Volume 13, Issue 3, Page 117-123, 2024 https://doi.org/10.46810/tdfd.1502471 Research Article

BİNGÖL UNIVERSITY

Türk Doğa ve Fen Dergisi
Turkish Journal of Nature and Science
www.dergipark.gov.tr/tdfd

Classification of Skin Diseases with Different Deep Learning Models and Comparison of the Performances of the Models

Ferdi DOĞAN[1*], Miktat AKTAŞ[2], Mehmet İsmail GÜRSOY[3]

[1] Adıyaman University, Faculty of Engineering, Computer Engineering, Adıyaman, Türkiye
[2] Fırat University, Faculty of Technology, Software Engineering, Elazig, Türkiye
[3] Adıyaman University, Technical Sciences Vocational School, Electrical Department, Adıyaman, Türkiye
Ferdi DOĞAN ORCID No: 0000-0002-9203-697X
Miktat AKTAŞ ORCID No: 0000-0002-0731-5668
Mehmet İsmail GÜRSOY SOYAD ORCID No: 0000-0002-2285-5160

**Corresponding author: fdogan@adiyaman.edu.tr*

(Received: 18.06.2024, Accepted: 25.08.2024, Online Publication: 26.09.2024)

Keywords
Deep learning,
Skin diseases classification,
Disease detection,
Dermatology,
Quality of life

Abstract: Classification of skin diseases is a important isssue for early diagnosis and treatment. The process of determining the disease by the specialist physician also delays the treatment process to be applied to the patient. Computer-aided diagnosis systems play an important role in early diagnosis and initiation of treatment by minimizing such processes. In this study, high-performance classification of skin lesions was performed by using Deep Learning models. Dataset was ISIC data set, dataset were expanded by using data augmentation techniques. In the images in this dataset, there are images of Actinic Keratosis, Dermatofibroma, Pigmented Benign Keratosis, Seborrheic Keratosis, Vascular Lesion skin diseases. The data set was classified by Deep Learning models by using the supervised learning method.. SequeezeNet, AlexNet, GoogleNet, Vgg-19, ResNet101, DenseNet201, ResNet-50, ResNet-18, Vgg-16 DL models were used for classification. To evaluate of classification success of Deep Learning models, confusion matrix and F1-score, precision, sensitivity and accuracy metrics obtained from the matrix were used. According to the F1-score, the most successful model is Vgg16 with 97.41%, while the highest accuracy rate obtained by ResNet18 with 98.06%. High success rate shows that such systems can be used for diagnosis and treatment processes.

Farklı Derin Öğrenme Modelleri ile Cilt Hastalıklarının Sınıflandırılması ve Modellerin Performanslarının Karşılaştırılması

Anahtar Kelimeler
Derin öğrenme,
Cilt hastalıkları sınıflandırması,
Hastalık tespiti,
Dermatoloji,
Yaşam kalitesi

Öz: Deri hastalıklarının sınıflandırılması erken tanı ve tedavi açısından önemli bir konudur. Hastalığın uzman hekim tarafından tespit edilmesi süreci aynı zamanda hastaya uygulanacak tedavi sürecini de geciktirmektedir. Bilgisayar destekli teşhis sistemleri bu süreçleri en aza indirerek erken teşhis ve tedaviye başlanmasında önemli rol oynamaktadır. Bu çalışmada Derin Öğrenme modelleri kullanılarak cilt lezyonlarının yüksek performanslı sınıflandırması yapılmıştır. Veri seti ISIC veri seti olup, veri arttırma teknikleri kullanılarak veri seti genişletilmiştir. Bu veri setindeki görsellerde Aktinik Keratoz, Dermatofibroma, Pigmentli Benign Keratoz, Seboreik Keratoz, Vasküler Lezyon cilt hastalıklarına ait görseller bulunmaktadır. Veri seti denetimli öğrenme yöntemi kullanılarak Derin Öğrenme modelleri ile sınıflandırılmıştır. Sınıflandırma için SequeezeNet, AlexNet, GoogleNet, Vgg-19, ResNet101, DenseNet201, ResNet-50, ResNet-18, Vgg-16 DL modelleri kullanılmıştır. Derin Öğrenme modellerinin sınıflandırma başarısını değerlendirmek için karışıklık matrisi ve matristen elde edilen F1 skoru, kesinlik, duyarlılık ve doğruluk metrikleri kullanılmıştır. F1 skoruna göre en başarılı model %97,41 ile Vgg16 olurken, en yüksek doğruluk oranı %98,06 ile ResNet18 tarafından elde edildi. Başarı oranının yüksek olması bu tür sistemlerin teşhis ve tedavi süreçlerinde kullanılabileceğini göstermektedir.

