Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi
Vol.10 No.1 January 2025, P-ISSN : 2502-3470, E-ISSN : 2581-0367

Skin Lesion Classification Using YOLOv11 on the HAM10000 Dataset

Islam Cahya Wicaksana[1], Ricardus Anggi Pramunendar[2], Galuh Wilujeng Saraswati[3], Gustina Alfa Trisnapradika[4]

[1,2,3,4]*Informatics Department, Universitas Dian Nuswantoro, Indonesia*

[1]wicakislamcahya@gmail.com(*)

[2,3,4][ricardus.anggi, galuhwilujengs, gustina.alfa]@dsn.dinus.ac.id

Received: 2024-12-02; Accepted: 2025-01-13; Published: 2025-01-21

Abstract— Skin cancer represents a significant global health concern due to its high mortality rate. Early and accurate detection is crucial but often hindered by the limitations of traditional diagnostic methods. This research applies the YOLOv11 algorithm for skin lesion classification directly from dermoscopic images using the HAM10000 dataset (10,015 images, 7 skin lesion classes). The primary objectives are to evaluate YOLOv11's performance in multi-class classification and assess the impact of data augmentation (rotation, horizontal flipping) in addressing class imbalance. The methodology involved two experiments: training YOLOv11 on the original and augmented datasets and comparing its performance with multi-stage architectures (VGG19 and ResNet50). Five pre-trained YOLOv11 models were tested using accuracy, precision, recall, and F1-score metrics. Results showed the YOLOv11x-cls model trained on the augmented dataset achieved the best performance among YOLOv11 models (accuracy 84.74%, precision 83.94%, recall 84.74%, F1-score 84.06%). However, VGG19 recorded the highest accuracy (89.68%). Data augmentation effectively improved model performance by mitigating class imbalance. This study also indicates that multi-stage architectures perform better in skin lesion classification tasks than single-stage architectures. The key contributions of this research are: (1) a comprehensive performance comparison of YOLOv11 with VGG19 and ResNet50 for skin lesion classification and (2) empirical validation of data augmentation's effectiveness in improving model performance. This study demonstrates that YOLOv11 can achieve competitive performance in skin lesion classification despite not surpassing the performance of multi-stage architectures.

Keywords— Skin Lesion Classification; YOLOv11; HAM10000 Dataset; Data Augmentation; Class Imbalance

I. INTRODUCTION

Skin lesions, including skin cancer, are among the most common types of cancer globally [1], thus highlighting the widespread distribution of skin lesions across multiple continents. Its primary cause is exposure to ultraviolet (UV) radiation from sunlight, which can damage the DNA of skin cells and trigger abnormal cell growth [2]. Additional risk factors include a history of sunburn and the use of tanning beds, which emit artificial UV radiation that significantly increases the risk of melanoma, a form of skin lesion [3], [4], [5]. These findings suggest that ultraviolet (UV) radiation exposure, including both solar and artificial sources like tanning beds, contributes to the development of skin cancer and that tanning beds constitute a significant source of this risk.

Globally, about 3 million new cases of non-melanoma skin cancer are reported annually [6]. Melanoma, a more aggressive type of skin cancer, accounts for around 125,000 new cases annually worldwide [7]. The known cancer types in skin lesions are melanocyte and non-melanocyte. Melanocytic lesions include melanoma (mel) and melanocytic nevi (nv). In contrast, non-melanocytic skin diseases include benign keratosis-like lesions (bkl), basal cell carcinoma (bcc), actinic keratosis (akiec), vascular lesions (vasc), and dermatofibroma (df). [8]. Among these, melanoma is the most deadly, with a survival rate of only 15% [9].

Dermoscopy is a diagnostic approach that does not require invasive procedures, enabling close examination of skin surface and subsurface structures [10]. While it is widely used to identify skin cancer symptoms, its effectiveness heavily depends on the dermatologist's experience, making it prone to subjectivity and diagnostic inaccuracies [11]. This underscores the necessity of advanced technologies, such as deep learning, to aid specialists in improving diagnostic accuracy. Recent innovations in deep learning have greatly enhanced the ability to diagnose skin cancer, particularly by classifying dermoscopic images of skin lesions. These methods offer improved accuracy, speed, and consistency compared to traditional diagnostic approaches.

The HAM10000 dataset is a well-known dataset for skin lesion classification, containing 10,015 dermoscopic images covering seven skin lesions, including actinic keratosis, vascular lesions, melanocytic nevi, seborrheic keratosis, basal cell carcinoma, dermatofibroma, and melanoma. The HAM10000 dataset is frequently used in studies due to its diversity, expert annotations, and accessibility, making it a standard benchmark for developing and evaluating machine learning models in dermatology research.

While many studies use HAM10000 and deep learning for skin lesion classification, the performance of single-stage CNNs, especially YOLOv11, is underexplored. Single-stage architectures like YOLOv11, with integrated detection and classification, offer the potential for detailed image processing. This study examines the viability of YOLOv11 in skin lesion classification by comparing it to multi-stage models, VGG19 and ResNet50, known for hierarchical feature extraction. We also analyze the impact of data augmentation on model performance, particularly in addressing class imbalance.

