Düzce University
Journal of Science & Technology

Research Article

Skin Type Detection with Deep Learning: A Comparative Analysis

Fatma Betül KARA ARDAÇ [a,*], Resul KARA [a], Seda Sakacı Çelik [b]

[a] *Duzce University, Dep. of Computer Engineering, Faculty of Engineering, Duzce,Turkey*
[b] *Seda Custom Skincare, 1125 Coast Village Rd., Santa Barbara California, USA*
* *Corresponding author's e-mail address*: *fatmabetulardac@duzce.edu.tr*
DOI:10.29130/dubited.930096

ABSTRACT

There are many factors that can change and affect appearance, including age and environment. Knowing the skin type helps to choose the products that best suit the needs of the skin and therefore the right skin care. Recently, the increasing demand for cosmetics and the scarcity of adequately equipped cosmetics have led the industry to artificial intelligence applications. Artificial intelligence applications can give highly accurate results in skin type classification. The aim of this study is to find the best classification model for skin type prediction in skin analysis data with deep learning. In accordance with this purpose; Two different deep learning architectures, CNN and LSTM were used. In order to find the best classification model, deep learning models were created by performing hyper parameter optimization with Tanh, ReLU and Sigmoid activation functions, Adam, SGD and RMSProp optimization algorithms and combinations of 50, 100, 500 epoch numbers. In experimental studies, the performance of the models varies according to the parameters, and it has been observed that the most successful deep learning model is the model consisting of a combination of CNN architecture, Adam optimization algorithm and Sigmoid activation function and 500 epochs with a accurate of 83.75%. The obtained accuracy result has a higher classification success compared to other architectures and shows that deep learning architectures can correctly classify skin type.

Key Words: *Classification, CNN, deep learning, LSTM, skin type*

Derin Öğrenme ile Cilt Tipi Tespiti: Karşılaştırmalı Bir Analiz

ÖZET

Yaş ve çevre dahil görünümü değiştirebilecek ve etkileyebilecek birçok faktör vardır. Cilt tipini bilmek, cildin ihtiyaçlarına en uygun ürünleri ve dolayısıyla doğru cilt bakımını seçmeye yardımcı olur. Son zamanlarda, kozmetik için artan talep ve yeterli donanıma sahip kozmetikçilerin azlığı, sektörü yapay zekâ uygulamalarına yöneltmiştir. Bu çalışmanın amacı, derin öğrenme ile cilt analizi verilerinde cilt tipi tahmini için en iyi sınıflandırma modelini bulmaktır. Bu amaç doğrultusunda; CNN ve LSTM olmak üzere 2 farklı derin öğrenme mimarisi kullanılmıştır. En iyi sınıflandırma modelini bulmak için Tanh, ReLU ve Sigmoid aktivasyon fonksiyonları, Adam, SGD ve RMSProp optimizasyon algoritması ve 50, 100, 500 epoch sayılarının kombinasyonları ile hiper parametre optimizasyonu yapılarak derin öğrenme modelleri oluşturulmuştur. Deneysel çalışmalarda, modellerin performansı parametrelere göre değişmekte olup en başarılı derin öğrenme modeli %83,75 başarı oranı ile CNN mimarisi, Adam optimizasyon algoritması, Sigmoid aktivasyon fonksiyonu ve 500 epoch kombinasyonundan oluşan modelin olduğu gözlemlenmiştir. Elde edilen doğruluk sonucu diğer mimarilere kıyasla daha yüksek bir sınıflandırma başarısına sahiptir ve derin öğrenme mimarilerinin doğru bir şekilde cilt tipi sınıflandırması yapabileceğini göstermektedir.

Anahtar Kelimeler: *Cilt tipi, CNN, derin öğrenme, LSTM, sınıflandırma*

Received: 24/09/2021, Revised: 01/05/2021, 24/06/2022

I. INTRODUCTION

Since every person is unique, the skin structure is different. It is important to know the skin type in order to take the best care of the skin. Knowing the skin and using the right applications and products can prevent eczema, psoriasis, rosacea, acne, itching, aging and allergic problems [1]. Skin type can be learned by performing a cosmetological skin analysis to recognize the skin. In cosmetological skin analysis, various measurements are made with very sensitive probes. In the measurements made, the pH of the skin, melanin condition, temperature, elasticity, hydration, oxidative stress, degree of acne tendency, wrinkle degree, pore clogging, sensitivity, brightness, size and depth of spots are obtained [2]. Tests performed with skin analysis devices in determining skin type are a complete professional skin assessment system. After the skin analysis, the skin type of the person is determined with the information of the cosmetologist. Different automatic approaches can be offered for skin type classification by uploading the skin analysis results to the computer and using the analysis results. The computer-aided skin type classification system can assist cosmetologists and clients in skin analysis.

