INTERNATIONAL JOURNAL OF APPLIED METHODS IN ELECTRONICS AND COMPUTERS

www.ijamec.org

International Open Access

Volume 11
Issue 03

September, 2023

*Research Article*

# *Benchmarking of ResNet models for breast cancer diagnosis using mammographic images*

***Hasan Serdar Macit*** [a,*], ***Kadir Sabanci*** [a]

[a]*Karamanoglu Mehmetbey University, Department of Electrical and Electronics Engineering, 70200 Karaman, Turkey*

ARTICLE INFO

*Article history:*
Received 13 February 2023
Accepted 11 July 2023

*Keywords:*
Breast cancer
Cancer diagnosis
CNN
ResNet models

ABSTRACT

Breast cancer is one of the cancer types with a high mortality rate worldwide. Early diagnosis is of great importance to reduce this mortality rate. Computer-aided early diagnosis systems enable doctors to make more precise and faster decisions. The Mammographic Image Analysis Society (MIAS) dataset was used in this study. The breast area was selected by masking in mammography images. The number of images was increased using data augmentation techniques. Mammography images were classified as normal, benign and malignant using four different ResNet models. The highest classification accuracy was achieved by using ResNet18 model with 93.83%. The accuracies obtained with ResNet50, ResNet101 and ResNet152 were 87.24%, 87.44% and 91.25% respectively.



## 1. Introduction

Cancer was identified as the second most common cause of death in 2018. This corresponds to 9.6 million deaths worldwide. By 2040, this number is expected to increase to 16.4 million [1]. According to the latest global cancer burden estimates from the International Agency for Research on Cancer (IARC), breast cancer now tops the list of the most frequently diagnosed cancers worldwide, surpassing lung cancer for the first time [2].

Everyone knows the importance of early diagnosis of cancer and that many types of cancer can be treated. Thanks to the developing technology, it is possible to get rid of the disease for people who are diagnosed with early cancer with the tests applied.

Attributes oppose concepts and abstractions that help us understand the change in data. It is difficult to extract high-level and abstract features from raw data. Deep Learning (DL) solves this problem by expressing more basic notations. DL allows computers to build more complex concepts from simple ones [3]. DL automates the task of predictions. DL helps to design a model that we can traverse out dataset, it automatically calculates all the features patterns that are important for analysis and predictions. DL is carried out with a wide variety of artificial neural networks. During the processing of data with this model, learning is carried out in all of the processing layers. Each new layer accepts the information obtained from the previous layer as output. DL determines which parameters should be given more weight by systematically analyzing the data we have. The learning process is completed more quickly and the product, technology, or service to be presented works much more successfully. Transfer Learning (TL) is a deep learning technique. In this technique, the model has already been trained [4].

In the literature, many studies have been carried out for the diagnosis of breast cancer using MIAS dataset. Below is information about some of them.

Taşdemir, Yengeç [5] used MIAS dataset and performed a classification study with breast cancer by textural attribute extraction and improved the images with histogram equalization method. They achieved 81%

* Corresponding author. E-mail address: 210801105@stu.kmu.edu.tr
DOI: 10.58190/ijamec.2023.39