Several studies have utilized the HAM10000 dataset for skin lesion classification and detection. For instance, Adebiyi A conducted research using the multi-modal ALBEF architecture to classify seven classes of skin lesions, achieving 94.11%

DOI : https://doi.org/10.25139/inform.v10i1.9310

accuracy [12]. Another study by Ingle Y employed the VGG16 architecture for the same classification task and reported 88.83% accuracy [13]. Additionally, a different study focused on a binary classification task—detecting benign and malignant lesions—using the YOLOv8 algorithm, which achieved precision and recall values of 78.2% and 81.6% [14].

In this research, we comprehensively explored the performance of YOLOv11 as a single-stage architecture for skin lesion classification on the HAM10000 dataset. This research investigated the extent to which YOLOv11 could achieve accurate and effective classification results in this task compared to multi-stage models such as VGG19 and ResNet50, known for their hierarchical feature extraction capabilities. By comparing these three architectures, this study aimed to provide in-depth insights into their performance in skin lesion classification and the effectiveness of data augmentation in addressing class imbalance.

The main contribution of this study is a comprehensive comparison of YOLOv11's performance against VGG19 and ResNet50 for skin lesion classification using the HAM10000 dataset, with particular emphasis on analyzing the impact of data augmentation on the performance of these models.

II. Research Methodology

This study implements YOLOv11, VGG19, and ResNet50 models, utilizing the HAM10000 dataset to classify various skin lesion types. The methodology encompassed several stages, such as data collection, image preprocessing involving segmentation, splitting dataset, resizing and augmentation, CNN modelling, and model performance evaluation, as depicted in Fig.1.

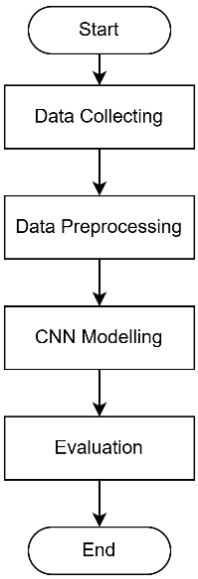

Fig.1. Flowchart Method

A. Data Collecting

The dataset used in this research is the HAM10000 dataset. This dataset represents 7 types of skin lesions and contains 10,015 dermatoscopic images. Dermatology experts annotated the dataset. The classes included are vascular lesions (vasc), actinic keratoses (akiec), melanocytic nevi (nv), benign keratosis-like lesions (bkl), basal cell carcinoma (bcc), dermatofibroma (df), and melanoma (mel). The distribution of images across these classes is visualized in Fig.2.

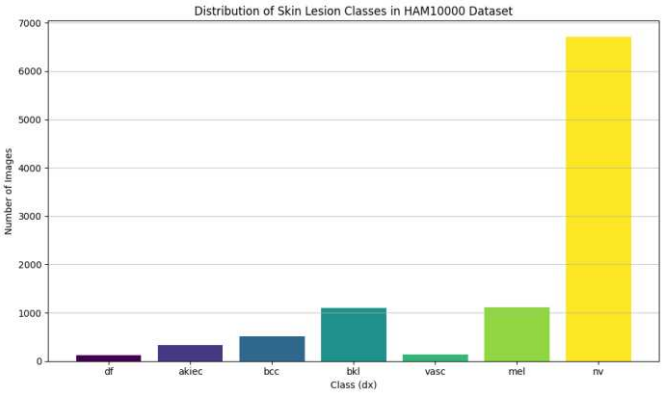

Fig.2. Class Distribution

The dataset contains a different number of images for each class. Dermatofibroma (df) has the smallest representation in the dataset with 115 images, followed by vascular lesions (vasc) with 142 images, actinic keratoses (akiec) with 327 images, and basal cell carcinoma (bcc) with 514 images. Benign keratosis-like lesions (bkl) are represented by 1,099 images, while melanoma (mel) accounts for 1,113. Melanocytic nevi (nv) has the largest representation, containing 6,705 images. This distribution highlights a significant class imbalance within the dataset.

The imbalance in class distribution within the dataset often leads to misclassification of minority classes, which are more likely to be incorrectly predicted than majority classes [15]. To solve this issue, this study conducts two distinct experiments. The first involves training the model using the original dataset without augmentation, and the second utilizes an augmented dataset to mitigate the effects of class imbalance.

B. Data Preprocessing

Before model training, the dataset underwent several preprocessing steps. Firstly, to focus on the skin lesion and eliminate irrelevant information such as surrounding skin, hair, and other noise, the provided binary masks in the HAM10000 dataset were used to segment each image. These masks are a binary representation, where pixels corresponding to the lesion are labelled 1, while the rest are labelled 0. Each image was segmented using the binary mask by multiplying it with its mask. Then, each segmented image was cropped automatically to focus on the lesional area using the bounding box information generated from the segmented non-black regions. Thirdly, all images were resized to 224x224 pixels for the VGG19 and ResNet50 models and 640x640 pixels for the YOLOv11 model. Then, the dataset was divided into three subsets using a stratified split, with a composition of 70% allocated for training, 20% for validation, and 10% for testing. Following this split, data augmentation techniques were exclusively applied to the training set to address the class

DOI : https://doi.org/10.25139/inform.v10i1.9310

imbalance and enhance the generalization ability of the models. The augmentation process included various transformations designed to increase the number of images in underrepresented classes, thereby minimizing class imbalance in the dataset [16]. Specifically, the augmentation techniques applied were rotation by -15 and 15 degrees vertical and horizontal flipping. The Nevus Melanocytic (nv) class, the majority class with 6705 images, was used as a standard for balancing other classes. This approach ensured no data or important information from the majority class was removed or reduced.

