



Research Article

Medical image classification with hybrid convolutional neural network models

Osman Doğuş Gülgün^{*1}, Hamza Erol²

¹Mersin University, Institute of Science, Computer Engineering Department, Mersin, Turkey

²Mersin University, Engineering Faculty, Computer Engineering Department, Mersin, Turkey

ABSTRACT

Keywords:

Deep learning
Medical diagnosis
Convolutional neural network
Data augmentation
Classification

Despite important developments in medicine and technology today, many people die due to false or late diagnosis. It is very important to identify the small details in the images that can be overlooked in the examinations made on medical images in terms of early diagnosis of the disease. Therefore, it is vital in some cases to provide early diagnosis by detecting the details in the images automatically by computer systems. In the study carried out, it was aimed to diagnose the disease through medical images by classifying different types of images. For this purpose, convolutional neural networks, which are among deep learning techniques, were evaluated together with different classifier models. In the applied hybrid model approach, feature extraction was obtained from medical images with the convolutional neural network model. The extracted features are used to train different classification models. In the continuation of the study, the performance results obtained from the classifier models are compared. Two different datasets including brain MR images and lung x-ray images were used in the training and testing of hybrid models. In the study, images were classified into two categories as malignant and benign tumors in order to detect images containing malignant tumors in MR images. In order to identify images with pneumonia, the images are similarly classified into two categories, healthy and pneumonia. At the end of the study, the performance results obtained from the model approaches were compared and the performance evaluation of the models was performed.

Hibrit evrişimli sinir ağı modelleri ile tıbbi görüntü sınıflandırması

ÖZ

Anahtar Kelimeler:

Derin öğrenme
Tıbbi teşhis
Evrişimli sinir ağı
Veri artırma
Sınıflandırma

Günümüzde tıp ve teknolojideki önemli gelişmelere rağmen, birçok kişi yanlış veya geç tanı nedeniyle hayatını kaybetmektedir. Tıbbi görüntüler üzerinde yapılan muayenelerde hastalığın erken teşhisi açısından gözden kaçırılacak görüntülerdeki küçük detayların belirlenmesi çok önemlidir. Bu nedenle, bazı durumlarda görüntülerdeki detayları bilgisayar sistemleri tarafından otomatik olarak tespit ederek erken teşhis sağlamak hayati önem taşımaktadır. Yapılan çalışmada, farklı görüntü tiplerini sınıflandırarak hastalığın tıbbi görüntülerle teşhis edilmesi amaçlanmıştır. Bu amaçla, derin öğrenme teknikleri arasında yer alan evrişimli sinir ağları, farklı sınıflayıcı modellerle birlikte değerlendirilmiştir. Uygulanan hibrid model yaklaşımında, evrişimli sinir ağı modeli ile tıbbi görüntülerden özellik çıkarımı elde edilmiştir. Çıkarılan özellikler farklı sınıflandırma modellerini eğitmek için kullanılır. Çalışmanın devamında, sınıflandırıcı modellerinden elde edilen performans sonuçları karşılaştırılmıştır. Hibrid modellerin eğitim ve testinde beyin MR görüntüleri ve akciğer röntgeni görüntüleri dahil olmak üzere iki farklı veri seti kullanılmıştır. Çalışmada MR görüntülerinde malign tümör içeren görüntüleri saptamak için görüntüler malign ve benign tümörler olarak iki kategoriye ayrıldı. Akciğer iltihaplanmalı görüntüleri tanımlamak için görüntüler benzer şekilde sağlıklı ve akciğer iltihaplanması olmak üzere iki kategoriye ayrılır. Araştırma sonunda model yaklaşımlarından elde edilen performans sonuçları karşılaştırılmış ve modellerin performans değerlendirilmesi yapılmıştır.

*Correspond Author

*(dogusgulgun@gmail.com) ORCID ID 0000-0003-1824-4401
(herol@mersin.edu.tr) ORCID ID 0000-0001-8983-4797

e-ISSN: XXXXXXXXXXXX

Arrival Date: 24/04/2020; Acceptanca Date: 11/05/2020

Journal of Computer Science and Technologies

1. INTRODUCTION

In our study, various diagnostic models are explained that can be used for diagnosis of diseases by classification. Hybrid convolutional neural network models that can be used to increase the accuracy rate in the classification of medical images have been described. The hybrid models described in the study were obtained by changing the original convolutional neural network architecture. In these models, different classifiers were used instead of artificial neural networks. Classifiers such as random forest and support vector machines were used as classifiers. Therefore, machine learning and deep learning techniques were evaluated together. In the continuation of the study, some useful deep learning techniques used to improve the performance of the models are also explained. In this study, the medical images of 2 different cases were diagnosed by classification. In the first case, images with pneumonia were detected within the lung x-ray images. Images are classified as "people with pneumonia" and "healthy". In the second case, brain MRI images with benign and malignant tumors were classified. Images are classified as binary, "benign" and "malignant", similar to the first case.

