



Dual Convolutional Neural Network for Skin Cancer Classification

Raya Sattar Shahadha¹, Belal Al-Khateeb^{2,*}

¹Informatics Institute for Postgraduate Studies, University of Information Technology & Communications, Iraq

²College of Computer Science and Information Technology, University of Anbar, Iraq

Emails: Rayasattar2016@gmail.com; belal-alkhateeb@uoanbar.edu.iq

Abstract

Skin cancer detection through deep learning is an evolving field, where convolutional neural networks (CNNs) have proven to be very effective in feature extraction. However, this approach still faces some limitations due to the use of data augmentation, It is the generation of artificial images. Which significantly increase the computational load without generate new clinically meaningful data and may introduce shadowed features. Therefore, this study aims to propose a new approach that use CNNs to extract important features from skin cancer medical images using the HAM 10000 dataset. The proposed approach involves training two different CNN architectures, extracting features from convolutional layers, and then use PCA to make the retrieved features less dimensional. In order to categorize skin cancer into seven different categories of skin lesions, the remaining features are then merged and fed into a classifier that uses neural networks. In comparison to earlier studies that employed CNN architectures on the same dataset, the results demonstrated that this method preserves significant information while improving computational efficiency and achieving superior classification performance. The suggested approach achieved 95.66% accuracy for multi-class classification.

Keywords: Skin Cancer; Convolutional Neural Network (CNN); Feature Extraction; Dual CNN; HAM 10000; Principal Component Analysis (PCA); Fully Connected Classifier

1. Introduction

The most prevalent cancers among this disease is skin cancer ranks among globally, and early detection via medical imaging can greatly enhance patient prognoses. Dermoscopy, a favoured imaging technique among dermatologists, enlarges the skin lesion surface, making its structure more discernible for examination. However, this method's effectiveness is heavily dependent on the visual acuity and experience of the physician, necessitating specialized training. These limitations have spurred the research community to seek advanced methods for visualizing and diagnosing melanoma. One such advancement is the computer-aided diagnosis (CAD) system, which aids in melanoma detection and offers a user-friendly interface for less experienced dermatologists. CAD systems can serve as a valuable second opinion in melanoma diagnosis [1,2]. While numerous machine learning techniques are employed in medical image analysis, deep learning has emerged as a superior method due to its high accuracy in overcoming image processing challenges and traditional machine learning limitations. Deep learning is particularly notable for its applications in medical image analysis, including the classification of melanomas, brain tumours, and eye diseases [3]. Remarkable success has shown with CNN in image classification tasks, including medical image analysis and among DL networks, CNN is among the most widely used and well-liked. Nevertheless, CNN, DL is quite well-liked these days, and its primary benefit over its forerunners is that it can identify significant traits automatically without human oversight, which contributes to its enormous popularity [4,5]. CNN is a powerful tool in medical image analysis, especially for detecting and classifying cancerous tumours by learning complex patterns and identifying subtle differences in medical images that humans might miss. CNNs can be trained to classify images into multiple categories, such as different types of skin lesions. Melanoma, benign keratosis-like lesions, basal cell carcinoma and other types of skin lesions. Its ability to

automatically extract relevant features and represent hierarchical features is particularly important in the medical field where distinguishing between types of lesions can influence treatment decisions. CNN designed of multiple consisted layers to adapt automatically learn the spatial hierarchy of features from input images. The main types of layers in CNN are convolutional layers that enable convolutional filters to be applied to the input image to extract features. These filters learn to recognize different patterns such as edges, textures and complex structures, pooling layers that reduce the dimensionality of feature maps while preserving important information, making the network more computationally efficient, and fully connected layers that combine extracted features to perform classification tasks [6,7]. As shown in Figure (1).