The primary objective was to enhance the representation of minority classes while retaining essential features of dominant classes, as reducing the number of images in majority classes could lead to the loss of critical information [17]. By adding this approach to the dataset, it was anticipated that the model's capacity to generalize across various types of lesions would be greatly improved, which would result in classification findings that were more reliable and balanced.

C. CNN Modelling

1) YOLOv11: In this research, we explored five different variants of the YOLOv11 algorithm to perform classification on seven skin lesion types from the HAM10000 dataset, such as YOLOv11n-cls, YOLOv11s-cls, YOLOv11m-cls, YOLOv11l-cls, and YOLOv11x-cls. These models, initialized with ImageNet pre-trained weights, vary in size, accuracy, and computational complexity. The lighter models, such as YOLOv11n-cls, are designed for faster inference with minimal computational requirements, while the larger models, like YOLOv11x-cls, prioritize accuracy with higher resource demands. YOLOv11, introduced in September 2024, is the most recent version in the YOLO series of models. Notable architectural advancements in YOLOv11 include the integration of the SPPF module to capture multi-scale information and the incorporation of the C2PSA block to improve focus on important regions within an image [16], which are modifications from the YOLOv8 design. YOLOv11 is a single-stage model designed for simultaneous object detection and classification. It comprises three main parts: the backbone for feature extraction, the Neck for feature fusion, and the Head for detection and classification, the YOLOv11 architecture, as illustrated in Fig.3.

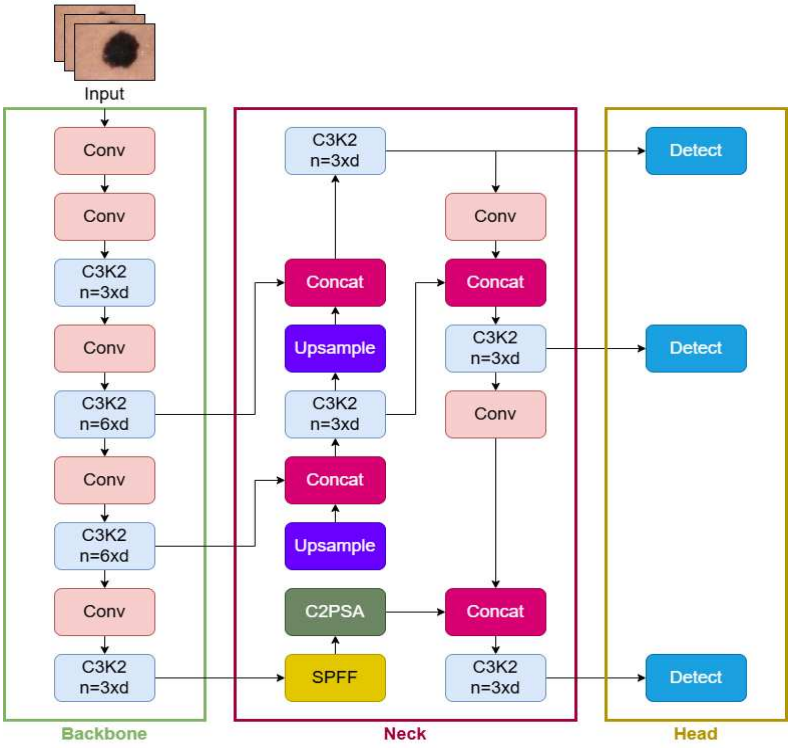

Fig.3. YOLOv11 Architecture

The backbone is responsible for extracting meaningful features from the input image. The input image (640x640x3) is passed through several convolutional layers and C3k2 blocks to down-sample the image and extract meaningful hierarchical features [17]. The feature map's spatial dimensions are progressively reduced, and the number of channels increases. The backbone consists of convolutional layers (Conv) and C3k2 blocks, progressively reducing the feature maps' spatial dimensions and increasing their depth. The C3k2 blocks are a crucial part of the backbone feature learning process, as they use two 3x3 convolution filters instead of one large filter to reduce the computational burden while learning highly detailed features. The backbone output is a set of multi-scale feature maps with different resolutions that are later used in the neck section. The

DOI : https://doi.org/10.25139/inform.v10i1.9310

operation of the C3k2 block is mathematically seen in Equation (1). Where F denotes the input feature map passed through the block while $Conv_{3x3}$ represents the application of a convolutional layer with a kernel size of 3x3 to half of the input feature map. The operation $Concat$ refers to concatenating the processed and unprocessed portions of the feature map, resulting in an output feature map. F_{C3k2}. This structure reduces computational redundancy and enhances feature representation efficiency.