Zhang et al. have proposed a synergistic deep learning model approach to eliminate deep learning techniques from performing poorly in some cases (Zhang et al., 2019). In the developed synergistic model, using two convolutional neural networks, these networks were jointly trained and learned through each other. ResNet-50 architecture was used in the design of convolutional neural networks. ResNet-50 is a 50-layer specific neural network architecture pre-trained and verified by testing its performance (He et al., 2016). The ResNet-50 model has learned high-level features for image classification because it was previously trained with the ImageNet data set, which is a fairly large data set containing millions of image data. If one of the convolutional neural networks used in the synergistic model developed is correctly classified and the other is misclassifying; The error made creates an extra effect for updating the parameters of the model that misclassifies. Therefore, this model is trained mainly on the mistakes made. Therefore, this model learns the classification errors made more effectively. This model developed in the study was evaluated using 4 different data sets and its performance was tested. According to the results obtained at the end of the study, the synergistic deep learning model created has managed to reveal the most successful performance results for each data set.

Oh et al. have proposed a computer-aided diagnostic system that can diagnose Parkinson's disease via EEG signals (Oh et al., 2018). EEG signal records of 20 Parkinson's patients and 20 healthy individuals were used to develop the diagnostic

system. Noises in the received EEG signals were filtered with amplitude and frequency filters to increase diagnostic performance. The researchers used the convolutional neural network architecture from deep learning techniques to classify signals, and proposed a 13-layer convolutional neural network model. The proposed model consists of 1 input layer, 4 (1 x 1) size convolution layers, 4 pooling layers and 1 fully connected layer. There is a 3-layer artificial neural network in the fully connected layer. In the model, 20 filters were used in the first convolution layer, 10 in the second and third convolution layers and 5 in the last convolution layer. In addition, dropout technique was used in the artificial neural network in the fully connected layer in order to avoid overfitting problem in the model. In this way, 50% of the neurons of the artificial neural network in the fully connected layer are disabled every iteration. The model proposed in the study achieved a good performance in diagnosing Parkinson's disease from EEG signal data, achieving 88.25% accuracy, 84.71% sensitivity and 91.77% specificity. The most important advantages of the study carried out compared to other studies in the literature are that the diagnosis of Parkinson is performed directly over EEG signals and there is no need for feature extraction.

Frid-Adar et al. proposed a deep learning-based diagnostic model approach for the classification of liver lesions via computed tomography images (Frid-Adar et al., 2018). They also proposed a data augmentation approach based on deep convolutional generative adversarial network (DCGAN) to improve the classification performance and reliability of their models. In the training of the proposed model, they used a data set containing 182 liver computed tomography images. The data set used contains images belonging to 3 different classes. In the first stage of the study, in order to increase the size of the data set, synthetic image data was produced with the proposed DCGAN model. Synthetic computed tomography images in size (64 x 64 x 1) were produced over random noises using the proposed DCGAN model. As a second data augmentation in addition to this data augmentation method, they also benefited from the classical data augmentation technique, which allows images to be displayed at different angles and at different distances to the model. After the data increase, the convolutional neural network model, another model proposed in the study, was classified with the images in the data set. The proposed convolutional neural network model consists of 3 convolution, 3 pooling and 1 fully connected layer. The artificial neural network in the fully connected layer consists of a total of 2 layers, the first with 256 neurons and the second with 3 neurons. In addition, in order to avoid overfitting problem, dropout technique is used in the fully connected layer. When using only classical data augmentation technique,

the proposed model achieved 78.6% sensitivity and 88.4% specificity values. In addition to the classical technique, using the additional data produced with the GAN model, the model has achieved 85.7% sensitivity and 92.4% specificity values.