Many studies on this subject have been carried out on machine learning [3-12], but machine learning methods in nonlinear situations are inadequate compared to deep learning methods. Deep learning [13] is a subfield of artificial intelligence that imitates the structure used by the human brain to process data and make decisions. The field of deep learning has grown after the creation of big data with digitalization. Deep learning algorithms offer a more realistic model for nonlinear problems that are difficult to solve. In a deep learning model, algorithms can determine whether a prediction is correct or not.

In this study, we show that deep learning architectures is successful in classifying skin analysis data [14-17]. Since artificial neural networks are used in nonlinear situations, it will not be possible to establish a linear logic between the values that their parameters will take. Therefore, in order to determine these values in the best way, we need to find the best values by making comparisons on the data set. With these parameters, an deep learning model can be created that will give the best accuracy. In the study; Two different deep learning architectures, Convolutional Neural Networks (CNN) and long short-term memory (LSTM) were used. For hyper parameter optimization, models were created with Adam, SGD and RMSProp optimization functions, Tanh, ReLU and Sigmoid activation functions, and combinations of 50, 100, 500 epoch numbers. The data set used was obtained from Seda Sakacı Cosmetology Center. As a result of 11,500 skin analyzes created by the center, the pH of the skin, melanin status, temperature, elasticity, hydration, oxidative stress, degree of acne propensity, degree of wrinkle, pore occlusion, sensitivity, shine, size of spots, depth and skin type. After applying data preprocessing techniques such as removing missing areas in the analysis results and eliminating duplicate data, a total of 11,266 data remained. Of the total data, 7887 was divided into training data, 1126 validation data, and 2253 test data. In experimental studies, the performance of the models varies according to the hyper parameter values, and it has been observed that the most successful deep learning model is the model consisting of a combination of CNN architecture, Adam optimization function and Sigmoid activation function and 500 epochs with a success rate of 83.75.

The content of the study, after the introductory part, the examination of the studies made with CNN and LSTM in the literature in the second part, the examination of the theoretical background of the study in the third part, the experimental results and developments in the 4th part, and finally the conclusion part in the 5th part.

II. LITERATURE REVIEW

Studies on skin type prediction discussed in this study in the literature have gained popularity in recent years. Among these studies, those using CNN and LSTM architectures were examined in detail, and it was guided both to guide the study and to determine the differences of the study.

In the study of Alarifi et al., they discuss three types of facial skin patches, namely skin classification techniques that use traditional machine learning and Convolutional Neural Networks to classify normal, blemishes and wrinkles. This study aims to carry out the basic study based on these three classes to provide the collective facial skin quality score. In this study, high-quality facial images of people of different ethnicities were collected to create a volume data set. Next, 100 × 100 resolution skin patches in three preset classes are outlined. With extensive parameter tuning, a series of computer vision experiments have been performed using both traditional machine learning and deep learning techniques for this three-class classification. Despite the limited data set, GoogLeNet surpassed the Support Vector Machine approach with 0.899 Accuracy, 0.852 F-Measure and 0.779 Matthews Correlation Coefficient [18].

In the study of Park et al., Two possible classifications of facial skin type were proposed using simple methods. From 662 healthy volunteers, sebum excretion rate (SER) on the forehead and cheek and skin surface patterns on the cheek, respectively, were examined using Sebutape® and skin replica. SER values measured from the forehead are 0.06-4.56 ng / cm2 / min and SER values measured from the cheek are 0.04-3.80 ng / cm2 / min. Of these data, five hundred skin types were classified according to SER: Low SER type, medium SER type, high SER type, combination-1 SER type, and combination-2 SER type. All twelve skin types are classified as pore size from star formation (SF), primary streaks (PL), secondary streaks (SL), and enlarged skin surface relief (SSR) of the cheek. New classifications of skin types according to SER and SSR have been proposed. It has been observed that SER and skin surface texture parameters (SF, PL and SL) decrease with age and pore size increases with age [19].