$$F_{C3k2} = Concat(F, Conv_{3x3}(F)) \quad (1)$$

The neck combines the feature maps from the backbone at different scales using up-sampling, concatenation, and additional convolutional layers. The Spatial Pyramid Pooling Fast (SPPF) module combines the multiple scales features of the *backbone* at a fixed scale. It allows the architecture to be trained with many objects of different sizes in mind [18]. The mathematical representation of the SPPF module is given in Equation (2). Where F mean is the input feature map, and $MaxPool_k$ applies for operations with kernel sizes of $k = 5$ and $k = 3$.

$$F_{SPPF} = Concat(MaxPool_{k=5}, MaxPool_{k=3}(F), F) \quad (2)$$

These pooling operations capture spatial information at different scales. The concatenation of the pooled results with the original feature map creates the final output F_{SPPF} feature map, which contains multi-scale contextual information. Then, the *C2PSA (Cross Stage Partial with Spatial Attention) Block* enables the model to use information from various scales to focus on the most important regions of the image during feature fusion [16]. The operation of the C2PSA block is expressed in Equation (3). Where X represents the input feature map. The block splits X into two pathways, X_1 and X_2. Each is processed through distinct convolutional or attention mechanisms. The concatenation of these pathways, $Concat(X_1, X_2)$ is passed through an attention mechanism. $Attention$, which assigns weights to regions of importance. This enables the model to focus on critical areas in the feature map, resulting in an output that emphasizes the most relevant regions for object detection.

$$C2PSA = Attention(Concat(X_1, X_2)) \quad (3)$$

By combining the C3k2 blocks, the SPPF modules, and the C2PSA blocks, YOLOv11 achieves an improved balance of computational efficiency and accuracy [19]. These enhancements enable YOLOv11 to make it an ideal solution for classifying skin lesions in the HAM10000 dataset.

2) VGG19: VGG19 is a classical multi-stage deep convolutional neural network (CNN) architecture renowned for its depth and the use of small 3x3 convolutional filters [20]. It is a feedforward architecture where the input data flows sequentially to the output without loops or skip connections. The main parts of VGG19 architecture include convolutional layers, pooling layers, and fully connected layers, each playing a different role in the feature extraction and classification process. The term multi-stage in this context refers to the architecture processes the input through several stages of convolution and pooling before the classification layer. The VGG19 architecture used in this research is shown in Fig.4.

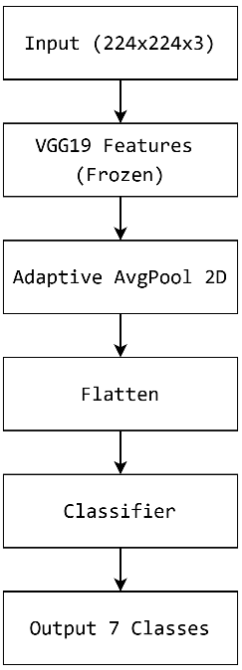

Fig.3. VGG19 Architecture

The VGG19 architecture used in this research consists of the VGG19 feature extraction network with initial layers (a max-pooling layer follows 16 convolutional layers arranged into blocks, each convolutional block) frozen. The feature maps from the frozen feature extraction layers are passed into an Adaptive Average Pooling Layer that downscales them. Then a fully connected layer is used to perform the final classification. These fully connected layers perform the final classification using the features extracted by previous layers. The final fully connected layer has an output size corresponding to the number of classes in the dataset. This research used a VGG19 model pre-trained on the ImageNet dataset. The initial layers of VGG19 were frozen to maintain the learned feature representations and decrease training time. This step was followed by training the fully connected layers to fit the model for the skin lesion classification task on the HAM10000 dataset.

3) ResNet50: ResNet50 is a powerful multi-stage deep convolutional neural network (CNN) architecture that employs shortcut connections, also known as skip connections, to facilitate the training of very deep neural networks and effectively address the vanishing gradient problem [21]. The architecture is based on residual blocks, the core components of ResNet50. In this research, ResNet50 was used with pre-trained weights on the ImageNet dataset. The ResNet50 architecture used in this research is presented in Fig.5. The ResNet50 architecture used in this research starts with an input layer that receives images of a specific dimension (224x224x3) followed by a convolutional layer that extracts simple initial

DOI : https://doi.org/10.25139/inform.v10i1.9310

features. The feature maps are then passed through multiple residual blocks organized in multiple stages. There are two types of residual blocks: the identity block, which is applied when input and output have the same dimensions.

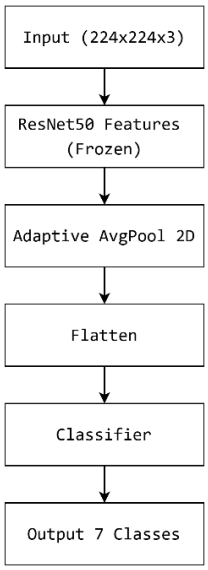

Fig.4. ResNet50 Architecture

It contains three convolutional layers within the block and a skip connection that adds the input of the block to the output of those convolutional layers. A convolutional block is applied when input and output have different dimensions. It contains three convolutional layers within the block and a convolutional layer on the skip connection that adjusts the input size. The skip connections enable the model to learn the residual mappings, preventing the vanishing gradients. A global Adaptive Average Pooling Layer is at the end of the feature extraction network, which reduces the feature map dimensions before flattening. This flattened feature is then passed into fully connected layers that perform classification. This research used a ResNet50 model pre-trained on the ImageNet dataset. To adapt this architecture to our specific task, the initial convolutional layers of ResNet50, which perform the feature extraction task, were frozen to maintain the pre-trained features. After that, the newly added fully connected layers for classifying seven different skin lesion classes in the HAM10000 dataset were fine-tuned. This transfer learning strategy allowed for more efficient training and boosted performance. Moreover, the ResNet50 and VGG19 models have different design principles, with VGG19 focusing on deep stacking of convolutional layers and ResNet50 emphasizing efficient training of deep neural networks using skip connections.