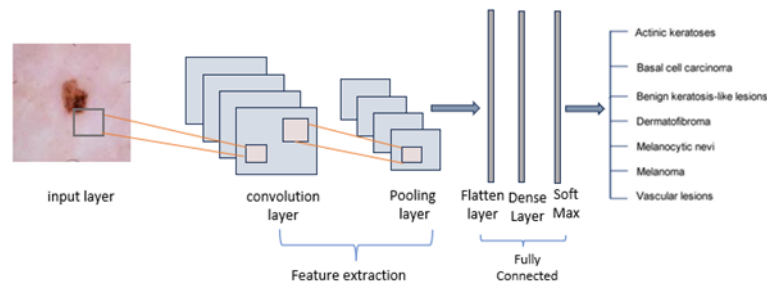


Figure 1. Basic layers in CNN architecture

CNNs are very effective in extracting features from medical images, including those for skin cancer detection. By leveraging the capabilities of automatic feature learning, hierarchical structure, and robustness, CNNs have become a cornerstone of medical image analysis. However, challenges such as data availability and computational requirements must be addressed to realize their full potential in clinical applications. Automatic feature learning is one of the main advantages of CNNs which is their ability to automatically learn and extract features from raw medical images. Unlike traditional methods that require manual feature engineering, CNNs learn relevant features during training. In addition to hierarchical feature representation, CNNs build a hierarchy of features, starting from low-level features such as edges and textures in the initial layers to high-level features. Such as specific shapes and patterns in deeper layers. This hierarchical representation is particularly effective in capturing complex structures in medical images. Features extracted from a trained CNN can be used to classify skin lesions in tasks such as segmentation, abnormality detection, and even other types of cancer detection [8,9]. A linear algebraic method for continuous attributes called principal component analysis (PCA) identifies new attributes, or principal components that are orthogonal (perpendicular) to one another capture the greatest range of variation in the data, and are linear combinations of the original qualities. Catches the greatest amount of residual variability. It is especially helpful in seeing patterns and in comprehending how various qualities are correlated with one another. PCA aims to identify new sets of dimensions, or qualities that more accurately reflect the data's variability. PCA is commonly used for dimensionality reduction [10].

In this research, we offer a strategy that uses dual CNN architectures in conjunction with Principal Component Analysis (PCA) and a fully connected classifier to extract and select salient features from skin cancer photos. The paper's remaining portions are the next section reviews related work. Section 3 provides a description of the datasets in depth. Section 4 describes the proposed approach's architecture and specifications. The experimental results are displayed in Section 5 along with a discussion of them and a comparison with alternative methods. Finally, conclusion and future works are in Section 5.

2. Related Work

The authors in [11] proposed a CNN architecture approach for skin cancer classification using a CNN algorithm with five layers that uses image pixels and diagnosis labels as input data. Following classification, the image is sent to a different model that estimates tumour severity using OpenCV color prediction. The performance of this model achieved a test accuracy of 86.68% using the HAM10000 dataset after increasing it, and these results lack accuracy, as the rest of the important metrics in medical research were not mentioned, such as precision and F1 score. In [12] a classification system is developed, including Grad-CAM and Grad-CAM++ to explain model decisions, with a classification accuracy of 81.24% using the HAM10000 dataset after augmentation. The accuracy was not high, and other important criteria were not mentioned, such as Precision and F1-score which determines the probability of false positives or false negatives, same as other important standards in medical fields such as precision and F1 score. [13] A proposed study uses a variety of deep learning and hybrid models to examine each model's performance in classifying cancer categories after pre-processing and data augmentation. Using the

