Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 12, 2025

HYBRID ATTENTION-GUIDED CAPSULE FRAMEWORK FOR AUTOMATED FACIAL ACNE SEVERITY CLASSIFICATION

Matee ur Rasool*[1], Abdul Salam[2], Ayesha Nazir[3], Talha Mushtaq[4]
[1*,2,4] *Department of Electronic Engineering, The Islamia University of Bahawalpur, Pakistan*
[3] *Department sir Sadiq Institute and Technology, The Islamia University of Bahawalpur,Pakistan*
engrmateeurrasool@gmail.com[1], abdulsalam73059@gmail.com[2], ayesshanazir9930@gmail.com[3], engrtalha57db@gmail.com[4]

DOI: https://doi.org/10.5281/zenodo.18141774

Keywords
HAG-CAP, acne severity classification, attention-guided feature selection, Capsule Networks, CBAM, DCGAN augmentation, deep learning, yolov6, yolov7, yolov8

Article History
Received on 20 September 2025
Accepted on 02 October 2025
Published on 24 December 2025

Corresponding Author: *
Matee ur Rasool*

Abstract
Automatic and precise evaluation of the severity of facial acne is an essential requirement of evidence-based clinical decision-making and to alleviate patient distress. Automatic and precise evaluation of the severity of facial acne is an essential requirement of evidence-based clinical decision-making and to alleviate patient distress. However, the modern deep learning methodologies are often challenged by constrained, massively imbalanced sets of data, thus lacking the ability to condense strong, perspective-free representations, which would provide credible severity stratification across heterogeneous imaging of patients. This paper presents the Hybrid Attention-Guided and Capsule-Based Pose Encoding (HAG-CAP) framework to the intelligent classification of acnes severity. VGG16 is combined with a sophisticated CBAM in this novel architecture to give priority to the salient lesions areas. Primarily, a Capsule Network is included to attain viewpoint-invariant feature encoding thus improving lesion discrimination. To enable generalization, the issues of data scarcity and class imbalance were alleviated using DCGAN-generated synthetic samples, thus increasing the number of annotated Roboflow samples to 2,500. The HAGCAP approach demonstrates better performance, which was confirmed by a thorough comparative analysis with the YOLOv6, YOLOv7 and YOLOv8 models. The overall accuracy was 99% with a precision and recall of 100%, and a strong kappa rate of 97.87%. The high reliability of the system justifies the fact that it can be implemented in light mobile health applications where scalable and automated dermatological decision support may be competently facilitated.

1. INTRODUCTION

Acne Clinically known as Acnes vulgaris, is a very common chronic inflammatory disease of the pilosebaceous unit. It is a spectrum of lesions of dermatologic disease, with comedones, inflammatory papules, and pustules, which usually appear on the face, and neck, as well as the back. The condition is mainly linked to the hormonal changes that are typical of adolescence, and it is seen in about 85 percent of people between the ages of 12 and 24. Acne vulgaris, often persists into adulthood, but the impact of acne vulgaris is not uniform as the conditions continue to have their effects well into the third and fourth decades of the lives of affected patients. This disorder is

one of ten most common illnesses among global populations with an approximate prevalence of 9.4 percent in the entire world. Being extremely prevalent and visible, this condition carries a significant burden, thus, generating significant psychosocial implications and negatively affecting the quality of life. This requires correct and prompt clinical management using sophisticated reliable methods.

Healthcare continues to evolve with the integration of artificial intelligence and mobile technologies, and dermatology is among the fields experiencing substantial impact. Acne vulgaris is now among the most prevalent skin diseases in the globe with individuals of all ages suffering the effects of the disease. Besides the evident cutaneous alterations such as scarring, acne has psychological pain (anxiety, depression, and lack of self-esteem). These psychosocial and clinical outcomes point to the need to identify it early and make sure that the severity is stratified appropriately so that the intervention would be timely and survive. Traditionally, acne evaluation has depended on manual inspection by dermatologists, involving visual examination and tactile assessment of the skin. While effective, these methods can be time-consuming, subjective, and inconsistent across practitioners [1]. The above obstacles result in inequities in the provision of dermatology services, particularly in under-serviced areas. With the current and increasing acne rates in the international market, solutions of AI-based diagnosis based on scalability, cost-efficiency, and reliability are in high demand [2].

The recent developments of deep learning have changed the way medical images are interpreted. The violation of lesion classification and detection in various dermatological tasks have been demonstrated by Convolutional Neural Networks (CNNs) and the YOLO (You Only Look Once) series of object detectors. However, there are still some vital challenges. First, the visual resemblance of types of lesions makes it difficult to distinguish the lesions. Second, the small size of data sets and the large imbalance in classes decrease the robustness and generalizability of models. Third, most of the current solutions use one architecture, e.g., standard CNNs or single You Only Look Only variants, without using mechanisms to promote feature discrimination or alleviate data constraints [3], [4].

In an attempt to resolve these challenges, this paper proposes a deep learning hybrid architecture in acnes severity prediction. The proposed model does not rely on

solitary CNN or YOLO models or models but instead, it entails using VGG16 alongside a Convolutional Block Attention Module (CBAM) and Capsule Networks (CapsNet), which is a combination of convolutional feature extraction and attention-based refinement and capsule-level hierarchical encoding. Deep Convolutional Generative Adversarial Networks (DCGANs) are additionally employed to produce variety of synthetic training samples to counter the lack of datasets as well as imbalance. Such a single pipeline enhances the lesion-level representation and increases resistance to real-world variability.

Another characteristic feature of this study is that it is congruent with mobile health (mHealth) implementation [5]. Dermatologic imaging is the only field that can be acquired remotely with smartphone devices and can be used to conduct tele-dermatology and evaluate the patient. This study contributes to the viability of AI-based acnes diagnosis in remote triage, telemedicine and self-monitoring systems by optimizing the model to achieve a balance between inference accuracy and computational efficiency [6], [7]. There is a possibility that the inclusion of automated severity grading into mobile interfaces will decrease access disparities and enable people to take charge of their skin condition.

The primary contributions and novelty of this work are summarized as follows:

- Development of a hybrid VGG16–CBAM–CapsNet architecture designed to enhance feature extraction, attention-driven refinement, and pose-aware representation for acne severity classification.
- Integration of DCGAN-based augmentation to counter dataset imbalance and improve minority-class representation without overfitting.
- Comparative evaluation against YOLOv6, YOLOv7, and YOLOv8 architectures to demonstrate performance gains over conventional single-model approaches.
- Design consideration for deployment in mobile Health environments, highlighting feasibility for smartphone-based inference and clinical decision support.

The paper is organized as: Section 2 presents the literature review. Section 3 describes the proposed methodology. Section 4 discusses experimental outcomes and analysis. Section 5 concludes the study and outlines its broader implications.

2. Related Work

Acne vulgaris is one of the common dermatological diseases with significant impact on the psychosocial and somatic health of the patients. Traditionally, the severity of the acne has been determined based on the manual examination of the dermatologists, which may be both subjective and time consuming on its part. Considering the need to have a system that is fast, scalable, and precise, deep-learning (DL) methods have emerged as an impressive tool to the automated classification of acnes severity. Yadav et al. (2022) added the Convolutional Block Attention Module (CBAM) to a convolutional neural network to detect acnes, and the results show a higher accuracy in the localization of lesions. However, exact definitions of acne severity are still challenging, and it can be attributed to the complexity and overlapping nature of the lesions.

H. Li [3] proposes the system of lesion detection by Faster R-CNN in the first stage and overall severity grading by Light-GBM in the second stage, which they call Acnes-Det. Even though mean accuracy was 0.85, mAP value of 0.54 was a weak localization of lesions, even with a dataset of 1,572 annotated smartphone images. Applied AIDDA to EfficientNet-b4 with 95.8 and generalizing over more than 89 percent accuracy and effectiveness of several inflammatory dermatoses, respectively. Although it was clinically relevant, it was not aimed at the acne specific severity [11], Faster R-CNN and R-FCN on the task of acnes detection, with mAP of 28.3, highlighting the possibility of automation, but the lack of performance to deploy it suggested a modified CNN using an improved Leaky ReLU activation function that reported 97.54 percent accuracy and performed better than SVM and baseline CNNs, however, due to the limited dataset, there is concern about the external validity of the results. In a study, Liao et al. [12], integrated YOLOv3, YOLOv4, and Mask R-CNN models to improve face skin symptom detection accuracy, achieving up to 60.38% mAP and elaborating improved robustness against image noise. Baharul et al. assembled a dataset of 420 dermatologist-labeled images in seven categories and used preprocessing methods to increase the random accessibility of the dataset, which is training convergence, but low scalability because of the limited dataset.

