All Sciences Proceedings
http://as-proceeding.com/

2nd International Conference on Engineering, Natural and Social Sciences

April 4-6, 2023 : Konya, Turkey

© 2023 Published by All Sciences Proceedings

https://www.icensos.com/

CANCER LESION CLASSIFICATION WITH GAN-BASED IMAGE AUGMENTATION METHOD FROM SKIN IMAGES

Emre Gören[1], Gökalp Çınarer [2*]

[1] Department of Computer Engineering, Faculty of Engineering, Yozgat Bozok University, Yozgat, Turkey
[2] Department of Computer Engineering, Faculty of Engineering, Yozgat Bozok University, Yozgat, Turkey

*(gokalp.cinarer@bozok.edu.tr)

Abstract – Skin cancer is one of the most common types of cancer in the world and an important public health problem. Rapid diagnosis and correct treatment are important in the treatment of skin cancer. However, the in-depth imaging and analysis required to diagnose skin cancer is a time-consuming process that requires expertise. Unfortunately, the lack of labeled data in the classification and the limited size of training datasets affect the performance of deep learning architectures and the correct diagnosis of the disease. In this study, cancer lesion classification was performed by using deep convolution networks with a GAN (Generative Adversarial Network) based data augmentation method for skin lesion classification. This model synthesizes new skin lesion images by analyzing skin lesion images in training data. The images synthesized by GAN are processed by the classification model in the same way as the real images and added to the training set for better classification performance. In all classification tasks, 3 different convolutional neural network architectures were used for classification and feature extraction. The classification accuracy of the architectures consisting of images obtained with the ISIC (International Skin Imaging Cooperation) Archive dataset is 86.00% for CNN, 90.81% for MobileNet, and 81.66% for ResNet-18, respectively. The performance of the models in the hybrid dataset consisting of the combination of synthetic images and the original primary dataset increased to 86.34% for CNN, 91.67% for MobileNet and 91.78% for ResNet-18. This study shows that GAN-based data enhancement models in skin lesion classification also make an important contribution to medical image analysis.

Keywords: Skin Cancer, Deep Learning, GAN, Artificial Intelligence, CNN

I. INTRODUCTION

Tumors formed by a rapid and abnormal increase in lesions on the skin are the basis of skin cancers. The most important and harmful type is melanoma, which is quite deadly [1]. Melanoma is an important type of cancer that also causes a lot of people to lose their lifes every year in the world. Lesion detection methods applied by dermatologists are used in the observation of skin cancer. The examination and reporting of the images obtained at this time are quite difficult and time-consuming processes. At the same time, while performing all these operations, both the error rate increases and it takes time. Deep learning algorithms are able to process very large data sets in a short time and analyze them with high performance with the technology that has developed in recent years. Convolutional neural networks (CNN) are the most popular in this field. When the literature is examined, there are many studies aimed at detecting benign and malignant tumors with CNN architecture. Daghrir et al. (2020) determined the boundaries of 640 dermoscopy images obtained from the ISIC archive using the morphological geometric active contour (morpGAC) method and extracted texture and color characteristics. They classified the obtained properties with the SVM algorithm and obtained an accuracy value of 71.8% [2]. Premaladha and Ravichandran (2016) performed segmentation by Otsu thresholding method after applying pre-processing to a total of 992 images in the Med-Node and PH2 data sets. They extracted the geometric properties of the relevant lesion area and tissue properties using the GLCM algorithm. The extracted features have been classified with deep neural networks. As a result of the study, they reached a 92.8% accuracy value [3]. Ismail et al. (2020) extracted the color characteristics of 1635 dermoscopy images taken from the ISIC archive, as well as the characteristics obtained by HOG, scale invariant property transformation (SIFT) and GIST methods. By classifying these characteristics with the Random Forest algorithm, they reached 67.5% accuracy [4].In the studies conducted, machine learning methods have been preferred in lesion detection.

