International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 12-No.2 (2026) pp. 474-485
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Deep Learning-Based Melanoma Classification Using Hybrid DCNN-LSTM and DCNN-BiLSTM Architectures

Mokhtaria Bekkaoui[1,2]*, Hafidha Sebbagh[3], Fekar Mohammed Riyadh El Mansour[4], Mohammed Merzoug[5], Mourad Hadjila[6], Mohammed M'hamedi[7,8]

[1]Department of First Cycle, Ecole Supérieure en Sciences Appliquées de Tlemcen ESSAT, Tlemcen, 13000, Algeria
[2]Manufacturing Engineering Laboratory of Tlemcen, University of Tlemcen, Algeria
* **Corresponding Author Email:** mokhtaria.bekkaoui@univ-tlemcen.dz- **ORCID:** 0009-0007-7625-9106

[3]Department of First Cycle, Ecole Supérieure en Sciences Appliquées de Tlemcen ESSAT, Tlemcen, 13000, Algeria
Email: sebbagh.hafidha@gmail.com - **ORCID:** 009-0009-0594-9198

[4]Department of Computer Science, Faculty of Science, University of Abou-Bakr Belkaid, LRIT Laboratory, Tlemcen, 13000, Algeria
Email: fekarmohammedriyadhelmansour@gmail.com - **ORCID:** 0000-0002-6554-3925

[5]Department of Telecommunications, Faculty of Technology,University of Abou-Bakr Belkaid, STIC Laboratory, Tlemcen,13000, Algeria
Email: mohammed.merzoug@univ-tlemcen.dz - **ORCID:** 0000-0002-0426-1668

[6]Manufacturing Engineering Laboratory of Tlemcen, University of Tlemcen, Algeria
Email: mourad.hadjila@univ-tlemcen.dz - **ORCID:** 0000-0002-6554-3925

[7]Department of First Cycle, Ecole Supérieure en Sciences Appliquées de Tlemcen ESSAT, Tlemcen, 13000, Algeria
[8]Department of Computer Science, Faculty of Science, University of Abou-Bakr Belkaid, LRIT Laboratory, Tlemcen, 13000, Algeria
Email: mohamed.mhamedi@univ-tlemcen.dz - **ORCID:** 0009-0003-7899-2185

Article Info:

DOI: 10.22399/ijcesen.5251
Received : 10 November 2025
Revised : 10 March 2026
Accepted : 18 May 2026

Keywords

Skin Cancer,
Melanoma,
Early Detection,
DCNN,
LSTM,
BiLSTM

Abstract:

Recent advancements in dermatology have improved the diagnosis and treatment of skin cancer. Early detection is particularly important, especially for melanoma, which is highly aggressive and can metastasize rapidly. However, diagnosing melanoma early is challenging due to its resemblance to atypical moles. While dermatologists assess lesions visually, microscopic examination is required for uncertain cases. The ABCD criteria are commonly used but may miss some melanomas. Dermoscopy offers greater accuracy and is preferred for reliable detection. AI and deep learning are revolutionizing dermatology, particularly in malignant melanoma diagnosis. This paper presents an AI system designed for accurate melanoma detection by incorporating a Deep Conventional Neural Network with transfer learning, data augmentation, and hybrid DCNN-LSTM and CNN-BiLSTM models. These techniques improve the performance of melanoma classification models. Experimental results demonstrate that the proposed methods surpass CNN approaches in accuracy, reaching 97.43%, specificity 99.6%, and, most importantly, sensitivity. This last metric, which represents the number of correctly identified malignant images, reaches 97.4% with the MobileNetV2-BiLSTM model.

1. Introduction

Dermatology, which encompasses the study of over 4,000 skin diseases, represents 15% to 30% of outpatient care consultations and calls on many resources for diagnosis, treatment, and aesthetics [1]. The rising prevalence of skin lesions, including eczema, psoriasis, and skin cancer, has driven advancements in dermatological treatments and technologies, notably through the integration of artificial intelligence [2]. Skin lesions can highlight both systemic and localized diseases, making certain diagnostics easier. However, the diagnostic process is often challenging, as different conditions

may produce similar lesions, while the same disease can manifest through various lesion types due to the limited range of cutaneous presentations [3, 4]. Skin cancers, including melanoma or non-melanoma, which is commonly used to designate both basal cell carcinoma (BCC) and squamous cell carcinoma (SCC) [5, 6]. Melanoma is a dangerous cancer that can lead to serious health complications if not diagnosed and properly treated early. Conversely, non-melanoma skin cancer has a favorable prognosis. BCC does not metastasize, whereas SCC can spread to the lymph nodes. It progresses in stages from a superficial lesion to skin invasion [6]. Conversely, malignant melanoma is one of the most serious types of skin cancers due to its high metastatic potential. It can quickly progress to organs such as the lungs, liver, or brain [7], which complicates treatment and significantly affects outcome.

Early diagnosis of melanoma provides effective surgical intervention, with a significant improvement in the chances of survival at five and ten years, which exceed 90% [8]. However, early detection remains challenging due to its resemblance to an atypical nevus. Dermatologists can diagnose melanoma by visual examination of the lesion, but if there is any uncertainty, microscopic analysis is required. The ABCD criteria [9] (Asymmetry, Border irregularity, Color variation, and large Diameter) are widely used to evaluate melanoma. Some forms of malignant melanoma, though, evade the ABCD criteria because of their similarity to benign moles [10].

Technological progress in medicine has given doctors advanced tools, such as deep learning and artificial neural networks with multiple layers. To detect skin cancer and improve the accuracy of malignant melanoma classification, enhancing early intervention and treatment outcomes. Although Deep Convolutional Neural Networks (DCNNs) are used to diagnose skin cancer with an accuracy comparable to that of specialist dermatologists, their training requires a large dataset of high-quality dermoscopic images. Datasets such as HAM10000 [11], BCN20000 [12], ISIC 2018 [13], and ISIC 2020[14], containing dermoscopic images of skin cancers, have become essential references for research into the classification of skin lesions, particularly skin cancer. A high level of diagnostic precision is crucial to confirm the detection of skin cancer and minimize medical errors. A major problem for accurate melanoma classification in the datasets cited above is class imbalance, where malignant lesions are significantly underrepresented compared to benign lesions. These limitations impact the performance of AI-driven diagnostic models. Several studies have examined the issue of class imbalance in skin cancer classification datasets. First approach [15] proposed merging the ISIC 2020 and ISIC 2019 datasets to increase sample diversity. The ISIC 2019 dataset consists of dermoscopic images categorized into nine diagnostic classes. These include pre-cancerous lesions such as actinic keratosis, as well as malignant cancers like melanoma, BCC, and SCC. The actinic keratosis is a precancerous lesion that can develop into SCC [16]. However, this diversity presents a significant challenge: by training a model on multiple lesion categories, the increased inter-class variability can potentially diminish its performance in detecting early-stage melanoma [17].