YOLOv5 used to classify acne into four severity levels with an accuracy of 96.45%, true positive of 93.59 and recall of 94.73. Nevertheless, the value of mAP was

relatively low (37.97 when using single-class and 26.50 when using multi-class), indicating that the multi-class stratification is hard. Filtered 8,400 images of acne by re-annotating 12% of the Flickr-Faces-HQ dataset and using segmentation-based filtering with CelebAMask-HQ, which has an F1 score of 60.84% and moderate generalizability [13]. Produced a mobile-oriented YOLOv4 model that has 91.25 percent accuracy and shows that it can be used in real-time to identify acne but not with the same accuracy across all types of lesions. Adopted by, YOLOv8 was reported to be used in the detection of psoriasis with an accuracy of 76.19% when using the Extra Large variant, confirming the presence of dermatological relevance but not a specificity to accesses [14] contrasted between traditional and deep models among which LR, SVM, RF, Inception V3, VGG16 and VGG19 were compared, and the best models were found to be Inception V3 with logistic regression with 99.5% accuracy but depending on the dataset [15].

Other cross-domain applications, like YOLOv3-CNN hybrid by Hasna Fadhilah et al. to diagnose skin cancer (96% static and 80% real-time accuracy), and the YOLO food classification (Kusuma and Soewito) (also exploit the adaptability) additionally show the importance of dataset design and task-specific optimization [16],[17]. Following table 1 summarizes the various studies and their limitation.

Researchers working on the analytical evaluation of facial acne often utilize either traditional CNN paradigms, YOLO-based object detectors or linear, single-stream architecture designs. Nevertheless, such methodologies lack strong attention mechanisms and, therefore, do not capture viewpoint-invariant features. Such methodologies often fail when the data is scarce, the classes are strongly imbalanced, the grading criterion is subjective, and the features are discriminatory, especially when acnes phenotypes resemble each other, or respond to changes in lighting, pose, and skin pigmentation. Also, the current methods emphasize detection more than the strict classification of severity and also limited by small and single-center datasets. Despite this, no approach has so far been able to effectively integrate attention-directed feature selection with capsule-based pose encoding a synthesis which is intractable without fine-grained severity analysis. To overcome these shortcomings, HAGCAP approach proposes incorporation of CBAM based hybrid attention, VGG16 hierarchical encoding and Capsule Networks in a manner that it is effective to extract spatial pose relations. To reduce

dataset limitations, utilized DCGAN-generated samples which not only increase the dataset but also diversify the dataset. The hybrid model described in this section contributes significantly to lesion discrimination and stratification of severity, which translates to a dependable diagnosis across the diverse situations. With the highest possible accuracy of 99%, 100% precision and kappa of 97.87%, the proposed methodology provides a more solid and generalizable model of acne severity classification than existing strategies possess.

The paper is organized as follows: the first part represents the workflow of the presented model, including data preparation, model training, evaluation, and deployment. Next, the following sections discuss the selection of data, the methods of preprocessing information, the selection of a network architecture, and training settings, hence, continuing the logical development of the discussion of a theoretical background and practical implementation.

Table 1. Tabular Comparison of Models for Acne/Skin Classification.

Paper Ref.	Technique	Applications	Result	Limitations
[1]	YOLOv4	Easy-to-use instruments for determining the kind of acne	YOLOv4 model achieved a remarkable 91.25% accuracy.	Combining YOLO with any existing Deep Learning method might improve identification performance.
[2]	CNN-based model (R-ASE)	Acne may vary from person to person, which can make precise diagnosis and evaluation difficult.	The test set showed that R-ASE outperformed the human dermatologists who participated in the research, with an RMSE of 0.482.	Using selfies as training photos and image meta data to supplement the limited number of labelled photographs may improve the model.
[3]	Light-based model (Canada)	Traditional acne diagnosis and grading are challenging due to their reliance on time-consuming, subjective visual inspection by dermatologist	The model achieved an accuracy rate of 85% and a mean average precision score of 54%.	Accuracy of severity estimation could be improved, and the dataset size was limited, potentially affecting model reliability.
[5]	Acne-ResNet model-based smartphone apps	Due to modest symptom fluctuations, conventional techniques of assessing the severity of face acne are problematic. Accurate diagnosis and treatment of common	A 94.56% success rate was attained by the Acne-ResNet based mobile application.	The dataset is limited in scope since it was only collected at one hospital, and the software isn't very user-friendly because users have to manually draw circles around acne lesions and submit photos.

		skin illnesses are complicated.		
[7]	CNN with residual architecture called AcneNet	Due to the similarities in appearance amongst acne types, physicians' traditional methods of evaluating the condition may be subjective and prone to mistake.	Acne Net's accuracy was over 94% across all categories, with a 99% success rate for one specific kind of acne.	Model architecture could not account for every kind of acne because of its design, and that it's dependent on the baseline dataset, which might limit the accuracy of predictions.
[9]	YOLOv8	Building a reliable and effective method for psoriasis picture classification	The model achieved 76.19% accuracy.	The current focus is on psoriasis, but future efforts will seek to expand to other dermatological illnesses, identify them in real-time, and work in tandem with dermatology specialists to increase accuracy.
[6]	EfficientNetB4 and CNN-based algorithm called AIDDA	The way things are now, dermatologists may have to use their subjective, time-consuming, and error-prone visual assessment skills.	EfficientNetB4 and CNN-based algorithms achieved diagnostic accuracy, sensitivity and specificity of 95.80%, 94.40%, and 97.20%, respectively.	Less experienced dermatologists may make the mistake of misdiagnosing psoriasis, eczema, or atopic dermatitis because of how similar the symptoms are.

3. Material and Methods

This paper proposes an end-to-end artificial intelligence pipeline that will be used to automatically classify the severity of facial acne. The approach starts with the careful dataset preparation, including the use of the Roboflow repository, and continues with the scrupulous image augmentation, annotation, and lesion location. The model proposed uses the latest object detection architectures, such as YOLOv6, YOLOv7, and YOLOv8, and a library of CNN architectures in order to provide precise acne severity classification. Standard metrics are used to measure model performance like confusion matrix, F1 score and precision-recall curves. The finalized model is finally implemented on a mobile app, which can provide the possibility to detect acne and evaluate its severity in real time. In the given Figure 1 - HAG-CAP Frame work is portrayed.

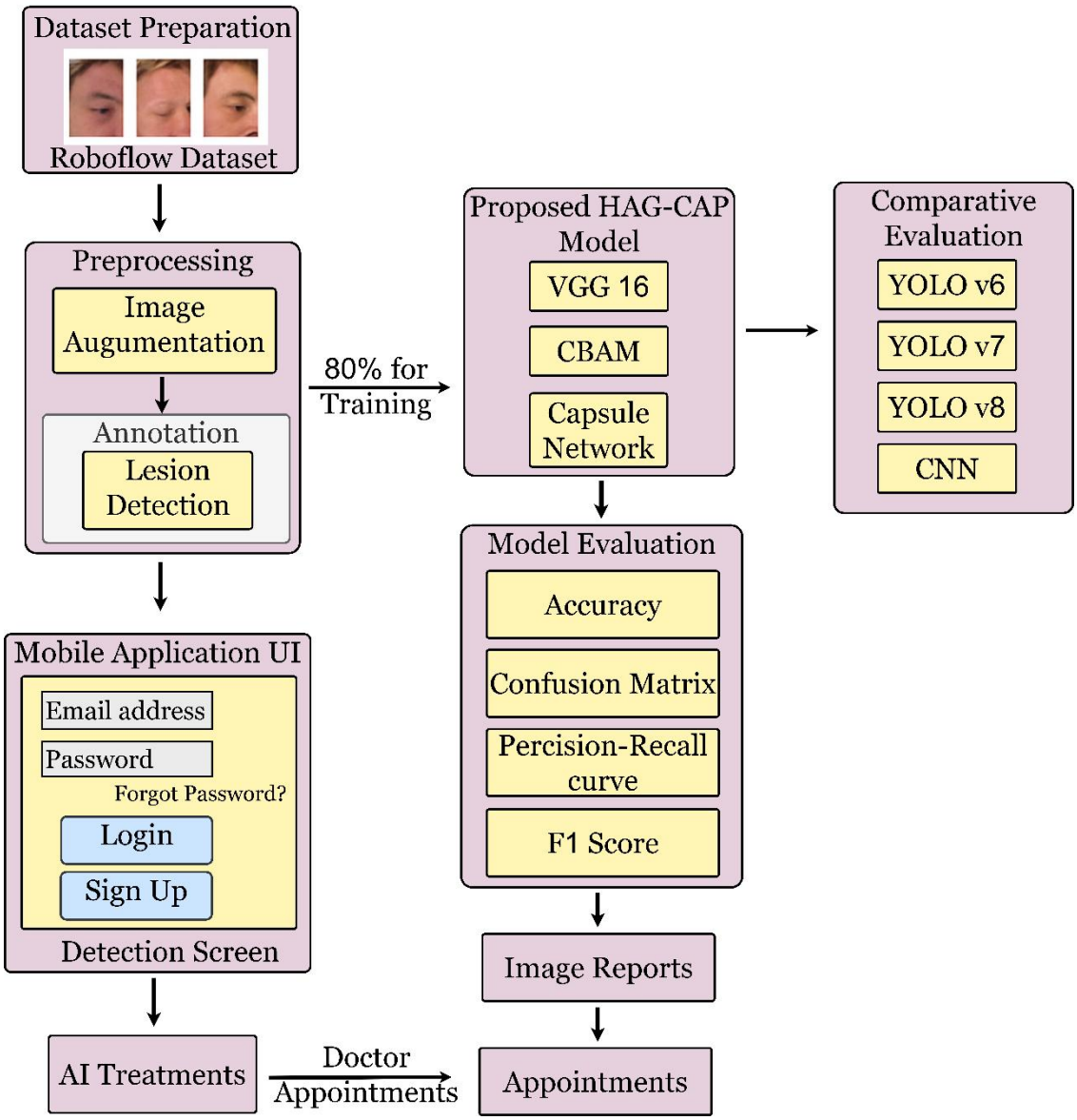


Figure 1. HAG-CAP Frame work.

