Celal Bayar University Journal of Science

Using Deep Learning Architectures for Skin Cancer Classification

Bafreen Mohammed[1*], Özkan İnik[2]

[1]Department of Computer Engineering , Tokat Gaziosmanpaşa University, Türkiye
[2]Department of Computer Engineering , Tokat Gaziosmanpaşa University, Türkiye
* bafreen771@gmail.com
* Orcid No: 0009-0008-1137-9307

Received: 10 July 2024
Accepted: 3 November 2024
DOI: 10.18466/cbayarfbe.1513945

Abstract

Since skin cancer is one of the most common types of cancer, prompt diagnosis is essential to successful treatment. Impressive performance in image-based classification tasks has been demonstrated by convolutional neural networks (CNNs), particularly in recent years. In this study, the proposed CNN model was applied to the ISIC skin cancer classification challenge. A proposed deep learning model and four popular deep CNN models (ResNet, GoogleNet, AlexNet, and VGG16) were used to classify the skin cancer images. High levels of accuracy on test data from the ISIC dataset were achieved by the proposed CNN model, according to experimental results. Preprocessing was performed on images with sizes of 64x64, 100x100, 224x224, and 128x128 pixels. The experimental results show that the proposed CNN model achieved the highest accuracy rate of 86.76% on 128x128 size images.

Keywords: Classification, Convolutional neural network (CNN), Deep learning (DL), Skin Cancer, ISIC dataset

1. Introduction

Skin cancer is a prevalent type of cancer that originates in the skin's epidermal layer. It accounts for one in three cancer cases globally[1]. At about 90.0%, exposure to ultraviolet light is one of the main causes of skin cancer[2]. It was one of the five illnesses that were prevalent in the US, particularly in an area with intense sunlight. 2019 saw almost 2490 females and 4740 males lose their lives to melanoma, translating to nearly 20 deaths per day in the US alone[3, 4]. In contrast, an estimated 6850 novel fatalities associated with melanoma were documented in 2020, comprising 2240 females and 4610 males[5]. Dermatologists most commonly use dermoscopic images, also known as epiluminescence light microscopy, to analyze pigmented skin lesions. Because of the similarities between the lesions and healthy tissues, a visual examination performed with the naked eye may contain errors in recognition [6-9]. Dermatologists' manual inspection is often difficult, subjective, and time-consuming, resulting in varied recognition accuracy depending on their workload and skill[9-12]. A deep-learning convolutional neural network (CNN) image classifier that outperformed 21 board-certified dermatologists in recognizing photos with malignant lesions was initially reported by Esteva et al. [13] in 2017. During training, the CNN dissected digital photos of skin lesions and created its own diagnostic standards for melanoma identification. Deep neural networks (CNN) have been used to demonstrate dermatologist-level skin cancer categorization in a number of follow-up articles [14-17].

In this work, we trained a skin cancer convolutional neural network by using open source ISIC dataset. The classification results of the proposed CNN were compared with the four-pretrained models. To avoid bias in the creation of the data set, we implemented the ImageDataGenerator object with the desired transformations, which balanced the overall dataset images to 2500 for each class of (actinic keratosis, basal cell carcinoma, dermatofibroma, melanoma, nevus, pigmented benign keratosis, seborrheic keratosis, squamous cell carcinoma, and vascular lesion) images divided into training, validation, and testing.

2. Materials and Methods

The materials and procedures used in this article to carry out the experiments are the main topics of this section. Section 2.1 describes ISIC datasets, whereas Section 2.2 explains the proposed Methodology. In Section 2.3, model performance and evaluation are presented.

2.1. Dataset

A dataset consisting of 2357 photographs of both benign and malignant oncological conditions, provided by the International Skin Imaging Collaboration (ISIC), was utilized. All photographs, except for melanomas and moles which are more prevalent in the photos, were categorized according to the ISIC classification. Each subgroup was then divided into an equal number of images[13]. Figure 1. displays a sample from the ISIC dataset .

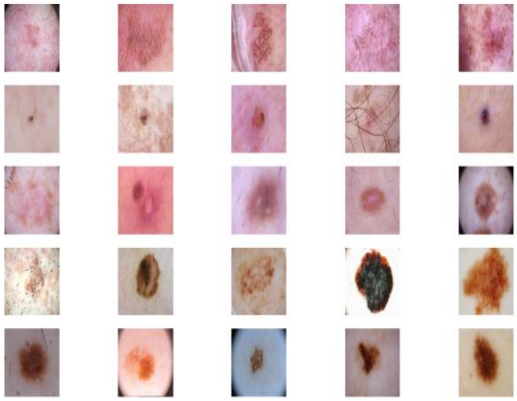

Figure 1. ISIC dataset sample

Table 1.Class-distribution of the ISIC Dataset

Class Name	Number of Images
actinic keratosis	130
basal cell carcinoma	392
Dermatofibroma	111
Melanoma	454
Nevus	373
Pigmented benign keratosis	478
seborrheic keratosis	80
Squamous cell carcinoma	197
vascular lesion	142
Total Number of Images	**2357**

