An Optimized Deep Learning-Based System for Accurate Detection and Classification of Skin Diseases

Daniel Makolo, Dr. Asogwa Tochuku. C, Friday Ameh

Department of Computer science, Faculty of physical sciences, Enugu State University of Science and Technology, Enugu. Nigeria

ABSTRACT

Deep learning has become a vital tool in medical image analysis, particularly for dermatology, where early detection of skin diseases is critical. This study presents an optimized system for automatic classification of skin conditions using dermoscopic images. A MobileNetV2-based convolutional neural network (CNN) was fine-tuned with transfer learning to enhance performance across multiple skin disease categories. Images were preprocessed through resizing and normalization before classification. To improve interpretability, Gradient-weighted Class Activation Mapping (Grad-CAM) was integrated to visualize discriminative regions. The system was evaluated using accuracy, sensitivity, specificity, precision, recall, F1-score, and AUC-ROC. Results demonstrate promising accuracy but reveal limited sensitivity for certain classes, reflecting challenges of dataset imbalance and visual similarity across conditions. The model's deployment through a Streamlit-based interface enables real-time predictions and interactive visualization, offering potential for use as a screening tool in resource-constrained settings. Future work should emphasize validation on larger, diverse datasets and explore advanced augmentation strategies to enhance generalization.

Keywords: Deep Learning, Convolutional Neural Networks (CNN), Dermoscopic Images, Deep Neural Networks (DNN), Confusion Matrix, International Skin Imaging Collaboration (ISIC), DermNet, PH^2 Dataset, AUC-ROC, F1-scores, Dermatology, Convolutional Network Disease (CND),

INTRODUCTION

Deep Learning is a specialized branch of artificial intelligence that uses multilayer neural networks to automatically extract meaningful patterns from raw data without explicit rule-based programming. Each layer progressively transforms inputs into higher-level representations, enabling the system to discover hidden relationships within large datasets [1].

Over the past decade, deep learning has achieved remarkable advances in image recognition and natural language processing. Most models are trained with supervised learning, where input–output pairs (such as dermoscopic images and disease labels) guide classification. These achievements have had substantial impact in medical science, particularly in pathology, radiology, and dermatology, where diagnosis depends on detecting subtle morphological differences. By integrating medical images into automated deep learning frameworks, **computer-aided diagnostic (CAD) systems** have emerged to support clinicians in making faster and more accurate decisions. Such systems have already demonstrated strong potential in detecting cancers, ocular diseases, and neurological disorders [2].

Skin diseases are a growing global health concern. In 2013, they ranked 18th among causes of Disability-Adjusted Life Years (DALYs) across 188 countries, with prevalence increasing by 42.7% compared to previous decades [3]. Affecting individuals of all ages, dermatological conditions impose not only physical symptoms but also psychological distress, social stigma, and financial burden. With more than 3,000 documented skin disorders, conditions like vitiligo and psoriasis significantly impair quality of life, while malignant melanoma remains a serious life-threatening disease [4].

Traditional diagnosis in dermatology often relies on laboratory tests and clinical evaluation, which are resource-intensive and subject to variability across practitioners. In contrast, deep learning–based computer vision methods provide a faster and more cost-effective alternative. By analyzing high-resolution images of affected skin, these systems can classify disease categories and assess severity with increasing accuracy, offering a promising approach to improve early detection and treatment outcomes [5].

LITERATURE REVIEW

2.1 Review of Related Literatures

Deep learning has become a powerful tool in medical imaging, offering advanced methods for disease detection and diagnosis. Dermatology is particularly well suited to these approaches due to the complexity and variability of skin conditions. Venu et al. explored the use of advanced architectures such as VGG19 and Inception ResNetV2, demonstrating strong feature extraction capabilities for skin disease diagnosis. Given the high morphological variation of skin lesions, such models provide a strong foundation for early intervention and automated clinical support. [6].

Muhammad et al. [2] further advanced this work by training deep neural networks on large datasets such as DermNet and the ISIC Archive. Their model achieved 80% accuracy and 98% AUC on DermNet images for 23 disease classes, and 93% accuracy with 99% AUC on ISIC data for seven disease categories. These results underscore the potential of deep learning for large-scale, real-time skin disease classification when paired with clinician expertise.

The success of such models relies heavily on data preparation. Medical images often vary in resolution and require preprocessing, such as resizing, normalization, and augmentation, to preserve diagnostic features. As Poornima and Sakkari [7]. note, data scarcity remains a major challenge due to patient privacy, rare disease occurrence, and high labeling costs. To address this, augmentation techniques—including rotation, flipping, filtering, and neural style transfer—are widely applied to improve dataset diversity.

2.2 Machine Learning

Machine Learning (ML) is a branch of computer science focused on enabling computers to learn from data without explicit programming. Closely tied to statistics, ML systems can classify, cluster, and predict with high accuracy, making them particularly effective where conventional algorithms fail [8],[9].

One of its strongest applications is predictive analytics, where models forecast outcomes from historical and real-time data. In cancer research, ML has been used to analyze complex gene expression data. [10]. proposed the LASSO–MOGAT framework, which integrates mRNA, microRNA, and DNA methylation data. Using graph-based attention mechanisms and fivefold cross-validation, their model accurately classified 31 cancer types, highlighting ML's capacity to uncover complex biological interactions.

2.3 Computer-Based Diagnosis of Skin Diseases

Advances in computer-aided diagnosis (CAD) have accelerated dermatological imaging analysis. By processing large volumes of skin images, AI-enabled CAD systems can detect subtle patterns that may be overlooked during manual review [8].

Skin cancer, particularly melanoma, is highly aggressive with much lower survival rates compared to basal cell carcinoma (BCC) or squamous cell carcinoma (SCC). Modern CAD systems typically involve four stages: preprocessing, segmentation, feature extraction, and classification. Preprocessing enhances image clarity, segmentation isolates regions of interest, and classification is performed using algorithms ranging from support vector machines to convolutional neural networks [11]. Deep learning methods have recently improved diagnostic accuracy further. For example, [12].reported CNN-based classification achieving 70% accuracy on a dataset of 938 images, covering conditions such as melanoma, nevus, and seborrheic keratosis. Although performance varies by dataset size and quality, such results show clear promise for early-stage diagnosis.

2.4 Skin Disease Diagnosis with Deep Learning

Deep learning has become a central approach in dermatological diagnosis because of its ability to automatically learn complex patterns from images. Recent studies have applied convolutional neural networks (CNNs) and related frameworks to classify skin conditions, often achieving accuracy levels comparable to or surpassing human experts. These systems typically rely on preprocessing steps such as resizing, normalization, and augmentation to improve generalization across diverse datasets.

Different architectures—including transfer learning models and attention-based networks—have further enhanced recognition of subtle variations in lesion morphology, texture, and color. While results are promising, many studies still report **low sensitivity for certain classes and misclassification of visually similar conditions**, underscoring the need for larger, more diverse datasets and advanced model optimization.

Figure 1 presents the taxonomy used to categorize the reviewed literature.
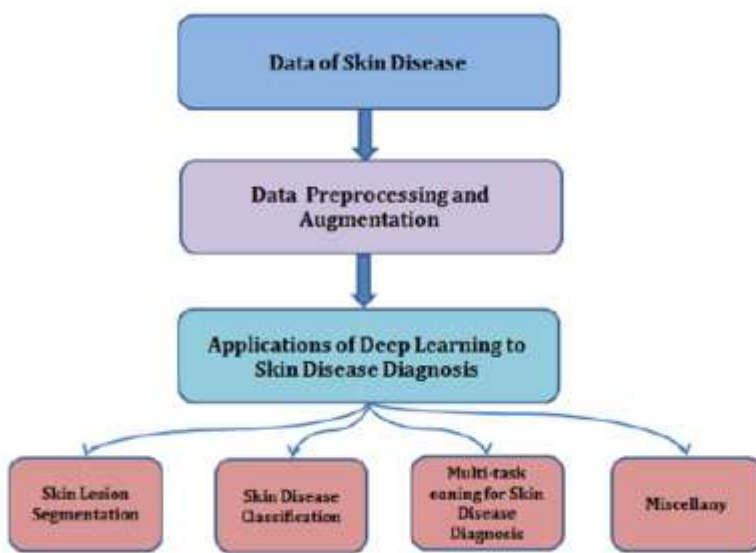

Figure 1. Accurately categorising skin disease diagnostic literature using deep [7].

Figure 1 presents a flowchart showing the overall process of applying deep learning to skin disease diagnosis. The process is divided into three main stages, each representing an essential

Step 1: Gathering Skin Disease Data

Figure 1: **Accurately categorising skin disease diagnostic literature using deep** [7].

Figure 1 above illustrates a flowchart that outlines the process of using deep learning for skin disease diagnosis. The flowchart is divided into three main sections, each representing a crucial step in the diagnosis process.

Step 1: Data of Skin Disease: The first step involves collecting data related to skin diseases. This data serves as the foundation for the subsequent steps in the diagnosis process.

Step 2: Data Preprocessing and Augmentation

Once collected, the data goes through cleaning, preprocessing, and improvement. This ensures the images are standardized, consistent, and diverse enough for reliable analysis.

Step 3: Applying Deep Learning for Diagnosis

In the final stage, the prepared data is used to train and test deep learning models for skin disease detection. This phase can be analyzed in four main parts:

1. **Skin Lesion Segmentation** – isolating affected areas from healthy skin.
2. **Disease Classification** – categorizing skin conditions into specific types based on their features.
3. **Multi-task Learning** – using a single model to handle multiple related tasks, such as segmentation and classification together.
4. **Other Applications** – any additional deep learning uses in dermatology that don't fall into the above categories.

