Performance Analysis of Deep Transfer Learning to Classify Skin Cancer Images

Omer Sabri Emeksiz*, Sakir Bingol*, Ibrahim Tinas*, Sude Bulut*, Sanem Uytan*, Müzeyyen Işıl Değirmencioğlu†, Zeynep Destan‡, Kadir Berk Yağar‡
*Department of Electrical and Electronics Engineering, Marmara University, Istanbul, Turkey
Emails: sakir.bingol@marmara.edu.tr, omeremeksiz@marun.edu.tr, ibrahimtinas@marun.edu.tr, sudebulut@marun.edu.tr, sanemuytan@marun.edu.tr
†Department of Industrial Engineering, Marmara University, Istanbul, Turkey
Email: misil@marun.edu.tr
‡Department of Computer Engineering, Marmara University, Istanbul, Turkey
Emails: zeynepdestan@marun.edu.tr, kadir.berk@marun.edu.tr

Abstract— The escalating global incidence of skin cancer underscores the urgent need for enhanced diagnostic methodologies that significantly improve early detection and treatment outcomes. This research focuses on the development and benchmarking of advanced artificial intelligence (AI) systems based on deep learning convolutional neural networks (CNNs) for the accurate classification of skin cancer using high-resolution dermatoscopic images. The comprehensive HAM10000 dataset, which includes various types of skin lesions, was utilized to meticulously evaluate the performance of multiple CNN models that had been both pre-trained and fine-tuned. A methodological framework was incorporated, utilizing transfer learning techniques to leverage the robust capabilities of pre-existing models from the extensive ImageNet database, thereby enhancing the classification efficiency of CNNs in identifying skin cancer lesions. Models such as Inception-v3, DenseNet-201, MobileNetV2, and others were assessed, each chosen for their specific advantages in handling image-based classification tasks efficiently. A range of performance metrics such as accuracy, F1 score, specificity, recall, and precision were employed to evaluate the models. These metrics are critical in medical diagnostic contexts where the reliability of the diagnosis can significantly impact patient management. It was found that some models achieve classification accuracies exceeding 98% on specific datasets, highlighting the substantial potential of AI to assist dermatologists in making more precise and timely diagnostic decisions. Additionally, the advantages of deploying deep learning technologies were emphasized in reducing the manual workload and the inherent subjectivity of traditional skin lesion analysis methods, such as visual inspection and histopathological examinations. Through automation of the detection and classification processes, AI systems can markedly reduce diagnostic errors and enhance overall patient care quality.

Index Terms—Skin Cancer, Deep Learning, Convolutional Neural Networks, Transfer Learning, Dermatoscopic Image Analysis

I. Introduction

Cancer, known for its high incidence and low survival rate, poses a serious risk to public health worldwide [1]. According to GLOBOCAN 2020 data, 19.3 million new cancer cases occur every year and 10 million people die as a result of cancer diseases [2]. Lung cancer continues to lead as the most common fatal cancer, with 1.8 million deaths annually [2]. Lung cancer is followed by skin cancer, is a widespread disease that affects a significant portion of the global population. According to the International Agency research on cancer in 2022, 325,000 new cases were detected in 2020 from melanoma, which is a rarer but more dangerous form of skin cancer due to its rapid spread, and 57,000 people passed away from this disease worldwide [3]. According to a study conducted by Siegel et al., in 2022, 97,920 new cases of melanoma in situ of the skin were detected in the United States [4]. According to the latest research by the American Cancer Society for 2024 predictions for melanoma in the United States, approximately 100,640 new melanomas will be diagnosed, including 59,170 in men and 41,470 in women, and approximately 8,290 people will be diagnosed with melanoma, including 5,430 men and 2,860 women. is expected to die [5].

