SAKARYA UNIVERSITY JOURNAL OF COMPUTER AND INFORMATION SCIENCES
http://saucis.sakarya.edu.tr/
e-ISSN: 2636-8129
Publisher: Sakarya University
Vol. 8, No. 2, 312-321, 2025
DOI: 10.35377/saucis..1582098
SAUCIS
SAKARYA UNIVERSITY

RESEARCH ARTICLE

Diagnosis of Lichen Sclerosus, Morphea, and Vasculitis Using Deep Learning Techniques on Histopathological Skin Images

Recep Güler[1], Zehra Karapınar Şentürk[2*], Mehmet Gamsızkan[3], Yunus Özcan[4]

[1] Düzce Vocational School, Düzce University, Düzce, Türkiye, ror.org/04175wc52

[2] Department of Computer Engineering, Düzce University, Düzce, Türkiye, ror.org/04175wc52

[3] Department of Pathology, Faculty of Medicine, Düzce University, Düzce, Türkiye, ror.org/04175wc52

[4] Department of Dermatology, Faculty of Medicine, Düzce University, Düzce, Türkiye, ror.org/04175wc52

Corresponding author:
Zehra Karapınar Şentürk,
Department of Computer Engineering,
Düzce University, Düzce, Türkiye
zehrakarapinar@duzce.edu.tr

Article History:
Received: 10.11.2024
Revised: 15.01.2025
Accepted: 12.04.2025
Published Online: 20.06.2025

ABSTRACT
Skin diseases are very common all over the world. The examination can be done by photographing the relevant area or taking a tissue sample to diagnose skin diseases. Examining tissue samples allows examination at the cellular level. This study discussed three skin diseases: lichen sclerosus, morphea, and cutaneous small vessel vasculitis (vasculitis). For this problem, which does not have an open-access dataset in the literature, a dataset consisting of histopathological images belonging to each class was created. Convolutional neural network models were created for this three-class classification problem, and their results were evaluated. In addition, in this problem where it is difficult to obtain sample images, the efficiency of transfer learning methods was evaluated with a limited number of examples. For this purpose, tests were performed with VGG16, ResNet50, InceptionV3, and EfficientNetB4 models, and the results were given. Among all the results, the accuracy value of the VGG16 model was 0.9755 and gave the best result. However, although the accuracy value was quite good, precision, recall, and f1-score metrics values were around 0.65. This shows deficiencies in how often the model correctly predicts the positive class and how well it predicts all positive examples in the dataset.

Keywords: Convolutional neural networks, Data augmentation, Transfer learning, Histopathology

1. Introduction

Skin diseases are the fourth most common non-fatal disease in 188 countries worldwide [1]. The frequency of skin diseases is increasing with the extension of human life and the effects of modern living conditions. In addition, many skin diseases are chronic and require long treatment periods and multiple physician check-ups. Considering the scarcity of trained specialist physicians, it is inevitable that there will be inadequacies and disruptions in the provision of health services for skin diseases from time to time.

Although the same skin disease can manifest itself in many different ways, most skin diseases can be diagnosed by a specialist at the first examination. However, when specialists have difficulty diagnosing, blood tests that will help are limited. At this point, the methods that help in making a diagnosis are histopathological examinations. Histopathology is the examination of tissue samples by an expert pathologist [2].

In these examinations, a small specimen taken from the skin tissue where the disease is located is examined under a microscope after a series of processes. The tissue architecture formed by the cells and their association with the changes that develop in these due to the diseases are determined. Experts have analyzed hematoxylin and eosin (H&E) stained sections of tissue samples, and it has emerged as the most effective method for examining histopathological images in the last century [3]. At this stage, a second bottleneck emerges in the health system, and service disruptions occur occasionally. The field of work of the pathology physicians who perform these examinations is wide, while the number of experienced physicians who focus on skin diseases is small. The decision support systems to be developed at this stage will reduce these disruptions.

Cite as: R. Güler, Z. Karapınar Şentürk, M. Gamsızkan & Y. Özcan. (2025) Diagnosis of Lichen Sclerosus, Morphea, and Vasculitis Using Deep Learning Techniques on Histopathological Skin Images. *Sakarya University Journal of Computer and Information Sciences*. vol. 8, no. 2, pp. 312-321, 2025 doi: 10.35377/saucis...1582098

Because of the similar appearance of skin diseases, it is not easy to distinguish them by the human eye, and it takes a long time to train experts [4]. For this reason, image processing, segmentation and machine learning studies have been carried out on dermatological images as auxiliary systems. The studies were generally conducted using color photographs of the diseased area, and machine learning and deep learning methods were preferred in processes such as classification and segmentation. A classification study was conducted with EfficientNetB4 using 13603 images containing 14 different skin diseases obtained from a hospital in China [4]. In a different study where photographic images were used for 59 diseases, including vasculitis and morphea, deep learning (ResNet18) based classification was performed [5]. A dataset consisting of skin images of five different skin diseases was tested on popular CNN networks such as ResNet50, DenseNet, MobileNet, Xception and comparative results were shared [6]. Lesion detection and classification were performed in the study, and color images were used to classify dermatological diseases [7]. Some of these studies used existing online datasets, while others created them. Table 1 compares some of these studies, summarizing the dataset's characteristics, the deep learning method used for classification, and the results obtained with the corresponding performance metric.