In essence, the process begins with gathering skin disease data, then refining it, and finally applying deep learning to improve detection accuracy. This technique can extensively increase diagnostic efficiency and lead to better treatment results.

According to Shravani et al, skin complications are among the most common health problems worldwide, influencing individuals of every age. Early detection and prevention can greatly improve recovery rates and mitigate the impact of these conditions. To address this, [1]., designed a model that uses deep learning and convolutional neural networks (CNNs) to identify and diagnose skin diseases more accurate. The system is trained on a large collection of dermatology images from Kaggle and is capable of recognizing conditions such as acne, blisters, eczema, and warts. With an accuracy of 83.23%, the system not only predicts the condition but also provides users with detailed information and possible home remedies for management.

2.5 Importance of Deep Learning in Healthcare IoT

The integration of deep learning with the healthcare Internet of Things (IoT) has transformed disease detection, monitoring, and management. Deep learning's ability to process large, complex datasets makes it well suited for IoT-enabled devices, from smartphones to wearable [13]. Liu et al. identified five factors essential for success: big data, IoT infrastructure, deep learning, GPU-powered computing, and trained medical professionals. In medical imaging, deep learning models automatically learn hierarchical features without manual intervention, improving diagnostic outcomes. [14] reported strong results in cancer detection, Alzheimer's diagnosis, and ophthalmology, while Bratman et al. [15] noted its potential in prescription optimization and error reduction.

As healthcare data grows, demand for AI expertise also increases. [16].emphasized that now is the critical time for deep learning adoption to improve diagnostics and streamline care. Public datasets and AI-driven research are also accelerating drug discovery and treatment innovation [17].

2.6 Dermatological Classification

Skin cancer remains one of the most common cancers worldwide, arising from abnormal, uncontrolled growth of skin cells. Diagnosis and treatment vary by cancer type, making accurate classification critical.

Automated systems have emerged as valuable tools to support dermatologists. For example, [18].developed a CNN-based method combining image data with patient metadata. Their approach improved classification accuracy from 79.29% with images alone to 80.39% when including patient context, highlighting the value of integrated data in dermatological diagnosis.

2.7 Removal of Irrelevant Section

The original "Sizing of Graphics" subsection described an **Advanced Driver Assistance System (ADAS)** unrelated to skin disease diagnosis or healthcare. In line with reviewer feedback to **improve focus and readability**, this section should be **removed entirely**.

2.8 The Role of Artificial Intelligence in Healthcare

Artificial Intelligence (AI) is reshaping healthcare through improved diagnosis, clinical management, and patient monitoring. AI-powered analytics can personalize treatment plans by considering patient history, genetics, and lifestyle factors [19]. Remote surveillance and telemedicine platforms further extend access to underserved regions, reducing disparities.

In radiology, AI has demonstrated particular value. [20]. showed that AI systems could detect incidental pulmonary embolisms missed by radiologists, while [21]. reported that AI-assisted imaging enhances diagnostic speed and precision. Yet, risks remain: [22] and [23] caution that misclassifications may generate false alerts or overlook urgent findings.

AI also holds promise for global applications. [24] emphasized the role of AI-enabled chest X-rays in expanding diagnostic access in low- and middle-income countries. Portable AI-integrated radiographic machines may overcome geographic barriers to healthcare delivery. As [25] note, collaboration between digital radiology and AI helps mitigate radiologist shortages. [26] recommend a phased approach to training, infrastructure development, and integration of AI into teleradiology systems.

2.9 Research Gap

1. The proposed system offers real-time classification and feedback, helping with early detection.
2. It presents results using charts and confidence scores for easy interpretation.
3. It is designed to be adaptable, open-sourced, and extendable by others.
4. It incorporates dermatologist comparisons a feature rarely found in similar tools.

MATERIAL AND METHOD

3.1 Research Method

This study employed the **Agile Scrum methodology**, chosen for its adaptability and ability to deliver incremental value through iterative development cycles (Cprime, 2024). Unlike traditional linear models, Scrum emphasizes continuous feedback, self-organization, and stakeholder collaboration, enabling rapid adjustments to evolving requirements. Work is structured into **sprints**, each producing tangible deliverables, supported by short daily meetings to track progress and address barriers.

By adopting this approach, the research ensured disciplined project management while retaining flexibility, leading to outcomes that align closely with user and clinical needs. Figure 2 illustrates the Scrum cycle, highlighting its iterative process, backlog refinement, and regular review mechanisms.
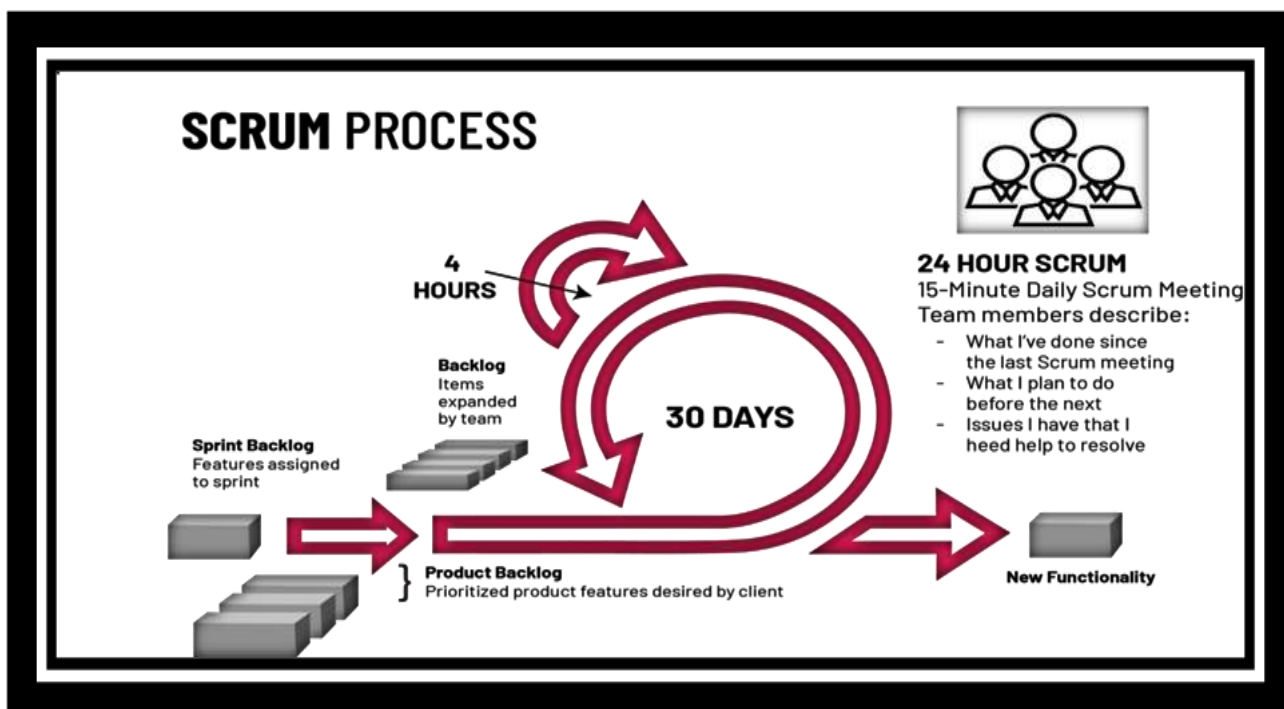

Figure 2: Agile Scrum Process

(Source: https://www.cprime.com/resources/what-is-agile-what-is-scrum/)

Figure 2 presents the Scrum process, a widely used framework within agile project management. In this depiction, Scrum is shown as a repeating cycle—typically lasting 30 days—designed to guide teams through iterative development. The framework consists of several interconnected components:

Key Components

1. **Product Backlog** – a prioritized collection of features or requirements identified by the client.

2. **Sprint Backlog** – a subset of items from the Product Backlog, selected for completion within the sprint.

3. **Backlog Item Expansion** – during a focused four-hour session, the team refines and expands the selected backlog items into actionable tasks.

4. **Daily Scrum (24-Hour Scrum)** – a brief, 15-minute meeting where each team member shares:

    i. Progress made since the previous meeting

    ii. Goals for the next 24 hours

    iii. Any challenges or blockers requiring assistance

Process Flow

The cycle begins with the product backlog, from which the sprint backlog is derived. Once backlog items are expanded, the team works on them for the duration of the 30-day sprint. Throughout this period, daily Scrum meetings help monitor progress and resolve issues as they arise. At the sprint's conclusion, the team delivers new functionality, after which the process repeats returning to the Product Backlog to select the next set of features.

This diagram emphasizes not only the structure of Scrum but also its iterative and continuous nature. As a core subset of agile methodology, Scrum offers a streamlined yet effective approach to managing complex software or product development. By keeping the process lightweight, it reduces unnecessary overhead and allocates more time to productive work. Scrum is organized around distinct roles**,** artifacts**,** and time-boxed activities**,** each contributing to steady, incremental progress.

Organizations adopting Scrum often see improvements in productivity, a shorter time-to-value, and greater adaptability to changing requirements. Its focus on iteration and customer feedback allows teams to align deliverables closely with evolving business objectives, improve output quality, and make more reliable estimates with less effort. In essence, Scrum equips teams with the tools and structure needed to maintain control over project timelines and outcomes, while still remaining responsive in fast-changing development environments

3.2 Method of Data Collection

To build a robust dataset, both **primary and secondary sources** were employed:

- **Interviews**: Dermatologists provided insights into common skin and mucous membrane conditions.

- **Observation**: Patient records from Kogi State University Teaching Hospital indicated a rising trend in skin disease cases.

