



Hybrid Convolutional Neural Network Architectures for Skin Cancer Classification

Emine Cengil^{1*}, Muhammed Yıldırım², Ahmet Çınar³

^{1*} Firat University, Faculty of Engineering, Department of Computer Engineering, Elazığ, Turkey, (ORCID: 0000-0003-4313-8694), ecengil@firat.edu.tr

² Firat University, Faculty of Engineering, Department of Computer Engineering, Elazığ, Turkey, (ORCID: 0000-0003-1866-4721), muyldrm@gmail.com

³ Firat University, Faculty of Engineering, Department of Computer Engineering, Elazığ, Turkey, (ORCID: 0000-0001-5528-2226), acinar@firat.edu.tr

(1st International Conference on Applied Engineering and Natural Sciences ICAENS 2021, November 1-3, 2021)

(DOI: 10.31590/ejosat.1010266)

ATIF/REFERENCE: Cengil, E., Yıldırım, M. & Çınar, A. (2021). Hybrid Convolutional Neural Network Architectures for Skin Cancer Classification. *European Journal of Science and Technology*, (28), 694-701.

Abstract

Skin cancer is a common form of cancer seen in humans. Like other diseases, early diagnosis of skin cancer is vital. In the study, deep learning architectures, which are popular machine learning algorithms, are used to classify skin cancer. In order to increase accuracy performance, hybrid structures are realized using K-Nearest neighbor (KNN), Support vector machine (SVM) and Decision tree (DT). After feature extraction using convolutional neural network, KNN, SVM and DT are applied separately for classification. While the KNN and SVM of the produced hybrid structures increase performance, the use of the decision tree has negatively affected the performance. After the training and validation processes with the seven-class skin cancer mnist: ham10000 dataset containing dermatological images, the validation accuracy and confusion matrix criteria of the architectures are reported. Eight different architectures are implemented. The highest accuracy is provided by the structure in which the last layer of Alexnet architecture is replaced by the SVM classifier.

Keywords: Skin Cancer, Classification, CNN, Deep Learning, Hybrid Structures, KNN, SVM, DT.

Cilt Kanseri Sınıflandırması için Hibrit Evrişimli Sinir Ağı Mimarileri

Öz

Deri kanseri, insanlarda görülen yaygın bir kanser türüdür. Diğer hastalıklarda olduğu gibi cilt kanserinde de erken teşhis hayati önem taşımaktadır. Çalışmada cilt kanserini sınıflandırmak için popüler makine öğrenmesi algoritmaları olan derin öğrenme mimarileri kullanılmaktadır. Doğruluk performansını artırmak için K-En yakın komşu (KNN), Destek vektör makinesi (SVM) ve Karar ağacı (DT) kullanılarak hibrit yapılar gerçekleştirilmektedir. Evrişimli sinir ağı kullanılarak öznetelik çıkarıldıktan sonra, sınıflandırma için KNN, SVM ve DT ayrı ayrı uygulanır. Üretilen hibrit yapıların KNN ve SVM'si performansı artırırken, karar ağacının kullanılması performansı olumsuz etkilemektedir. Dermatolojik görüntüleri içeren yedi sınıflı cilt kanseri mnist:ham10000 veri seti ile yapılan eğitim ve doğrulama işlemlerinden sonra mimarilerin doğrulama doğruluğu ve karmaşıklık matrisi kriterleri raporlanır. Sekiz farklı mimari uygulanmaktadır. En yüksek doğruluk, Alexnet mimarisinin son katmanının SVM sınıflandırıcısı ile değiştirildiği yapı tarafından sağlanmaktadır.

Anahtar Kelimeler: Deri Kanseri, Sınıflandırma, ESA, Derin Öğrenme, Hibrit Yapılar, KNN, SVM, DT.

* Corresponding Author: ecengil@firat.edu.tr

1. Introduction

Image classification is a machine learning problem that can be related to many topics and interests most areas. Medicine is one of these fields. Classification of medical images is a challenging task. Accurate diagnosis and detectability are critical for a general skin cancer prevention policy, given the aging population and limited health resources. Skin cancer is a common type of disease that can lead to death if not treated. As with all diseases, early and accurate detection of skin cancer is very important.

Skin cancer, the most frequent human malignancy, is detected clinically, with the first screening followed by dermoscopic examination, and if necessary, histological testing. Because of the fine grained heterogeneity in the manifestation of skin lesions, automatic classification of the lesions utilizing images is a challenging task [1]. Computer Aided Diagnosis systems can help physicians make better decisions. Machine learning is widely and effectively used in computer assisted diagnosis systems.

