

CLASSIFICATION OF SKIN BURN IMAGES USING MACHINE LEARNING

by

**Anıl Demirok
Eren Çetin
Akın Ateş**

CSE4197 / CSE4198 Engineering Project report submitted to Faculty of Engineering
in partial fulfillment of the requirements for the degree of

BACHELOR OF SCIENCE

Supervised by:

Assoc.Prof.Dr. Mustafa Ağaoğlu

Marmara University, Faculty of Engineering

Computer Engineering Department

2025

Copyright © Group members listed above, 2025. All rights reserved.

CLASSIFICATION OF SKIN BURN IMAGES USING MACHINE LEARNING

by

**Anıl Demirok
Eren Çetin
Akın Ateş**

CSE4197 / CSE4198 Engineering Project report submitted to Faculty of Engineering
in partial fulfillment of the requirements for the degree of

BACHELOR OF SCIENCE

Supervised by:

Assoc.Prof.Dr. Mustafa Ağaoğlu

Sign

Marmara University, Faculty of Engineering

Computer Engineering Department

2025

Copyright © Group members listed above, 2025. All rights reserved.

ABSTRACT

Burn injuries remain a serious global health issue, causing approximately 180,000 deaths annually, with the highest incidence in low- and middle-income countries where access to specialized medical care is limited. Accurate and timely assessment of burn severity is essential for determining the appropriate course of treatment and improving patient outcomes. However, traditional diagnostic methods rely heavily on clinical expertise, which may not always be available or consistent across practitioners. To address this gap, this project presents the development of an automated system that uses machine learning and image processing techniques to detect and classify burn wounds.

The primary objective of the project is to accurately identify the degree of burns—categorized as first, second, or third degree—and estimate the total affected body surface area. Various machine learning models, methods, and algorithms were evaluated to determine the most effective approach, using a curated dataset composed primarily of images from individuals with medium and lighter skin tones. The system processes user-submitted images to detect burned regions, classify the severity, and compute the extent of the injury.

A web-based application was developed to provide a user-friendly interface for healthcare professionals and patients, enabling easy image uploads and real-time diagnostic results. The system demonstrates high accuracy in both burn classification and affected area estimation, enhancing the speed and reliability of clinical decision-making. While the model is currently optimized for specific skin tones due to dataset constraints, it highlights the potential of scalable, AI-driven diagnostic tools in medical practice.

This project contributes meaningfully to the field of intelligent healthcare systems by improving access to care, accelerating diagnostics, and ensuring greater consistency in burn evaluation. Future developments may focus on expanding dataset diversity and integrating the system into mobile platforms to further increase accessibility and usability worldwide.

TABLE OF CONTENTS

ABSTRACT.....	3
TABLE OF CONTENTS.....	4
LIST OF FIGURES.....	6
LIST OF TABLES.....	7
1. INTRODUCTION.....	1
1.1 Problem Description and Motivation.....	1
1.2 Main Goal and Objectives of the Project.....	2
1.2.1 To accurately identify the burn degree using machine learning techniques.....	2
1.2.2 To precisely detect and quantify the burned area.....	2
1.2.3 To develop a web application.....	2
2. DEFINITION OF THE PROJECT.....	3
2.1 Scope of the Project.....	3
2.2 Success Factors.....	5
2.3 Professional Considerations.....	7
2.3.1 Methodological Considerations.....	7
2.3.2 Realistic Constraints.....	8
2.3.3 Legal Considerations.....	9
2.4 Related Work.....	10
2.4.1 Automatic Segmentation and Degree Identification in Burn Color Images.....	10
2.4.2 Segmentation and Classification with Artificial Intelligence.....	10
2.4.3 Deep Learning-Assisted Burn Wound Diagnosis.....	10
2.4.4 Burn Image Segmentation and Depth Diagnosis Framework.....	11
3. SYSTEM DESIGN AND SOFTWARE ARCHITECTURE.....	13
3.1 Project Requirements.....	13
3.1.1 Functional Requirements.....	13
3.1.2 Nonfunctional Requirements.....	13
3.2 System Design.....	15
3.2.1 UML Use case Diagram(s) for the main use cases.....	15
3.2.2 UML Class and/or Database ER diagram(s).....	15
3.2.3 User Interface.....	16
3.2.4 Test Plan.....	18
3.3 Software Architecture.....	19
4. TECHNICAL APPROACH AND IMPLEMENTATION DETAILS.....	20
4.1 Technical Approach and Implementation Details.....	20
4.1.1 Background Removal.....	20
4.1.2 Targeted Preprocessing for Burn Region Isolation.....	22
4.1.3 Burn Classification Using RegNetY-080.....	23
4.2 Alternative Approach (Ensemble Learning).....	25
5. SOFTWARE TESTING.....	29
5.1 Implementation Phase.....	29

5.2 Testing Phase.....	30
6. BENEFITS AND IMPACT.....	32
6.1 Scientific Impact.....	32
6.2 Economic, Commercial and Social Impact.....	32
6.3 Potential Impact on New Projects.....	33
7. CONCLUSION AND FUTURE WORK.....	34
7.1 Conclusion.....	34
7.2 Future Work.....	34
REFERENCES.....	36

LIST OF FIGURES

Figure 3.1: Use Case Diagram.....	15
Figure 3.2: Uml Class Diagram.....	16
Figure 3.3: Information Page.....	17
Figure 3.4: Image Upload Page.....	17
Figure 3.5: Result Page.....	18
Figure 3.6: General Flow of Project.....	19
Figure 4.1: Normal Image.....	21
Figure 4.2: Background Removed Image.....	21
Figure 4.3: Flow of Preprocessing.....	22
Figure 4.4: Background Removed Image.....	23
Figure 4.5: Preprocessed Image.....	23
Figure 4.6: The Flow of Burn Classification.....	24
Figure 4.9: The Flow of Ensemble Model.....	28
Figure 5.1: Accuracy Comparison of Projects.....	29
Figure 5.3: Epoch versus Accuracy.....	31

LIST OF TABLES

Table 2.1: Comparison of Related Works.....	13
Table 4.7: The result of Model.....	25
Table 4.8: The weights of models.....	27
Table 4.10: Evaluation of Ensemble Model.....	28
Table 5.2: Comparison of Models.....	31

1. INTRODUCTION

1.1 Problem Description and Motivation

Human skin, a complex organ composed of multiple layers, serves as the body's first line of defense against environmental threats. It plays a vital role in thermoregulation, sensation, and immune response. When the skin is damaged by burns, these essential functions are disrupted, sometimes with life-threatening consequences. Burn injuries vary significantly in severity. First-degree burns affect only the outer epidermis and cause mild symptoms like redness and pain. Second-degree burns effect deeper into the dermis, resulting in blistering, swelling, and intense pain and Third-degree burns are the most severe, destroying all layers of skin and often affecting underlying tissues. These more severe burns may also damage nerve endings, causing reduced pain perception and complicating diagnosis.