1. INTRODUCTION

Skin diseases are one of the most common diseases in the world[1]. Diagnosis and treatment of skin diseases are quite difficult due to their complexity. In order to diagnose these diseases, comprehensive tests should be performed and examined[2]. The experience of the doctor and obtaining pathological results delay the diagnosis process of the patient considerably. Diagnosing and diagnosing the patient's condition is vital in some cases. Accelerating such processes is very important for medical processes [3]. Many known skin diseases are visually similar to each other. While some skin diseases are benign, some skin diseases can be malignant. In both cases, early diagnosis and treatment are valuable to improve the patient's quality of life. Skin lesions appearing in different parts of the body can have different meanings. The treatment processes of these skin lesions, which are very similar to each other, can be different from each other. When the specialist suspects the lesion, he performs a visual inspection. However, it is not possible to perform visual inspection for all lesions [4-7]. There is no accepted definitive diagnostic method for visual inspection.

Image processing and it's applications have become increasingly common with the increase in the high computational capacity of computer technologies,. It is used in many fields such as production, industry, defense and medicine[8, 9]. In recent years, computer-aided diagnosis systems related to medical diagnosis and treatment have become quite widespread. The systems with the highest rate in the performance of computer vision systems are artificial intelligence supported systems. Especially DL-based systems are used very popularly nowadays [10]. Deep Learning (DL) can be defined as a deep artificial intelligence system used especially on images. Thanks to its high depth artificial neural network layer structure, it gives quite successful results in image classification and detection processes [11]. The layers and classification process used in a DL architecture are shown in Figure 1.

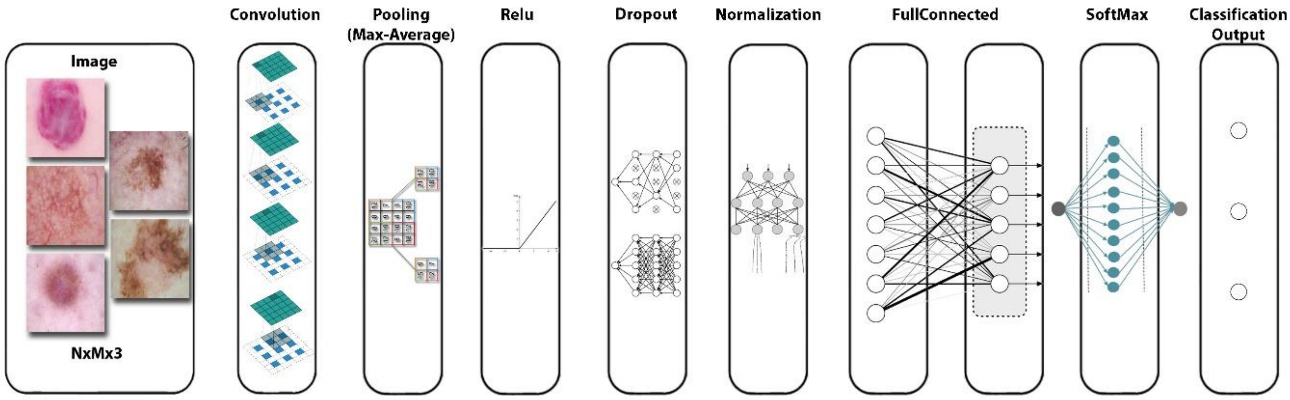


Figure 1. An example DL model and layers used in the model[12].