accuracy. Alruwaili and Gouda [6] used ResNet50 to distinguish between malignant and benign cancer in their study. In addition to that, augmented the images for stable learning. After these techniques achieved 89.5% accuracy. Krishna and Rajabhushanam [7] used MIAS dataset. They obtained 92% accuracy rates in the AlexNet. Görgel, Sertbas [8] used the segmented region of interest (ROI) method in their study and then extracted the features by using a wavelet transform. Then, the Support Vector Machine (SVM) was employed as the classifier. By employing these methods, achieved 91.4% accuracy for the classification. Zhang, Wang [9] used a novel computer-aided diagnosis system for detecting abnormal cases. ROI was used for this. They proposed SVM and k-nearest neighbors methods achieved 92.16%. Zhang, Wu [10] carried out preprocessing to deaden noises, enhanced images and removed background and pectoral muscles the MIAS dataset. Subsequently, extracted the features and implemented 10-fold cross-validation for analysis. Their experimental results showed 92.52% accuracy. Vedalankar, Gupta [11] proposed the image augmentation technique in their study. These researchers used AlexNet and extracted the MIAS dataset features and implemented to SVM. It is reported that, they achieved maximum value 92% accuracy. Sha, Hu [12] used image noise reduction, grasshopper optimization algorithm and optimized feature extraction. By applying these methods to the MIAS dataset, compared the simulation results with ten different state-of-the-art methods to analyze the proposed system's efficiency and showed that the proposed method has 92% accuracy. Civcik [13] removed the pectoral muscles MIAS dataset. Enhanced the images. He proposed that he achieved 91.3% accuracy in classifying the enhanced images in the hospital dataset and 82% accuracy in classifying the enhanced images in the MIAS dataset.

In this study, the MIAS image dataset obtained from the Kaggle site [14] was analyzed by extracting the features of benign, normal and malignant tumors using DL method. Breast cancer diagnosis was made with ResNet models using the images in the MIAS dataset. The classification accuracies obtained with four different ResNet models were compared with each other. The contributions of this research are summarized below:

•Performing breast cancer diagnosis with ResNet models.

•Discussing the results obtained using four different ResNet models

•The proposed method is compared with other state-of-the-art models.

## 2. Dataset

The MIAS dataset contains a total of 322 mammogram medical images [14]. These images were collected by J. Suckling from the Royal Marsden Hospital in London [15]. Each image has 1024 X 1024 pixels. Every image portable gray map (PGM) format. The MIAS dataset includes three classes, 209 images are normal cases, 61 images are benign cases, and 52 images are malignant cases. This dataset also contains a report by radiologists on the tumor types and locations. Figure 1 presents sample images and labels of the dataset.

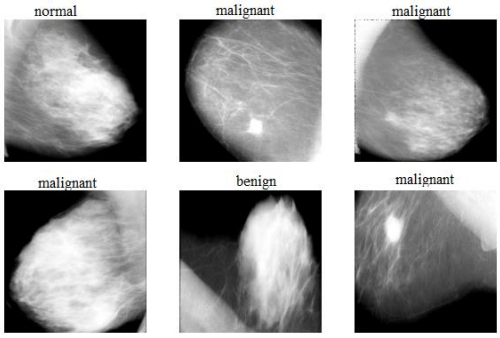


**Figure 1**. Sample images randomly selected from the dataset

Data label details are indicated in Table 1. This table provides detailed information on the images such as image name, background texture, tumor type class, severity of abnormality, X coordinates, Y coordinates and radius. MIAS has no missing data. Radius was given as an approximate value.

**Table 1.** MIAS dataset labels

| Columns | Labels name |
|---|---|
| 1st column | Image name |
| 2nd column | Background texture |
| 3rd column | Tumor type class |
| 4th column | Severity of abnormality |
| 5th column | X coordinates |
| 6th column | Y coordinates |
| 7th column | Radius approximate value |

## 3. Dataset Pre-Processing

Some pre-processing methods have been applied to increase success in this section. PGM format converted to PNG format. The unwanted areas on the images were cropped by the masking methods. Labelme annotations tool [16] was used for the masking of breast images. The masked breast area is presented in Figure 2.

The performance of DL largely depends on size datasets. Learning from small datasets is a major challenge. However, collecting large datasets is time-consuming and expensive.


**Figure 2.** Masked sample breast images

There is a solution to this problem by enlarging the existing data by duplicating it [17]. With this duplicating, additional training examples are created by changing the existing data. In this study, the medical image data is augmented by applying the techniques of changing the current position of the image (shifting), rotating the image with an angle (transform 90, 180, 270 degrees), horizontal flipping, gamma effect and deformation (salt and pepper) techniques. An original image and augmented images are presented in Figure 3.

