

SKINVGG-NET: A MODIFIED AND FINE-TUNED VGG19-BASED DEEP LEARNING ARCHITECTURE FOR SKIN CANCER CLASSIFICATION

Maysaa R. Naeemah
Department of Physics,
College of Science, Mustansiriyah University,
Baghdad, Iraq.
maysaa_raba@uomustansiriyah.edu.iq

Mohammed Y. Kamil
Department of Physics,
College of Science, Mustansiriyah University,
Baghdad, Iraq.
m80y98@uomustansiriyah.edu.iq

Submitted: Mar, 28, 2025 **Revised:** Jun, 03, 2025 **Accepted:** Jun, 19, 2025

Abstract: Skin cancer, one of the most common and potentially fatal cancers, requires early and correct diagnosis to improve patient outcomes. While dermatologists possess extensive diagnostic expertise, recent studies have shown that Convolutional Neural Networks (CNNs) can often surpass human performance in multiclass skin lesion classification, owing to their ability to extract subtle and complex features that may be overlooked during manual examination. In this study, SkinVGG-Net, an enhanced deep learning framework based on the VGG19 architecture is proposed. The SkinVGG-Net uses transfer learning and an advanced fine-tuning strategy, along with a comprehensive data preprocessing pipeline that includes image normalization, resizing, and extensive data augmentation to address the class imbalance challenges. The SkinVGG-Net model achieves 90.95% accuracy on HAM10000 dataset, with high performance across seven types of skin cancer. These results demonstrate the capability of CNN-based systems to provide consistent, accurate diagnostic support and highlight their promise as assistive tools in clinical settings with greater precision than traditional visual assessment methods.

Keywords: Deep learning, transfer learning, skin cancer, AI in healthcare, medical imaging.

I. INTRODUCTION

Skin cancer is the most common form of cancer worldwide and it requires early diagnosis to reduce mortality [1]. It results from the abnormal proliferation of epidermal cells, leading to either benign or malignant tumors. Benign tumors consist of normal cells that multiply without disrupting the functions of surrounding healthy tissues [2], whereas malignant tumors, such as melanomas, involve abnormal cell growth that interferes with normal cellular function and proliferate uncontrollably [3].

Traditionally, physical examination and biopsy are employed to diagnose skin cancer, wherein tissue samples from suspicious areas are analyzed in a laboratory [4]. This diagnostic process, though accurate, typically requires about a week to yield definitive results [5]. While the conventional approach is time-consuming, early detection can lead to a treatment success rate of approximately

98% [6]. To expedite diagnosis and improve accessibility, automated detection using Deep Learning (DL) methods has gained prominence [7].

Deep CNNs can be particularly suitable for analyzing skin lesion images as they perform very well in the fields of image recognition, object detection, and semantic segmentation making them [8-9]. Recent studies have explored the application of advanced CNN architectures for classifying dermoscopic images of skin lesions [10]. A GPU-based training is employed in [11] to accelerate the learning process while enhancing image quality through preprocessing and applying data augmentation techniques such as horizontal and vertical flipping to mitigate class imbalance [12-13].

In [14], a deep learning approach for skin cancer classification highlights the potential of CNN models on standard datasets. An integrated diagnostic model in [15] combined lesion segmentation and classification stages. The segmentation phase employed a full-resolution CNN to isolate lesion boundaries, followed by classification using CNN architectures like ResNet50, Inception-v3, Inception-ResNet-v2, and DenseNet201, with CNNs demonstrating strong performance.

A three-step super-pixel segmentation method was proposed in [16] to enhance lesion boundary detection. Following segmentation, ResNet50 was applied to the processed images, and features were further optimized using an enhanced grasshopper optimization algorithm. The optimized features were then classified using a Naive Bayes (NB) classifier. Similarly, the study in [17] tackled imbalanced multiclass classification by applying image preprocessing techniques such as augmentation, hair removal, and resizing to meet model-specific input requirements. They implemented transfer learning and fine-tuned various EfficientNet models (B0 to B7) to provide better performance. A multi-label skin cancer classification system was developed in [18] using a seven-point checklist combined with ProtoPNet architecture.