A dataset consisting of images of skin diseases, which is the input image in Figure 1, is put as input data to a sample DL model. Then, feature extraction is performed via convolutional layer. In the next layers, the image size used as input for the network is reduced by the polling process. Relu is used for activation function [13-15]. Probability of memorization of the network is reduced and some of the random weights in the network are discarded in the dropout layer,. The image data is standardized via normalization layer. Neurons in a fully connected layer are connected to neurons in the previous layer. This layer combines all the features (attributes) learned by previous layers in the image to identify larger patterns. The Softmax layer is the entropy layer and it makes probabilistic estimation and it is estimated which class the input image belongs to via the Classification layer [16,17].

9 DL models used for the classification of skin diseases. These models are SequeezeNet, AlexNet, GoogleNet, Vgg-19, ResNet101, DenseNet201, ResNet-50, ResNet-18, Vgg-16 [18-23]. These were used in the study because they were the most commonly encountered deep learning models in the literature reviews .

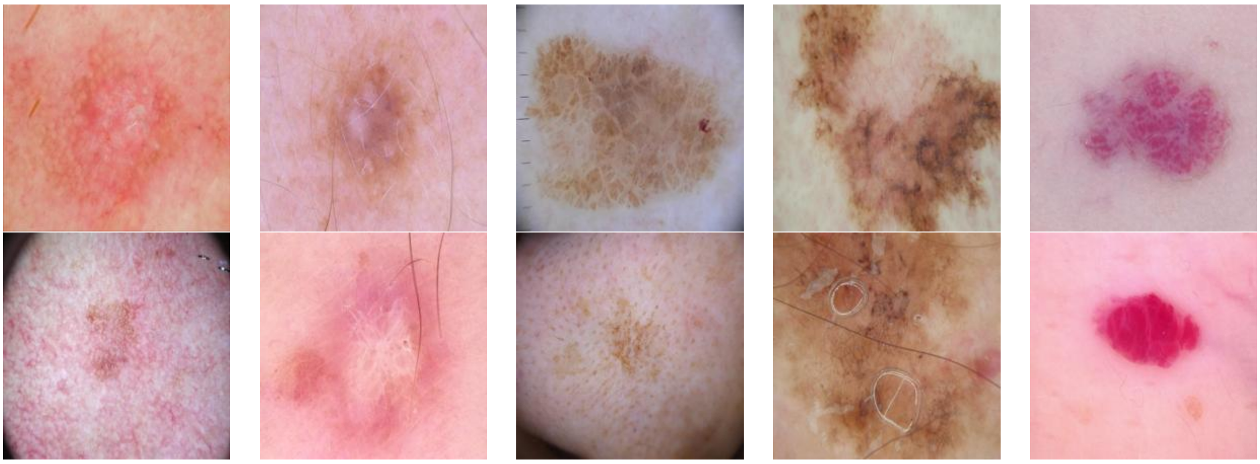
2. RELATED WORK

Classification of skin diseases is important for early diagnosis and early intervention of the disease [24]. Studies on this subject will make the doctor's work easier. It will provide early diagnosis and early intervention. On the other hand, it will lead the patients to start the treatment due to early diagnosis. When the studies are examined, the classification process was carried out using different datasets. Images obtained from different datasets were tried to be classified with different algorithms. Classification studies of skin diseases basically consist of two main parts. The first is classification using traditional methods. The second is classification using DL architectures. In traditional methods, features of the disease are determined and feature extraction is performed. These features are searched on the image and results are produced accordingly. In DL models, the processes are carried out by the layers in the model and the results are produced [25-27].