2.2. Methodology

Deep neural networks can be understood mathematically as functions with millions of freely variable parameters, or weights. The intensities of the pixels in an input image are transferred to a probability of a class label when these weights are modified for a specific image classification task. Training these functions necessitates a large number of images for which the class is known due to the enormous number of free parameters. The function's output is computed for each image, compared to the specified class label, and then the weights are gently adjusted to lower the error. With only the pixel intensities of each image as input, the function "learns" how to accurately predict the class labels through numerous repetitions of this method for every image in the training set. The function shows considerable generality in predicting the class labels for unknown images, thanks to the use of training data that accurately describe the potential input space. CNNs, which have a particular architecture, were used in this work. In normal neural networks, all pixel dependencies influence all weights, with the exception of the first layers. CNNs, on the other hand, first combine nearby local pixels to identify local features before combining them to create global features. Faster training and less complex models are the outcomes of this restriction on local connections. For this reason, CNNs have been successfully applied to many different problems [18-20] because they create different representations at different layers.

The proposed convolutional neural network (CNN) model, which is intended for image classification tasks, was presented in this paper. The model begins with an input layer that requires images to have three color channels and four different resolutions, those of 64x64, 100x100, 128x128 and 224x224 pixels. The next three convolutional layers use max-pooling to decrease spatial dimensions and progressively increase filter depth (32, 64, and 128). After every convolutional layer, batch normalization is used to speed up and stabilize training. After that, a global average pooling layer is used to further reduce spatial dimensions. Dropout layers for regularization are then added after two dense layers with 128 and 64 units, respectively, and rectified linear unit (ReLU) activation functions. The last layer uses softmax activation for multi-class output probabilities and is a dense layer with nine units, which correspond to the number of classes in the classification problem. For optimization, the Adam optimizer is selected, having an epsilon of 1e-07 and a learning rate of 0.001. Accuracy is the evaluation metric, and the model is created using the categorical crossentropy loss function. With the help of regularization approaches, this architecture seeks to extract hierarchical features from input images for efficient classification, improving generalization performance.

The methodology of the proposed method is given in Table 2. As can be seen in the table, the imbalance between the classes in the dataset is eliminated in the first stage. Data augmentation methods were used to eliminate data imbalance. After balancing the dataset,

the dataset is divided into three parts: training, testing and validation. The CNN models run on the training and validation data were then tested with the test data and the final result was obtained. Four different state-of-the-art models and an original new model were developed and used as CNN models.

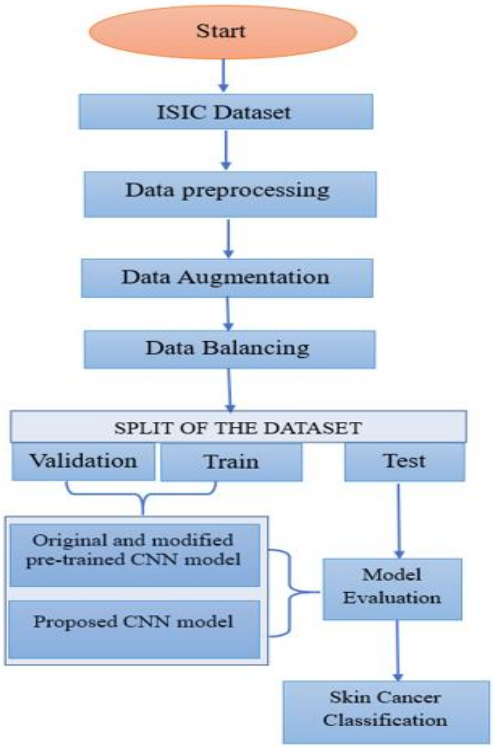

Figure 2. Proposed methodology diagram

2.2.1. Data balancing

Three techniques were employed by early machine learning researchers to address the classification problem with imbalanced data: class weight balancing [21, 22], under sampling [23, 24], and oversampling [25, 26]. Oversampling [25, 26]causes the data for categories with fewer photos to be duplicated. Oversampling does have one drawback, though: the duplicate image is identical to the original, making it useless for feature learning. Second, under sampling [23] eliminates data in categories with a higher number of photographs; however, this has the drawback of potentially removing a large number of images with useful attributes. The final step in the class weight balancing process is to multiply the loss weights by constants, each of which has an inverse relationship to the quantity of data in each category. The learning rate step may get too big to converge as a result of the class weight balancing; this is more noticeable in highly unbalanced data [27]. As a result, the suggestion in this article is to provide data for categories with fewer images. In order to enable the learning process to converge, the image generation approach [28] not only preserves the original data's features, but also adds random vectors to prevent overfitting. As shown in Figure 3. The original image data was resized to 64, 100, 128, and 224 dimensions, and the resized images were used after the dataset was balanced using augmentation techniques. Finally, the diversity of created images is increased by employing data augmentation [36] approaches.