In the literature, there are also those working on web-based or mobile applications to reduce the workload of physicians and ensure image processing and results are delivered directly to the user. With the developed smartphone application, a DenseNet161-based network was designed to classify 40 skin diseases, such as eczema, melanoma, and lichen sclerosus [1]. A different study presented a web application where users of 5 skin diseases could learn the predicted disease result by uploading skin photos [6]. However, the long processing time of the mentioned web application and the fact that it only gives results for a limited number of diseases can be stated as the points that need to be completed in the study [6].

Another method used in addition to dermatological photographs to examine skin diseases in more detail is histopathological examination of tissue samples. Experts can examine Histopathological images directly under a microscope, or images can be transferred to a computer using digital microscopes. Another way to transfer histopathological images to a computer is to use whole slide imaging (WSI) devices. Thus, images can be recorded at high resolution (100000x100000 pixels, etc.). After the positive effects of deep learning methods on objectivity and efficiency in cancer diagnosis with full slide imaging emerged, many studies have been conducted on diagnosing or segmenting diseases such as skin cancer, prostate cancer, lung cancer, etc. [3]. Due to the large size of the images obtained from full slide imaging, applying them directly as input to CNN networks is impossible. When the size of the images is reduced to a suitable resolution for CNN networks, since there will be losses in image features, extracting smaller patches from large images and using them as input increases efficiency [8], [9].

In convolutional neural networks (CNN) and deep learning models, the size of the dataset, its correct labeling and distribution are important for solving image classification problems with high accuracy. Especially since the datasets consisting of medical images are small and labeling the images is costly, the number of images in the dataset can be increased using data augmentation methods. It is known that data augmentation methods used in computer vision studies contribute positively to the overall performance and that the combined use of several data augmentation methods generally increases the performance [10], [13]. Data augmentation methods have been examined in different CNN networks for melanoma classification. It has been shown that the best results are obtained by using geometric and color transformation methods together [11]. Among the data augmentation methods, synthetic data creation also uses the features of the existing data. Experimental studies have been conducted on melanoma, histopathological images, and MRI images related to the study. A new method is presented to obtain new data by style transfer using the texture and color features of the images in the existing data set [12]. It has also been shown that overfitting can be prevented in CNN networks by using data augmentation methods [13], [14]. Data augmentation is usually performed on the training dataset, and as a result, accuracy, precision, sensitivity and F1-score values are expected to increase. In addition to expanding the dataset with data augmentation, class imbalances can also be eliminated.

Table 1. Selected deep learning-based skin image classification studies

Ref.	Dataset	Classifier	Performance measure
1	15418 images for 40 skin disease	DenseNet161	Overall accuracy 0.769
4	13603 images for 14 skin disease	Custom CNN based on EfficientNet-B4	Overall accuracy 0.948
5	70196 images for 59 skin disease	ResNet18	Overall accuracy 0.579
6	18692 images for 5 skin disease	ResNet50, Inception-V3, Inception-ResNet, DenseNet, MobileNet, Xception	MobileNet accuracy 0.960 Xception accuracy 0.979
7	505 images for 5 skin disease	DenseNet121	Overall accuracy 0.952

A generalized model and many images are needed to create a deep-learning model with high accuracy. The transfer learning method eliminates the need for computational resources and creates models that can produce faster results [6]. When machine learning-based studies in cancer diagnosis are examined, attempts are made to increase the prediction accuracy by using data augmentation and transfer learning methods. In classifying benign and malignant skin lesions with CNN networks, data preprocessing and data augmentation methods are used to remove image artifacts and correct imbalances in the data set [15]. ResNet, Inception, VGG, and DenseNet networks are frequently used in the transfer learning approach, and their results are compared. With the different architecture, filter size, and parameter number values of these deep learning networks, the

accuracy values obtained in medical image classification vary depending on the feature of the image [16], [17]. High accuracy values can be obtained with transfer learning methods to detect and classify skin and breast cancer [17], [18]. Learning processes can be accelerated with transfer learning methods in the segmentation of cancer tissue [19]. In addition, the transfer learning method is not used directly for classification but for feature extraction, and the obtained features can be used in a different CNN network [19], [20], [21].