In a study, the images of 8 classes in the ISIC 2019 dataset were classified using a DL model with transfer learning. In the study, results were obtained by using a model whose weights were determined with GoogleNet. The skin diseases included in the study were classified as melanoma, melanocytic nevus, basal cell carcinoma, benign keratosis, actinic keratosis, dermatofibroma, vascular lesion, and squamous cell carcinoma. The percentages of accuracy, sensitivity, specificity and precision respectively were 94.92%, 79.8%, 97% and 80.36% [28]. Another study sought to classify skin lesions as cancerous or non-cancerous images. In the study, the skin lesion was tried to be classified via AlexNet DL model. In a study using a two-stage method, an accuracy rate of 78% was achieved in cancerous and non-cancerous skin images [29]. In another study, skin diseases in the HAM10000 dataset were classified by using the DenseNet DL model [30]. AlexNet DL model was used in the study for the classification of skin diseases in the ISIC 2018 dataset, which consists of seven classes. The dataset of melanoma, melanocytic nevus, basal cell carcinoma, actinic keratosis, benign keratosis, dermatofibroma, and vascular lesion images yielded 98.70% accuracy, 95.60% sensitivity, 99.27% specificity, and 95.06% accuracy [31].

3. MATERIAL AND METHOD

In this study, images belonging to 5 classes in the ISIC dataset containing skin diseases were classified. These classes were determined as dermatofibroma, pigmented benign keratosis, actinic keratosis, vascular lesion, seborrheic keratosis. DL from multilayer neural networks, one of the artificial intelligence methods, was used for the classification of skin diseases. 9 different DL models were used for classification.



a)Actinic Keratosis b)Dermatofibroma c)Pigmented Benign Keratosis d)Seborrheic Keratosis e)Vascular Lesion

Figure 2. Example images used for classification

As seen in Figure 2, the sample images in the dataset are difficult to distinguish, so the diagnosis must be made by a specialist physician and evaluated in the light of pathological findings The images in each class were augmented by quadrupling with image reproduction techniques. The number of images in each class is given in

Table 1. Number of Image in the dataset used for classification

	Actinic Keratosis	Dermatofibroma	Pigmented Benign Keratosis	Seborrheic Keratosis	Vascular Lesion
Number Of Original Images	114	95	462	77	139
Number Of Augmented Images	456	380	1848	308	556
Total number of Images for Training				3548	

Used DL models and the detailed data of these models are given in Table 2. Table 1. It is known that training is better as the number of images increases in deep learning models. For this reason, the number of samples in the dataset was increased by 4 times. As data augmentation methods, images were applied as reflection, inversion, 90-degree rotation, and 180-degree rotation.

Table 2. Features of used DL models[12]

Model	Number of Layers	Number of Connections	Depth	Number Of Parameter	Top-1 Error rate	Top-5 Error rate
AlexNet	25	-	8	61m	36.7	15.4
VGG16	41	-	16	138m	25.6	8.1
VGG19	47	-	19	144m	25.5	8
GoogleNet	144	170	22	7m	-	6.67
ResNet18	72	79	18	11.7m	30.43	10.76
ResNet50	177	192	50	25.6m	22.8	6.71
ResNet101	347	379	101	44.6m	21.75	6.05
SqueezeNet	68	75	18	1.2m	41.90	19.58
DenseNet201	708	805	201	20m	21.46	5.54

The DL models in Table 2 are frequently encountered in the literature. Each model is at different depths and has different number of layers. The error rates of the models and the number of parameters are important. As the number of parameters increases, the area occupied by the memory increases. Although the structure of each model is different, they can have their own special layer blocks. The dataset used for the classification of skin lesions was carried out using supervised learning. The weights of the models were determined by training the images in each class in the network. Weights were optimized by using transfer learning and classification performances were revealed.

The same hyperparameters were used for each of the deep learning models in Table 2. These hyperparameters are detailed in Table 3.