Death rates from melanoma, the deadliest form of skin cancer, declined rapidly from 2013 to 2017, falling by about 6% to 7% per year, largely due to advances in treatment [5]. In the development of skin cancer treatment, as in other cancers, enhancing early diagnosis, which is a critical phase of treatment in terms of saving the patient's life, has been effective. The most commonly used traditional techniques for detecting skin cancer include visual and histopathological analysis of lesions [6]. While visual assessment is based on clinical observation that takes into account the visual features of the lesion, histopathological examination involves laboratory examination, usually using a biopsy to examine a sample of the lesion. In addition, features of the lesions such as pigmentation, vascularity and regression are monitored by dermatologists using devices such as dermatoscopes [7]. An additional instance of skin cancer detection includes utilizing confocal microscopy, a technique enabling the analysis of skin layers without the need for extracting samples [8] [9]. However, these methods, which are effective in the detection and diagnosis of skin diseases, increase the risks and workloads of subjectivity in analysis and the margin of human error resulting from examining complex lesions for long hours. Technological advances in artificial intelligence and machine learning have made significant contributions to reducing the margin of error in the detection and classification of skin cancer and reducing the resources allocated for this work [10].

Table 1 presents an encapsulation of the studies conducted, showcasing the methodologies and outcomes employed by various researchers in the realm of skin cancer classification. The methodologies used and the results obtained shed light on the roadmap in this article. In the study conducted by Allizadeh and Mahloojifar, an accuracy rate of 94.11% was obtained in classifying skin lesions into seven categories with Normal Bayes and SVM-based algorithms using the HAM10000 data set, which we also use [11]. In Roy et al.'s study, a precision value of 98% was obtained by using YOLOv2, a convolutional neural network, for melanoma detection [12]. In Fujisawa et al.'s study, an accuracy of 76.5% was achieved by using the deep convolutional neural networks GoogLeNet in the classification of benign and malignant lesions obtained from the University of Tsukuba Hospital [13]. Udrea et al. achieved a 95.8% accuracy rate in the ISIC Database when classifying skin lesions with the help of a model combining a conditional generative adversarial network SVM [14]. Giavina-Bianchi et al. An accuracy of 89.3% was achieved with the model developed by applying convolutional networks (VGG16) and KNN algorithms for melanoma detection on ISIC2019 and PH2 datasets [15]. In a study conducted by Samy Bakheet, Shtwai Alsubai, Aml El-Nagar and Abdullah Alqahtani for use in detecting melanoma in dermatoscopy images, a 94% accuracy was achieved while developing a Computer Aided Diagnosis (CAD) system using SVM, KNN and Gentle AdaBoost machine learning models. Remarkable results were obtained with a high accuracy rate [16]. Esteva et al. investigated the application of deep learning algorithms, specifically convolutional neural networks (CNNs), to classify skin cancer. Their model, trained on a vast dataset of over 140,000 dermatoscopic images, achieved accuracy comparable to board-certified dermatologists in distinguishing malignant melanoma from benign lesions. In a study using nine-fold cross-validation, the CNN achieved 72.1% overall accuracy for a 3-class disease partition (benign, malignant, non-neoplastic lesions) [17]. Ali et al. proposed a highly accurate DCNN model (93.16% training, 91.93% testing) for classifying skin lesions as benign or malignant. Their model outperformed other techniques like AlexNet and ResNet on the HAM10000 dataset [18].Togacar et al. introduced a novel model merging auto-encoder, spiking, and CNN architectures. They utilized the International Skin Imaging Collaboration (ISIC) dataset. The study employed MobileNetV2, incorporating residual blocks and spiking networks, for training and classifying both the original and structured datasets. The model achieved a remarkable classification accuracy of 95.27% [19]. Hosny et al. used transfer learning, utilizing a pre-trained deep learning network (AlexNet), in their investigation. They conducted classification on three distinct types of lesions, employing fine-tuning techniques. To address dataset imbalances, data augmentation methods were employed. The model achieved an impressive accuracy of 98.61% on the PH2 dataset [20]. In conclusion, our comprehensive review of the literature demonstrates the notable success of various machine learning models in skin cancer classification, highlighting significant advancements in the analysis of dermoscopic images within this field.