Conversely, the ISIC 2020 dataset focuses exclusively on the aggressive form of skin cancer, named melanoma, allowing for greater model specialization. In another study [18], multiple ISIC dataset archives were used, but the model was trained on only 243 melanoma images, leading to reduced accuracy. This limitation arose because the dataset lacked sufficient malignant melanoma samples, which are crucial for improving classification performance, particularly in detecting subtle variations within this category.

This paper proposes a robust and adaptable model for accurately classifying malignant melanoma. To achieve this, various DCNN architectures are explored using transfer learning and data augmentation methods, with a focus on hybrid models that combine Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BiLSTM) with DCNN. Specifically, DCNN-LSTM and DCNN-BiLSTM. These architectures enhance performance and enable high-precision early detection of skin cancer, potentially improving melanoma detection and patient survival rates.Our contributions focus on several key aspects. First, we address class imbalance during data preprocessing by applying data augmentation techniques. Second, we introduce two models based on transfer learning, leveraging the pre-trained VGG16 and MobileNetV2 models. To further optimize performance, we integrate the strengths of DCNNs with recurrent networks (LSTM and BiLSTM) in a hybrid approach. Finally, our models are rigorously evaluated on the ISIC 2020 dataset, ensuring a robust validation of our methodology.

The rest of the paper is organized as follows: Section 2 presents the related work, providing an overview of existing studies on melanoma classification and deep learning models. Section 3 outlines the methodology, starting with the dataset and data preprocessing, which covers data cleaning and augmentation with splitting. It then introduces the hybrid DCNN-LSTM/BiLSTM architecture and

concludes with results and discussions, analyzing performance metrics. Section 4 provides an evaluation and comparative study, benchmarking the proposed approach against related research methods. Finally, Section 5 concludes this paper, summarizing key findings and suggesting future research projects.

2. Related Works

Numerous studies have explored automated skin cancer detection using deep learning. Traditional methods relied on handcrafted features, but recent advancements in CNNs and hybrid architectures, such as CNN-LSTM and CNN-BiLSTM, have significantly improved classification accuracy. Researchers have tackled challenges like class imbalance, dataset diversity, and feature extraction to enhance model performance.

Mahmoud et al. [15] employed various CNN model architectures configured across four distinct layers. Additionally, a transfer learning approach is adopted, leveraging a robust pre-trained model, like the ResNet50. This model has a consistently higher performance than other architectures in transfer learning, achieving an accuracy rate of 92.98%.

Abir et al. [18] focused on noise reduction, ROI extraction, and data enhancement techniques for data pre-processing. They utilized a ResNet-based CNN for lesion classification, achieving an accuracy of 93%. Expanding on advanced neural networks, Qasim Gilani et al. [19] introduced spiking-deep neural networks (S-DNN) with a surrogate gradient descent method to classify melanoma and non-melanoma classes from the ISIC 2019 dataset. Their proposed model, named Spiking VGG-13, achieved an accuracy of 89.57%.

Yashwant S. Ingle [20] compared the classification of skin cancer using a custom CNN versus VGG16 with transfer learning on the HAM10000 dataset. The results indicated that the VGG16 pre-trained model, augmented by the inclusion of more dense layers and dropout regularization layers, performed better than the CNN model with 89% accuracy.

Imran et al. [21] improved skin cancer classification with EfficientNetB0 and Ant Colony Optimization (ACO) as feature selection. Their CB-SVM model reached more than 98% accuracy on a self-created ISIC dataset.

Kaur et al. [22] aimed to produce a lightweight and less complex DCNN for efficient melanoma classification. The study utilized images from the ISIC 2016, ISIC 2017, and ISIC 2020 datasets. The model was tested and evaluated using accuracy, precision, sensitivity (Recall), specificity, and F1-score metrics. The classifier attained accuracies equal to 81.41%, 88.23%, and 90.42% on the ISIC 2016, ISIC 2017, and ISIC 2020 datasets, respectively.

The proposed method in [23] is based on lung cancer classification using a dataset of 3600 images with a resolution of 224×224 pixels, divided into two classes. A CNN with fully connected layers (FC layers) was employed, resulting in an accuracy of 86.23% with efficient computations. Lastly, in [24], the authors addressed the issue of class imbalance in skin cancer classification by applying data pre-processing and data augmentation techniques. The pre-trained VGG-19 and MobileNetV2 architectures were fine-tuned using transfer learning with the SIIM-ISIC 2020 dataset, yielding a model with 95.16% accuracy in image feature extraction and classification.

Table 1 summarizes related works, detailing their approaches, datasets, and results to provide a concise overview of skin cancer classification methods.

3. Methodology

Dermoscopy is vital for evaluating and classifying skin cancer, aiding in early detection and the differentiation of benign and malignant lesions. Our methodology comprises four phases: dataset preparation, preprocessing, feature extraction, and fine-tuning pre-trained models to optimize performance on the ISIC 2020 dataset.

3.1. Dataset

The proposed approach is based on the authorized SIIM-ISIC 2020 dataset, which contains 33126 dermoscopic images. This dataset exhibits a substantial class imbalance, consisting of 32542 benign lesions and only 584 malignant melanomas. An overview of the dataset's characteristics is provided in Table 2. The significant disparity in class distribution, visually emphasized in Figure 1, is a crucial consideration for model training and is addressed in this work.

3.2. Data preparation

In the context of machine learning or any predictive process, the "target" refers to the variable we aim to predict or classify. One noteworthy peculiarity of the used ISIC 2020 dataset is the significant class imbalance, as shown in Figure 1. In Figure 2, the circular chart depicting the distribution of “patient_id” reveals that each patient has one or more images of benign lesions, with a maximum of 114 images per patient. As a result, some patients have a considerably higher number of images than

the average, while others have only one. To address this, an initial preprocessing step is required to eliminate duplicate images, reducing the number of unique benign lesion images to 6271 [24].