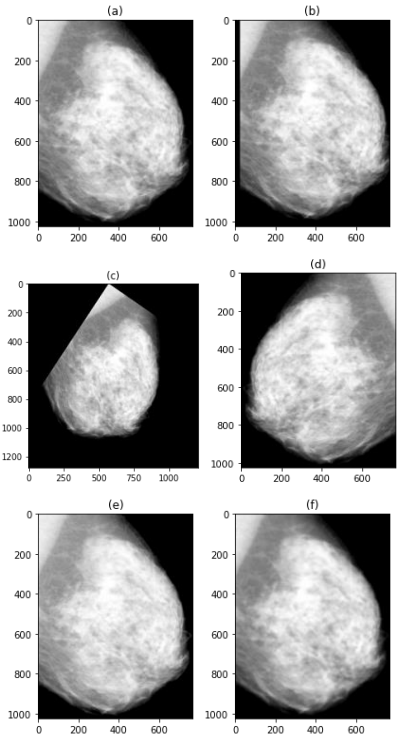


**Figure 3.** The original image of sample of malignant image (a) and augmented images samples: (a) original, (b) shifting, (c) rotating, (d) flipping, (e) gamma, (f) deformation

In this study, data augmentation and re-sampling methods were combined. In order to balance the dataset, the amounts were adjusted by applying the data augmentation technique and re-sample technique to less frequent samples.

After these techniques, dataset was balanced, and data annotations folder was created for the training and testing. One of the biggest challenges when developing a DL model is to avoid overfitting the model to the dataset. The challenge arises when the model learns the combination of weights that performs well on the training data but fails to generalize when given images that the model has never seen. This is known as overfitting. When applying a model to be used in real world, we may want to predict how it will behave after training. This is where the test set comes into play. A random portion of the original dataset intended to represent data not used for training is the test set. So model can predict how it will treat unseen data [18]. The image dataset annotation folder view is presented in Figure 4. After data augmentation and re-sampling, 540 images were formed for each of the malignant, benign, and normal cases. A total 1620 images were created. It is seen that the dataset was balanced. In this study, all images were initially resized to 224 x 224 pixels, and then classification is performed using ResNet models. 80% of the images in the dataset were used for training of model, 20% were used for testing. Then the training folder is again divided into 80% training and 20% validation. It is also common to keep portions of the training set for validation [19]. The validation set is used for fine-tuning the model performance such as for choosing hyper parameters or regularization parameters in the model.

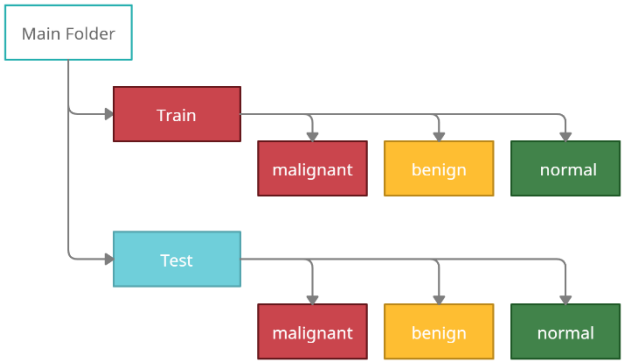


**Figure 4.** Test and training folder of the image dataset

## 4. Model

ResNet is an innovative neural network first introduced by Kaiming She, Xiangyu Zhang, Shaoqing Ren and Jian Sun in a 2015 research paper titled “Deep Residual Learning for Image Recognition” [20]. It is a winner of the ILSVRC 2015 image classification competition [21]. An example of ResNet architecture [22] is indicated in Figure 5. Different ResNet architectures are used to classify the dataset, the number of layers ranges from 18 to 152 [23].