This study uses histopathological images to present a computer-aided diagnosis of lichen sclerosus, morphea and vasculitis diseases. Different CNN network structures were trained and tested using medical image data augmentation and transfer learning. Unlike studies in the literature using skin image datasets, our study is one of the first studies in the field using histopathological images to classify related diseases using machine learning. The main contributions of this study are as follows:

- A new histopathological dataset was created to diagnose lichen, morphea and vasculitis diseases and deep learning was used for classification.
- The success of data augmentation and transfer learning methods was demonstrated on the dataset.

The rest of the paper is structured as follows: section 2 presents the dataset's characteristics, data preprocessing stages and the methods mentioned in the CNN networks used. In section 3, under the results and discussion, the results of the training and testing processes are presented in detail (accuracy, precision, sensitivity and F1-score and confusion matrix) in a table. Finally, in section 4, the results are presented in the conclusion.

2. Methods

The proposed method for classifying histopathological images consists of three main parts (Fig. 1). The first part is transferring the tissue sample to the digital environment and creating the data set. The second part is the resizing process to bring the images to the appropriate input size for the CNN network to be used with data preprocessing. Following this process, data sets consisting of different samples were created with data augmentation to see the effect of the data set size on success. In the last part, training and testing processes were performed with a custom CNN and deep learning models (ResNet, VGG, Inception, EfficientNet) found in the literature.

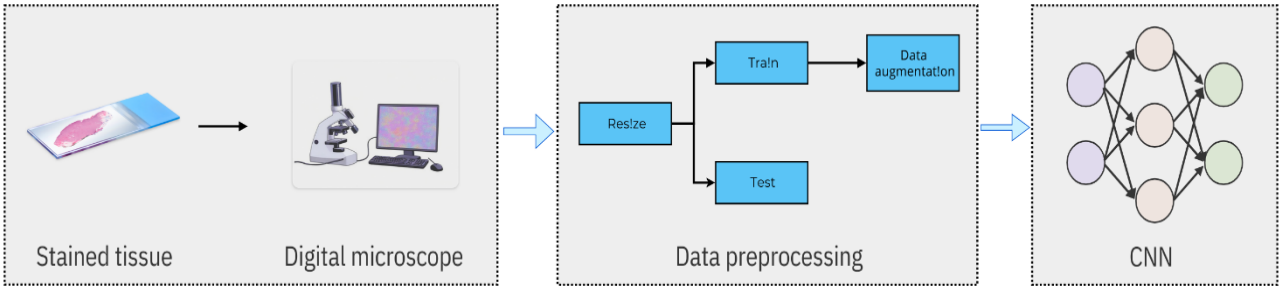

Figure 1. The general structure of the study

2.1 Dataset

The specimen obtained for histopathological examinations is first reduced to appropriate dimensions and embedded in paraffin. A 3-5 µm thick section is taken from this prepared specimen. This section, fixed on a glass called a slide, is stained using different dyes and is made ready for viewing under a microscope after a series of processes.

Usually, the first sample is prepared with hematoxylin-eosin stain, and subsequent stain selections are determined by the tissue being processed and the disease being investigated. When an examination with a different stain, a new section is taken from the tissue previously stored in paraffin and stained. Considering the size of the sections, the newly taken section usually looks similar to the previous one when viewed with the naked eye, but it is never the same.

The dyes used are usually determined by the region intended to be examined. Among the diseases we study, lichen sclerosus and morphea are diseases in which the structure of proteins called collagen is disrupted. The Masson Trichrome dye we use allows examination by staining areas where collagen is present more visibly. In patients with cutaneous vasculitis, since collagen is not affected, it is expected that no abnormal pattern will occur.

This study consists of 147 microscopic skin images of three diseases (lichen sclerosus, morphea, vasculitis) created by Düzce University Faculty of Medicine, Department of Pathology specialist doctors. After the tissue samples were stained with Masson Trichrome dye by the specialists, they were saved as 1920x1080 JPG files using a digital microscope. The distribution of diseases constituting the data set is given in Table 2. The images that comprise the data set do not contain personal information.

Table 2. Dataset properties

Disease	Number of Images
Lichen sclerosus	51
Morphea	42
Vasculitis	54

Convolutional neural networks are deep learning structures that consist of a multi-layered structure and require a large amount of input data to make the right decision. Although it is possible to access many different data sets from different fields of study today, it has been observed that medical data sets, in particular, are not sufficient in number or not open access. In addition, labeling images by experts in creating a medical data set is disadvantageous in terms of time and cost.

2.2 Data Preprocessing

The images that make up the dataset are large enough to be used in an artificial neural network model. Although it is possible to work with large-sized images and produce better results by preventing data loss, they are not preferred since they require a lot of processing power and have time-cost disadvantages. Smaller images, such as 224x224, are generally used in the literature. Thus, less processing power is required for matrix calculations. In addition, the squareness of the input image allows convolution, padding, and resizing operations to be performed more efficiently.