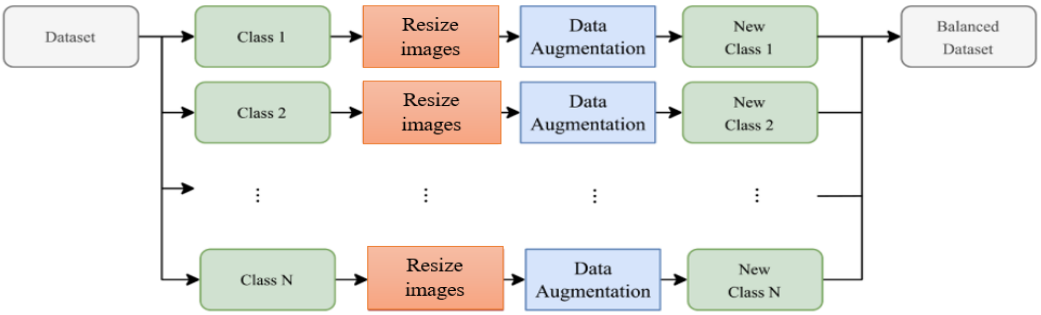

Figure 3. Flowchart of the dataset balancing

2.2.2. Data Augmentation

More and more diverse training data sets are known to enhance the performance of deep learning neural networks [29]. The process of creating artificially altered copies of images from training data is known as data augmentation, which also enlarges the training dataset by combining the original training data with the altered versions of the images. In this work, we employed rotation, flipping, shearing, and zooming as the augmentation functions. The angle of random rotation was between 0 and 20 degrees. There was a 0.2 random zooming range. The shearing range was 0.2, meaning that one axis would lengthen by 20% while the other would remain intact. It is extremely improbable that the legion will disappear from the image due to this slight degree of rotation, shearing, and zooming because it is located in the center of the image in both datasets. A few arbitrary enhanced pictures were taken from the

training dataset as shown in Table 2. ISIC Dataset Summary after augmentation

Table 2. ISIC Dataset Summary after augmentation

ISIC	Train images	Validation images	Test images	Total images
actinic keratosis	631	167	202	1000
basal cell carcinoma	631	167	202	1000
dermatofibroma	634	163	203	1000
melanoma	649	149	202	1000
nevus	633	164	203	1000
pigmented benign keratosis	647	153	200	1000
seborrheic keratosis	660	153	187	1000
squamous cell carcinoma	645	164	191	1000
vascular lesion	645	164	204	1000

2.3 Model Performance and Evaluation

This parameter is a measure of the statistical correctness of the prediction as defined in Equation-1[30]. Relying on this parameter alone can be sometimes misleading in evaluating the performance of a predictor because of its dependence on both the FP and FN. This means that two models can have the same accuracy, while one has high FP and low FN, and the other one has the opposite, i.e., low FP and high FN. Thus, the first model can be preferred to the other one, due to having a low FN for the sensitive medical scenario, which may not be decided only from the accuracy values of the models.

True Positives (TP): Occurrences where a positive outcome was predicted, and the actual result was also positive.
True Negatives (TN): Occurrences where a negative outcome was predicted, and the actual result was also negative.
False Positives (FP): Occurrences where a positive outcome was predicted, but the actual result was negative.
False Negatives (FN): Occurrences where a negative outcome was predicted, but the actual result was positive. [31].
Accuracy is defined as the ratio of the total number of input samples to the number of right predictions[31].

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

Precision is defined as the number of correct positive results divided by the total number of positive outcomes that the classifier predicted[31].
Precision and recall are measured with the statistical correctness of the final prediction as defined in Equation-2 and Equation-3, respectively. Precision measures among the predicted images for a particular class and how many of them are actually of that class. On the other hand, recall measures the total number of images from a particular class and what fraction of that are correctly classified as images from that class.

Recall is the number of correct sure results divided by the total number of conjugate samples—that is, all the samples that should have been classified as sure[31].

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

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

F1-score, It also goes by the name "harmonic mean," which seeks to strike a balance between recall and precision. It works well on an unbalanced dataset and requires both false positives and false negatives for computation[31]. F1-score, defined in Equation-4 is known as weighted average of recall and precision.

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

Additionally, we observe that the precision, recall, and F1 score range in values from 0 to 1. A model performs better for a certain classification job the greater its precision, recall, and F1 score values [32].

3. Results and Discussion

3.1. Result of proposed CNN model with image size 224x224

For the purpose of the performance analysis, the dataset is split into three sets, one of which is the training set that is used to train the model. To prevent overfitting, the model's hyperparameters are adjusted using the validation set. And lastly, the testing set, which is used to evaluate the overall performance of the model. The training loss and validation loss gradually decreasing across the epochs indicates that the model is learning and generalising successfully. After epoch 50, the validation loss stops decreasing, suggesting that the model has reached its peak performance. It is important to keep in mind that the appropriate number of epochs can vary depending on the size of the dataset and the complexity of the model. **Error! Reference source not found.** shows the proposed model's with image size

Celal Bayar University Journal of Science
Volume 20, Issue 4, 2024, p 82-91
Doi: 10.18466/cbayarfbe.1513945

(224x224) training and validation accuracy over a range of epochs, and Figure 5. shows the model's loss analyses over the same number of epochs. Figure 6. show the confusion matrix of proposed model with the same image size.