An imbalanced class distribution can negatively influence the performance of classification models during both training and evaluation tests. We addressed this by applying image data augmentation to the malignant class, using TensorFlow [25], which makes available the ImageDataGenerator class [26]. This allowed us to introduce transformation parameters such as rotation range, width shift range, shear range, horizontal, and vertical_flip, which effectively increased the diversity of the dataset and improved the robustness and generalisation of the model, which is crucial when dealing with lesion variations in dermatology.

Figure 3 summarizes the data preprocessing pipeline, which included removing duplicates from the benign class, splitting the dataset, and performing data augmentation on the malignant melanoma class to fix the class imbalance problem. Before data augmentation, we split the dataset into a training set (60%), a validation set (20%), and a test set (20%) to ensure rigorous model development and evaluation. At the end of the data augmentation phase, the images from the benign and malignant classes are combined to form the final dataset. This dataset consists of 6,271 benign images and 4796 malignant images distributed as follows: 4680 for the training set, and the validation set, and 116 original images for the test set.

Reducing the number of benign images to balance the classes is not an adequate solution, as deep learning models [27] require a large volume of data to ensure reliable and efficient classification. Hence, it won't train well only on a low number of images and will cause inevitable overfitting and limited generalization. In our case, it is not feasible to train the model directly with the original images, as the dataset includes only 352 malignant melanoma images in the training set and just 116 in the validation set. This is a significantly smaller number than the number of benign images (After removing duplicates, 4 271 images were used for training and validation set). The effect of data augmentation on a few melanoma sample photos is shown in Figure 4. The augmented lesion samples, which were produced by transformations like rotation (0.5), width shifts (0.4), shear transformations (0.1), and flipping horizontally and vertically, are displayed next to the original lesion samples.Our first goal is to use data augmentation techniques to resolve the problem of data imbalance. We assessed the images produced by the image generator using the LPIPS metric, which produced a score of 0.4, to make sure they were of a quality that was satisfactory when compared to the reference images. Better perceptual quality is indicated by lower values for the LPIPS metric, ideally around 0.2 [28, 29].

Furthermore, LPIPS values of about 0.4 are typically interpreted as indicating satisfactory, or moderate-to-good, similarity rather than poor performance in the context of medical image analysis, especially in the analysis of skin cancer [30].

3.3. Hybrid DCNN- LSTM/BiLSTM Network Architecture

In the proposed approach, we utilize a hybrid architecture that integrates a DCNN with an LSTM or BiLSTM layer. This architecture demonstrates the combination of Convolutional neural network features extraction power with LSTM's ability to capture temporal or sequential dependencies. The result is a more robust model capable of handling both spatial and sequential data, improving performance in tasks that involve time-series or sequence-based patterns.

Figure 5 presents the Hybrid VGG16-LSTM/BiLSTM and MobileNetV2-LSTM/BiLSTM Architectures. It begins by loading VGG16 [31] and also MobileNetV2 [32], pre-trained models, and removing their final classification layer to use them as a feature extractor. The input shape (224, 224, 3) defines the input image size as 224x224 pixels with 3 color channels. We have either a Global Average Pooling layer in the output of the convolution layers, this layer converts the features that have been extracted into a one-dimensional vector. The LSTM/BiLSTM layer is then added, reshaping the feature vector to be compatible with LSTM processing, thereby forming a structure of (batch_size, time_steps, and features). Finally, L2 regularization is applied to help prevent overfitting.

The output layers of the LSTM/BiLSTM are then passed to three additional dense, fully connected layers (FC layers), which perform the final classification. Each FC layer is accompanied by a dropout layer to regularize the model and avoid the overfitting problem. The final layer applies a Sigmoid activation function to produce a binary probability output, indicating the predicted class: 0 is benign and 1 is malignant melanoma.

3.4. Evaluation Method and Criteria

Model performance was evaluated using a comprehensive set of metrics: accuracy (ACC), specificity (SPE), sensitivity (SEN), and

precision (PRE), as defined in Equations (1) to (4). These metrics were derived from the confusion matrix [33], a fundamental tool for classification assessment that compares actual and predicted labels. The matrix encapsulates four key outcomes: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), which collectively provide a detailed breakdown of model predictive behavior.

$$ACC = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$SPE = \frac{TN}{TN+FP} \quad (2)$$

$$SEN = \frac{TP}{TP+FN} \quad (3)$$

$$PRE = \frac{TP}{TP+FP} \quad (4)$$

4. Results and Discussion

The proposed models were developed using the Python language with the TensorFlow and Keras frameworks. The training configuration employed a batch size of 64 and a learning rate of 0.0001. To address potential class imbalance, we utilized a Binary Focal Cross-Entropy loss function in conjunction with the Adam optimizer.

4.1. Training and validation results

The loss curves for the VGG16-LSTM and VGG16-BiLSTM hybrid models over 100 training epochs are shown in Figures 6 and 7, respectively. There are two curves on each plot: one for the training set and one for the validation set. In both cases, the training loss drops quickly from the first few iterations, which shows that the model is converging well on the training data. The validation loss, on the other hand, acts differently for each architecture, showing that their ability to generalize is different. Figures 8 and 9 show the loss curves for the MobileNetV2-LSTM and MobileNetV2-BiLSTM models, respectively.

The VGG16-LSTM model exhibited a steady decline in training loss, decreasing from approximately 0.5 to 0.1 over 100 epochs. However, it displayed clear signs of overfitting, indicated by a sharp increase in validation loss after 20 epochs following an initial period of stability. This suggests the model memorized the training data, but has failed to generalize to new unseen instances. In contrast, the VGG16-BiLSTM model demonstrated a different learning pattern. Although its training loss decreased more rapidly (from 1.75 to 0.25), it also showed signs of overfitting, with validation loss decreasing initially, stabilizing, and then rising slightly with fluctuations. Despite this imperfect generalization, its behavior indicates a more favorable trade-off between learning and validation performance compared to the VGG16-LSTM model.The MobileNetV2-LSTM and MobileNetV2-BiLSTM models demonstrated similar training dynamics, characterized by a rapid decline in loss during the initial epochs and stabilization after approximately ten epochs. The strong concordance between the training and validation curves signifies that the models learned efficiently, avoided overfitting, and maintained a high capacity for generalization.