4) Hyperparameter Tuning: Optimization for the performance of the models, hyperparameter tuning was performed. The hyperparameter values for all models (YOLOv11, VGG19, and ResNet50) are summarized in Table I.

TABLE I
TEST DATASET CLASS DISTRIBUTION

Hyperparameter	Values
Optimizer	AdamW
Initial Learning Rate	0.01, 0.001, 0.0001
Batch Size	32
Epochs	100

All models were trained using the AdamW optimizer, known for its adaptive learning rate capabilities and effectiveness in training deep neural networks. We explored three initial learning rates, such as 0.01, 0.001, and 0.0001. These values were selected based on initial experimentation and prior literature, where smaller learning rates have proven useful on tasks requiring fine-tuning, and larger learning rates are beneficial for the initial training. A batch size of 32 was used across all models in this research to fit within the available computing resources, where we utilized two Nvidia T4 GPUs with 15GB of memory each. The number of epochs was fixed at 100, based on initial observations on the convergence of the models. This configuration was carefully chosen to balance computational resources and optimal model performance.

D. Evaluation

This section describes the evaluation metrics used to assess the performance of all the models (YOLOv11, VGG19, and ResNet50) on the skin lesion classification task, using a separate testing set different from the training and validation sets. We employed precision, recall, F1-score, and accuracy, calculated for each class. These class-specific metrics were combined using a weighted average approach to obtain the overall performance scores. By applying these, the performance of each model, trained on both raw and augmented datasets, was systematically evaluated.

1) Precision is the ratio of true positives to the total number of true and false positives, as expressed in Equation (4).

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)} \quad (4)$$

2) Recall: The proportion of actual positive cases correctly recognized is measured by this metric and is defined in Equation (5).

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)} \quad (5)$$

3) F1-Score: The F1-score, a metric combining precision and recall through their harmonic mean, is defined in Equation (6).

$$F1\ Score = \frac{Precision \cdot Recall}{Precision + Recall} \quad (6)$$

4) Accuracy: This metric determines the percentage of correctly classified samples among the total samples, with its definition provided in Equation (7).

$$Accuracy = \frac{True\ Positives\ (TP) + True\ Negatives\ (TN)}{Total\ Samples} \quad (7)$$

DOI : https://doi.org/10.25139/inform.v10i1.9310

III. Result and Discussion

This section presents and discusses the results obtained from the skin lesion classification experiments using three different CNN architectures: YOLOv11, VGG19, and ResNet50. The model performance was evaluated on two sets of experiments, the first set of results using the original (raw) dataset and the second using the augmented dataset.

A. Data Segmentation Results

The masks provided in the HAM10000 dataset were used to segment each image, focusing on the skin lesion itself and eliminating irrelevant information such as surrounding skin, hair, and other noise, resulting in better focus on the lesion area. Fig.6 presents the sample of this masking and cropping process applied to an image taken from a nevus melanocytic class.

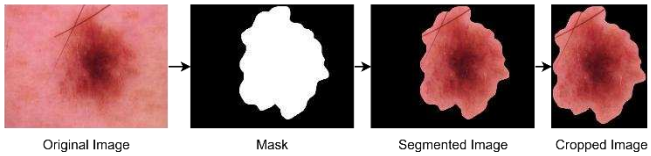

Fig.5. Sample of Segmentation Result

B. Data Augmentation Results

This research used augmentation approaches to improve the representation of minority classes and overcome the imbalanced dataset problem. In Fig.7, we see the outcome of applying augmentations to a single image that belongs to the Melanoma (mel) class.

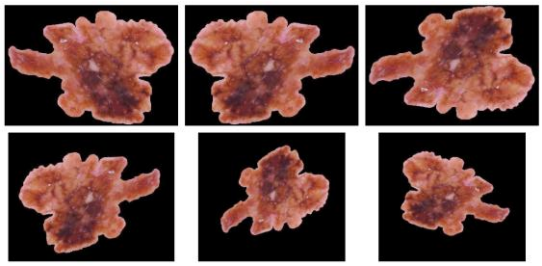

Fig.6. Augmentation Sample on Melanoma

Table II informs the distribution of train images across classes in the original and augmented datasets. The augmentation process significantly increased the number of train images for minority classes while keeping the nv class unchanged.