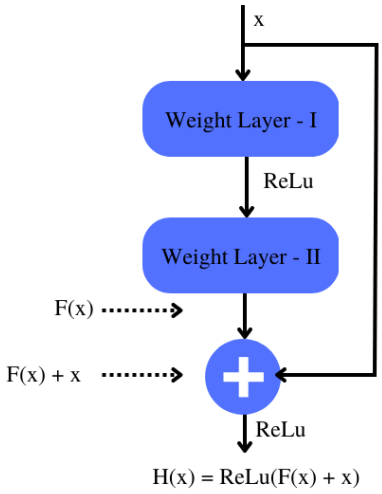


**Figure 5.** This flowchart demonstrates the ResNet architecture [24]

ResNet network resembles networks with convolution, pooling, activation and fully connected stacked layers. The

only structure to convert a simple network to a residual network is the identity link between the layers. A flowchart presenting steps of classification with ResNet model is indicated in Figure 6.

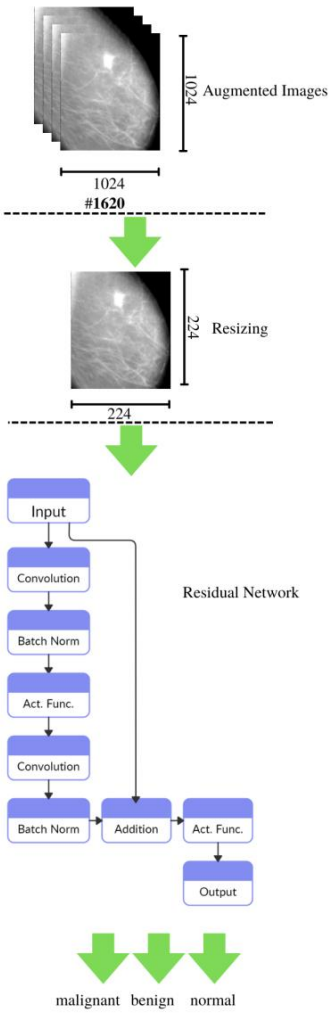


**Figure 6.** Flow chart of classification steps with ResNet

The Sequential model was used in this study. Sequential is a parameter that allows us to create a model layer in Keras. ImageNet is a standard for image classification. Models used in ImageNet classification competitions are measured for performance. Therefore, it provides a standard measure of how good a model is. Most common TL models use imageNet weights [25]. We used Keras, TensorFlow, PyTorch libraries and Python in this approach.

A high-speed graphics processing unit (GPU) was used for deep learning models. The GPU need was met using compute platform by Google Colaboratory (Google Colab). Colab is an interactive, fully cloud-based, collaborative programming platform on Artificial Intelligence (AI) and DL projects.

Normalization is to transform data so that is scaled to [0,1] arrange. Standardization is to transform data so that it has zero mean and unit variance [26]. In this approach, raw data are normalized and standardized for ResNet18 model. Each ResNet model was trained for 100 epochs, the learning range (Lr) was chosen as 0.001 and batch size was chosen as 32. The early stopping model was applied and patience 15 was applied to the model, and the values obtained at each step of this approach were collected (see Table 2). When the difference between validation loss and training loss starts to widen during training, it means that the model is overfitting or learning noisy. In this case the training is stopped by early stop function [27].

**Table 2.** The values of early stopping in the training process

| Transfer Learning | Restoring Model Weights from the best Epoch | Early Stopping | Early Stopping Patience |
|---|---|---|---|
| ResNet18 | N/A | N/A | N/A |
| ResNet50 | 16 Epoch | 31 Epoch | 15 |
| ResNet101 | 25 Epoch | 40 Epoch | 15 |
| ResNet152 | N/A | N/A | 15 |

The activation function rectified linear unit (ReLu) is used for the hidden layers. Softmax is used for the output layer activation. Adam optimization method is used to update the weights of the network. Categorical cross entropy is used for loss function. Images were presented to the network. All the experiments were repeated.

