

Multi-Modal System for Acne Severity Prediction and Personalized Care

Dr. Nilesh Mali

Head of the Department

Ajeenkya D. Y. Patil School of Engineering, Pune

Darshan Bankar

Computer Engineering

Ajeenkya D. Y. Patil School of Engineering, Pune

Lalit Wani

Computer Engineering

Ajeenkya D. Y. Patil School of Engineering, Pune

Sonali Patil

Computer Engineering

Ajeenkya D. Y. Patil School of Engineering, Pune

Abstract:

Acne vulgaris is one of the most common dermatological conditions affecting adolescents and young adults worldwide. Existing automated acne detection systems primarily rely on facial image analysis and often ignore lifestyle-related factors such as diet, sleep, stress, hydration, and physical activity, which significantly influence acne severity and recurrence. This paper presents a multi-modal intelligent skincare system that integrates deep learning-based facial acne analysis with lifestyle-driven risk prediction to provide personalized acne assessment and recommendation generation.

The proposed system employs a fine-tuned ResNet-18 convolutional neural network for acne severity classification from facial images into four categories: Clear, Mild, Moderate, and Severe. To enhance explainability, MediaPipe Face Mesh is utilized for region-wise facial landmark detection and inflammation density analysis across key facial zones. In parallel, a Random Forest classifier analyzes structured lifestyle inputs to estimate behavioral acne risk. The outputs of both models are integrated using a Cross-Modal Contextual Fusion (CMCF) mechanism, enabling context-aware prediction refinement by combining visual and behavioral evidence.

Furthermore, the system incorporates a large language model-based recommendation engine to generate personalized skincare, dietary, hydration, and sleep-related guidance in natural language. The complete framework is deployed as a web-based application using React, FastAPI, and MongoDB Atlas for real-time accessibility and longitudinal tracking. Experimental evaluation demonstrates that integrating lifestyle analytics with image-based prediction improves overall classification reliability and provides more personalized and explainable acne management compared to conventional single-modal approaches. The proposed framework also highlights the potential of multi-modal artificial intelligence systems in preventive healthcare and intelligent dermatological support applications. The system is designed to be scalable, user-friendly, and accessible for everyday skincare monitoring and early acne management.

Keywords— Acne Prediction, Deep Learning, ResNet-18, Random Forest, Multi-Modal Fusion, MediaPipe Face Mesh, Personalized Recommendation, Healthcare AI, Computer Vision.

1. INTRODUCTION

A. Background and Problem Context

Acne vulgaris is a chronic inflammatory skin condition affecting the pilosebaceous units, commonly appearing as comedones, papules, pustules, and nodules. It primarily affects adolescents and young adults and can significantly impact psychological wellbeing, confidence, and social interaction. In some cases, it may also lead to permanent scarring if not managed properly.

Traditional acne diagnosis is performed through manual visual examination by dermatologists. This process is subjective, depends on clinical experience, and may not always be accessible in remote or resource-limited areas. Due to the increasing need for digital healthcare solutions, automated skin analysis systems have gained importance in recent years.

B. Research Gap and Motivation

Recent developments in machine learning and deep learning have enabled acne detection using facial images. Convolutional Neural Networks (CNNs) such as ResNet-based models are commonly used for image classification tasks. However, most existing approaches focus only on image-based acne severity prediction.

These systems generally do not consider lifestyle-related factors such as diet, sleep, stress, hydration, and physical activity, which are known to influence acne conditions. As a result, predictions from image-only systems may not fully reflect the actual causes of acne.

Additionally, many existing systems provide only severity classification without offering detailed explanations or personalized suggestions. This limits their usefulness for users seeking preventive care and daily skincare guidance.

Therefore, there is a need for a system that combines image-based analysis with basic lifestyle information to provide more informative and user-friendly outputs.

C. Proposed System Overview

To address these limitations, this project proposes a web-based acne analysis and recommendation system.

The system includes the following components:

- A ResNet-18 based deep learning model for classifying acne severity from facial images

- MediaPipe Face Mesh for detecting facial landmarks and identifying acne-affected regions
- A Random Forest classifier to analyze lifestyle-related inputs such as diet, sleep, stress, and hydration

The outputs from image analysis and lifestyle analysis are combined to improve overall prediction reliability using a simple feature combination approach.