TABLE II
TRAIN SET IMAGE COUNTS AFTER AUGMENTATION

Class	Original Count	Augmented Count
df	80	400
akiec	228	1140
bcc	359	1795
bkl	769	3845
vasc	99	495
mel	779	3895
nv	4693	4693

C. Models Performance Evaluation

The capability of all models (YOLOv11, VGG19, and ResNet50) was measured using the testing set, which was reserved exclusively for testing, ensuring that the reported metrics represent the models' ability to generalize on unseen data. The results presented for each model were obtained with the best configuration found after several experiments using three different initial learning rates (0.01, 0.001, and 0.0001), and the testing set performance showed that a learning rate of 0.0001 generally produces the best result for all models. Therefore, the performance metrics detailed in the following sections were obtained from experiments that exclusively employed a learning rate 0.0001.

1) Results Trained on the Raw Dataset: For each model that was trained with the raw dataset that is shown in Table III, performance measures such as accuracy, precision, recall, and F1-score were calculated.

TABLE III
PERFORMANCE METRICS TRAINED USING THE RAW DATASET

Model	Accuracy	Precision	Recall	F1-Score
YOLOv11x-cls	0.6851	0.6849	0.6841	0.6891
YOLOv11l-cls	0.6743	0.6745	0.6734	0.675
YOLOv11m-cls	0.6725	0.6702	0.669	0.6517
YOLOv11s-cls	0.6592	0.6635	0.653	0.6412
YOLOv11n-cls	0.6491	0.6351	0.6252	0.6305
VGG19	0.7912	0.7236	0.7391	0.7212
ResNet50	0.7743	0.7591	0.7743	0.7424

The experiment using the raw dataset showed that the YOLOv11x-cls model achieved the best performance among the different YOLOv11 variants, with an accuracy of 68.51% and an F1-score of 68.91%. While the other YOLOv11 variants had slightly inferior results, these findings do indicate the potential of the YOLOv11 architecture for classifying skin lesions. It is important to note that in this experiment, VGG19 had the best performance in terms of accuracy (79.12%) and F1-score (72.12%), followed by ResNet50 with an accuracy of 77.43% and F1-score of 74.24%. However, the YOLOv11 models exhibited relatively low recall, with the highest recall value being 68.41%, obtained by the YOLOv11x-cls model. Thus, the potential of YOLOv11 for skin lesion classification needs further examination, specifically with a more comprehensive approach to data augmentation.

2) Results Trained on the Augmented Dataset: In the second experiment, the model's performance was evaluated using the same test dataset as in the raw dataset experiment, as presented in Table IV.

TABLE IV
PERFORMANCE METRICS TRAINED USING THE AUGMENTED DATASET

Model	Accuracy	Precision	Recall	F1-Score
YOLOv11x-cls	0.8474	0.8394	0.8474	0.8406
YOLOv11l-cls	0.8315	0.8236	0.8312	0.8249
YOLOv11m-cls	0.8112	0.8035	0.8110	0.8047
YOLOv11s-cls	0.7920	0.7845	0.7920	0.7856
YOLOv11n-cls	0.7713	0.7639	0.7713	0.7651
VGG19	0.8968	0.88	0.8965	0.8720
ResNet50	0.8751	0.8542	0.8751	0.8421

DOI : https://doi.org/10.25139/inform.v10i1.9310

The results show a significant improvement in all metrics for all models when trained using an augmented dataset. The YOLOv11x-cls model exhibited the most prominent performance gain following data augmentation, with accuracy increasing from 68.51% to 84.74% and recall improving from 68.41% to 84.74%. This indicates the substantial potential of this single-stage architecture for skin lesion classification tasks. Despite this, the VGG19 model achieved the best overall performance on the augmented dataset, with an accuracy of 89.68% and an F1-score of 87.20%, suggesting the advantage of multi-stage architectures. These results demonstrate that data augmentation significantly improves the performance of all tested models, underscoring the importance of this technique in addressing class imbalance and data limitations within the skin lesion dataset. The confusion matrix in Fig.8. details the classification results obtained by the YOLOv11x model, specifically after its training on the augmented dataset.