## 5. Results and Discussion

The classification accuracies obtained with ResNet models are given in Table 3. The classification performance metrics achieved with ResNet models are presented in Table 4. The highest classification accuracy was achieved by using ResNet18 model with 93.83%. The accuracies obtained with ResNet50, ResNet101 and ResNet152 were 87.24%, 87.44% and 91.25% respectively.

The confusion matrix (CM) of ResNet18 is presented in Figure 7. Receiver Operating Characteristics (ROC) analysis for ResNet18 is presented in Figure 8.

**Table 3.** The accuracy values obtained with ResNet models

| Model | Accuracy |
|---|---|
| ResNet18 | 93.83 |
| ResNet50 | 87.24 |
| ResNet101 | 87.44 |
| ResNet152 | 91.25 |

**Table 4.** The performance metrics obtained with ResNet models

| Performance Metrics | ResNet18 | ResNet50 | ResNet101 | ResNet152 |
|---|---|---|---|---|
| Loss | 0.2476 | 0.4770 | 0.4410 | 0.4778 |
| Precision | 0.9400 | 0.8164 | 0.8136 | 0.8699 |
| Recall | 0.9400 | 0.7962 | 0.8086 | 0.8672 |
| AUC | 0.9500 | 0.9430 | 0.9545 | 0.9606 |
| F1-Score | 0.9400 | 0.8061 | 0.8110 | 0.8685 |

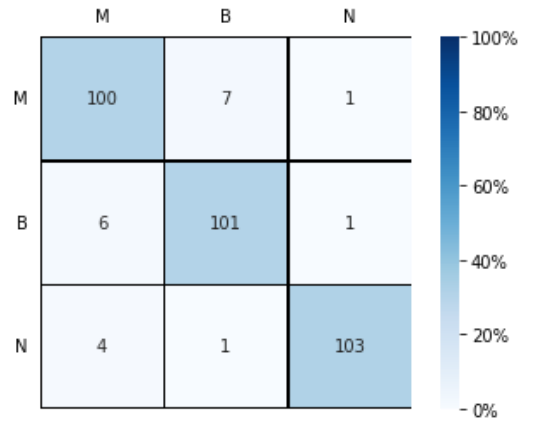


**Figure 7.** Confusion matrix obtained with the ResNet18 model

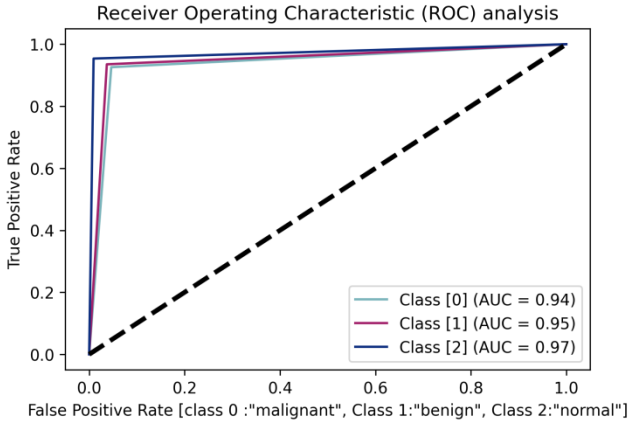


**Figure 8.** ROC curve obtained with the ResNet18 model

CM has four components True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). Accuracy can be measured from TP, FP, TN and FN values [28]. In addition, Accuracy, Loss, Precision, Recall, and F1 score are calculated according to the results of classifications and presented in Table 4. The formulas of the calculated metrics are given between Eqs. (1)-(5) [29].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \text{X}100 \quad (1)$$

$$\text{Loss} = (100 - \text{Accuracy})/100 \quad (2)$$

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

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{F1} - \text{score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (5)$$

When the CM was examined, ResNet18 model found 100 of 108 malignant samples as true positives. The same model detected 101 of 108 samples as benign and 103 of 108 normal samples as true positives. Also, this configuration has minimal validation loss between the training and validation. A receiver operating characteristic (ROC) that follows the diagonal line x=y produces false positive results at the same rate as true positive results. Because of that, we expect a diagnostic test with reasonable accuracy to have a ROC in the upper left triangle above the x=y line as seen in Figure 8. The area under the ROC curve (AUC) is a global measure of the ability of a test to discriminate whether a specific condition is present or not. If an AUC rates is 1.0, the model represents a test with prefect performance [30]. It is seen that the ROC curve is close to 1.0 (see Figure 8).