Burn injuries represent a serious public health issue globally, particularly in low- and middle-income countries. According to the World Health Organization (WHO)[1], approximately 180,000 deaths occur each year as a result of burns. Many of these fatalities could be prevented with timely and accurate medical care. However, in regions with limited healthcare infrastructure, burn victims often experience delays in diagnosis and treatment due to a shortage of trained medical professionals. Even in better-resourced settings, the complexity of accurately determining burn depth can result in misdiagnosis, which in turn can lead to inappropriate or delayed treatment, prolonging recovery and increasing the risk of complications.

The primary motivation behind this project is to help address these diagnostic and treatment challenges through the use of image analysis and machine learning. By automating the assessment of burn severity, we aim to support healthcare professionals especially in areas with limited access to specialists with faster, more reliable diagnostic tools. This can enable earlier intervention, improved treatment outcomes, and reduced mortality in severe burn cases.

In this project, we developed a system that analyzes user-submitted images to detect the presence of burns. If burns are identified, image processing techniques are applied to

isolate affected areas. These features are then evaluated using a trained machine learning model to determine the degree of the burn. This approach aims to reduce diagnostic error and support timely medical intervention, making it a valuable tool for clinical and remote healthcare settings.

1.2 Main Goal and Objectives of the Project

The primary goal of this project is to develop a machine learning–based system capable of accurately determining the burn degree and isolating the total affected area in burn wound victims.

Objectives:

1.2.1 To accurately identify the burn degree using machine learning techniques

Develop and train a machine learning model capable of classifying burn wounds into first, second, or third degree with diagnostic accuracy comparable to or better than expert clinicians.

1.2.2 To precisely detect and quantify the burned area

Implement image processing methods to segment burn wounds from images and calculate the total affected area, a key factor in determining the severity of burns and guiding treatment protocols.

1.2.3 To develop a web application

Build a user-friendly web application that allows users to upload images of burn wounds and receive immediate diagnostic feedback, including burn classification and affected area measurements.

2. DEFINITION OF THE PROJECT

2.1 Scope of the Project

This project focuses on the development of an automated, machine learning–based system for the detection and classification of skin burns. The system is designed to assist medical professional and standard users by providing fast and reliable assessments of burn severity using input images. The primary functionality of the system includes identifying burned areas on the skin and classifying them into first, second, or third-degree burns. Additionally, it will distinguish healthy skin from burned regions to support accurate severity analysis.

In-Scope Components:

- **Burn Classification:** Implementation of a machine learning model to classify burns into first, second, or third degree categories.
- **Burned Area Identification:** Application of image processing techniques to segment and quantify burned skin regions.
- **Healthy Skin Detection:** Differentiation between affected (burned) and unaffected (healthy) skin areas to support diagnosis.
- **Web Application Development:** Design and deployment of a web-based user interface that allows users to upload images and receive diagnostic outputs, including burn degree.

Out-of-Scope Components:

- **Other Skin Conditions:** The system will not diagnose or classify any skin condition other than burns(rashes, infections, or dermatological diseases).
- **Burn Treatment or Medication Recommendation:** The system does not provide treatment suggestions or prescribe medications. Its purpose is limited to detection

and classification.

- **Support for All Skin Tones:** The model is trained primarily on images of medium to lighter skin tones. It is not currently optimized for accurate performance on darker skin tones.
- **Hardware Integration or Optimization:** The project does not involve hardware optimization, deployment on edge devices, or partnerships with hardware vendors. The development will be limited to software and algorithmic components.
- **Mobile or Cross-Platform Deployment:** While the system is designed as a web application, mobile app development or cross-platform integration is not included within the current scope.

Assumptions:

- **High-Quality Image Input:** It is assumed that users will upload clear, high-resolution, noise-free images of burn wounds for optimal performance.
- **Only Burn Image Input:** It is assumed that users will upload only burn wound images.
- **Controlled Lighting Conditions:** The system assumes that images are captured under adequate lighting conditions to ensure visible differentiation between healthy and burned skin.
- **Skin Tone Limitation:** The training dataset predominantly consists of images from individuals with lighter to medium skin tones. As such, performance assumptions are based on these skin types.

- **User Expertise:** It is assumed that users (healthcare professionals or trained staff) have basic knowledge of how to capture medically relevant images and interpret preliminary automated results.

Constraints:

- **High-Performance GPU Requirement:** The training and execution of the model require access to a high-performance NVIDIA GPU to ensure efficiency and speed, particularly for deep learning inference.

- **Framework Compatibility and Resource Usage:** The software is designed using machine learning frameworks that are compatible with most systems but benefit from enhanced GPU, memory, and storage for optimal performance.

- **Dataset Limitations:** Due to the dataset's skin tone bias, the model's accuracy may be reduced for underrepresented groups, particularly individuals with darker skin tones.

2.2 Success Factors

Success Factor for Objective 1.2.1: To Accurately Identify the Burn Degree

- **Key Performance Indicators:**
 - The machine learning model must achieve at least 90% accuracy on a validated test dataset.
 - The model must also reach at least 90% recall, ensuring that actual burn cases (particularly second and third degree burns) are correctly identified and not missed.
 - Evaluation metrics such as confusion matrix, F1-score, and precision will also be used to confirm that the model performs reliably across all three burn degree classes.

- **Success Criterion:**

- A burn classification model that meets or exceeds the 90% accuracy and recall threshold is considered a success for this objective.

Success Factor for Objective 1.2.2: To Accurately Detect the Burned Area

- **Key Performance Indicators:**

- The system should accurately segment the burned regions in an image.
 - Target performance is between 85% to 90% accuracy in area estimation compared to ground truth (expert-labeled or clinically validated annotations).

- **Success Criterion:**

- If the system consistently achieves burn area calculation accuracy within the 85%–90% range.

Success Factor for Objective 1.2.3: To Develop a Web Application

- **Key Performance Indicators:**

- The web application must be intuitive and usable by non-technical users such as patients or medical staff with minimal training.
 - Usability testing will be conducted with a sample of target users to assess user satisfaction, ease of use, and clarity of instructions.
 - A success rate of at least 90% task completion (e.g., image upload and interpretation of results) among test users is the target benchmark.

- **Success Criterion:**

- The application is considered successful if users can confidently upload images and understand the system's output, with minimal support or confusion, and if user feedback indicates general ease of use and usefulness.

2.3 Professional Considerations

2.3.1 Methodological Considerations

This project followed a structured engineering process, using widely accepted tools and standards for software development, research, and collaboration:

- **Source Control: Git and GitHub**

- All source code, model files, and documentation were version-controlled using Git, and hosted on GitHub to ensure collaborative development and version tracking.

- **Communication & Collaboration:**

- WhatsApp and Google Meet were used for real-time communication and regular meetings.
- Google Drive was used for storing and sharing research materials, reports, and datasets.

