Turkish Journal of Science & Technology
16(2), 187-195, 2021

Research Paper

Classification of Skin Cancer Images with Convolutional Neural Network Architectures

Muhammed YILDIRIM[1*], Ahmet ÇINAR
[1] Computer Engineering Department, Engineering Faculty, Firat University, Elazığ, Turkey
[2] Computer Engineering Department, Engineering Faculty, Firat University, Elazığ, Turkey
[*1] 171129205@firat.edu.tr, [2] acinar@firat.edu.tr

(Geliş/Received: 28/01/2021; Kabul/Accepted: 02/05/2021)

Abstract: The skin, in which our body is completely covered, both provides the heat balance of our body and protects our body against external factors. Skin cancer, which occur as a result of the uncontrolled proliferation of cells on the skin surface, are one of the most common types of cancer in the world. Early detection of skin cancer means early treatment of the disease. With early diagnosis, patients can be cured earlier and mortality rates can be reduced. The hardest part of skin cancer diagnosis is that skin lesions are very similar to each other. Therefore, it is of great importance that skin cancer can be diagnosed and classified as benign or malignant tumor. In this study, Convolutional Neural Network were used to determine whether skin cancer is benign or malignant. Separate results are obtained with Alexnet, Resnet50, Densenet201 and Googlenet. Then the performance rates of the models used have been compared. The highest accuracy rate was achieved with the Resnet50 model with 83.49%. Then, the results were obtained with the proposed hybrid model. In the proposed hybrid model, the accuracy rate was 84.11%. This rate is an important value for early diagnosis and treatment of the disease.

Key words: CNN, Classification, Deep Learning, Image Processing, Skin Cancer.

Evrişimsel Sinir Ağları Mimarileri ile Cilt Kanseri Görüntülerinin Sınıflandırılması

Öz: Vücudumuzun tamamen kaplandığı deri hem vücudumuzun ısı dengesini sağlar hem de vücudumuzu dış etkenlere karşı korur. Cilt yüzeyindeki hücrelerin kontrolsüz çoğalması sonucu ortaya çıkan cilt kanserleri, dünyada en sık görülen kanser türlerinden biridir. Deri kanserlerinin erken teşhisi, hastalığın erken tedavisi anlamına gelir. Erken tanı ile hastalar daha erken tedavi edilebilir ve ölüm oranları azaltılabilir. Cilt kanseri teşhisinin en zor kısmı, cilt lezyonlarının birbirine çok benzemesidir. Bu nedenle cilt kanserinin iyi huylu veya kötü huylu tümör olarak teşhis edilip sınıflandırılabilmesi büyük önem taşımaktadır. Bu çalışmada, Evrişimsel Sinir Ağı ağları, cilt kanserinin iyi huylu veya kötü huylu olup olmadığını belirlemek için kullanılmıştır. Alexnet, Resnet50, Densenet201 ve Googlenet ile ayrı ayrı sonuçlar elde edilmiştir. Daha sonra kullanılan modellerin performans oranları karşılaştırılmıştır. En yüksek doğruluk oranı % 83,49 ile Resnet50 modelinde elde edilmiştir. Daha sonra önerilen hibrit modelle sonuçlar elde edilmiştir. Önerilen hibrit modelde doğruluk oranı %84.11 olmuştur. Bu oran, hastalığın erken teşhis ve tedavisi için önemli bir değerdir.

Anahtar kelimeler: CNN, Sınıflandırma, Derin Öğrenme, Görüntü işleme, Cilt Kanseri.

1. Introduction

The skin, which covers almost all of our body, can deteriorate over time and lead to various types of cancer [1]. Skin cancer have become one of the most common types of cancer in the world [2]. Early diagnosis of these types of cancer is of great importance for the treatment process [3]. When tumor is called, generally 2 different types of tumors are considered. The first of these are benign tumors and the second is malignant tumors [4]. Benign tumors put the person in more visual and functional problems. Malignant tumors can put a person's life at risk [5]. There are many deaths in the world due to malignant tumor.