3.1. Proposed Model

This paper identifies a complete AI pipeline to the automated detection of facial acne severity, including steps such as dataset preparation to mobile implementation. The first operations are defined by the Roboflow dataset, where the data is subjected to a strict preprocessing step, state-of-the-art image augmentation, scrupulous annotation, and lesion detection mechanisms, thus improving the quality and the heterogeneity of the data. After splitting the data into portions, the models of classification are used, such as the YOLOv6, YOLOv7, YOLOv8, and different CNN architecture, to identify acne pattern and categorize its

level of severity. Their performance is strictly evaluated by the measures of confusion matrix, F1-score, precision recall curve, and recall curve hence are reliable and strong.

At the end of training, the model is integrated into a mobile application interface and thus acnesiform eruptions can be detected and staged in real-time. The app allows users to post their pictures where the AI engine analyses them, and generates detailed diagnostic feedback and personalized treatment advice. An appointment module is also included in the system, thus promoting a smooth integration between the users and dermatologists to consult them later as needed. This combined approach supports automated examination and follow-up of the clinic, which makes the platform especially susceptible to scalable mobile health software.

3.2 Datasets

The deep learning models also rely on datasets to provide the visual variability required to train the models in pattern recognition and feature learning. Available public dermatological data on websites like Roboflow, Kaggle, and Medical ImageNet have been extensively used in the literature. In the present research, the annotated image data was acquired in the Roboflow repository, where samples of various skin diseases were labeled, including acne. It originally had 1,981 images that were augmented to 2,500. The designated density of the dataset allowed performing successful supervision in the process of training, which allowed mapping the types of lesions and the severity segments correctly. It has also made sure that the model was subjected to real-life patterns in dermatological settings and could be optimized.

Table 2. Disease Dataset Descriptions.

S. No	Disease	Description
1	Papules	690 images
2	Dark spots	651 images
3	Pustules	389 images
4	White Head	337 images
5	Black Heads	240 images
6	Nodules	163 images
7	Normal	30 images
8	Total	2500 images

3.2.1 Dataset Pre-Processing

Dataset preprocessing is applied to ensure data quality, consistency, and diversity before model training. Augmentation techniques and systematic dataset partitioning to improve generalization, mitigate overfitting, and maintain balanced class representation for reliable acne lesion detection.

3.2.1.1 Dataset Augmentation

To enhance the model against rare and hidden visual variability, data augmentation was taken as a fundamental enhancement measure. This is done by using a set of targeted transformations on existing images that were intended to artificially enlarge the diversity of the datasets. Random flipping, rotation, translation, scaling, and contrast enhancement are the techniques used to create alternate visual representations of the patterns of the same lesions. The network was trained by exposing the model to these variations and was able to learn to generalize to lighting variations, orientations and spatial distortions. Consequently, the initial set of 1,981 images was increased to 2,500 samples, so that the coverage by classes would be better. This not only helped lower the chance of overfitting but also led to increased adaptability to the real-world conditions in which there exists a more heterogeneous distribution of the skin tone, the lesion location and the quality of images.

3.2.1.2 Dataset Splitting

To achieve the uniformity of evaluation and proper training functionality, two configurations of a dataset split are applied. The samples used in the main division are divided into 80 percent training, 10 percent validation, and 10 percent validation and 10 percent testing. This proportional allocation facilitated maximized learning whilst having a representative sample to do unbiased evaluation.

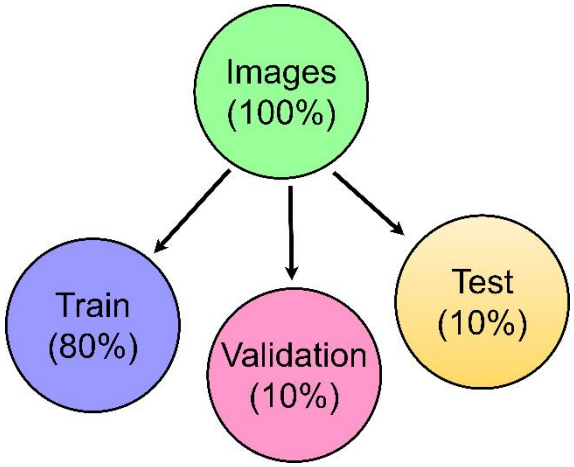


Figure 2. Data partition overview.

3.3 Detection Models and Techniques Employed

This section provides an exhaustive description of detection models and calculational processes adopted in the current investigation to specifically identify and question acne lesions. The inclusion of advanced deep-learning designs and the optimization of training schemes, in its turn, prominently enhance the detection of lesions, the derivation of features, and the severity classification. The current discussion focuses on YOLO-based models, laying out the architectural virtues and elaborating on the justification rationale behind the adoption of this method to enable the support of reliable, real-time dermatological evaluation. Datasets are a foundation to increase the reliability and generalizability of deep learning systems; they contain the visual grounding on which patterns of space, hierarchical dependencies, and lesion specific features can be identified. These principles are essential in proper categorization and identification of dermatological disorders, and by grant, they are fundamental to the creation of a stringent system of acne severity estimation.

3.3.1 YOLOv6

The trade-off between speed and accuracy in YOLO architectures has contributed to their prominent adoption in real-time visual detection. YOLOv6 introduces several architectural enhancements over earlier versions, including a clear separation between the classification and bounding box regression heads. This untangled structure enables all the tasks to be optimized independently, increasing accuracy when localizing lesions [18].

YOLOv6 employs stacked convolutional layers to extract multi-level feature representations, detailed and semantic information. The Rep-PAN neck aggregates these multi-scale features to enable precise detection of lesions with different dimensions. Further improvements have been made as anchor-free training, SimOTA label assignment and SIoU loss function to enhance bounding box regression. All of these design decisions minimize computational costs and increase the detection performance [19].

With its hardware-efficient architecture, YOLOv6 achieves effective real-time inference even under limited computational resources. Nonetheless, its architectural efficiency does not have extensive support regarding fine-grained and multi-class lesion classification of dermatological lesions. This requires comparative evaluation with more advanced versions to determine its suitability for acne severity analysis [20]. The model also designs neck

and backbone through Efficient backbone further optimizing its performance to be lightweight. overall architecture of YOLOv6, which is composed of three main components: the backbone, the feature pyramid network (FPN), and the detection heads. The backbone is responsible for extracting multi-scale feature representations from the input image, while the FPN aggregates these features across different layers to improve detection of objects of varying sizes. The detection heads then predict class probabilities, bounding box coordinates, and abjectness' scores using specialized convolutional layers and corresponding YOLO loss functions. This modular design enhances both accuracy and real-time performance.

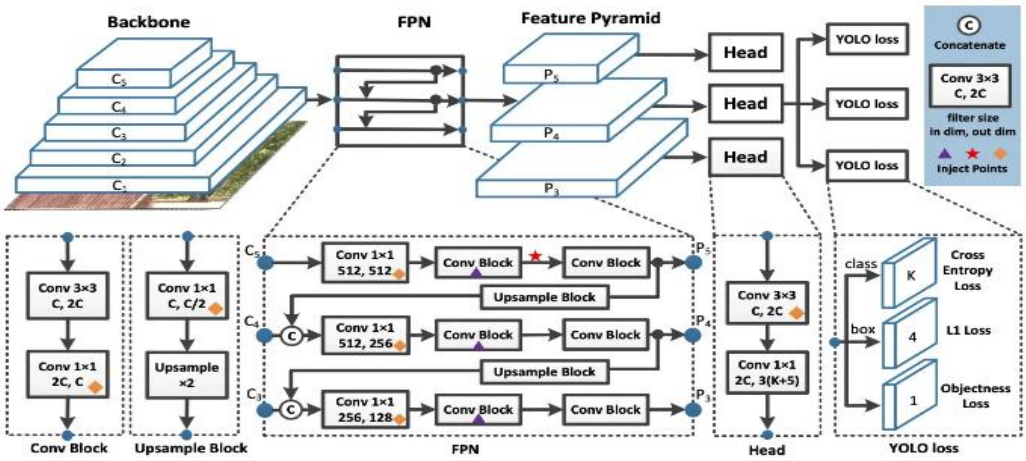


Figure 3. Architecture of YOLOv6 illustrated by Krishnapriya [1].