TABLE I
LITERATURE REVIEW

Authors	Year	Accuracy Results	Methodology	Dataset
Allizadeh and Mahloojifar	2018	94.11%	Normal Bayes and SVM	HAM10000
Roy et al.	2018	98.00%	YOLOv2	PH2
Fujisawa et al.	2019	76.50%	DCNN GoogLeNet	University of Tsukuba Hospital
Udrea et al.	2020	95.80%	CGAN-SVM	ISIC Database
Giavina-Bianchi et al.	2021	89.30%	VGG16 and KNN	ISIC2019 and PH2
Samy Bakheet et al.	2023	94.01%	SVM, KNN, AdaBoost	MED-NODE
Esteva et al.	2020	72.19%	DCNN GoogLeNet	ISIC and Stanford Hospital
Togacar et al.	2021	95.27%	DCNN MobileNetV2	ISIC
Ali et al.	2021	91.93%	DCNN	HAM10000
Hosny et al.	2019	98.61%	DCNN AlexNet	PH2

In this study, it is aimed to develop reliable deep learning convolutional neural network-based artificial intelligence systems that classify skin cancer using skin cancer images to facilitate skin cancer classification and to perform performance analysis between these models. This involves the utilization of deep learning convolutional neural networks (CNNs), which are specialized architectures adept at extracting features from images for classification tasks [21]. Additionally, transfer learning is employed, leveraging pre-trained models like those from ImageNet, to adapt and fine-tune the CNNs for improved performance in the skin cancer classification task [22]. Our research facilitates comparisons among different architectures through comprehensive performance analysis. Furthermore, we introduce supportive tools that can significantly assist dermatologists in making accurate early diagnoses with reduced effort and error, thereby mitigating the necessity for surgical interventions.

II. System Model

Convolutional Neural Networks (CNNs) are widely utilized for the analysis of visual imagery owing to their advanced ability to recognize patterns. These applications span across various fields such as natural language processing, recommendation systems, and image and video identification. A CNN architecture comprises several layers, each serving distinct functions as illustrated in the image. Convolutional layers utilize learnable filters to extract feature maps, followed by pooling layers that reduce feature dimensions while retaining essential information. Fully connected layers then integrate these features to classify input into output classes. This layering allows the network to recognize varied patterns, and backpropagation optimizes weights to minimize prediction errors, enhancing the CNN's effectiveness across applications [23]. Figure 1 illustrates the general structure of CNN.

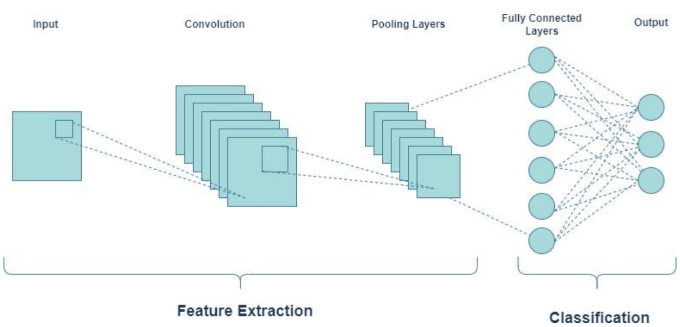

Fig. 1. General Structure of CNN.

Deep learning algorithms excel at uncovering hidden structures within data through hierarchical learning [24]. Transfer learning enhances this capability by leveraging pre-trained models to boost performance on new tasks, particularly when data is limited. By initially training a model on a large dataset to reveal structural relationships, transfer learning enables the transfer of this knowledge to new tasks, even with differing data distributions. This approach is especially potent when labeled data is scarce, as it harnesses insights from a broader context to achieve superior results [25]. In contrast, traditional machine learning starts from scratch on a specific dataset for a given task. Transfer learning stands out by enhancing performance through the transfer of knowledge from a source task to a related target task. As shown in Figure 2, initially trained on a large source dataset, the model's weights are fine-tuned using a target dataset, thereby reusing learned features and accelerating learning, a crucial advantage when data is limited. This iterative process not only improves efficiency but also facilitates better results, demonstrating the adaptability and effectiveness of transfer learning across various domains [26].