The reason why the detection of tumors is difficult is that different skin lesions are very similar to each other [6]. Because of this similarity, detection of these tumors becomes very difficult. Dermatoscopy is the most common method used by physicians to detect these cancer tumors [7]. In this method, an examination is made with a device with a very high visual resolution and the relevant region is examined by enlarging. In this process, the professional experience of the doctor and the images obtained from dermatoscopy are of great importance [8].

In this study, CNN models, which are very popular in recent years, are used. After the Alexnet architecture won the large-scale visual recognition (ILSVRC) competition in 2012, CNN models have become increasingly important [9]. With the developing technology, deep learning architectures have started to achieve successful results in large data sets [10]. Since skin cancer is a common and deadly disease, computer-aided systems are of

[*] Corresponding author: 171129205@firat.edu.tr. ORCID Number of authors: [1]0000-0003-1866-4721, [2] 0000-0001-5528-2226

great importance in combating this disease. CNN networks need to be trained with large amounts of data so that this disease can be diagnosed early and the treatment process can be initiated [11]. For the classification of skin cancer data, segmented images of the skin cancer data are used and the models are trained and tested.

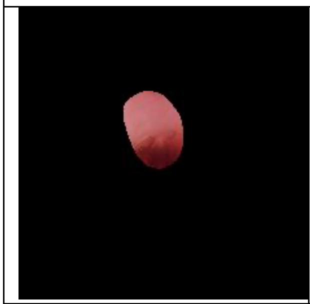
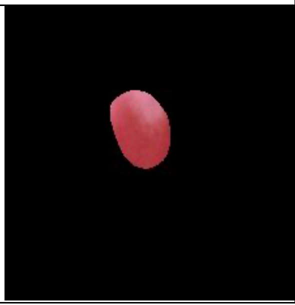
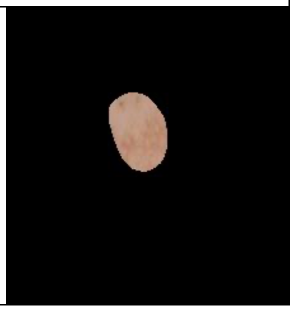
In this paper, Densenet201 architecture [12], Alexnet architecture [13], Googlenet architecture [14] and Resnet50 architecture [15] are used to classify skin cancer data.

The data set used in the rest of the article is examined. Then the architectures used are discussed. After that, the application and the results obtained are shared. Finally, the conclusion section is prepared.

2. Description of the Skin Cancer Dataset

For the classification of skin cancer data, segmented images of the skin cancer dataset [16] on Kaggle's page were used. This data set consists of 2 classes. The first class is the benign class; the second class is the malignant class. While benign class refers to benign tumor, malignant refers to malignant tumor. Although benign type tumors cause more visual and functional problems, malignant type tumors can lead to death. Examples from the data set used in the study and the number of images used are given in table 1.

Table 1. Examples belonging to the benign and malignant class.

Benign (1440 images in total)		Malignant(1197 images in total)	
			

3. Models

Cnn architectures Alexnet, Resnet50, Densenet201 and Googlenet are used to classify data belonging to benign and malignant classes. When designing Cnn architectures, we can classify the parameters to be considered as kernel size, batch-size, number of cycles, number of layers, activation function, learning rate, dropout and the size of the data set [17]. The characteristics of the computer used also affect the training and testing time of the model. The architectures and training values used in this study are given in table 2.

Table 2. Training values of models.