HAM10000 dataset after augmentation, the suggested CNN-LSTM-SVM hybrid model produced the best result when compared to the models of its competitors, with an accuracy of 88.24%. This study's precision rate of 88% and F1 score of 88%, which indicate the possibility of false positives or false negatives, indicate that its accuracy is fairly low. The work in presented a lightweight CNN model for multilayer skin lesion categorization. Using a median filter, the noise from the dermatoscopy images is first minimized. Following that, two standard models, ResNet50 and VGG16, were created, and image-degraded dermatoscopy was used to train the suggested CNN. It achieved an accuracy of 89.71 and Precision of 89.24 and F1-score 89.20 which is higher than the two standard models using HAM10000 dataset after augmentation here, the accuracy is not high enough for medical research, and there is a not insignificant rate of occurrence Potential for false positives or false negatives. In the work [14], A CNN network based on spatial attention has been imported together with a pre-trained DenseNet-201. The retrieved features from both networks were merged to get the best forecast. With an overall accuracy of 82.576%, the experimental findings demonstrate that the HAM10000 dataset works remarkably well after augmentation. The obtained accuracy is low, and it was not mentioned that other crucial measurement standards in medical research, like precision and F1-score, are also significant factors that impact the probability of false positives or false negatives. The work in [15] compared four deep learning techniques RNN (Recurrent Neural Networks), CNN, ResNet50, and Xception. The dataset used is HAM10000 with data augmentation. The accuracy using CNN, RNN, ResNet50, and Xception was 72%, 69%, 93%, and 79%. Respectively with ResNet50 performance is better. Although the accuracy here is somewhat high, however, there is a significant chance of false positives. [16]This paper presents a comparative study on classifying skin lesions using Convolutional Neural Networks (CNN) and Random Forest classifiers, alongside a real-time simulation for detecting skin cancer. The HAM10000 dataset, which contains a variety of images representing seven different skin lesion types, was utilized in this study. After pre-processing the images to remove noise and artefacts, the images underwent segmentation through the Active Contours without Edges (ACWE) method. Feature extraction was conducted using the ABCDT technique, with texture analysis performed using the Gray Level Co-Occurrence Matrix (GLCM) and Fractal Dimension Texture Analysis (FDTA). The CNN classifier achieved an accuracy of 91.97%, while the Random Forest classifier reached 89.82%. Real-time simulation of skin cancer detection using the trained models showed that the CNN model outperformed the Random Forest classifier. However, despite the high accuracy of the CNN, other evaluation metrics such as recall or F1 score were not reported, which is important for a comprehensive assessment of the model in the medical field. [17]A CNN model was developed to classify seven types of skin lesions in the HAM10000 dataset, achieving a classification accuracy of 91.51%. The model was integrated into a web application and underwent evaluation by seven expert dermatologists in two stages. In the first stage, it was determined that the model could accurately diagnose skin lesions with an accuracy of 90.28% in practical scenarios. During the second stage, the model was able to correct the experts' misdiagnoses with an accuracy of 11.14%. The classification report indicated that accuracy was skewed towards the lesion type (nv), while other evaluation metrics for each lesion, such as Precision, Recall, and F1-score, were relatively low. These metrics were not reported for the system overall.

Data augmentation was used in all previous works. However, there were many limitations, such as having a weakness in accuracy, or using accuracy only as a measure of the efficiency of the method while neglecting the rest of the other measures in the confusion matrix, such as accuracy and F1-score, or a weakness in these measures, which It leads to a significant error rate, resulting in a false positive or false negative diagnosis.

3. Methodology

3.1 Dataset

The HAM10000 dataset ("Human against Machine with 10,000 training images"), a sizable collection of dermoscopy images frequently utilized in dermatology and medical image analysis, was used. 10,015 photos of pigmented skin lesions from two different locations—the Cliff Rosendahl Skin Cancer Clinic in Queensland and the Department of Dermatology at the Medical University of Vienna in Austria—were collected over the course of 20 years to create this dataset. Australia. Using digital cameras. Because it was collected across different populations and different clinics, this helped ensure the diversity and generalizability of this dataset. It is publicly available and accessible for research purposes [21,22]. The HAM10000 suffers from a class imbalance, as shown in the figure (2) with some types of lesions being more common than others, and there is significant heterogeneity in this dataset because it includes images from different devices and of varying quality, which can pose challenges to developing robust algorithms [18].

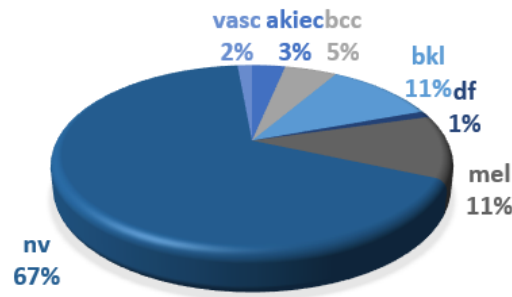


Figure 2. Distribution of the seven diagnostic categories

3.2 Proposed approach

The proposed approach is dual CNN with different architectures. The idea behind using multiple architectures was to capture a variety of features, each consisting of four convolutional layers. Features are extracted using convolutional layers followed by activation function after each convolutional layer there is a max pooling layer, at this stage, features are extracted and entered into flatten layer in each algorithm. PCA is applied to the features extracted. The features resulting from applying PCA to the features of both algorithms are then combined, and finally a Fully Connected Classifier is used for classification. Following the combination of the features obtained by applying PCA to the features of both techniques, a Fully C. Using two layers (Dense), with Dropout layer in between, followed by a (SoftMax) layer, to classify the features into seven types of skin lesions, as shown in the figure (3), and the parameters of dual CNN model in table (1).