- **Research Standards:**

- Scientific literature was reviewed using Google Scholar, IEEE Xplore, and ScienceDirect, adhering to academic citation and referencing norms.

- **Programming Language and Libraries:**

- Python was used as the main language for algorithm development, with TensorFlow, PyTorch, and Keras as the main machine learning libraries.

- **Visualization and Documentation Tools:**

- Matplotlib, Seaborn, and Plotly were used for visualizing training progress and evaluation metrics.
- Draw.io was used to create system architecture diagrams and flowcharts.

- **Modeling Standards:**

- Model evaluation used standard metrics such as Accuracy, Recall, F1 Score, and Confusion Matrices.
- UML diagrams were used to depict the architecture of the web application and data flow.

2.3.2 Realistic Constraints

This project was carried out under a range of real-world constraints, which were identified and addressed as follows:

Economical

- The project has minimal environmental impact. It involves software development and does not include any hardware components or processes that contribute to environmental degradation.

Environmental

- The project has minimal environmental impact. It involves software development and does not include any hardware components or processes that contribute to environmental degradation

Ethical

- The project complies with ethical standards concerning data usage. In the event that real patient images are utilized, data is anonymized and handled under strict data privacy regulations.
- All research and experimentation are conducted responsibly, ensuring fairness and transparency in model predictions.

Health and Safety

- No direct physical safety risks are associated with this software-based project.
- Indirectly, the system supports improved health outcomes by facilitating more accurate and timely burn assessments.

Sustainability

- Although the system itself does not directly address sustainability, the software's design focuses on long-term usability and potential for integration into healthcare infrastructure.
- Future mobile app versions could contribute to remote and sustainable healthcare delivery.

Social

- The system empowers healthcare professionals and potentially even non-experts by providing a fast, accessible diagnostic tool.
- It promotes equitable healthcare access in underserved regions and contributes to improved patient outcomes through early intervention.

2.3.3 Legal Considerations

- **Data Protection:** The system is designed in compliance with KVKK (Kişisel Verilerin Korunması Kanunu – Turkish Personal Data Protection Law). Any medical or personal data used is anonymized and stored securely.
- **Medical Compliance:** While the current project is a prototype and not a certified medical device, future commercial deployment would require approval from health authorities such as TÜSEB or the Turkish Ministry of Health.

➤ **Licensing:** All used libraries and frameworks are open-source, under licenses such as MIT, Apache 2.0, or GNU GPL, ensuring legal use within academic and prototype settings.

➤ **Web Platform Use:** The web application follows standard internet privacy practices, with secure data transmission (HTTPS) and proper user interaction design.

2.4 Related Work

2.4.1 Automatic Segmentation and Degree Identification in Burn Color Images

This study [2] applied Cr-Transformation, Luv-Transformation, and Fuzzy C-Means (FCM) for segmenting burn areas. Classification was done using Bayes, k-NN, and SVM, with SVM showing the best results. Our approach builds on this by using similar color transformations but enhances segmentation precision through additional features like kurtosis, hue, and skewness.

2.4.2 Segmentation and Classification with Artificial Intelligence

In [3], YOLOv7 was used to detect and classify burns in a mobile app using a dataset of 21,018 images, achieving 75.12% test accuracy.

Unlike this work, which uses black-box deep models, our project prioritizes interpretability with color-based techniques, offering greater transparency and control.

2.4.3 Deep Learning-Assisted Burn Wound Diagnosis

This paper [5] used U-Net and Mask R-CNN to detect burn wounds and calculate %TBSA pixel-wise. Mask R-CNN achieved high segmentation accuracy (Dice = 0.9496) and low TBSA deviation.

Inspired by this, we evaluated multiple ML models to identify the most accurate and interpretable solution.

2.4.4 Burn Image Segmentation and Depth Diagnosis Framework

In [6], deep learning was used to segment burns and assess burn depth. The framework achieved IOU scores of 0.8467 (burn vs. non-burn) and 0.5144 (depth classification).

In contrast, our system avoids edge-based methods like Gaussian blur and focuses on a multi-model, color-based segmentation strategy more suited to visible burn regions.

Related Work	2.4.1	2.4.2	2.4.3	2.4.4	Our Project
Image Preprocessing Techniques	Cr-Transformation, Luv-Transformation, FCM	Histogram of Oriented gradients (HOG)	Crop - resize	Gaussian Blue, sharpening, pseudo labeling	Cr-Transformation, Luv-Transformation, FCM
Size of The Data Set	Not mentioned	21,018 images	2,591 images	1,200 images	5,676 images
Segmentation Method	Feature Extraction(h-space, contrast, homogeneity)	Edge Detection Based Segmentation	Semantic and instance	Encode and Decode network	Feature Extraction (Hue, Skewness, Kurtosis)
Machine Learning Model	SVM	Yolov7	U-net and ResNet	ResNet, HrNet	RegNetY-080
Application	No	Yes	No	No	Yes
Accuracy	89.29%	75.12%	92.93%	90.32%	94.11%

Table 2.1: Comparison of Related Works

3. SYSTEM DESIGN AND SOFTWARE ARCHITECTURE

3.1 Project Requirements

3.1.1 Functional Requirements

- **Image Uploading:**

The system must allow users to upload burn images via a web interface.

- **User Approval:**

Users must be informed and must approve consent before the classification process begins.

- **Background Removal:**

The system must remove the background of the uploaded image using models like RemBG or U2Net, then compare results and choose the most suitable output.

- **Burn Area Extraction:**

The system must detect and separate the burned area from the image.

- **Image Classification:**

The processed image must be classified using the trained RegNetY-080 deep learning model.

- **Result Display:**

The classification result (e.g., 1st, 2nd, or 3rd-degree burn) must be displayed on a result page.

- **End-to-End Flow Automation:**

The entire process from uploading to result rendering must flow without manual intervention.

3.1.2 Nonfunctional Requirements

- **Usability:**

The interface must be user-friendly and intuitive for users with no technical background.

- **Performance:**

The system should provide classification results within a reasonable time.

- **Accuracy:**

The classification accuracy must meet a minimum threshold (>90%).

- **Scalability:**

The system should support batch processing or future expansion for classifying multiple images simultaneously.

- **Model Comparison Robustness:**

The background removal module must reliably compare outputs from different models and choose the best result.

- **Data Privacy:**

User-uploaded images must be processed securely without being stored unless user consent is obtained.

- **Cross-Platform Compatibility:**

The application must be accessible from desktop and mobile browsers.