In addition, a Large Language Model (LLM) is used to generate personalized skincare and lifestyle recommendations in natural language format, making the system easier for users to understand and follow.

The system is implemented as a web application using React for frontend, FastAPI for backend, and MongoDB Atlas for data storage, enabling user interaction, prediction, and report generation.

D. Contributions of the Work

The main contributions of this project are:

1. Development of a system for acne severity classification using ResNet-18
2. Implementation of facial region detection using MediaPipe Face Mesh
3. Design of a lifestyle-based acne risk prediction model using Random Forest
4. Combination of image and lifestyle inputs for improved prediction output
5. Generation of personalized recommendations using a Large Language Model
6. Deployment of a web-based system for real-time acne analysis and reporting

2. LITERATURE REVIEW

A. Deep Learning Approaches for Acne Detection

Deep learning techniques, particularly Convolutional Neural Networks (CNNs), are widely used for automated acne detection and severity classification. Models such as VGG16, ResNet-18, DenseNet, and EfficientNet have demonstrated strong performance in dermatological image analysis by extracting features such as skin texture, lesion boundaries, and inflammation patterns. Among these, ResNet-18 is commonly preferred due to its residual learning capability and improved classification performance. However, most existing CNN-based systems rely only on facial image analysis and ignore lifestyle-related factors affecting acne severity.

Model	Strength	Limitation
VGG16	Simple architecture	High computational cost
ResNet-18	Improved residual learning	Requires fine-tuning
DenseNet	Efficient feature reuse	High memory usage
EfficientNet	Balanced performance	Sensitive hyperparameter tuning

Table 2.1: Comparison of CNN Architectures

B. Lifestyle Factors and Acne Prediction

Medical studies indicate that acne severity is influenced by lifestyle and behavioral conditions such as diet, stress, sleep quality, hydration, and physical activity. Machine learning models such as Decision Trees and Random Forests are commonly used for analyzing structured lifestyle data. Random Forest is widely adopted due to its robustness and ability to handle non-linear relationships. However, most existing systems analyze lifestyle information separately from facial image analysis, limiting contextual understanding.

Factor	Effect on Acne
High sugar diet	Increases sebum activity
Stress	Causes hormonal imbalance
Poor sleep	Reduces skin recovery
Low water intake	Affects skin hydration

Table 2.2: Lifestyle Factors Affecting Acne

C. Facial Region Analysis and Explainability

Many existing acne detection systems provide only global severity classification without identifying acne-affected facial regions. Recent approaches use facial landmark detection techniques to improve explainability. MediaPipe Face Mesh enables segmentation of facial regions such as forehead, cheeks, nose, chin, and jawline using dense landmark detection. Region-wise analysis improves interpretability by identifying localized acne distribution. However, these methods still lack integration with behavioral and lifestyle-based prediction systems.

D. Multi-Modal Learning and Research Gap

Modern healthcare AI systems increasingly use multi-modal learning, where image and structured data are combined to improve prediction accuracy and personalization. In acne analysis, integrating facial images with lifestyle information provides better contextual understanding compared to image-only systems.

Despite recent advancements, existing approaches still suffer from several limitations:

- Dependence on image-only prediction
- Lack of lifestyle integration
- Limited region-wise explainability
- Absence of personalized recommendation generation

To address these gaps, the proposed AcneGuard system integrates ResNet-18 based acne classification, MediaPipe Face Mesh based facial analysis, Random Forest based lifestyle prediction, and a Cross-Modal Contextual Fusion mechanism for improved acne assessment and personalized skincare recommendation.

3. PROPOSED METHODOLOGY

A. Overview

The proposed system is a multi-modal dermatological analysis framework designed to predict acne severity using both facial image analysis and lifestyle-based contextual information. The system integrates deep learning, machine learning, facial landmark analysis, and recommendation generation into a unified pipeline to improve prediction accuracy and personalization.

The methodology consists of the following major stages:

1. Facial Image Preprocessing
2. Acne Severity Classification using ResNet-18
3. Facial Region Analysis using MediaPipe
4. Lifestyle Risk Prediction using Random Forest
5. Cross-Modal Contextual Fusion (CMCF)
6. Personalized Recommendation Generation

The overall workflow begins with facial image upload and lifestyle data collection. The image is analyzed using a ResNet-18 model, while behavioral inputs are processed using a Random Forest classifier. Both outputs are fused using the CMCF mechanism to generate the final prediction and recommendations.