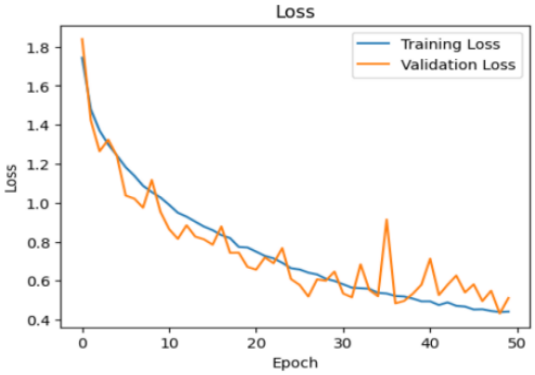

Figure 4. Convergence graph of the training and accuracy loss values obtained by the proposed CNN model in the training phase on the 224x224 image size dataset.

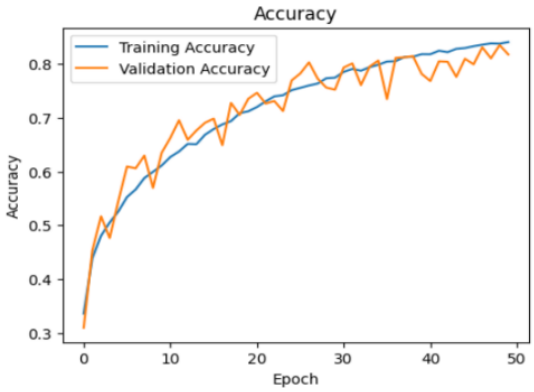

Figure 5. Convergence graph of the training and validation accuracy values obtained by the proposed CNN model in the training phase on the 224x224 image size dataset.

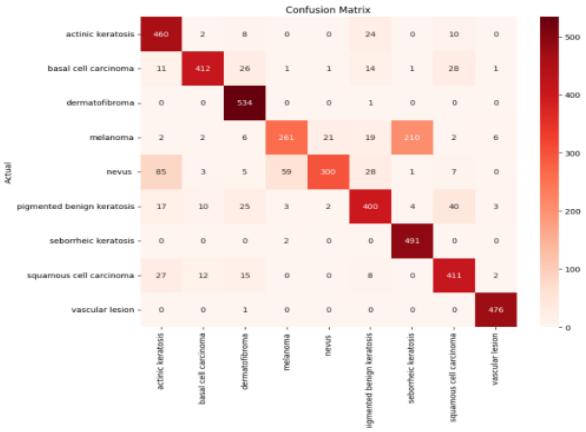

Figure 6. The confusion matrix obtained by the proposed CNN model on the 224x224 test dataset.

3.2. Result proposed CNN model with image size 128x128

After reducing the image to 128x128 pixels, the proposed CNN model result increased the accuracy of our model. Following this stage, our accuracy was 86.76%. Figure 7 shows the loss convergence graph obtained by the model during the training phase, and Figure 8 shows the accuracy convergence graph. Figure 9 shows the confusion matrix obtained by the model on the test data. When looking at the confusion matrix, it was seen that the model incorrectly labeled the "melanoma" class the most.

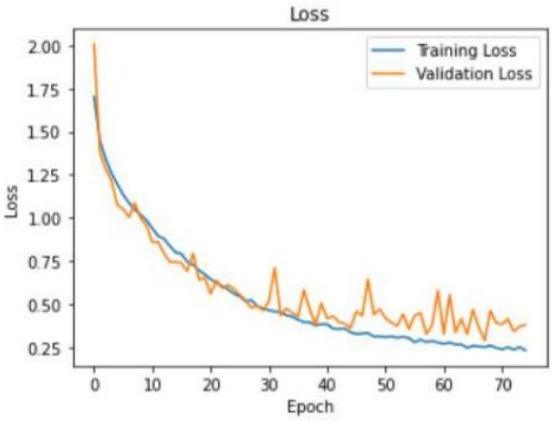

Figure 7. Convergence graph of the training and accuracy loss values obtained by the proposed CNN model in the training phase on the 128x128 image size dataset.

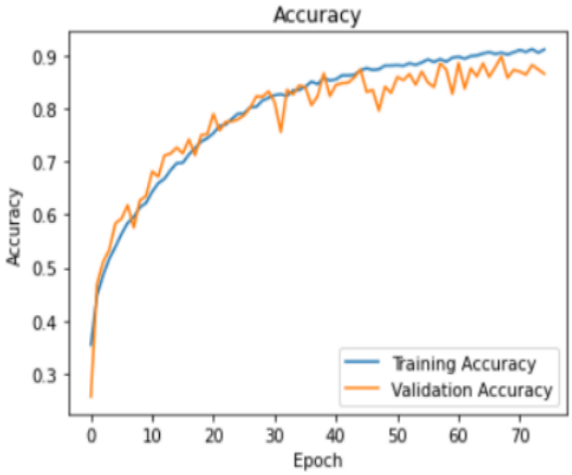

Figure 8. Convergence graph of the training and validation accuracy values obtained by the proposed

Celal Bayar University Journal of Science
Volume 20, Issue 4, 2024, p 82-91
Doi: 10.18466/cbayarfbe.1513945

CNN model in the training phase on the 128x128 image size dataset.