3.2 System Design

3.2.1 UML Use case Diagram(s) for the main use cases

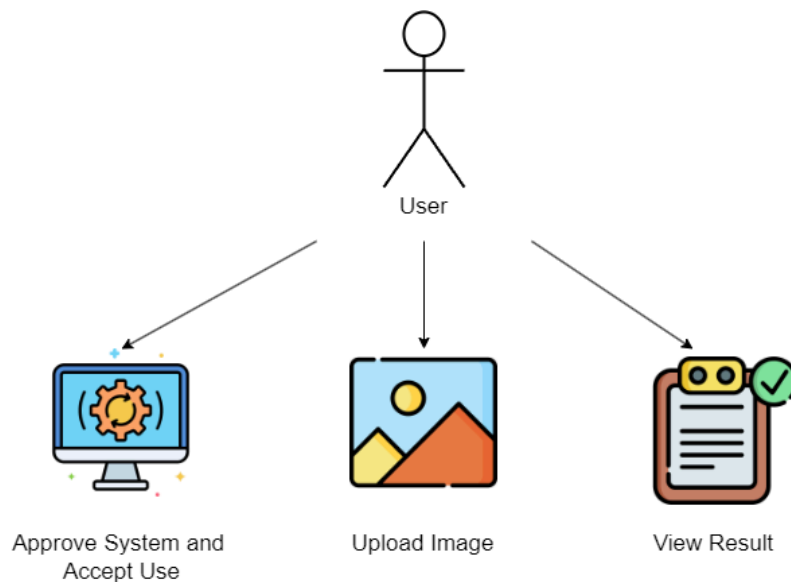


Figure 3.1: Use Case Diagram

3.2.2 UML Class and/or Database ER diagram(s)

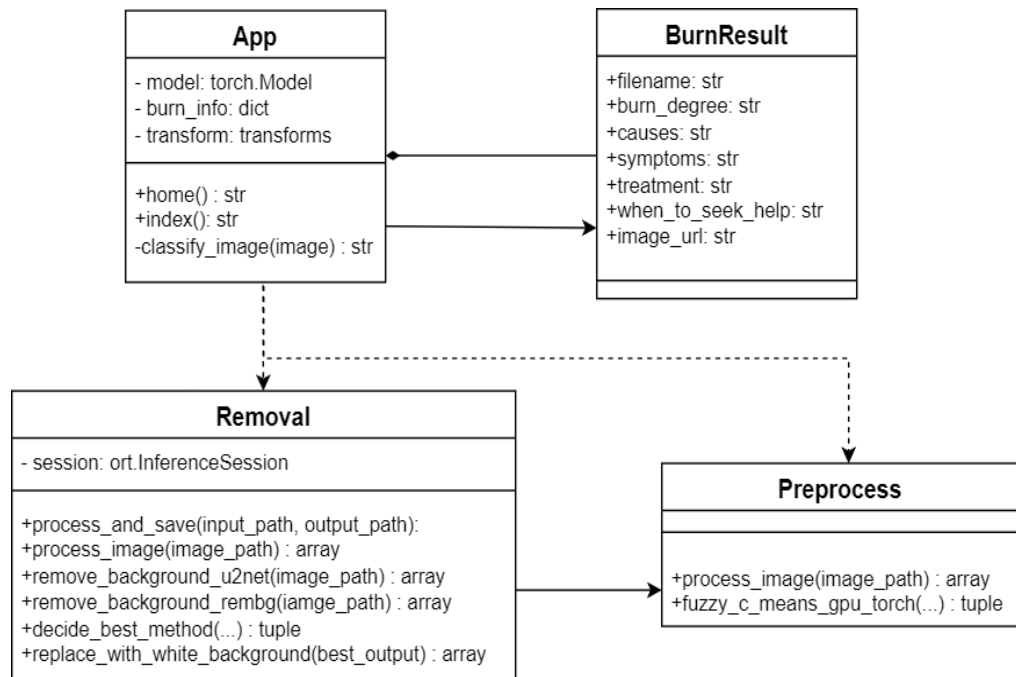


Figure 3.2: Uml Class Diagram

3.2.3 User Interface

The system includes a simple and user-friendly web interface composed of three main pages, each designed to guide the user throughout the image classification process in a clear and structured manner.

- **Information Page:**

The first page informs the user about how to properly use the system. It highlights important instructions to obtain accurate results, such as:

- Uploaded images must contain a burn.
- Uploaded images should not contain excessive noise.
- Uploaded images should be clear and not overly bright.

Once the user reads and approves this information, they can proceed to the upload page. The page is shown Figure 3.3 .

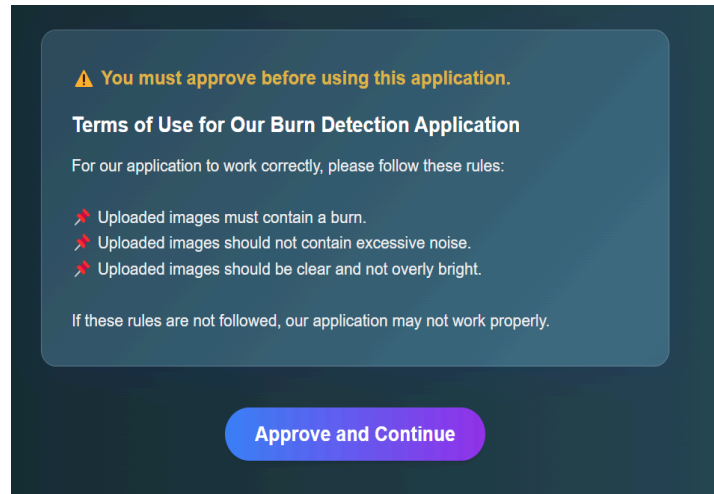


Figure 3.3: Information Page

- **Image Upload Page:**

On this page, users can upload one or multiple images. A file selection interface allows multi-file input, and once the files are uploaded, users can click on the filenames to preview the selected images. This helps ensure that the correct files have been uploaded before initiating analysis. The image upload page is shown in Figure 3.4 .

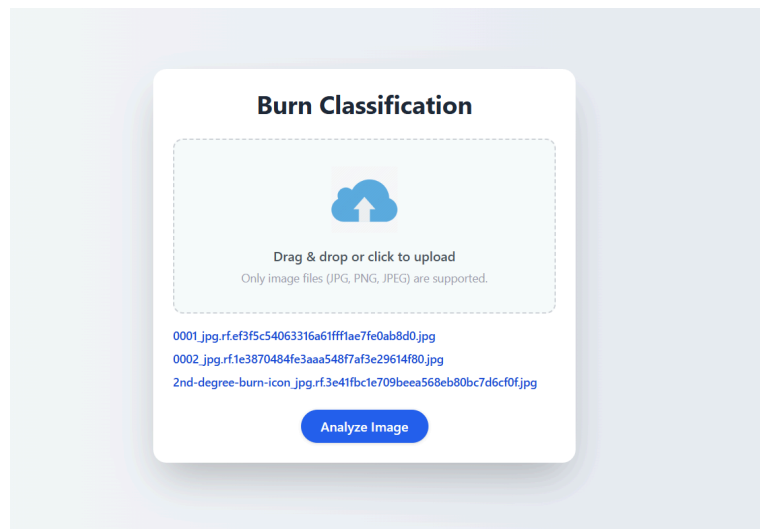


Figure 3.4: Image Upload Page

- **Result Page:**

After the images are uploaded, the user can click the “Analyze Image” button. This triggers the background removal, burn area extraction, and classification processes. Once completed, the system displays the classification results, indicating the burn severity level (e.g., 1st, 2nd, or 3rd degree) for each uploaded image in a structured and readable format. The page includes two additional buttons to enhance user experience:

- **Show Burn Info:**

This button provides informative content based on the detected burn degree. It includes clinical details such as typical symptoms, possible treatments, and recommended medical response procedures. This feature aims to increase user awareness and provide guidance on next steps.