VGG16-based models suffer from overfitting and require regularization, though VGG16-BiLSTM is more stable than VGG16-LSTM. Conversely, MobileNetV2-based models generalize well, evidenced by rapid convergence and aligned loss curves. Despite a higher initial validation loss, MobileNetV2-BiLSTM does not consistently outperform MobileNetV2-LSTM, necessitating further comparison using metrics like accuracy.

As shown in Figure 10, the MobileNetV2-based hybrid models achieve higher classification accuracy than their VGG16-based counterparts. This superior performance is attributed to their faster convergence and stronger generalization capabilities, which mitigate the overfitting that limits the VGG16 models. The results indicate that MobileNetV2 provides a more effective foundation for feature extraction and temporal modeling when integrated with LSTM or BiLSTM layers. Consequently, the following section presents test results to further validate the two MobileNetV2-based models.

4.2. Evaluation Test and Comparative Study

This section evaluates the model's classification accuracy through an analysis of the test results and corresponding confusion matrices. As a summarized performance overview of a classification model, the confusion matrix serves as a fundamental tool for quantitative assessment [33]. Figure 11 displays the confusion matrices for both the MobileNetV2-LSTM and MobileNetV2-BiLSTM models, enabling a direct visual comparison of their predictive behavior.

From these matrices, key classification metrics were derived. Specificity (TN rate), which reflects the model’s ability to correctly identify benign cases, reached 98.6% for the MobileNetV2-LSTM model and 99.6% for the MobileNetV2-BiLSTM model [33, 34]. More critically, sensitivity (true positive rate)—indicating the correct identification

of malignant melanoma cases—was 94.8% for MobileNetV2-LSTM and 97.4% for MobileNetV2-BiLSTM. Although the latter demonstrates improved performance, it still exhibited a false negative rate of 2.6%, corresponding to three missed malignant cases, suggesting a residual risk in clinical application.

Table 3 also provides a summary of the results on the ISIC 2020 dataset, comparing the performance of the proposed methods with state-of-the-art DNN approaches for malignant melanoma classification. We evaluated the test results using four key metrics previously mentioned. Overall, by leveraging data augmentation and combining the strengths of MobileNetV2 with LSTM/BiLSTM, our approach outperforms existing methods in detecting malignant lesions.

In comparison, other models, while effective, show slightly lower performance, especially in terms of sensitivity, such as those of [15] (95.08% sensitivity) and [22] (90.4% for all metrics). All the proposed models, particularly with models such as those by [24], MobileNetV2-LSTM, and MobileNetV2-BiLSTM, display strong accuracy, precision, and specificity values, demonstrating their ability to correctly identify benign images. However, for detecting the majority of malignant melanoma cases, only the hybrid MobileNetV2-BiLSTM-based model stands out with an optimal sensitivity of 97.4%.

***Table 1**. Overview of Related Deep Learning Models for Skin Cancer Classification.*

Reference	Methodology	Dataset	Main Results
Mahmoud et al. [15] (2024)	CNN architectures with transfer learning (Efficient CNN and deep transfer learning)	ISIC 2019, ISIC 2020	Accuracy: 92.98%, Sensitivity: 92.46%, Precision: 95.08%
Abir et al. [18] (2024)	Data preprocessing (noise reduction, ROI cutting, data augmentation), ResNet CNN	ISIC archives	Accuracy: 93%, F1 score: 90%, Sensitivity: 91%, Precision: 89%
Qasim Gilani et al. [19] (2023)	Spiking Deep Neural Networks with Surrogate Gradient Descent	ISIC 2019	Accuracy: 89.57%, F1 score: 90.07%
Yashwant S. Ingle [20] (2024)	Custom CNN vs VGG16 with transfer learning	HAM10000	VGG16 Accuracy: 89%, F1 score: 89%
Imran et al. [21] (2024)	EfficientNetB0 with Ant Colony Optimization (ACO) for feature selection	Self-created ISIC archives	Accuracy: >98%
Kaur et al. [22] (2022)	Lightweight DCNN for melanoma classification	ISIC 2016, 2017, 2020	Accuracy: 90.42%, Sensitivity: 90.3%, Precision: 90.4%
Atta et al. [23] (2022)	CNN with fully connected layers for skin cancer classification	ISIC 2016	Accuracy: 86.23%
M'hamedi et al. [24] (2024)	Data preprocessing and augmentation with transfer learning (VGG-19, MobileNetV2)	ISIC 2020	Accuracy: 95.16%, Sensitivity: 90.83%, Specificity: 99.2%, AUC: 97.57%, Precision: 99.06%

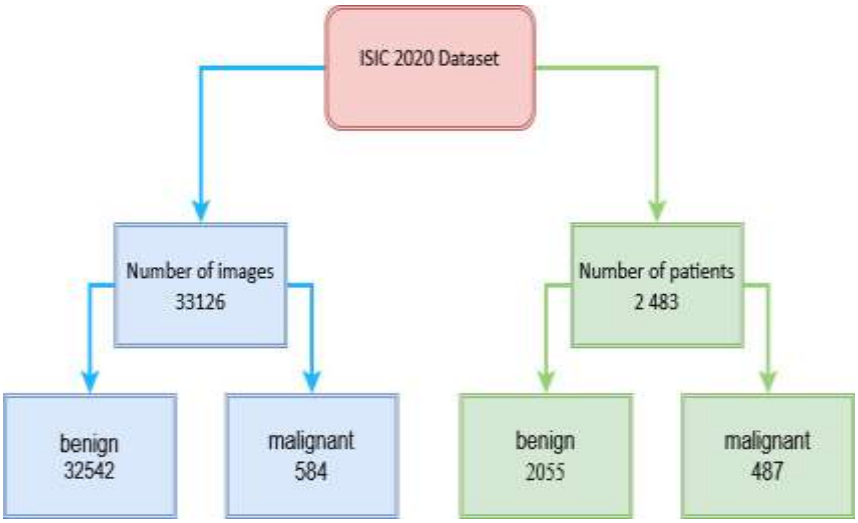


***Figure 1.** Class Distribution in the ISIC 2020 Dataset*

Table 2: *Key Features of the ISIC 2020 dataset.*

Attribute	Description
"image_name"	The file name of the image
"patient_id"	The unique identifier of the patient
"age_approx"	The approximate age of the patient
"anatom_site_general_challenge"	The anatomical location of the image on the body
"diagnosis"	Label indicating whether the lesion is malignant or benign
target	Binary target variable (0 = benign, 1 = confirmed melanoma)