Bejnordi et al. proposed a diagnostic model based on the convolutional neural network architecture from deep learning techniques to diagnose breast cancer disease through breast biopsy slide images (Bejnordi et al., 2017). In the training of the model proposed in the study, a data set containing 646 breast biopsy images was used. The model proposed in the study consists of two parts and accordingly, the diagnosis is carried out in two stages. The first convolutional neural network used in the first stage is classified as the epithelial, fat and stroma regions in the images. The model used in the first stage is based on the VGG-Net network architecture and consists of 11 layers. In this section, 3x3 size filters are used in the convolutional neural network. In addition, the ReLU activation function was used in the convolution layers, and (2x2) pooling layers were used after each convolution layer. In the second stage, a different second convolutional neural network model was used to classify the stroma regions as normal or cancer related stroma. The architecture of the second model is based on the VGG-Net architecture as in the first model and consists of 16 layers. Using this model, feature extraction is performed from the stroma regions. Attribute outputs from the second model are then classified into a normal or cancer-related stroma at the last stage using a random forest classifier. Accordingly, breast cancer occurrences are detected by the model in biopsy images. According to the results obtained at the end of the study, the first CNN model, which constitutes the first part of the proposed model, has managed to classify the tissue sections in the images as fat, epithelium and stroma with a 95.5% accuracy rate. In the second stage, the second model used to determine whether the stroma sections are cancer-related or not has managed to accurately predict the classes of stroma sections in the images with an accuracy rate of 92.0%.

Martinez-Murcia et al. proposed a deep learning-based model to diagnose Alzheimer's disease via MRI images (Martinez-Murcia et al., 2019). They proposed a convolutional autoencoder model in their studies to obtain high-level attributes to be used for classifying images. The performance of the autoencoder model proposed in the study was evaluated with the ADNI data set. ADNI dataset is an open source dataset developed for the diagnosis of Alzheimer's disease via MRI images. The proposed autoencoder model consists of encoder and decoder parts. While the encoder part of the model consists of 6 convolution layers and 1 pooling layer, the decoder part has 5 reverse convolution layers. They compared the different results obtained by changing the number of neurons

in the convolution layers. According to the results obtained, the proposed model can determine the images of people with Alzheimer's disease with 84% sensitivity rate. Therefore, the classification of the images in this way by obtaining important attributes with the convolutional autoencoder models produced promising results in terms of diagnosis.

Indraswari et al. proposed an advanced deep learning model in their studies to perform the segmentation of 3D images (Indraswari et al., 2019). Three different data sets were used to evaluate the proposed model. The first two of the data sets were used to perform brain tumor segmentation, and the third data set was used to perform dental segmentation over the jaw images of individuals. A convolutional autoencoder architecture consisting of encoder and decoder parts has been proposed to perform segmentation. First of all, images are divided into axial, coronal and sagittal components and 2D image component is obtained for each plane. Important attributes were obtained by applying convolution and deconvolution operations to the image components. Important features were obtained by applying convolution and deconvolution operations to the image components. Attributes extracted from axial, coronal, and sagittal slices were combined after the final pooling and size reduction phase to obtain the main input vector to be used in the model. By applying the convolution process to the combined input information, the attributes to be used in the classification of the images are extracted from the data. A special cost function is also proposed for this model used in the study. The proposed cost function adds a certain weight to the classes to take into account the probability of all classes in the classification. At the end of the study, despite the insufficient data sets, they managed to achieve good performance results thanks to the model they proposed and the cost function they used.

Shahzadi et al. proposed a deep learning model that performs the type detection of tumors in the brain via MRI images (Shahzadi et al., 2018). In the study, brain MRI images from individuals with glioma, a common type of brain tumor, were used. Glioma type tumors were classified as high grade (HG) and low grade (LG). In the model proposed for this purpose, deep convolutional neural network and long short-term memory (LSTM) network architectures are used together. In the model, significant features were extracted from the images using the VGG-16 convolutional neural network architecture and then these features were used to train the LSTM network. In this way, using the trained model, glioma cases were classified in volumes by high grade (HG) or low grade (LG). During the training phase, 80% of image data was used to train the proposed model and the remaining 20% was used to test the proposed model. According to the results, the model managed to accurately predict the class of 84% of the test data.

In the study, feature extraction was also performed by using AlexNet and ResNet architectures as the convolutional neural network, but the best result was obtained with VGG-16 architecture with 84% accuracy rate.