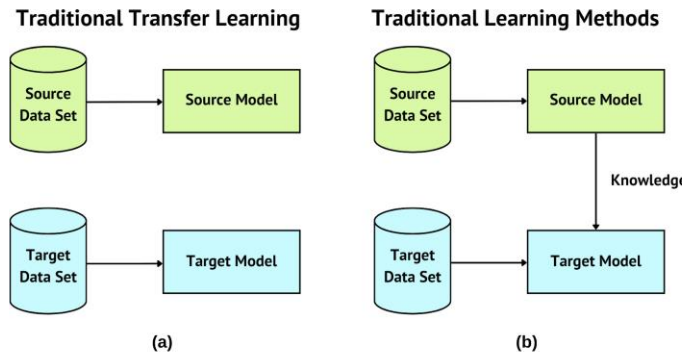

Fig. 2. General Structure of Transfer Learning.

A. Deep Learning Models

Understanding the distinctions among pre-trained models for performance analysis studies offers a starting point for our endeavors. In this study, we delve into the characteristics of several models, listing them below and detailing their structures and layer counts. Our aim is to facilitate informed decision-making in the application of deep learning methodologies by providing a thorough review and comparison of the advantages and disadvantages inherent to each neural network architecture.

- Inception-v3, the third version of the Inception models, enhances its predecessors by using three filter sizes in parallel convolution layers, is 48 layers deep, and is pre-trained on ImageNet with an input size of [299 299 3] [27].
- DenseNet-201, with its 201-layer architecture and [224 224 3] image input size, achieves a significant increase in depth over other models by using shorter connections between layers near the input/output to enhance feature propagation and reuse, thereby drastically reducing the number of parameters [28].
- MobileNetV2 is designed with 53 layers, utilizing a [224 224 3] input size, and employs inverted residuals and linear bottlenecks alongside ReLU6 activation functions, focusing on efficiency and performance optimization without squeeze-and-excitation blocks, making it suitable for mobile devices and pre-trained on the ImageNet database [29].
- MobileNetV3, with 53 layers and a [224 224 3] input size, incorporates AutoML optimizations, h-swish activations, and squeeze-and-excitation blocks, enhancing efficiency and performance for mobile devices, pre-trained on ImageNet [30].
- ResNet-101, featuring 101 layers and a [224 224 3] input size, provides enhanced capacity for modeling complex functions due to its deeper structure, making it ideal for high-performance tasks requiring detailed visual understanding and is trained on extensive datasets like ImageNet [31].

- ResNet-50, with 50 layers and a [224 224 3] input size, offers efficient computation and fast inference, utilizing powerful residual learning for complex feature capture on datasets like ImageNet [32].
- The Xception model, 71 layers deep with [299 299 3] input, uses depthwise separable convolutions to enhance Inception modules, trained on ImageNet [33].
- The Inception-ResNet-v2 model, combining Inception architecture and residual connections for faster training, is 164 layers deep and trained on ImageNet with a [299 299 3] input size [34].
- NASnetMobile, the mobile version of the Neural Architecture Search Network (Nasnet), uses reinforcement learning to optimize cell structures in convolutional networks during training on specific datasets, and is pre-trained on ImageNet requiring a [224 224 3] image input size [35].
- EfficientNet-b0, the baseline of the EfficientNet series up to b7, uses compound scaling and is pre-trained on ImageNet, requiring [224 224 3] input size [36].