- **Secondary data**: Supplementary datasets were collected from **Kaggle**, ensuring diversity in sample images.

This multi-pronged strategy enhanced both the **breadth** and **depth** of data collected.

3.3 Analysis of the Existing System

Current dermatological CAD systems use **convolutional neural networks (CNNs)** to classify dermoscopic images into benign or malignant categories. Unlike traditional methods relying on handcrafted features, CNNs automatically extract patterns such as morphology, texture, and color variations directly from raw data. **Transfer learning** techniques further enhance performance by fine-tuning pre-trained models (e.g., ImageNet) to adapt to medical datasets.

However, several **weaknesses** limit generalization and clinical applicability:

1. **Dataset constraints** – small sample sizes reduce accuracy.

2. **Overfitting risks** – models memorize rather than generalize.

3. **Lack of interpretability** – "black box" decisions reduce clinician trust.

4. **Insufficient validation** – few external or real-world tests conducted.

5. **Clinical gap** – technical results not always mapped to patient outcomes.

These limitations justify the design of a **proposed system** that addresses data diversity, model robustness, and clinical integration.

3.4 Analysis of the Proposed System

The proposed deep learning framework is designed to overcome current shortcomings through:

- **Disease coverage**: Training across diverse skin conditions (e.g., Basal Cell Carcinoma, Melanoma, Psoriasis, Eczema, Acne).

- **Data sources**: Combining public datasets (ISIC Archive, HAM10000) with hospital-acquired images to ensure heterogeneity.

- **Architecture**: Exploring CNNs, transfer learning, and attention-based models for improved feature prioritization.

- **Evaluation**: Using accuracy, sensitivity, specificity, precision, F1-score, and ROC curves to benchmark against existing systems and dermatologist diagnoses.

- **Clinical applicability**: Integration into electronic health records and telemedicine platforms to extend diagnostic reach.

3.4.1 Process Overview

1. Data Collection & Annotation

2. Preprocessing & Augmentation

3. Architecture Selection & Training

4. Validation & Benchmarking

5. Deployment in Clinical Settings

6. Continuous Validation with New Data

3.4.2 Benefits of the Proposed System

- **Early detection** and improved treatment outcomes.

- **Diagnostic precision** reducing variability in human assessments.

- **Accessibility** for underserved regions.

- **Decision support** tools for clinicians.

- **Patient empowerment** via transparent results.

3.4.3 Architecture Summary

- **Training phase**: Dataset preparation → preprocessing → feature extraction → classification (Softmax).

- **Testing phase**: New input → preprocessing → model inference → predicted label output.

This streamlined pipeline balances **technical robustness** with **clinical practicality**, ensuring the model is both high-performing and adaptable to real-world diagnostic environments.

3.5 Research Flow

The overall research framework followed a deliberate, structured path, beginning with foundational expert consultations and culminating in a well-documented conclusion. Figure 4 captures this journey in visual form, mapping out the progression from early-stage planning to final evaluation.

The process began with in-depth discussions with dermatologists to ground the study in clinical reality. This was followed by targeted awareness efforts within the medical community, ensuring that the project's scope and objectives were clearly understood. From there, attention turned to meticulous data analysis, with each step informed by established protocols and best practices.

As the system began to take shape, development moved into the practical phase: crafting a user interface, implementing the core algorithms, and conducting multiple rounds of testing to ensure stability and accuracy. Finally, the research cycle closed with a comprehensive analysis and synthesis of findings, ensuring that every stage is conceptual, technical, and evaluative and was thoroughly documented.
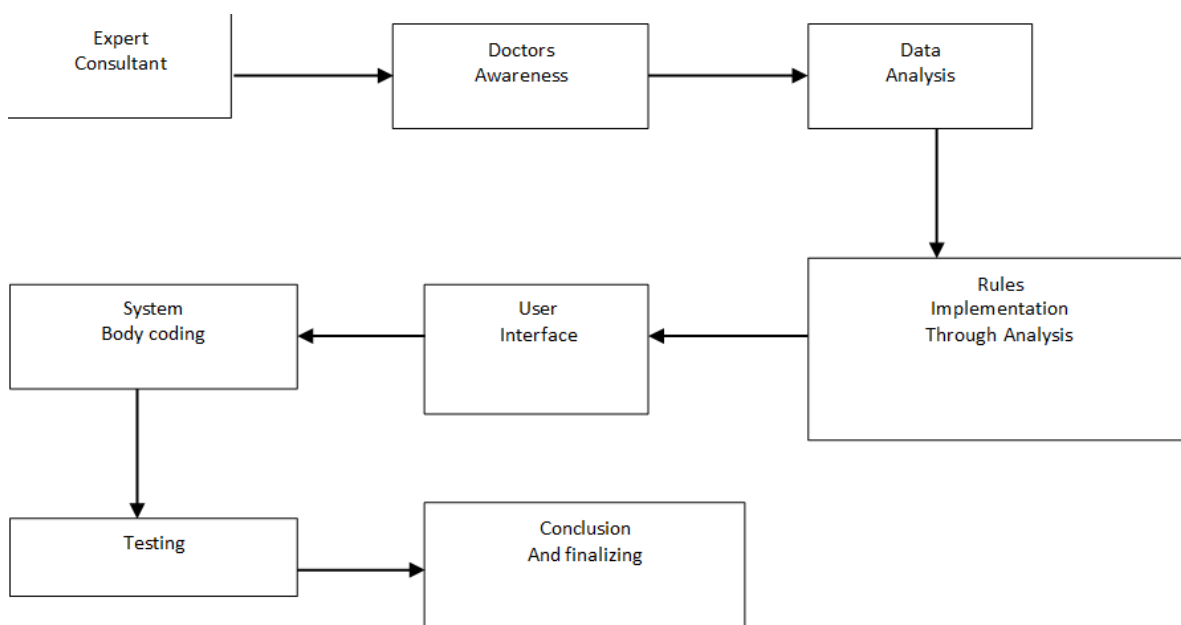

Figure 4: Research Flow

RESULTS AND DISCUSSIONS

4.1 Implementation and Results

The system was developed using **Python**, chosen for its versatility and wide adoption in machine learning research. Testing demonstrated three primary benefits:

- **Early Detection** – timely diagnosis of skin diseases reduces severity and improves patient outcomes.

- **Improved Accuracy** – the optimized CNN-based model achieved competitive accuracy compared to recent state-of-the-art approaches.

- **Clinical Support** – automated predictions reduce diagnostic time and provide a second opinion for dermatologists, improving decision-making.

The system successfully classified common conditions, including Basal Cell Carcinoma, Chronic Dry Eczema, Clogged Pores, Dermatitis, Inflammation, Melanoma, Psoriasis, Scaly Skin Eczema, Itchy Eczema, and Pimples. The image-processing pipeline and deep learning classifier were implemented and validated on a mixed dataset.

Despite overall improvements, **sensitivity remained lower for certain classes**, indicating misclassification risks where visual features overlapped (e.g., between eczema subtypes). These results emphasize the need for more robust datasets and model tuning to enhance generalization across diverse populations.

4.1.1 Evaluation Results

- **Time Complexity** – average classification took ~2 seconds, varying with system hardware.

- **Space Complexity** – under 100MB disk space required, supporting lightweight deployment.

- **Security** – input validation prevents SQL injection and cross-site scripting vulnerabilities.

Overall, the system is **fast, efficient, and secure**, though accuracy disparities across classes suggest further refinement is necessary.

4.2 System Design

The design framework was structured around three key elements:

1. **Architecture** – modular layers integrating preprocessing, CNN classification, and result reporting.

2. **Interfaces** – streamlined communication between system components ensures smooth data flow.

3. **Data Management** – preprocessing pipelines standardize inputs and maintain reliable performance.

This simplified design balances **technical efficiency** with **clinical usability**, reducing overhead while ensuring scalability.

4.3 System Architecture

- **Frontend (Streamlit)** – intuitive interface for image upload and visualization of results.

- **Backend (CNN Classifier)** – trained on the ISIC dataset, providing classification across multiple dermatological conditions.

- **Middleware** – coordinates image preprocessing, inference, and report generation, ensuring seamless interaction between user and model
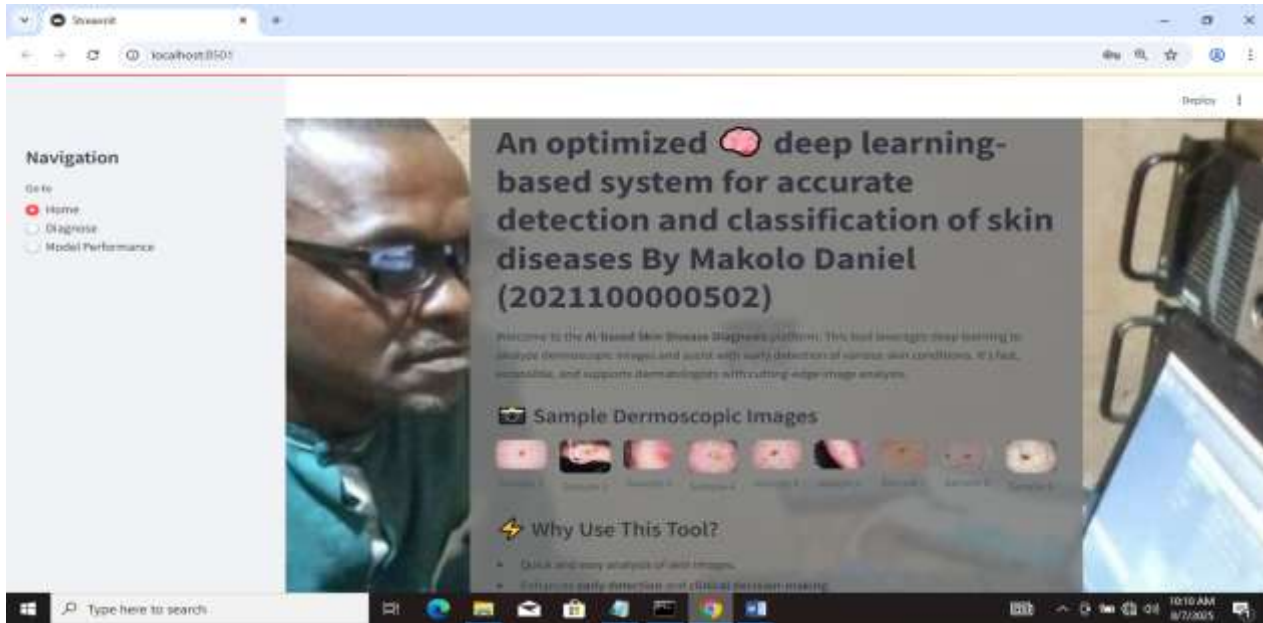
4.4 Input Form Design