Table 3. Hyperparameters used for training

Hyperparamers	Value
MaxEpoch	50
MiniBatchSize	32
Solver	SGDM
InitialLearnRate	0.01
ValidationFrequency	50
Momentum	0.9
LearnRateDropFactor	0.1
LearnRateDropPeriod	10
L2Regularization	0.0001

80% of the image in each class is reserved for the training of the network, and 20% for the test. All images were randomly selected during the training. Each model was trained three times and tested after each training. Average of three training and testing results were used for calculating final results. Accuracy, Sensitivity, Precision and F1-score metrics were used for evaluation result criteria. The equations for these metrics are Equation 1, Equation 2, Equation 3, Equation 4 below.

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

$$Sensivity(Recall) = \frac{TP}{(TP+FN)} \tag{2}$$

$$\text{Precision} = \frac{TP}{(TP+FP)} \tag{3}$$

$$\text{F} - \text{Score} = 2 * \frac{(\text{Precision}*\text{Recall})}{(\text{Precision}+\text{Recall})} \tag{4}$$

3. RESULTS

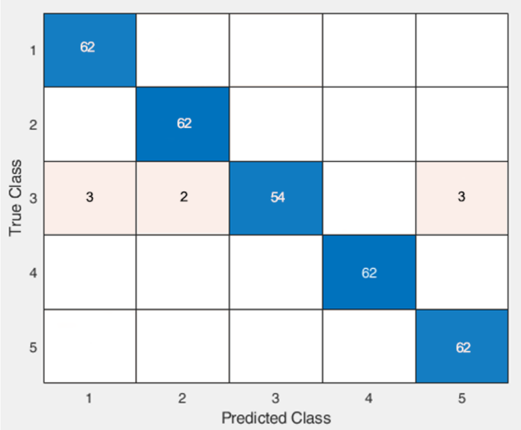
The detailed results obtained in the application for the classification of skin diseases are given in Table 4. In the table, the results of the DL models used for calssification these 5 different type of skin diseases are compared in terms of accuracy, sensivity, precision and F1-scores metrics. The metrics results show that the performance of DL models is quite high. 2810 images were used for training and 710 images were used for testing.

Table 4. Classification evaluation result table of DL models

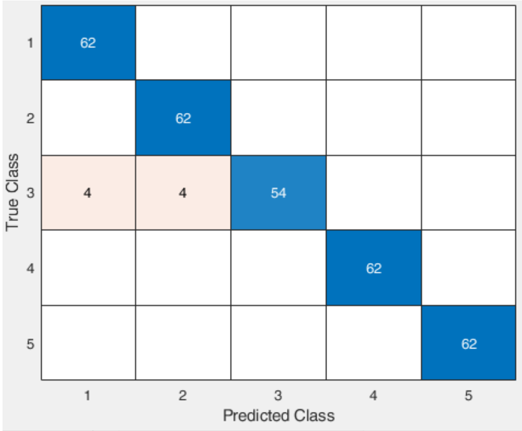
	Model	Image Size	Number of Classes	Accuracy	Sensitivity	Precision	F1 Score
1	SequeezeNet	227*227	5	95.48	95.19	95.17	95.16
2	AlexNet	227*227	5	94.84	95.48	95.8	95.5
3	GoogleNet	224*224	5	95.00	95.8	95.97	95.74
4	Vgg-19	224*224	5	97,1	96.77	96.82	96.69
5	ResNet101	224*224	5	97.41	97.09	97.08	97.08
6	DenseNet201	224*224	5	96,61	97.09	97.13	97.1
7	ResNet-50	224*224	5	97.74	97.42	97.52	97.36
8	ResNet-18	224*224	5	98.06	97.42	97.58	97.37
9	Vgg-16	224*224	5	97.50	97.42	97.48	97.41

SequeezeNet, AlexNet, GoogleNet, Vgg-19, ResNet101, DenseNet20, , ResNet-50, ResNet-18, Vgg-16 were trained with images of different input sizes. Detailed information in the training is given in Table 4. Accuracy, sensitivity, precision and f1 evaluation metrics were used for the obtained results. The results were presented according to these evaluation metrics. It is seen that the results obtained from deep learning models have 94.84% and 98.5% accuracy and f1 score between 95.16 and 97.41. When the results are examined, the model with the highest F1 score among the given models is Vgg-16. It is seen that ResNet-18 has the highest result in the results obtained according to the accuracy rate. It was found that the results obtained from these two models are quite close to each other.