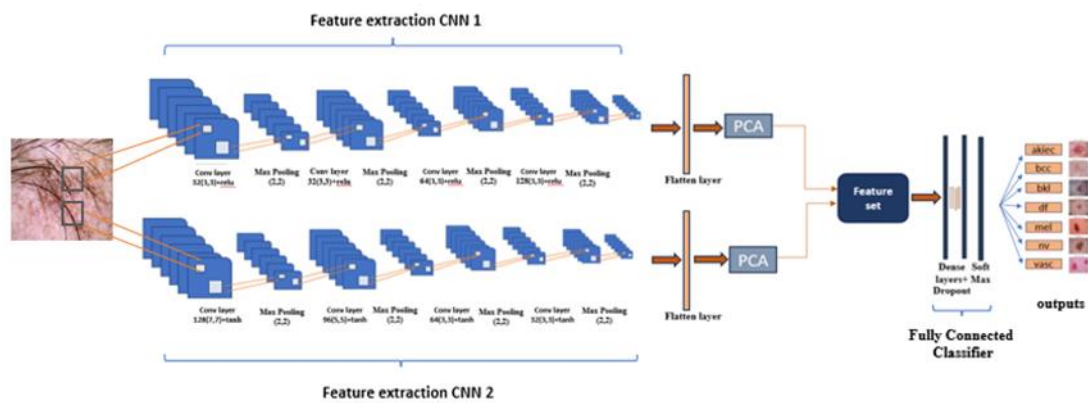


Figure 3. Architecture of the proposed approach (Dual CNN)

Table 1: The Parameters of the CNN model

CNN architecture	Layer	Kernel	Filter	activation
First CNN	Conv2D	(3, 3)	32	relu
	MaxPooling2D	(2, 2)		
	Conv2D	(3, 3)	64	relu
	MaxPooling2D	(2, 2)		
	Conv2D	(3, 3)	128	relu
	MaxPooling2D	(2, 2)		
	Conv2D	(3, 3)	256	relu
	Flatten			

Second CNN	Conv2D	(7,7)	256	tanh
	MaxPooling2D	(2, 2)		
	Conv2D	(5, 5)	128	tanh
	MaxPooling2D	(2, 2)		
	Conv2D	(3, 3)	64	tanh
	MaxPooling2D	(2, 2)		
	Conv2D	(3, 3)	32	tanh
	MaxPooling2D	(2, 2)		
	Flatten			
Fully Connected Classifier	Dense		256	relu
	Dropout 0.2			
	Dense		128	
	SoftMax7			relu

4. Result and discussion

A number of criteria, including accuracy, precision, recall, validation loss accuracy, loss validation, accuracy, and F1-score, were used to evaluate the classification's performance. Accuracy was determined by dividing the number of correctly predicted cases by the total number of samples.

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

The number of successfully anticipated positive cases divided by the total of correctly predicted positive cases and mistakenly predicted positive cases yields the precision. When false positives are more concerning than false negatives, this statistic is very useful.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Recall, or sensitivity, is calculated by dividing the number of correctly predicted positive cases by the total number of actual positive cases. This metric is particularly important when false negatives are of greater concern than false positives.

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

The F1 score balances the trade-off between precision and recall, which typically inversely affect each other when one is improved. It is the harmonic mean of precision and recall, providing a comprehensive evaluation of both metrics. The F1 score ranges from 0 to 1, with 1 being the optimal value and 0 the worst [10].

$$\text{F1 - score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

The classification report for the dual CNN, as presented in Table 2, demonstrates the effectiveness of the proposed approach in the multi-class classification of skin cancer lesions. According to the report, the highest classification accuracy was achieved for melanocytic nevi (nv), with a precision 0.98, recall 1.00, and F1 score 0.99. The lowest performance was observed for actinic keratoses (bcc), with a precision of 0.79, recall of 0.83, and an F1 score of 0.81. Table (2) Classification report for dual CNN