In image classification problems, the model's ability to achieve better results is directly related to the size of the data set. Using data augmentation methods, it aims to increase the numerical increase of the image in the data set and its diversity. In this way, in addition to better learning, the overfitting problem can be reduced [10]. It has been shown in studies that the use of data augmentation increases the performance of deep learning models. Data augmentation methods include geometric transformation, color space transformations, Kernel filters, image deletion, image fusion, style transfer, and GAN-based image generation [14].

In this study, geometric transformation-based data augmentation methods were used. Flipping is obtaining a mirror image of the image horizontally or vertically (Figure 2b). Rotating is rotating the image around itself by a certain degree (Figure 2c). Scaling can be applied by zooming in or out of the image (Figure 2d).

Since the microscopic histopathological images have large dimensions (1920x1080), they were resized during preprocessing. Resizing the images at this stage can also cause data loss. A total of four data sets were created: two data sets (A and B) to see the effect of image resizing on the performance of the CNN model used and two data sets (C and D) to see and compare the performance of data augmentation processes (Table 2). First, the data set was divided into 85% for training and 15% for testing. Then, the data allocated for training was divided again into 85% for training and 15% for validation during the training of the CNN model.

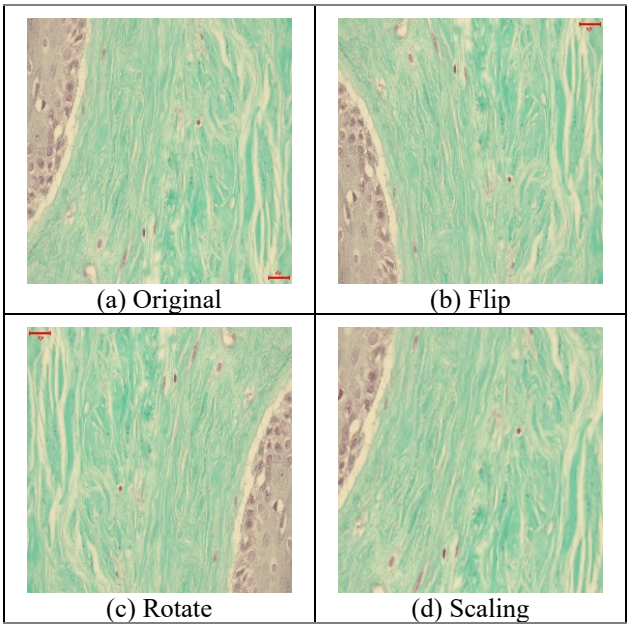

Figure 2. Geometric transformations

After separating the dataset into training and testing, the dataset clusters whose features are given in the Table 3 were created. Dataset A was created by resizing the original images to 256x256 pixels, and dataset B was created by resizing the original

images to 512x512 pixels. Then, the training data was increased approximately 5 and 10 times with the previously mentioned geometric data augmentation methods, and dataset clusters C and D were created, respectively.

Table 3. Created dataset clusters (LS=Lichen Sclerosus, M=Morphea, V=Vasculitis)

Dataset	Image Size	Training		Test		Total Number of Images
A	256x256	122	LS:42 M:35 V:45	25	LS:9 M:7 V:9	147
B	512x512	122	LS:42 M:35 V:45	25	LS:9 M:7 V:9	147
C	256x256	675	LS:225 M:225 V:225	25	LS:9 M:7 V:9	700
D	256x256	1350	LS:450 M:450 V:450	25	LS:9 M:7 V:9	1375

2.3. CNN Models

The number and success of deep learning studies are increasing day by day. Convolutional neural networks are a sub-element of deep learning, which is frequently used, especially in classification and image processing studies [8], [12]. Convolutional neural networks consist of convolution, pooling and fully connected layer sections. CNN models can be designed with different sizes and features to perform different tasks.

Deep learning models require large data sets and the ability to include quite complex operations to extract the most important and useful features from the data set for classification. One disadvantage of working with complex models and large data sets is that training a deep-learning model takes a long time. In medical image processing, there is usually a problem with small data sets and images not being labeled correctly. Transfer learning is used to overcome these problems in deep learning. Transfer learning uses knowledge from models trained on large datasets for different tasks and datasets [22], [23]. With transfer learning, models trained using large datasets are used especially to perform tasks with a small number of images [24]. AlexNet is one of the first large-scale convolutional neural network models developed. AlexNet won the ImageNet large-scale image recognition challenge in 2012, drawing attention to the success of CNN models in image processing [25]. In 2014, VGG and GoogLeNet (Inception) achieved better results on the ImageNet dataset with deeper CNN models [26], [27]. Considering that deeper networks are more difficult to train and that accuracy does not always increase even with increasing network depth, the ResNet model, which was developed, facilitated the training of deep networks and won the ImageNet challenge in 2015 [28]. These models were studied on the ImageNet dataset, which consists of 1,281,167 training images, 50,000 validation images, and 100,000 test images of 1000 object classes [29].