In the present study, SkinVGG-Net, a CNN-based skin lesion classification system is proposed based on the VGG19 architecture. The model is trained using the HAM10000 dataset, which includes seven skin cancer types. To overcome the significant class imbalance inherent in the dataset, an issue that negatively impacts multiclass classification—data augmentation techniques were applied. Furthermore, multiple combinations of hyperparameters were explored to identify the optimal training configuration, enhancing the model's classification performance.

II. MATERIALS AND METHODS

VGGNet, a CNN architecture introduced in 2014 by the Visual Geometry Group [19] for image classification tasks, has gained widespread popularity due to its simplicity. The architecture has two primary variants: VGG16 and VGG19, which differ in the number of layers and overall depth. Among them, VGG19 offers a deeper network structure compared to VGG16, providing enhanced feature extraction capabilities by increasing the depth of the network. In this study, the VGG19 architecture is adapted and modified to improve accuracy for skin cancer classification. While VGG19 delivers superior performance due to its greater depth, it also contains significantly more parameters than VGG16. It has approximately 143 million which makes training more computationally intensive. The original VGG19 architecture, composed of 19 layers, is illustrated in Figure 1(a) and the proposed SkinVGG-Net is shown in Figure 1(b).

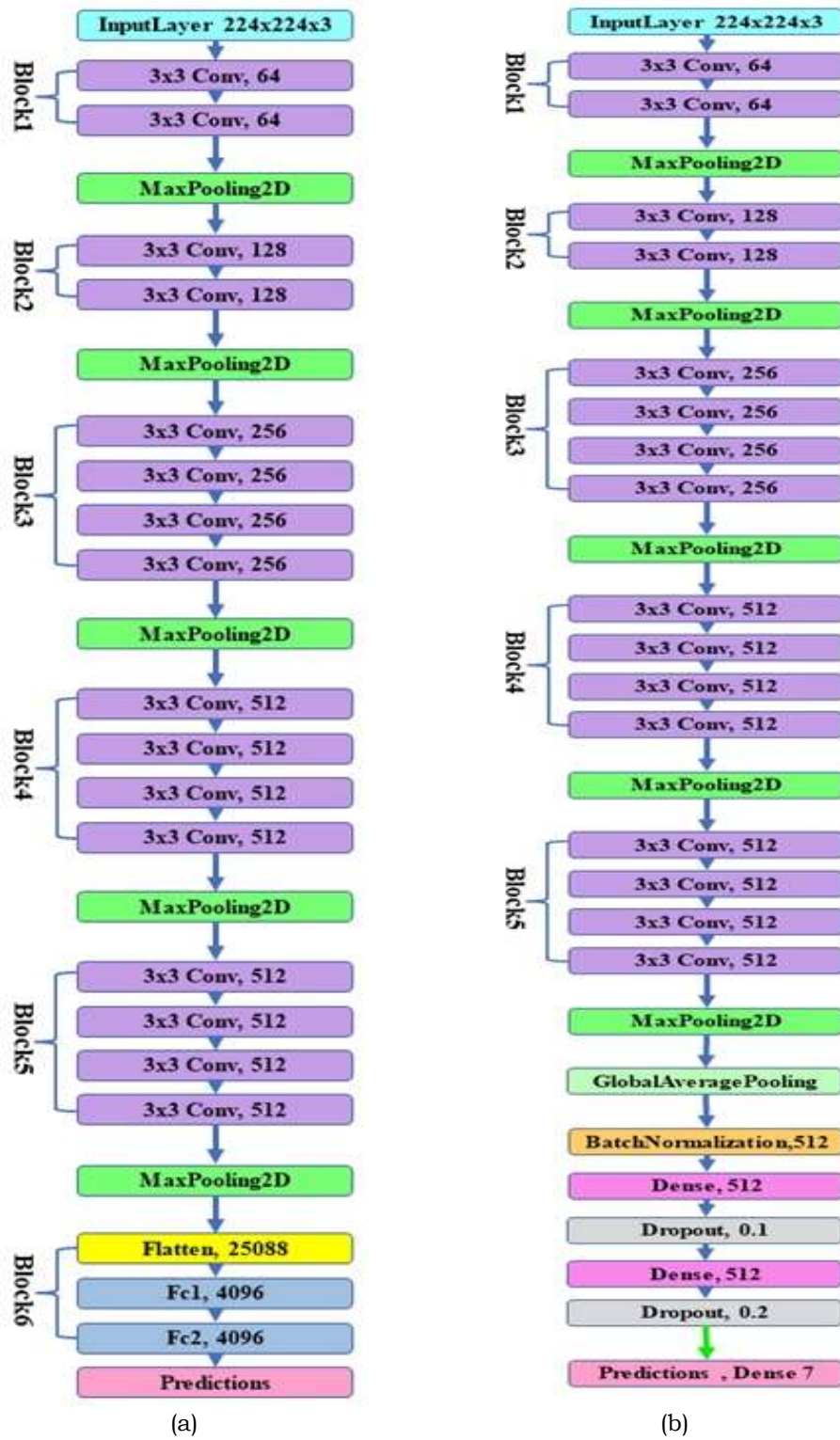


Fig. 1 (a) Conventional VGG19 model (b) Proposed SkinVGG-Net

A. Pre-processing

Although raw data can be directly fed into neural networks, proper data preparation is crucial for enhancing model performance. To enable effective feature extraction, several preprocessing techniques were applied. In accordance with the input requirements of the proposed model, all images in the dataset, originally sized at 450×600 pixels (RGB) were resized to 224×224 pixels using bilinear interpolation, a method that maintains a balance between computational efficiency and image quality.

A significant challenge in medical image analysis is class imbalance, where certain disease categories contain far fewer samples than others. This imbalance can impair a model's ability to generalize, often leading to bias toward more frequent classes. To address this, three main strategies such as data sampling, data augmentation, and class weighting are employed. Among these, data augmentation was particularly effective. Without requiring additional data collection, this technique artificially expands the training set by generating slightly modified versions of existing images [20]. It also serves as a regularization method to prevent overfitting [21]. A diverse set of augmentation techniques was applied, including random rotations (up to 360°), zoom transformations (up to a factor of 0.20), horizontal shifts (up to 20% of the image size), vertical shifts, horizontal and vertical flips, shearing, color jitter, and color space alterations [22].