Since Deep Learning methods rely on the support of labeled data, which is rare in the medical field, researchers recommend using data augmentation techniques to increase the learning ability of the model [5],[6]. In general, simple image changes such as scaling, rotation, flipping and scrolling are the most common methods of enlarging image data. When the data enhancement methods applied recently are examined, studies in which preprocessions such as image resizing [7][8], dog filter, thresholding, masking [9], noise filtering, CLAHE, contrast, edge detection [3] are applied in the data are seen. These classical methods have become a standard process of the educational process of networks [10]. However, these changes made in the original data set do not affect the study success very much. On the other hand, the proposal of data synthesis as a method of enlarging a complex data set has emerged. Variational autoencoders (VAES) [10] and generative adversarial networks (GANS) [11] are the two most straightforward methods for data creation. Although their training is challenging and unbalanced, vaes are capable of producing pictures that are realistic and sharp. Gans demonstrate their usefulness in medical imaging [11],[12], but there hasn't been any thorough investigation into the picture synthesis of skin lesions. However, systematic studies and developed methods on skin lesion imaging remain insufficient. The use of traditional methods such as visual examination, scanning in a clinical setting, dermoscopic image analysis, histopathological examination of a skin lesion by biopsy requires a high degree of skill, concentration and takes time [13]. The main contributions of this article can be listed as follows:

A WGAN-GP based data augmentation model has been proposed in the analysis of skin cancer images. The recommended modle is effective in creating skin lesion images with high-quality and unique values. Three different architectures were used in the classification of the images and the results of each architecture were compared with the results obtained with a hybrid dataset. GAN model, which is seen that the performance of the classification model increases even more in the detection of skin cancer. The results obtained with popular CNN architectures are presented in a comparative tabular form in the study. It makes an important contribution to the accurate detection of skin cancer images.

The advanced parts of this article are as follows. The datasets and architectures used in the study are given in the materials and methods section in Section 2. The performance of this Gan model and the results obtained Part 3.it is also given with detailed tables and figures. Finally, Part 4.in the discussion section, a comparison was made with similar studies.

II. MATERIALS AND METHOD

A. Dataset

The study utilized a dataset consisting of 3297 dermoscopic skin cancer images created by the ISIC Archive [14] and shared on the Kaggle platform. The dataset is categorized into two classes: benign and malignant. It contains 1800 benign and 1497 malignant skin cancer images, all of which are 24-bit RGB format images with 82 dpi and a uniform size of 224 pixels [15]. Some skin images from this dataset are provided in Figure 1.

In the study, 660 test images and 2637 train images were used. The test set includes 300 malignant and 360 benign images. The train set includes 1197 malignant and 1440 benign images.

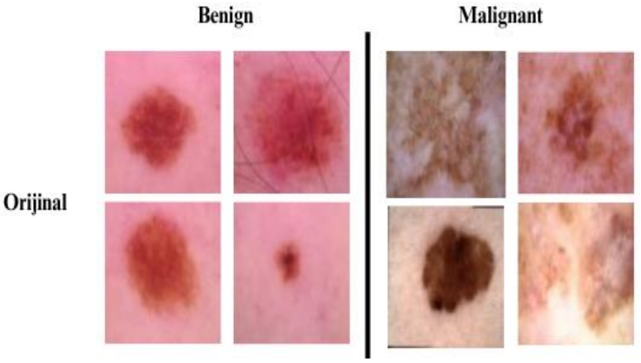


Fig. 1. Some images in the ISIC dataset

B. GAN Based Data Augmentation

Data augmentation and sampling techniques are commonly used on imbalanced datasets to address the issue of insufficient data. Since their introduction, GANs have emerged as one of the most notable developments in deep learning methods [11]. Preprocessing is applied to images to improve classification accuracy. The images, which have a size of 224x224, are preprocessed by resizing them to 64x64 to increase the quantity of data. To improve the performance of deep learning architectures, GAN processing is applied to images to enhancement the number of images in the dataset. In this study, unlike traditional GAN models, data augmentation was performed using the WGAN-GP algorithm.

C. WGAN-GP

There are two basic network structures in gan algorithms which area discriminator and generator. The generator network aims to simulate real images from the training set using noise vectors as input. The discriminator tries to distinguish between synthetic (fake) images and real training images [16]. The design of the discriminator and generator in the WGAN-GP algorithm is the same as GAN and WGAN algorithms. [17]. Unlike traditional GANs, WGAN-GP optimizes a Wasserstein distance metric, which requires the outputs of its training steps to be at a unit distance. This feature allows WGAN-GP to produce more diverse and realistic outputs by preventing the problem of mode collapse, where a certain portion of the generated model is repeatedly produced, thus reducing the diversity of outputs. Another feature of WGAN-GP is the use of gradient penalty technique during training [18]. This technique involves penalizing the output of the gradient penalty discriminator model when it deviates from 1. This process prevents irregularities in the discriminator model's training and ensures a more stable training process. In WGAN-GP, the gradients of the discriminator model are calculated frequently and updated during training. Therefore, we used WGAN-GP in this study. The critic neural network model is seen in Table 1.