Table 4. Features of the created CNN models

CNN Model	Number of Convolution Layer	Activation Function	Total Parameter
Model 1	2	ReLU	134239171
Model 2	2	ReLU	134241603
Model 3	3	ReLU	33612803
Model 4	3	ReLU	31548419

This study created four CNN models to detect diseases from microscopic skin images. The features of these CNN models are given in Table 4. ReLU was used as each model's activation function, and the MaxPooling (2x2) operation was performed after the convolution process. In addition, the dropout functions used were used to prevent memorization during training. Finally, the classification result for three diseases was obtained with the SoftMax output layer.

Table 5. Features of the used pre-trained models

CNN Model	Number of Total Layer	Activation Function	Total Parameter
VGG16	16	ReLU	138.4M
ResNet50	50	ReLU	25.6M
InceptionV3	189	ReLU	23.9M
EfficientNetB4	258	Softmax	19.5M

Using the transfer learning method, VGG16, ResNet50, InceptionV3, and EfficentNetB4 pre-trained models were also tested on the same datasets. The features of the pre-trained networks used are given in Table 4.

3. Results and Discussion

The previously mentioned datasets were tested on a total of eight networks. The studies were performed with Keras and TensorFlow on Google Colab. The program was run for the first time for each network, and learning was performed for 50 epochs. While evaluating the networks, Accuracy, Precision, Recall (Sensitivity), and F1-score metrics were calculated.

Table 6. Test Results

Data Set	CNN Model	Validation Accuracy	Precision	Recall	F1 score
A	Model 1	0.5714	**0.70**	**0.64**	**0.64**
	Model 2	0.3333	0.13	0.36	0.19
	Model 3	0.4762	0.49	0.44	0.45
	Model 4	0.4286	0.13	0.36	0.19
	VGG16	0.7619	0.64	0.60	0.60
	ResNet50	**0.8095**	0.57	0.60	0.57
	InceptionV3	0.5238	0.39	0.28	0.28
	EfficientNetB4	0.7143	0.48	0.44	0.45
B	Model 1	0.5238	0.66	0.60	0.61
	Model 2	0.3333	0.13	0.36	0.19
	Model 3	0.5714	**0.78**	**0.72**	**0.70**
	Model 4	0.3810	0.13	0.36	0.19
	VGG16	**0.8095**	0.59	0.56	0.52
	ResNet50	**0.8095**	0.67	0.64	0.65
	InceptionV3	0.4762	0.32	0.40	0.34
	EfficientNetB4	0.6190	0.53	0.52	0.52
C	Model 1	0.5392	0.58	0.52	0.51
	Model 2	0.7647	0.32	0.44	0.37
	Model 3	0.6569	0.67	0.56	0.57
	Model 4	0.5392	0.50	0.48	0.48
	VGG16	0.9314	0.62	0.60	0.59
	ResNet50	**0.9608**	**0.81**	**0.80**	**0.79**
	InceptionV3	0.5784	0.41	0.40	0.35
	EfficientNetB4	0.8922	0.56	0.48	0.45
D	Model 1	0.6520	0.20	0.28	0.23
	Model 2	0.7892	0.56	0.52	0.46
	Model 3	0.4755	**0.72**	0.60	0.61
	Model 4	0.7598	0.48	0.48	0.46
	VGG16	**0.9755**	0.69	0.68	0.68
	ResNet50	0.9608	**0.72**	**0.72**	**0.71**
	InceptionV3	0.7157	0.52	0.52	0.51
	EfficientNetB4	0.9118	0.55	0.48	0.49

Precision, Recall, and F1-score values were calculated for the performance measurement of the test results. Accuracy gives the correct result rate among all measured results (Equation 1). Precision gives the proportion of correctly predicted among all correctly labeled examples (Equation 2). Sensitivity gives the correct prediction rate among all examples (Equation 3). The F1-score value is the harmonic mean of precision and sensitivity values (Equation 4).

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

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

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

The performance of each model on the created datasets is presented in Table 6. When the evaluation is made according to the validation process results, it is seen that for dataset A, the best accuracy value among the created CNN models belongs to Model 1 with 0.5714 and ResNet50 from the pre-trained networks with 0.8095. For dataset B, the best accuracy value among the created CNN models belongs to Model 3 with 0.5714 and ResNet50 from the pre-trained networks with 0.8095. Considering that no data augmentation was performed on datasets A and B, it is seen that the results obtained for the dataset containing a small number of images are not sufficient for the created CNN models. However, the results of the pre-trained networks were better than expected.

Table 5 shows that especially pre-trained networks have much better accuracy values for the C and D data sets created by applying data augmentation. For dataset C, the best accuracy value among the created CNN models belongs to Model 2 with 0.7642 and ResNet50 from the pre-trained networks with 0.9608. For dataset D, the best accuracy value among the created CNN models belongs to Model 2 with 0.7892 and VGG16 from the pre-trained networks with 0.9755.