The comparison of the analysis studies performed with the breast cancer MIAS dataset in the literature is given in Table 5. It is seen that breast cancer analysis can be accomplished successfully in the proposed model. Accuracy can be increased by using different and hybrid deep learning models.

**Table 5.** Comparison of the methods and accuracies with related works.

| Related Works | Model | Accuracy |
|---|---|---|
| Taşdemir, Yengeç [5] | CNN | 81% |
| Alruwaili and Gouda [6] | ResNet50 | 89.5% |
| Krishna and Rajabhushanam [7] | AlexNet | 92% |
| Görgel, Sertbas [8] | SWT-SVM | 91.4% |
| Zhang, Wang [9] | SVM | 92.16% |
| Zhang, Wu [10] | MLP | 92.52% |
| Vedalankar, Gupta [11] | AlexNet | 92% |
| Sha, Hu [12] | CNN | 92% |
| Civcik [13] | ANN | 82% |
| Proposed model | ResNet18 | 93.83% |

Traditional machine learning algorithms were used in some of the studies in Table 5. In some studies, the AlexNet model with less depth was used. In the proposed method, higher accuracy was obtained by using four different ResNet models.

## 6. Conclusions

In recent years, researchers have been using DL methods on MR images in the classification of breast cancer and other cancer types. High-accuracy medical imaging systems support specialist in diagnosis and treatment. In this study, DL algorithms were employed to classify the images in the MIAS dataset. As a result of the training and test experiments, it was seen that the ResNet18 model architecture had the highest accuracy rate. ResNet18 output performed other ResNet architectures. Different data augmented techniques can be applied to the images in the MIAS dataset and the images in the dataset can be further reproduced.

This model can be used in experiments to determine the best dataset. The ResNet18 model is the best choice for MIAS dataset. This proposed system has a low computational complexity and a faster processing structure. Other ResNet models compared; it seems that the performance of this approach is high. The proposed model is applicable and reliable in clinical practice.

Although the proposed method is successful, different methods based on deep learning will be proposed for the diagnosis of breast cancer in future studies. The number of images in the dataset can be increased using different data augmentation techniques. As it is known, the success of deep learning largely depends on the number of labeled data. In addition, classification success can be increased by using different CNN models.

## References


[1] Chauvet, C., *Giving Tuesday*, in *International Agency for Research on Cancer*. 2020.

[2] Weiderpass, E., *Global cancer estimates from IARC*, in *International Agency for Research on Cancer*. 2021.

[3] Goodfellow, I., Y. Bengio, and A. Courville, *Deep learning*. 1 ed. Vol. 1. 2016: Massachussets Institute of Technology. 801.

[4] Hussain, M., J.J. Bird, and D.R. Faria, *A Study on CNN Transfer Learning for Image Classification*, in *18th Annual UK workshop on computational intelligence*. 2018.

[5] Taşdemir, S., et al., *Ysa kullanılarak mamogramlardan dokusal öznitelik tabanlı meme kanseri ilgi bölgesi sınıflandırılması.* Mühendislik Bilimleri ve Tasarım Dergisi, 2020.

[6] Alruwaili, M. and W. Gouda, *Automated Breast Cancer Detection Models Based on Transfer Learning.* Sensors (Basel), 2022. **22**(3).

[7] Krishna, T.H. and C. Rajabhushanam, *Mammography image breast cancer detection using deep transfer learning.* Advances and Applications in Mathematical Sciences, 2021(20:7): p. 1187-1196.