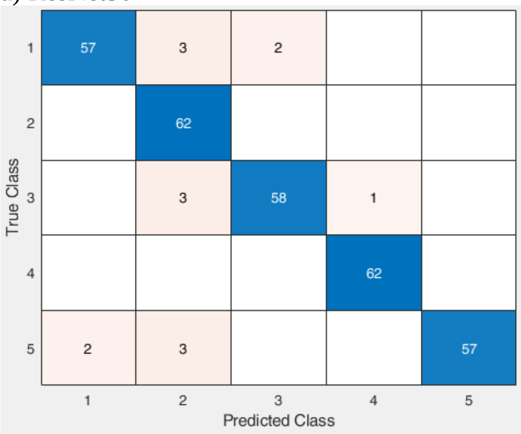
Even though there are small performance differences between the models, it has been observed in the literature studies that the performances of different deep learning models vary depending on the dataset and the study area. It is known in the literature that the Resnet architecture is used more for the dataset and medical images used in the study.



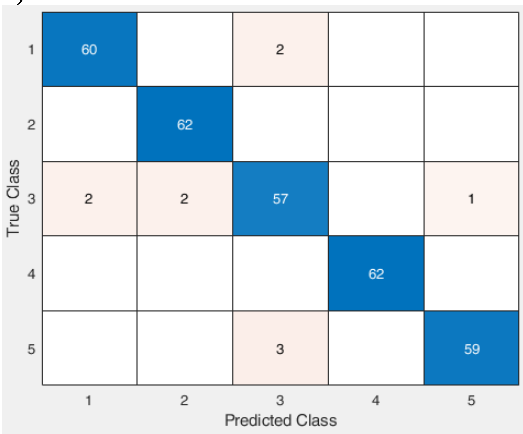
a) ResNet50



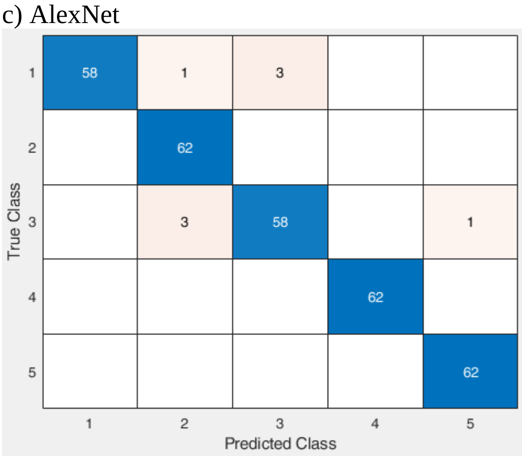
b) ResNet18



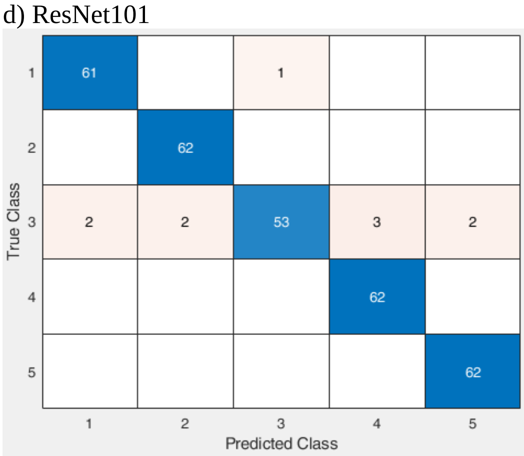
c) AlexNet



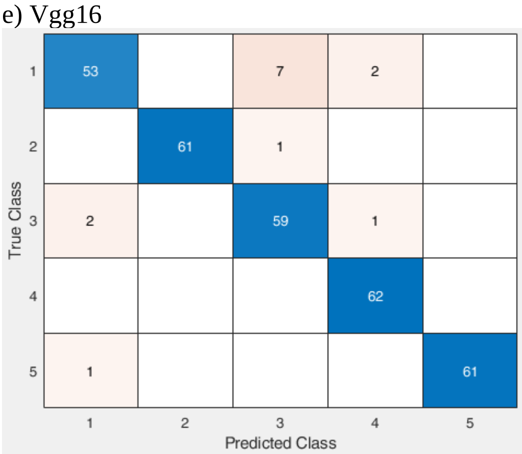
d) ResNet101



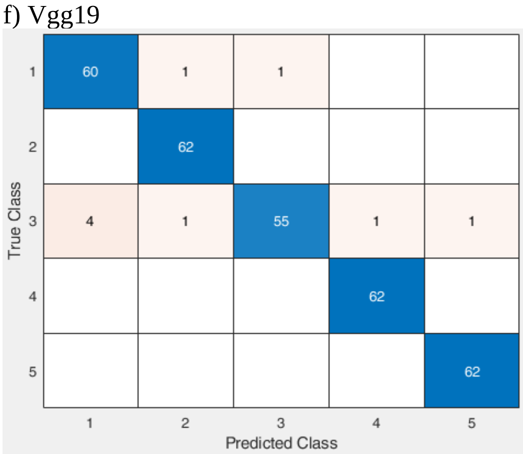
e) Vgg16



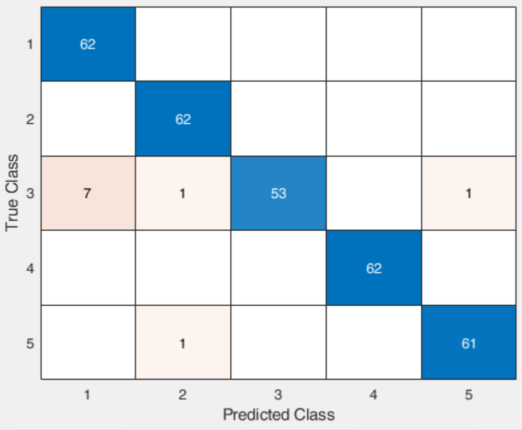
f) Vgg19



g) SequezeeNet



h) DenseNet201



i) GoogleNet

Figure 3. Confusion Matrix from DL models

In the figures shown in Figure 3, there are Confusion Matrices for each model. These matrices show the number of true and false classified images obtained in each model. According to the confusion matrix results, the highest results are obtained by Vgg-16 and ResNet18.

REFERENCES

