■ RESEARCH ARTICLE

AITA
ARTIFICIAL INTELLIGENCE THEORY AND APPLICATIONS

Detection of Benign and Malignant Skin Cancer from Dermoscopic Images using Modified Deep Residual Learning Model

Mustafa Said Kartal [a], **Özlem Polat** [b†]

[a] Faculty of Medicine, Sivas Cumhuriyet University, Sivas, Turkey
[b] Department of Mechatronics Engineering, Sivas Cumhuriyet University, Sivas, Turkey
[†] ozlem.polat@cumhuriyet.edu.tr, corresponding author

RECEIVED JUNE 22, 2022
ACCEPTED SEPTEMBER 14, 2022

CITATION Kartal, M. S., Polat, Ö. (2022). Detection of Benign and Malignant Skin Cancer from Dermoscopic Images using Modified Deep Residual Learning Model, *Artificial Intelligence Theory and Applications, 2*(2), p. 10-18.

Abstract

Skin cancer is caused by the uncontrolled proliferation of cells on the skin surface due to damaged DNA structures of them and is among the most common cancer types in the world. If malignant skin cancer is not detected early, it can result in death. For this reason, early and high accuracy detection of skin cancer is important in terms of increasing the chance of survival of patients. In this study, ResNet101 architecture, which is one of the deep residual learning models, is suggested for the detection of malignant skin cancer from dermoscopy images. ResNet is a short name for a residual network. The model was trained and tested on a dataset of 3297 dermoscopic images from the ISIC 2017 archive. This archive was published by the International Skin Imaging Collaboration (ISIC) as a dataset of dermoscopy images. As a result of 10 experiments, average 90,67% and maximum 91,36% accuracy values were reached. In this study, a better performance was obtained compared to previous studies using the same dataset in the literature. In conclusion, the proposed approach has promise in the field of medicine and can help dermatologists diagnose skin cancer.

Keywords: skin cancer; classification; deep residual learning; pattern recognition; machine learning

1. Introduction

Skin cancers, also called tumoral skin lesions, occur due to abnormal proliferation of cells on the skin. One of the most important reasons for this is that the skin is more exposed to ultraviolet rays due to the thinning of the ozone layer [1]. Squamous cell carcinoma, basal cell carcinoma, and melanoma are a few types of skin cancer. The most common of these is basal cell carcinoma, followed by squamous cell carcinoma. The most dangerous skin cancer is melanoma, although it is not very common, its lethality is higher than other cancer types [2].

It is difficult to detect skin cancers because of differences in skin textures [2]. Clinically, skin cancers are detected using the ABCDE technique. Here, A, B, C, D and E denote the asymmetry, border, color, diameter and enlargement of the lesion, respectively. With

Artificial Intelligence Theory and Applications, ISSN: 2757-9778. ISBN : 978-605-69730-2-4 © 2022 University of Bakırçay

this technique, the diagnosis is made using the naked eye. However, since most skin cancers look alike, they are more likely to make mistakes. Another method is the surgical removal of the tissue, also called biopsy, and its pathological examination. With this method, definitive diagnosis is made, but this is also a time-consuming process [3]. For this reason, dermatologists often use the dermoscopy technique to detect skin lesions. With this technique, a magnified image is obtained by using a special magnification tool in order to examine the lesion area on the skin in more detail. These images are examined manually by dermatologists to determine whether there is cancer or not. The accuracy of the detection depends on the dermatologist's experience [4]. Since this application has a high risk of error and is time consuming, studies have been started with machine learning techniques in recent years to help experts in detecting skin cancer faster and with higher accuracy. In particular, deep learning techniques have been used quite frequently in the field of medicine in recent years, as their success in various fields has been proven.

There are many studies in the literature on the detection of skin cancer, the multiple classification of cancer types [3,5,6], and also the binary classification as benign and malignant [7-16]. Various datasets have been brought to the literature to be used in these studies. The most widely used dataset is [17] compiled from the ISIC archive. ISIC is an organization working with academia and industry to facilitate the application of digital skin imaging to help reduce melanoma mortality. Nadipineni [3] used deep learning models to classify eight types of skin cancer identified in the ISIC 2019 archive and achieved a maximum accuracy of 88,6%. A similar study was carried out by Kassem et al. [5], and a success of 94,92% was achieved for eight classes with the GoogleNet model. Sun et al. [6] left one class out of the ISIC 2019 dataset and classified seven types of skin cancer with their EfficientNet models with 89,5% success.