There are many studies in the literature for the diagnosis of skin cancer. O. Yildiz [2] proposes a system that can diagnose skin cancer type melanoma automatically. Designed for the diagnosis of melanoma, it was made with a deep neural network model called C4Net. It showed high classification success. H. Chang [3] proposes a new transfer learning-based deep network for the diagnosis of melanoma, a type of skin cancer. The network achieves high performance in the dataset applied. To identify melanoma in color images captured by dermoscopy equipment, E.I.Unlu et al. [4] employ a deep learning-based approach. The designed method makes use of the International Skin Imaging Collaboration database.

N. Codella et al. [5] propose a deep learning, sparse coding, and SVM learning system for melanoma detection in dermoscopy pictures. Unsupervised learning within the field and the transfer of features from the field of natural images minimize the requirement for annotated data in the target task to learn good features, which is one of the advantages of the proposed technique.

M. Yildirim et al. [6] classified skin cancer into two classes. For this, they combined the features they extracted using the Alexnet and Densenet201 models. Finally, they classified the dataset with SVM.

K. E. Purnama et al. [7] present a system for teledermatology that may be used to classify and identify skin illnesses. Using a Deep Learning algorithm, this system classifies skin diseases in dermoscopic images. In this study, MNIST HAM10000 is divided into seven skin disease classes that fall under the category of skin cancer. MobileNet v1 and Inception v3 are two pre-trained CNN models used in the image classification process.

Convolutional Neural Networks are used by K. Pai et al. [8] to predict and classify seven different skin lesion types. For real-time usage of the algorithm, which can forecast the three most likely skin lesion types for a given image, a website is being established. Studies in the literature show that deep learning classifiers are a promising candidate for detecting skin cancer.

The remaining part of the paper is carried out as follows. Section 2 presents the theoretical background, talking about convolutional neural network structure, dataset and hybrid structures. Hybrid structures describe SVM, KNN, softmax, and DT which are used for classifier. Section 3, presents the details of performance metrics and experimental results. Finally we concluded the paper and discussion the future work in section 4.

2. Theoretical Background

2.1. Data Set

Dermoscopic images are a convenient source for automatically diagnosing pigmented skin lesions [9].

In this study, Scin Cancer Mnist Ham10000 dataset is used for the classification task of skin cancer. Dataset contains pigmented lesions from different populations.

Table 1. The number of images contained by classes in the dataset

Total	akiec	bcc	bkl	df	mel	nv	vasc
10015	327	514	1099	115	1113	6705	142

Dataset consists of seven different classes. Classes are as follows:

nv: Melanocytic nevi

mel: Dermatofibroma

bkl: Benign keratosis-like lesions

bcc: Basal cell carcinoma

akiec: Actinic keratoses

vasc: Vascular lesions

df: Dermatofibroma

Not all classes have the same number of data, and the distribution of classes is as in Table 1. Some images of dataset containing dermoscopic images are given in Fig.1.