In the study of Kumagai et al., Various physiological parameters of the skin were measured in order to develop a scientific method for classifying skin types. 80 women, ages 20 to 59, participated in the one-year study. Skin type self-assessment questionnaires were administered six times to each and they were asked to answer questions about the nature of their skin. Parameters such as skin surface lipid amount, transepideramal water loss (TWL), skin surface morphology, conversion ratio of glutamic acid to pyrrolidone carboxylic acid (% PCA) were measured simultaneously with the self-estimation questionnaires given. It has been found that skin type is subjectively classified according to independent dryness, oilness and sensations. The first was found to correspond to TWL, skin surface morphology and% PCA, and the second to skin surface lipid. While the traditional skin classification procedure mainly depends on the level of skin surface lipids, this study confirmed the existence of other parameters related to dryness [20].

Pham et al. proposed a classification model with a CNN network for skin lesion classification. The proposed classification system is evaluated using the largest public skin lesion test dataset containing 600 tests and 6,162 training data. The proposed model result is archived with AUC (89.2% vs. 87.4%), AP (73.9% vs. 87.4%). 71.5% and ACC (89.0% vs. 87.2%). Additionally, they investigated the effect of each data augmentation on the three classifiers and observed that the performance of each classifier was affected differently by each amplification and had better results compared to traditional methods [21].

Dong has developed a recognition system based on deep education. System CNN and RNN architectures were tested on the dermatology dataset. Based on conducted experiments, user best class call using CNN, an AUROC of 81.6% per patient; this result is roughly 5% higher than seeing a similar comparison across the same number of data [22].

Srinivasu et al's proposed model, based on MobileNet V2 and the LSTM approach, proved efficient for skin classification and detection with minimal computational power and effort. The result is 85.34% accuracy on real-time data from Kaggle compared to other methods [23].

Ahmed et al. proposed a hybrid classification method with CNN and BLSTM for the skin type recognition system. First, they extracted profound features from skin disease facial images. Next, sequential features among the input data are learned using a binary BLSTM network with binary BLSTM by maximum pooling, both the feature matrix and transpose forward and backward term LSTM latent states, dense, fully connected (FC) merge to make an entry into a layer. Finally, classification was made with the siftmax classifier. The proposed method achieved the best average accuracy of 91.73% in skin classification compared to state-of-the-art skin classification methods [24].

Elashiri et al., in their study for skin lesion classification; First, the dataset is collected and preprocessed by contrast enhancement technique via "histogram equalization". After preprocessing, segmentation of images is done by Fuzzy C Means segmentation (FCM). Also, segmented images assigned as input for deep feature extraction using Resnet50, VGG16, and Deeplabv3. The features are obtained from the last layer of these three techniques and combined. These combined features are provided to the feature conversion stage through weighted feature extraction performed by Hybrid Squirrel Butterfly Search Optimization (HSBSO). The converted features are exported to Modified Long Short Term Memory (MLSTM) where architectural optimization is done by the same HSBSO to produce the final classified output. Through performance analysis, the proposed HSBSO provided 26.8% better security than DNN, 18.8% better than CNN, 4.77% better than SVM, and 56.33% higher than LSTM in classification performance under dataset 2. Therefore, it was concluded that the proposed skin disease classification model using the proposed HSBSO and the development of the designed MLSTM outperform conventional skin disease classification models [25].

Skin type determination studies in the literature are based on image processing and statistical information. In our study, different from the studies in the literature, classification with deep learning was made using the data set obtained from experimental studies.

III. THEORIC BACKGROUND

A. DEEP LEARNING

Deep learning (DL) is a new machine learning method derived from artificial neural networks that can automatically extract features from main data [26,27]. In most cases, unlike classical machine learning methods, it does not require data preprocessing [28]. Instead, DL learns a combination of lower-level features and characteristic hierarchies with higher hierarchical features [29]. Thus, DL is successfully used in solving complex, high dimensional problems. There are many deep learning algorithms [30]: CNN [31], LSTM [32], Gated Recurrent Unit Neural Networks (GRU) [33], Recurrent Neural Networks (RNN) [34].

B. CONVOLUTIONAL NEURAL NETWORK

CNN network [35] is a deep learning architecture whose idea was put forward by Yann LeCun in 1988, whose improvements continued until 1998 and whose first name was LeNet. In the CNN architecture, the first layers consist of cascading convolution and maximum pooling layers. The next layers correspond to fully connected traditional multi-layer networks.