Skin cancer types can be grouped as benign and malignant. Melanoma is one of the malignant cancers. If melanoma is detected early, the chance of cancer spreading to surrounding tissues and resulting in death is reduced. For this reason, studies in the literature have focused on determining whether skin cancer is malignant or not. Among them, Nusraddinov [7] trained the InceptionV3 transfer learning model for binary classification on 2750 dermoscopic images taken from the ISIC 2017 archive and achieved 70% success as a result of the tests he made. There are also studies that apply different deep learning models in the detection of benign and malignant cancers using ISIC archive images [8-10].

In the literature, there are also studies [11-13] using the HAM10000 dataset [18] compiled from the ISIC 2018 archive. In this dataset, there are a total of 10015 dermoscopic images, of which 1954 are malignant and 8061 are benign. Ergun and Kılıç [11] used AlexNet, DenseNet12, ResNet18, ResNet34, SqueezeNet and VGG16 models for classification. In their tests with augmented data, they obtained the best performance from the ResNet34 model with an accuracy of 87,5%. Öztürk and İçer [12] obtained an accuracy value of 77,8% on the preprocessed dataset using feedforward neural networks. Ameri [13] using the Alexnet model has achieved 84% classification performance on the HAM10000 dataset.

Another dataset consisting of 3297 dermoscopic images compiled from the ISIC archive was used in the studies in [14-16]. Demir [14] obtained an accuracy of 88,35% by using MobileNetv2 and Support Vector Machines for feature extraction and classification, respectively. Farooq et al. [15] preprocessed on the same dataset and increased the performance of classifying benign and malignant skin cancer to 86% with the Inception model. In the study conducted by Soylu and Demir in 2021 [16], the performance reached 89,89% by using the Darknet-19 model.

In this study, the detection of benign and malignant skin cancer from dermoscopic images was made with the modified ResNet101 model from deep residual networks. The proposed model is trained and tested on the dataset used in [14-16] to evaluate its performance. The study has a novelty in terms of using the ResNet101 model modified by adding different layers to the end. The contribution of the study to the literature can be explained as follows: Benign and malignant skin cancers were classified with a maximum accuracy of 91.36%, and this result is better than the study performances using the same dataset in the literature. The study shows promise in helping experts decide on the classification of skin cancers.

2. Materials and Methods

2.1. Dataset

In this study, a two-class Kaggle skin cancer dataset [17] consisting of 3297 dermoscopic images compiled from the ISIC archive was used as a dataset. The dataset contains 1800 and 1497 dermoscopic images, benign and malignant, respectively. Each image is in jpeg format and has a pixel size of 224×224. In addition, the dataset is presented on Kaggle by creating training and test sets at a rate of 80% and 20%, respectively, and, used in this study as presented in order to make a meaningful comparison with previous studies using the same dataset. Sample dermoscopic images of benign and malignant skin cancer in the dataset are shown in Figure 1.

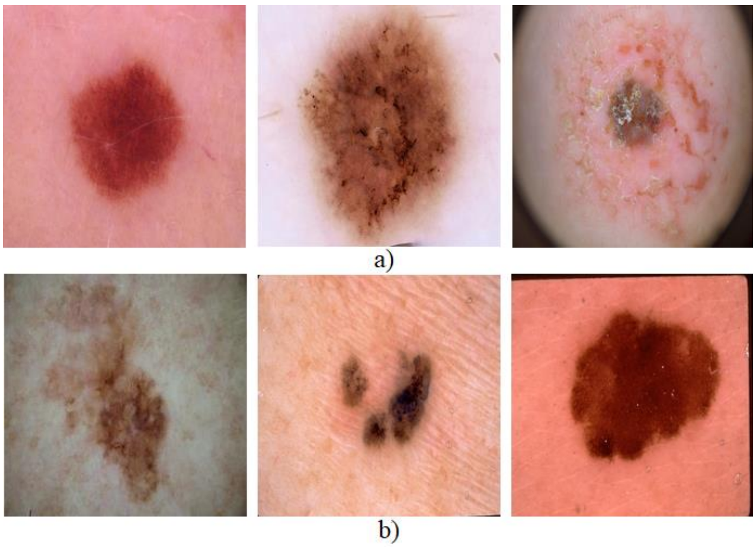

Figure 1. Sample dermoscopy images of both classes in the Kaggle skin cancer dataset compiled from the ISIC archive a) benign b) malignant.