3.3.2 YOLOv7

YOLOv7 is a significant improvement in real-time object detection, which is more accurate and faster to infer than the previous versions. It has architectural stability and a robust convergence behavior unlike many state-of-the-art detectors and is able to achieve competitive performance without using external datasets or without auxiliary training pipelines. Another important improvement of YOLOv7 is that it focuses on optimization of training instead of radical redesigning of the structure. It presents a trainable bag-of-freebies, a set of augmentation and regularization techniques that enhance the accuracy of detection with no augmentation or regularization cost. These mechanisms not only significantly lower the computation requirements (almost 50 percent) and shrink the model parameters

(more than 40 percent) but do not improve the performance.

YOLOv7 is especially beneficial in medical imaging on a time-limited basis due to its combination of accuracy, speed, and efficiency. With its lower resource usage due to the nature of dermatological tasks, like the detection of acne lesions, it can be run on GPUs, edge devices, and even on a mobile platform [21]. This qualifies it as a good contender in scalable mHealth solutions in which the speed of the inferences and the reliability of the diagnostic results are critical.

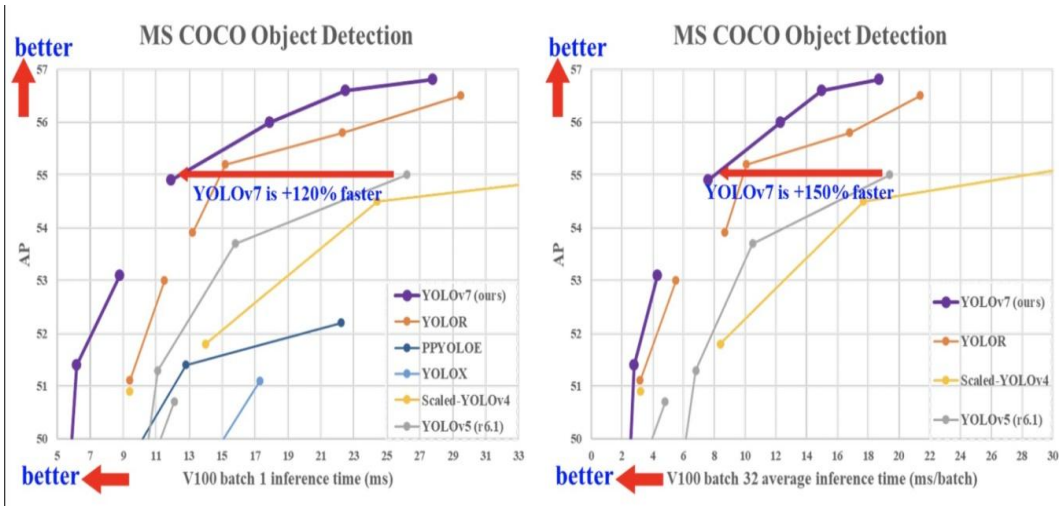


Figure 4. Graphs of YOLOv7.

3.3.3 YOLOv8

The latest addition to the Ultralytics family is YOLOv8, which introduces an advanced architecture and enhanced deployment options to the YOLO framework. It is built as a continuation of YOLOv5, and is focused on long-term maintainability and modularity, to allow continuous improvement by the community. YOLOv8 has a simplified computational representation, compared to previous versions, that minimizes the architectural complexity and enhances the detection accuracy and generalization on a variety of imaging tasks.

One of the most salient methodological improvements represented by YOLOv8 is its non-reliance on the paradigm involving

the use of anchors by replacing the need to use anchor boxes, the model can attain a significant improvement in the localization efficiency; without the predefined anchor boxes, the model is able to obtain considerable upgrading in localization performance through it. Besides, that decoupling the classification and regression heads, the model attains a more specialized representation of each sub-task. Such high-quality data-augmentation algorithms which are applied and a significant enhancement to its ability to handle variability of the domain and to address imbalance of the dataset. The model's modular design makes it highly scalable, supporting deployment on both high-performance GPUs and edge devices, and ensuring compatibility with diverse hardware environments.

Despite the fact that formalized in a peer-reviewed publication, YOLOv8 has the advantage of a rich open-source documentation and community improvement. Its performance in terms of efficiency, scalability and accuracy of detection makes it especially applicable to dermatological use cases like detecting acne lesions [22],[23]. Lightweight computation and real-time inference abilities of the model make it a good candidate in mHealth integration, where trustworthy classification has to run on limited processing means. Figure 4 shows the YOLOv8 architecture adapted from Sapkota et al. [2] consisting of a backbone, a neck (feature fusion layers), and multiple detection heads. The backbone is responsible for extracting hierarchical feature maps at different scales using components like Conv, C2f blocks, and the SPPF module. These characteristics are then passed on to the neck where up sampling, concatenation and further C2f layers provide an opportunity to aggregate features multi-scale. Lastly, the head detects objects at various resolutions, resulting in the areas of bounding box regression, class prediction, and the abjectness scores. The design is accurate, fast and scalable among model variations.

With the consideration of the influential and powerful changes to the model, which is self-evident. A major advancement in YOLOv8 is its box-free design. Instead of using predefined boxes to predict object positions, it directly predicts the centers of objects. This puts an end. Traditional YOLO models relied on anchor boxes that are predefined rectangular templates used to estimate object locations and sizes. Although effective for standard datasets but these anchors mostly failed to align with objects in custom datasets, reducing detection accuracy. YOLOv8 gets around this limitation by directly anticipating the

object centers and dimensions in actual coordinates, eliminating dependency on predefined shapes. This anchor-free system increases adaptability, simplifies training, and improves overall detection precision. As a bonus, using anchor-free detection streamlines the Non-Maximum Suppression (NMS) post-processing phase that filters detections following inference by reducing the number of box predictions. In Figure 4, we can see the YOLOv8 architecture. Like YOLOv5, YOLOv8 relies heavily on the training schedule in addition to the architectural improvements. The use of online image augmentation by YOLOv8 during training is noteworthy. To improve its capacity to learn objects in novel settings, with partial occlusion, and against different backdrops, the model is subjected to slightly altered copies of the training pictures at the end of each session. One method that is utilized to increase the model's flexibility is mosaic augmentation, which is sewing four photos together [24].