Figure 1. Some example images from dataset

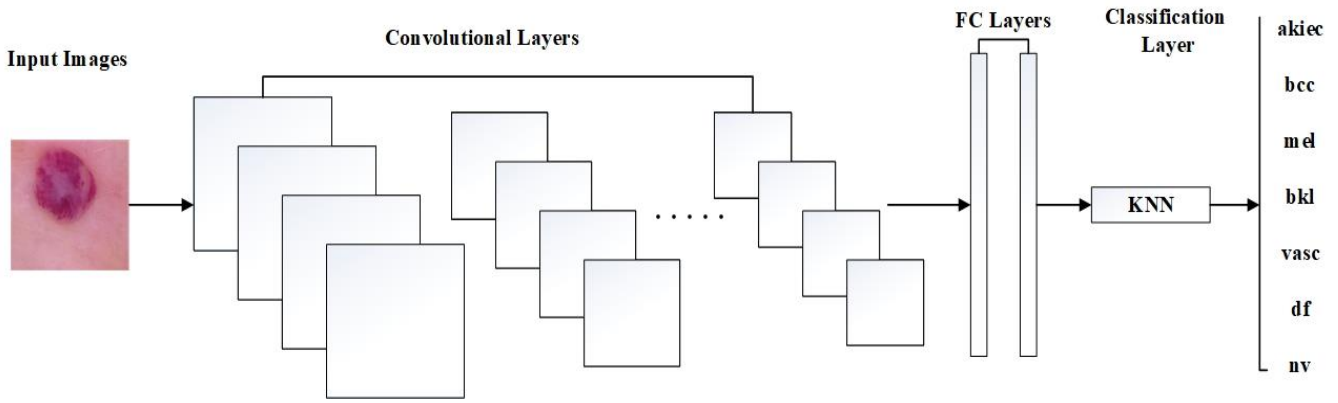


Figure 2. Structure of the hybrid model

2.1.1. Convolutional Neural Network

CNN is one of the most impressive forms of artificial neural network architectures (ANN). ANNs are made up of a large number of interconnected compute nodes that are spread in a distributed way to learn from the input and improve the final output. CNNs are built up of neurons that learn to optimize themselves. In this respect, they are similar to ANNs [10].

CNNs are designed with the idea that the input will be images. This focuses on the architecture that should be built in the most appropriate manner to handle a specific type of data. There are different type of layers in CNNs. Convolutional layers, pooling layers, and fully connected layers are the three types. A CNN architecture is formed by stacking these layers.

2.1.2. Structures of Hybrid Architectures

CNNs contain convolution, pooling, and fully connected layer. For classification process, it has a classification function such as softmax in its last layer. In addition, structures where the classification layer is realized with other machine learning algorithms are also preferred for classification problems. Hybrid structures mostly affect performance. In the study, CNN structure used KNN, SVM and DT algorithms for classification task. Figure 2, gives an example of the architectures used. The figure shows our hybrid structure where the softmax classifier in the last layer of Alexnet is removed and the KNN classifier is added instead.

Softmax: Multinomial Logistic Regression is another name for Softmax. It's a mathematical term that's used a lot, especially in probability theory [11]. With the rapid advancement of computer vision, it is widely employed in deep learning. Softmax is utilized in the last layer of the convolutional neural network architectures as a standard for classification problems [8].

$$\sigma(x_j) = \frac{e^{x_j}}{\sum_i e^{x_i}} \tag{1}$$

The Softmax classifier gives a possibility for each class. For the same dataset, the probability sum for all vectors is 1. Softmax is calculated with the equation (1).

Decision Tree: DT is a traditional classifier that is defined as the sample space's recursive component. The nodes in the decision tree form a rooted tree, which is a directed tree with no incoming edges and a "root" node. There is one incoming edge for each of the other nodes. A node having outgoing edges is known as an internal or test node. The leaves refer to the other nodes.

Each inner node in a DT divides the sample region into two or more subdomains, each depending on a different function of

the input attribute values. Each test takes into account a single feature in the simplest and most typical instance, and the sample area is partitioned by the feature's value. In numerical characteristics, the condition refers to a range [12]. The tree structure is formed by branching, starting from the feature that has the highest impact at the decision stage. This tree structure created using the training set is then tested with test data. While creating the model, entropy calculation is commonly used to determine the order of importance of the features. Figure 3 gives an illustration of the decision tree structure.

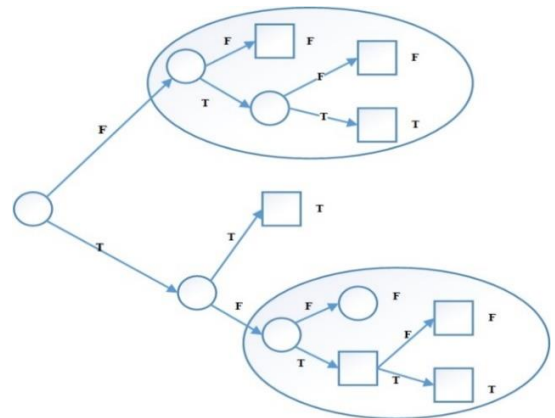


Figure 3. Illustration of DT

K Nearest Neighbour: The KNN algorithm assumes that similar things are close together. When deciding which class to forecast for a new observation, the algorithm determines which points in the training set are sufficient by selecting the data points closest to the new observation and taking the most common. This is why it is called the k Nearest Neighbour algorithm [13].

The procedure is applied on data from a sample set with specific classes. According to the existing data, the distance of the new data that will be included in the sample data set is computed, and k nearby neighborhoods are examined. Euclidean, manhattan and minkowski distance functions are generally used for distance calculations.

Support Vector Machine: SVMs were originally designed for binary classifications. They are also used in the solution of multi-classification problems. There are many methods for extending SVMs to multi-class problems. An "i" class SVM output is as in formula (2). The predicted class is as in equality (3).

$$a_i(x) = W^T x \tag{2}$$

$$arg_i max a_i(x) \tag{3}$$

The estimation functions of Softmax and SVM are the same. Softmax and multi-class SVMs are identical except for their aim, which is defined by all weight matrices W. The Softmax layer decreases cross entropy or maximizes the chance of log, while SVMs strive to discover the biggest margin between various data points. [14].