To ensure class balance, the category with the fewest original samples was augmented until each of the seven lesion classes contained 500 images. This resulted in a final training set of 3,500 images, evenly distributed across all lesion types. This balancing process was essential for preventing the model from developing a bias toward overrepresented classes and promoting fair learning across all categories.

B. SkinVGG-Net model

Figure 1(a) shows the base architecture which is the conventional VGG-19 model. For the proposed SkinVGG-Net, the final three fully connected layers of VGG-19 were removed. In their place, the following layers such as global average pooling, batch normalization layer, two dense layers (each with 512 units) interleaved with dropout layers (with rates of 0.1 and 0.2, respectively) were added. Finally, a dense output layer was added to classify the seven categories of skin diseases, as depicted in Figure 1(b).

Using pre-trained weights during model training significantly enhances the proposed model's generalization ability and diagnostic performance. The model accepts input images of size $224 \times 224 \times 3$, aligning with standard dimensions used in many deep learning applications. To adapt the model for skin disease classification, fine-tuning was employed. A common technique that involves retraining a pre-trained model on a new dataset. This approach is particularly effective when working with limited data by using the knowledge learned from large datasets which contain thousands of image classes such as ImageNet.

C. Hyper-parameter tuning

In order to achieve better performance, the CNN hyperparameters such as batch size, learning rate, optimization algorithm, image resolution, batch normalization, and data augmentation strategies should be tuned carefully. To find the configuration that would work best for the SkinVGG-Net model, a number of tests were carried out. Among the optimization algorithms tested, Adam demonstrated superior performance compared to SGD. The best results discussed

in the results section were obtained by training the SkinVGG-Net model uses a learning rate of 0.001, a batch size of 32 for 80 epochs. To enhance training efficiency and mitigate overfitting, two Keras callbacks such as ModelCheckpoint and EarlyStopping were employed. The model with the highest validation accuracy is saved for further testing using ModelCheckpoint and EarlyStopping was used to terminate training when the validation loss showed no improvement over five consecutive epochs. These techniques effectively improved the model's generalization capability and reduced unnecessary training time. Table 1 summarizes the optimized hyperparameters of the proposed SkinVGG-Net model, determined through validation-based optimization.