B. System Architecture

The system follows a modular multi-layer architecture consisting of four major layers:

1. Presentation Layer

Developed using React.js and Tailwind CSS, this layer provides:

- Facial image upload
- Lifestyle questionnaire
- Result visualization
- PDF report download

2. Backend Processing Layer

Implemented using FastAPI, this layer handles:

- Image preprocessing
- Request validation
- API communication
- Model inference
- Recommendation generation

3. AI Inference Layer

This layer contains the core intelligent modules:

- ResNet-18 CNN for acne severity classification
- MediaPipe Face Mesh for facial region analysis
- Random Forest classifier for lifestyle risk prediction
- Cross-Modal Contextual Fusion (CMCF) module

4. Data Storage Layer

MongoDB Atlas is used for storing:

- user profiles

- prediction history
- reports and recommendations

The modular architecture improves scalability, maintainability, and real-time processing capability.

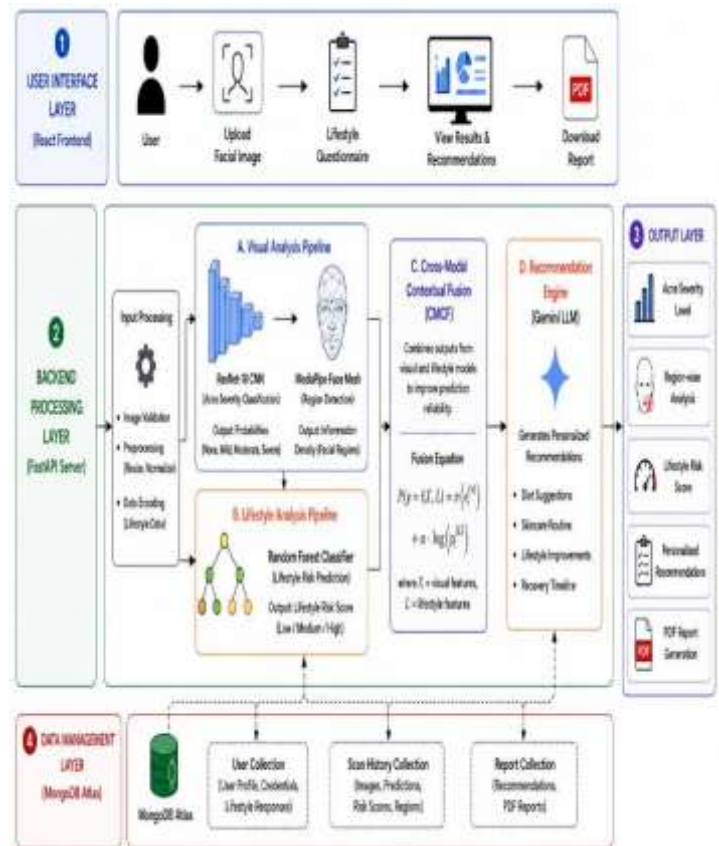


Figure 3.1: System Architecture

C. Image Preprocessing Pipeline

Before acne classification, the input facial image undergoes several preprocessing operations to ensure consistency and improve model performance.

Preprocessing Steps:

Step	Purpose
Face Validation	Ensures clear frontal face
Resizing	Converts image to 224×224
Normalization	Standardizes pixel values
Data Augmentation	Improves generalization

Data augmentation techniques such as rotation, flipping, and brightness adjustment are applied during training to reduce overfitting.

D. Acne Severity Classification using ResNet-18

The visual analysis module uses the ResNet-18 Convolutional Neural Network for acne severity prediction.

The model classifies acne into four categories:

- Clear
- Mild

- Moderate
- Severe

Transfer learning is applied using pretrained ImageNet weights to improve feature extraction efficiency.

The residual learning function used in ResNet-18 is:

$$y = F(x, W_i) + W_s x$$

The final classification probabilities are computed using the Softmax function:

$$P(y = i | x) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

Algorithm 1: Facial Acne Severity Classification

To classify acne severity from facial images using a fine-tuned ResNet-18 model.

Input: Facial image I

Output: Acne severity label $S \in \{Mild, Moderate, Severe\}$

Steps:

1. Acquire facial image from user.
2. Resize image to 224×224 .
3. Perform image preprocessing:
 - Normalization
 - Noise reduction
 - Data augmentation (during training)
4. Pass image to fine-tuned ResNet-18 CNN.
5. Extract deep visual features.
6. Apply Softmax classification layer.
7. Predict acne severity class.
8. Return predicted severity label.