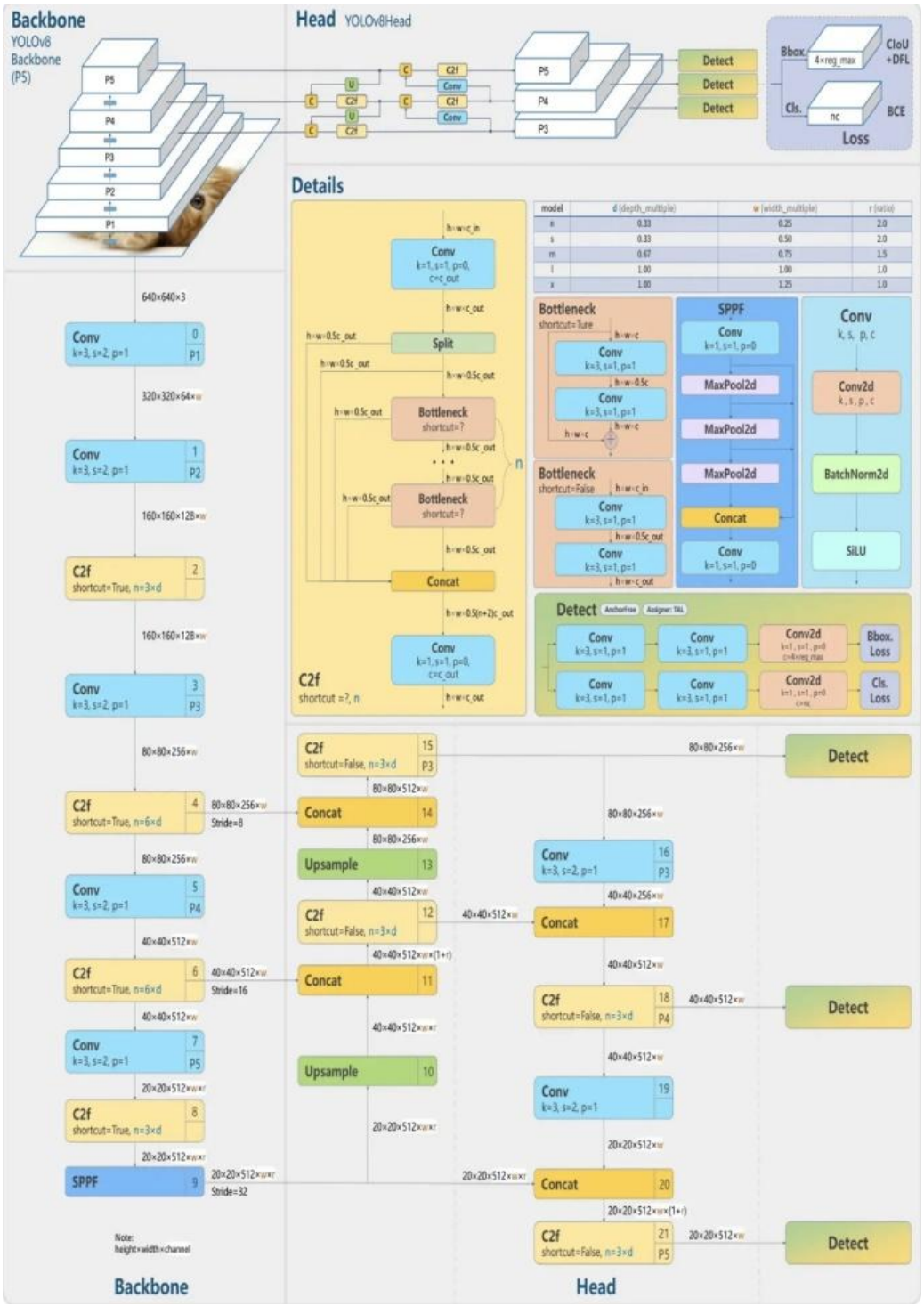


Figure 5. Architecture of YOLOv8 [2].

3.4 CNN Architecture and Supporting Tools

The CNNs are the staple of the procedure of feature extraction in the proposed model. Based on the human visual apparatus, this network takes advantage of the various linked layers of neurons to identify spatial hierarchies in image data. The architecture

consists of a series of convolutional steps that are followed by pooling processes making the model afford the ability to capture the finer details of acnes lesions. Conventionalized to the goal of classification, the CNN utilizes the activation of Rectified Linear Unit (ReLU) together with a Softmax layer to produce multi-class probability distributions. In the experimental paradigm, Python based tools were used to train the model, in particular, Google Colab, which provided easy access to the GPU acceleration; simultaneously, a mobile app interface was developed in Flutter, and Firebase supported secure data storage and real-time synchronization.

3.4.1 Convolution Neural Network

Convolutional Neural Networks (CNNs) are inspired by the structure and function of the human visual system. CNNs mimic its architecture by using several layers of linked artificial neurons. These cells analyze the data received by them, e.g., pixels of an image by means of mathematical operations to obtain pertinent features. The output of each layer is known as an activation map and highlights various features of the image. Through the cascading of these layers, convolutional neural networks (CNNs) columns offer accurate prediction and pattern learning. The forward propagation is computed as the equation below:

$$Z=X*F \tag{1}$$

An asterisk (*) is a common mathematical symbol for convolution. Assuming that X is the input picture and F is the filter

Equation for linear transformation of data is:

$$Z = WT.X + b \tag{2}$$

here, X is the input, W is weight, and b is bias is a constant.

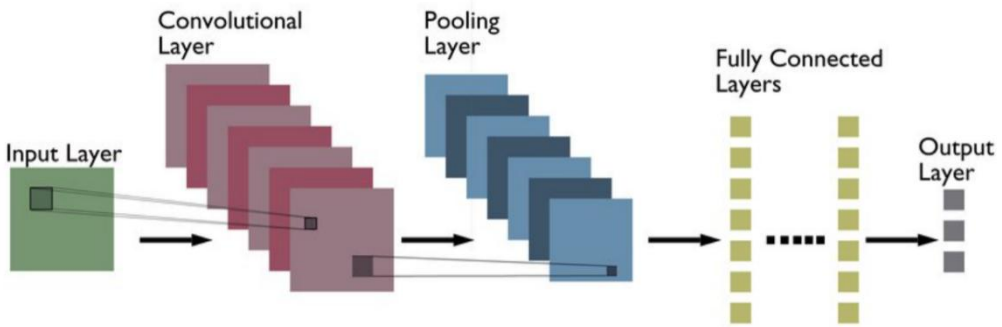


Figure 6. Architecture of CNN Model by Hayat et al. [3].

Figure 6 presented by Hayat [3] depicts the fundamental architecture of a Convolutional Neural Network (CNN), illustrating the sequential flow of data from raw input to final prediction. This starts at the input layer whereby the image or data sample is given. It then goes through convolutional layers, which learn to get local spatial features through learnable filters. They are then followed by pooling layers, which decrease the spatial dimension and preserve the valuable information and assist in controlling overfitting and enhancing computational efficiency. The features are extracted and flattened, after which high-level reasoning and classification take place in the fully connected layers. Lastly, the prediction is made by the output when the features are learned.

3.4.2 Tools

The analysis of the proposed techniques was conducted in Python using Google Colab, while the mobile application was developed using Flutter with Firebase serving as the backend. These tools were selected to ensure efficient experimentation, scalability, and deployment across both research and mobile health contexts.

3.4.2.1 Google Colab

For this study experimental research utilized the google colab notebook which is free of source with free limitations of GPU and TPU. Google Colab is a free, cloud-based environment that extends the functionality of Jupyter Notebook. It does not need any local installation and offers performance to GPU and TPU resources that are essential in training deep learning models using the power of nanoparticle engines. Colab, like Google Docs, is also compatible with carrying out real-time collaboration and it supports the use of major machine learning libraries (TensorFlow, PyTorch and Keras). These characteristics render it especially appropriate to reproducibility and trial-and-error experimentation in research.

3.4.2.2 Flutter

Flutter is a user interface (UI) system written in cross-platform language that was created by Google to enable programmers to code an application using a single codebase, which operates on both iOS and Android. It exposes the native device functionality such as cameras, which are crucial in obtaining dermatological pictures. As it ensures similarity in the user interface across platforms and offers to support high-performance rendering, Flutter is a high-performance choice to develop healthcare applications that should be responsive and scalable. This tool help the civilians to detect dermatological picture in real time to detect disease which help finding the exact

lesion area to help more proficient treatment outcomes of patient.

3.4.2.3 Firebase

Firebase, also developed by Google, is a cloud-based backend platform offering services such as authentication, real-time databases, cloud storage, and hosting. Its support to Flutter makes it easy to develop mobile applications, as it allows handling securely the mobile data, users, and scalable storage options. In this project, the Firebase was utilized to provide user authentication, image storage in a secure way, and real-time updates to ensure that the mobile app would perform well in many-use cases.

4. Results and Discussion

This section presents the experimental workflow and analytical procedures applied to the Roboflow dataset. Each model was evaluated using widely recognized diagnostic metrics, including confusion matrices, F1 curves, precision–recall curves, and recall curves. The section concludes with a comparative interpretation of model performance based on these metrics.

4.1 First Experimental Results

In the first experiment, the data were split into three parts, of which 80% was to be used in training, 10 percent in testing and 10percent in validation. Facial Acnes Detection Dataset was marked and tested on four classifier types CNN, YOLOv8, YOLOv7 and YOLOv6. Confusion matrices, F1 score curves, recall curves, and precision recall curves were used to evaluate their performance to obtain the class-wise and the overall prediction reliability. while learning convergence and loss behavior were visualized in above figure YOLOv8, YOLOv7, YOLOv6, and CNN, respectively.

4.2 Optimizing Hyper-parameters

The hyper-parameter tuning was performed to enhance the generalization and maximize the classification. There was a grid search strategy which was used in order to test the combinations of batch size, learning rate and the number of epochs. The batch sizes were 32, and the learning rate was kept to 0.001 as it best fitted on models. Adam optimizer was chosen as it has the ability to adjust its learning rate that is best suited to classifying multiclass dermatological images. The convergence behavior of this tuning was optimum, and better detection accuracy was attained in all experimental models.

The experiments utilized the SoftMax layer in classifier with categorical cross-entropy loss, expressed as:

$$CategoricalCross - Entropy = -i \sum yi \log(pi) \quad (3)$$

where y_i signifies the actual probability distribution function across cases for the i^{th} instance and pi indicates the estimated probability distribution function across different classes for the i^{th} instance.