2.2. Deep Learning

Deep learning is one of the machine learning techniques and has been used frequently in recent years due to its high performance in various fields. Convolutional neural networks (CNN), a kind of neural network, are used in deep learning. The concept of convolution was first defined by LeCun et al. [20]. The two-dimensional convolution operation consists of multiplying the pixel values of two matrices with the same indices and then adding these products [21]. Networks that use the convolution process are called CNNs. CNNs are frequently used in medicine and many fields such as image classification and segmentation, audio, text and video processing. CNNs consist of several layers as briefly outlined below:

- Input layer: In this layer, data is given as input to the network.
- Convolution layer: It is the layer where the filter matrix and the image matrix are convoluted and the feature map is obtained at the output.
- Rectified linear unit (ReLU) layer: It is the layer that transforms the linear output of the previous layer into a non-linear structure using an activation function.
- Pooling layer: It is the layer used to reduce the size of the feature map given as input to the next layer.
- Flattening layer: It is the layer where the two-dimensional feature map is converted to a one-dimensional vector.
- Fully connected layer: It is the layer where the classes of the images are determined by labeling them.

2.3. Modified ResNet101 Architecture

Deep Residual Networks, or ResNet for short, is an improved form of CNN. The ResNet approach seeks to address CNN networks' degradation issue. The degradation issue appears as deep neural networks start to converge. The efficiency hits saturation as the network's layer count rises, but after that it tends to decrease rapidly [22]. ResNet adds short-cut links between layers to overcome this issue. It can be said that ResNet employs the jump link (short-cut link) concept. Short-cut links skip one or more layers. The convolution layers are cascaded but at the same time add the original input to the output of the convolution block. Short-cut links do not add extra parameters or computational complexity to the network [22]. However, bottleneck blocks are used to speed up training in ResNet architecture. ResNet101 is a network of 101 layers trained on the ImageNet dataset; and it uses convolution layers as Conv(1×1), Conv(3×3) and Conv(1×1).

In this study, the ResNet101 model was modified and used in order to increase the performance. The weights in the first eight layers of the 101-layer model are frozen and the remaining layers are arranged as trainable. In addition, two convolution layers consisting of 64 filters with the size of 3×3 were added to the end of the model, followed by a pooling layer of 2×2 and two fully connected layers consisting of 4096 nodes. Two-node Softmax layer is added to the end of the model for classification purposes. Modifications to the ResNet101 model were made by trial and error in order to obtain high-level features. After various trials, a modified model with better classification performance than the original Resnet101 model was obtained. The block diagram of the model used is shown in Figure 2. While the layers on the left of Figure 2 show the original ResNet101 architecture, the layers added to the end of the model and shown on the right represent the change made in the model. Also, the pseudocode of the training, testing and evaluation stages of the modified ResNet101odel is shown in Figure 3.