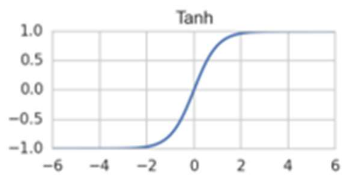
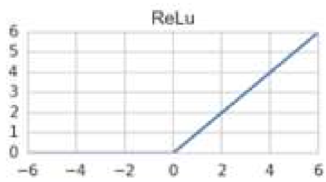
C. LONG SHORT TERM MEMORY NETWORK

LSTM networks were developed by Hochreiter and Schmidhuber in 1997.has been introduced [36]. LSTM architecture has 3 gates as input, forget and output, block input, Hard Fault Loop, output activation function and surveillance connections. The output of the block is repeatedly connected to the input of the block and all its gates. Surveillance connections and forget gate are not present in the first developed architecture. A forget gate has been added to reset the LSTM's own state, and watch connections have been added to make it easier to learn the exact timings.

D. ACTIVATION FUNCTION

Activation functions: It is used to transmit the output value of neurons in one layer to the next layers. The threshold value needs to be determined to decide whether this output value is to be transmitted to other layers. Because the value of the information in an artificial neural cell may be in the range (−∞, +∞) and the neuron may not know the limits of the true value. Therefore, activation functions are needed to decide whether the neuron should be active or not. Thus, it will be able to control the output value produced by a neuron and decide whether external connections will actively see the neuron. Since artificial neural networks are mostly used in nonlinear classifications, the activation function is usually chosen as a nonlinear function. While the back propagation algorithm is used in the learning process of the architecture, it is important for the speed of the algorithm to use an activation function whose derivative can be easily calculated since the derivative of the activation function is also used [37]. Table 1 shows the graphs of the activation functions used in the study.

Table 1. *Activation functions and properties used in the study*

Activation Function	Graphic	Explanation
Tanh		Tanh function produces values between -1 and +1 for each of the input values [38].
ReLU		In this function, the output is zero when the input value is below zero, but if the input value is above zero, the output is equal to the input value and a linear relationship occurs with the dependent variable [39].
Sigmoid		The sigmoid function is defined for y values between 0 and 1, and it converges to 0 before the x=0 axis and then to 1 [40].

E. OPTIMIZATION ALGORITHMS

In deep learning applications, the absolute minimum value of the error function must be found in order to achieve the best result of the learning process. This process is carried out using optimization algorithms. Optimization is the method used to minimize the error, that is, the difference between the output value produced by the network and the actual value. Gradient descent is one of the most used methods for optimization of artificial neural networks. Various algorithms based on the gradient descent method are RMSProp, Adagrad, Adam, Nadam [41]. Table 2 shows the properties of the optimization algorithms used in the study.

***Table 2**. Optimization algorithms and properties used in the study*

Optimization Algorithm	Explanation
Adam	Adam [42] is another method that computes adaptive learning rates for each parameter.
SGD	SGD [43] in contrast performs a parameter update for each training example x(i) and label y(i).
RMSProp	RMSprop [44] is divides the learning rate by an exponentially decaying average of squared gradients. Hinton suggests γ to be set to 0.9, while a good default value for the learning rate η is 0.001.

F. DEEP LEARNING LIBRARIES

There are many ready-made libraries and APIs (Application Programming Interface) with different features developed by various universities and companies to make machine learning and deep learning practical and easy. It contains many libraries in its deep learning structure. Libraries suitable for the subject to be studied must be installed on the computer. Each of these libraries has different functions. There are multiple deep learning libraries available in the Python programming language. Libraries used in this study are Keras and TensorFlow [45,46]. Some of the deep learning libraries that can be used with the Python programming language are listed in Table 3.

***Table 3**. Deep Learning Libraries for Python*

Library Name	Developer	Using Area
Theano	MILA Lab	It is a Python library that enables defining, optimizing and evaluating multidimensional arrays and math expressions [47].
TensorFlow	Google	It enables efficient numerical calculations with data flow graphics [48].
Keras	Google	The Keras library is a Python library written on Tensorflow and Theano as a top layer, enabling easier model development. It works with CPU and GPU and supports CNN and RNN combinations [49].
Mxnet	Amazon	It is a high-level library, since it is a polyglot, it offers solutions for teams sharing models in different languages. Another advantage is that it supports distributed computing [50].