- **Process Another Image:**

This button allows the user to return to the Image Upload Page and start the process again with new images.

The result page is shown in Figure 3.5 .

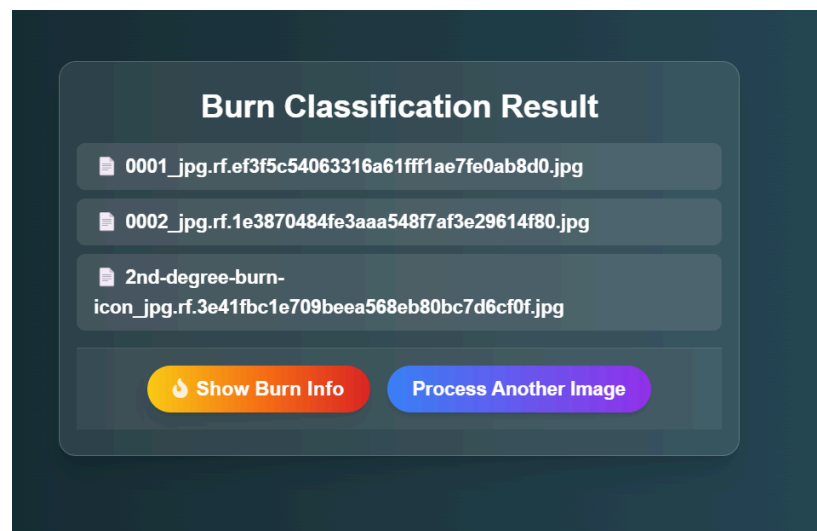


Figure 3.5: Result Page

3.2.4 Test Plan

The testing process will include both functional and non-functional test cases to ensure the reliability and accuracy of the burn classification system. The proposed test plan includes:

- **Unit Tests:**

Each module (e.g., background removal, image preprocessing, classification) will be tested individually using Python-based unit testing tools such as pytest.

- **Integration Tests:**

After individual modules are validated, the entire pipeline will be tested to verify smooth data flow and model interaction.

- **User Interface Tests:**

Basic UI functionalities like file upload, result display, and error handling will be tested manually.

- **Performance Tests:**

The response time for classification and background removal will be measured, with a target of <10 seconds per image.

- **Validation Tests:**

Model predictions will be compared to ground-truth labels using metrics such as accuracy, precision, recall, and F1-score.

3.3 Software Architecture

The general flow of the project is shown below in Figure 3.6 .

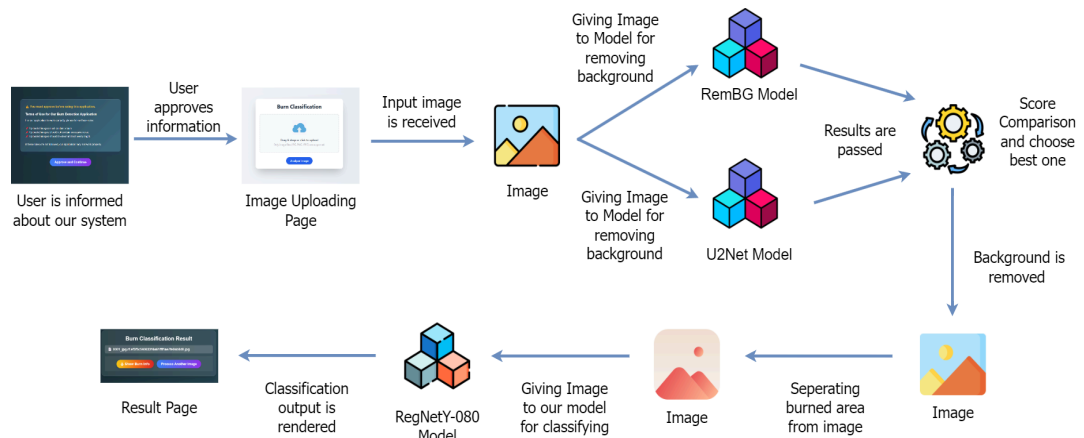


Figure 3.6: General Flow of Project

4. TECHNICAL APPROACH AND IMPLEMENTATION DETAILS

4.1 Technical Approach and Implementation Details

The proposed system for automated skin burn classification leverages a multi-stage pipeline combining background removal, targeted preprocessing, and deep learning-based classification. Each component of the pipeline was carefully designed and optimized to ensure robust performance across diverse image conditions. The complete workflow is structured into three primary phases: (1) Background Removal, (2) Burn Region Isolation via Preprocessing, and (3) Classification Using a Custom Deep Learning Architecture. The implementation was carried out using Python, PyTorch, Albumentations, and supporting image processing libraries.

4.1.1 Background Removal

Accurate background removal is a crucial preprocessing step, ensuring that classification focuses exclusively on the region of interest, namely the skin and burn area. Two distinct pretrained models were integrated into the pipeline for background subtraction:

- **RemBG:** A general-purpose background removal model based on U-2-Net, designed for object segmentation across diverse domains.
- **U-2-Net (Body-focused variant):** A saliency-aware deep neural network tailored for salient object detection, particularly effective at isolating human body parts.

Model Selection via Quality Assessment

To determine the superior background removal output per image, we devised a comprehensive evaluation framework based on the following four criteria:

- **Structural Similarity Index (SSIM):** Measures the visual similarity between the original and the masked image, preserving overall structure and luminance consistency.

- **Edge Similarity via Canny Detection:** Assesses alignment of high-frequency features between the original and masked image by applying Canny edge detectors and comparing edge maps.

- **Foreground Pixel Density:** Evaluates the proportion of non-background (non-transparent) pixels remaining after segmentation to ensure the foreground is preserved.

- **Histogram Difference (Pixel Intensity Discrepancy):** Computes pixel-wise histogram differences between original and masked images, ensuring color consistency.

Additionally, a saliency-based check was introduced to bypass background removal entirely if the input image exhibited no clearly distinguishable foreground. This mechanism avoids introducing artifacts when segmentation is likely to fail or provide negligible benefits.

The final output of this phase was the optimal background-removed image, selected adaptively per input. Transparent regions were post-processed by replacing them with a white background to maintain consistency in image format and facilitate downstream processing.