B. Performance Metrics

Four criteria were used to assess the models' performance: accuracy, F1 score, recall, and precision. The ratio of genuine positives to all photos classified as positive (including false positives) is known as precision. The ratio of true positives to all relevant variables (i.e., actual positives) is recall, also known as sensitivity. The F1 score indicates the classification accuracy in unbalanced datasets and is calculated as the harmonic mean of recall and precision. The ratio of true positives for each class to the total number of instances—that is, all the photos in the testing set—defines accuracy [37]. The following is a definition of the five measures:

$$Precision = \frac{TP}{(TP + TF)} \tag{1}$$

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

$$F1 = 2 \times \frac{Recall \times Precision}{Recall + Precision} \tag{3}$$

$$Accuracy = \frac{\sum_{i}^{No.\ of\ Classes} TP_i}{No.\ of\ testing\ images} \tag{4}$$

The number of images that have been accurately detected for each class is known as true positives (TP). The amount of photos that have been mistakenly classified as belonging to a different class is known as false positives (FP). False negatives (FN) on the other hand indicate the number of photos that the classifier was unable to accurately identify.

C. Dataset Overview

The HAM10000 dataset, a cornerstone in skin cancer detection research, comprises 10,000 dermatoscopic images sourced from diverse demographic and geographic backgrounds. These images are categorized into seven distinct types, representing a range of skin lesions, both benign and malignant, making it an invaluable tool for training and evaluating machine learning models in dermatology. Curated from leading dermatology research centers and public archives, the dataset's extensive variety enhances its applicability across different populations, ensuring that models developed are effective in real-world scenarios. Each image is meticulously annotated with detailed lesion information, verified by experienced dermatologists, ensuring high diagnostic accuracy and reliability. Additionally, metadata on patient demographics and lesion localization are included, which support deeper diagnostic insights and more targeted algorithm training. This robust dataset not only facilitates the development of advanced diagnostic models but also enhances the precision and efficiency of machine learning algorithms in classifying diverse skin lesions, reinforcing our commitment to utilizing advanced technology for improving early skin cancer detection and treatment outcomes.

D. Methodology

In our study, we assessed deep learning models on the HAM10000 dataset using Google Colaboratory's powerful GPUs, the NVIDIA Tesla K80 and T4. This configuration made it possible to train and infer models efficiently, which is crucial for meeting the computing demands of the dataset. We used TensorFlow, which is renowned for its scalability in production situations, and PyTorch, which we chose for its dynamic computing graphs, to implement our models. In order to evaluate the models' capacity to manage class imbalances in medical image analysis and offer insights into their usefulness in dermatoscopic diagnoses, we concentrated on important metrics such as accuracy and F1-score. Furthermore, Google Colaboratory's versatility made it easier to integrate sophisticated analytical tools and libraries, which improved our process. This comprehensive analysis highlights how cutting-edge deep learning methods can greatly increase the accuracy of dermatoscopic picture classification in therapeutic contexts.

TABLE II
SKIN LESION CLASSIFICATION ACCURACY RESULTS

Model	Mean Accuracy	Max. Accuracy	Min. Accuracy	Standard Deviation
Inception-v3	85.64%	91.22%	78.85%	4.10%
DenseNet-201	96.15%	98.04%	93.41%	1.60%
MobileNet-v2	88.22%	92.34%	84.47%	2.30%
MobileNet-v3	87.53%	92.50%	82.17%	3.10%
ResNet-101	94.91%	97.25%	92.35%	1.50%
ResNet-50	84.31%	90.34%	79.91%	3.40%
Xception	85.66%	90.75%	80.73%	3.20%
Inception-ResNet-v2	92.33%	96.98%	87.01%	2.80%
NASNet-Mobile	75.67%	83.52%	70.27%	4.50%
EfficientNet-B0	86.05%	91.47%	81.49%	3.00%

III. RESULTS AND DISCUSSION

The comparative analysis of various deep learning models for skin cancer classification reveals notable differences in