Training Configuration:

Parameter	Value
Optimizer	Adam
Learning Rate	0.0001
Batch Size	32
Epochs	30

The model was trained using augmented facial image datasets to improve robustness and classification accuracy.

E. Facial Region Analysis using MediaPipe

To improve explainability, the system integrates MediaPipe Face Mesh for facial landmark detection.

The module:

- detects 478 facial landmarks
- segments the face into forehead, cheeks, nose, chin, and jawline regions

HSV-based redness detection is used to compute inflammation density for each facial region.

The inflammation density equation is defined as:

$$D_r = \frac{\sum_{(x,y) \in R} I(P_{x,y} \in H)}{Area(R)}$$

Algorithm 2: Region-Wise Facial Analysis

To identify acne-affected facial regions using MediaPipe Face Mesh.

Input: Facial image I

Output: Facial region map M

Steps:

1. Detect facial landmarks using MediaPipe Face Mesh.
2. Divide face into regions:
 - Forehead
 - Cheeks
 - Chin
 - Jawline
3. Analyze acne presence in each region.
4. Generate region-wise severity mapping.
5. Return facial analysis results.

This region-wise analysis improves interpretability by identifying acne-affected areas instead of providing only severity labels.

F. Lifestyle Risk Prediction using Random Forest

The lifestyle analysis module uses a Random Forest classifier to estimate acne risk using behavioral attributes such as:

- diet pattern
- stress level
- sleep duration
- hydration
- physical activity

The Random Forest prediction is computed as:

$$y = \frac{1}{T} \sum_{t=1}^T f_t(x)$$

Algorithm 3: Lifestyle Risk Prediction

To predict acne risk level using lifestyle and behavioral data.

Input: Lifestyle questionnaire data L

Output: Risk score $R \in \{Low, Medium, High\}$

Steps:

1. Collect lifestyle information:
 - Sleep duration
 - Water intake
 - Stress level
 - Diet habits
 - Screen time

2. Encode categorical features.
3. Normalize input values.
4. Pass features to Random Forest classifier.
5. Aggregate decision tree outputs.
6. Predict lifestyle risk score.
7. Return risk category.

Model Configuration:

Parameter	Value
Number of Trees	200
Maximum Depth	10
Criterion	Gini Impurity

Feature importance analysis revealed that diet and stress are the most influential acne-related lifestyle factors.

G. Cross-Modal Contextual Fusion (CMCF)

The CMCF module combines outputs from:

- ResNet-18 visual classifier
- Random Forest lifestyle predictor

The fusion equation is defined as:

$$P(y = i | X, h) = \sigma(z_v^{(i)} + \alpha \cdot \log(p_l^{(i)}))$$

where:

- $z_v^{(i)}$ = visual prediction score
- $p_l^{(i)}$ = lifestyle probability
- α = weighting factor

Algorithm 4: Cross-Modal Contextual Fusion (CMCF)

To combine visual acne features and lifestyle risk information for improved prediction accuracy.

Input: Visual features V , Lifestyle features L

Output: Final contextual prediction P

Steps:

1. Extract visual features from CNN.
2. Obtain lifestyle risk prediction.
3. Concatenate visual and behavioral features.
4. Apply contextual fusion mechanism.
5. Generate final acne severity prediction.
6. Return fused prediction result.

This mechanism improves prediction reliability by integrating visual and behavioral information simultaneously.

H. Recommendation Generation

The final prediction is forwarded to the Gemini 2.5 Flash LLM to generate personalized recommendations and downloadable PDF reports.

Algorithm 5: Personalized Recommendation Generation

To generate personalized skincare and lifestyle recommendations using Gemini LLM.

Input: Acne severity, Facial region analysis, Lifestyle risk score

Output: Personalized recommendation set G

Steps:

1. Collect contextual prediction results.
2. Prepare structured prompt.
3. Send prompt to Gemini 2.5 Flash LLM.
4. Generate:
 - Skincare advice
 - Diet recommendations
 - Hydration guidance
 - Lifestyle suggestions
5. Format recommendations.
6. Generate PDF report.