Afshar et al. proposed a classification model based on capsule networks architecture with a different deep learning approach to perform brain tumor type detection via MRI images (Afshar et al., 2018). In the study, a data set containing 3064 MRI images belonging to 233 patients with one of three different tumor types was used. It is aimed to reduce the processing time and increase the model performance by rescaling the images in the data set to the size (64 x 64). The scaled images are first processed in the convolution layer containing 56 filters and then transferred to the main capsule layer. In this section, 8 feature maps (24 x 24) are obtained and the obtained features are transferred to the last capsule layer. The last capsule layer contains 3 capsules in size 16 for 3 tumor classes. Different attempts were made by changing the layer and capsule numbers of the capsule network architecture used. According to the best results obtained at the end of the study, the proposed model estimated that 86.56% of the test data. The study is important in terms of proposing a different approach in the field of image classification with deep learning.

Korolev et al. suggested a model using deep learning techniques to perform the detection of Alzheimer's disease and different neurological diseases via MRI images (Korolev et al., 2017). In the study, binary classification of images belonging to 4 different categories was performed by using a part of ADNI data set that is open to public access. In the classification of MR images, two different model approaches were used, the first one is classical convolutional neural network and the other one is modern residual neural network model. Performance comparison of these two models was made. The first model used is a classical convolutional neural network architecture and consists of 21 layers. The second model used is the modern ResNet architecture that won the Imagenet competition in 2015. According to the results obtained at the end of the study, 79% accuracy rate was obtained with 21-layer convolutional neural network model in the classification of individuals with or without Alzheimer's from the categories. In the same task, with the ResNet model, 80% accuracy rate was achieved.

Khobragade et al. proposed a deep learning-based diagnostic model for the detection of different lung diseases (Khobragade et al., 2016). In the study, the classification of tuberculosis, pneumonia and lung cancer diseases, which are three important lung diseases, were dealt with on chest x-ray images. In order to classify the chest X-ray images into 3 different categories, a 4-step model approach has been proposed. In the first stage, the images are subjected to some

preprocesses and the high-pass filter reduces the noise in the images. In the second stage, lung segmentation is performed on the images and the limits of the lungs in the images are determined. For this purpose, density-based edge detection technique is used. In the third stage after lung segmentation, important geometrical features such as region circumference, equivalent diameter, irregularity index, and statistical features such as standard deviation and entropy are extracted for classification by image processing. In the fourth stage, an artificial neural network consisting of 3 hidden layers was trained and tested using the features. In this way, the model was evaluated. According to the results obtained after evaluating the model, the model successfully managed to estimate the class of 92% of the test data.

Varshni et al. proposed a diagnostic model approach based on deep learning and machine learning methods for the automatic determination of pneumonia in the lung over chest x-ray images (Varshni et al., 2019). The data set used in the study was created by making use of a large-scale data set containing 112,120 images. In the data set created, there are 1431 images for people with and without pneumonia. Therefore, a data set consisting of a total of 2862 images selected from 112,120 images was used in the study. In the study, feature extraction was performed with different convolutional neural network architectures such as Xception, VGG-16, ResNet-50, DenseNet-121, DenseNet-169. In the next step, the important features extracted were used to train and test different classification models such as artificial neural network, support vector machines, naive bayes, nearest neighborhood and random forest. Therefore, the study includes a comprehensive performance comparison from different combinations of convolutional neural network architectures and different classifier models. According to the results obtained in the study, the most successful result was obtained by classifying the features extracted with DenseNet-169 convolutional neural network model with the support vector machines classifier.

This study consists of "Introduction", "Materials", "Methods", "Results" and "Conclusions and Suggestions" sections. In the "Introduction" section, the literature review is included and the study is explained. In the "Materials" section, two different image datasets are described, which are used as materials in the study. In the "Methods" section, the deep learning models used in the classification of image data are explained in detail. In the "Results" section, all calculation and classification results obtained from the deep models used are included. In the "Conclusion and Suggestions" section, the results of two different cases were evaluated and the performance comparison of the models was made.

2. MATERIALS

In this part of the study, 2 different data sets that will be used as a material for testing the created models will be explained. Therefore, the classification models created in this study were used to diagnose two different diseases through medical images. The first data set used in the study was used to establish the pneumonia disease diagnosis model, and the second data set was used to create the brain tumor diagnosis model.