TABLE III
F1 SCORE, PRECISION, AND RECALL METRICS

Model	F1 Score	Precision	Recall
Inception-v3	0.86	0.87	0.85
DenseNet-201	0.97	0.98	0.96
MobileNet-v2	0.89	0.90	0.88
MobileNet-v3	0.88	0.89	0.87
ResNet-101	0.95	0.96	0.94
ResNet-50	0.85	0.86	0.84
Xception	0.86	0.87	0.85
Inception-ResNet-v2	0.93	0.94	0.92
NASNet-Mobile	0.76	0.77	0.75
EfficientNet-B0	0.87	0.88	0.86

their performance metrics, as summarized in Table 2 and Table 3. DenseNet-201 and ResNet-101 exhibit the highest mean accuracy rates of 96.15% and 94.91%, respectively, demonstrating robustness with high F1 scores, precision, and recall. DenseNet-201 achieved a mean accuracy of 96.15%, with a maximum accuracy of 98.04% and a minimum accuracy of 93.41%, showing consistency and reliability in performance. Similarly, ResNet-101 shows a mean accuracy of 94.91%, a maximum accuracy of 97.25%, and a minimum accuracy of 92.35%. MobileNet-v2 and MobileNet-v3 also perform well, particularly notable for their balance between accuracy and computational efficiency, making them suitable for mobile and resource-constrained applications.

In terms of precision, DenseNet-201 leads with a score of 0.98, indicating a high true positive rate and minimal false positives, while ResNet-101 follows closely with a precision of 0.96. The recall rates for DenseNet-201 and ResNet-101 are 0.96 and 0.94, respectively, underscoring their effectiveness in correctly identifying positive cases. This balance is critical in medical applications where both false positives and false negatives can have serious implications. The F1 scores for DenseNet-201 and ResNet-101 are 0.97 and 0.95, respectively, reinforcing their robustness across different performance metrics. Conversely, NASNet-Mobile shows significantly lower accuracy and higher variability, suggesting potential limitations in handling complex lesion classifications.

The standard deviation values provide insights into the consistency of the models, with DenseNet-201 and ResNet-101 showing low standard deviations, indicating stable performance across different datasets. The variability observed in models such as NASNet-Mobile underscores the need for targeted optimizations to enhance their effectiveness. For mobile applications, MobileNet-v2 and MobileNet-v3 offer a good balance between performance and efficiency. These findings highlight the significant potential of models like DenseNet-201 and ResNet-101 in achieving high accuracy and robust performance, thereby supporting early diagnosis and treatment and emphasizing the critical role of deep learning in advancing skin cancer diagnostics.

IV. CONCLUSION AND FUTURE WORK

This study highlights the significant potential of deep learning models such as DenseNet-201 and ResNet-101 in the classification of skin lesions, which have demonstrated high accuracy and robust performance metrics. The variability observed in the performance of models like NASNet-Mobile underscores the importance of targeted optimizations to enhance their effectiveness.

Looking forward, it is imperative to explore pre-image processing techniques, including noise reduction and image enhancement, to improve the quality of input images. This approach is expected to bolster the models' ability to accurately classify complex skin lesions. Additionally, experimenting with hybrid models that leverage the unique strengths of various architectures could lead to improvements in both accuracy and computational efficiency.

Practical application and validation of these models in clinical settings are essential next steps. This will not only test their efficacy but also provide critical feedback for further refinement. The ultimate goal is to equip dermatologists with reliable, non-invasive diagnostic tools that can facilitate early detection and treatment of skin cancer, thus improving patient care and outcomes.

