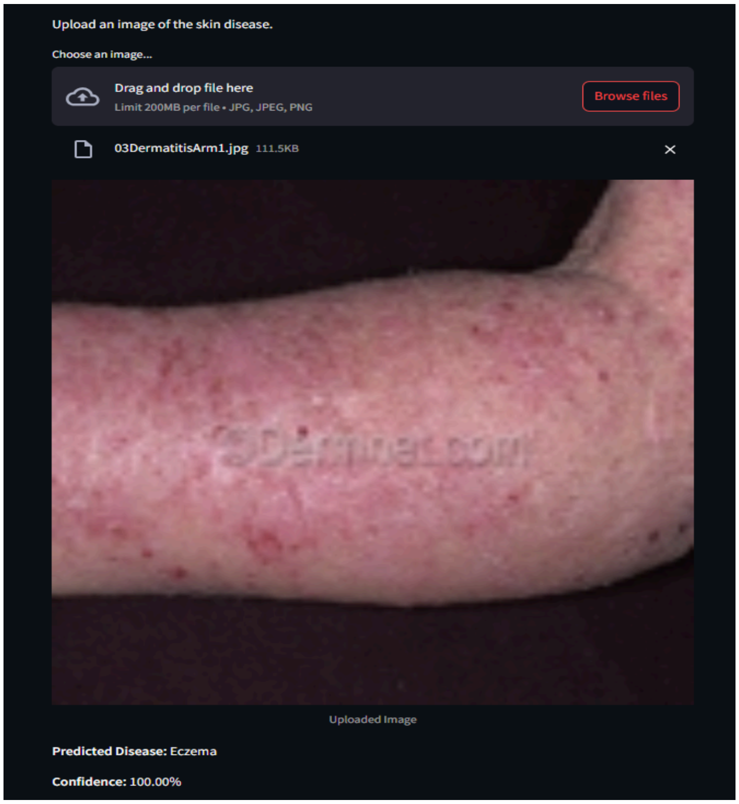

Fig. 11 Upload an image

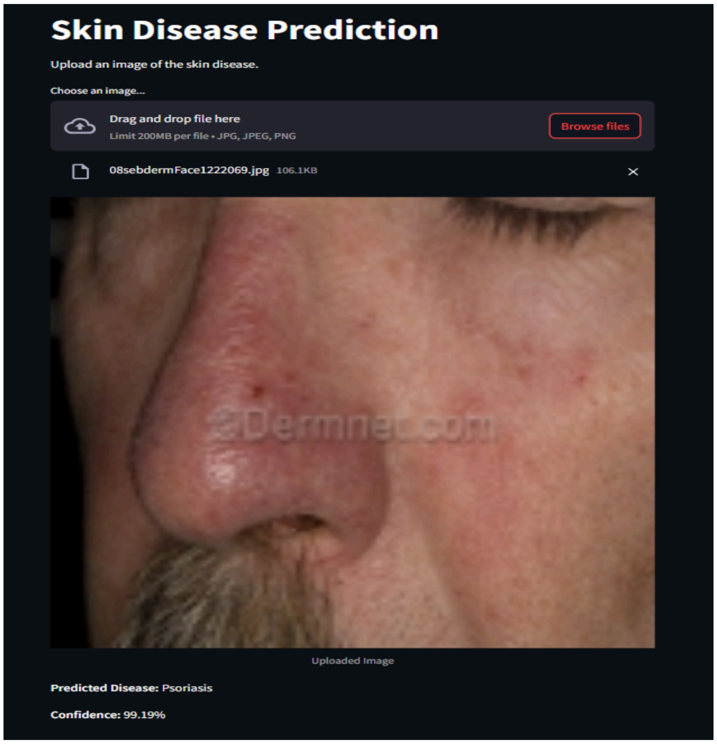

Fig. 12 Prediction of disease

4.5 Algorithms Utilized in the Existing Systems

Various algorithms are commonly used in existing skin disease detection systems, each offering unique advantages for different aspects of image classification, text analysis, and audio processing. These algorithms play a crucial role in ensuring accurate predictions and reliable results.

- **Convolutional Neural Networks (CNNs)**: CNNs are widely used in medical image analysis, particularly for skin disease detection. These networks excel at identifying visual patterns in images, making them highly effective for recognizing different types of skin conditions such as psoriasis, eczema, and melanoma. CNNs process input images through multiple layers, including convolutional, pooling, and fully connected layers, to extract important features that aid in disease classification.
- **Support Vector Machines (SVMs)**: Some existing systems employ SVMs for skin disease classification. SVMs are known for their high accuracy in binary classification tasks and are often used when the number of classes is limited. SVMs work by finding the optimal hyperplane that separates data points of different classes, making them a robust choice for simpler classification problems.
- **K-Nearest Neighbors (KNN)**: KNN is a simpler, yet effective, algorithm used in some dermatological systems. It classifies skin diseases by comparing the input image to similar images in the training dataset. KNN calculates the "distance" between the input and training data points and assigns the class label based on the majority vote from its "neighbors."
- **Random Forest**: The Random Forest algorithm is also widely used in skin disease detection systems. This ensemble learning method builds multiple decision trees and merges them to produce a more accurate and stable prediction. Random Forest is particularly useful when the data has a mix of categorical and continuous features. It is known for handling large datasets and providing reliable classification results.
- **Natural Language Processing (NLP) Algorithms**: For systems that accept text input, NLP algorithms like **Tokenization**, **Stemming**, and **Lemmatization** are used to process and understand symptom descriptions provided by users. These algorithms extract meaning from the text input, converting it into a format that can be used by machine learning models for disease prediction.
- **Audio Feature Extraction**: In systems that utilize audio inputs, techniques such as **Mel-Frequency Cepstral Coefficients (MFCCs)** are applied to extract relevant features from audio recordings. MFCCs capture the essential characteristics of speech or sound, which can then be processed by machine learning models to detect disease-related symptoms described verbally.