3. Application and Results

In this paper, we exploit deep learning's structure to create new learning-based inference and decision techniques that meet desired properties like robustness and interpretability. In the study, it is aimed to classify the images of skin cancer in the most accurate way. The network is trained with a seven-class database to verify. The system is carried out in matlab environment.

3.1. Performance Metrics

There are many performance metrics utilized to check the success of the classification. Sensitivity, specificity, F-score accuracy, and Matthews correlation coefficient are the metrics used to measure classification performance [15].

True Positive (TP): The predicted positive data is estimated correctly. The main values are the diagonal.

True Negative (TN): Data known as negative was estimated as negative. It is the sum of all values in the confusion matrix, except for the row and column of the corresponding class.

False Positive (FP): Data known as negative was estimated as positive. It is the sum of all values in the relevant column except TP for each class.

False Negative (FN): Data known as positive is estimated as negative. It is the sum of all values in the relevant row except TP for each class.

Table 2. Confusion matrix for multi-class classification

Classes	C1	C2	C3	C4	C5
C1	x1	x6	x11	x16	x21
C2	x2	x7	x12	x17	x22
C3	x3	x8	x13	x18	x23
C4	x4	x9	x14	x19	x24
C5	x5	x10	x15	x20	x25

In the confusion matrix of the five-class classification in Table 2, for the C1 class;

$$TP=x1$$

$$TN= \text{Sum of all values except values in row and column C1}$$

$$FP=x2+x3+x4+x5$$

$$FN=x6+x11+x16+x21$$

Performance metrics can be calculated in multi-classification operations with TP, TN, FP and FN values.

Accuracy: Accuracy is calculated by dividing the overall number of correct estimations by the total number of correct estimates in the dataset.

Error Rate (ERR): It is obtained by dividing the misclassified values into all values.

Sensitivity (SN): It's derived by dividing the overall positive number by the number of true positive predictions. It's also referred to as recall.

Specificity (SP): It is evaluated by dividing the number of correct negative predictions by the total number of negatives. This is also referred true negative rate.

Precision (PREC): It is calculated by dividing the number of true positive estimates by the total number of positive estimates.

False Positive Rate (FPR): It is obtained by dividing the number of false positive estimates by the overall of negatives.

F-Score: It is the harmonic average of the PREC and SN values.

Matthews Correlation Coefficient: It is a correlation coefficient calculated utilizing all four values in the confusion matrix.

3.2. Experimental Results

In the study, the metrics mentioned above are used to test the performance of our methods. Firstly, Alexnet and Resnet18 architectures, which are convolutional neural network architectures that give high accuracy in image processing problems, were used for the training of our dataset.

Alexnet is an architecture proposed by Khrizevsky et al. in 2012. There are 11 layers in the network. The presence of a large number of layers in the network improves feature extraction. Furthermore, the amount of factors has a negative impact on performance. The convolution layer is AlexNet's initial layer. Following the maximum pooling and normalization layers is the convolution layer. Softmax is the final layer, which does the classification [16].

Figure 4 gives the confusion matrix of the validation accuracy made with Scin Cancer Mnist Ham10000 dataset. 20% of the dataset is used for validation. The accuracy result of the model was derived from dividing the sum of TP and TN values by the sum of TP, TN, FP and FN values. Alexnet provided 77.16% validation accuracy. Calculation as in (4).

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} = \frac{2318}{3004} = 77.16 \quad (4)$$

akiec	16	29	36	1	4	12	
bcc	3	100	26	1	1	20	3
bkl	3	18	216		5	87	1
df	1	9	8	4	1	11	
mel	1	5	58		34	233	3
nv	1	19	68	3	1	1914	5
vasc	1	1	1			6	34
	akiec	bcc	bkl	df	mel	nv	vasc

Predicted Class

Figure 4. Confusion matrix of Alexnet

Loss and accuracy curves obtained as a result of Alexnet training are given in the Figure 5.

given in Table 3. The highest performance was achieved with the architectures using SVM.

Resnet18 architecture was presented in 2015 by kaiming et al. Increasing the network depth does not work by bringing layers together. Deep architectures are challenging to train due to the worsening destructive gradient. Because the gradient spreads back to the before layers, repeated multiplication may make the gradient infinitely small. Architecture focuses on the need to overcome the gradient problem by increasing the depth of the Networks [17]. Figure 6, shows the accuracy and loss curves of Resnet18. The structure provided 74.44% validation accuracy.