G. DATASET

In this section, the creation phase of the data set used in the study, the data preprocessing processes, the dependent variables in the data set and the values and definitions of the independent variable are explained.

Figure 1. *Performance and loss graph of the most successful artificial neural network*

The data set used in the study was taken from Seda Sakacı Cosmetology Center. Cosmetological skin analysis is performed by the center for customers. The device seen in Figure 1 in cosmetological skin analysis makes measurements with very sensitive probes. Measurements are as follows: skin pH, melanin status, temperature, elasticity, hydration, oxidative stress, acne proneness, wrinkle degree, pore occlusion, sensitivity, brightness, size and depth of spots are measured. After the skin analysis, the person's cosmetology identity, that is, the skin type, is revealed. 11,500 skin analyzes measured by the centre; There are 13 independent variables as skin pH, melanin status, temperature, elasticity, hydration, oxidative stress, acne proneness, wrinkle degree, pore clogging, tenderness, brightness, blemish size, and depth. The dependent variable is skin type. In Table 4, the explanations and values of the variables belonging to the data set are given. 11,500 analyzes were subjected to data preprocessing. First of all, data with missing fields and repetitive data were removed. Then the text data was digitized. Thus, our data set was formed with 11,266 analyzes remaining from 11,500 analyzes. Then 70% of the dataset for training, 10% for validation and the remaining 20% is reserved as test data.

Table 4*. Values that data set variables can take and descriptions of the variables*

Variable Name	Values	Description
Skin pH	0-14	PH value means the alkaline or acidic value of the skin.
Melanin status	0-100 mm^2	It is the amount of melanin, the pigment that gives the skin its color.
Elasticity	0-10	Elasticity is the skin's ability to stretch. In other words, it is the youth state of the skin.
Hydration	0-10	Hydration is the moisture state of the skin.

Oxidative stress	0-62	Oxidative stress is a condition in which the skin looks aged.
Acne pronesess	0-3	The tendency to acne is the sebum quality of the skin.
Wrinkle degree	0-10	Wrinkle degree is the amount of wrinkles on the skin.
Pore clogging	0-1	It is the clogging of the pores on the skin due to too much sebum or environmental factors.
Tenderness	0 (none) 1 (tension) 2 (abnormal tingling) 3 (burning) 4 (stinging) 5 (pain) 6 (itching)	Sensitivity is a condition in which the skin reacts faster than it would to the same attacks under similar conditions.
Brightness	0-3	Brightness is the condition of the skin looking healthy.
Blemish size	0-100 mm^2	It is the size of the spots on the skin.
Depth	0-100 mm^2	It is the depth of the spots on the skin.
Skin type	0 (normal) 1 (dry) 2 (oily) 3 (sensitive) 4 (complicated)	It is the skin type that comes out as a result of the measurements.

IV. EXPERIMENTAL RESULT AND DISCUSSION

It has been widely tested that deep neural networks work faster. For this, it has been seen that the use of GPUs instead of CPUs affects the performance and training time of the network [50-55]. In this study, all experiments were evaluated on GPUs in Colab and using Keras deep learning library. Hyperparameter optimizations are critical both for the performance of the network and for benchmarking studies. The evaluation metric used in this study is the accuracy of the network.

The results obtained from the experiments we performed on the skin analysis data set with CNN and LSTM architecture using Adam, SGD, RMSProp optimization algorithms, Tanh, ReLU, Sigmoid activation functions and 50, 100, 500 epochs are given in Table 5. After the layers are created and the network structure is completed, the training options of the created network must be specified. The most used training options; size of data set, mini-batch size, learning speed and momentum coefficient, optimization algorithm selection, epoch number, weight and activation function. Softmax activation function is used in the output layer of each model created. Learning rate 0.001 was used by default values of optimization algorithms.