REFERENCES

[1] N. Harbeck and M. Gnant, "Breast cancer," Lancet, vol. 389, no. 10074, pp. 1134–1150, 2017. [Online]. Available: doi:10.1016/S0140-6736(16)31891-8

[2] H. Sung et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA Cancer J Clin, vol. 71, no. 3, pp. 209–249, 2021. [Online]. Available: doi:10.3322/caac.21660

[3] International Agency for Research on Cancer, "Skin cancer," International Agency for Research on Cancer, 2022. [Online]. Available: https://www.iarc.who.int/cancer-type/skin-cancer/

[4] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2022," CA: A Cancer Journal for Clinicians, vol. 72, no. 1, pp. 7–33, 2022.

[5] American Cancer Society, "Facts & Figures 2024," Atlanta: American Cancer Society, 2024.

[6] S. Jiang, H. Li, and Z. Jin, "A visually interpretable deep learning framework for histopathological image-based skin cancer diagnosis," IEEE J Biomed Health Inform, vol. 25, pp. 1483–1494, 2021. doi: 10.1109/JBHI.2021.3052044

[7] S.K. Parsons et al., "Noninvasive Diagnostic Techniques for the Detection of Skin Cancers," 2011.

[8] J. Malvehy and G. Pellacani, "Dermoscopy, confocal microscopy and other non-invasive tools for the diagnosis of non-melanoma skin cancers and other skin conditions," Acta Derm Venereol, vol. 97, pp. 22–30, 2017. doi: 10.2340/00015555-2720

[9] V. Narayanamurthy et al., "Skin cancer detection using non-invasive techniques," RSC Adv, vol. 8, pp. 28095–28130, 2018. doi: 10.1039/C8RA04164D

[10] M.J. Iqbal et al., "Clinical applications of artificial intelligence and machine learning in cancer diagnosis: looking into the future," Cancer Cell Int, vol. 21, no. 270, 2021. [Online]. Available: https://doi.org/10.1186/s12935-021-01981-1

[11] S. M. Alizadeh and A. Mahloojifar, "A Mobile Application for Early Detection of Melanoma by Image Processing Algorithms," in 2018 25th National and 3rd International Iranian Conference on Biomedical Engineering (ICBME), Qom, Iran, 2018, pp. 1-5, doi: 10.1109/ICBME.2018.8703575.

[12] S. S. Roy, A. U. Haque, and J. Neubert, "Automatic diagnosis of melanoma from dermoscopic image using real-time object detection," in 2018 52nd Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, 2018, pp. 1-5.

[13] Y. Fujisawa, Y. Otomo, Y. Ogata, Y. Nakamura, R. Fujita, Y. Ishitsuka, et al., ”Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumor diagnosis,” Br J Dermatol, vol. 180, pp. 373-381, 2019, doi: 10.1111/bjd.16924.

[14] A. Udrea, G. Mitra, D. Costea, E. Noels, M. Wakkee, D. Siegel, et al., ”Accuracy of a smartphone application for triage of skin lesions based on machine learning algorithms,” J Eur Acad Dermatol Venereol, vol. 34, pp. 648-655, 2020, doi: 10.1111/jdv.15935.

[15] M. Giavina-Bianchi, R. M. de Sousa, V. ZdA Paciello, W. G. Vitor, A. L. Okita, R. Prôa, et al., ”Implementation of artificial intelligence algorithms for melanoma screening in a primary care setting,” PLoS ONE, vol. 16, e0257006, 2021, doi: 10.1371/journal.pone.0257006.

[16] S. Bakheet, S. Alsubai, A. El-Nagar, and A. Alqahtani, “A Multi-Feature Fusion Framework for Automatic Skin Cancer Diagnostics,” Diagnostics, vol. 13, no. 1474, 2023, doi: 10.3390/diagnostics13081474.

[17] Esteva, Andre, et al. ”A Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks.” Nature, vol. 542, no. 7639, 2020, pp. 115-118, 2017, doi: 10.1038/nature21056

[18] Ali, M. S., Miah, M. S., Haque, J., Rahman, M. M., & Islam, M. K., “An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models. Machine Learning with Applications”, 5, 2021, 100036. doi: 10.1016/J.MLWA.2021.100036.