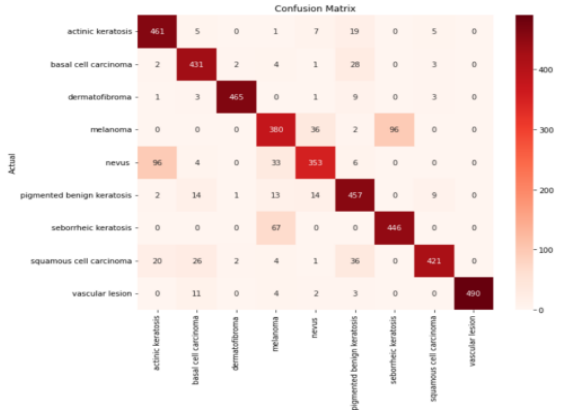

Figure 9. The confusion matrix obtained by the proposed CNN model on the 128x128 test dataset.

3.3. Result proposed CNN model with image size 64x64

The proposed CNN model was trained on a 64x64 dataset. The loss convergence graph in the training phase is given in Figure 10, and the accuracy convergence graph is given in Figure 11. When looking at the convergence graphs, the model showed its best performance approximately from the 35th epoch, and its success continued horizontally after this epoch. The confusion matrix obtained by the model on the test data is given in Figure 12. When looking at the confusion matrix, it was seen that the model incorrectly labeled the "actinic keratosis" class the most.

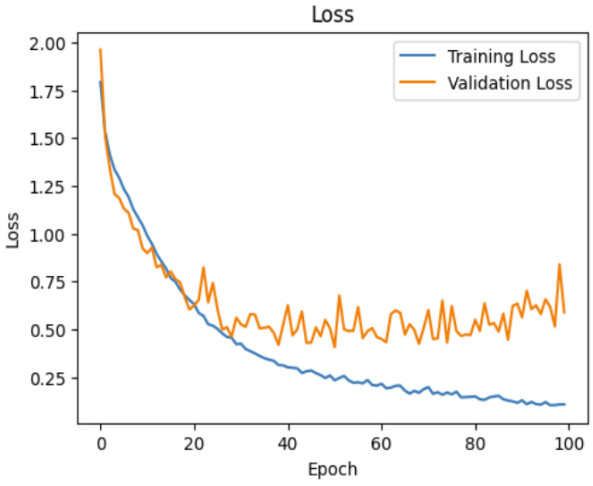

Figure 10. Convergence graph of the training loss and validation loss values obtained by the proposed CNN model in the training phase on the 64x64 image size dataset.

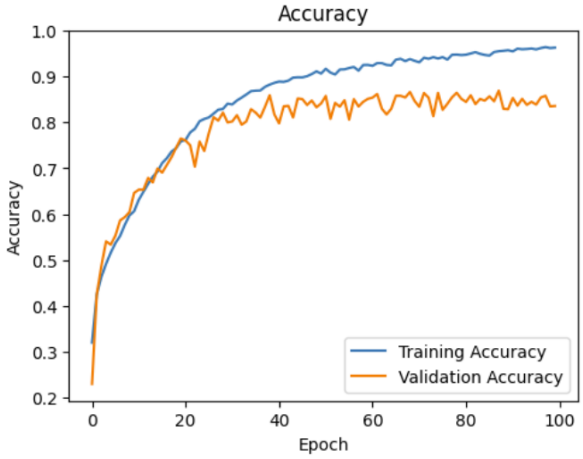

Figure 11. Convergence graph of the training and validation accuracy values obtained by the proposed CNN model in the training phase on the 64x64 image size dataset

Figure 12. The confusion matrix obtained by the proposed CNN model on the 64x64 test dataset.

3.4. Result proposed CNN model with image size 100x100

The loss convergence graph of the CNN model trained on the 100x100 dataset in the training phase is given in Figure 13, and the accuracy convergence graph is given in Figure 14. When the convergence graphs are examined, the model showed its best performance approximately from the 36th epoch, and its success continued horizontally after this epoch. The confusion matrix obtained by the model on the test data is given in Figure 15. When looking at the confusion matrix, it was seen that the model incorrectly labeled the "melanoma" class the most.

Celal Bayar University Journal of Science
Volume 20, Issue 4, 2024, p 82-91
Doi: 10.18466/cbayarfbe.1513945

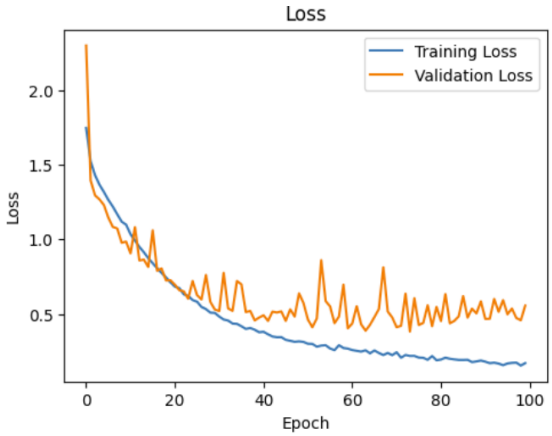

Figure 13. Convergence graph of the training and validation loss values obtained by the proposed CNN model in the training phase on the 100x100 image

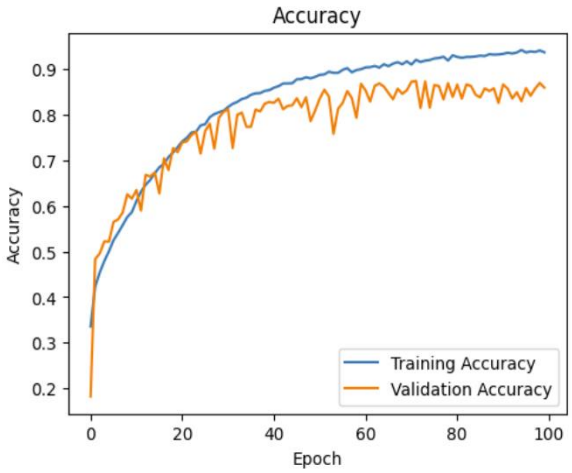

Figure 14. Convergence graph of the training and validation accuracy values obtained by the proposed CNN model in the training phase on the 100x100 image size.