***Table 5**. Deep learning algorithms training accuracy result*

			Epoch(%)		
Deep Learning Algorithm	**Activation Function**	**Optimization Algorithm**	**50**	**100**	**500**
CNN	Tanh	Adam	61.44	65.57	80.21
		SGD	72.19	78.21	80.64
		RMSProp	62.25	68.96	79.56
	ReLU	Adam	62.92	68.25	72.51
		SGD	66.52	66.76	78.16
		RMSProp	66.11	72.78	78.52
	Sigmoid	Adam	69.29	73.96	83.75
		SGD	67.6	71.22	82.89
		RMSProp	67.3	70.92	80.1
LSTM	Tanh	Adam	47.01	60.69	67
		SGD	43.45	51.32	52.26
		RMSProp	43.10	50.3	65.22
	ReLU	Adam	43	52.14	68.9
		SGD	44.43	52.5	56.7
		RMSProp	44	51	66
	Sigmoid	Adam	40.3	41.5	45.98
		SGD	41.32	44.74	45.63
		RMSProp	43.10	50.3	65.22

The performance of the models was evaluated according to the accuracy of the test data. It has been seen that the most successful model is CNN network, Adam optimization algorithm, Sigmoid activation function and 500 epoch values. The best model created completed the training in 300 seconds, and it was determined that it achieved 83.75% accuracy in training data, 83.3% accuracy in test data, and the test took 2 seconds. Accuracy and loss graph of the deep learning model with the highest success rate in the study is shown in Figure 2.

Figure 2. *Performance and loss graph of the most successful artificial neural network*

V. CONCLUSION

Deep neural networks have proven their functionality in many areas and, thanks to their previous great success, are still being experimented with on a wide variety of topics. Since there are extremely popular architectures for the implementation of deep neural networks, which one provides the best performance is a question to be explored. For this purpose, in this study, CNN and LSTM architectures were compared with different hyperparameters for the classification of skin analysis data. Thus, it is aimed to find the most successful model for classification. The accuracy evaluation metric is used for performance comparison of deep neural network architectures. According to the experimental result, models created with CNN were found to be a better choice compared to LSTM in terms of both accuracy and time. The CNN network Adam was the most successful model with sigmoid and 500 epoch parameters, with an accuracy of 83.75%.

As a future study, it is aimed to make a comparison study again to make skin analysis from human pictures and to select the most suitable model for skin analysis. In addition, a more comprehensive comparison can be made for this study and other deep neural network architectures and other deep neural network platforms can be included for experiments.

V. REFERENCES

[1] Seda S. “Kozmetolojik Cilt Analizi.” ScientificAnalysis. http://www.sedasakaci.com/tr/ScientificAnalysis (Access Date: 12.04.2021).

[2] M. H. Ravnbak, “Objective determination of Fitzpatrick skin type.,” *Danish Medical Bulletin*, vol. 57, no. 8, p. B4153, Aug. 2010.

[3] S. Kolkur and D. Kalbande, “Survey of texture based feature extraction for skin disease detection,” *2016 International Conference on ICT in Business Industry &Amp; Government (ICTBIG)*, Jan. 2016.

[4] Kolkur, M. S., Kalbande, D. R. and Kharkar, V. “Machine learning approaches to multi-class human skin disease detection.,” *International Journal of Computational Intelligence Research*, 14(1), 1-12, Jan 2018.

[5] P. R. Hegde, M. M. Shenoy, and B. H. Shekar, “Comparison of Machine Learning Algorithms for Skin Disease Classification Using Color and Texture Features,” *Advances in Computing and Communications*, Sep. 2018.

[6] S. Bhadula, S. Sharma, P. Juyal, and C. Kulshrestha, “Machine Learning Algorithms based Skin Disease Detection,” *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 2, pp. 4044–4049, Dec. 2019.

[7] V. Chaurasia and S. Pal, “Skin Diseases Prediction: Binary Classification Machine Learning and Multi Model Ensemble Techniques,” *Research Journal of Pharmacy and Technology*, vol. 12, no. 8, p. 3829, Aug. 2019.

[8] V. Chaurasia, “Machine learning algorithms using binary classification and multi model ensemble techniques for skin diseases prediction,” *International Journal of Biomedical Engineering and Technology*, vol. 34, no. 1, p. 57, Jan. 2020.

[9] S. Chan, V. Reddy, B. Myers, Q. Thibodeaux, N. Brownstone, and W. Liao, “Machine Learning in Dermatology: Current Applications, Opportunities, and Limitations,” *Dermatology and Therapy*, vol. 10, no. 3, pp. 365–386, Apr. 2020.

[10] A. X. Du, S. Emam, and R. Gniadecki, “Review of Machine Learning in Predicting Dermatological Outcomes,” *Frontiers in Medicine*, vol. 7, Jun. 2020.