Models	MiniBatchSize	MaxEpochs	ValidationFrequency	InitialLearnRate
Alexnet Resnet50 Densenet201 Googlenet	16	5	6	10^{-4}

In deep learning applications, processing all data in the data set at the same time is a costly process in terms of time and memory. Because in every iteration of learning, back propagation is done. This is an important process in terms of time and cost. To solve this problem, it is necessary to divide the data sets into small groups. In this way, giving more than one input piece by piece is called minibatch. In this study, the reason for choosing the minibatch size value 16 is related to the computer used. Since the learning of the network becomes stable after 5 epoch values, the epoch value of 5 was chosen for the architectures used. In order to increase the readability of the accuracy and loss curves, the ValidationFrequency value of 6 was chosen. InitialLearnRate value is $10^{-4.}$

3.1. Alexnet

Published by LE Cun in 1988, LeNet architecture is one of the first articles published for deep learning. But when Alexnet won the large-scale visual recognition (ILSVRC) competition in 2012, deep learning architectures began to be heard around the world. The victory of Alexnet architecture in this competition led to the start of a popular era in deep learning. With this architecture, the error rate in pattern recognition has been reduced from 26% to 15%. Relu is used as the activation function in the model. This model is designed to classify 1000 objects. The filters used here are 11 x 11 in size. Alexnet architecture consists of 25 layers in total [18].

3.2. Resnet50

The Resnet50 model was the winner of the ImageNet competition in 2015. The error rate of the Resnet50 architecture in this competition was 3.6%. While the ResNet50 architecture classifies images with an error rate of 3.6%, people can classify the image with an average error rate of 5-10%. This ratio is a great success for the Resnet50 architecture. The Resnet50 architecture is one of the first architectures to use BatchNormalization for the normalization process. Relu is used as the activation function in this architecture. It has a deeper structure than previously designed architectures [19].

3.3. Densenet201

DenseNet201 provides the merged feature maps produced by all previous layers as input to the next layer. In this way, it is aimed to make the training process easier. It is aimed to use the parameters more efficiently in the Densenet201 architecture. This allows access to all feature maps produced in layers prior to the deeper layers in the network. Avaragepooling has been preferred for the pooling process in the Densenet201 architecture. Relu was used as the activation function and Batchnormalization was used for the normalization process [20].

3.4. Googlenet

This model is the winner of the ILSVRC competition held in 2014.The model with an error rate of 5.7% gave successful results in data sets. The Googlenet model, which has a depth of 22 layers, has a 144-layer structure. This model is one of the convolution and CNN architectures that move away from the common layers in an ordered structure. The start-up modules used here have reduced memory and power usage. [21].

3.5. Proposed Hybrid Model

In the proposed hybrid model, features obtained from the Resnet50 and Densenet201 architectures are concatenated. These features are taken from the layer before the softmax layer. Later, these concatenated features were classified in the SVM classifier. A rough representation of the proposed hybrid model is given in figure 1.

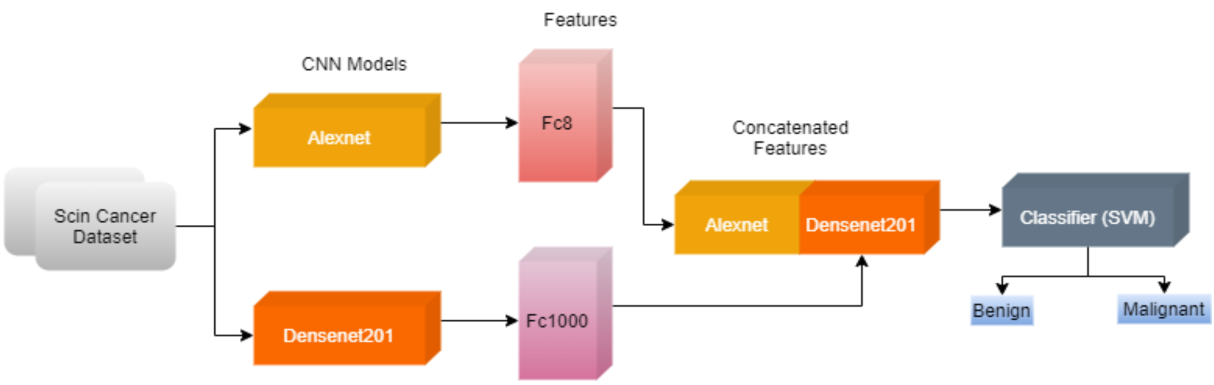


Figure 1. Proposed Hybrid Model.