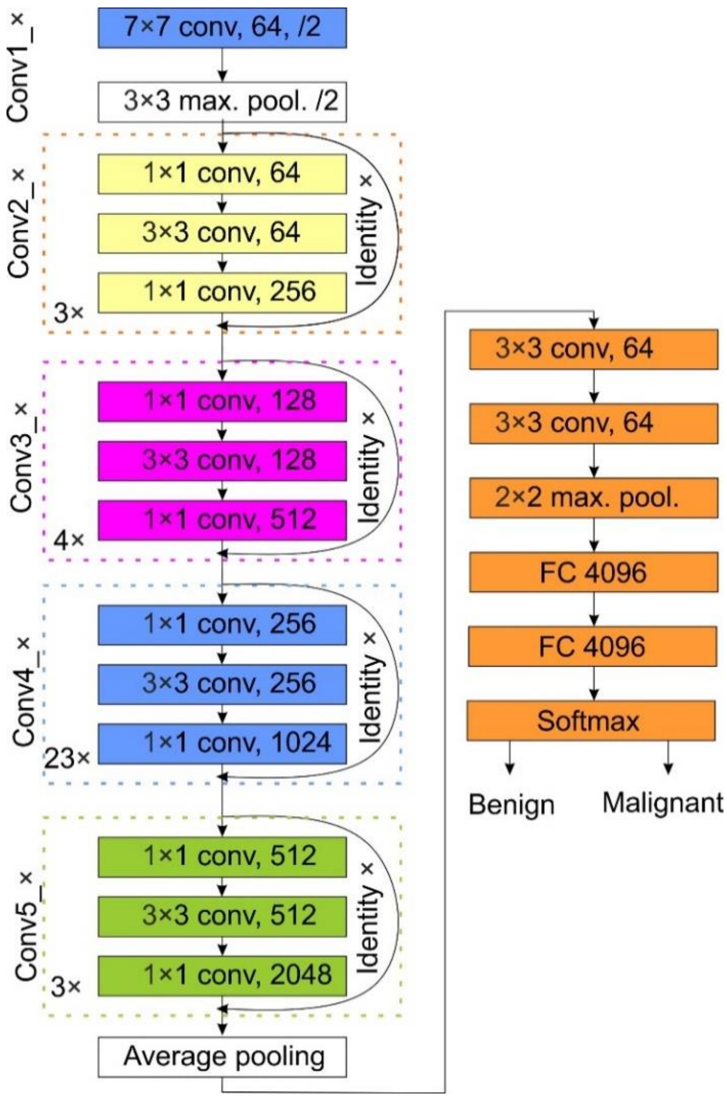

Figure 2. Modified ResNet101 architecture.

Pseudocode for training stage	Pseudocode for test & evaluation stage
Input: - Dermoscopic train images and labels (train_x, train_y)	Input: - Dermoscopic test images and labels (test_x, test_y) - Model: model from training stage
1. **Resize** images (224x224) 2. **Define** deep learning model (ResNet101) 3. **Modify** the model by adding model.add(Conv2D(64, (3x3))) model.add(Conv2D(64, (3x3))) model.add(MaxPooling2D(2x2)) model.add(Dense(4096)) model.add(Dense(4096)) 3. **Set** the hyperparameters Batch size: 32 Epoch: 50 Optimizer: Adam Learning Rate: 0.0001 4. **Train** the model model.fit(train_x, train_y)	1. **Load** the model 2. **for** i =1 : 10 model.evaluate(test_x, test_y) pred_y=model.predict(test_x) acc = accuracy_score(test_y, pred_y) **Calculate** sensitivity, precision, f1-score **end** 3. **Calculate** average accuracy
Output: model	Output: accuracy, sensitivity, precision f1-score

Figure 3. Pseudocode for training and test stages of the modified ResNet101 model

3. Experimental Results and Discussion

In this study, to classify benign and malignant skin cancer, ResNet-based models (ResNet50, ResNet50V2, ResNet101, ResNet101V2, ResNet152, ResNet152V2) were trained and tested with 2637 and 660 dermoscopic images, respectively. The focus is on this model, as the best results were obtained with ResNet101. The experimental results were compared with the actual results determined by the experts in terms of accuracy, sensitivity, precision and f1-score. Equations for the metrics used to evaluate the performance of the model are shown in Equations 1, 2, 3 and 4.

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

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

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

$$F1 - score = 2 \times \frac{Sensitivity \ \times Precision}{Sensitivity + Precision} \tag{4}$$

Here, TP, TN, FP and FN are used instead of true positive, true negative, false positive and false negative, respectively. The accuracy value is widely used in the evaluation of classification performance. However, other metrics have also been calculated because they are of great importance in medical applications. In the training of ResNet models, ReLU as activation function, Adam as optimization algorithm with a learning rate of 0.0001 were used due to their success in the applications in the literature. The batch size was determined as 32 and the epoch value was used as 50, since there was not much change in the accuracy value for more epoch values. A computer equipped with an AMD Ryzen 9 5900HX processor and a 16 GB Nvidia GeForce RTX3080 graphics card with 32 GB DDRAM was used to perform the experiments. The proposed models are trained and tested in Python using Keras and Tensorflow libraries. Experiments for the evaluation of the models were repeated 10 times, the results of the experiments are shown in Table 1. Accordingly, the best performance was obtained with the ResNet101 model as an average of 90,67% and a maximum of 91,36%. The confusion matrix for the best performing experiment is shown in Figure 4.