[19] Togacar, M., Comert, Z., & Ergen, B., “Intelligent skin cancer detection applying autoencoder, MobileNetV2 and spiking neural network”, Chaos, Solitons & Fractals, 144, 110714, 2021, doi:10.1016/J.CHAOS.2021.110714.

[20] Hosny, K. M., Kassem, M. A., & Foaud, M. M., “Skin Cancer Classification using Deep Learning and Transfer Learning.”, 2018 9th Cairo International Biomedical Engineering Conference, CIBEC 2018 - Proceedings, 90–93., 2019, doi:10.1109/CIBEC.2018.8641762.

[21] I. Goodfellow, Y. Bengio, and A. Courville, ”Deep Learning,” MIT Press, 2016.

[22] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, ”ImageNet: A large-scale hierarchical image database,” in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255.

[23] S. Ali, J. Li, Y. Pei, M. S. Aslam, Z. Shaukat, and M. Azeem, ”An Effective and Improved CNN-ELM Classifier for Handwritten Digits Recognition and Classification,” Symmetry, vol. 12, no. 1742, Oct. 2020. DOI: 10.3390/sym12101742.

[24] Yoshua Bengio, “Proceedings of ICML Workshop on Unsupervised and Transfer Learning”, PMLR 27:17-36, 2012.

[25] F. Zhuang et al., ”A Comprehensive Survey on Transfer Learning,” in Proceedings of the IEEE, vol. 109, no. 1, pp. 43-76, Jan. 2021, doi: 10.1109/JPROC.2020.3004555.

[26] Combs, K.; Lu, H.; Bihl, T.J., “Transfer Learning and Analogical Inference: A Critical Comparison of Algorithms, Methods, and Applications.” Algorithms 2023, 16, 146, doi: doi.org/10.3390/a16030146.

[27] Long D. Nguyen, Dongyun Lin, Zhiping Lin, Jiuwen Cao, ”Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation,” in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP).

[28] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, and M. Kaur, ”Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning,” Journal of Biomolecular Structure and Dynamics, 2020, doi: 10.1080/07391102.2020.1788642.

[29] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. -C. Chen, ”MobileNetV2: Inverted Residuals and Linear Bottlenecks,” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 4510-4520, doi: 10.1109/CVPR.2018.00474.

[30] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, ”Searching for MobileNetV3,” arXiv preprint arXiv:1905.02244, Nov. 2019.

[31] K. He, X. Zhang, S. Ren, and J. Sun, ”Deep Residual Learning for Image Recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.

[32] H. Wang, Y. Li, and B. Buckles, ”A Variety of Deep Learning Models to Classify Disaster Scene Videos,” Proceedings of TheIIER International Conference, Houston, United States of America, 23rd-24th September 2021, pp. 29-33.

[33] Y. Liu, L. Zhang, Z. Hao, Z. Yang, S. Wang, X. Zhou, and Q. Chang, ”An Xception model based on residual attention mechanism for the classification of benign and malignant gastric ulcers,” Scientific Reports, vol. 12, 15365, 2022, doi: 10.1038/s41598-022-19639-x.

[34] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, ”Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning,” in Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), 2017.

[35] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le, ”Learning Transferable Architectures for Scalable Image Recognition,” in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 8697-8710.

[36] S. Gang, N. Fabrice, D. Chung, and J. Lee, ”Character Recognition of Components Mounted on Printed Circuit Board Using Deep Learning,” Sensors, vol. 21, 2921, 2021, doi: 10.3390/s21092921.

[37] S. Das, M. Singh, A. Sharma, S. Kumar, and S. A. Akbar, ”An Improved Deep Convolutional Neural Network Model for Predicting the Stability of Car-following Behavior,” IEEE Access, vol. 7, pp. 92556-92565, 2019. DOI: 10.1109/ACCESS.2019.2924045.