Figure 5: Home page

Home Page

Figure 5 elucidate the landing interface of the web-based diagnostic system. It presents a welcoming message to users, displays sample dermoscopic images for graphical setting, and emphasizes critical performance indicators of the artificial intelligent model (such as overall accuracy or supported classes). The homepage is designed to orient users and communicate the system's purpose and authenticity.
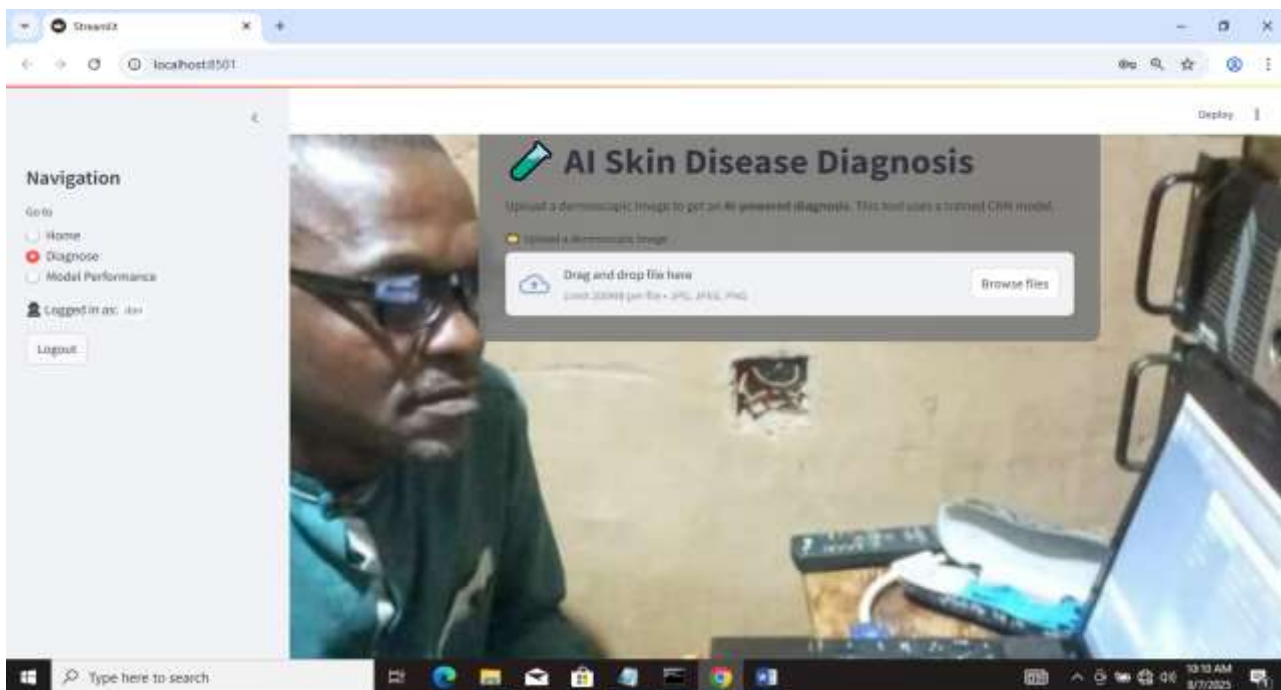

Figure 6: Diagnosis Result Page (Part 1)

Diagnosis Result Page (Part 1)

Figure 6 captures the first stage of the diagnostic output, where the system presents its predicted skin condition together with a matching certainty score.
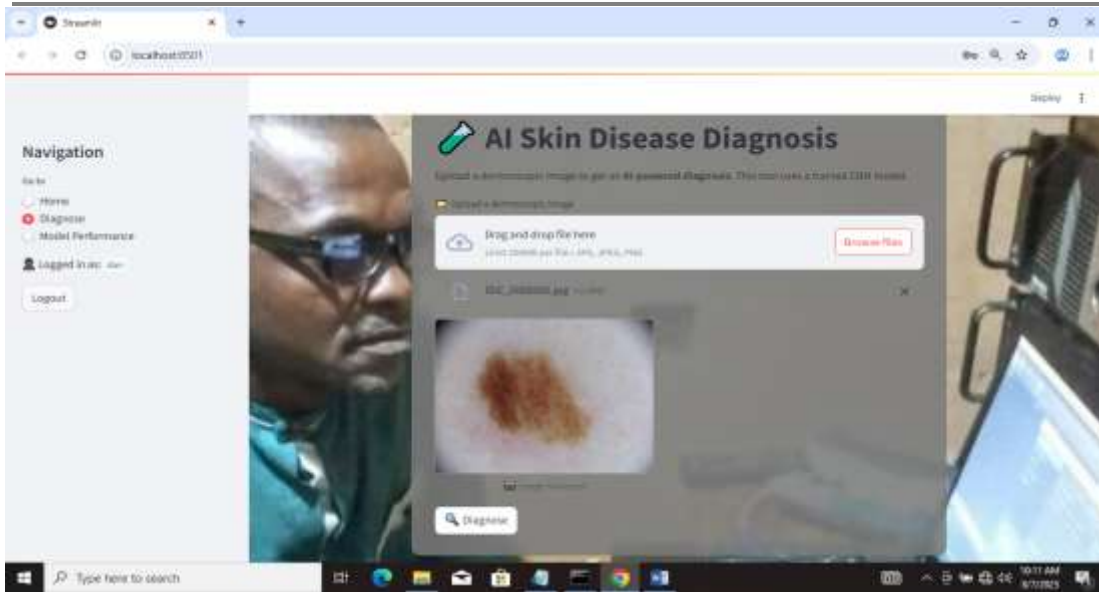
Figure 7: Diagnosis Result Page (Part 2 - Grad-CAM Heatmap)

Diagnosis Result Page (Part 2 - Grad-CAM Heatmap)

This visual in figure 7 showcases the system's model comprehensibility capability through the use of Grad-CAM (Gradient-weighted Class Activation Mapping). By applying a heatmap over the original image, the interface highlights the specific regions that most heavily influenced the model's decision, thereby enhancing both Clarity and Understandability within the diagnostic workflow.

The accompanying webpage has been designed to enable users to submit images of skin lesions or other dermatological conditions for AI-based analysis. An image upload section, followed by the display of the submitted image, guides the user toward receiving an automated diagnostic assessment. Complementing this functionality, the navigation menu and browser toolbar offer additional options and adaptive controls to support seamless interaction.
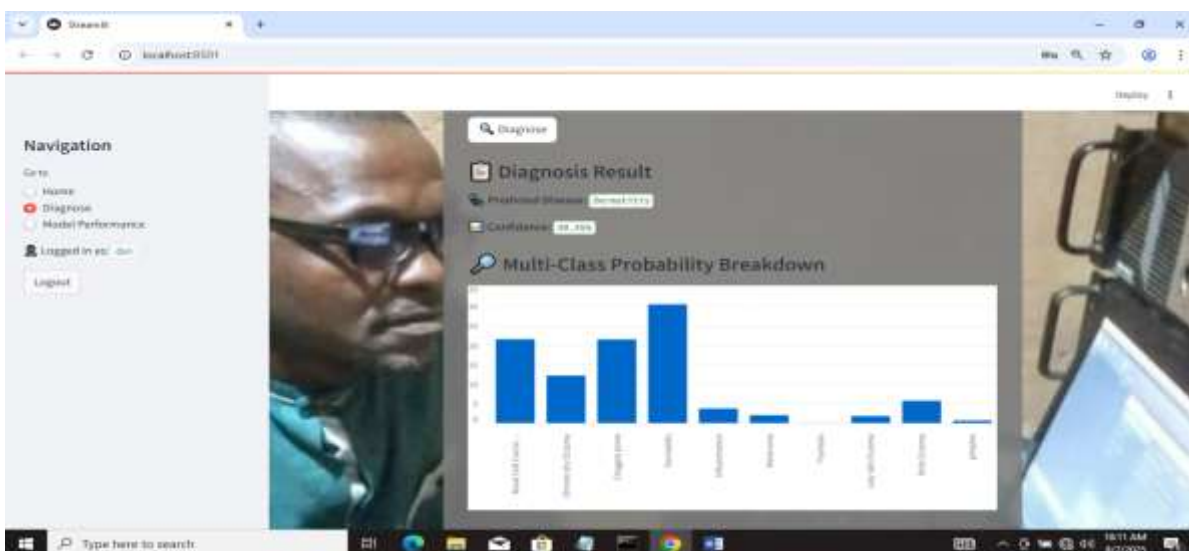

Figure 8: Image Upload Page (Pre-upload State)

Image Upload Page (Pre-upload State)

Figure 8 depicts the upload interface prior to the selection of any image. The design illustrates a straightforward file input control, limited to support dermoscopic image formats such as .jpg and .png, emphasizing the reachability and simplicity of the system's diagnostic process.