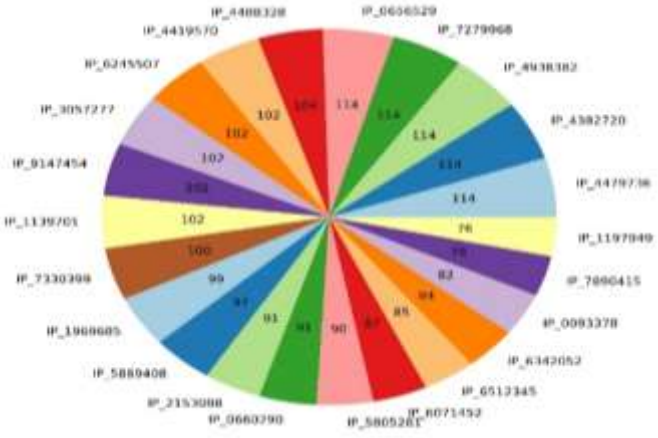


Figure 2. *Features visualization: distribution of "image_name" per "patient_id" (for counts > 75)*

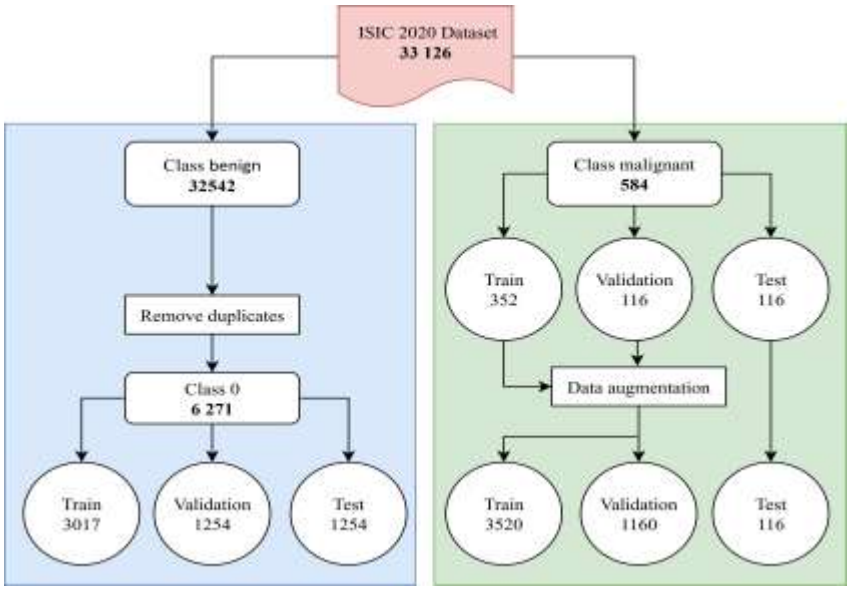


Figure 3. *Data preparation process*

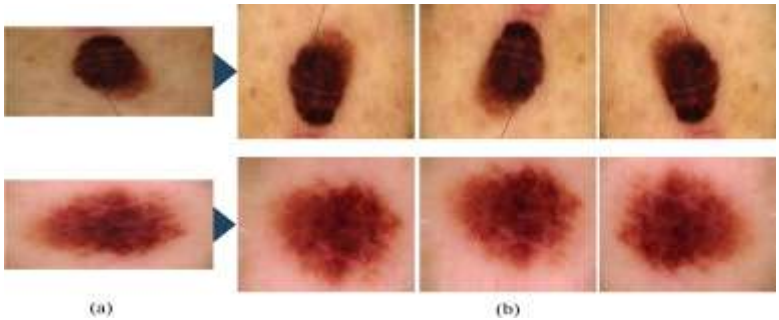


Figure 4. *Data Augmentation Applied to Melanoma Images (a): Original images, and (b): Augmented images*

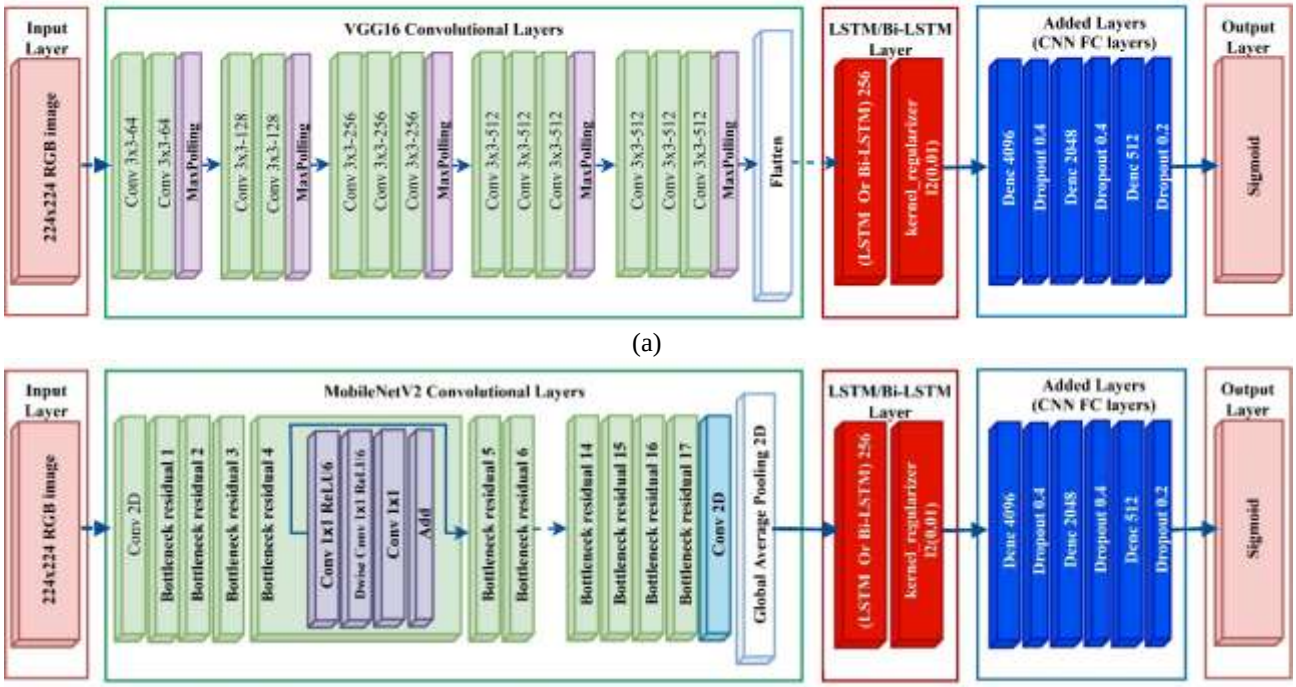


Figure 5. *Illustration of the proposed Hybrid Architectures: (a): VGG16- LSTM/BiLSTM (b): MobileNetV2-LSTM/BiLSTM*