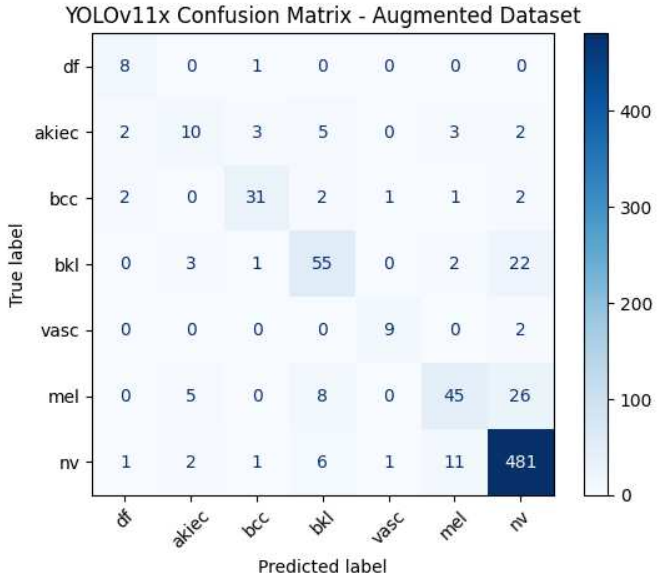

Fig.7. YOLOv11x Confusion Matrix

The confusion matrix reveals that the model exhibited the lowest recall for the 'akiec,' 'bkl,' and 'mel' classes, frequently misclassifying these lesions into other classes. For instance, several 'bkl' lesions were often classified as 'nv'. This suggests that the model may struggle to differentiate patterns among these classes, likely due to visual similarities in some features, despite the application of data augmentation. To further evaluate the YOLOv11x-cls model's ability to discriminate between classes, we present the Receiver Operating Characteristic (ROC) curves and the Area Under the Curve (AUC) values in Fig.9. The ROC curves, shown in Fig.9, exemplify the model's strong ability to discriminate among the various classes. With a weighted average AUC of 0.9504, the YOLOv11x-cls model demonstrates exceptional overall performance. Class-wise AUC scores also highlight good performance, with the df class attaining the highest score of 0.9972 and vasc the lowest at 0.9628.

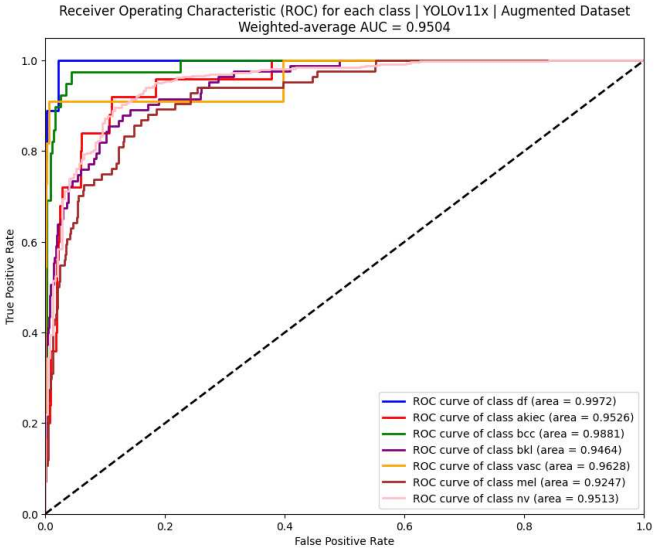

Fig.8. ROC AUC YOLOv11x

For broader contextualization, Table V a comparison of the performance of this study and relevant research, outlined in the introduction, that used the HAM10000 dataset for skin lesion classification.

TABLE V
COMPARATIVE RESULTS ACROSS RELATED STUDIES

Research	Model	Accuracy
Adebiyi A [12]	ALBEF	94.11%
Ingle Y [13]	VGG16	88.83%
S. Ćirković [14]	YOLOv8	81.6%
This research	YOLOv11x	84.74%
This research	VGG19	89.68%
This research	ResNet50	87.51%

The findings reveal that the YOLOv11x model, employing our method, attains a competitive and noteworthy level of accuracy. Although it does not achieve the performance levels of multi-stage models like VGG, ResNet, and the multi-modal system of Adebiyi A et al., YOLOv11x outperforms the YOLOv8 model used in prior studies [14]. Therefore, this study offers empirical support for applying YOLOv11x to skin lesion classification and highlights the effectiveness of data augmentation for improving its performance.

IV. CONCLUSION

This study focused on exploring the performance of the single-stage architecture YOLOv11 in skin lesion classification using the HAM10000 dataset, comparing it with the multi-stage architectures VGG19 and ResNet50, and investigating the impact of data augmentation on model performance. The results demonstrated that data augmentation significantly improved the accuracy of the YOLOv11x-cls model, from 68.51% to 84.74%. However, VGG19 achieved the highest performance of all models, with an accuracy of 89.68%. This study highlights that while single-stage architectures offer computational efficiency, multi-stage architectures with

DOI : https://doi.org/10.25139/inform.v10i1.9310

hierarchical feature extraction yield better classification performance.

The main contribution of this research lies in providing empirical validation of YOLOv11's performance in the context of skin lesion classification, as well as confirming the significance of data augmentation in enhancing model performance. These findings indicate that although single-stage architectures are efficient in processing, multi-stage architectures maintain superiority in classification task performance. The YOLOv11x-cls model performed superior to the YOLOv8 model reported in prior studies. Furthermore, utilizing the current methodology, the performance gap between our YOLOv11x-cls and multi-stage models was comparatively small. Further research is recommended to explore more advanced models and to develop more innovative augmentation techniques to address challenges within skin lesion datasets.

ACKNOWLEDGMENT

The author gratefully thanks Universitas Dian Nuswantoro (UDINUS) for the support and resources that greatly facilitated the successful completion of this research.

REFERENCES

[1] S. Imani, G. Roozitalab, mahdieh Emadi, A. Moradi, P. BEHZADI, and P. Jabbarzadeh Kaboli, “The Evolution of BRAF-Targeted Therapies in Melanoma: Overcoming Hurdles and Unleashing Novel Strategies,” *Front Oncol*, vol. 14, Nov. 2024, doi: 10.3389/fonc.2024.1504142.

[2] Z. Hayder, J. Alkufaishi, D. Amer, and A. A. Rahi, "Review of the current knowledge on the Types, pathogenesis, and prevention of Carcinoma Occurrence," 2024.