4.6 Project Scheduling & Tracking using Timeline / Gantt Chart

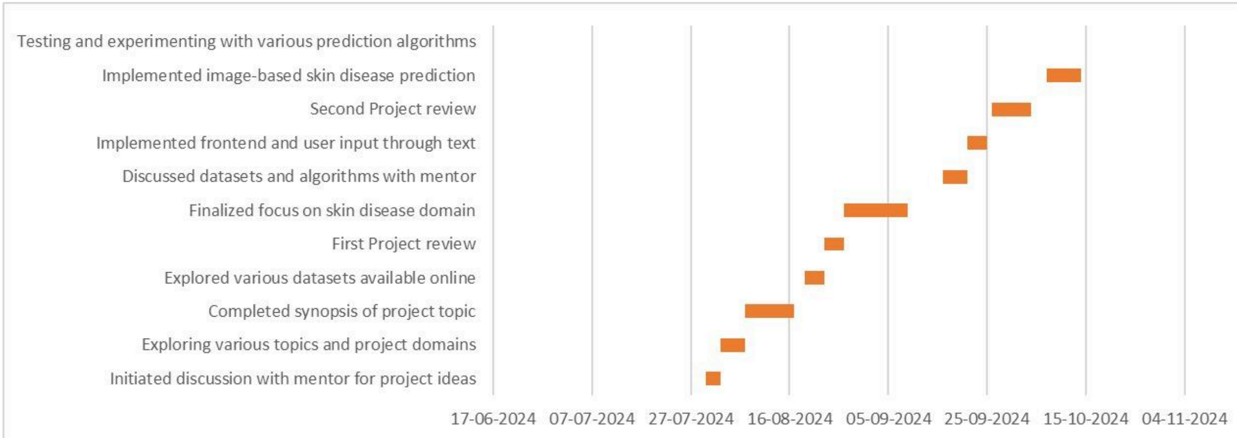

Fig. 13 Gantt Chart

The gantt chart outlines the project's timeline, tasks and dependencies visually. Each task is represented by a bar indicating its duration, and milestones mark significant project points. It helps track progress, identify the delays, and ensure timely completion.

Chapter 5: Proposed Results and Discussions

This chapter presents the anticipated results of the WellnessInsight system, focusing on its performance in terms of efficiency, accuracy, and sensitivity analysis. The discussion includes expected outcomes, evaluation metrics, and visual representations of system performance.

5.1. Determination of Efficiency

The efficiency of WellnessInsight will be determined by measuring the time taken for image, text, and audio inputs to be processed and return predictions. The system should process inputs swiftly, with minimal delays, ensuring a smooth user experience. Efficiency will be measured by calculating the average response time for different inputs across multiple test cases.

Expected Outcome: The system should demonstrate fast prediction times, typically under a few seconds per input.

5.2. Determination of Accuracy

The accuracy of the system is critical in determining the reliability of the predictions for psoriasis, eczema, and melanoma. The prediction model's accuracy will be assessed using metrics such as:

- Confusion Matrix: To show correct and incorrect predictions.
- Precision, Recall, and F1-Score: To analyze prediction performance in more detail.
- Overall Accuracy: The percentage of correct predictions across all test cases.

Expected Outcome: The model is expected to achieve high accuracy, likely above 85%, based on the training of the Random Forest and CNN models.

5.3. Reports on Sensitivity Analysis

Sensitivity analysis will focus on the system's ability to correctly predict diseases based on variations in input data. This involves determining how sensitive the model is to small changes in input, such as variations in image quality, text description, or audio recordings. Sensitivity analysis helps to identify the model's robustness and adaptability across different data inputs.

Expected Outcome: The system should show reasonable sensitivity, meaning small variations in input should not significantly affect prediction results.

5.4. Graphs: Accuracy vs. Time

Graphs plotting accuracy over time will illustrate the system's learning curve and performance improvements as it trains on more data. The graph will show how the system's accuracy improves after each iteration during training and testing phases.

- X-axis: Represents time or the number of epochs (iterations).
- Y-axis: Represents the accuracy of the predictions.

Expected Outcome: The graph should show a positive trend, with accuracy increasing as the system is trained further and stabilizing at a high level over time.