TABLE. 1 Hyperparameters of the SkinVGG-Net model

Training parameter	Value
Optimizer	Adam
Learning rate	0.001
Batch size of training data	32
Loss function	cross entropy
β_1 , β_2 , ϵ	0.9, 0.999, 0.1
Split dataset	70% training 30% testing
Early stopping	Monitor = 'loss', patience = 5
Number of epochs	80

III. RESULTS AND DISCUSSIONS

In this section, the results of the proposed SkinVGG-Net model and the findings in the context of existing work were discussed. It begins with an overview of the data set used for training and evaluation, followed by a description of the performance metrics employed to assess model effectiveness. Then, the results achieved by SkinVGG-Net across different lesion categories are reported. Finally, a comparative study is conducted to benchmark the performance of the proposed model against existing deep learning architectures.

A. Dataset Description

The application of DL methods typically requires access to large datasets to achieve robust and accurate results. However, obtaining a sufficiently large and well-annotated dataset of skin cancer images remains a critical challenge. The limited availability of annotated medical images poses a significant barrier to training effective DL models. To address this limitation, the HAM10000 ("Human Against Machine with 10000 training images") dataset [23], a publicly available collection of 10,015 dermoscopic images, is utilized.

The dataset comprises images collected from Austrian and Australian patients and includes a diverse representation of skin lesions: 514 Basal Cell Carcinomas (BCC), 6,705 Melanocytic Nevi (NV), 142 Vascular Lesions (VASC), 1,113 Melanomas (MEL), and 2,197 images of other lesion types, including 1,099 Benign Keratoses (BKL), 327 Actinic Keratoses / Intraepithelial Carcinomas (AKIEC), and 115 Dermatofibromas (DF). All images are 24-bit RGB, uniformly formatted at 96 DPI, and sized at 600×450 pixels. Representative examples of each lesion type are illustrated in Figure 2. The dataset covers seven diagnostic categories, namely: NV, MEL, BKL, BCC, AKIEC, VASC, DF. This comprehensive and balanced dataset serves as a solid foundation for training and evaluating DL-based skin lesion classification models.