2.1. Pneumonia Data Set

The first data set used in the study was created for the diagnosis of pneumonia on lung x-ray images, and the data set consists of a total of 5840 x-ray images (Kermany and Goldbaum, 2018). This data set consists of x-ray images of one and five-year-old pediatric patients taken from Guangzhou Women's and Children's Medical Center. All X-ray images taken were taken as part of routine clinical care of patients ("Chest X-Ray Images", 2019). In order to perform accurate analysis of chest x-ray images with computer systems, images with low quality or undetectable diagnosis were identified and removed from the data set, thereby ensuring quality control of the images. In order to determine the classification of the received images and to classify the classification models to be used in the classification of the images, the diagnosis of the data set images was first made by two specialist doctors. The data set was also checked by a third specialist to take into account any possible erroneous evaluations of the specialist physicians who evaluated the data.

While 4265 of the images in the data set are lung images of people with pneumonia, the remaining 1575 are images of healthy people without pneumonia. Therefore, the data set contains x-ray images of two different classes. 75% of the data set images were used to train the diagnostic models created and 25% were used to test the models. Therefore, 4380 images were used in the training of models, and 1460 images were used in testing the models. Before the image data was presented as input to the model, some pre-processes were applied to the image data in order to increase the classification performance of the models. The applied pre-processes are described in the following sections. An example image of the data set described is given in Figure 1.

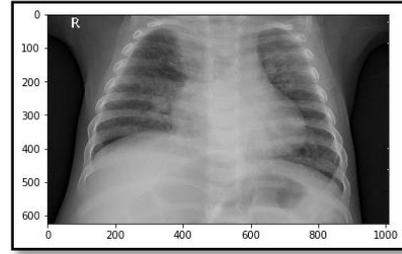


Figure 1. A sample chest x-ray image in the data set ("URL-1").

2.2. Brain Tumor Data Set

The second data set used in the study was created for brain tumor detection and the data set consists of 253 brain MRI images (Chakrabarty, 2019). 153 of the MRI images included in the data set are images of patients with malignant (malignant) tumors, while 98 of them are images of benign (benign) tumors ("Brain MRI Images", 2019). Therefore, the data set consists of MRI images of two different classes.

189 images, which constitute 75% of the total images and randomly selected, were used to train the diagnostic models created, and the remaining 64 images were used to test the models. Therefore, 25% of the total images were used to test the model. In order for the created models to perform higher, a number of pre-processes such as normalization and resizing were applied to these images. Details of these pre-processes will be discussed in detail in the following sections. An example image of the data set described is shown in Figure 2.

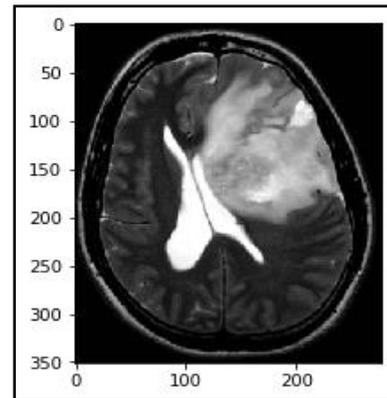


Figure 2. A sample brain MRI image in the data ("URL-2").

3. METHODS

In this section, deep learning models used to diagnose two different diseases are explained over two different medical image data sets. In addition, different deep learning techniques applied in models used to increase model performance are also discussed in this section. 3 different model designs were used in the classification of data sets. Classification procedures were carried out using deep learning models with deep convolutional neural network, convolutional neural network with random forest classifier, and convolutional neural network with support vector machine classifier.

3.1. Deep Convolutional Neural Network Model

A convolutional neural network model was first used to diagnose two different diseases based on image data. The keras library of the python programming language was used in the design of the convolutional neural network model. In order for the images to be perceived and classified by computer systems, they must first be converted into a digital format that the computer can perceive. For this reason, firstly, image matrices containing the pixel values of the images to be used must be obtained. In the model used, the images of the datasets previously described are resized and image matrices in the size of the images (64x64) are obtained.

After obtaining the image matrices, in addition to these matrices, normalization process is applied. In this way, matrix pixel values are provided to take a value between 0 and 1 (Goodfellow et al., 2016). Thanks to the normalization pre-process, the distorted pixel values in the image matrices are filtered out. In this way, the classification performance of the models used has been increased. When classifying images in convolutional neural networks, it is necessary to obtain classifying features that affect the class of images via image matrices. While the models used determine the class of the test images in the test phase, they determine the class of the images depending on the similarity and difference of these attributes. Therefore, the resulting image matrices are subjected to convolution with filters of a certain size in the convolution layer in order to obtain the features from the images, so that the features to be used in the classification are obtained. In the convolutional neural network model used in the study carried out, the image matrices are subjected to convolution with (5x5) filters.