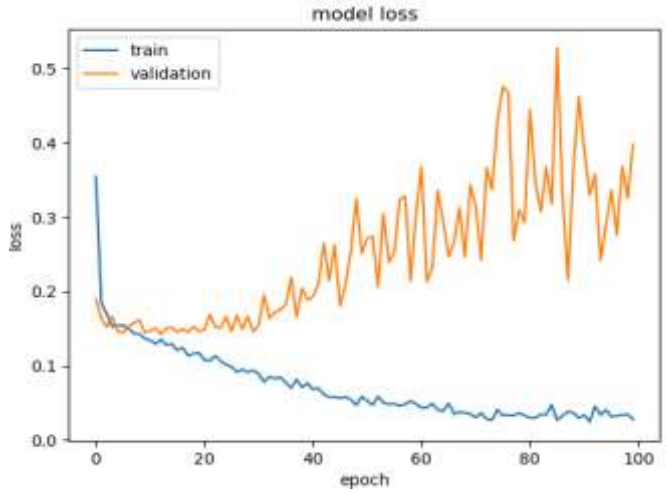


Figure 6. *Loss vs. Epochs of the hybrid VGG16-LSTM model*

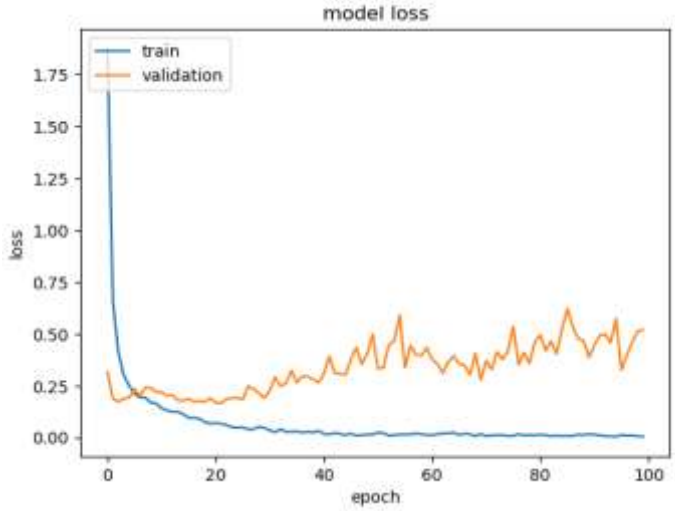


Figure 7. *Loss vs. Epochs of the hybrid VGG16-BiLSTM model*

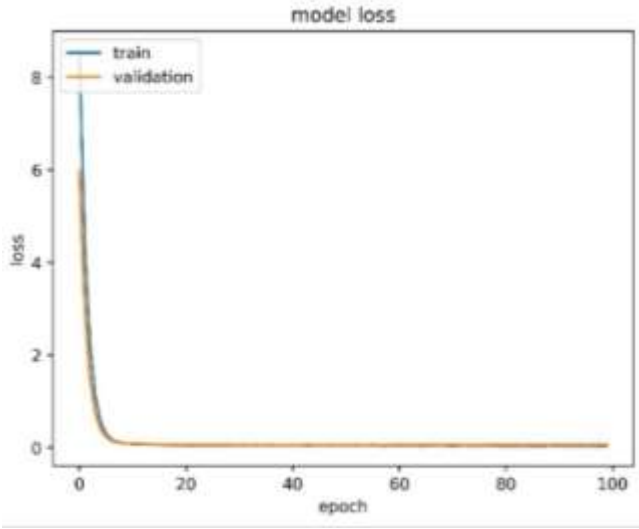


Figure 8. *Loss vs. Epochs of the hybrid MobileNetV2-LSTM model*

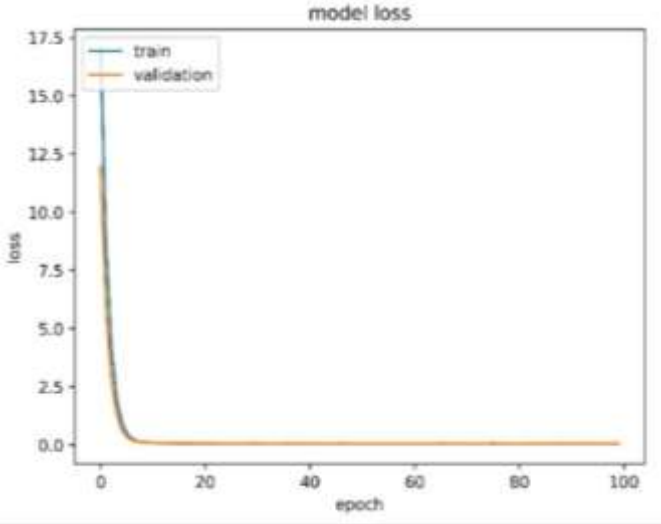


Figure 9. *Loss vs. Epochs of the hybrid MobileNetV2-BiLSTM model*

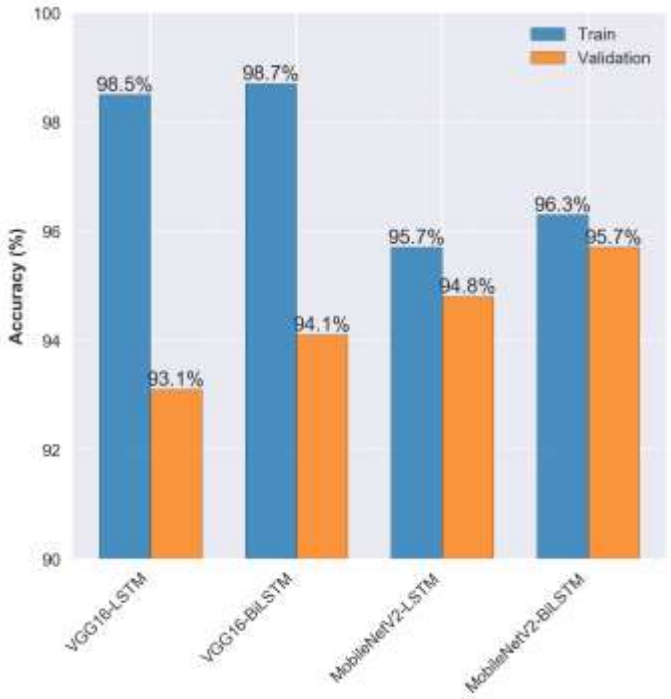
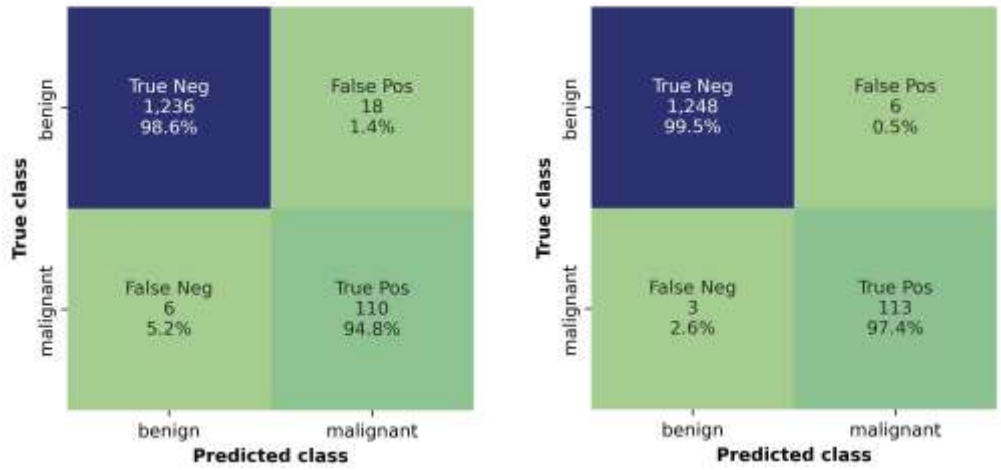


Figure 10. *Training Vs. Validation Accuracy by Model*



(a) (b)

Figure 11. *Confusion matrix for (a): MobileNetV2-LSTM, and (b): MobileNetV2-BiLSTM*

Table 3: *State-of-the-Art Performance Comparison of Different Models for Melanoma Detection.*

	ACC (%)	**PRE (%)**	**SEN (%)**	**SPE (%)**
Kaur et al. [22]	90.4	90.4	90.4	--
M'hamedi et al. [24]	95.16	99.06	90.83	99.20
Abir et al. [18]	93	89	91	--
Mahmoud et al. [15]	92.98	92.46	95.08	--
MobileNetV2-LSTM	**95.67**	**92.91**	**94.80**	**98.60**
MobileNetV2-BiLSTM	**97.43**	**94.96**	**97.4**	**99.6**

5. Conclusions

This paper has focused on utilizing CNNs and transfer learning with the pre-trained VGG16 and MobileNetV2 models for melanoma image classification. The LSTM/BiLSTM hybrid models based on VGG16 have been prone to overfitting and have trouble generalizing to new data. But the hybrid MobileNetV2-LSTM/BiLSTM architecture has demonstrated outstanding results in these evaluations. By incorporating techniques like data augmentation, these methods have effectively combined the strengths of MobileNetV2 and LSTMs, achieving exceptional performance across all metrics. As a result, this approach has provided a highly robust solution for melanoma diagnosis, enhancing the quality of patient care. Sensitivity, or the True Positive (TP) rate, has been particularly crucial in cancer detection, as malignant melanoma has often