4. Experimental Results

In this study, CNN architectures are used to classify the skin cancer data set as benign and malignant. CNN architectures used in these paper are the most widely used in recent years. Cnn architectures are widely used, especially in biomedical image processing [23].

Some values must be calculated in order to perform performance evaluation of data classified with CNN architectures [24]. An example of a confusion matrix, which presents a sort of summary of the success of the model used, is given in Table 3. Performance criteria of CNN models are calculated using a Confusion matrix.

Table 3. Confusion matrix obtained in models.

	Benign	**Malignant**
Benign	TruePositive	FalsePositive
Malignant	FalseNegative	TrueNegative

TP: The benign image is correctly predicted and placed in the benign class.

FN: The image belonging to the class of malignant is guessed incorrectly and placed in the benign class.

FP: Image belonging to the benign class are predicted as malignant.

TN: The malignant image is correctly predicted and placed in the malignant class.

F-Measure value in equation 1, Precision value in equation 2, Specifity value in equation 3, **Sensitivity** value in equation 4, False Positive Rate (FPR) value in equation 5, False Discovery Rate (FDR) value in equation at 6, the False Negative Rate (FNR) value is given in equation 7 and the Accuracy value is given in equation 8. Confusion matrix was used to calculate all these values.

$$F-measure = \frac{2*Precision*Recall}{Precision+Recall} \ (1), \ Precision = \frac{TP}{TP+FP} \ (2), \ Specificity \ \ (TNR) = \frac{TN}{TN+FP} \ (3)$$

$$Sensitivity \ (TPR) = \frac{TP}{TP+FN} \ (4), \ False \ Positive \ Rate \ (FPR) = \frac{FP}{FP+TN} \ (5),$$

$$False \ Discovery \ Rate \ (FDR) = \frac{FP}{FP+TP} \ (6) \ , False \ Negative \ Rate \ (FDR) = \frac{FN}{FN+TP} \ (7)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \ (8)$$

Alexnet architecture was first used to classify benign and malignant images of the skin cancer data set. The accuracy and loss curves of the benign and malignant classes obtained using the Alexnet architecture are given in figure 2.

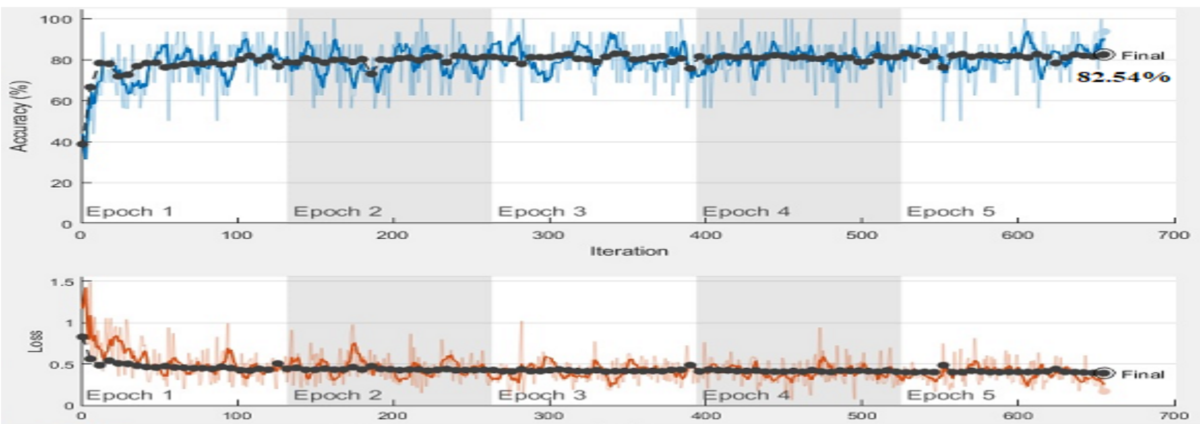


Figure 2. Alexnet's accuracy and loss curves.