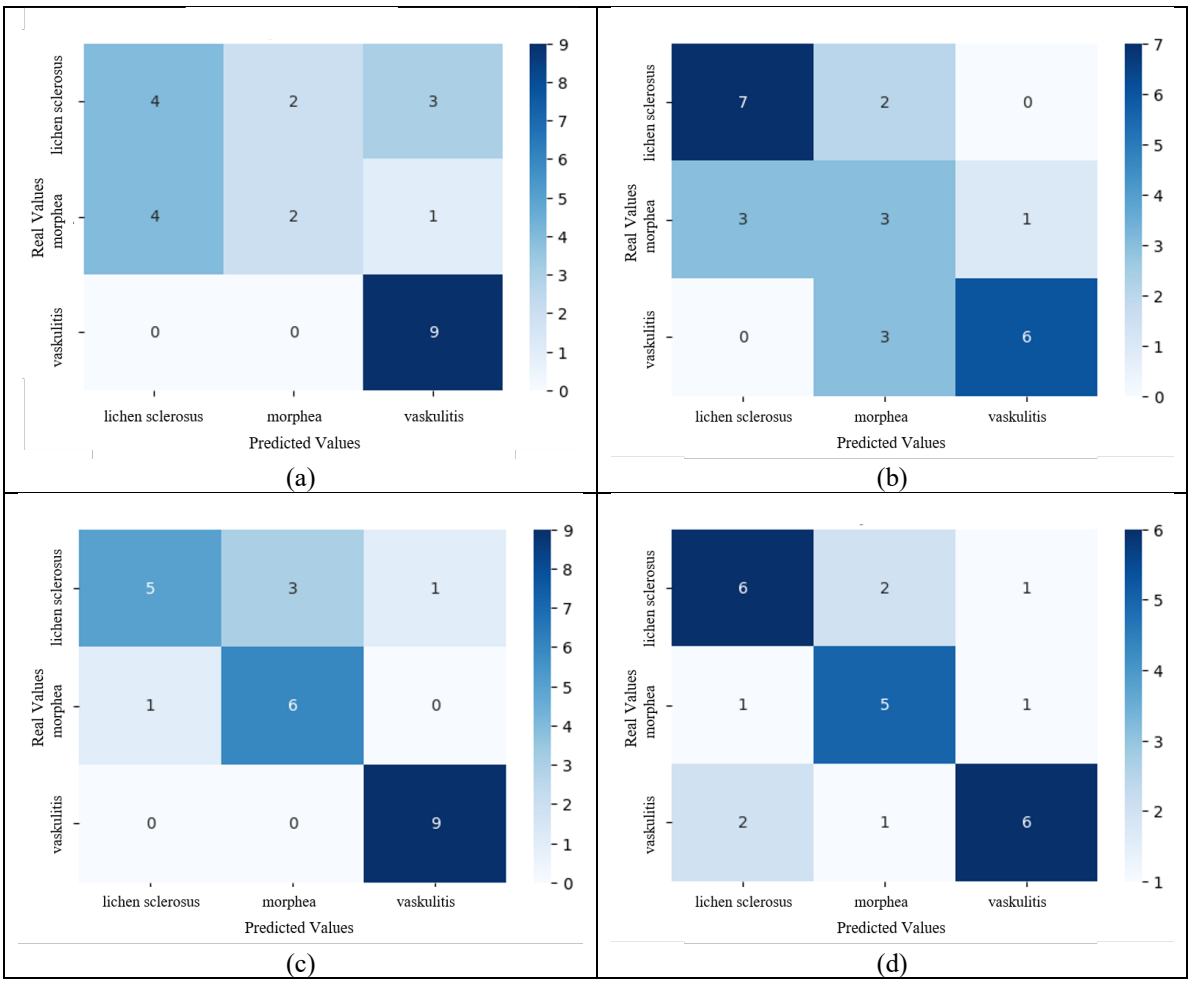

Figure 3. Confusion matrix for the best models concerning datasets A, B, C, and D

The model's performance can also be visualized by creating a Confusion Matrix based on the test results. Confusion matrices for the best models according to the datasets are given in Figure 3. The confusion matrices are given in Figure 3a for the ResNet50 model for dataset A, Figure 3b for the ResNet50 model for dataset B, Figure 3c for the ResNet50 model for dataset C, and Figure 3d for the VGG16 model for dataset D, respectively.

Although the accuracy values obtained due to the training and validation processes of the models are quite good, the precision, sensitivity, and F1-score values obtained as a result of the test are lower than the accuracy value. This situation shows that the model is inadequate in correctly predicting the positive class. Data augmentation was performed for the images in the data set for training and validation, thus increasing the diversity and preventing the overfitting problem. However, to see the model's performance on real data and obtain reliable results, it is important not to perform data augmentation on the test data

set. Since the number of images in the created data set is limited, the small amount of data allocated for testing causes the precision, sensitivity, and F1-score values more sensitive.