In this study, hybrid architectures were obtained by using KNN, SVM, and DT classification algorithms and Resnet18, Alexnet architectures. Validation accuracy rates of these structures are as

Table 3. Accuracy of models

Model Name	Accuracy
Alexnet	0.7716
Alexnet+tree (H1)	0.6302
Alexnet+knn (H2)	0.7563
Alexnet+svm (H3)	0.7780
Resnet18 (H4)	0.7444
Resnet18+tree (H5)	0.6065
Resnet18+knn (H6)	0.7417
Resnet18+svm (H7)	0.7623

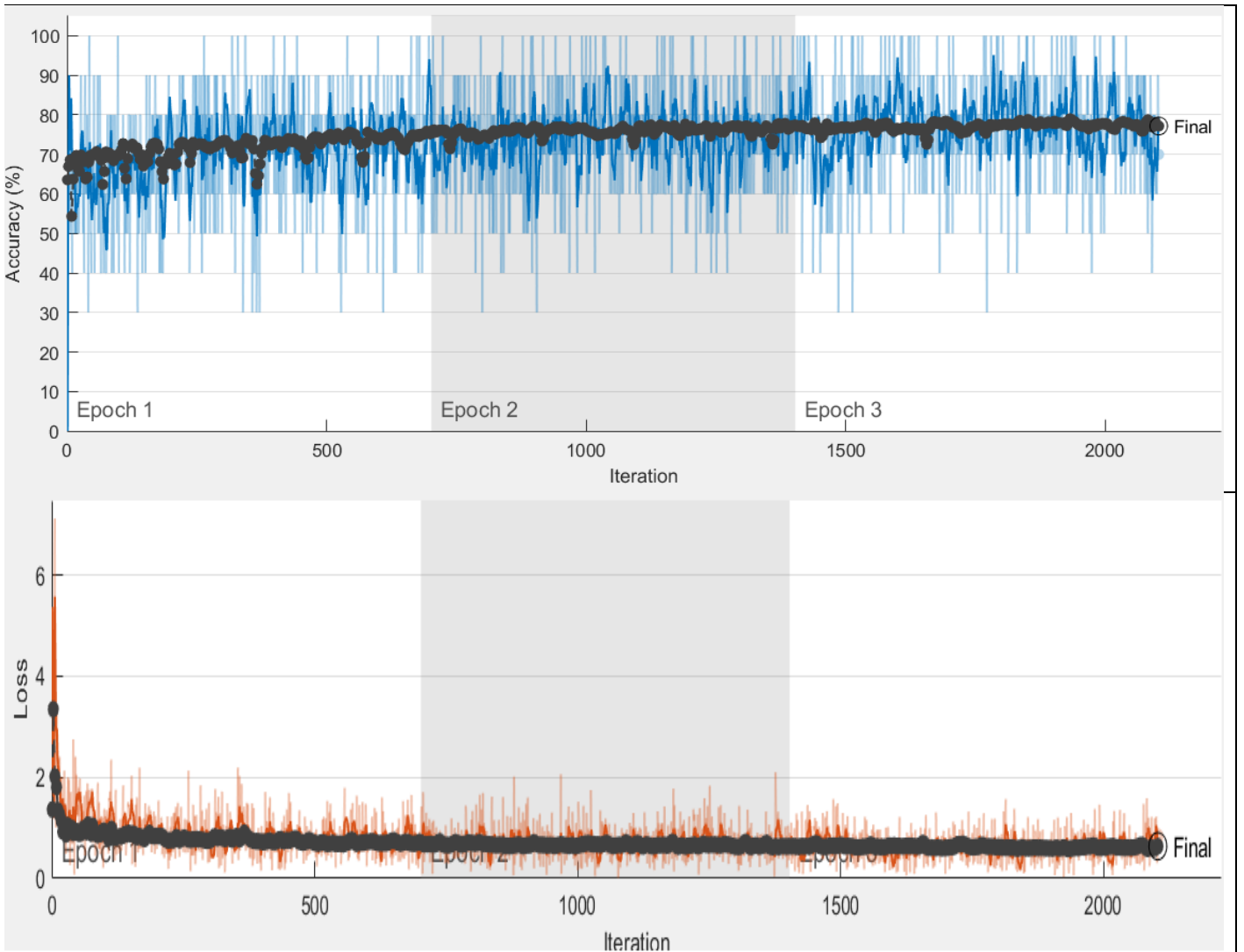


Figure 5 Accuracy and Loss curves of the Alexnet

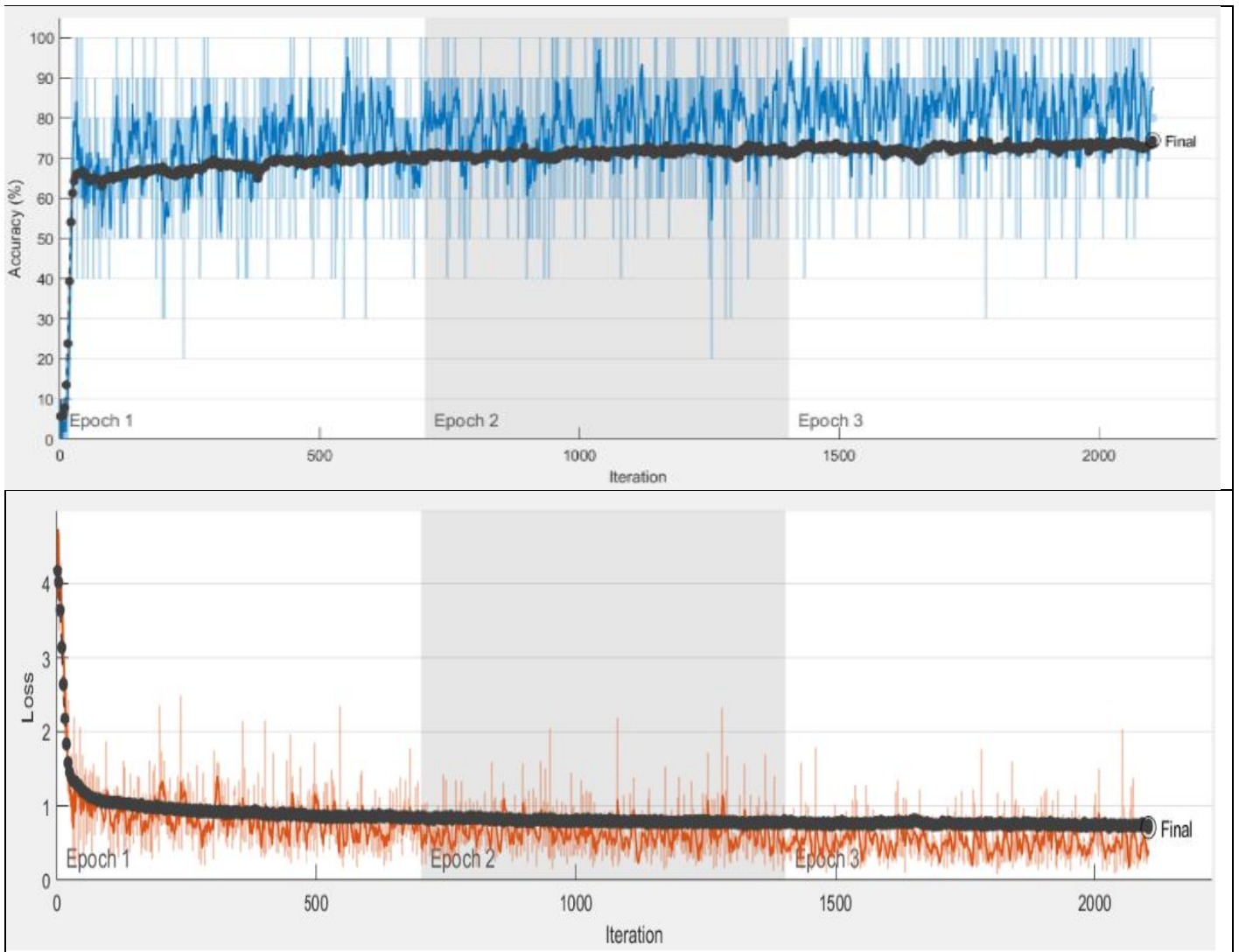


Figure 6. Accuracy and Loss curves of the Resnet18

In the models where we use pretrained architectures Alexnet and Resnet18, as classifier, respectively; we use DT, KNN and SVM. In this way, 6 different architectures are obtained. We perform the training and validation of the models in the same datasets and environment. We use 80% of the Scin Mnist Ham10000 dataset for training and the remaining 20% for validation. Models were compiled with graphics processing unit support. For model training, the learning rate is determined as $1e-4$ and momentum as 0.9.