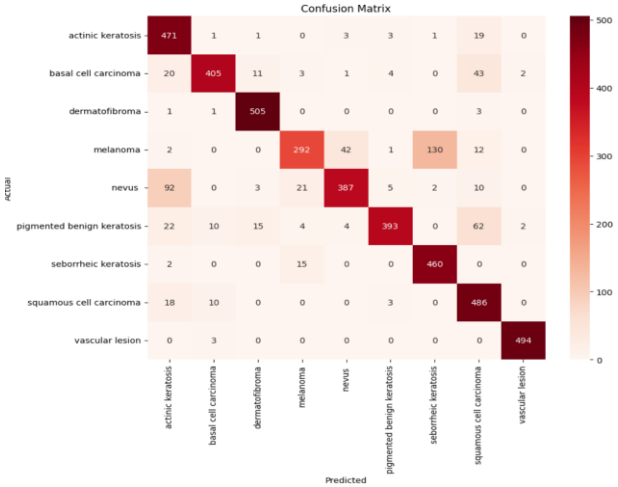

Figure 15. The confusion matrix obtained by the proposed CNN model on the 100x100 test dataset.

3.5. Result of Used Original Popular CNN Model

We first trained a number of well-known Convolutional Neural Network (CNN) models, including InceptionV3, AlexNet, ResNet, and InceptionV3, on the ImageNet dataset in order to tackle our classification task. For our 9-class challenge, the early findings did not yield great accuracy, even with their outstanding feature extraction skills. In order to address this, we modified each of these models before using them again for our classification task. Specifically, we added more thick layers and dropout layers to enhance performance and avoid overfitting.

During preprocessing, we used data augmentation approaches to improve the model's generalization capabilities across various skin conditions and reduce class imbalances within the ISIC dataset. Applying a number of transformations, such as rotation, shifting the width and height, shearing, zooming, and horizontal flipping, was required for this. By enhancing the existing photos, these methods guaranteed that each class was adequately represented in addition to adding diversity to the training data. Next, we made sure the dataset was carefully balanced by limiting the number of images to 1000 for each class. After that, several models were trained and assessed using the balanced and expanded dataset. The accuracy comparison of these models is shown in Table, which also highlights how well they classify various skin disorders.

The parameter numbers of the models in the study are given in Table 3. When the table is examined, it is seen that the proposed CNN model has fewer parameters than all other models.

Table 3. Parameter comparison of the proposed CNN model and the state-of-the-art CNN models.

Models	Number of Parameters	Model Size
Modified AlexNet	46803035	178.54 MB
Modified VGG16	8266761	31.54 MB
Modified ResNet50	24756644	94.44 MB
Modified InceptionV3	48250537	184.06 MB
Original AlexNet	71988013	274.61 MB
Original VGG16	136362825	520.18 MB
Original ResNet50	25695113	98.02 MB
Original InceptionV3	23910185	91.21 MB
Proposed CNN Model (64, 100, 128,224)	**764041**	**465.04 KB**

The comparison of the results obtained in this study with the competitor studies is given in Table 4. The table shows the references of the studies on the ISIC dataset and the accuracy rates obtained in these studies. It is seen that some studies have achieved a higher success rate than the proposed method. However, since the training and test datasets used in the studies are not known, the accuracy values are not an absolute measure of success. Another parameter of model success is model computational complexity. In this context, the proposed CNN model obtained an accuracy value close to other models with a very low number of parameters.

Table 4. Review of the studies performed on the ISIC dataset and comparison of the results of these studies with the proposed method.

Reference	Model	Accuracy (%)
Albahar [33]	CNN and Novel Regularizer	97.49
Sanketh et al. [34]	CNN	98
Daghrir, Tlig et al. [35]	Hybrid approach	85.5
Vipin, Nath et al. [36]	U-Net	88.7
Rahi, Khan et al. [37]	CNN network with CGG16, RESNET50, DENSNET50	90
Jojoa Acosta et al. [38]	CNN combined with pretrained ResNet152	90.4
Rahi et al. [37]	CNN	76
Yu et al. [39]	CNN	94.9
Majtner et al. [40]	HCF+CNN	82.6
Li and Shen [41]	CNN	85.7
Mahbod et al. [42]	CNN	90.5
Zhang et al. [43]	CNN	87.4
Amin et al. [44]	CNN	99.0
Mahbod et al. [45]	CNN	96.6
Kwasigroch et al. [46]	HCF+CNN	77.0
Hameed et al. [47]	HCF+CNN	96.5
Khan et al. [48]	CNN	93.4
Mporas et al. [49]	ML+ANN	74.3
Khan et al. [50]	CNN	92.8
Pereira et al. [51]	HCF	90.0
Khan et al. [52]	CNN	90.7
Proposed Work	Proposed CNN Model_64	84.27
	Proposed CNN Model _100	86.5
	Proposed CNN Model_128	86.76
	Proposed CNN Model _224	83.22
	Modified AlexNet	89.56
	Original AlexNet	73
	Modified VGG16	86.50
	Original VGG16	64
	Modified ResNet50	88.94
	Original ResNet50	53
	Modified InceptionV3	90.56
	Original InceptionV3	64