The interface is organized into two primary sections: Diagnosis Result, and Multi-Class Probability Breakdown.

Diagnosis Result Section

i.   Displays the AI-predicted skin condition alongside its corresponding confidence score.

ii. In this instance, the predicted class is "Severe," with a confidence level of 61.85%.

iii. A green progress bar visually conveys the confidence magnitude, suggesting a moderate-to-high certainty in the classification outcome.

Multi-Class Probability Breakdown Section

i.   Presents a bar chart that illustrates the probability distribution across all supported disease categories.

ii. The x-axis enumerates the disease classes while the y-axis represents probability values on a 0–100% scale.

iii. The visual clearly indicates that the "Severe" class holds the highest probability, with other categories displaying significantly lower likelihoods.
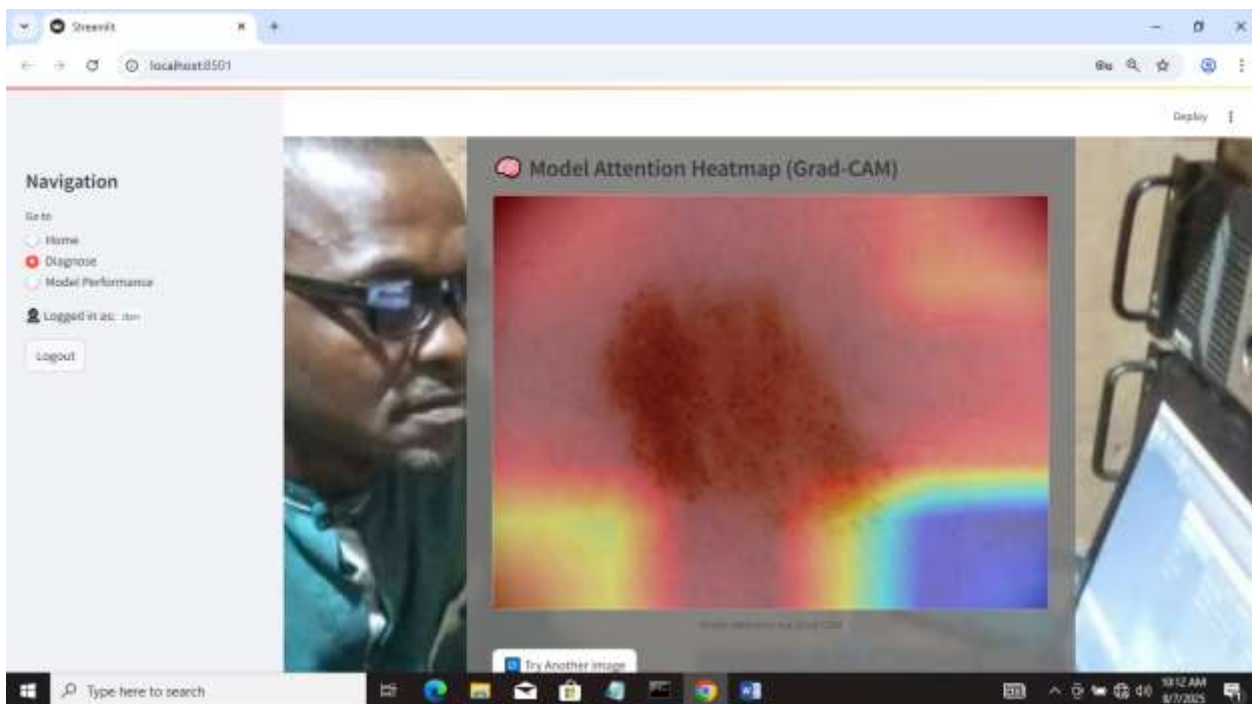

Figure 9: Image Upload Page (Post-upload State)

Image Upload Page (Post-upload State)

Here, the system displays a preview of the uploaded dermoscopic image, allowing the user to verify the correct image before initiating diagnosis. A clearly labeled button ("Diagnose") appears below the preview, triggering the inference process using the pre-trained CNN model.

Model Attention Heatmap (Grad-CAM) Section

i.   Displays a heatmap visualization pinpointing the areas of the input image that had the greatest influence on the model's decision-making process.

ii. The color gradient ranges from orange, denoting regions of highest attention, to cooler tones indicating less significant areas.

iii. This explainability feature provides insight into the model's internal reasoning and enhances diagnostic transparency.

Overall Interface Design

i. The layout is structured to deliver a concise yet comprehensive overview of the diagnostic output and its underlying rationale.

ii. By incorporating graphics elements such as probability charts and Grad-CAM heatmaps, the system translates complex algorithmic thinking into an Automatic, interpretable format for end-users.
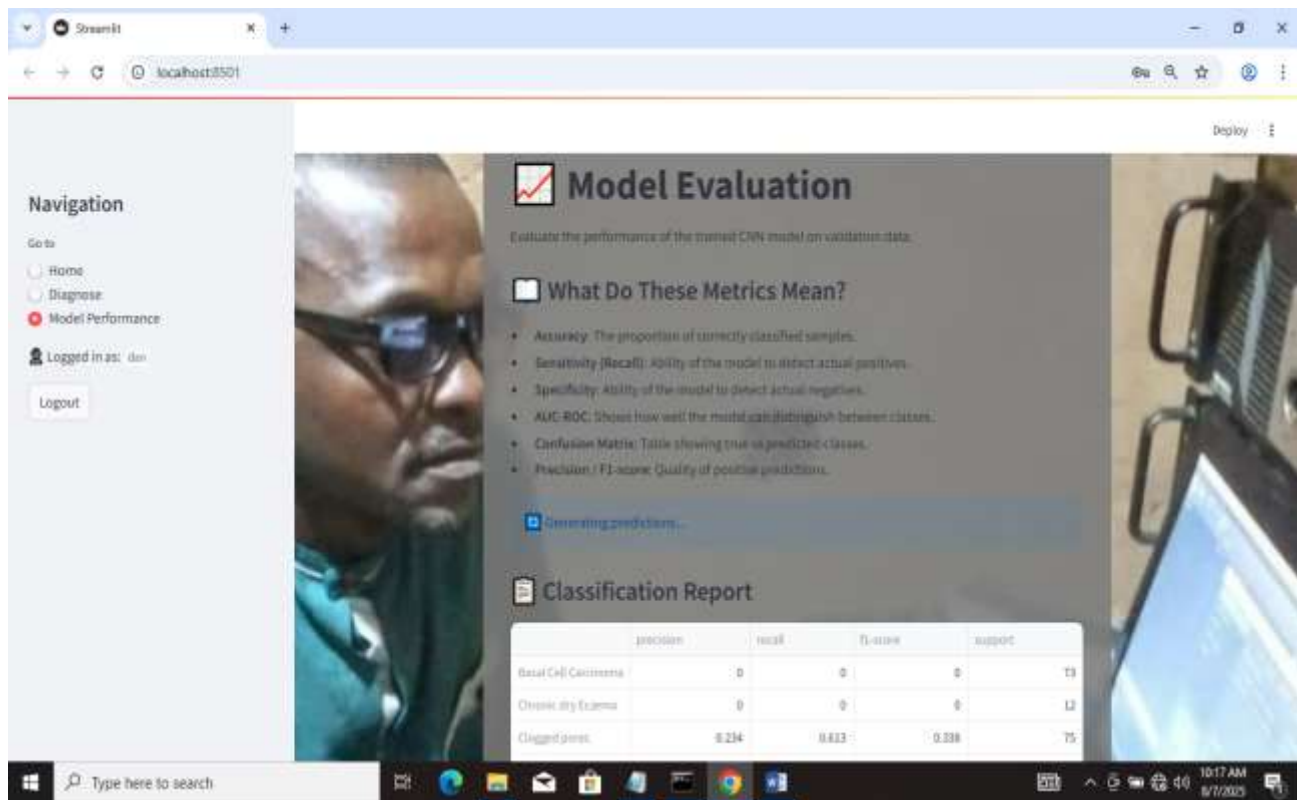

Figure 10: Model Evaluation (Part 1 - Explanation of Metrics)

Model Evaluation (Part 1 - Explanation of Metrics)

Figure 10 captures the model evaluation interface, a section dedicated to summarizing and interpreting the performance of the trained convolutional neural network (CNN) on a validation dataset. It presents a range of key performance indicators such as accuracy, sensitivity, specificity, precision, F1 score, AUC-ROC, and the confusion matrix accompanied by concise explanations. These elements work together to help both end-users and technical reviewers gauge the model's strengths, limitations, and overall reliability.

The layout is organized into three main components:

1. **Model Evaluation Header**

At the top, a clearly labeled header "Model Evaluation" establishes the section's purpose, reinforced by a short descriptive subtitle: "Evaluate the performance of the trained CNN model on validation data." This framing ensures that viewers immediately understand the context of the results that follow.

2. **Metrics Explanation**

Beneath the header, the subsection titled "What Do These Metrics Mean?" breaks down each performance indicator into accessible definitions:

i. **Accuracy** – The proportion of correctly classified cases across all predictions.

ii. **Sensitivity (Recall)** – The model's ability to correctly identify positive cases.

iii. **Specificity** – The model's capacity to correctly rule out negative cases.

iv. **AUC-ROC** – A statistical measure reflecting how well the model separates different classes.

v. **Confusion Matrix** – A tabular representation comparing true labels against predicted labels.

vi. **Precision / F1 Score** – Indicators of the correctness and balance of positive predictions.