Table 1. The critical neural network model

Layer No.	Layer Type	Output Shape	Number of Parameters
0	Conv2d	(batch_size, 64, 64, 64)	3,072
1	InstanceNorm2d	(batch_size, 64, 64, 64)	128
2	LeakyReLU	(batch_size, 64, 64, 64)	0
3	Conv2d	(batch_size, 128, 32, 32)	131,200
4	InstanceNorm2d	(batch_size, 128, 32, 32)	256
5	LeakyReLU	(batch_size, 128, 32, 32)	0
6	Conv2d	(batch_size, 256, 16, 16)	524,544
7	InstanceNorm2d	(batch_size, 256, 16, 16)	512
8	LeakyReLU	(batch_size, 256, 16, 16)	0
9	Conv2d	(batch_size, 512, 8, 8)	2,097,664
10	InstanceNorm2d	(batch_size, 512, 8, 8)	1,024
11	LeakyReLU	(batch_size, 512, 8, 8)	0
12	Conv2d	(batch_size, 1, 5, 5)	8,193

D. Discriminator

The Discriminator takes a tensor of size [batch_size, channels, 64, 64] as input. The batch_size express to the number of images for a batch, and the channels are organized based on the number of input channels (e.g. 3 for RGB images). The Discriminator applies a series of convolutional layer to extract feature from input tensor. Finally applies a convolutional layer to generate a scalar value that evaluates the realism of the input image as input to the critic.

In the first convolutional layer there are 64 filters and their size 4x4 with a stride of two and padding of one. After this batch normalization and a LeakyReLU activation function comes with a slope of 0.2. Subsequent convolutional layers aim to reduce the spatial resolution of the feature maps while increasing the number of filters. The last convolutional layer has a single filter of size 4x4 with a stride of 1 and is designed to generate a scalar value that evaluates the realism of the input image.

Batch normalization was used in intermediate layers to normalize the activations for each example separately, without relying on the distribution of examples within the batch. The LeakyReLU activation function was used to prevent the "dead ReLU" problem and stop weight updates. The Discriminator neural network model parameters are provided in Table 2.

Table 2. Discriminator neural network model

Layer (type)	Output Shape	Parameter
Conv2d-1	[-1, 224, 112, 112]	10,976
InstanceNorm2d-2	[-1, 224, 112, 112]	448
LeakyReLU-3	[-1, 224, 112, 112]	0
Conv2d-4	[-1, 448, 56, 56]	1,606,080
InstanceNorm2d-5	[-1, 448, 56, 56]	896
LeakyReLU-6	[-1, 448, 56, 56]	0
Conv2d-7	[-1, 896, 28, 28]	6,423,424
InstanceNorm2d-8	[-1, 896, 28, 28]	1,792
LeakyReLU-9	[-1, 896, 28, 28]	0
Conv2d-10	[-1, 1792, 14, 14]	25,691,904
InstanceNorm2d-11	[-1, 1792, 14, 14]	3,584
LeakyReLU-12	[-1, 1792, 14, 14]	0
Conv2d-13	[-1, 1, 11, 11]	28,673
Total params: 33,767,777	**Trainable params:** 33,767,777	**Non-trainable** params: 0

E. Generator

A GAN creates a competition between a model that generates fake images (the Generator) and a model that can distinguish between the fake and real images (the Discriminator). The Generator takes a random noise vector (z) of the same size as the noise vector and transforms it into an image of size [batch_size, channels, 64, 64]. Here, channels are 3 for RGB images. After transforming the input vector into an image using a series of transpose convolutional layers, the Generator model increases the size of the image by increasing the spatial resolution of the feature maps (from 4x4 to 64x64) and reducing the number of the filters. After each transpose convolutional layer, it follows a BatchNorm2d for normalization and a LeakyReLU activation function. Batch normalization normalizes the layer output and prevents mode collapse. Also, the LeakyReLU activation function helps to prevent the vanishing gradient problem. In the last transpose convolutional layer, the number of channels is set to the same as the input channel count, and the output is scaled to the range of [-1,1] using the Tanh activation function. This allows the Generator model to bring the pixel values of its outputs into a suitable range for images. The Generator model can be used with the forward method and is called with a random noise vector (z) of the noise vector size. The Generator neural network model is given in Table 3.