Initially, the original ResNet101 model was used and 88,94% classification success was achieved. In order to increase the performance, the ResNet101 model was modified and according to the results of the experiment, it was seen that the performance increased to 91,36%.

Table 1. Performance results of ResNet models.

Residual Learning Models	Maximum Accuracy (%)	Average Accuracy (%)
ResNet50	89,70	87,35
ResNet0V2	86,67	86,00
ResNet101	**91,36**	**90,67**
ResNet101V2	87,58	86,90
ResNet152	88,79	87,71

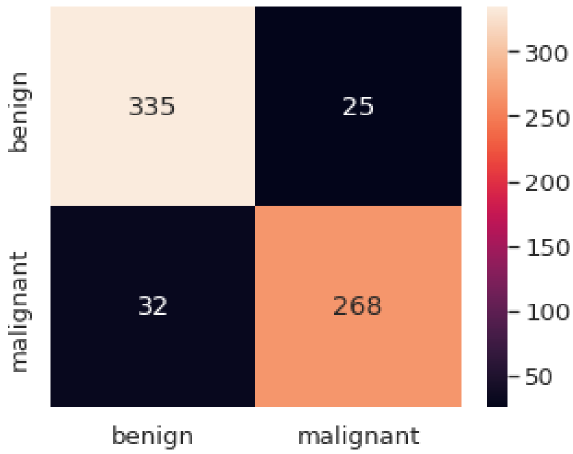

Figure 4. Confusion matrix related to modified ResNet101 model. The rows show the actual classes and the columns show the classes predicted by the model.

The results of the modified ResNet101 model used in the classification of benign and malignant skin cancer in the study were compared with the results of other studies using the same dataset in the literature. The comparison result is presented in Table 2. Looking at the Table 2, it is seen that other studies also use deep transfer learning models; and the proposed model has a superior classification performance than other models.

Table 2. Comparison table related to studies using the same dataset.

Method	**Accuracy**	**Sensitivity**	**Precision**	**F1-score**
MobileNetV2+SVM [14]	88,35	0,88	0,90	0,87
InceptionV3 [15]	86,00	0,89	0,80	0,84
DarkNet19 [16]	89,89	n.i.	n.i.	n.i.
Modified ResNet101	91,36	0,91	0,91	0,91

n.i.: No information

4. Conclusion

Skin cancer is among the most common cancers in the world. Melanoma, which is one of the malignant types, is the most lethal, although it is very rare. However, when diagnosed early, the survival rate is around 90%. Therefore, early detection of malignant types of skin cancer and initiation of appropriate treatment will increase the chances of survival of patients.

In this study, dermoscopic images of benign and malignant skin cancers were classified using the modified ResNet101 model from deep residual networks. When the original ResNet101 model was used, 88,94% success was achieved. In order to increase the performance, the model has been modified as indicated in Figure 2 and the classification performance has increased to 91,36%. The classification performance obtained in the proposed study is higher than that of the studies using the same dataset in the literature. This is promising in that the proposed model will assist professionals in medical decision making.

In future studies, it is planned to increase the performance by further developing the model. Deep learning gives very good results in image classification, but the training process takes a long time. For this reason, in future studies, efforts will be made to achieve high performance in shorter times. It will also be attempted to generalize the model to classify different medical data.

References

[1] Saba, T., Khan, M.A., Rehman, A. & Marie-Sainte, S.L. (2019). Region extraction and classification of skin cancer: A heterogeneous framework of deep CNN features fusion and reduction. J. Med. Syst., 43(9), 289.

[2] Afza, F., Sharif, M., Khan, M.A., Tariq, U., Yong, H.-S. & Cha, J. (2022). Multiclass skin lesion classification using hybrid deep features selection and extreme learning machine. Sensors, 22(3), 799. doi: 10.3390/s22030799

[3] Nadipineni, H. (2020). Method to classify skin lesions using dermoscopic images. arXiv 2008.09418.