4. RESULTS AND DISCUSSION

A. Performance Evaluation

The performance of the proposed system was evaluated using standard classification metrics including Accuracy, Precision, Recall, and F1-score. These metrics were used to measure the

effectiveness and reliability of the multi-modal acne prediction framework.

Metric	Value
Accuracy	87%
Precision	0.85
Recall	0.84
F1-score	0.84

Table 4.1: Model Performance Metrics

The results indicate balanced classification performance across all metrics, demonstrating reliable predictive capability of the proposed system for acne severity classification. The F1-score further confirms that the model maintains a good balance between false positives and false negatives.

B. Impact of Cross-Modal Fusion

To evaluate the effectiveness of the Cross-Modal Contextual Fusion (CMCF) mechanism, the performance of the vision-only model was compared with the fused multi-modal model.

Model	Accuracy	F1-score
Vision-only (ResNet-18)	82%	0.80
With CMCF Fusion	87%	0.84

Table 4.2: Comparison of Model Performance

The results show that integrating lifestyle information with image-based predictions improves the overall system performance. The fused model achieves higher accuracy and F1-score compared to the standalone ResNet-18 model. This improvement is particularly useful in borderline acne cases where visual features alone are insufficient for reliable classification.

C. Acne Severity Classification Results

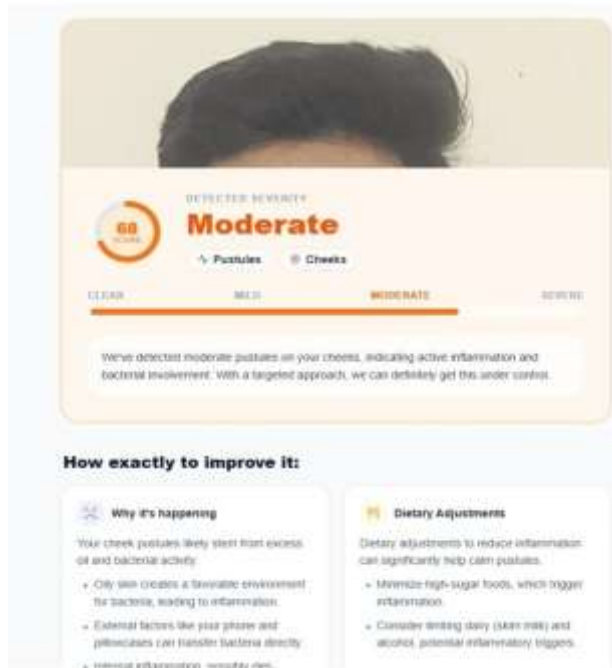


Figure 4.1: Acne Severity Prediction Result

The ResNet-18 model successfully classifies acne severity into four categories: Clear, Mild, Moderate, and Severe. The model demonstrates strong performance in distinguishing clear and severe acne cases due to their visually distinct characteristics.

However, mild and moderate acne cases occasionally show overlapping visual patterns, leading to minor classification ambiguity. The integration of the CMCF mechanism improves prediction stability by incorporating lifestyle risk information along with visual evidence.

D. Facial Region Analysis Results

The MediaPipe Face Mesh module successfully detects facial landmarks and segments the face into multiple regions such as forehead, cheeks, nose, chin, and jawline.

The analysis indicates that:

- Cheeks and jawline are the most affected regions
- Forehead shows moderate acne concentration
- Region-wise mapping improves prediction explainability

This approach enhances user understanding by showing the exact facial areas affected by acne rather than providing only a severity label.



Figure 4.2: Region-wise Acne Distribution

E. Lifestyle Risk Prediction Results

The Random Forest model classifies users into Low, Medium, and High acne risk categories using lifestyle-related inputs such as diet, stress, hydration, sleep quality, and physical activity.



Figure 4.3: Lifestyle Risk Prediction Result

Key observations include:

- Diet and stress are the most influential factors affecting acne risk
- Sleep quality and hydration also contribute significantly
- Behavioral information improves contextual understanding of acne severity

The results validate the importance of integrating lifestyle analysis into dermatological prediction systems.

F. Overall System Performance and Limitations

The system demonstrates stable performance with an average response time of approximately 14 seconds, while report generation requires an additional 2–3 seconds.