The confusion matrix and success rates obtained as a result of the classification process performed by the Alexnet model in the skin cancer dataset are given in table 4. When table 4 is examined, it is seen that 220 out of 288 benign data were placed correctly and 68 were incorrectly placed. Likewise, 215 of 239 malignant data were placed in the correct class, while 24 were placed in the wrong class. The accuracy rate of the model is 82.54%.

Table 4. Alexnet's confusion matrix and success rates achieved.

	Benign	Malignant
Benign	**220**	68
Malignant	24	**215**

F-Measure	Precision	Specifity	Sensitivity	FPR	FDR	FNR	Accuracy
82.71%	76.39%	75.97%	90.16%	24.03%	23.61%	9.84%	82.54%

Another model used to classify benign and malignant images belonging to the skin cancer data set is Densenet201. The accuracy and loss curves of the benign and malignant classes obtained using the Densenet201 architecture are given in figure 3.

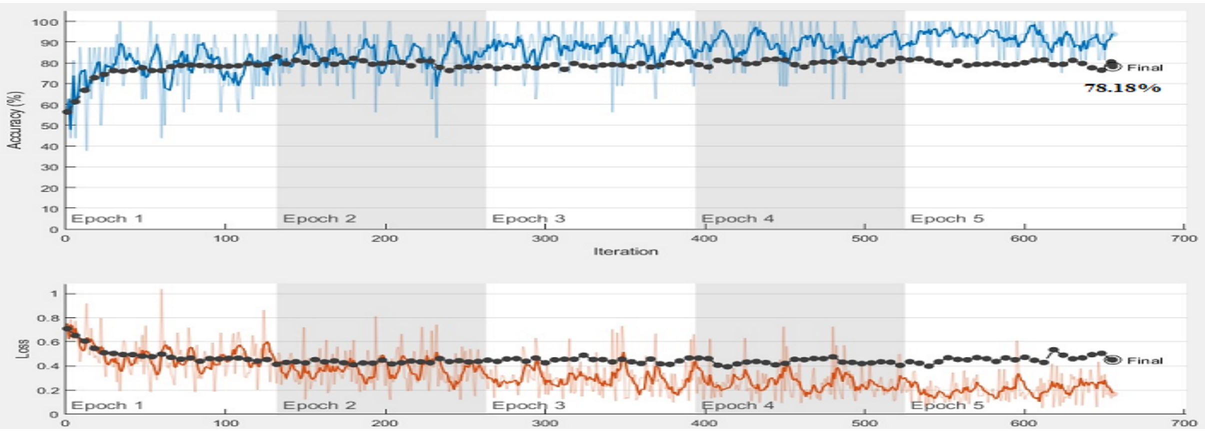


Figure 3. Densenet201's accuracy and loss curves.

The confusion matrix and success rates obtained as a result of the classification process performed by the Densenet201 model in the skin cancer data set are given in table 5. When table 5 is examined, it is seen that 227 of 288 benign data were placed correctly and 61 were placed incorrectly. Likewise, 185 of 239 malignant data were placed in the correct class, while 54 were placed in the wrong class. The accuracy rate of the model is 78.18%

Table 5. Densenet201's confusion matrix and success rates achieved.

	Benign	Malignant
Benign	**227**	61
Malignant	54	**185**

F-Measure	Precision	Specifity	Sensitivity	FPR	FDR	FNR	Accuracy
79.79%	78.82	75.20%	80.78%	24.80%	21.18%	19.22%	78.18%

Another model used to classify benign and malignant images belonging to the skin cancer data set is Googlenet. The accuracy and loss curves of the benign and malignant classes obtained using the Googlenet architecture are given in figure 4.