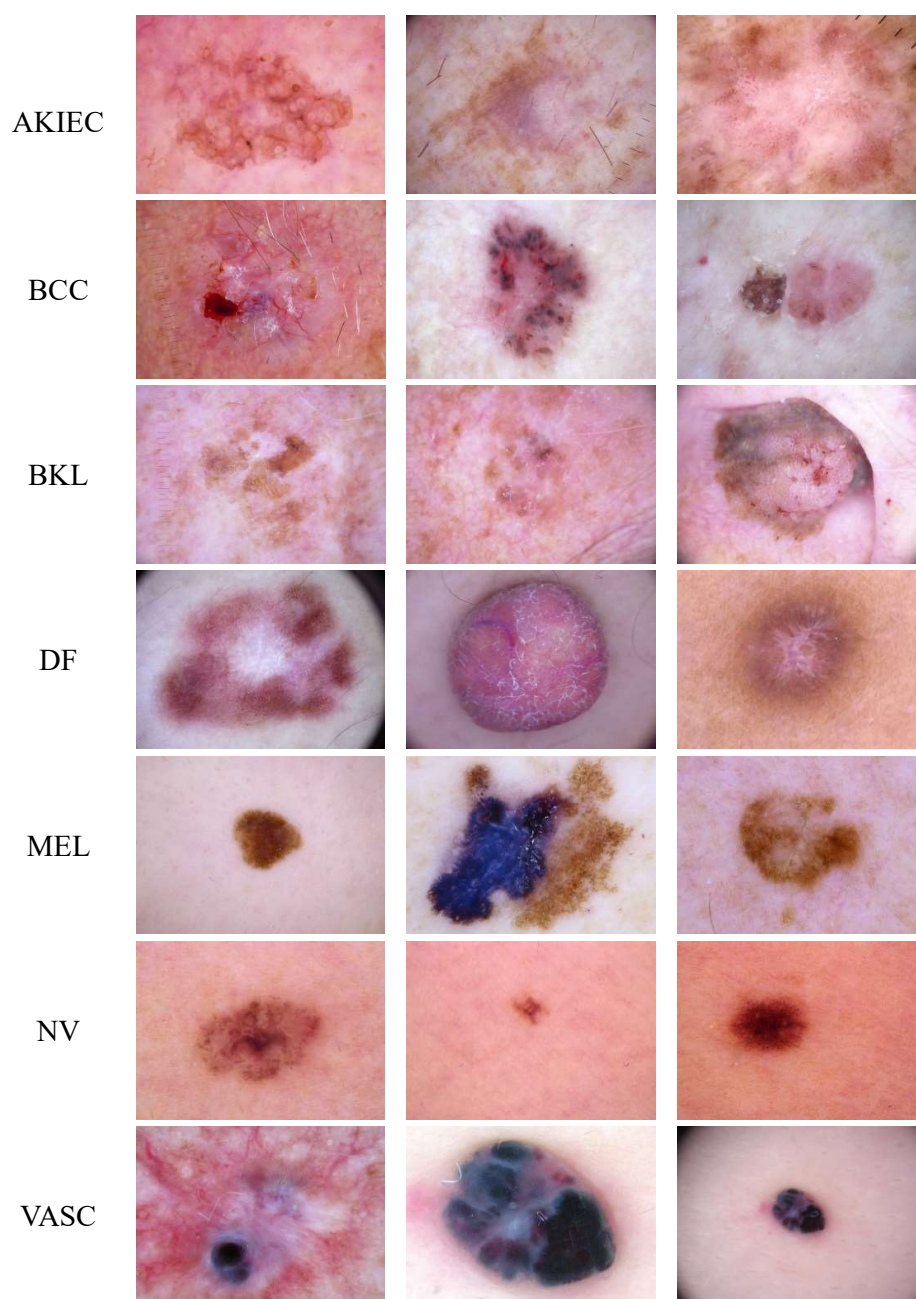


Fig. 2 Sample images from the dataset

B. Evaluation Criteria

The performance measures such as precision, accuracy, recall, and F1-score are popular metrics used to assess the performance of the model. The definitions of these measures are as follows:

$$precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F1 - score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (3)$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

where the numbers TP, TN, FP, and FN stand for true positives, true negatives, false positives, and false negatives, respectively.

C. Performance of SkinVGG-Net

Figure 3 shows the obtained confusion matrix of the proposed SkinVGG-Net model.