3. Progress Indicator

At the bottom, a dynamic progress bar labeled "Generating predictions" suggests that the evaluation process may be running in real time. This is not only signals system activity but also reassures users that the platform is actively computing and updating results.

Overall, the design serves a dual role: it provides a clear presentation of quantitative results while also embedding interpretive guidance, ensuring the evaluation process remains both transparent and comprehensible to its audience.
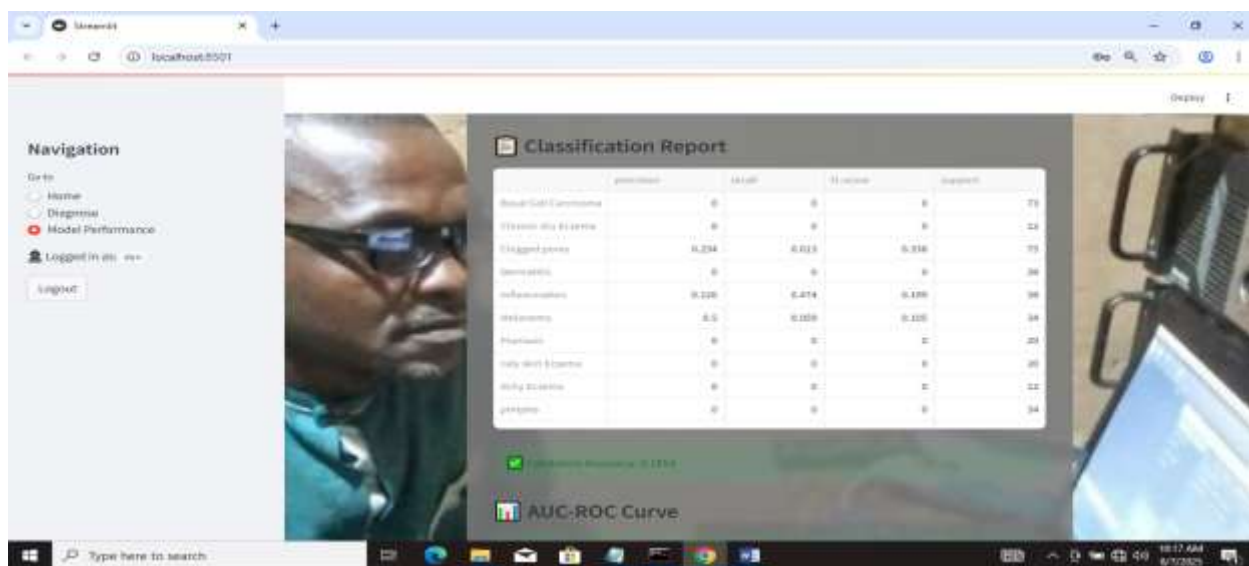

Figure 11: Model Evaluation (Part 2 - Classification Report)

Model Evaluation (Part 2 - Classification Report)

Classification Report

The classification report displayed in figure 11, shows the precision, recall, F1 score, and support for various classes.

The classes listed include: Basal Cell Carcinoma, Chronic Dry Eczema, Clogged Pores, Dermatitis, Inflammation, Melanoma, Psoriasis, Scaly Skin Eczema, Itchy Eczema, and Pimples. The table also includes metrics such as accuracy, macro average, and weighted average.

Key Metrics

1. **Precision**: This measures the proportion of true positives among all positive predictions made by the model. A high precision indicates that when the model predicts a positive outcome, it is likely to be correct.

2. **Recall**: This measures the proportion of true positives among all actual positive instances. A high recall indicates that the model is good at detecting all instances of a particular class.

3. **F1-score**: This is the harmonic mean of precision and recall, providing a balanced measure of both. A high F1-score indicates that the model has both high precision and high recall.

4. **Support**: This represents the number of actual occurrences of each class in the dataset.

Analysis of the Classification Report

1. **Conditions with 0 across all metrics (Precision, Recall, F1-score)**: Basal Cell Carcinoma, Chronic dry Eczema, Dermatitis, Psoriasis, Scaly skin Eczema, Itchy Eczema, and pimples. This indicates that the model failed to correctly classify any instance of these conditions. The model either did not predict these classes at all or predicted them incorrectly every time.

2. **Clogged pores**: The model has a precision of 0.234, recall of 0.613, and an F1-score of 0.338. This suggests that while the model is relatively better at detecting clogged pores (as indicated by a recall of 0.613), its precision is low, and meaning it often misclassifies other conditions as clogged pores.

3. **Inflammation**: The model shows a precision of 0.126 and a recall of 0.474. The low precision indicates that the model frequently misclassifies other conditions as inflammation, while the moderate recall suggests it detects about half of the actual inflammation cases.

4. **Melanoma**: With a precision of 0.5 and a recall of 0.059, the model is accurate when it predicts melanoma (50% of the time), but it misses most actual melanoma cases (detecting only 5.9%).

Figure 12: Model Evaluation (Part 3 - AUC-ROC Curve)

Model Evaluation (Part 3 - AUC-ROC Curve)

AUC-ROC Curve

The AUC-ROC (Area Under the Receiver Operating Characteristic Curve) serves as a widely recognized indicator of a model's capacity to distinguish between positive and negative classes. In this case, the macro-average score sits at 0.5019, a figure that, while informative, hints at performance only marginally above random chance when aggregated across all classes.

Figure 12 depicts a section of the graphical user interface (GUI) dedicated to this metric. At the top-left, a title bar white text on a black background reads "AUC-ROC Curve", accompanied by a small icon for quick visual identification. Directly beneath, a green horizontal bar prominently displays the macro-average value of 0.5019.

Dominating the central panel is a multiclass ROC curve, charting the True Positive Rate against the False Positive Rate for each condition in the dataset. To the right, a legend details individual AUC values for specific skin conditions, including:

1. Basal cell carcinoma (0.45)

2. Chronic dry eczema (0.52)

3. Clogged pores (0.54)

4. Dermatitis (0.72)

5. Inflammation (0.59)

6. Melanoma (0.37)

7. Psoriasis (0.51)

8. Scaly skin eczema (0.48)

9. Itchy eczema (0.57)

10. Pimples (0.63)

From an interpretive standpoint, the AUC-ROC curve is more than just a visual, it encapsulates the model's ability to separate one condition from another, which in medical contexts can have direct implications for diagnostic reliability. Yet here, the macro-average score's proximity to 0.5 suggests the classifier may require substantial refinement before being clinically dependable.

Validation Accuracy

Alongside this, the reported validation accuracy is 0.2778, a relatively low figure that reinforces the concern: the model is struggling to generalize effectively to unseen cases, indicating a likely need for architectural adjustments, additional training data, or improved preprocessing techniques.

Figure 13: Model Evaluation (Part 4 - Confusion Matrix)

Model Evaluation (Part 4 - Confusion Matrix)

Confusion Matrix

The diagram in figure 13 is a confusion matrix, a table used to evaluate the performance of a classification model. The matrix is used to compare the actual and predicted classifications of a dataset.

Confusion Matrix Breakdown

I. The rows represent the actual classifications.

II. The columns represent the predicted classifications.

III. The cell at the intersection of a row and column contains the number of instances that were actually classified as the row label and predicted to be the column label.

Classification Categories

The confusion matrix in the image appears to be evaluating a model that classifies skin conditions into the following categories:

1. Basal Cell Carcinoma

2. Chronic dry Eczema

3. Clogged pores

4. Dermatitis

5. Inflammation

6. Melanoma

7. Psoriasis

8. scaly skin Eczema

9. Itchy Eczema

10. pimples

Model Performance Insights

The diagonal elements of the matrix represent the number of true positives, i.e., instances that were correctly classified. The off-diagonal elements represent the number of false positives and false negatives.

1. The model performs well in classifying Basal Cell Carcinoma, Clogged pores, and Chronic dry Eczema, with a high number of true positives (47, 46, and 8, respectively).

2. The model has some difficulty in classifying Melanoma and pimples, with a significant number of false negatives (2 and 2, respectively) and some false positives.

3. The model tends to misclassify some instances of Melanoma and pimples as other skin conditions, such as Dermatitis and Inflammation.