[8] Görgel, P., A. Sertbas, and O.N. Uçan, *Computer-aided classification of breast masses in mammogram images based on spherical wavelet transform and support vector machines.* The Journal of Knowledge Engineering, 2015. **32**(1): p. 155-164.

[9] Zhang, Y.-D., et al., *Computer-aided diagnosis of abnormal breasts in mammogram images by weighted-type fractional Fourier transform.* Advances in Mechanical Engineering, 2016. **8**(2).

[10] Zhang, Y., et al., *Smart detection on abnormal breasts in digital mammography based on contrast-limited adaptive histogram equalization and chaotic adaptive real-coded biogeography-based optimization.* SIMULATION, 2016. **92**(9): p. 873-885.

[11] Vedalankar, A.V., S.S. Gupta, and R.R. Manthalkar, *Addressing architectural distortion in mammogram using AlexNet and support vector machine.* Informatics in Medicine Unlocked, 2021. **23**: p. 100551.

[12] Sha, Z., L. Hu, and B.D. Rouyendegh, *Deep learning and optimization algorithms for automatic breast cancer detection.* International Journal of Imaging Systems and Technology, 2020. **30**(2): p. 495-506.

[13] Civcik, L., *Görüntü zenginleştirme ve hücresel ysa kullanarak meme kanseri teşhisi*, in *Electical - Electronic*. 2013, Selcuk University: Konya p. 154.

[14] Mader, K.S., *Mias mammography*. 2017.

[15] Suckling, J., et al. *The Mammographic Image Analysis Society digital mammogram database*. 1994.

[16] Wada, K. *Image polygonal annotation with python*. 2018; Available from: https://github.com/wkentaro/labelme.

[17] Mitrano, P. and D. Berenson, *Data augmentation for manipulation*, in *Robotics: Science and Systems*. 2022: New York City, NY, USA.

[18] Igerata, A., *How to split a tensorflow dataset into train, validation, and test sets.* Towards Data Science, 2021.

[19] Joseph, V.R. and A. Vakayil, *SPlit: An Optimal Method for Data Splitting.* Technometrics, 2022. **64**(2): p. 166-176.

[20] Boesch, G., *Deep residual networks (ResNet, ResNet50)*, in *Deep Learning*. 2022.

[21] Tsang, S.-H. *Resnet-winner of ilsvrc 2015 image classification, localization, detection*. 2018 [cited 2022 29th Sep.]; Available from: https://towardsdatascience.com/review-resnet-winner-of-ilsvrc-2015-image-classification-localization-detection-e39402bfa5d8.

[22] Sachan, A. *Detailed guide to understand and implement resnet*. Learn Machine Learning, AI & Computer vision 2019 [cited 2022 29th Sep.]; Available from: https://cv-tricks.com/keras/understand-implement-resnets.

[23] LE, K., *A quick overview of resnet models*. 2021, MLearning.ai.

[24] He, K., et al., *Deep Residual Learning for Image Recognition.* 2015.

[25] Gerry, P. *Why are keras models instantiated with imagenet weights only?* 2020; Available from: https://stackoverflow.com/questions/62739893/why-are-keras-models-instantiated-with-imagenet-weights-only.

[26] Pattanayak, S., *Pro deep learning with tensorflow*. 2017: Apress. 412.

[27] Çarkacı, N., *Derin öğrenme uygulamalarında başarım iyileştirme yöntemleri.* Medium, 2017.

[28] Das, A., et al., *Classification of succulent plant using convolutional neural network*, in *2nd International Conference on Cyber Security and Computer* 2020, Daffodil International University.

[29] Koklu, M., et al., *A CNN-SVM study based on selected deep features for grapevine leaves classification.* Measurement, 2022. **188**: p. 110425.

[30] Hoo, Z.H., J. Candlish, and D. Teare, *What is an ROC curve?* Emergency Medicine Journal, 2017. **34**(6): p. 357-359.