Table 3. Generator neural network model

Layer (type)	Output Shape	Param eter
1-ConvTranspose2d	[-1, 1792, 4, 4]	5,736,192
2-BatchNorm2d	[-1, 1792, 4, 4]	3,584
3-ReLU	[-1, 1792, 4, 4]	0
4-ConvTranspose2d	[-1, 896, 8, 8]	25,691,008
5-BatchNorm2d	[-1, 896, 8, 8]	1,792
6-ReLU	[-1, 896, 8, 8]	0
7-ConvTranspose2d	[-1, 448, 16, 16]	6,422,976
8-BatchNorm2d	[-1, 448, 16, 16]	896
9-ReLU	[-1, 448, 16, 16]	0
10-ConvTranspose2d	[-1, 224, 32, 32]	1,605,856
11-BatchNorm2d	[-1, 224, 32, 32]	448
12-ReLU	[-1, 224, 32, 32]	0
13-ConvTranspose2d	[-1, 3, 64, 64]	10,755
14-Tanh	[-1, 3, 64, 64]	0
Total params: 39,473,507	**Trainable params:** 39,473,507	**Non-trainable** params: 0

F. Wasserstein Distance

The Wasserstein distance is a measure that quantifies the difference between two probability distributions. This distance is commonly used in image, video, and natural language processing to measure the similarities between two probability distributions. WGAN learns whether the generator is performing well or not. For WGAN the generator learns and the gradient is smoother everywhere, and better even if it does not produce good images [19].

With the GAN-based data augmentation technology, a total of 1767 photos were synthesized, including 883 benign and 884 malignant skin lesion images, for the architecture. The images in the test sets are unique compared to the ones in the training set.

According to the newly created data set, there are 460 test images and 2219 training images for benign, and 404 test images and 1981 training images for malignant. The images obtained with the WGAN-GP data augmentation model are shown in Figure 2.

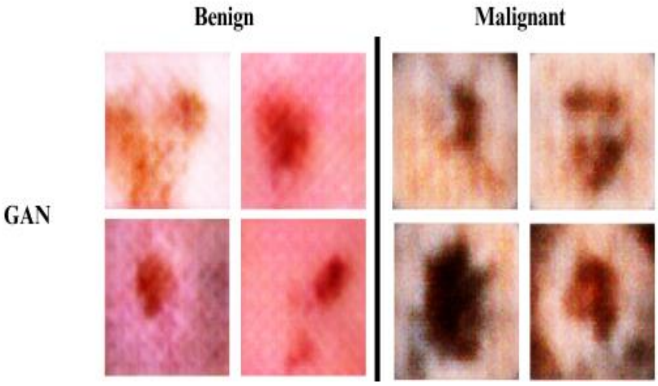


Fig. 2. Examples of images obtained with the WGAN-GP data augmentation model

G. Deep Learning Classification Architectures

Deep learning is an algorithm architecture based on artificial neural networks. It is a different architecture from traditional neural networks, having more hidden layers. In deep neural networks, input values are generated by passing through activation functions of certain weights. Thus, the correct optimization of the network is targeted and the error value is tried to be minimized. In this study, deep learning architectures used are briefly described.

H. CNN

CNN architecture is particularly effective for algorithmic problems such as recognition and image processing. This architecture accepts data as 2 or 3-dimensional and typically performs classification as output. While processing image data, this architecture learns by using the local structure of the data, particularly observing similarities between different images.

This converts the information obtained from previous layers into higher-level features and learns more complex features at each level.

İ. ResNet

ResNet, also known as "Residual Networks" in Turkish, is a neural network architecture developed by authors [20] in 2015. It demonstrated excellent performance and won first place in the ImageNet competition. In convolutional neural network architectures, increasing the number of layers also increased the error rate. However, ResNet architecture, which has many more layers compared to other architectures, was able to overcome this issue with residual blocks. In residual blocks, the input x is passed through a convolutional operation and ReLU activation function, and then the original x input is added to the result of the F(x) function to obtain the $H(x) = F(x) + x$ result. The input image size of the network is 64x64, and the filter size used in convolutional layers is fixed at 3x3. ReLU activation function is used, and maximum pooling is applied in pooling layers. Softmax function is used for classification [20]. ResNet architecture has 18-34-50-101-152 layer versions that are commonly used for image classification and object recognition problems.