Chapter 6: Plan of action for the next semester

a. Work done till date

- Developed a clear understanding of the WellnessInsight project, which focuses on AI-driven skin disease prediction. The system is designed to predict psoriasis, eczema, and melanoma based on image and text inputs, aiming to bridge gaps in early diagnosis and care accessibility.
- Functional and non-functional requirements have been defined, including core system capabilities such as image upload, disease prediction, text analysis, and system performance goals.
- The conceptual architecture, modular design, and data flow diagrams have been created, defining a layered structure and the main components, including user interaction, data processing, and the recommendation engine.
- Tool & Techniques Implementation: Tools such as TensorFlow, Keras, OpenCV, and Google Colab were utilized to build and train the initial models, while preprocessing techniques for images, text, and audio were implemented to improve prediction accuracy.

b. Plan of action for project II

- Implement and test GANs (Generative Adversarial Networks) for data augmentation to enhance the training dataset.
- Integrate audio analysis for skin disease prediction and ensure seamless switching between input modes (image, text, and audio).

Chapter 7: Conclusion

In summary, **WellnessInsight** is a skin disease prediction system designed to provide quick and accurate results for conditions like psoriasis, eczema, and melanoma. By utilizing image, text, and audio inputs, along with advanced algorithms such as CNNs and Random Forest, it offers an accessible and efficient solution for early diagnosis. The system's functional and non-functional requirements, along with the tools and techniques used, ensure that it meets user needs while maintaining performance and scalability. Ultimately, **WellnessInsight** addresses a critical gap in dermatological care by helping users identify skin diseases and connect with nearby healthcare professionals for further assistance.

Chapter 8. References

[1] Nirupama and Virupakshappa, "Survey on Classification of Skin Diseases Using Machine Learning Techniques," 2024 3rd International conference on Power Electronics and IoT Applications in Renewable Energy and its Control (PARC), Mathura, India, 2024, pp. 135-140, doi: 10.1109/PARC59193.2024.10486701. keywords: {Surveys;Renewable energy sources;Machine learning algorithms;Reviews;Psychology;Machine learning;Prediction algorithms;Skin diseases;Identification;Classification;Machine Learning;Artificial intelligence},

[2] J. Sathya and A. Kalaivani, "Machine Learning Based Skin Disease Detection Using Semi-Automatic Image Segmentation," 2024 International Conference on Smart Systems for Electrical, Electronics, Communication and Computer Engineering (ICSSEECC), Coimbatore, India, 2024, pp. 215-219, doi: 10.1109/ICSSEECC61126.2024.10649509. keywords: {Image segmentation;Machine learning algorithms;Machine learning;Nearest neighbor methods;Feature extraction;Skin;Classification algorithms;Skin Disease diagnosis;Image segmentation;Convolutional Neural Networks;Machine Learning;Image Classification},

[3] N. H. Sany and P. Chandra Shill, "Image Segmentation based Approach for Skin Disease Detection and Classification using Machine Learning Algorithms," 2024 International Conference on Integrated Circuits and Communication Systems (ICICACS), Raichur, India, 2024, pp. 1-5, doi: 10.1109/ICICACS60521.2024.10498287. keywords: {Support vector machines;Pathogens;Machine learning algorithms;Machine learning;Skin;Medical diagnosis;Medical diagnostic imaging;Skin Disorder;SVM;K-NN;Decision Tree;Image Processing;Machine Learning},

[4] R. K. Kaushal, G. R. Hemalakshmi, N. B. Prakash, S. Maruthai, K. K. Gupta and N. L, "Machine Learning and Cloud-based Mobile App for Real Time Skin Cancer Prediction," 2023 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Kirtipur, Nepal, 2023, pp. 629-636, doi: 10.1109/I-SMAC58438.2023.10290392. keywords: {Analytical models;Feature extraction;Wavelet analysis;Iron;Skin;Real-time systems;Discrete wavelet transforms;Skin Cancer;Mobile Application;Data Pre-Processing;Discrete Wavelet Transform;Naïve Bayes;MIT App Inventor},

[5] D. H. Patil, M. Pawar, M. Jaiswal, P. Rane and S. Jagtap, "Lumpy Skin Disease Prediction Using Machine Learning.," 2023 4th IEEE Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2023, pp. 1-5, doi: 10.1109/GCAT59970.2023.10353350. keywords: {Support vector machines;Image recognition;Cows;Feature extraction;Skin;Agriculture;Vaccines;Random Forest Classifier;Lumpy skin disorders;Neural Networks;LSD;F1},

[6] S. Gupta, A. Panwar and K. Mishra, "Skin Disease Classification using Dermoscopy Images through Deep Feature Learning Models and Machine Learning Classifiers," IEEE EUROCON 2021 - 19th International Conference on Smart Technologies, Lviv, Ukraine, 2021, pp. 170-174, doi: 10.1109/EUROCON52738.2021.9535552. keywords: {Temperature sensors;Deep learning;Pollution;Malignant tumors;Neural networks;Feature extraction;Skin;Machine Learning;Skin disease;Classification;Benign tumor;Melanoma},

Chapter 9. Appendix

a. List of Figures