4. Conclusion

In this paper, a novel CNN model working on RGB images for skin cancer classification is presented. In addition, ResNet, GoogleNet, AlexNet, and VGG16 deep learning models are used by reorganizing them by changing some of their layers. The models are tested using ISIC datasets, with image enhancement in pre-processing utilizing the Image class from the PIL or Pillow library to open and resize images, and np.asarray to translate the resized images into NumPy arrays. Data augmentation techniques are also used. The proposed novel CNN model achieved 84.27%, 86.51%, 86.76%, and 83.22% accuracy rates for different resolutions of 64x64, 100x100, 128x128, and 224x 224 pixel image sizes, respectively. The proposed model has very few parameters and its success metrics are comparable to other studies.

Furthermore, in this study compared a set of pre-trained modified CNN models using both original and customized models to compare their parameters and performance. After doing a comparative analysis, updated VGG16 achieved 86.50%, AlexNet 89.56%, ResNet 88.94%, and InceptionV3 90.56%. In future work, we would like to add more different images and revisions to the training dataset in order to improve the model's ability to generalize to varying skin tones and appearance changes. Optimizing the current model, attempting more complex architectures, or experimenting with different transfer learning algorithms or pre-trained models will all yield better results.

References

1. Leiter, U., U. Keim, and C. Garbe, *Epidemiology of skin cancer: update 2019.* Sunlight, Vitamin D and Skin Cancer, 2020: p. 123-139.

2. Narayanamurthy, V., et al., *Skin cancer detection using non-invasive techniques.* RSC advances, 2018. **8**(49): p. 28095-28130.

3. Singer, S., et al., *Gender identity and lifetime prevalence of skin cancer in the United States.* JAMA dermatology, 2020. **156**(4): p. 458-460.

4. Trager, M.H., et al., *Biomarkers in melanoma and non-melanoma skin cancer prevention and risk stratification.* Experimental dermatology, 2022. **31**(1): p. 4-12.

5. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2018.* CA: a cancer journal for clinicians, 2018. **68**(1): p. 7-30.

6. Jones, O., et al., *Dermoscopy for melanoma detection and triage in primary care: a systematic review.* BMJ open, 2019. **9**(8): p. e027529.

7. Phillips, M., et al., *Detection of malignant melanoma using artificial intelligence: an observational study of diagnostic accuracy.* Dermatology practical & conceptual, 2020. **10**(1).

8. Vestergaard, M., et al., *Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting.* British Journal of Dermatology, 2008. **159**(3): p. 669-676.

9. Carli, P., et al., *Addition of dermoscopy to conventional naked-eye examination in melanoma screening: a randomized study.* Journal of the American Academy of Dermatology, 2004. **50**(5): p. 683-689.

10. Al-Masni, M.A., 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, 2020. **190**: p. 105351.

11. Hasan, M.K., et al., *Dermo-DOCTOR: A framework for concurrent skin lesion detection and recognition using a deep convolutional neural network with end-to-end dual encoders.* Biomedical Signal Processing and Control, 2021. **68**: p. 102661.

12. Hasan, M.K., et al., *DSNet: Automatic dermoscopic skin lesion segmentation.* Computers in biology and medicine, 2020. **120**: p. 103738.

13. Esteva, A., et al., *Dermatologist-level classification of skin cancer with deep neural networks.* nature, 2017. **542**(7639): p. 115-118.

14. Marchetti, M.A., et al., *Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images.* Journal of the American Academy of Dermatology, 2018. **78**(2): p. 270-277. e1.

15. Haenssle, H.A., et al., *Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists.* Annals of oncology, 2018. **29**(8): p. 1836-1842.

16. Brinker, T.J., et al., *A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task.* European Journal of Cancer, 2019. **111**: p. 148-154.

17. Brinker, T.J., et al., *Skin cancer classification using convolutional neural networks: systematic review.* Journal of medical Internet research, 2018. **20**(10): p. e11936.

18. İnik, Ö., et al., *A new method for automatic counting of ovarian follicls on whole slide histological images based on convolutional neural network.* Computers in biology and medicine, 2019. **112**: p. 103350.

19. Celik, M. and O. Inik, *Development of hybrid models based on deep learning and optimized machine learning algorithms for brain tumor Multi-Classification.* Expert Systems with Applications, 2024. **238**: p. 122159.

20. Inik, O., et al., *Prediction of Soil Organic Matter with Deep Learning.* Arabian Journal for Science and Engineering, 2023. **48**(8): p. 10227-10247.