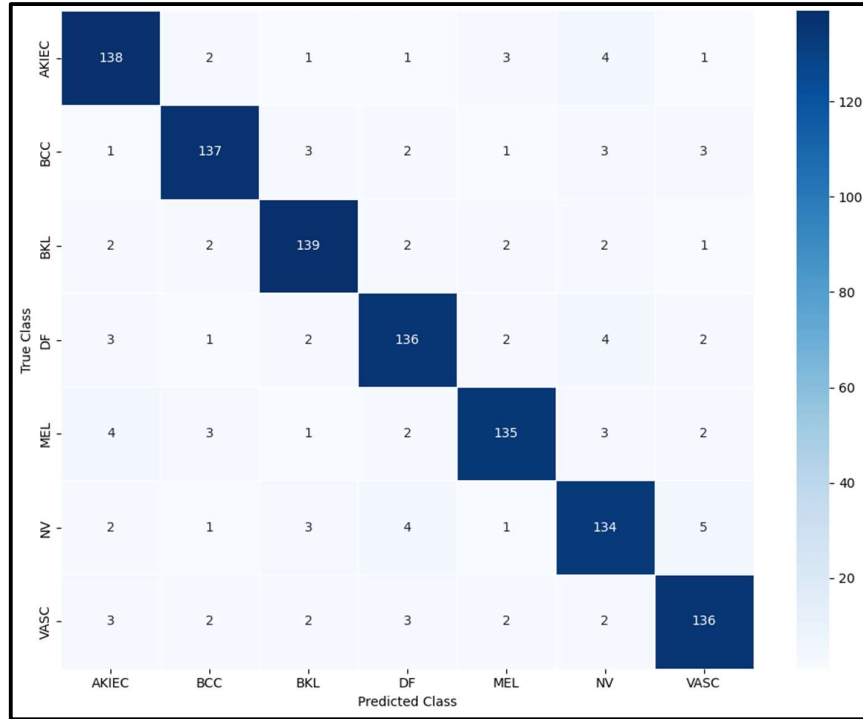


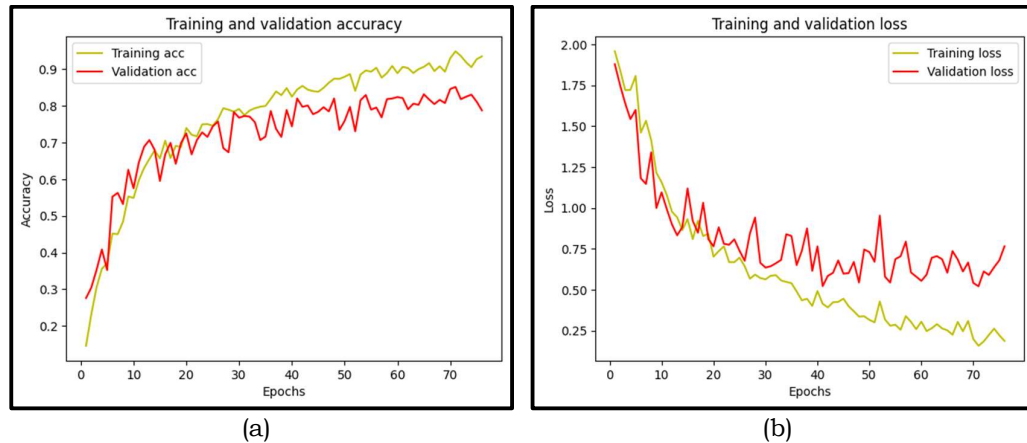
Fig. 3 Confusion matrix of SkinVGG-Net

To quantitatively evaluate the proposed SkinVGG-Net model's performance in classifying skin lesions, commonly used standardized metrics in medical image classification discussed in the previous section were employed. Given that the dataset comprises seven distinct lesion categories, the classification task follows a multi-class approach. The evaluation results of the SkinVGG-Net model for this multi-class classification are summarized in Table 2.

TABLE. 2 Proposed SkinVGG-Net model's performances

Classes	Precision	Recall	F1-score	Accuracy
AKIEC	90.20	92.00	91.09	92.00
BCC	92.57	91.33	91.95	91.33
BKL	92.05	92.67	92.36	92.67
DF	90.67	90.67	90.67	90.67
MEL	92.47	90.00	91.22	90.00
NV	88.16	89.33	88.74	89.33
VASC	90.67	90.67	90.67	90.67
Average	90.97	90.95	90.96	90.95

It can be seen from Table 2 that the proposed SkinVGG-Net model achieved an average accuracy of 90.95%, with closely aligned values for precision (90.97%), recall (90.95%), and F1-score (90.96%), indicating a balanced and reliable performance across metrics. Among all classes, BKL showed the highest accuracy (92.67%) and F1-score (92.36%), reflecting the model's strong capability in distinguishing this class. MEL, was detected with a precision of 92.47% and an F1-score of 91.22%, showcasing the model's high sensitivity and specificity in identifying malignant lesions. BCC and AKIEC also demonstrated robust performance, with accuracy above 91%. Though NV has the highest number of samples in the dataset, it had slightly lower metrics possibly due to intra-class variability and visual similarity with other lesion types. The diagnosis of NV using imaging and deep learning methods is especially difficult due to the high variability in shape, colour, and size within this class. This diversity hinders the model's ability to learn consistent features, often resulting in misclassification. Other less frequent classes such as DF and VASC were classified with identical precision, recall, and F1-scores (90.67%), indicating stable performance even in lower-represented categories. Overall, these results confirm the effectiveness of SkinVGG-Net in handling class imbalance and achieving high classification accuracy in multiclass skin lesion diagnosis. Figure 4 shows the accuracy and loss curves of SkinVGG-Net for both training and validation sets.

**Fig. 4 (a) Accuracy of SkinVGG-Net (b) Loss of SkinVGG-Net**

A detailed examination of Figure 4 reveals several notable trends. The relatively small gap between the curves in Figure 4 suggests that the model generalizes well and that overfitting has been effectively mitigated. This is likely to be due to the integration of dense layers and dropout regularization, which enhances the model's ability to learn meaningful features without considering the

noise in the training data. The validation accuracy occasionally exceeds the training accuracy, and vice versa due to several factors. The dropout and batch normalization may cause temporary dips in training performance due to their stochastic nature, while the validation set may yield more stable and occasionally higher accuracy if it has less noisy or more uniform skin images. On the other hand, instances where training accuracy surpasses validation accuracy may indicate brief periods of overfitting, which are later corrected as the regularization mechanisms take full effect during training.