[1] Hameed, N., A.M. Shabut, and M.A. Hossain. Multi-class skin diseases classification using deep convolutional neural network and support vector machine. in 2018 12th International Conference on Software, Knowledge, Information Management & Applications (SKIMA). 2018. IEEE.

[2] Manoorkar, P., D. Kamat, and P. Patil. Analysis and classification of human skin diseases. in 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT). 2016. IEEE.

[3] Hu, Y., et al., Metabolic syndrome and skin diseases. Frontiers in endocrinology, 2019. 10: p. 788.

[4] Flohr, C., & Hay, R. (2021). Putting the burden of skin diseases on the global map. British Journal of Dermatology, 184(2), 189-190.

[5] Anand, V., Gupta, S., & Koundal, D. (2022). Skin disease diagnosis: challenges and opportunities. In Proceedings of Second Doctoral Symposium on Computational Intelligence: DoSCI 2021 (pp. 449-459). Springer Singapore.

[6] Saunte, D. M., Gaitanis, G., & Hay, R. J. (2020). Malassezia-associated skin diseases, the use of diagnostics and treatment. Frontiers in cellular and infection microbiology, 10, 112.

[7] Outerbridge, C. A., & Owens, T. J. (2023). Nutritional management of skin diseases. Applied Veterinary Clinical Nutrition, 345-383.