been challenging to differentiate from benign cases like moles. Maximizing this rate has been essential to reducing medical errors and saving lives. The MobileNetV2-BiLSTM model has outperformed related techniques, achieving a sensitivity of 97.4%, reinforcing its effectiveness for melanoma case detection. Moving forward, future research should explore deep reinforcement learning to optimize decision-making and enhance model adaptation using real-time clinical feedback.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The

data are not publicly available due to privacy or ethical restrictions.

- **Use of AI Tools:** The author(s) declare that no generative AI or AI-assisted technologies were used in the writing process of this manuscript.

References

[1] A. Alani, M. Sadlier, A. Uddin, C. Hackett, B. Ramsay, and K. Ahmad, "An analysis of inpatient dermatologic consultations at University Hospital Limerick: Inadequate infrastructure leads to acute skin failure," *Irish Journal of Medical Science*, vol. 186, no. 2, pp. 305–307, 2014. https://doi.org/10.1007/s11845-016-1424-8

[2] L. C. Parish, "Dermatologic therapy for the 21st Century," *Clinics in Dermatology*, vol. 18, no. 2, pp. 147, 2000. https://doi.org/10.1016/S0738-081X(00)00117-6

[3] E. M. Miró and N. P. Sánchez, "Cutaneous manifestations of infectious diseases," in *Atlas of Dermatology in Internal Medicine*, New York: Springer, 2011, pp. 77–119. https://doi.org/10.1007/978-1-4614-0688-4_7

[4] **National Institute of Arthritis and Musculoskeletal and Skin Diseases,** "Skin Diseases," National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), [Online]. Available: https://www.niams.nih.gov/health-topics/skin-diseases. [Accessed: Feb. 01, 2025].

[5] D. S. Rigel, R. J. Friedman, L. M. Dzubow, D. S. Reintgen, J.-C. Bystryn, and R. Marks, *Cancer of the Skin*, 2nd ed. Philadelphia, USA: Elsevier Saunders, 2005.