J. MobileNet

The first mobile computer sense model in TensorFlow, the MobileNet model [21] is intended for use in mobile apps. MobileNet's use of depthwise separable convolutions greatly organized the size of parameter when compared to networks using organized convolutions of the same deep. This produces thin deep neural networks and makes it possible to separate convolutional cores.

K. Classification

In the study, two different classifications were carried out. For all classification tasks, hybridized GAN data with the ISIC Archive dataset were used. Processed images and raw original images were used as two different datasets. The total of 5064 images obtained were used as 864 test and 4200 train data. The training, validation, and test images in both datasets consist of the same images. Preprocessing was applied to images to increase classification accuracy. The images with a size of 224x224 were resized to 64x64 after preprocessing. The GAN technique was applied to prevent overfitting and improve performance of the trained models. The images in the test sets and validation sets are unique images from the training set.

The CNN, ResNet-18, and MobileNet architectures were used for classification and feature extraction processes. The architectures were trained from scratch without using transfer learning approach.

Deep learning applications were carried out on Google Colab. The iteration number for the GAN-trained model was set to 1000 and batch size 32. For MobileNet, the iteration number was set to 50 with a batch size of 64, for CNN the iteration number was set to 50 with a batch size of 32, and for ResNet-18 the iteration number was set to 100 and batch size to 32. The Adam optimization algorithm was used for the models' optimization and parameter updates, while the ReLU activation function was used in convolutional layers. Classification performance of the architectures was compared with both preprocessed and unprocessed original data for the validation and test sets. The classification performance of the best-performing architectures for each classification task was demonstrated through the loss and accuracy curves during training, ROC curve, and confusion matrix.

L. *Evaluation Criteria*

In the classification process, confusion matrix values were taken into account. Since a binary comparison analysis was carried out in the study, accuracy, recall, f1 score, and precision values were calculated by comparing the real values of the images with the predicted values using a 2x2 confusion matrix. Metrics that measure the performance of the classifier were calculated using the TP, TN, FN, and FP values obtained from the confusion matrix, and the performances of the algorithms were compared and analyzed.

III. Results

In the study, new hybrid datasets are created by adding modifications to the ISIC Archive dataset. Each dataset is trained with three architectures. Table 4. contains the training results of the ISIC Archive dataset. The MobileNet architecture has the highest success rate of 0.9081.

A hybrid dataset is created by combining the GAN-generated dataset with the ISIC Archive dataset. The results of the hybrid dataset training are shown in Table 5. MobileNet architecture is the most successful architecture in the hybrid dataset with a success rate of 0.9167. When compared to the training results of the original dataset, it can be seen that the hybrid dataset training results are more successful.

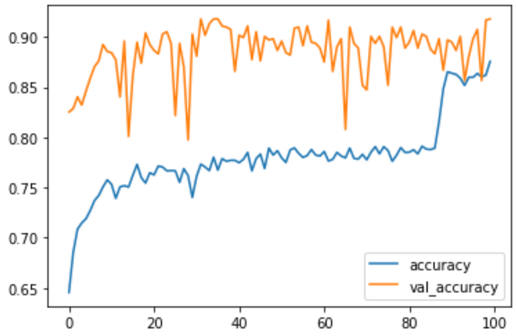
Table 4. Results of models trained with ISIC Archive dataset

Model	Dataset	Accuracy	Loss	Val Acc
CNN	ISIC Archive	0.8674	0.2564	0.8900
MobileNet	ISIC Archive	0.9081	0.0732	0.9200
ResNet-18	ISIC Archive	0.8166	0.4212	0.9100