[8] Arı, A., O.F. Alcin, and D. Hanbay, Brain MR image classification based on deep features by using extreme learning machines. Biomedical Journal of Scientific and Technical Research, 2020. 25(3).

[9] Toraman, S., T.B. Alakus, and I. Turkoglu, Convolutional capsnet: A novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks. Chaos, Solitons & Fractals, 2020. 140: p. 110122.

[10] Alakus, T.B. and I. Turkoglu, Comparison of DL approaches to predict COVID-19 infection. Chaos, Solitons & Fractals, 2020. 140: p. 110120.

[11] Aktaş, M., F. Doğan, and İ. Türkoğlu. Analysis Of Cracks In Photovoltaic Module Cells From Electroluminescence Images By DL. in 1st International Conference on Computing and Machine Intelligence. 2021.

[12] Dogan, F. and I. Turkoglu, Comparison of DL models in terms of multiple object detection on satellite images. Journal of Engineering Research, 2021.

[13] Das, B., & Toraman, S. (2022). Deep transfer learning for automated liver cancer gene recognition using spectrogram images of digitized DNA sequences. Biomedical Signal Processing and Control, 72, 103317

[14] Demir, F., Sengur, A., Ari, A., Siddique, K., & Alswaitti, M. (2021). Feature mapping and deep long short term memory network-based efficient approach for Parkinson's disease diagnosis. IEEE Access, 9, 149456-149464.

[15] Çalışan, M., Gündüzalp, V., & Olgun, N. Evaluatıon of U-Net and Resnet Archıtectures for Biomedıcal Image Segmentatıon. International Journal of 3D Printing Technologies and Digital Industry, 7(3), 561-570.

[16] Sunnetci, K. M., Kaba, E., Beyazal Çeliker, F., & Alkan, A. (2023). Comparative parotid gland segmentation by using ResNet-18 and MobileNetV2 based DeepLab v3+ architectures from magnetic resonance images. Concurrency and Computation: Practice and Experience, 35(1), e7405.

[17] Doğan, F. and İ. Türkoğlu, Derin öğrenme modelleri ve uygulama alanlarına ilişkin bir derleme. Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 2019. 10(2): p. 409-445.

[18] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2016). Inception-v4, inception-resnet and the impact of residual connections on learning. arXiv preprint arXiv:1602.07261.

[19] 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.

[20] 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).

[21] 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).

[22] Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360

[23] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[24] Selçuk, T., Beyoğlu, A., & Alkan, A. (2022). Automatic detection of exudates and hemorrhages in low-contrast color fundus images using multi semantic convolutional neural network. Concurrency and Computation: Practice and Experience, 34(6), e6768.
[25] Zhou, H., et al., Mean shift based gradient vector flow for image segmentation. Computer Vision and Image Understanding, 2013. 117(9): p. 1004-1016.
[26] Sadri, A.R., et al., Segmentation of dermoscopy images using wavelet networks. IEEE Transactions on Biomedical Engineering, 2012. 60(4): p. 1134-1141.
[27] Xie, F. and A.C. Bovik, Automatic segmentation of dermoscopy images using self-generating neural networks seeded by genetic algorithm. Pattern Recognition, 2013. 46(3): p. 1012-1019.
[28] Kassem, M.A., K.M. Hosny, and M.M. Fouad, Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning. IEEE Access, 2020. 8: p. 114822-114832.
[29] Kaymak, S., P. Esmaili, and A. Serener. DL for two-step classification of malignant pigmented skin lesions. in 2018 14th Symposium on Neural Networks and Applications (NEUREL). 2018. IEEE.
[30] Carcagnì, P., et al. Classification of skin lesions by combining multilevel learnings in a DenseNet architecture. in International conference on image analysis and processing. 2019. Springer.
[31] Hosny, K.M., M.A. Kassem, and M.M. Fouad, Classification of skin lesions into seven classes using transfer learning with AlexNet. Journal of digital imaging, 2020. 33(5): p. 1325-1334.