[6] C. Karimkhani, L. N. Boyers, R. P. Dellavalle, and M. A. Weinstock, "It's time for 'keratinocyte carcinoma' to replace the term 'nonmelanoma skin cancer'," *Journal of the American Academy of Dermatology*, vol. 72, no. 1, pp. 186–187, 2015. https://doi.org/10.1016/j.jaad.2014.09.036

[7] M. Ernst and A. Giubellino, "The current state of treatment and future directions in cutaneous malignant melanoma," *Biomedicines*, vol. 10, no. 4, pp. 822, 2022. https://doi.org/10.3390/biomedicines10040822

[8] C. M. Balch, A. C. Buzaid, S. J. Soong, et al., "Final version of the American Joint Committee on Cancer Staging System for cutaneous melanoma," *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, vol. 19, no. 16, pp. 3635–3648, 2001. https://doi.org/10.1200/JCO.2001.19.16.3635

[9] C. Papageorgiou, D. Ioannides, Z. Apalla, E. Vakirlis, E. Lazaridou, E. Sotiriou, and A. Lallas, "Dermoscopy of difficult-to-diagnose melanomas," *Serbian Journal of Dermatology and Venereology*, vol. 8, no. 3, pp. 121–127, 2016. https://doi.org/10.1515/sjdv-2016-0011

[10] C. Longo, R. Pampena, E. Moscarella, et al., "Dermoscopy of melanoma according to different body sites: Head and neck, trunk, limbs, nail, mucosal and acral," *Journal of the European Academy of Dermatology and Venereology*, vol. 37, no. 9, pp. 1718–1730, 2023. https://doi.org/10.1111/jdv.19221

[11] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, no. 1, pp. 1–9, 2018. https://doi.org/10.1038/sdata.2018.161

[12] C. Hernández-Pérez, M. Combalia, S. Podlipnik, et al., "BCN20000: Dermoscopic lesions in the wild," *Scientific Data*, vol. 11, no. 1, pp. 641, 2024. https://doi.org/10.1038/s41597-024-03387-w

[13] N. Codella, V. Rotemberg, P. Tschandl, et al., "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the International Skin Imaging Collaboration (ISIC)," *arXiv preprint*, 2019. https://arxiv.org/abs/1902.03368

[14] V. Rotemberg, N. Kurtansky, B. Betz-Stablein, et al., "A patient-centric dataset of images and metadata for identifying melanomas using clinical context," *Scientific Data*, vol. 8, no. 1, pp. 34, 2021. https://doi.org/10.1038/s41597-021-00815-z

[15] H. Mahmoud, O. A. Omer, S. Ragab, H. Esmaiel, and M. Abdel-Nasser, "Classifying melanoma in ISIC dermoscopic images using efficient convolutional neural networks and deep transfer learning," *Traitement du Signal*, vol. 41, no. 2, pp. 679–691, 2024. https://doi.org/10.18280/ts.410211

[16] A. Fuchs and E. Marmur, "The kinetics of skin cancer: Progression of actinic keratosis to squamous cell carcinoma," *Dermatologic Surgery*, vol. 33, no. 9, pp. 1099–1101, 2007. https://doi.org/10.1111/j.1524-4725.2007.33224.x

[17] R. Pilarczyk and W. Skarbek, "On intra-class variance for deep learning of classifiers," *Foundations of Computing and Decision Sciences*, vol. 44, no. 3, pp. 285–301, 2019. https://doi.org/10.2478/fcds-2019-0015

[18] S. I. Abir, S. Shoha, M. M. Hossain, et al., "Deep learning-based classification of skin lesions: Enhancing melanoma detection through automated preprocessing and data augmentation," *Journal of Computer Science and Technology Studies*, vol. 6, no. 5, pp. 152–167, 2024. https://doi.org/10.32996/jcsts.2024.6.5.13

[19] S. Q. Gilani, T. Syed, M. Umair, and O. Marques, "Skin cancer classification using deep spiking neural network," *Journal of Digital Imaging*, vol. 36, no. 3, pp. 1137–1147, 2023. https://doi.org/10.1007/s10278-023-00776-2

[20] Y. Ingle and N. F. Shaikh, "Deep learning for skin cancer classification: A comparative study of CNN and VGG16 on HAM10000 dataset," *Communications on Applied Nonlinear Analysis*, vol. 31, no. 4s, pp. 490–499, 2024. http://dx.doi.org/10.2139/ssrn.4787463

[21] T. Imran, A. S. Alghamdi, and M. S. Alkatheiri, "Enhanced skin cancer classification using deep learning and nature-based feature optimization," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12702–12710, 2024. https://doi.org/10.48084/etasr.6604

[22] R. Kaur, H. GholamHosseini, R. Sinha, and M. Lindén, "Melanoma classification using a novel deep convolutional neural network with dermoscopic images," *Sensors*, vol. 22, no. 3, pp. 1134, 2022. https://doi.org/10.3390/s22031134

[23] A. Atta, M. A. Khan, M. Asif, G. F. Issa, R. A. Said, and T. Faiz, "Classification of skin cancer empowered with a convolutional neural network," in Proc. International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates, 2022. https://doi.org/ 10.1109/ICCR56254.2022.9995928

[24] M. M'hamedi, M. Merzoug, M. Hadjila, and A. Bekkouche, "Enhancing melanoma skin cancer classification through data augmentation," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 22, no. 5, pp. 1209, 2024. http://doi.org /10.12928/telkomnika.v22i5.26106

[25] TensorFlow Core, "Load and preprocess images," [Online]. Available: https://www.tensorflow.org /tutorials/load_data/images. [Accessed: Feb. 05, 2025].

[26] TensorFlow Core**,** "Data augmentation," [Online]. Available: https://www.tensorflow.org /tutorials/images/data_augmentation. [Accessed: Feb. 05, 2025].

[27] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting unreasonable effectiveness of data in deep learning era," in Proc. IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017. https://doi.org/10.1109/iccv.2017.97

[28] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, USA, 2018. https://doi.org/10.1109/CVPR.2018.00068

[29] M. Aziz, U. Rehman, M. U. Danish, and K. Grolinger, "Global-Local Image Perceptual Score (GLIPS): Evaluating photorealistic quality of AI-generated images," *IEEE Transactions on Human-Machine Systems*, vol. 55, no. 2, pp. 223–233, 2025. https://doi.org /10.36227/techrxiv.173933251. 15884096/v1

[30] M. A. Farooq, A. Abaid, I. Ullah, and P. Corcoran, "A comparative study on diffusion sampling methods across diverse medical imaging modalities," In Proc. Asian Conference on Computer Vision (ACCV), 2024, Hanoi, Vietnam. https://doi.org/10.1007/978-981-96-2641-0_13

[31] U. Muhammad, W. Wang, S. P. Chattha, and S. Ali, "Pre-trained VGGNet architecture for remote-sensing image scene classification," In Proc. 24th International Conference on Pattern Recognition (ICPR), IEEE, 2018, Beijing, China. https://doi.org/10.1109/ICPR.2018.8545591

[32] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, Salt Lake City, USA. https://doi.org/ 10.1109/CVPR.2018.00474

[32] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009. https://doi.org /10.1016/j.ipm.2009.03.002

[33] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, USA: MIT Press, 2016.

[34] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. London, United Kingdom: The MIT Press, 2012.