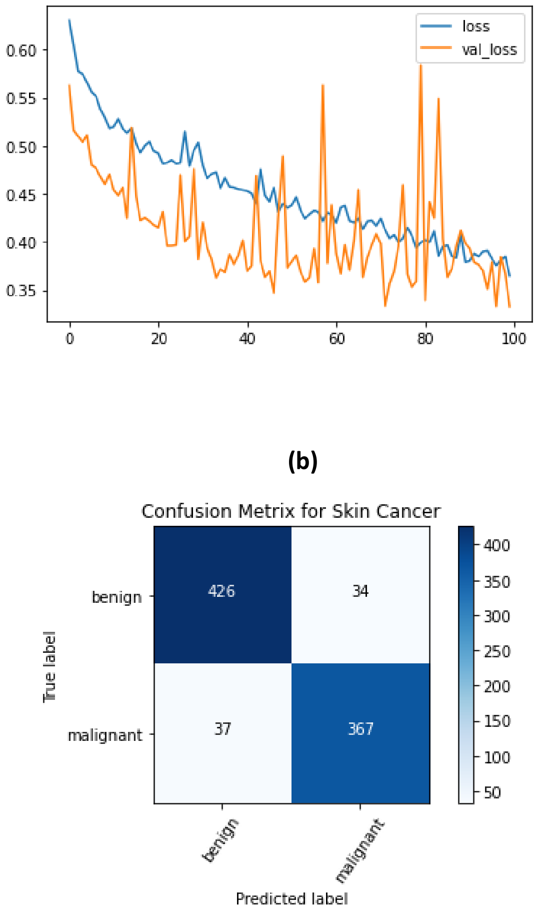
Table 5. Results of models trained with ISIC Archive + Gan dataset

Model	Dataset	Acc	Loss	Val Acc
CNN	ISIC Archive + GAN	0.8634	0.2821	0.8700
MobileNet	ISIC Archive + GAN	0.9167	0.1829	0.9200
Resnet-18	ISIC Archive, +GAN	0.9178	0.3300	0.9100

In the study conducted with the hybrid data set, the accuracy, loss graphs and confusion matrix values of our training results of the Resnet model are given in Figure 3.



(a)



(b)

(c)

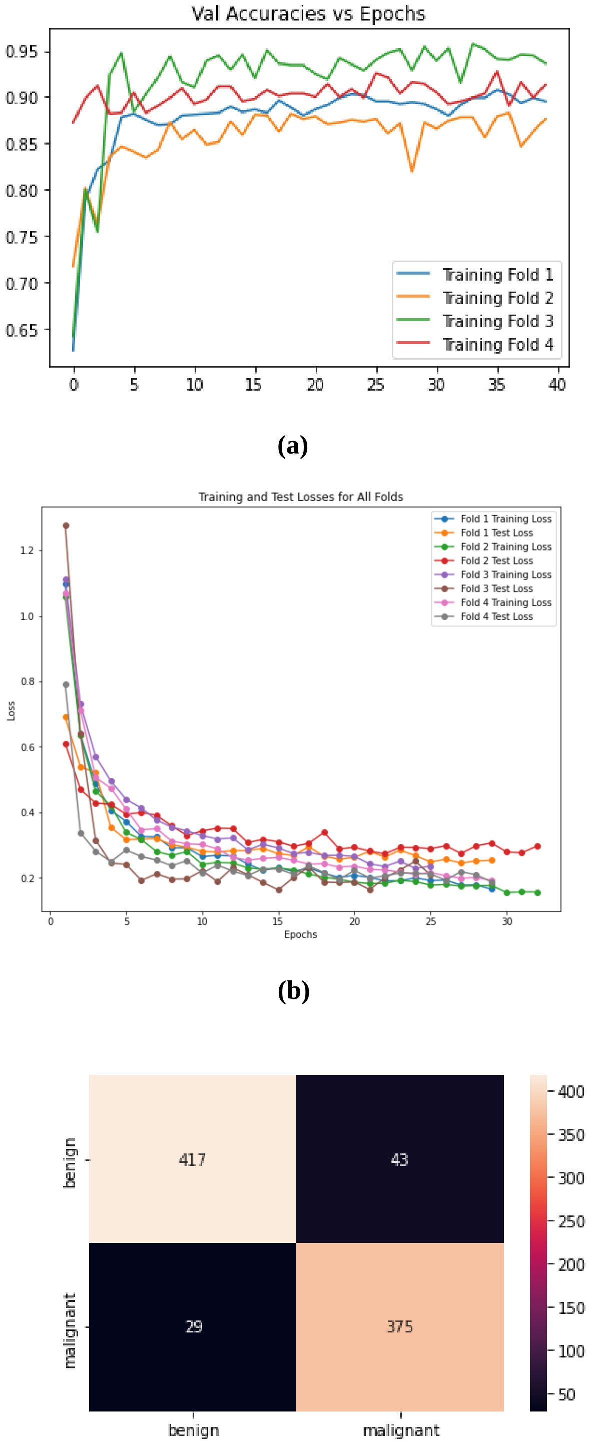
Fig. 3. ResNet-18's (a) accuracy value (b) loss value (c) confusion matrix value

The results of the Resnet architecture in the study with the hybrid data set are given in Table 6.