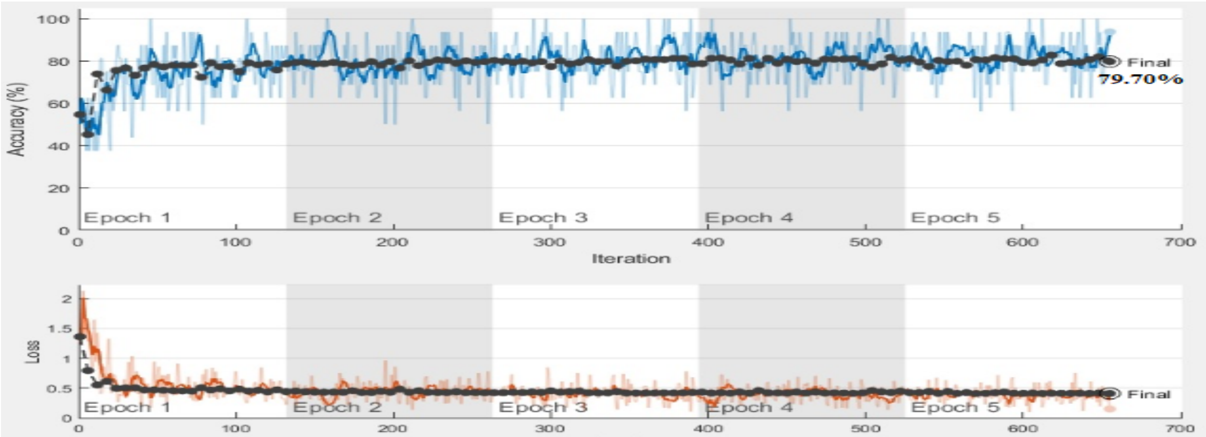


Figure 4. Googlenet’s accuracy and loss curves.

The confusion matrix and success rates obtained as a result of the classification process performed by the Googlenet model in the skin cancer data set are given in table 6. When table 6 is examined, it is seen that 204 out of 288 benign data were placed correctly and 84 were placed incorrectly. Likewise, 216 out of 239 malignant data were placed in the correct class, while 23 were placed in the wrong class. The accuracy rate of the model is 79.70%.

Table 6. Googlenet's confusion matrix and success rates achieved.

	Benign	**Malignant**
Benign	**204**	84
Malignant	23	**216**

F-Measure	**Precision**	**Specifity**	**Sensitivity**	**FPR**	**FDR**	**FNR**	**Accuracy**
79.22%	70.83%	72%	89.87%	28%	29.17%	10.13%	79.70%

Resnet50 is the model that classifies benign and malignant images of the skin cancer data set with the highest accuracy. Accuracy and loss curves of benign and malignant classes obtained using the Resnet50 architecture are given in figure 5.

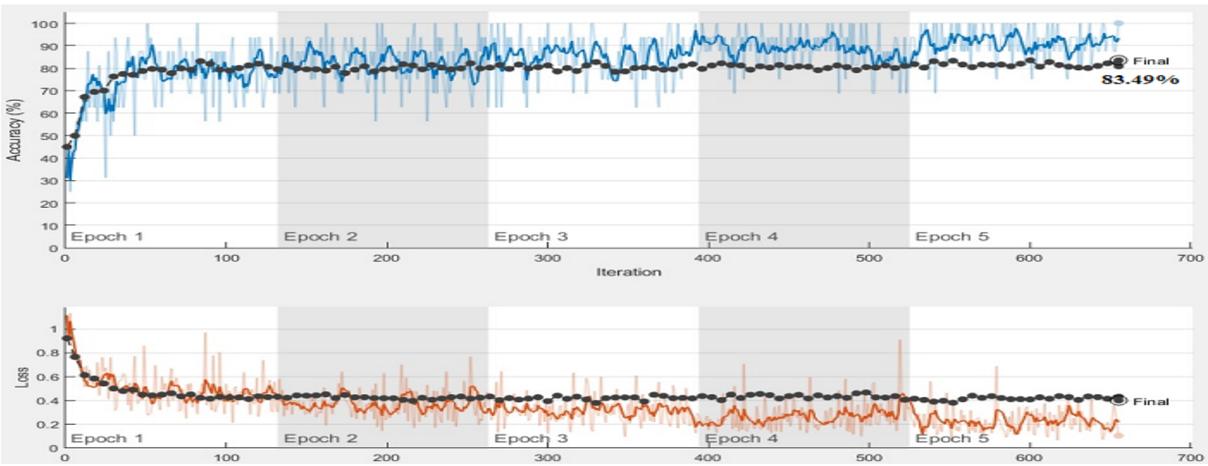


Figure 5. Resnet50’s accuracy and loss curves.