Figure 4.1: Normal Image

Background
→
Removal



Figure 4.2: Background Removed

4.1.2 Targeted Preprocessing for Burn Region Isolation

Following background subtraction, a targeted preprocessing pipeline was applied to isolate burn areas from healthy skin and surrounding tissue. The procedure involves color space transformations and custom clustering-based segmentation.

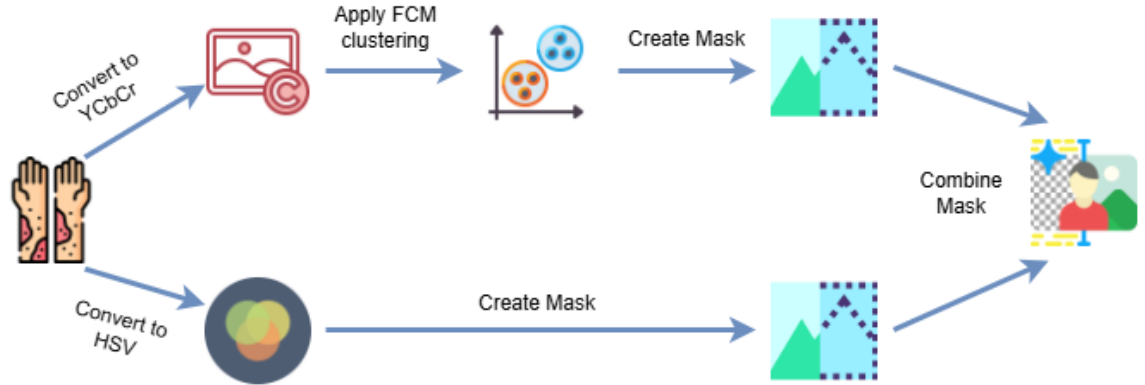


Figure 4.3: Flow of Preprocessing

Conversion to YCbCr and FCM Clustering

- The input image was converted from the RGB to YCbCr color space, leveraging the Y (luminance) and Cr (chrominance-red) channels.
- Bright regions, typically corresponding to glare or lighting artifacts, were suppressed by masking high-intensity values in the Y channel.
- A custom Fuzzy C-Means (FCM) clustering algorithm was applied on the Cr channel to segment skin tones and differentiate healthy skin from potentially affected areas.
- The FCM algorithm employed a soft membership strategy, which allowed flexible separation of ambiguous pixels, improving segmentation of transitional burn zones.

HSV-Based Detection of Severely Burned Areas

- To further enhance detection of severely burned (darkened) regions, the image was transformed into the HSV (Hue, Saturation, Value) color space.
- Pixels exhibiting low saturation and low brightness (value) were heuristically identified as severely burned areas, typically corresponding to third-degree burns or necrotic tissue.
- These identified regions were binarized into a separate mask and merged with the FCM-derived segmentation mask to create a comprehensive burn region mask.

Final Masking

The merged mask was used to:

- Retain only the identified burn areas in the image.
- White out all other regions, including unaffected skin and residual background elements, resulting in an image focused entirely on the burn site.



Figure 4.4: Background Removed Image

Preprocess
→

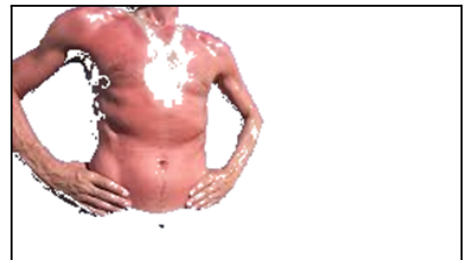


Figure 4.5: Preprocessed Image

These standardized and cleaned images were passed to the classification model.

4.1.3 Burn Classification Using RegNetY-080

The classification backbone was constructed using a custom-trained RegNetY-080 architecture, a modern convolutional neural network optimized for efficient scaling and high representational capacity.

Architectural Overview

- RegNetY introduces meta-parameterized scaling functions that determine the network's depth, width, group width, and bottleneck ratio.
- The Y-variant includes Squeeze-and-Excitation (SE) blocks, which improve feature recalibration by modeling interdependencies between channels, helping the model focus on more informative features—particularly valuable for medical image analysis.

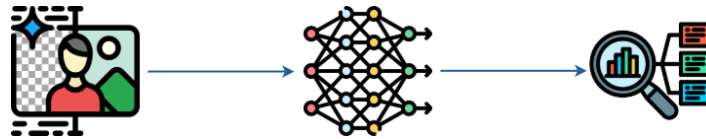


Figure 4.6: The Flow of Burn Classification

Training Configuration

- The model was trained from scratch using PyTorch, with no reliance on pretrained weights, to adapt fully to the specific features of burn images.
- **Albumentations** was employed for data augmentation, incorporating:
 - Random rotations
 - Brightness/contrast adjustments
 - Horizontal and vertical flips
 - Elastic transformations

- To address class imbalance (first, second, and third degree burns), a `WeightedRandomSampler` was implemented, ensuring equitable representation of all classes during training.

Optimization Strategy

- **Loss Function:** Cross-entropy with **label smoothing** to mitigate overconfidence and improve generalization.

- **Learning Rate Scheduler:** **Cosine annealing** was adopted to progressively reduce the learning rate, helping the model escape local minima and converge smoothly.

- **Early Stopping:** Monitored validation accuracy and loss to terminate training early if overfitting was detected.

This final classifier takes the masked, preprocessed image as input and outputs a prediction corresponding to one of the three burn severity classes.

To evaluate the model's performance, early stopping was employed alongside a cosine annealing scheduler to optimize the training process. Accuracy, precision, recall, and F1-score were calculated, and the comprehensive results are presented in Table 4.7.

Accuracy	94.11%
Precision	93.67%
Recall	94.23%
F1-Score	93.74%

Table 4.7: The result of Model

4.2 Alternative Approach (Ensemble Learning)

A review of the relevant literature revealed that ensemble learning has not been widely explored in this specific context. Therefore, this project also employs an ensemble learning approach to improve the classification performance for skin burn degree prediction.

Ensemble learning is a methodology that aims to achieve more accurate and stable predictions by combining multiple machine learning models. The fundamental assumption is that multiple weak learners can collectively form a strong predictor. This approach helps reduce variance, bias, and overfitting issues that may occur in individual models.

There are three types of ensemble learning. In the project, a soft voting ensemble classifier has been developed. In soft voting, different types of models work together and their predictions are combined with averaging probability. In the project, the classifier combines five different base models:

- KNN – for simplicity and local decision boundaries and euclidean is used for distance metric.
- Random Forest – for robust performance through bagging and decision trees
- Support Vector Machine (SVM) – for handling complex, non-linear decision surfaces
- XGBoost (Extreme Gradient Boosting) – a boosting based model known for high performance
- LightGBM (Light Gradient Boosting Machine) – a lightweight gradient boosting model optimized for speed and accuracy.