Table 6. ResNet-18’s Classification Report

Measure	**Value**	**Formuls**
Sensitivity	0.9201	TP / (TP + FN)
Specificity	0.9152	TN / (FP + TN)
Precision	0.9261	TP / (TP + FP)
Accuracy	0.9178	(TP + TN) / (P + N)
F1 Score	0.9231	2TP / (2TP + FP + FN)

In the study conducted with the hybrid data set, the accuracy and loss graphs and confusion matrix values of our training results of the MobileNet model are given in Figure 4.



(a)

(b)

(c)

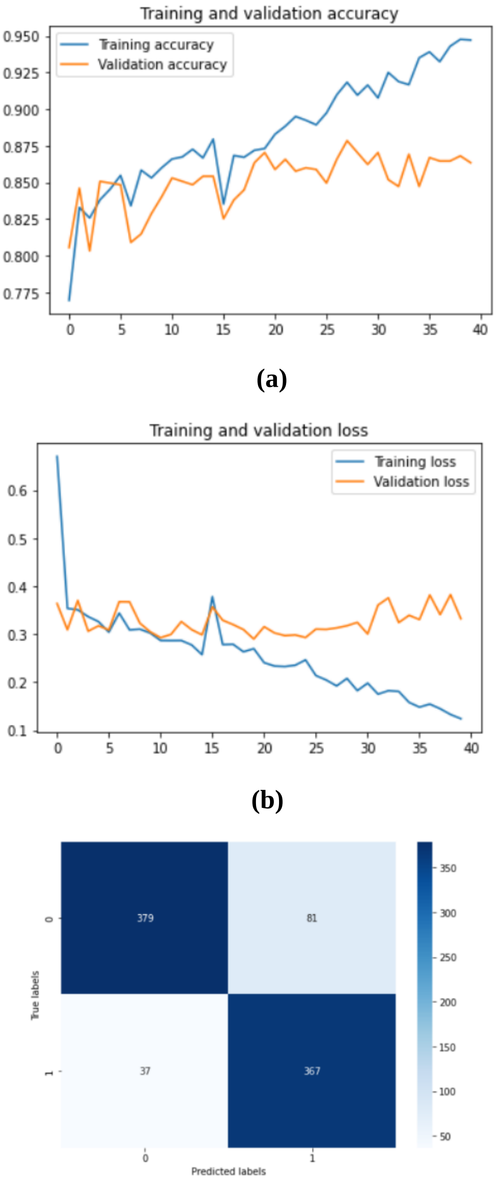
Fig. 4. MobileNet’s (a) accuracy value (b) loss value (c) confusion matrix value

The results of the Mobilenet architecture in the study with the hybrid data set are given in Table 7.

Table 7. MobileNet's Classification Report

Measure	Value	Formuls
Sensitivity (recall)	0.9350	TP / (TP + FN)
Specificity	0.8971	TN / (FP + TN)
Precision	0.9065	TP / (TP + FP)
Accuracy	0.9167	(TP + TN) / (P + N)
F1 Score	0.9205	2TP / (2TP + FP + FN)

In the study conducted with the hybrid data set, the accuracy and loss graphs and confusion matrix values of our training results of the CNN model are given in Figure 5.



(a)

(b)

(c)

Fig. 5. CNN's (a) accuracy value (b) loss value (c) confusion matrix value

The results of the CNN architecture in the study with the hybrid data set are given in Table 8.

Table 8. CNN Classification Report

Measure	Value	Formuls
Sensitivity	0.9111	TP / (TP + FN)
Specificity	0.8192	TN / (FP + TN)
Precision	0.8239	TP / (TP + FP)
Accuracy	0.8634	(TP + TN) / (P + N)
F1 Score	0.8653	2TP / (2TP + FP + FN)

IV. DISCUSSION

In the field of artificial intelligence, hybrid datasets are used in various areas, especially in image processing applications. Models trained with hybrid datasets have been found to produce better results compared to those trained with normal datasets, especially when using the CNN MobileNet and ResNet architectures. Studies have shown that models trained with hybrid datasets achieve higher accuracy than models trained with normal datasets. It is seen that hybrid datasets contain a larger variety and range of patterns due to the combination of synthetic and real data, which helps the model learn better and produce more accurate results.