Additionally, according to the experimental results, although we obtained high accuracy values in certain models, other parameters are relatively low. As we have seen in the literature, the main reason for the low results is a class imbalance in the data set or the nature of the data set. There are no class imbalances in the data sets we used, both for the original and the data augmented data set. However, when the complexity matrices are examined in more detail, it is seen that lichen sclerosus and morphea diseases are predicted more incorrectly than vasculitis instead of each other. While there is a deterioration in the collagen structure in lichen sclerosus and morphea diseases, vasculitis disease does not affect the collagen structure. The number of convolution layers can be increased for a deep learning model that produces better predictions, or a different model can be developed to distinguish lichen sclerosus and morphea diseases better.

4. Conclusion

This study applied classification with machine learning methods for three different skin diseases: Lichen Sclerosus, Morphea, and Vasculitis. Testing operations were performed on eight different models using transfer learning with CNN models and different deep learning models created within the scope of the study. Unlike studies on the classification of skin diseases with machine learning, histopathological images were used in this study for the first time. Since cells and tissues are detailed in histopathological images, detecting related diseases through features extracted from these images provides more reliable and accurate results than diagnostic methods performed with non-invasive images. With the proposed approach, even if the effect of the disease on the skin is not yet seen, the disease status can be revealed with the changes observed at the cellular level.

According to the test results, better results were obtained in both CNN networks and transfer learning methods on data sets with data augmentation. Among all the results, although the accuracy value of the VGG16 model was obtained as the highest with 0.9755, it is seen that the overall performance of the ResNet50 model is good on all data sets. It has been shown with experimental results that the classification performance can be increased by using transfer learning in small-scale data sets.

Our future work will continue to evaluate the reliability of the classification performed with real and artificially augmented images by increasing the number of images in the dataset and obtaining better results by fine-tuning the parameters in transfer learning.

References

[1] R. Pangti *et al.*, "A machine learning-based, decision support, mobile phone application for diagnosis of common dermatological diseases," *Journal of the European Academy of Dermatology and Venereology*, vol. 35, no. 2, pp. 536–545, Feb. 2021, doi: 10.1111/JDV.16967.

[2] O. Jimenez-Del-Toro *et al.*, "Analysis of Histopathology Images: From Traditional Machine Learning to Deep Learning," *Biomedical Texture Analysis: Fundamentals, Tools and Challenges*, pp. 281–314, Jan. 2017, doi: 10.1016/B978-0-12-812133-7.00010-7.

[3] G. Litjens *et al.*, "Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis," *Sci Rep*, vol. 6, no. 1, p. 26286, May 2016, doi: 10.1038/srep26286.

[4] C. Y. Zhu *et al.*, "A Deep Learning Based Framework for Diagnosing Multiple Skin Diseases in a Clinical Environment," *Front Med (Lausanne)*, vol. 8, p. 626369, Apr. 2021, doi: 10.3389/FMED.2021.626369/BIBTEX.

[5] M. Tanaka *et al.*, "Classification of a large-scale image database of various skin diseases using deep learning," *Int J Comput Assist Radiol Surg*, vol. 16, no. 11, pp. 1875–1887, Nov. 2021, doi: 10.1007/S11548-021-02440-Y/FIGURES/8.

[6] R. Sadik, A. Majumder, A. A. Biswas, B. Ahammad, and M. M. Rahman, "An in-depth analysis of Convolutional Neural Network architectures with transfer learning for skin disease diagnosis," *Healthcare Analytics*, vol. 3, p. 100143, Nov. 2023, doi: 10.1016/J.HEALTH.2023.100143.

[7] E. Goceri, "Deep learning based classification of facial dermatological disorders," *Comput Biol Med*, vol. 128, p. 104118, Jan. 2021, doi: 10.1016/J.COMPBIOMED.2020.104118.

[8] L. Duran-Lopez, J. P. Dominguez-Morales, A. F. Conde-Martin, S. Vicente-Diaz, and A. Linares-Barranco, "PROMETEO: A CNN-Based Computer-Aided Diagnosis System for WSI Prostate Cancer Detection," *IEEE Access*, vol. 8, pp. 128613–128628, 2020, doi: 10.1109/ACCESS.2020.3008868.

[9] Y. Wu *et al.*, "Recent Advances of Deep Learning for Computational Histopathology: Principles and Applications," *Cancers (Basel)*, vol. 14, no. 5, p. 1199, Feb. 2022, doi: 10.3390/cancers14051199.

[10] A. Mumuni and F. Mumuni, "Data augmentation: A comprehensive survey of modern approaches," *Array*, vol. 16, p. 100258, Dec. 2022, doi: 10.1016/J.ARRAY.2022.100258.