21. King, G. and L. Zeng, *Logistic regression in rare events data.* Political analysis, 2001. **9**(2): p. 137-163.

22. Zhu, M., et al., *Class weights random forest algorithm for processing class imbalanced medical data.* IEEE Access, 2018. **6**: p. 4641-4652.

23. Han, H., W.-Y. Wang, and B.-H. Mao. *Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning*. in *International conference on intelligent computing*. 2005. Springer.

24. He, H. and E.A. Garcia, *Learning from imbalanced data.* IEEE Transactions on knowledge and data engineering, 2009. **21**(9): p. 1263-1284.

25. LemaÃŽtre, G., F. Nogueira, and C.K. Aridas, *Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning.* Journal of machine learning research, 2017. **18**(17): p. 1-5.

26. Ramentol, E., et al., *Smote-rs b*: a hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using smote and rough sets theory*. Knowledge and information systems, 2012. **33**: p. 245-265.

27. Dong, Q., S. Gong, and X. Zhu, *Imbalanced deep learning by minority class incremental rectification.* IEEE transactions on pattern analysis and machine intelligence, 2018. **41**(6): p. 1367-1381.

28. Mariani, G., et al., *Bagan: Data augmentation with balancing gan.* arXiv preprint arXiv:1803.09655, 2018.

29. Cubuk, E.D., et al. *Autoaugment: Learning augmentation strategies from data*. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019.

30. Çelik, M. and Ö. İnik, *Detection of monkeypox among different pox diseases with different pre-trained deep learning models.* Journal of the Institute of Science and Technology. **13**(1): p. 10-21.

31. Ali, M.S., et al., *An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models.* Machine Learning with Applications, 2021. **5**: p. 100036.

32. Chanda, D., et al., *DCENSnet: A new deep convolutional ensemble network for skin cancer classification.* Biomedical Signal Processing and Control, 2024. **89**: p. 105757.

33. Albahar, M.A., *Skin lesion classification using convolutional neural network with novel regularizer.* IEEE Access, 2019. **7**: p. 38306-38313.

34. Sanketh, R.S., et al. *Melanoma disease detection using convolutional neural networks*. in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*. 2020. IEEE.

35. Daghrir, J., et al. *Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach*. in *2020 5th international conference on advanced technologies for signal and image processing (ATSIP)*. 2020. IEEE.

36. Vipin, V., et al. *Detection of melanoma using deep learning techniques: A review*. in *2021 international conference on communication, control and information sciences (ICCISc)*. 2021. IEEE.

37. Rahi, M.M.I., et al. *Detection of skin cancer using deep neural networks*. in *2019 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*. 2019. IEEE.

38. Jojoa Acosta, M.F., et al., *Melanoma diagnosis using deep learning techniques on dermatoscopic images.* BMC Medical Imaging, 2021. **21**: p. 1-11.

39. Yu, L., et al., *Automated melanoma recognition in dermoscopy images via very deep residual networks.* IEEE transactions on medical imaging, 2016. **36**(4): p. 994-1004.

40. Majtner, T., S. Yildirim-Yayilgan, and J.Y. Hardeberg. *Combining deep learning and hand-crafted features for skin lesion classification*. in *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. 2016. IEEE.

41. Li, Y. and L. Shen, *Skin lesion analysis towards melanoma detection using deep learning network.* Sensors, 2018. **18**(2): p. 556.

42. Mahbod, A., et al., *Fusing fine-tuned deep features for skin lesion classification.* Computerized Medical Imaging and Graphics, 2019. **71**: p. 19-29.

43. Zhang, J., et al., *Attention residual learning for skin lesion classification.* IEEE transactions on medical imaging, 2019. **38**(9): p. 2092-2103.

44. Amin, J., et al., *Integrated design of deep features fusion for localization and classification of skin cancer.* Pattern Recognition Letters, 2020. **131**: p. 63-70.

45. Mahbod, A., et al., *Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification.* Computer methods and programs in biomedicine, 2020. **193**: p. 105475.

46. Kwasigroch, A., M. Grochowski, and A. Mikołajczyk, *Neural architecture search for skin lesion classification.* IEEE Access, 2020. **8**: p. 9061-9071.

47. Hameed, N., et al., *Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques.* Expert Systems with Applications, 2020. **141**: p. 112961.

48. Khan, M.A., et al., *Developed Newton-Raphson based deep features selection framework for skin lesion recognition.* Pattern Recognition Letters, 2020. **129**: p. 293-303.

49. Mporas, I., I. Perikos, and M. Paraskevas. *Color models for skin lesion classification from dermatoscopic images*. in *Advances in Integrations of Intelligent Methods: Post-workshop volume of the 8th International Workshop CIMA 2018, Volos, Greece, November 2018 (in conjunction with IEEE ICTAI 2018)*. 2020. Springer.

50. Khan, M.A., et al., *Pixels to classes: intelligent learning framework for multiclass skin lesion localization and classification.* Computers & Electrical Engineering, 2021. **90**: p. 106956.

51. Pereira, P.M., et al., *Skin lesion classification enhancement using border-line features–The melanoma vs nevus problem.* Biomedical Signal Processing and Control, 2020. **57**: p. 101765.

52. Khan, M.A., et al., *Skin lesion segmentation and multiclass classification using deep learning features and improved moth flame optimization.* Diagnostics, 2021. **11**(5): p. 811.