D. Comparative Analysis

The accuracy of the models served as the primary evaluation metric. It is crucial to acknowledge that direct comparisons with prior research are difficult to conduct due to the differences in methodologies, datasets, and model architecture. Table 3 shows a comprehensive summary of state-of-the-art models.

TABLE. 3 Comparison with state-of-the-art models

References	Method	Precision	Recall	F1-score	accuracy
[11]	Alex Net	84.21	81.25	82.31	80.45
	ResNet50	93.73	92.53	92.74	90.08
	VGGNet16	90.70	90.32	90.61	88.36
	VGGNet19	88.55	88.82	88.38	88.70
	Xception	90.19	90.57	90.41	90.30
[15]	Inception-v3		77.88	77.84	88.05
	ResNet-50		81.00	81.28	89.28
	Inception- ResNet-v2	-	78.18	78.46	87.74
	DenseNet-201		79.57	79.47	88.70
[16]	ResNet-50	-	-	-	85.80
[17]	EfficientNetB4	88.00	88.00	87.00	87.91
Proposed	SKINVGG-NET	90.97	90.95	90.96	90.95

The proposed SkinVGG-Net model outperforms widely used deep learning models in literature. When compared to classical architectures such as AlexNet (80.45%) and VGGNet19 (88.70%), SkinVGG-Net demonstrates superior accuracy. It also surpasses EfficientNetB4, reported at 87.91% accuracy in [17], as well as Inception-ResNet-v2 (87.74%) and DenseNet201 (88.70%) from [15]. Although ResNet50 in [11] achieved slightly higher precision and recall, the SkinVGG-Net maintains a more balanced trade-off across all evaluation metrics. These results validate the effectiveness of SkinVGG-Net's architectural modifications, such as dropout regularization, dense layers, and data augmentation, in handling multi-class classification with high precision and generalization capability.

Although data augmentation techniques were employed to increase the size of the database and variability of training data, this enhancement is still insufficient to fully address the scarcity of labeled medical images. Acquiring high-quality annotated datasets in the medical domain is inherently difficult, as it requires expert interpretation and labeling by specialist physicians, making large-scale data collection both time-consuming and resource intensive.

IV. CONCLUSION

Early diagnosis and treatment of skin cancer significantly reduce mortality rates and greatly improve patient survival outcomes, underscoring the critical importance of early detection. Given the remarkable success of deep CNNs across various domains, it is unsurprising that they also hold significant promise for

applications in healthcare. The obtained results from SkinVGG-Net model shows that transfer learning can substantially enhance the capabilities of pre-trained models by enabling them to effectively recognize image features and classify skin lesions with greater precision than human experts. By using pre-trained weights of VGG-19 model and fine-tuning a modified VGG19 architecture, the proposed SkinVGG-Net strikes an optimal balance between architectural depth and computational simplicity. The model consistently delivers strong performance in skin image classification and has the potential to be a reliable and efficient diagnostic support tool in clinical practice.

Funding Statement: This research received no external funding.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