[11] A. Hekler *et al.*, “Superior skin cancer classification by the combination of human and artificial intelligence,” *European Journal of Cancer*, vol. 120, pp. 114–121, Oct. 2019.

[12] A. S. Adamson and A. Smith, “Machine Learning and Health Care Disparities in Dermatology,” *JAMA Dermatology*, vol. 154, no. 11, p. 1247, Nov. 2018.

[13] Y. LeCun, Y. Bengio, and G. E. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[14] Patnaik, S. K., Sidhu, M. S., Gehlot, Y., Sharma, B., & Muthu, P. “Automated skin disease identification using deep learning algorithm,” *Biomedical & Pharmacology Journal*, 2018.

[15] B. Zhang *et al.*, “Opportunities and Challenges: Classification of Skin Disease Based on Deep Learning,” *Chinese Journal of Mechanical Engineering*, vol. 34, no. 1, Nov. 2021.

[16] A. Verma, S. Pal, and S. Kumar, “Classification of Skin Disease using Ensemble Data Mining Techniques,” *Asian Pacific Journal of Cancer Prevention*, vol. 20, no. 6, pp. 1887–1894, Jun. 2019.

[17] K. Mahajan, M. Sharma, and L. Vig, “Meta-DermDiagnosis: Few-Shot Skin Disease Identification using Meta-Learning,” *Computer Vision and Pattern Recognition*, Jun. 2020.

[18] J. S. Alarifi, M. S. Goyal, A. K. Davison, D. Dancey, R. Khan, and M. H. Yap, “Facial Skin Classification Using Convolutional Neural Networks,” *Lecture Notes in Computer Science*, pp. 479–485, Jul. 2017.

[19] S.-H. Park, Y. H. Kim, J. Kim, and S.-H. Kang, “Two possible classifications of facial skin type by two parameters in Korean women: sebum excretion rate (SER) and skin surface relief (SSR),” *Skin Research and Technology*, vol. 5, no. 3, pp. 189–194, Aug. 1999.

[20] H. Kumagai *et al.*, “Development of a scientific method for classification of facial skin types.,” *Journal of S C C J*, vol. 19, no. 1, pp. 9–19, Sep. 1985.

[21] T.-C. Pham, C. M. Luong, M. Visani, and V.-D. Hoang, “Deep CNN and Data Augmentation for Skin Lesion Classification,” *Lecture Notes in Computer Science*, pp. 573–582, Mar. 2018.

[22] Dong, H., & Farid, F. “A Deep learning based patient care application for skin cancer detection,” *https://www.researchsquare.com/article/rs-1582255/v1*, 2022.

[23] P. N. Srinivasu, J. G. SivaSai, M. Ijaz, A. K. Bhoi, W. Kim, and J. D. Kang, “Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM,” *Sensors*, vol. 21, no. 8, p. 2852, Apr. 2021.

[24] B. Ahmad, M. Usama, T. Ahmad, S. Khatoon, and C. M. Alam, “An ensemble model of convolution and recurrent neural network for skin disease classification,” *International Journal of Imaging Systems and Technology*, vol. 32, no. 1, pp. 218–229, Oct. 2021.

[25] M. A. Elashiri, A. Rajesh, S. N. Pandey, S. K. Shukla, S. Urooj, and A. Lay-Ekuakille, “Ensemble of weighted deep concatenated features for the skin disease classification model using modified long short term memory,” *Biomedical Signal Processing and Control*, vol. 76, p. 103729, Jul. 2022.

[26] N. Rusk, “Deep learning,” *Nature Methods*, vol. 13, no. 1, p. 35, Dec. 2015.

[27] G. M. Farinella *et al.*, “Deep Learning for Health Informatics,” *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 4–21, Jan. 2017.

[28] Y. Li, X. Nie, and R. Huang, “Web spam classification method based on deep belief networks,” *Expert Systems With Applications*, vol. 96, pp. 261–270, Apr. 2018.

[29] Y. Bengio, “Learning Deep Architectures for AI,” *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, Jan. 2009.

[30] A. K. Shrestha and A. Mahmood, “Review of Deep Learning Algorithms and Architectures,” *IEEE Access*, vol. 7, pp. 53040–53065, Apr. 2019.

[31] H. P. Phan, F. Andreotti, N. Cooray, O. Chen, and M. De Vos, “Joint Classification and Prediction CNN Framework for Automatic Sleep Stage Classification,” *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 5, pp. 1285–1296, May 2019.