VGG16 is a CNN model with 16 layers, including 13 convolutional layers and 3 fully connected layers. It uses 3x3 convolutional layers with ReLU activation and 2x2 max-pooling layers to extract features. After flattening the feature maps, it passes through 3 fully connected layers with 4096 neurons each, followed by a final layer mapped to the number of classes using Softmax for multiclass classification.

4.3 Evaluation Metrics

To figure out the efficacy of the proposed hybrid model, a strong assessment framework was built around a confusion matrix as depicted in Figure 7. By this confusion matrix (CM), crucial performance indicators such as recall, F1-score, specificity, accuracy, precision, and kappa scores given below are calculated. The accuracy (%), which measures the model's overall correctness in classification, is calculated as:

$$CategoricalCross - Entropy = -i \sum yi \log(pi) \quad (4)$$

The above equation calculates the proportion of total correct predictions both positive and negative to the total number of predictions made. It is one of the most commonly used evaluation metrics for classification models.

The precision (%) of the model indicates how accurate it is at classifying positive instances, as computed by:

$$\Pr e\,cision = \frac{TP}{TP + FP} \quad (5)$$

Table 3. Comparison of Different Pre-Trained Models, Sota Models, Proposed Hybrid Model Without Augmentation and Various Augmentation Techniques (Batch Size=64, Learning Rate=0.001).

Acne dataset	Models	Epochs	Precision (%)	Recalling (%)	F1-score (%)	Accuracy (%)	Specificity (%)	Kappa score (%)
Un-augmentation	VGG16	100	72.00	78.00	73.00	62.00	99.08	41.02
	MobileNetV2	100	100	100	98.00	93.00	87.31	87.31
	InceptionV3	100	100	92.00	97.00	93.00	100	92.62

	ResNet50	100	100	88.00	93.00	55.00	100	15.01
	InceptionResNetV2	100	88.00	92.00	84.00	96.00	96.63	69.90
	ResNetCBAM [52]	15	97.00	97.00	93.00	93.00	97.53	91.57
	HAG-CAP VGG-16 (Proposed model)	15	**97.00**	**98.00**	**98.00**	**96.00**	**98.55**	**93.69**
Traditional augmentation	VGG16	100	72.00	78.00	70.00	66.00	100	38.42
	MobileNetV2	100	86.00	89.00	87.00	75.00	98.24	55.53
	InceptionV3	100	100	85.00	81.00	71.00	98.24	54.17
	ResNet50	39	49.00	100	66.00	69.00	0	45.92
	InceptionResNetV2	32	97.00	98.00	95.00	90.00	98.45	91.61
	ResNetCBAM [52]	15	97.00	98.00	95.00	90.00	98.52	91.61
	HAG-CAP VGG16 (Proposed model)	15	**100**	**98.00**	**97.00**	**97.00**	**100**	**94.71**
DCGAN augmentation	VGG16	5	100	98.00	88.00	68.00	95.67	70.00
	MobileNetV2	5	100	100	90.00	77.00	87.09	77.09
	InceptionV3	5	100	98.00	92.00	87.00	94.71	84.00
	ResNet50	5	100	98.00	66.00	49.00	0	45.92
	InceptionResNetV2	5	100	98.00	98.00	93.00	100	96.84
	ResNetCBAM [52]	5	100	98.00	97.00	94.00	100	96.84
	HAG-CAP VGG16 (Proposed model)	5	**100**	**100**	**99.00**	**100**	**100**	**97.87**

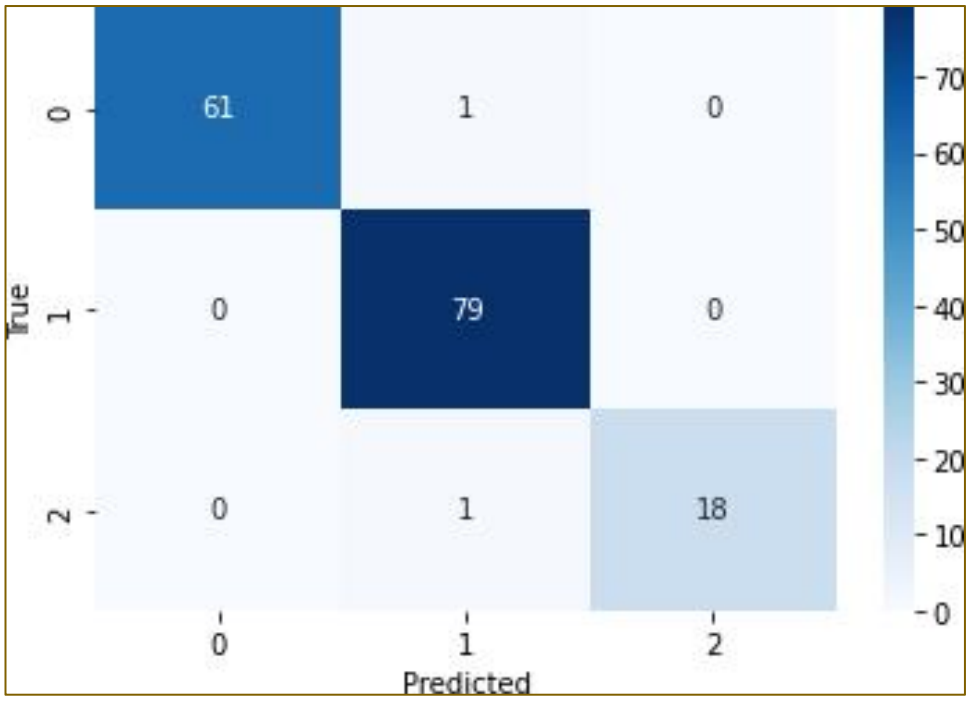


Figure 7. Confusion Matrix representation of severity level of classes 0,1 and 2 respectively.

Figure 7 shows the confusion matrix, which is employed in assessing the classification ability of the model about three classes (0, 1, and 2). The matrix will compare the real labels (rows) and the predicted labels (columns), and it will give a clear picture of accurately and incorrectly classified cases. The diagonal values indicate the figures of the correct prediction of each of the classes, whereas off-diagonal values denote misclassification. This representation is used to evaluate the accuracy of a model and determine the performance of that model by the classes and how the model can confuse one class with another.

Recall (%) measures the model's ability to accurately detect individuals with the condition and is given by:

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

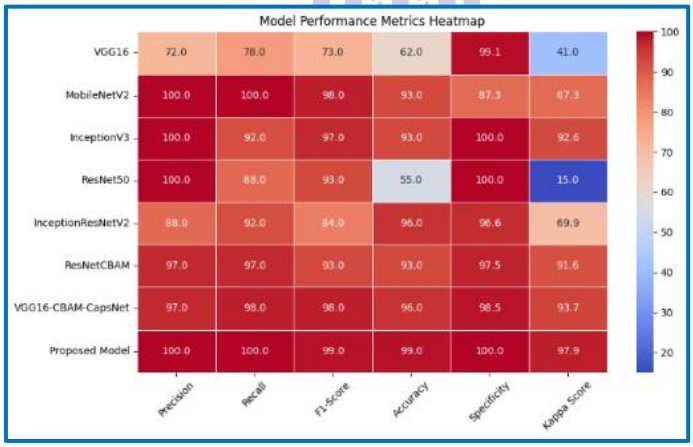
F1-score (%), a balanced metric that harmonizes recall and precision, is expressed as:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

Specificity (%) evaluates how well the model can identify individuals who are free of disease and is portrayed by:

$$Specficity = \frac{TN}{TN + FP} \quad (8)$$

Here, True Positive (TP) correctly identified positive samples, True Negative (TN) correctly identified negative samples, False Positive (FP) incorrectly identified positive samples, and False Negative (FN) incorrectly identified negative samples. To provide a clearer and more detailed comparison of the classification performance across different models, a heatmap visualization is presented below in Figure 8. The heatmap outlines the main performance indicators Precision, Recall, F1-Score, Accuracy, Specificity and Kappa-



Score of a variety of tested architectures, and pre-trained models such as VGG16 and MobileNetV2. Precision and

Recall measure the effectiveness of the models at identifying positive instances, both of which have MobileNetV2 scoring 100 percent. F1- Score (precision -recall ratio) of MobileNetV2 is 98.0 indicating a high level of harmonic accord. Accuracy, or the percentage of correct predictions in general, is 100 on proposed model and MobileNetV2. Measuring the correct classification of negative samples, specificity, is especially high in MobileNetV2, with values over 90. The Kappa statistic, which tests the consistency above chance, ranks both MobileNetV2 and the Proposed Model at the top layer where the results are 100 per cent. Through the application of color intensity to encode these measures, the heatmap highlights those models that showed a higher performance in a more pronounced way, namely those that are MobileNetV2, and it places those with relatively lower performance scores in a

demarcation area, e.g. VGG16.

Figure 8. Heatmap representation of experimental models.