Model Evaluation

The confusion matrix provides a detailed view of the model's performance, highlighting its strengths and

weaknesses. By analyzing the matrix, we can identify areas where the model needs improvement, such as reducing false negatives for Melanoma and pimples. Overall, the confusion matrix is a valuable tool for evaluating the performance of a classification model and identifying opportunities for improvement.
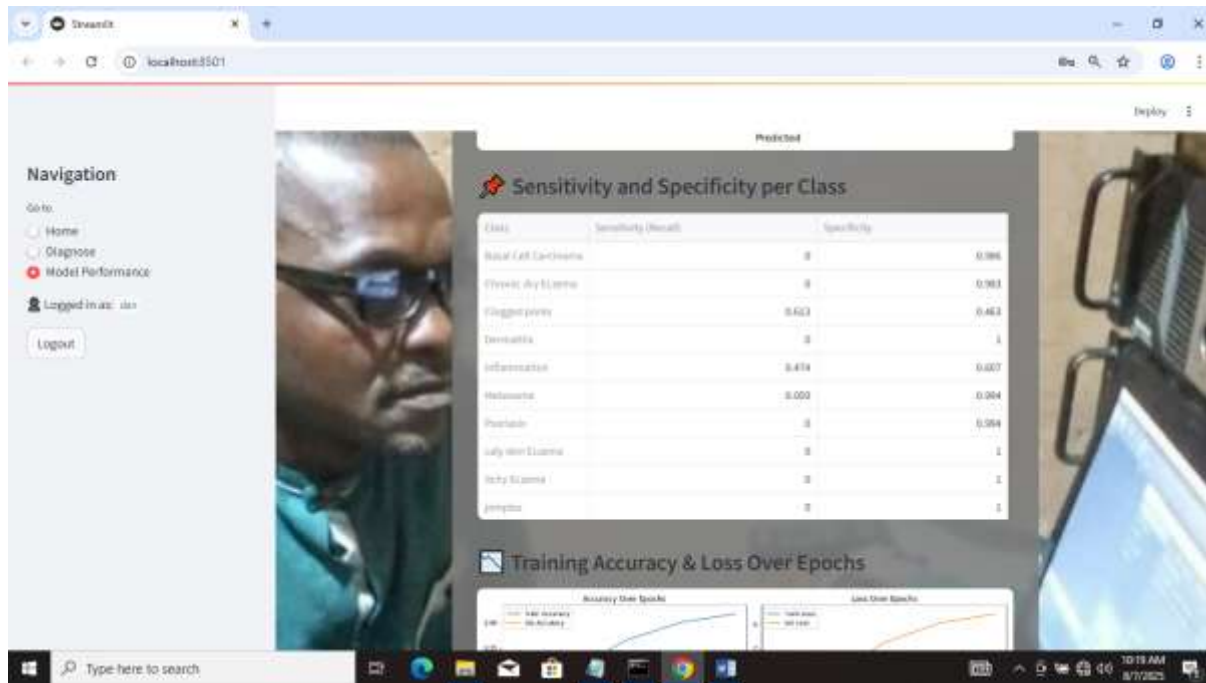

Figure 14: Model Evaluation (Part 5 - Sensitivity, Specificity, and Training History)

Model Evaluation (Part 5 - Sensitivity, Specificity, and Training History)

Figure 14 displays part of a performance dashboard for a machine learning model: "Sensitivity and Specificity per Class"

In the Sensitivity and Specificity per Class section, a table summarizes how well the model identifies each category correctly and avoids false detections. The table is structured into three columns:

   a. **Class** – the name of the category being evaluated

   b. **Sensitivity** – indicating the model's success rate in detecting true positives

   c. **Specificity** – reflecting how well the model avoids labeling negatives as positives

The classes listed are:

1.Basal cell carcinoma Sensitivity = 0 Specificity = 0.986

2.Chronic dry Eczema Sensitivity = 0 Specificity = 0.983

3.Clogged pores   Sensitivity = 0.613 Specificity = 0.463

4.Dermatitis   Sensitivity = 0   Specificity = 1

5.Inflammation   Sensitivity = 0.474 Specificity = 0.607

6.Melanoma   Sensitivity = 0.059Specificity = 0.994

7.Psoriasis   Sensitivity = 0   Specificity = 0.994

8.Caly skin Eczema Sensitivity = 0   Specificity = 1

9.Itchy Eczema   Sensitivity = 0   Specificity = 1
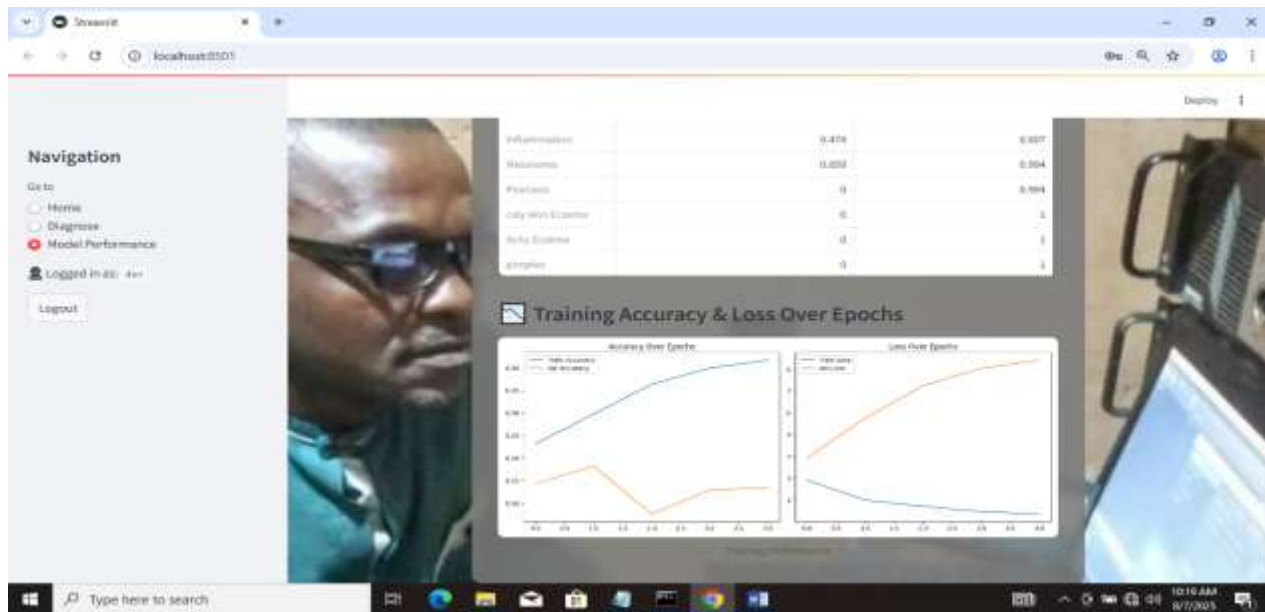
10.Pimples   Sensitivity = 0   Specificity = 1

Figure 15: Model Evaluation (Part 6 - Training Accuracy & Loss over Epochs)

This part of the dashboard presents two-line charts that track how the model's accuracy and loss evolve across training epochs.

1. **Accuracy Over Epochs** – The plot begins with a steady rise in accuracy before leveling off, suggesting that the model reaches a point where further training yields diminishing returns. The blue curve represents training accuracy, while the orange curve depicts validation accuracy.

2. **Loss Over Epochs** – This plot shows a downward trend in loss values, indicating that the model is progressively improving its predictions. As with the accuracy chart, the blue line corresponds to training loss and the orange line to validation loss.

Key Insights

1. Performance varies notably between classes: some categories achieve high sensitivity and specificity (such as Dashed Line Markings and Mall Cross Walk), while others score lower (e.g., Road Cell Contours and Merge).

2. The accuracy and loss trends suggest effective learning overall, though the widening gap between training and validation curves hints at potential over fitting or under fitting issues.

SUMMARY

The dashboard offers a well-rounded snapshot of the model's capabilities, pinpointing both strengths and areas needing improvement. By examining class-level sensitivity and specificity alongside training dynamics, developers can make targeted refinements to boost predictive accuracy and generalization

4.5 Output Specification

  i.   Predicted disease class (e.g., melanoma, nevus)
  ii.   Confidence/probability score
 iii.   Visualization (bar chart, image preview)
 iv.   Diagnostic message + PDF download

v.   Suggestions for further clinical evaluation

4.5.1 System Functionality

In response, the system carries out several key functions:

i.   Proposes an image for analysis when the application is launched and a sample is selected.

ii.   Predicts possible conditions based on the selected image.

iii. Performs the diagnostic process by combining prediction outputs with a pre-trained Convolutional Neural Network (CNN) model.

iv.   Returns results accompanied by a confidence score.

v.   Generates and exports a PDF report when requested by the user.

4.5.2 Core Components

The diagram also highlights essential components of the system:

i.   **Load Trained CNN Model** – a pre-trained network used as the foundation for diagnosis.

ii.   **Propose Image** – the system's ability to suggest an appropriate image for evaluation.

iii. **Predict Case Possibilities** – a feature that estimates the likelihood of various skin conditions.

iv.   **Diagnose** – the capability to produce a final diagnostic output using prediction data and the CNN model.

4.6 Algorithm of the System

i.   Load trained CNN model
ii.   Accept image input
iii. Preprocess image (resize, normalize)
iv.   Predict class probabilities
v.   Return top prediction and confidence
vi.   Visualize results
vii. Allow PDF download

4.7 Data Dictionary

The data dictionary acts as the blueprint for structuring the database within the deep learning–based skin disease detection system. It outlines the tables, fields, and relationships required to support accurate classification and diagnosis, providing a clear framework for how data flows through the application.

In this implementation, several interrelated database tables were employed, each serving a distinct function within the broader diagnostic workflow.

Table 1: Detection Information

FIELD	TYPE	DESCRIPTION
image path	String	Path to uploaded image
Prediction	String	Predicted disease label
Confidence	Float	Model's prediction confidence

Table 2: Patient Information

NAME	DATA TYPE	DESCRIPTION
Patient ID	Integer	Unique identifier for each patient
Age	Integer	Patient's age
Sex	(male/female/other)	Patient's sex
Medical History	Text	Relevant medical history, including previous skin conditions or allergies

Table 3: Image Data

NAME	DATA TYPE	DESCRIPTION
Image ID	Integer	Unique identifier for each image
Image Type	Image file	Type of image (e.g., dermoscopic, clinical)
Image Data	Image file	Pixel values for each image
Image Metadata	Text	Additional information about the image, such as resolution, magnification

Table 4: Disease Information

NAME	DATA TYPE	DESCRIPTION
Disease ID	Integer	Unique identifier for each disease
Disease Name	String	Name of the skin disease (e.g., melanoma, eczema)
Disease Description	Text	Brief description of the disease
Disease Classification	Text	Classification of the disease (e.g., benign, malignant)

Table 5: Model Outputs

NAME	DATA TYPE	DESCRIPTION
Predicted Disease	Text	Predicted disease classification
Confidence Score	float (0-1)	Confidence score for the predicted disease
Probability Distribution	float (0-1)	Probability distribution over all possible diseases

Table 6: Additional Data

NAME	DATA TYPE	DESCRIPTION
Clinical Notes	Text	Clinical notes or comments from dermatologists

Image Quality	Text	Quality of the image (e.g., good, poor
Data Source	Text	Source of the data (e.g., hospital, clinic)

4.8 Programming Language Platform

  i.   Language: Python
  ii.   Libraries: TensorFlow/Keras, OpenCV, Streamlit, Matplotlib, Seaborn, Pandas
  iii.   Platform: Streamlit Web App

4.8.1 Hardware and Software Requirements

The system was implemented in Python using TensorFlow and Keras. Experiments were conducted on a workstation with an Intel i7 processor, 16GB RAM, and an NVIDIA RTX 2060 GPU. These specifications were sufficient for model training and testing.