[32] K. Cho *et al.*, "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," *arXiv (Cornell University)*, Jan. 2014.

[33] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv (Cornell University)*, Dec. 2014.

[34] R. Socher, C. J.-F. Lin, C. Manning, and A. Y. Ng, "Parsing Natural Scenes and Natural Language with Recursive Neural Networks," *International Conference on Machine Learning*, pp. 129–136, Jun. 2011.

[35] Y. LeCun *et al.*, "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, Dec. 1989.

[36] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[37] A. V. Savchenko, "Probabilistic Neural Network With Complex Exponential Activation Functions in Image Recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 2, pp. 651–660, Feb. 2020.

[38] F. Godin, J. Degrave, J. Dambre, and W. De Neve, "Dual Rectified Linear Units (DReLUs): A replacement for tanh activation functions in Quasi-Recurrent Neural Networks," *Pattern Recognition Letters*, vol. 116, pp. 8–14, Dec. 2018.

[39] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," *International Conference on Machine Learning*, pp. 807–814, Jun. 2010.

[40] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals, and Systems*, vol. 2, no. 4, pp. 303–314, Dec. 1989.

[41] E. M. Dogo, O. Afolabi, N. Nwulu, B. Twala, and C. Aigbavboa, "A Comparative Analysis of Gradient Descent-Based Optimization Algorithms on Convolutional Neural Networks," *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, Dec. 2018.

[42] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv (Cornell University)*, Jan. 2015.

[43] Y. LeCun, L. Bottou, G. Orr, and K.-R. Müller, "Efficient BackProp," *Lecture Notes in Computer Science*, pp. 9–50, Jan. 1998.

[44] D. S. Yook, H. W. Lee and I. C. Yoo, "A survey on parallel training algorithms for deep neural networks," *The Journal of The Acoustical Society of Korea*, vol. 39, no, 6, pp. 505-514, Nov. 2020.

[45] D. Quang, Y. Guan, and S. C. Parker, "YAMDA: thousandfold speedup of EM-based motif discovery using deep learning libraries and GPU," *Bioinformatics*, vol. 34, no. 20, pp. 3578–3580, Oct. 2018.

[46] B. A. Erickson, P. Korfiatis, Z. Akkus, T. L. Kline, and K. A. Philbrick, "Toolkits and Libraries for Deep Learning," *Journal of Digital Imaging*, vol. 30, no. 4, pp. 400–405, Aug. 2017.

[47] A. Parvat, J. Chavan, S. Kadam, S. Dev, and V. Pathak, "A survey of deep-learning frameworks," *2017 International Conference on Inventive Systems and Control (ICISC)*, Jan. 2017.

[48] E. J. Choi, L. Junyeong, and I. Han, "Deriving adoption strategies of deep learning open source framework through case studies," *Journal of Intelligence and Information Systems*, vol. 26, no. 4, pp. 27–65, Jan. 2020.

[49] B. A. Erickson, P. Korfiatis, Z. Akkus, T. L. Kline, and K. A. Philbrick, "Toolkits and Libraries for Deep Learning," *Journal of Digital Imaging*, vol. 30, no. 4, pp. 400–405, Aug. 2017.

[50] Z. Wang, M. Yan, J. Chen, S. Liu, and D. Zhang, "Deep learning library testing via effective model generation," *Foundations of Software Engineering*, Nov. 2020.

[51] Y. Wang, G.-Y. Wei, and D. J. Brooks, "Benchmarking TPU, GPU, and CPU Platforms for Deep Learning," *arXiv (Cornell University)*, Jul. 2019.

[52] T. Chen *et al.*, "TVM: an automated end-to-end optimizing compiler for deep learning," *Operating Systems Design and Implementation*, pp. 578–594, Oct. 2018.

[53] E. Buber and B. Diri, "Performance Analysis and CPU vs GPU Comparison for Deep Learning," *International Conference on Control, Engineering & Information Technology*, Oct. 2018.

[54] W. Dai and D. Berleant, "Benchmarking Contemporary Deep Learning Hardware and Frameworks: A Survey of Qualitative Metrics," *HAL (Le Centre Pour La Communication Scientifique Directe)*, Dec. 2019.

[55] S. I. Baykal, D. Bulut, and O. K. Sahingoz, "Comparing deep learning performance on BigData by using CPUs and GPUs," *Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, Apr. 2018.