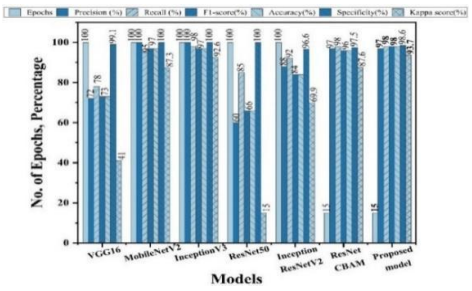
4.4 Proposed Model for Evaluating Performance with SOTA Methods

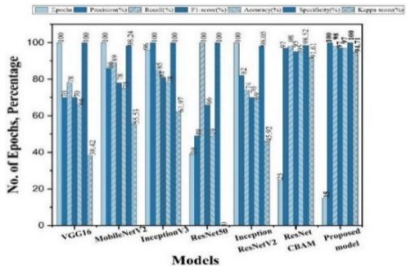
The presented subsection gives a comparative analysis between several pre-trained models, a state-of-the-art (SOTA) architecture, and the proposed hybrid framework. The comparison has been done on the original (not-augmented) dataset and augmented datasets. Two augmentation methods are adopted: (i) traditional methods, such as rotation, flipping and zooming, and (ii) state-of-the-art augmentation, using a Deep Convolutional Generative Adversarial Network (DCGAN).

Figure 9 shows the development of the acne dataset through the pipeline of the experiment. To evaluate, one split 80:20 was done and 801 images (80 percent) were trained and 198 images (20 percent) were tested. Both augmentation methods were used to increase the size of the training set by a factor of three, which allowed a stronger model training.

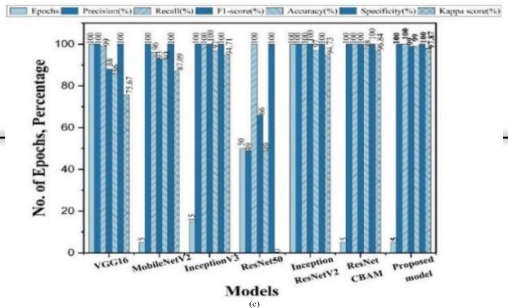
This design enabled comparative evaluation of pre-trained networks, the SOTA model, and the proposed hybrid model to be compared under the same circumstances, thus giving the proposed model a strict benchmark to evaluate classification performance [25],[26]. A more detailed analogy between the proposed hybrid model with the various pre-trained and SOTA models on various acnes datasets is presented in Table 4 which is also visually represented in the accompanying Figure 9. This paper examined three dataset settings, i.e., un-augmented and two augmentation strategies, to evaluate how the traditional methods influence the efficacy of models.

In the case of the un-augmented dataset, the proposed hybrid model classification accuracy (96%) is the highest compared to





(b)



(c)

that of all the other pre-trained and SOTA DL models [27]. The proposed hybrid model also attains impressive performance in terms of the other parameters, achieving higher recall (98%), F1-score (98%), and kappa score (93.69%) than did the other existing models. MobileNetV2 and InceptionV3 achieved higher precision (100%) and specificity (100%).From the above-mentioned Table 3, we can also observe that upon applying the conventionally augmented technique, the proposed model exhibits superior performance than the un-augmented technique and achieves higher precision (100%), kappa score (94.71%), specificity (100%), and accuracy (97%). Comparing the performance of the proposed hybrid model against pre-trained models in the conventional augmentation dataset, it out-performed and achieved the highest recall (98%), F1-score (97%), kappa score (94.71%), and accuracy (97%). Furthermore, we tested the efficacy of the proposed hybrid model on an augmented dataset generated using the DCGAN, and the evaluation metrics further improved to precision (100%), recall (100%), F1-score (99%), accuracy (99%), specificity (100%), and kappa score (97.87%).

Figure 9. Pictorial Representation of Various Models' Performance Comparisons on (a) Un-Augmented Datasets (b) Traditional Augmented Datasets and (c) DCGAN Augmented Datasets.

It is worthwhile that the proposed model bolsters the system's robustness, ensuring adaptability to diverse future test data scenarios. Additionally, the hybrid model consistently out- performs the five pre-trained and SOTA models across all the metrics of evaluation, both with and without augmentation [28].

Table 4. Performance Evaluation of Different Pre-Trained Models in Combination with the Proposed Model [3].

Hybrid models	Epochs	Accuracy (%)	Precision (%)	Recalling (%)	F1-score (%)	Specificity (%)	Kappa score (%)
HAG-CAP(VGG16+CBAM+CapsNet)	5	99.00	100	100	99.00	100	97.87
MobileNetV2+CBAM+CapsNet	5	97.00	100	99.00	98.00	100	95.74
InceptionV3+CBAM+CapsNet	20	97.00	100	100	100	100	95.78
InceptionResNetV2+CBAM+CapsNet	10	49.00	49.00	100	66.00	100	0

Table 5. The proposed HAG-CAP model integrates the VGG16 backbone, CBAM, and CapsNet components. The model is trained for five epochs with three batch sizes represented with accuracy metrics, including Training Accuracy (Ta), Validation Accuracy (Va) inline with Training Loss (Tl), and Validation Loss (Vl).

	Batch Size 32				Batch Size 64				Batch Size 128			
No. of Epochs	TA (%)	VA (%)	TL	VL	TA (%)	VA (%)	TL	VL	TA (%)	VA (%)	TL	VL

5	45.71	49.38	0.9668	0.9182	45.71	49.38	0.9904	0.9534	45.71	49.38	1.0197	0.9887
10	80.50	90.62	0.4573	0.3307	80.97	86.87	0.4940	0.4274	45.71	49.38	0.9376	0.9025
15	99.38	96.25	0.0396	0.1537	98.60	96.88	0.0567	0.1554	91.89	90.62	0.2733	0.2866
20	99.53	95.63	0.0105	0.1868	99.22	96.25	0.0159	0.1623	99.53	95.00	0.0145	0.1561
25	99.53	96.88	0.0096	0.1513	99.38	96.88	0.0079	0.1778	99.53	95.63	0.0084	0.1789

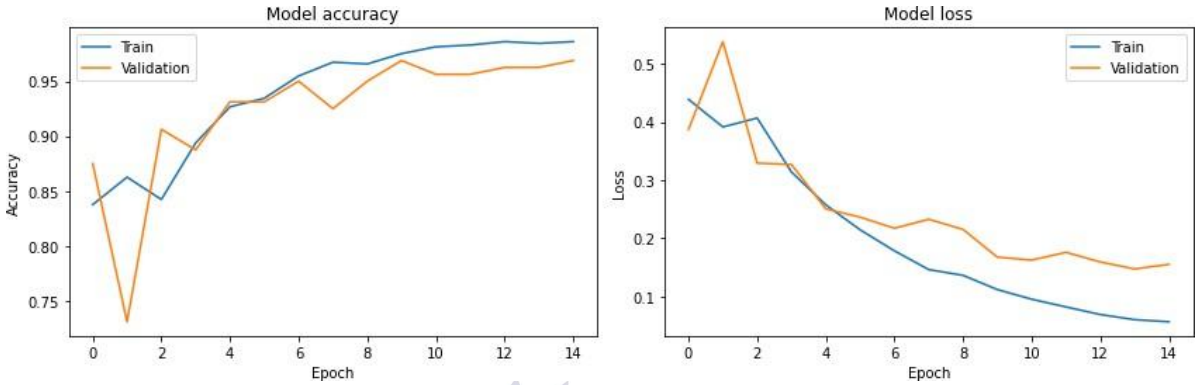


Figure 10. Training and Validation Performance Graphs for the Proposed HAG-CAP Model.

The training and validation performance of the proposed hybrid model with respect to various epochs are demonstrated in Figure 10. The accuracy plot shows that both training and validation data sets follow a steady upward trajectory and therefore, the learning and generalization process is successful. On the same note, the loss curve indicates a gradual decrease over time; the training loss decreases at a faster rate and the validation loss follows the same pattern. Combined, these graphs indicate the potential of the model to converge effectively and the limited overfitting as well as good performance in predicting.

MobileNetV2, InceptionV3, and InceptionResNetV2 examine its impact on the accuracy of classification [29]. Table 4 provides an analysis of these various model combinations in detail. This is able to note that VGG16 CBAM Caps Net is more favorable in terms of precision (100%), kappa score (97.87%), specificity (100%), and accuracy (99%) compared to other combinations of pre-trained models in our proposed model.

4.5 Hyper-parameter Tuning and Its Impact on Model Performance

Hyper-parameter tuning is a crucial step in deep learning model development, as it directly affects convergence speed, model

stability, and overall predictive accuracy. Among the most influential hyper-parameters are batch size and learning rate, which control how the model modifies its weights during training. While a larger batch size can improve computational efficiency, smaller batches frequently result in faster convergence, sometimes toward the direction of a local rather than a global optimum [30]. Similarly, the learning rate regulates the step size during gradient descent; excessively high values can cause divergence, while very low values may hinder convergence or result in suboptimal solutions.