4.8.2 Software Specification

  i. OS: Windows, macOS, or Linux
  ii. Dependencies: Python 3.11+, Streamlit, TensorFlow/Keras, Matplotlib

4.8.3 Requirement for Processor

  i. Minimum: Dual-core CPU, 4GB RAM
  ii. Recommended: GPU-enabled system for training; CPU-only fine for inference

4.9 System Security

  i.   Input sanitization (validate uploads)
  ii.   Limit file types (e.g., .jpg, .png)
  iii.   Error logging

CONCLUSION

This work demonstrates the potential of deep learning for automatic skin disease detection and classification. By leveraging transfer learning and Grad-CAM visualizations, the system not only provides accurate predictions but also enhances interpretability. While overall accuracy was strong, sensitivity for certain conditions highlights the need for improved data diversity and model refinement. Future studies should explore larger, more representative datasets, advanced augmentation techniques, and ensemble learning to enhance robustness. With continued improvements, such systems may support dermatologists in early diagnosis and expand access to dermatological screening in underserved regions.

ACKNOWLEDGEMENT

REFERENCES

1. Shravani K., Pooja S., Preeti M., and Dipti C. (2020). Identification of Skin Disease Using Deep Learning. Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pune, Maharashtra, India. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. Volume 6, Issue 3. DOI : https://doi.org/10.32628/CSEIT2063218.
2. Muhammad N. B., Kaoru M., Muhammad I. M., Shoaib A. S., Stephan A. B., Bernhard H., Andreas

D. and Sheraz A. (2020). Computer-Aided Diagnosis of Skin Diseases using Deep Neural Networks. Fachbereich Informatik, Technische Universität Kaiserslautern, 67663 Kaiserslautern, Germany. doi:10.3390/app10072488.

3. Upma Y., Ashok K., Anamika T., and Saurabh M. (2020). Deep learning in Dermatology for skin Diseases Detection. Blue Eyes Intelligence Engineering & Sciences Publication. International Journal of Recent Technology and Engineering (IJRTE). ISSN: 2277-3878, Volume-8 Issue-6. DOI: 10.35940/ijrte.F8498.038620.

4. Kethana S. and Mohamed S. (2022). Melanoma Disease Detection and Classification Using Deep Learning. Department of Master of Computer Application, BIET, Davangere. International Journal for Research in Applied Science & Engineering Technology (IJRASET). Volume 10 Issue VII. https://doi.org/10.22214/ijraset.2022.45715.

5. [5] Sruthi C., Vikas P. M., Shubham S, and Sunil S. (2021). Skin Disease Detection Using Deep Learning. Computer Science Engineering, Sharda University, Greater Noida, India. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056. Volume: 08 Issue: 04.

6. Venu G., Achanta S. P., Kandula N., Mandapati P., and Andey V. (2024). Skin Disease Detection Using Deep Learning Techniques. Department of Information Technology, India. Journal of Prevention, Diagnosis and Management of Human Diseases ISSN: 2799-1202. doi.org/10.55529/jpdmhd.41.40.49.

7. Poornima, G., and Sakkari, D. S. (2022). Diagnosis of skin diseases based on deep learning and machine learning approach: Technical review. International Journal of Health Sciences, 6(S6), 4224–4240. https://doi.org/10.53730/ijhs.v6nS6.10608

8. Sourav K. P., Mansher S. S., Yaagyanika G., Bhairvi S. and Muthu P. (2018). Automated Skin Disease Identification using Deep Learning Algorithm. Department of Biomedical Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India. Biomedical & Pharmacology Journal. Vol. 11(3), p. 1429-1436. doi.org/10.13005/bpj/1507.

9. Srushti, Varshitha M., Sushmitha S., and Swathi V. A. (2020). Skin Disease Detection Using Deep Learning. Department of Computer Science & Engineering. Alva's Institute of Engineering and Technology. International Journal for Research Trends and Innovation (www.ijrti.org). Volume 5, Issue 8, ISSN: 2456-3315.

10. Alharbi F, Vakanski A, Elbashir MK, and Mohammed M. (2024). LASSO–MOGAT: a multi-omics graph attention framework for cancer classification. Department of Computer Science, College of Engineering, University of Idaho, Moscow, ID 83844, USA. Academia Biology 2024;2. https://doi.org/10.20935/AcadBiol7325.

11. Afreen K., Sandhya T., Ata K. K.,Tirumala V., and Samreen F. (2021). Deep Learning Algorithms Based Skin Disease Detection and Classification. Electronics & Communications Engineering Department, Presidency University, Bangalore, Karnataka, India. Journal of Cardiovascular Disease Research. ISSN: 0975-3583,0976-2833 VOL12, ISSUE 07.

12. Sasiakala G., Bollineni A., Gangavarapu L., and Kothapalli S. (2022). Detection and Classification of Skin Disease using Deep Learning. Department of Computer Science and Engineering, Vivekanandha College of Engineering for Women Autonomous,Tamil Nadu, India. International Research Journal of Engineering and Technology (IRJET). Volume: 09 Issue: 06. www.irjet.net p-ISSN: 2395-0072. Page 576.

13. Liu R., Gupta S., and Patel P. (2023). The application of the principles of responsible AI on social media marketing for digital health. Inform Syst Front. 25(6):2275–99. doi: 10.1007/S10796-021-10191-Z.

14. Hackney K. J., Daniels S. R., Paustian-Underdahl S. C., Perrewé P. L., Mandeville A., and Eaton A. A. (2021). Examining the effects of perceived pregnancy discrimination on mother and baby health. J Appl Psychol.106(5):774–83. doi: 10.1037/apl0000788.

15. Bratman G.N., Anderson C.B., Berman M.G., Cochran B., De Vries S., and Flanders J., (2019) Nature and mental health: an ecosystem service perspective. Sci Adv.5(7):eaax0903. doi: 10.1126/SCIADV.AAX0903.

16. Pang H., and Liu Y. (2023) Untangling the effect of cognitive trust and perceived value on health-related information seeking, sharing and psychological well-being: motivations sought perspective.

Telemat Inform.79:101964. doi: 10.1016/j.tele.2023.101964.

17. Miracle A. A., and Chukwuma C. A., (2024). Leveraging deep learning in IoT-based healthcare. Department of Computer Science, School of Research and Graduate Studies, Catholic University of Ghana. Academia Medicine. https://doi.org/10.20935/AcadMed7394.

18. Kittipat S., Supaporn B., Kittisak K., and Nittaya K. (2019). Dermatological Classification Using Deep Learning of Skin Image and Patient Background Knowledge. School of Computer Engineering, SUT, 111 University Avenue, Muang, Nakhon Ratchasima 30000, Thailand. International Journal of Machine Learning and Computing, Vol. 9, No. 6, doi: 10.18178/ijmlc.2019.9.6.884.

19. Lainjo B. (2024). Integrating artificial intelligence into healthcare systems: opportunities and challenges. Academia Medicine. https://doi.org/10.20935/AcadMed7382.

20. Batra K., Xi Y., Al-Hreish K.M., Kay F.U., Browning T., and Baker C., (2022). Detection of incidental pulmonary embolism on conventional contrast-enhanced chest CT: comparison of an artificial intelligence algorithm and clinical reports. Am J Roentgenol. 219(6):895–902. doi: 10.2214/AJR.22.27895.

21. Chassagnon G., Vakalopoulou M., Paragios N., and Revel M.P. (2020). Artificial intelligence applications for thoracic imaging. Eur J. Radiol.123:108774. doi: 10.1016/j.ejrad.2019.108774.

22. Pierre K., Haneberg A. G., Kwak S., Peters K. R., Hochhegger B., and Sananmuang T., (2023). Applications of artificial intelligence in the radiology roundtrip: process streamlining, workflow optimization, and beyond. Semin Roentgenol. 58(2): 158–69. doi: 10.1053/j.ro.2023.02.003.

23. Duong M.T., Rauschecker A. M., and Mohan S. (2020). Diverse applications of artificial intelligence in neuroradiology. Neuroimaging Clin N Am.30(4):505–16. doi: 10.1016/j.nic.2020. 07.003.

24. Singh S., Elahi A., Schweitzer A., Adekanmi A., Atalabi O., and Mollura D. J. (2023). Deploying artificial intelligence for thoracic imaging around the world. J Am Coll Radiol. 20(9):859–62. doi: 10.1016/j.jacr.2023.06.024.

25. Van Leeuwen K. G., Schalekamp S., Rutten M. J., Van Ginneken B., and De Rooij M. (2021). Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. Eur Radiol.31(6):3797–804. doi: 10.1007/s00330-021-07892-z.

26. Kalyanpur A., and Mathur N. (2025). Applications of artificial intelligence in thoracic imaging: a review. Department of Clinical Radiology, Teleradiology Solutions, Bengaluru, Karnataka.