Before incorporating XGBoost and LightGBM models into the ensemble, a hyperparameter optimization process has been conducted using Optuna, an efficient framework for automated hyperparameter tuning. The objective is to maximize the individual performance of these models on the given dataset prior to integration. After identifying the best-performing hyperparameters through multiple optimization trials, both XGBoost and LightGBM were trained with these tuned parameters. The final optimized models were then included in the ensemble classifier to ensure that each base learner contributed at its most effective capacity.

After integrating the optimized base models, multiple weight combinations were systematically tested to determine the most effective configuration for the ensemble classifier. Each model's contribution to the final prediction was adjusted through soft voting weights. And final weights are shown in Table 4.8 .

KNN	1
Random Forest	2
SVM	2
XGBoost	2
LightGBM	3

Table 4.8: The weights of models

While the ensemble model achieved high performance in identifying third-degree burns, it occasionally struggled to accurately distinguish between first-degree and second-degree burns. This limitation is primarily due to the visual similarity of these burn types, particularly in terms of color distribution and texture, which can cause overlapping feature representations in the input space.

To address this challenge, a two-stage classification approach has been adopted. In the first stage, the ensemble classifier makes an initial prediction across all three classes. If the predicted class is either 1 or 2, a secondary classifier based on CatBoost is employed to refine the decision. This model is specifically trained to distinguish between only class 1 and class 2, allowing it to focus on subtle differences between them.

This hierarchical decision strategy significantly improved the classification accuracy for borderline cases, leveraging both the generalization power of the ensemble and the precision of a focused binary classifier. The main flow of this is shown in Figure 4.9 .

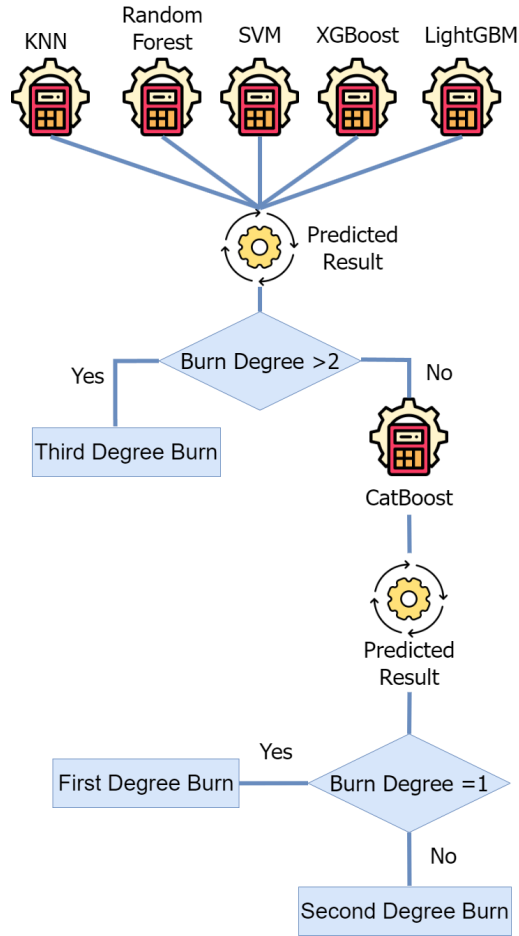


Figure 4.9: The Flow of Ensemble Model

To evaluate the model performance, a GroupKFold cross-validation strategy with 5 folds has been applied. Accuracy, Precision, Recall and F1-Score are calculated and results are shown in Table 4.10 .

Accuracy	82.66%
Precision	84.15%
Recall	82.26%
F1-Score	82.84%

Table 4.10: Evaluation of Ensemble Model

5. SOFTWARE TESTING

5.1 Implementation Phase

The implementation phase involved developing a machine learning-based system for automated burn classification and TBSA (Total Body Surface Area) estimation. The project was implemented using Python, leveraging libraries such as TensorFlow, Keras, PyTorch, and OpenCV for model development and image processing.

We developed multiple models, tested different algorithms, and fine-tuned hyperparameters using the curated dataset of 5,676 burn images. The models were trained to classify burns into first, second, and third degrees and to estimate the affected body area. The implementation also includes a web application built with Flask (backend) and HTML/CSS/JS (frontend), allowing users to upload images and receive instant analysis.

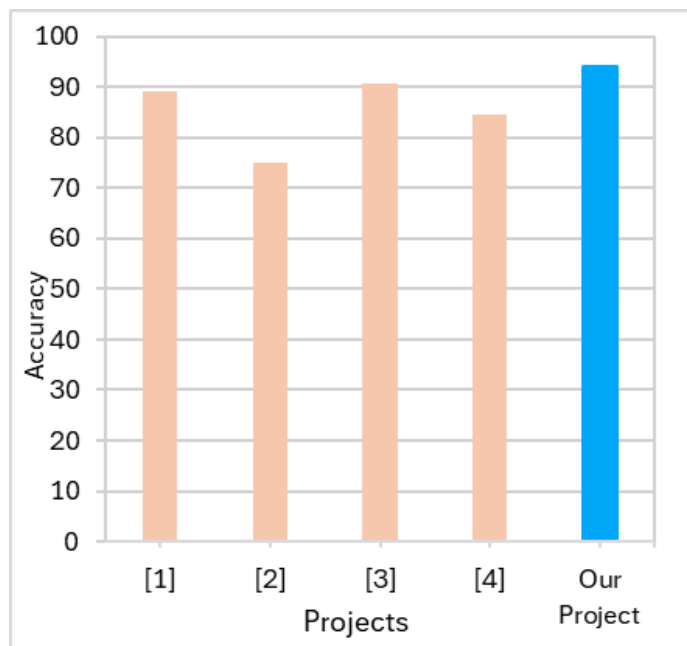


Figure 5.1: Accuracy Comparison of Projects

5.2 Testing Phase

Testing focused on evaluating both the model's classification accuracy

Test Tools and Frameworks

- Google Colab and local GPU systems were used for training and testing.
- Scikit-learn, TensorBoard, and Matplotlib were used to evaluate and visualize model performance.
- Draw.io and Python-based visualization libraries (e.g., Seaborn) were used for architecture diagrams and result presentation.

Experimental Setup

- **Dataset Split:** 80% training, 10% validation, 10% testing.

Evaluation Metrics

- **Accuracy**
- **Precision**
- **Recall**
- **F1-score**

Results

Traditional machine learning algorithms yielded the lowest accuracy. Ensemble learning, which combines multiple models and selects the best-performing one, achieved an accuracy of up to 82.66%. A custom CNN with tuned parameters performed slightly better, reaching 84.11%. However, our best-performing model, RegNetY-080, achieved a significantly higher accuracy of 94.11%.