The confusion matrix and success rates obtained as a result of the classification process performed by the Resnet50 model in the skin cancer data set are given in table 7. When table 7 is examined, it is seen that 253 of

288 benign data were placed in the correct class and 35 were placed in the wrong class. Likewise, 187 of 239 malignant data were placed in the correct class, while 52 were placed in the wrong class. The accuracy rate of the model is 83.49%.

Table 7. Resnet50's confusion matrix and success rates achieved.

	Benign	Malignant
Benign	**253**	35
Malignant	52	**187**

F-Measure	Precision	Specifity	Sensitivity	FPR	FDR	FNR	Accuracy
85.33%	87.85%	84.23%	82.95%	15.77%	12.15%	17.05%	83.49%

Among the models used in this study, the highest accuracy rate was obtained with the Resnet50 architecture, while the lowest accuracy rate was obtained in the Densenet201 architecture. The accuracy values of the models used are given in table 8.

Table 8. Accuracy values of pre-trained CNN models.

Resnet50	Alexnet	Googlenet	Densenet201
83.49%	82.54%	79.70%	78.18%

When the features obtained from the four pre-trained CNN architectures used in the study were classified in the SVM classifier, the accuracy values in table 9 were obtained.

Table 9. CNN models + SVM classifier.

Resnet50 + SVM	Alexnet + SVM	Googlenet + SVM	Densenet201 + SVM
82.6%	82.8%	81.5%	82.7%

Finally, the confusion matrix obtained in the proposed hybrid model and the performance metrics of the proposed hybrid model are given in table 10.

Table 10. Proposed Hybrid model's confusion matrix and performance metrics.

	Benign	Malignant
Benign	**1194**	246
Malignant	173	**1024**

F-Measure	Precision	Specifity	Sensitivity	FPR	FDR	FNR	Accuracy
85.07%	82.92%	80.63%	87.34%	19.37%	17.18%	12.66%	**84.11%**

While the proposed hybrid model correctly classified 1194 of 1440 Benign images, it misclassified 246 of them. While the proposed hybrid model correctly classified 1024 out of 1197 Malignant images, 173 were misclassified. The accuracy rate of the model was 84.11%.

5. Conclusion

Skin cancer is one of the most common types of cancer that occurs as a result of uncontrolled proliferation of cells on the skin surface. It is difficult to diagnose because different skin lesions are very similar. Early detection of cancer cells is very important for the treatment process. For this reason, computer aided systems were used to diagnose skin cancer in this study. Resnet50 architecture achieved the highest performance among pre-trained networks with an accuracy of 83.49%. In the proposed hybrid model, the accuracy value obtained is 84.11%. With the proposed hybrid model, it is aimed to alleviate the workload of the experts and shorten the diagnosis time. Thanks to computer-aided systems, individual mistakes that can be made by experts are prevented. In subsequent studies, it is aimed to increase the success rate of the model by developing different models on the subject.

References

[1] Kabashima, K., Honda, T., Ginhoux, F., & Egawa, G. (2019). The immunological anatomy of the skin. Nature Reviews Immunology, 19(1), 19-30. https://doi.org/10.1038/s41577-018-0084-5

[2] American Cancer Society: Cancer facts and figures, Aug 2018, [online] Available: https://www.cancer.org/content/dam/cancer-org/research/cancer-factsand-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-andfigures-2018.pdf.

[3] N. C. F. Codella et al., "Deep learning ensembles for melanoma recognition in dermoscopy images", IBM Journal of Research and Development, vol. 61, pp. 5:1-5:15, 2017.