Table 2: Performance scores (dual CNN)

lesion	precision	recall	f1-score
akiec	0.84	0.94	0.89
bcc	0.79	0.83	0.81
bkl	0.96	0.85	0.90
df	0.93	0.92	0.93

mel	0.94	0.96	0.94
nv	0.98	1.00	0.99
vasc	0.97	0.84	0.90
macro avg	0.92	0.90	0.91
Weighted avg	0.96	0.96	0.96

We also calculated the amount of accuracy, verification, accuracy loss, and verification loss to confirm the model’s performance. The figures below show that the model’s performance was good, as the verification accuracy reached 0.9651 and the verification loss was 0.3282. Figure (4) confusion matrix for dual CNN. Figure (5) shows the accuracy and verification and figure (6) amount of accuracy loss, and the amount of loss.

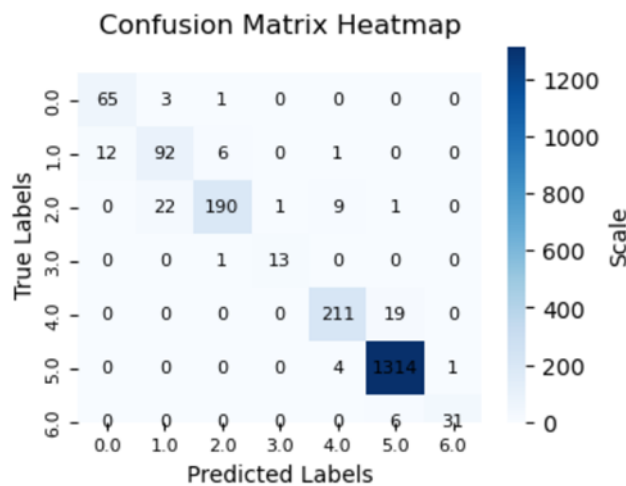


Figure 4. Confusion matrix for dual CNN

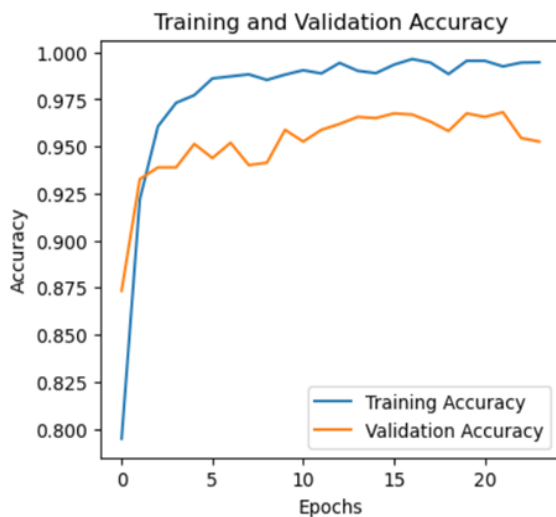


Figure 5. Accuracy and validation

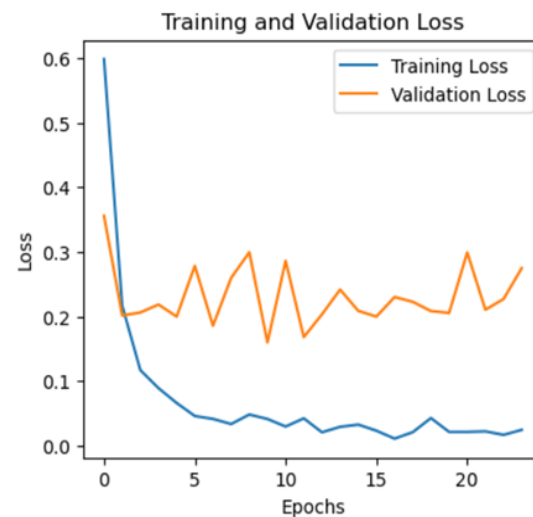


Figure 6. Accuracy loss and validation loss

5. Comparative Analysis

Compared with previous studies using the HAM10000 dataset, our method showed improved classification accuracy. This can be attributed to the effectiveness of feature extraction and dimensionality reduction techniques used, which demonstrates the effectiveness of the proposed method. Instead of relying on data augmentation techniques, which is what all the compared studies did, data augmentation techniques such as zooming, rotating, scaling, or flipping can provide features that are not naturally present in medical images. For example, rotating an

image of a skin lesion can create orientations that are rare or not present in real cases, which may lead to a misaligned model. Learn incorrect patterns. That is, increasing data may provide shaded features, and the research results demonstrated that image features extracted from convolutional layers in CNN networks are more useful in classifying skin lesions specific to skin cancer and are more able to generalize to real-world conditions table (3) Comparison of the proposed methodology with previous studies that used the same type of data (HAM 10000). The comparison shows the difference in the measures of accuracy, Precision, F1 score.