Model	Accuracy	Precision	Recall	F1 Score
SVM	72.11%	72.35%	72.12%	71.71%
KNN	72.96%	73.96%	73.23%	72.99%
Ensemble Learning	82.66%	84.15%	82.26%	82.84%
Custom CNN	84.11%	83.87%	82.67%	83.79%
Our Model	94.11%	93.67%	94.23%	93.74%

Table 5.2: Comparison of Models

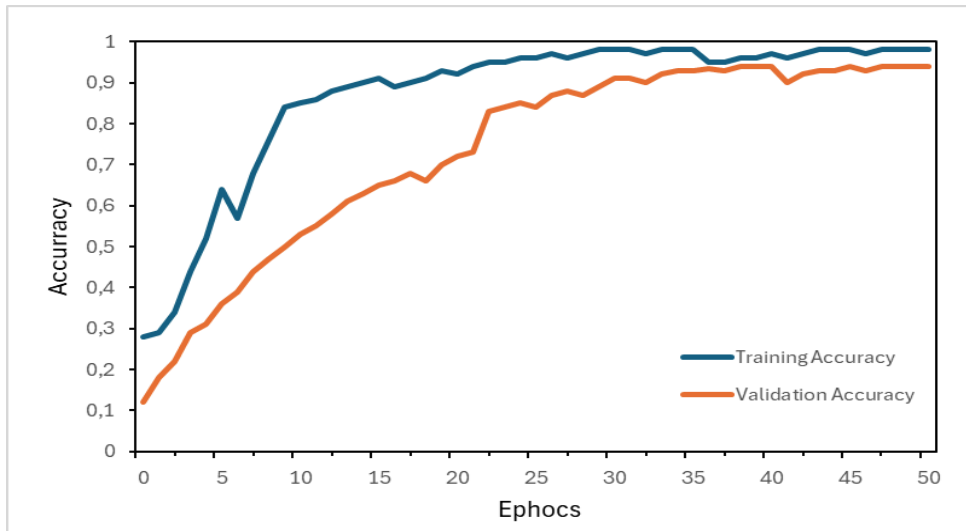


Figure 5.3: Epoch versus Accuracy

6. BENEFITS AND IMPACT

6.1 Scientific Impact

- **Advancement of AI in Medical Diagnostics:** The project showcases how deep learning and image processing techniques can be effectively applied to solve real-world medical problems.

- **Research Contribution:** The methodology, training process, and results could serve as the basis for **future academic publications** or conference presentations in the fields of medical imaging, machine learning, and digital health.

- **Data and Model Sharing:** The trained model and annotated datasets (if made available) could support further academic research in dermatology and computer-aided diagnosis.

6.2 Economic, Commercial and Social Impact

- **Improved Healthcare Access:** By automating burn assessment, the system can support healthcare workers in low-resource and rural areas, helping to bridge gaps in specialist availability.

- **Enhanced Patient Outcomes:** Faster and more accurate diagnosis can lead to earlier intervention and better recovery rates, improving patient quality of life.

- **Cost Efficiency:** Reducing diagnostic errors and unnecessary referrals can lower overall healthcare costs, particularly in overstretched systems.

- **Commercial Potential:** The project could evolve into a commercial product or service (e.g., SaaS for hospitals or NGOs), potentially forming the basis for a startup or being integrated into existing telemedicine platforms.

6.3 Potential Impact on New Projects

The success of this project could serve as a pioneering model for future initiatives:

- **Inspiration for Other AI-Driven Medical Tools:** This system may encourage similar applications of machine learning in diagnosing other conditions (diabetic ulcers, skin cancers, or infections).
- **Interdisciplinary Collaboration:** The project opens opportunities for partnerships between AI developers, medical researchers, public health organizations, and humanitarian groups.
- **Scaling Across Use Cases:** The framework and architecture used can be adapted or extended to support other **image-based diagnostic tools** in telehealth and emergency care.

7. CONCLUSION AND FUTURE WORK

7.1 Conclusion

In this project, we successfully developed an automated system for classifying burn images using machine learning techniques. The system accurately identifies the degree of burns—first, second, or third—based on image input and estimates the affected body surface area. Unlike previous studies that primarily relied on publicly available datasets or simple classifiers, we created a custom dataset from scratch due to the lack of suitable open-source alternatives. This dataset was enhanced through data augmentation techniques, resulting in a total of 5,676 labeled burn images.

Throughout the development phase, we experimented with a range of machine learning algorithms and pre-trained models. Our final system achieved a classification accuracy of 94.11%, demonstrating its reliability and clinical relevance. Notably, our project is among the first to apply ensemble learning techniques in burn classification, which contributed significantly to performance improvements.

In addition to the machine learning model, we developed a web application to ensure the system is accessible to both healthcare professionals and the general public. This user-friendly platform allows users to upload burn images and receive immediate classification results and affected area estimates, potentially supporting early diagnosis and treatment planning in real-time.

7.2 Future Work

Several areas have been identified for improvement and expansion in future iterations of this project:

- **Mobile Application Development:** To enhance accessibility and usability, particularly in remote or low-resource regions, we plan to develop a mobile version of the application.

- **Dataset Expansion and Diversity:** We aim to increase the size of the dataset and ensure better representation of darker skin tones to improve model fairness and generalization.

- **Multilingual and Offline Support:** Expanding language support and enabling offline functionality can further increase the tool's global usability.

- **Integration with Medical Systems:** Future versions may integrate with hospital databases or electronic health records for streamlined clinical workflows.

- **Real-Time Video Analysis:** Enhancing the system to analyze video input for continuous monitoring of wound healing over time.

REFERENCES

- [1] “Burns”, WorldHealth Organization: <https://www.who.int/news-room/fact-sheets/detail/burns>. Date accessed: 17 October 2024
- [2] Wantanajittikul K, Auephanwiriyakul S, Theera-Umpon N, Koanantakool T. 2012. “Automatic segmentation and degree identification in burn color images”, BMEiCON-2011, 1-5.
- [3] Yıldız M, Sarpdağı Y, Okuyar M, Yildiz M, Çiftci N, Elkoca A, et al. 2024. “Segmentation and classification of skin burn images with artificial intelligence: Development of a mobile application”, Burns, 50(4), 966-979.
- [4] Yadav DP, Sharma A, Singh M 2019. “Feature extraction-based machine learning for human burn diagnosis from burn images”, Med Imaging Diagn Radiol. , 7, 1-9.
- [5] Che Wei Chang, MD; Feipei Lai, et al 2021. “Deep Learning–Assisted Burn Wound Diagnosis: Diagnostic Model Development Study,” *JMIR Medical Informatics*, 9(12)
- [6] Brenda Rangel-Olvera, Roberto Rosas-Romero, 2024. “Detection and Classification of Skin Burns on Color Images Using Multi-Resolution Clustering and the Classification of Reduced Feature Subsets,” *Multimedia Tools and Applications*, Springer Nature, Volume 83(54925–54949).