[4] Taroni, P., Paganoni, A. M., Ieva, F., Pifferi, A., Quarto, G., Abbate, F., ... & Cubeddu, R. (2017). Non-invasive optical estimate of tissue composition to differentiate malignant from benign breast lesions: A pilot study. Scientific reports, 7(1), 1-11. https://doi.org/10.1038/srep40683.

[5] Wylie, J. D., Jenkins, P. A., Beckmann, J. T., Peters, C. L., Aoki, S. K., & Maak, T. G. (2018). Computed tomography scans in patients with young adult hip pain carry a lifetime risk of malignancy. Arthroscopy: The Journal of Arthroscopic & Related Surgery, 34(1), 155-163. https://doi.org/10.1016/j.arthro.2017.08.235

[6] Yap, J., Yolland, W., & Tschandl, P. (2018). Multimodal skin lesion classification using deep learning. Experimental dermatology, 27(11), 1261-1267.https://doi.org/10.1111/exd.13777

[7] Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. In Proceedings of COMPSTAT'2010(pp. 177-186). Physica-Verlag HD.

[8] Deng, L.,& Yu, D. (2014). Deep learning: methods and applications. Foundations and Trends® in Signal Processing, 7(3–4), 197- 387.

[9] Espinosa, J. E., Velastin, S. A., & Branch, J. W. (2017, November). Vehicle detection using alex net and faster R-CNN deep learning models: a comparative study. In International Visual Informatics Conference (pp. 3-15). Springer, Cham. https://doi.org/10.1007/978-3-319-70010-6_1

[10] Singh, A. K., Ganapathysubramanian, B., Sarkar, S., & Singh, A. (2018). Deep learning for plant stress phenotyping: trends and future perspectives. Trends in plant science, 23(10), 883-898. https://doi.org/10.1016/j.tplants.2018.07.004

[11] Yildirim, M., Cinar, A. (2020). A deep learning based hybrid approach for COVID-19 disease detections. Traitement du Signal, 37(3): 461-468. https://doi.org/10.18280/ts.370313.

[12] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[13] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[14] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

[15] 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).

[16] URL-1, https://www.kaggle.com/naim99/segmented-images-of-the-skin-cancer-dataset, Last Accessed Date:03.02.2021

[17] Yildirim, M., Çinar, A. (2019). Classification of white blood cells by deep learning methods for diagnosing disease. Revue d'Intelligence Artificielle, Vol. 33, No. 5, pp. 335-340. https://doi.org/10.18280/ria.330502.

[18] Liu, Z. (2020). Soft-shell shrimp recognition based on an improved AlexNet for quality evaluations. Journal of Food Engineering, 266, 109698. https://doi.org/10.1016/j.jfoodeng.2019.109698

[19] Zahisham, Z., Lee, C. P., & Lim, K. M. (2020, September). Food Recognition with ResNet-50. In 2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET) (pp. 1-5). IEEE. doi: 10.1109/IICAIET49801.2020.9257825.

[20] Carcagnì, P., Leo, M., Cuna, A., Mazzeo, P. L., Spagnolo, P., Celeste, G., & Distante, C. (2019, September). Classification of skin lesions by combining multilevel learnings in a DenseNet architecture. In International Conference on Image Analysis and Processing (pp. 335-344). Springer, Cham. https://doi.org/10.1007/978-3-030-30642-7_30

[21] Al-Qizwini, M., Barjasteh, I., Al-Qassab, H., & Radha, H. (2017, June). Deep learning algorithm for autonomous driving using googlenet. In 2017 IEEE Intelligent Vehicles Symposium (IV) (pp. 89-96). IEEE. doi: 10.1109/IVS.2017.7995703.
[22] Çinar, A., Yıldırım, M. (2020). Detection of tumors on brain MRI images using the hybrid convolutional neural network architecture. Medical Hypotheses, 139: 109684. https://doi.org/10.1016/j.mehy.2020.109684
[23] Çınar, A., Yıldırım, M., & Eroğlu, Y. (2021). Classification of Pneumonia Cell Images Using Improved ResNet50 Model. Traitement du Signal, 38(1), 165-173. https://doi.org/10.18280/ts.380117