Table 3: Comparing the proposed methodology with previous studies that used the same type of data (HAM 10000).

Ref	Technique	Accuracy	Precision	Recall	F1 score
[11]	CNN	86.68%	NA	NA	NA
[12]	CNN +Grad-CAM++.	82%	NA	NA	NA
[13]	Customized CNN	86%	86%	86%	86%
	CNN-SVM	87%	87%	87%	87%
	CNN-RandomFores	87%	87%	87%	87%
	ResNet50	78.82%	79%	79%	79%
	DenseNet121	83.1%	83%	83%	83%
	VGG19	79.85%	80%	80%	80%
	CNN-LSTM	84.45%	84%	84%	84%
	CNN-LSTM-SVM	88.24%	88%	88%	88%
[14]	VGG16	88.80%	88.24%	88.48	88.32%
	ResNet50	86.62%	86.94%	88.24	86.44%
	CNN	89.71%	89.28%	89.20	89.20%
[15]	DenseNet-201+CNN	82.57%	NA	NA	NA
[16]	CNN	72%	87 %	72%	85%
	RNN	69%	72 %	88%	81%
	ResNet50	93%	89 %	93%	93%
	Xception	79%	89 %	89%	87%
[17]	CNN	91.97%	NA	NA	NA
	Random Forest	89.82%			
[18]	CNN	91.28	NA	NA	NA
Proposed approach	Dual CNN	95.66%	96%	NA	96%

6. Conclusion and future work

This study demonstrates that using multiple CNN architectures for feature extraction followed by PCA and fully connected classifier is a powerful method for extracting and selecting important features from skin cancer medical images. This approach not only enhances the efficiency of subsequent analyses, but also ensures that important information necessary for accurate diagnosis is retained. The results achieved in this study show a significant improvement compared to previous methods, highlighting the effectiveness of the proposed methodology. Future research will focus on integrating this feature extraction and classification methodology into a full automated diagnostics pipeline. Additionally, evaluating the performance of this approach across diverse datasets will be essential to validate its generalizability and robustness in real-world clinical settings. The proposed method demonstrated the possibility of overcoming the problem of high imbalance in datasets without the need to use data augmentation techniques as is done in all previous studies, which in the end are artificially generated images. Also, problems specific to skin cancer images were overcome, such as the problem of hair covering skin based on CNN algorithms filters and layers without the need to perform pre-treatment operations for such problems.