[4] Bafounta, M.L., Beauchet, A., Aegerter, P. & Saiag, P. (2001). Is dermoscopy (epiluminescence microscopy) useful for the diagnosis of melanoma? Results of a meta-analysis using techniques

adapted to the evaluation of diagnostic tests. Arch Dermatol., 137(10), 1343-1350. doi: 10.1001/archderm.137.10.1343
[5] Kassem, M.A., Hosny, K.M. & Fouad, M.M. (2020). Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning. In IEEE Access, 8, 114822-114832. doi: 10.1109/ACCESS.2020.3003890
[6] Sun, Q., Huang, C., Chen, M., Xu, H. & Yang, Y. (2021). Skin lesion classification using additional patient information. BioMed Research International, vol. 2021, Article ID 6673852, 6 pages. doi: 10.1155/2021/6673852
[7] Nusraddinov, T. (2018). Skin lesion classification using deep convolutional neural network and HSV color space. Master Thesis, İstanbul Technical University.
[8] Yilmaz, E. & Trocan, M. (2021). A modified version of GoogLeNet for melanoma diagnosis. Journal of Information and Telecommunication, 5(3), 395-405. doi: 10.1080/24751839.2021.1893495
[9] Romero Lopez, A., Giro-i-Nieto, X., Burdick, J. & Marques, O. (2017). Skin lesion classification from dermoscopic images using deep learning techniques. 13th IASTED International Conference on Biomedical Engineering (BioMed), pp. 49-54. doi: 10.2316/P.2017.852-053.
[10] Kaur, R., GholamHosseini, H., Sinha, R. & Lindén, M. (2022). Melanoma classification using a novel deep convolutional neural network with dermoscopic images. Sensors, 22(3), 1134. doi: 10.3390/s22031134
[11] Ergün, E. & Kılıç, K. (2021). Derin öğrenme ile artırılmış görüntü seti üzerinden cilt kanseri tespiti. Black Sea Journal of Engineering and Science, 4(4), 192-200. doi: 10.34248/bsengineering.938520
[12] Öztürk, E. & İçer, S. (2021). Classification of dermoscopy images with feed forward neural network, decision trees and random forest. International Journal of Multidisciplinary Studies and Innovative Technologies, 5(2), 129-135.
[13] Ameri, A. (2020). A deep learning approach to skin cancer detection in dermoscopy images. Journal of Biomedical Physics & Engineering, 10(6), 801-806. doi: 10.31661/jbpe.v0i0.2004-1107
[14] Demir, F. (2021). Derin öğrenme tabanlı yaklaşımla kötü huylu deri kanserinin dermatoskopik görüntülerden saptanması. Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 33(2), 617-624. doi: 10.35234/fumbd.900170
[15] Farooq, M.A., Khatoon, A., Varkarakis, V. & Corcoran, P.M. (2019). Advanced deep learning methodologies for skin cancer classification in prodromal stages. AICS.
[16] Soylu, E. & Demir, R. (2021). Development and comparison of skin cancer diagnosis models. Avrupa Bilim ve Teknoloji Dergisi, (28), 1217-1221. doi: 10.31590/ejosat.1013910
[17] Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H. et al. (2018). Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the International Skin Imaging Collaboration (ISIC), in: Biomedical Imaging (ISBI 2018), IEEE 15th International Symposium on, IEEE, pp. 168–172.
[18] Tschandl, P., Rosendahl, C. & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data 5(1), 1-9.
[19] Fanconi, C. (2022). Skin cancer: malignant vs. benign. https://www.kaggle.com/fanconic/skin-cancer-malignant-vs-benign, (Last access date: May 20, 2022).
[20] LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W. & Jackel, L.D. (1989). Backpropagation applied to handwritten zip code recognition. Neural Comput., 1(4), 541-551. doi: 10.1162/neco.1989.1.4.541
[21] Goodfellow, I., Bengio, Y. & Courville, A. (2016). Deep Learning. MIT Press.
[22] He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778. doi: 10.1109/CVPR.2016.90.

Acknowledgement

The summary of this study was presented at the 3rd International Congress on Artificial Intelligence in Health, held between 11-13 May 2022.