Furthermore, models trained with hybrid datasets have been found to have less overfitting problems compared to models trained with normal datasets. Overfitting occurs when the model overfits to the training data and performs poorly on test data. Models trained with hybrid datasets have a greater diversity of data due to the combination of synthetic and real data, which reduces the overfitting problem.

Future studies aim to increase accuracy values by using different data augmentation techniques in new deep learning architectures. Additionally, the performance of models trained with different datasets will be evaluated.

REFERENCES

[1] Afza, F., Sharif, M., Khan, M.A., Tariq, U., Yong, H.-S. & Cha, J. (2022). Multiclass skin lesion classification using hybrid deep features selection and extreme learning machine. Sensors, 22(3), 799. doi: 10.3390/s22030799

[2] Daghrir, J., Tlig, L., Bouchouicha, M., Sayadi, M. (2020). 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) (pp. 1-5). IEEE.

[3] Premaladha, J., Ravichandran, K. S. (2016). Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms. Journal of medical systems, 40(4), 1-12.

[4] Ay, İ. C. (2021). COVID-19 Pandemisinin Türkiye'nin İhracatı Üzerine Etkileri İçin Bir Analiz. JOEEP: Journal of Emerging Economies and Policy, 6(1), 272-283.

[5] Ayan, E., & Ünver, H. M. (2018, April). Data augmentation importance for classification of skin lesions via deep learning. In 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT) (pp. 1-4). IEEE.

[6] Zhang, J., Xie, Y., Wu, Q., & Xia, Y. (2018, September). Skin lesion classification in dermoscopy images using synergic deep learning. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 12-20). Springer, Cham.

[7] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. nature, 542(7639), 115-118

[8] Lopez, A. R., Giro-i-Nieto, X., Burdick, J., & Marques, O. (2017, February). Skin lesion classification from dermoscopic images using deep learning techniques. In 2017 13th IASTED international conference on biomedical engineering (BioMed) (pp. 49-54). IEEE.

[9] Daghrir, J., Tlig, L., Bouchouicha, M., Sayadi, M. (2020). 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) (pp. 1-5). IEEE.

[10] Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing, 321, 321-331.

[11] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

[12] Kazeminia, S., Baur, C., Kuijper, A., van Ginneken, B., Navab, N., Albarqouni, S., & Mukhopadhyay, A. (2020). GANs for medical image analysis. Artificial Intelligence in Medicine, 109, 101938.

[13] Xie Y, Zhang J, Xia Y, Shen C. 2020. A mutual bootstrapping model for automated skin lesion segmentation and classification. IEEE Transact on Medical Imag, 39(7): 2482-2493.

[14]. Kaggle, "Skin Cancer: Malignant vs. Benign" (Public access processed Skin cancer images from ISIC Achieve) https://www.kaggle.com/fanconic/skin-cancer-malignant-vs-benign.

[15]. Codella, N., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S., Gutman, D., ... & Halpern, A. (2019). Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368*.

[16] Goodfellow, I., NIPS 2016 Tutorial: Generative Adversarial Networks, arXiv:1701.00160, (2016).

[17] Lee, J., & Lee, H. (2022). Improving SSH detection model using IPA time and WGAN-GP. Computers & Security, 116, 102672.

[18] Gao, X., Deng, F., & Yue, X. (2020). Data augmentation in fault diagnosis based on the Wasserstein generative adversarial network with gradient penalty. Neurocomputing, 396, 487-494.

[19] Arjovsky, M., Chintala, S., & Bottou, L. (2017, July). Wasserstein generative adversarial networks. In International conference on machine learning (pp. 214-223). PMLR.

[20] Öztürk, Ş., & Özkaya, U. (2020). Skin lesion segmentation with improved convolutional neural network. Journal of digital imaging, 33, 958-970.

[21]. A. G. Howard, M. Zhu, and B. Chen, "Mobilenets: efficient convolutional neural networks for mobile vision applications," 2017, https://arxiv.org/abs/1704.04861.