In the present study, the proposed hybrid model is trained using batch sizes of 32, 64, and 128 and across 5, 10, 15, 20, and 25 epochs to evaluate the influence of these parameters. Multiple learning rate values are also explored to determine the most stable and accurate configuration. Table 6 summarizes the outcomes of these experiments, including training accuracy (TA), validation accuracy (VA), training loss (TL), and validation loss (VL). The results indicate that the optimal configuration was obtained at 15 epochs with a batch size of 64, achieving peak TA and VA values of 98.60% and 96.88%, respectively. Under this configuration, the lowest training loss (0.0567) and validation loss (0.1554) were also recorded.

The learning curves of this setup are shown in Figure 9. The first accuracies were about 87% (training) and 70% (validation) but tended to rise with further epochs. After the 14th epoch, the accuracy of both converged at 98.6 and 96.88 percent. The loss curves followed a steady decrease, beginning at 0.4 (training) and 0.6 (validation), decreasing to 0.2 and 0.25 in the sixth and seventh epochs, respectively, and finally, the loss curves reached a value of 0.0567 and 0.1554, respectively. The findings indicate that the hybrid model converges well without any signs of overfitting or under-fitting.

In addition, Table 6 reports performance for three acne severity classes on the test dataset using F1-score, recall, specificity, and precision. Notably, Level_2 achieved an F1-score of 97%, specificity of 100%, and precision of 100%, indicating highly reliable classification. Meanwhile, Level_0 attained the highest recall (98%), reflecting the model's strong sensitivity to early-stage acne detection.

Table 6. Performance Metrics for Acne Class in the Proposed HAG-CAP Model: Recall, Precision, Specificity, and F1-Score Values.

Dataset Class	Precision (%)	Recalling (%)	F1-score (%)	Specificity (%)

Level_0	94.00	98.00	96.00	95.91
Level_1	99.00	96.00	97.00	98.75
Level_2	100.00	95.00	97.00	100.00

5. Conclusion

Machine learning has emerged as a significant research domain over the past two decades, with considerable progress made in the application of deep learning models to dermatology. This thesis presented a comparative analysis of YOLO and CNN algorithms for facial acne detection, aiming to improve predictive accuracy in a challenging and clinically relevant task. Prior studies have demonstrated the potential of these models, yet further refinement remains necessary to achieve consistently reliable outcomes.

In the proposed study, the Roboflow facial acne dataset was utilized to evaluate the performance of CNN, YOLOv6, YOLOv7, and YOLOv8 models under multiple data splits. Performance was assessed using confusion matrices, recall, precision–recall curves, and F1-scores. Findings have shown that generally the productivity of YOLOv8 became the highest, because, however, under some circumstances, the results of YOLOv6, YOLOv7, and CNN were comparable. From the two possible split datasets, the one with an 80/10/10 split will give the most balanced performance, and the final application would thus be run with such a split.

Overall, the results confirm that the HAG-CAP framework can be effectively applied to facial acne detection. Moreover, this computational methodology demonstrates the potential for extension to other dermatological conditions, ensuring a scalable and adaptable foundation for future diagnostic applications.

References

A. Quattrini, C. Boër, T. Leidi, and R. Paydar, “A deep learning-based facial acne classification system,” Clin. Cosmet. Investig. Dermatol., vol. 15, pp. 851–857, 2022.

Q. T. Huynh et al., “Automatic acne object detection and acne severity grading using smartphone images and artificial intelligence,” Diagnostics, vol. 12, no. 8, p. 1879, 2022.

H. Li et al., “Deep skin diseases diagnostic system with dual-channel image and extracted text,” Front. Artif. Intell., vol. 6, 1213620, 2023.

J. Wang et al., "A cell phone app for facial acne severity assessment," Appl. Intell., vol. 53, no. 7, pp. 7614–7633, 2023.

H. Wu et al., "A deep learning, image-based approach for automated diagnosis for inflammatory skin diseases," Ann. Transl. Med., vol. 8, no. 9, 2020.

N. Yadav et al., "HSV model-based segmentation driven facial acne detection using deep learning," Expert Syst., vol. 39, no. 3, e12760, 2022.

H. Wen et al., "Acne detection and severity evaluation with interpretable convolutional neural network models," Technol. Health Care, vol. 30, pp. S143–S153, 2022.

J. Wang et al., "A novel automatic acne detection and severity quantification scheme using deep learning," Biomed. Signal Process. Control, vol. 84, p. 104803, 2023.

M. B. Islam et al., "Acne vulgaris detection and classification: A dual integrated deep CNN model," Informatica, vol. 47, no. 4, 2023.

M. S. Junayed et al., "ScarNet: Development and validation of a novel deep CNN model for acne scar classification with a new dataset," IEEE Access, vol. 10, pp. 1245–1258, 2021.

R. Pangti et al., "A machine learning-based, decision support, mobile phone application for diagnosis of common dermatological diseases," J. Eur. Acad. Dermatol. Venereol., vol. 35, no. 2, pp. 536–545, 2021.

Y. H. Liao, P. C. Chang, C. C. Wang, and H. H. Li, "An optimization-based technology applied for face skin symptom detection," Healthcare, vol. 10, no. 12, p. 2396, 2022.

A. Garg, H. Agrawal, S. M. Satapathy, and M. U. Khan, "Automated detection and diagnosis of skin-lesion using transfer learning based YOLOv7 approach," Algorithms Intell. Syst., pp. 393–398, 2023.

R. Nersisson, T. J. Iyer, A. N. Joseph Raj, and V. Rajangam, "A dermoscopic skin lesion classification technique using YOLO-CNN and traditional feature model," Arab. J. Sci. Eng., vol. 46, no. 10, pp. 9797–9808, 2021.

H. M. Ünver and E. Ayan, "Skin lesion segmentation in dermoscopic images with combination of YOLO and grabcut algorithm," Diagnostics, vol. 9, no. 3, p. 72, 2019.

B. Aldughayfiq, F. Ashfaq, N. Z. Jhanjhi, and M. Humayun, "YOLO-based deep learning model for pressure ulcer detection and classification," Healthcare, vol. 11, no. 9, p. 1222, 2023.

M. Abbas et al., "Enhanced skin disease diagnosis through convolutional neural networks and data augmentation techniques," J. Comput. Biomed. Inform., vol. 7, no. 1, pp. 87–106, 2023.

K. Singh, K. P. Singh, and M. Y. Khan, "Investigation and optimization of process parameters in the electrical discharge machining process," Future Technol., vol. 4, no. 2, pp. 22–29, 2025.

D. F. Sittig and H. Singh, "Recommendations to ensure safety of AI in real-world clinical care," JAMA, 2024.

F. Deutsch et al., "Biplex quantitative PCR to detect transcriptionally active HPV16 from patient saliva," BMC Cancer, vol. 24, no. 1, 2024.

Y. Yashu, V. Kukreja, P. Srivastava, A. Garg, and S. Hariharan, "AcneAI+: Revolutionizing dermatology through advanced machine learning," in Proc. IDC-IoT, 2024.

N. Hameed et al., “Mobile-based skin lesions classification using convolution neural network,” 2023.

N. Mangshor and N. A. M. Isa, “Acne type recognition for mobile-based application using YOLO,” J. Phys. Conf. Ser., vol. 1962, no. 1, p. 012041, 2021.

P. C. Kusuma and B. Soewito, “Multi-object detection using YOLOv7 object detection algorithm on mobile device,” J. Appl. Eng. Technol. Sci., vol. 5, no. 1, pp. 305–320, 2023.

P. Chen et al., “AI-Skin: Skin disease recognition based on self-learning and wide data collection through a closed-loop framework,” Inf. Fusion, vol. 54, pp. 1–9, 2020.

Y. Sun, X. Li, and X. Zhang, “HRT-YOLO: A transformer-based high-resolution representation model for face flaw detection,” J. Phys. Conf. Ser., vol. 2258, no. 1, p. 012032, 2022.

A. Sankar et al., “Retracted: Utilizing generative adversarial networks for acne dataset generation in dermatology,” BioMedInformatics, vol. 4, no. 2, 2024.

N. Yadav, A. Alfayeed, A. Khamparia, B. Pandey, D. N. Thanh, and S. Pande, “HSV model-based segmentation driven facial acne detection using deep learning,” Expert Syst., vol. 39, no. 3, e12760, 2022.

R. Yadav, A. Jain, and S. Sharma, “Acne detection care system using deep learning,” in Proc. ICRITO, 2024.

M. Chen et al., “AI-Skin: Skin disease recognition based on self-learning and wide data collection through a closed-loop framework,” Inf. Fusion, vol. 54, pp. 1–9, 2020.