The confusion matrices in the first and second rows given in Fig. 7 respectively show the results of the architectures created with Alexnet, Resnet18.

Performance measures are calculated for vasc, one of the 7 classes used in the classification task, from the confusion matrix of the given Alexnet + Svm (H3) architecture. The values of the metrics we have chosen as error rate, sensitivity, specificity, precision, FPR, MCC and F-score are presented in Table 4.

Table 4. Performance measures for vasc class in H3

Performance Metric	Value
Error Rate	0.005
Sensitivity	0.70
Specificity	0.99
Precision	0.85
FPR	0.001
MCC	0.77
F-Score($\beta = 0.5$)	0.81

There are other studies in the literature that use the public dataset we use. Comparisons of methods made using Scin Cancer mnist ham10000 dataset are as given in Table 5.

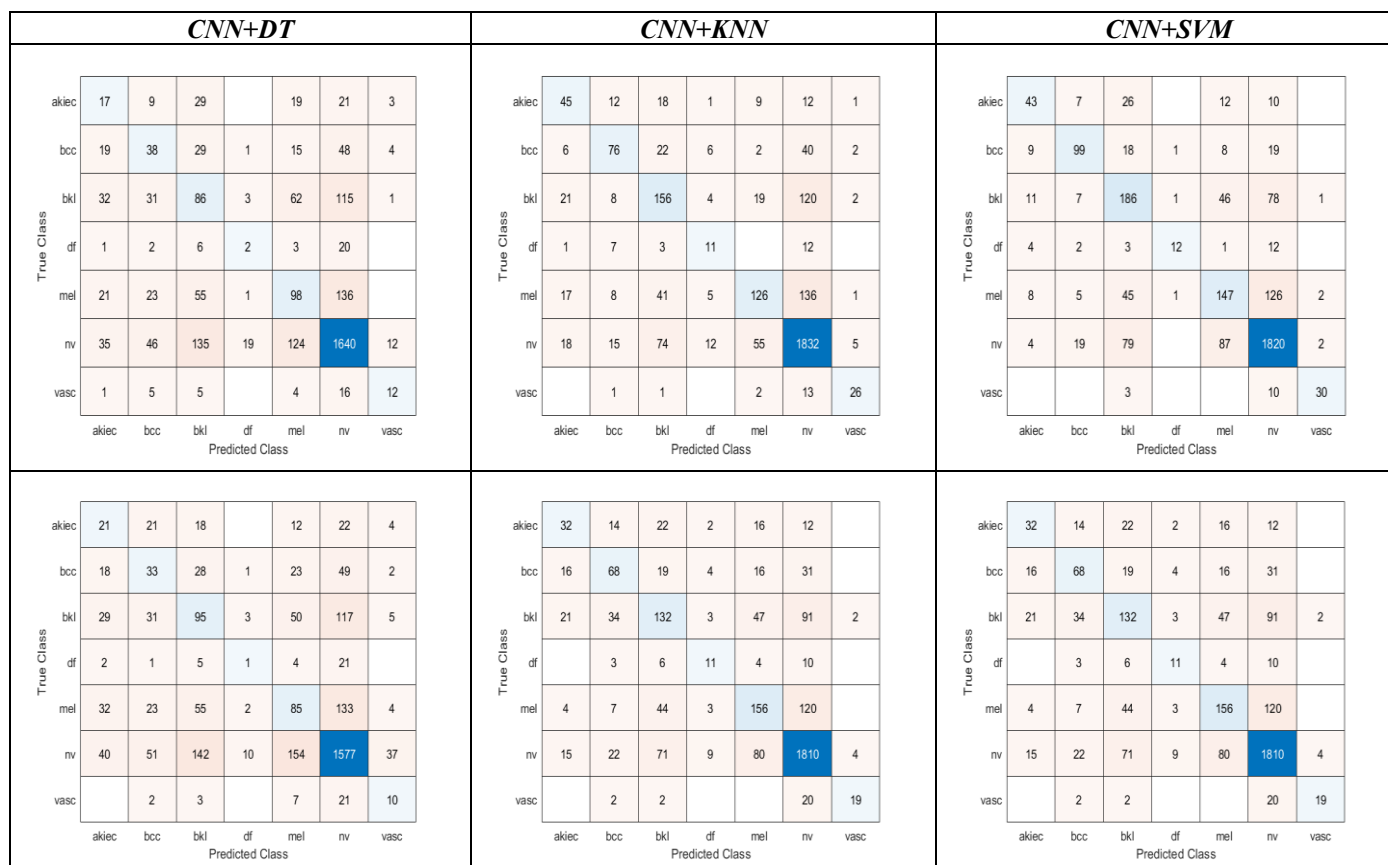


Figure 7. Confusion matrix of hybrid

Table 5. Studies on skin mnist ham10000

Authors	Methods	Class Number	Accuracy
Purnama et al.[7]	Inception V3	7	72 %
Purnama et al.[7]	MobileNet V1	7	58 %
Murugan et al.[18]	Resnet	4	61%
Murugan et al.[18]	VGG19	4	59%
Murugan et al.[18]	InceptionV3	4	64%
Proposed Method	H3	7	77 %
Proposed Method	H7	7	76 %