[3] J. Reimann *et al.*, "A Process Evaluation of the Skin Cancer Prevention Act (Tanning Beds): A Survey of Ontario Public Health Units," *J Community Health*, vol. 44, no. 4, pp. 675–683, Aug. 2019, doi: 10.1007/s10900-019-00658-1.

[4] K. Shehzad *et al.*, "A Deep-Ensemble-Learning-Based Approach for Skin Cancer Diagnosis," *Electronics (Switzerland)*, vol. 12, no. 6, Mar. 2023, doi: 10.3390/electronics12061342.

[5] C. Dessinioti and A. J. Stratigos, "An Epidemiological Update on Indoor Tanning and the Risk of Skin Cancers," Nov. 17, 2022, *NLM (Medline)*. doi: 10.3390/curroncol29110699.

[6] S. S. Chaturvedi, K. Gupta, and P. S. Prasad, "Skin Lesion Analyser: An Efficient Seven-Way Multi-class Skin Cancer Classification Using MobileNet," in *Advanced Machine Learning Technologies and Applications*, A. E. Hassanien, R. Bhatnagar, and A. Darwish, Eds., Singapore: Springer Singapore, 2021, pp. 165–176.

[7] P. Agrahari, A. Agrawal, and N. Subhashini, "Skin Cancer Detection Using Deep Learning," in *Futuristic Communication and Network Technologies*, A. Sivasubramanian, P. N. Shastry, and P. C. Hong, Eds., Singapore: Springer Nature Singapore, 2022, pp. 179–190.

[8] A. Naeem, T. Anees, M. Fiza, R. A. Naqvi, and S. W. Lee, "SCDNet: A Deep Learning-Based Framework for the Multiclassification of Skin Cancer Using Dermoscopy Images," *Sensors*, vol. 22, no. 15, Aug. 2022, doi: 10.3390/s22155652.

[9] Y. Dong, L. Wang, S. Cheng, and Y. Li, "FAC-Net: Feedback attention network based on context encoder network for skin lesion segmentation," *Sensors*, vol. 21, no. 15, Aug. 2021, doi: 10.3390/s21155172.

[10] B. S. Ankad, S. V Smitha, and V. R. Koti, "Basic Science of Dermoscopy," *Clinical Dermatology Review*, vol. 4, no. 2, 2020.

[11] H. M. Ünver and E. Ayan, "Skin lesion segmentation in dermoscopic images with combination of yolo and grabcut algorithm," *Diagnostics*, vol. 9, no. 3, Sep. 2019, doi: 10.3390/diagnostics9030072.

[12] A. Adebiyi *et al.*, "Accurate Skin Lesion Classification Using Multi-modal Learning on the HAM10000 Dataset," May 31, 2024. doi: 10.1101/2024.05.30.24308213.

[13] Y. S. Ingle and N. Faiz Shaikh, "Deep Learning for Skin Cancer Classification: A Comparative Study of CNN and Vgg16 on HAM10000 Dataset," 2024.

[14] S. Ćirković and N. Stanić, "Application of the YOLO algorithm for Medical Purposes in the Detection of Skin Cancer," in *10th International Scientific Conference Technics, Informatic, and Education*, University of Kragujevac, Faculty of Technical Sciences, Čačak, 2024, pp. 83–88. doi: 10.46793/TIE24.083C.

[15] S. Fotouhi, S. Asadi, and M. W. Kattan, "A comprehensive data level analysis for cancer diagnosis on imbalanced data," Feb. 01, 2019, *Academic Press Inc.* doi: 10.1016/j.jbi.2018.12.003.

[16] R. Khanam and M. Hussain, "YOLOv11: An Overview of the Key Architectural Enhancements," Oct. 2024.

[17] M. A. R. Alif, "YOLOv11 for Vehicle Detection: Advancements, Performance, and Applications in Intelligent Transportation Systems," Oct. 2024.

[18] Y. Wan, H. Wang, L. Lu, X. Lan, F. Xu, and S. Li, "An Improved Real-Time Detection Transformer Model for the Intelligent Survey of Traffic Safety Facilities," *Sustainability*, vol. 16, no. 23, p. 10172, Nov. 2024, doi: 10.3390/su162310172.

[19] N. Jegham, C. Y. Koh, M. Abdelatti, and A. Hendawi, "Evaluating the Evolution of YOLO (You Only Look Once) Models: A Comprehensive Benchmark Study of YOLO11 and Its Predecessors," Oct. 2024.

[20] V. Rajinikanth, A. N. J. Raj, K. P. Thanaraj, and G. R. Naik, "A customized VGG19 network with concatenation of deep and handcrafted features for brain tumor detection," *Applied Sciences (Switzerland)*, vol. 10, no. 10, May 2020, doi: 10.3390/app10103429.

[21] D. Theckedath and R. R. Sedamkar, "Detecting Affect States Using VGG16, ResNet50 and SE-ResNet50 Networks," *SN Comput Sci*, vol. 1, no. 2, Mar. 2020, doi: 10.1007/s42979-020-0114-9.

DOI : https://doi.org/10.25139/inform.v10i1.9310