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

- [1] A. Naeem, M. S. Farooq, A. Khelifi, and A. Abid, "Malignant Melanoma Classification Using Deep Learning: Datasets, Performance Measurements, Challenges and Opportunities," *IEEE Access*, vol. 8, pp. 110575–110597, 2020, doi: 10.1109/ACCESS.2020.3001507.
- [2] L. Talavera-Martínez, P. Bibiloni, A. Giacaman, R. Taberner, L. J. D. P. Hernando, and M. González-Hidalgo, "A novel approach for skin lesion symmetry classification with a deep learning model," *Comput. Biol. Med.*, vol. 145, no. March, 2022, doi: 10.1016/j.compbiomed.2022.105450.
- [3] A. A. Mukhlif, B. Al-Khateeb, and M. A. Mohammed, "An extensive review of state-of-the-art transfer learning techniques used in medical imaging: Open issues and challenges," *J. Intell. Syst.*, vol. 31, no. 1, pp. 1085–1111, 2022, doi: 10.1515/jisys-2022-0198.
- [4] L. Alzubaidi et al., *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*, vol. 8, no. 1. Springer International Publishing, 2021. doi: 10.1186/s40537-021-00444-8.
- [5] S. Malaiarasan, R. Ravi, D. R. Maheswari, C. Yesubai Rubavathi, M. Ramnath, and V. Hemamalini, "Towards Enhanced Deep CNN for Early and Precise Skin Cancer Diagnosis," *Proc. 1st IEEE Int. Conf. Netw. Commun. 2023, ICNWC 2023*, pp. 1–7, 2023, doi: 10.1109/ICNWC57852.2023.10127521.
- [6] A. Lembhe, P. Motarwar, R. Patil, and S. Elias, "Enhancement in Skin Cancer Detection using Image Super Resolution and Convolutional Neural Network," *Procedia Comput. Sci.*, vol. 218, pp. 164–173, 2022, doi: 10.1016/j.procs.2022.12.412.
- [7] D. Raval and J. N. Undavia, "A Comprehensive assessment of Convolutional Neural Networks for skin and oral cancer detection using medical images," *Healthc. Anal.*, vol. 3, no. May, p. 100199, 2023, doi: 10.1016/j.health.2023.100199.
- [8] M. M. Taye, "Theoretical Understanding of Convolutional Neural Network :," *Mdpi Ag*, vol. 11, no. 3, p. 52, 2023.
- [9] W. Chen, X. Tan, J. Zhang, G. Du, Q. Fu, and H. Jiang, "A robust approach for multi-type classification of brain tumor using deep feature fusion," *Front. Neurosci.*, vol. 18, no. February, pp. 1–14, 2024, doi: 10.3389/fnins.2024.1288274.
- [10] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining (New International Edition)*, Pearson Education Limited, England, Essex CM20 2JE, September. 2014.
- [11] A. Mankawade, A. Bodhankar, A. Mahajan, D. Prasad, S. Mahajan, and R. Dhakalkar, "Skin Cancer Detection and Intensity Calculator using Deep Learning," *2023 Int. Conf. Adv. Technol. ICONAT 2023*, pp. 1–4, 2023, doi: 10.1109/ICONAT57137.2023.10080001.
- [12] K. Mridha, M. M. Uddin, J. Shin, S. Khadka, and M. F. Mridha, "An Interpretable Skin Cancer Classification Using Optimized Convolutional Neural Network for a Smart Healthcare System," *IEEE Access*, vol. 11, no. April, pp. 41003–41018, 2023, doi: 10.1109/ACCESS.2023.3269694.
- [13] I. N. Nawrin, T. T. Trina, R. Rahman, and A. A. Rasel, "A Comparative Analysis of Deep Learning and Hybrid Models to Diagnose Multi-Class Skin Cancer," *2023 Int. Conf. Next-Generation Comput. IoT Mach. Learn. NCIM 2023*, no. June, pp. 1–6, 2023, doi: 10.1109/NCIM59001.2023.10212670.
- [14] M. R. Uddin and T. I. Mahmud, "Dense-par-AttNet: An Attention Based Deep Learning Model for Skin Lesion Classification by Transfer Learning Approach," *4th IEEE Int. Conf. Artif. Intell. Eng. Technol. IICAIET 2022*, pp. 1–5, 2022, doi: 10.1109/IICAIET55139.2022.9936758.
- [15] P. Singh, M. Kumar, and A. Bhatia, "A Comparative Analysis of Deep Learning Algorithms for Skin Cancer Detection," *Proc. - 2022 6th Int. Conf. Intell. Comput. Control Syst. ICICCS 2022*, no. Iccics, pp. 1160–1166, 2022, doi: 10.1109/ICICCS53718.2022.9788197.
- [16] T. K. Kumar and I. Naga Himanshu, "Artificial Intelligence Based Real-Time Skin Cancer Detection," *2023 15th Int. Conf. Comput. Autom. Eng. ICCAE 2023*, pp. 215–219, 2023, doi: 10.1109/ICCAE56788.2023.10111099.
- [17] O. Sevli, "A deep convolutional neural network-based pigmented skin lesion classification application and experts evaluation," *Neural Comput. Appl.*, vol. 33, no. 18, pp. 12039–12050, 2021, doi: 10.1007/s00521-021-05929-4.
- [18] 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," *Neurosci. Informatics*, vol. 2, no. 4, p. 100034, 2022, doi: 10.1016/j.neuri.2021.100034.