REFERENCES

- [1] Z. Rahman, M.S. Hossain, M.R. Islam, M.M. Hasan and R.A. Hridhee, "An approach for multiclass skin lesion classification based on ensemble learning," *Informatics in Medicine Unlocked*, vol. 25, 2021, Article. 100659.
- [2] M.R. Naeemah and M.Y. Kamil, "Advances in deep learning for skin cancer diagnosis," *Advance Sustainable Science, Engineering and Technology*, vol. 6, no. 4, 2024, pp. 02404019-01-02404019-011.
- [3] D. Keerthana, V. Venugopal, M.K. Nath and M. Mishra, "Hybrid convolutional neural networks with SVM classifier for classification of skin cancer," *Biomedical Engineering Advances*, vol. 5, 2023, Article. 100069.
- [4] J.C. Manning, X. Chu, J. Boza, R. Ayanga, H. Muwando, R. Lukande, M. Horenstein, T. Maurer, E. Cesarman, A. Semeere, J. Martin and D. Erickson, "Skin biopsy processing for rapid molecular diagnosis and histopathologic interpretation: Application to Kaposi sarcoma in East Africa," *Infectious Agents and Cancer*, vol. 20, no. 1, 2025, Article. 38.
- [5] N. deRosa, "Evolving role of sentinel lymph node biopsy for melanoma, merkel cell, and squamous cell carcinoma," *Surgical Clinics of North America*, vol. 105, no. 3, 2025, pp. 483–500.
- [6] T.A. Putra, S.I. Rufaida and J.S. Leu, "Enhanced skin condition prediction through machine learning using dynamic training and testing augmentation," *IEEE Access*, vol. 8, 2020, pp. 40536–40546.
- [7] S. Serte and H. Demirel, "Gabor wavelet-based deep learning for skin lesion classification," *Computers in biology and medicine*, vol. 113, 2019, Article. 103423.
- [8] N.F. Lattoofi, I.F. Al-sharuee, M.Y. Kamil, A.H. Obaid, A.A. Mahidi, A.A. Omar and A.K. Saleh, "Melanoma skin cancer detection based on ABCD rule," *1st International Scientific Conference of Computer and Applied Sciences*, 2019, pp. 154–157.
- [9] N. Ashfaq, Z. Suhail, A. Khalid, N. Sarwar, A. Irshad, O. Yaman, A. Alubaidi, F.M. Ahmed and F.A. Almalki, "SkinSight: Advancing deep learning for skin cancer diagnosis and classification," *Discover Computing*, vol. 28, no. 1, 2025, Article. 63.
- [10] P. Natha, S.P. Tera, R. Chinthaginjala, S.O. Rab, C.V. Narasimhulu and T. H. Kim, "Boosting skin cancer diagnosis accuracy with ensemble approach," *Scientific Reports*, vol. 15, no. 1, 2025, Article. 1290.

- [11] S.H. Kassani and P.H. Kassani, "A comparative study of deep learning architectures on melanoma detection," *Tissue and Cell*, vol. 58, 2019, pp. 76–83.
- [12] B. Ozdemir and I. Pacal, "A robust deep learning framework for multiclass skin cancer classification," *Scientific Reports*, vol. 15, no. 1, 2025, Article. 4938.
- [13] M.I. Khaleel and A. Lakizadeh, "Skin cancer diagnosis using hybrid deep pre-trained convolutional neural networks," *IAES International Journal of Artificial Intelligence*, vol. 14, no. 3, 2025, pp. 2291–2301.
- [14] M.A. Kadampur and S. Al Riyae, "Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images," *Informatics in Medicine Unlocked*, vol. 18, 2020, Article. 100282.
- [15] M.A. Al-Masni, D.H. Kim and T.S. Kim, "Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification," *Computer methods and programs in biomedicine*, vol. 190, 2020, Article. 105351.
- [16] F. Afza, M. Sharif, M. Mittal, M.A. Khan and D.J. Hemanth, "A hierarchical three-step superpixels and deep learning framework for skin lesion classification," *Methods*, vol. 202, 2022, pp. 88–102.
- [17] K. Ali, Z.A. Shaikh, A.A. Khan and A.A. Laghari, "Multiclass skin cancer classification using EfficientNets—a first step towards preventing skin cancer," *Neuroscience Informatics*, vol. 2, no. 4, 2022, Article. 100034.
- [18] H. Afra, I. Shehana and D.A. Thanuja, "Multi-label prototype based interpretable machine learning for melanoma detection," *International journal of advances in signal and image sciences*, vol. 8, no. 1, 2022, pp. 40–53.
- [19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [20] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman and P.R. Pinheiro, "CovidGAN: Data augmentation using auxiliary classifier GAN for improved Covid-19 detection," *IEEE Access*, vol. 8, 2020, pp. 91916–91923.
- [21] M.K. Hasan, M.A. Ahamad, C.H. Yap and G. Yang, "A survey, review, and future trends of skin lesion segmentation and classification," *Computers in Biology and Medicine*, vol. 155, 2023, Article. 106624.
- [22] M. Goyal, T. Knackstedt, S. Yan and S. Hassanpour, "Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities," *Computers in Biology and Medicine*, vol. 127, 2020, Article. 104065.
- [23] 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, 2018, pp. 1–9.