[11] F. Perez, C. Vasconcelos, S. Avila, and E. Valle, "Data augmentation for skin lesion analysis," *Lecture Notes in*

Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11041 LNCS, pp. 303–311, 2018, doi: 10.1007/978-3-030-01201-4_33/FIGURES/3.

[12] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," *2018 International Interdisciplinary PhD Workshop, IIPhDW 2018*, pp. 117–122, Jun. 2018, doi: 10.1109/IIPHDW.2018.8388338.

[13] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/J.GLTP.2022.04.020.

[14] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019, doi: 10.1186/S40537-019-0197-0/FIGURES/33.

[15] M. S. Ali, M. S. Miah, J. Haque, M. M. Rahman, and M. K. Islam, "An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models," *Machine Learning with Applications*, vol. 5, p. 100036, Sep. 2021, doi: 10.1016/J.MLWA.2021.100036.

[16] A. Saber, M. Sakr, O. M. Abo-Seida, A. Keshk, and H. Chen, "A Novel Deep-Learning Model for Automatic Detection and Classification of Breast Cancer Using the Transfer-Learning Technique," *IEEE Access*, vol. 9, pp. 71194–71209, 2021, doi: 10.1109/ACCESS.2021.3079204.

[17] M. Fraiwan and E. Faouri, "On the Automatic Detection and Classification of Skin Cancer Using Deep Transfer Learning," *Sensors 2022, Vol. 22, Page 4963*, vol. 22, no. 13, p. 4963, Jun. 2022, doi: 10.3390/S22134963.

[18] H. Aljuaid, N. Alturki, N. Alsubaie, L. Cavallaro, and A. Liotta, "Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning," *Comput Methods Programs Biomed*, vol. 223, p. 106951, Aug. 2022, doi: 10.1016/J.CMPB.2022.106951.

[19] S. Hosseinzadeh Kassani, P. Hosseinzadeh Kassani, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Deep transfer learning based model for colorectal cancer histopathology segmentation: A comparative study of deep pre-trained models," *Int J Med Inform*, vol. 159, p. 104669, Mar. 2022, doi: 10.1016/J.IJMEDINF.2021.104669.

[20] N. Ahmad, S. Asghar, and S. A. Gillani, "Transfer learning-assisted multi-resolution breast cancer histopathological images classification," *Visual Computer*, vol. 38, no. 8, pp. 2751–2770, Aug. 2022, doi: 10.1007/S00371-021-02153-Y/FIGURES/21.

[21] A. Ben Hamida *et al.*, "Deep learning for colon cancer histopathological images analysis," *Comput Biol Med*, vol. 136, p. 104730, Sep. 2021, doi: 10.1016/J.COMPBIOMED.2021.104730.

[22] P. Kora *et al.*, "Transfer learning techniques for medical image analysis: A review," *Biocybern Biomed Eng*, vol. 42, no. 1, pp. 79–107, Jan. 2022, doi: 10.1016/J.BBE.2021.11.004.

[23] M. Iman, H. R. Arabnia, and K. Rasheed, "A Review of Deep Transfer Learning and Recent Advancements," *Technologies 2023, Vol. 11, Page 40*, vol. 11, no. 2, p. 40, Mar. 2023, doi: 10.3390/TECHNOLOGIES11020040.

[24] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: a literature review," *BMC Medical Imaging 2022 22:1*, vol. 22, no. 1, pp. 1–13, Apr. 2022, doi: 10.1186/S12880-022-00793-7.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. Burges, L. Bottou, and K. Q. Weinberger, Eds., Curran Associates, Inc., 2012.

[26] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, Sep. 2014, Accessed: Sep. 06, 2024.

[27] C. Szegedy *et al.*, "Going Deeper with Convolutions," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June-2015, pp. 1–9, Sep. 2014, doi: 10.1109/CVPR.2015.7298594.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 770–778, Dec. 2015, doi: 10.1109/CVPR.2016.90.

[29] "ImageNet," https://www.image-net.org/.

Author(s) Contributions

Recep Güler contributed to design, data collection, data analysis, writing, material support and literature review. Zehra Karapınar Şentürk contributed to design, data analysis, writing, material support and literature review. Mehmet Gamsızkan contributed to data collection, data analysis, material support. Yunus Özcan contributed to data collection, data analysis, writing, material support.

Acknowledgments

The authors would like to thank Research Assistant Sümeyye Güneş Takır for creating the dataset used in this study.

Conflict of Interest Notice

The authors declare that there is no conflict of interest regarding the publication of this paper.

Ethical Approval and Informed Consent

It is declared that during the preparation process of this study, scientific and ethical principles were followed, and all the studies benefited from are stated in the bibliography.

Availability of data and material

Not applicable.

Artificial Intelligence Statement

No artificial intelligence tools were used while writing this article.

Plagiarism Statement

This article has been scanned by iThenticate ™.