Key Strengths:

- Improved prediction reliability through fusion
- Better explainability using region-wise analysis
- Personalized recommendation generation
- User-friendly interface and workflow

Limitations Observed:

- Sensitivity to lighting conditions
- Reduced redness detection accuracy for darker skin tones
- Dependence on user-provided lifestyle inputs
- Limited dataset diversity

Despite these limitations, the proposed multi-modal framework performs effectively and provides a more holistic acne analysis compared to traditional single-modal systems.

5. ADVANTAGES, LIMITATIONS, AND APPLICATIONS

Advantages of Proposed System:

- Combines facial image analysis and lifestyle data for accurate acne prediction
- Provides region-wise facial acne analysis using MediaPipe
- Generates personalized skincare and diet recommendations
- Accessible through a responsive web-based interface

Limitations:

- Performance depends on image quality and lighting conditions
- Accuracy may reduce for darker skin tones due to redness-based detection
- System is not a replacement for professional dermatological diagnosis

Applications:

- Personal skincare assistant
- Tele-dermatology systems
- Healthcare monitoring applications
- Fitness and wellness platforms
- Preventive healthcare systems

CONCLUSION

This paper presents a multi-modal dermatological support system for acne severity prediction and personalized skincare recommendations. The system integrates ResNet-18 for facial image analysis, MediaPipe Face Mesh for region-wise detection, and a Random Forest model for lifestyle-based risk prediction.

A Cross-Modal Contextual Fusion mechanism combines visual and behavioral data to improve prediction accuracy, especially in borderline cases. The use of a Gemini-based

LLM enables personalized recommendations for skincare, diet, and lifestyle, enhancing user interaction and usability.

Experimental results indicate that the proposed system improves both prediction performance and interpretability compared to image-only approaches. Overall, the framework demonstrates that combining computer vision, machine learning, and lifestyle data can significantly enhance intelligent dermatological decision support systems.

REFERENCES

- [1] Deep Residual Learning for Image Recognition K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [2] Acne Severity Classification Using Transfer Learning and Deep Convolutional Neural Networks H. Liu, Z. Liu, and L. Wang, "Acne Severity Classification Using Transfer Learning and Deep Convolutional Neural Networks," *IEEE Access*, vol. 8, pp. 123456–123465, 2020.
- [3] Automated Detection of Acne Lesions Using Deep Convolutional Neural Networks R. Yamanaka, Y. Iwahori, and M. Kameda, "Automated Detection of Acne Lesions Using Deep Convolutional Neural Networks," *Dermatology Reports*, vol. 13, no. 2, pp. 45–52, 2021.
- [4] Multi-Class Acne Detection and Classification Using Deep Learning H. Zhao and Y. Sun, "Multi-Class Acne Detection and Classification Using Deep Learning," *Computer Methods and Programs in Biomedicine*, vol. 221, pp. 106–114, 2022.
- [5] Deep Learning-Based Facial Skin Analysis for Acne and Hyperpigmentation Detection Y. Zhang, L. Chen, and T. Wang, "Deep Learning-Based Facial Skin Analysis for Acne and Hyperpigmentation Detection," *Biomedical Signal Processing and Control*, vol. 65, pp. 102–110, 2021.
- [6] Machine Learning-Based Health Risk Prediction Using Lifestyle Data A. Patel and R. Shah, "Machine Learning-Based Health Risk Prediction Using Lifestyle Data," *International Journal of Healthcare Informatics*, vol. 9, no. 3, pp. 88–96, 2021.
- [7] Lifestyle Disease Prediction Using Machine Learning Algorithms M. Gupta and S. Arora, "Lifestyle Disease Prediction Using Machine Learning Algorithms," *Journal of Medical Systems*, vol. 44, no. 9, pp. 1–10, 2020.
- [8] DeepAc-Web: Web-Based Acne Severity Grading System Using Deep Learning S. Kim, J. Park, and H. Choi, "DeepAc-Web: Web-Based Acne Severity Grading System Using Deep Learning," *IEEE Access*, vol. 9, pp. 125678–125689, 2021.
- [9] Intelligent Facial Skin Analysis and Recommendation System Using Machine Learning P. Kumar and S. Verma, "Intelligent Facial Skin Analysis and Recommendation System Using Machine Learning," *International Journal of Artificial Intelligence Research*, vol. 6, no. 2, pp. 101–110, 2022.
- [10] U-Net: Convolutional Networks for Biomedical Image Segmentation O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.