4. Conclusions

In the study, it is aimed to classify skin cancer using cnn networks. Classification is made with Alexnet and Resnet architectures, which are frequently used cnn architectures. Hybrid structures are created to increase the performance rate. Basically, by taking Alexnet and Resnet architectures, decision tree, knn and svm structures are used for classification task instead of softmax classifier in the last layer. In this way, six different hybrid architectures were created.

The two hybrid architectures created with Svm also give the highest results and the accuracy of standard Alexnet and Resnet18 increases in the rate of performance. On the other hand, the structures created with the decision tree have a decreasing effect on my performance. It is seen that the solution of the problem is improved with the proposed architectures by making comparisons with the previous studies in order to classify the data in the Skin Cancer Mnist Ham10000 dataset into seven classes.

Studies to classify medical diseases aim to assist doctors and patients. Skin cancer is a serious disease that can result in death. It is possible for doctors to make better decisions through computer-based systems. A method with high accuracy performance provides confidence in the diagnosis of the disease. At this point, the proposed method will be useful by improving existing methods. Studies are thought to be done to advance the method in the future.

References

- [1] Öztürk, Ş., & Özkaya, U. (2020). Skin lesion segmentation with improved convolutional neural network. *Journal of digital imaging*, 33(4), 958-970.
- [2] Yıldız, O. (2019). Melanoma detection from dermoscopy images with deep learning methods: A comprehensive study. *Journal of the Faculty of Engineering and Architecture of Gazi University*, 34(4), 2241-2260.
- [3] Chang, H. (2017). Skin cancer reorganization and classification with deep neural network. *arXiv preprint arXiv:1703.00534*. Bejan, A. (2015). Constructal thermodynamics. *Constructal Law & Second Law Conference, Parma*, pp. S1-S8.

- [4] Ünlü, E. I., Çınar, A. (2018). Classification of skin images with respect to melanoma and nonmelanoma using the deep neural network.
- [5] Codella, N., Cai, J., Abedini, M., Garnavi, R., Halpern, A., & Smith, J. R. (2015). Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images. In *International workshop on machine learning in medical imaging* (pp. 118-126). Springer, Cham.
- [6] YILDIRIM, M., & ÇINAR, A. (2021). Classification of Skin Cancer Images with Convolutional Neural Network Architectures. *Turkish Journal of Science and Technology*, 16(2), 187-195.
- [7] Purnama, I. K. E., Hernanda, A. K., Ratna, A. A. P., Nurtanio, I., Hidayati, A. N., Purnomo, M. H., ... & Rachmadi, R. F. (2019). Disease Classification based on Dermoscopic Skin Images Using Convolutional Neural Network in Teledermatology System. In *2019 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)* (pp. 1-5). IEEE.
- [8] Pai, K., & Giridharan, A. (2019, October). Convolutional Neural Networks for classifying skin lesions. In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)* (pp. 1794-1796). IEEE.
- [9] Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, 5(1), 1-9.
- [10] O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.
- [11] Wolfe, J., Jin, X., Bahr, T., & Holzer, N. (2017). Application of softmax regression and its validation for spectral-based land cover mapping. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 455.
- [12] Maimon, O., & Rokach, L. (Eds.). (2005). *Data mining and knowledge discovery handbook*.
- [13] Papernot, N., & McDaniel, P. (2018). Deep k-nearest neighbors: Towards confident, interpretable and robust deep learning. *arXiv preprint arXiv:1803.04765*.
- [14] Tang, Y. (2013). Deep learning using linear support vector machines. *arXiv preprint arXiv:1306.0239*.
- [15] Cengil, E., & Cinar, A. (2018, September). A deep learning based approach to lung cancer identification. In *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)* (pp. 1-5). IEEE.
- [16] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 1097-1105.
- [17] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [18] Murugan, S., & Verwillow, A. DeepDerm: Detection of Cancerous Skin Lesions Through Deep Learning.