As a result of the convolution process, the size of the original image matrix is reduced. The reduction of the image matrix size increases the processing speed of the models used and provides an advantage in a temporal sense (LeCun et al., 1989). However, this decrease in size also leads to a certain level of information loss (Aghdam and Heravi, 2017). Padding technique can be used in

such models to prevent loss of information due to the decrease in the size of the image matrix. When this technique is applied, the size of the image matrices presented as an input to the convolution process does not change and as a result of the process, a matrix of the same size is obtained as the output (Krizhevsky et al., 2012). In this way, the problem of any loss of information from the image data as a result of the convolution process is prevented (Qian et al., 2016). In the model used in the study, the same padding technique was used. The number of filters used in the convolution layer can vary depending on the designer and images.

After the convolution process, the image matrices in the convolution layer are included in an activation function. In this study, relu activation function was used at this stage. The main reason why relu activation function is used in the model is the complexity of this function. The use of this activation function performs better in detecting complex nonlinear shapes when convolutional neural networks detect distinctive features in images (Chollet, 2017). After this process, the image matrices are transferred to the pooling layer by the model. In this layer, the number of pixels is reduced by changing the image size. In this way, it is ensured that the performance and speed of the model is increased and the problem that the model called overfitting produces wrong results based on memorization during the learning phase is also solved (Goodfellow et al., 2016). In the model used in this section, the max pooling technique, one of the pooling methods, is used in the pooling layer.

As a result of passing the image matrices through the first convolution layer, the model learns low-level features in the first place through the data. In the model used, the image matrices are passed through convolution and pooling layers two more times. This allows the model to learn mid-level and high-level features in images. The model's learning of more complex high-level features also significantly increases classification performance. In the model created, 30 filters in the first convolution layer and 60 filters in the second and third convolution layers were used. As the layer depth increases, the model can learn more complex features in a better way. After these steps, the attributes of the processed image matrices are transferred to the fully connected layer by the model.

The fully connected layer is the section where the classification of the images is performed depending on the similarity and difference of the attributes. Therefore, the diagnosis of the test images in the data set will be made in this section. The model carries out the classification process and thus the diagnosis of the disease through an artificial neural network in this layer. However, since the artificial neural network in this layer cannot process the image data in matrix format, the image matrix data in question must be given to the artificial neural network in vector format.

Therefore, in this layer, the image matrices are converted into vectors by the flatten operation, and the data of the images are provided as input to the artificial neural network in this layer. Vectors carrying feature data are transferred to an artificial neural network with 500 neurons. Depending on the training data provided as input, the weights of the artificial neural network are cyclically rearranged, and as a result, the artificial neural network in this layer gains the ability to classify through features and thus diagnose diseases. The model trained on the training data is then tested on the image data allocated for the testing phase in the data set, and the classification accuracy percentage, total error and performance of the model in question are tested. Sigmoid function was used as activation function in the artificial neural network at this stage.

In this type of neural networks used in the fully connected layer, as described earlier, the problem of overfitting can be seen. In this problem, the model outputs erroneous classification results based on memorization based on the excessive learning of the data by the model. Therefore, dropout technique was used in this layer as an additional step to prevent this problem. Dropout technique is a regulation method used to train the artificial neural network in this layer (Srivastava et al., 2014). With this technique, half of the neurons are disabled in different iterations with different possibilities during data processing in the artificial neural network. Thus, in each iteration, a certain number of weights are changed instead of all the weights of the artificial neural network in this layer (Chen et al., 2015). Thus, the effect of any image given to the model on the neural network weights is limited. This prevents the network from performing incorrect classifications due to over learning. With the dropout technique applied in the model described in this section, 40% of neurons in the neural network are disabled in every iteration.

Another technique that is applied in order to provide better training on image data and to improve classification performance during the training of the model is the data augmentation technique. The data augmentation technique is applied to the images used to train the model in the data set. In this technique, the images in the image dataset are given to the model both in their original formats and by making some changes (Mikołajczyk et al., 2018). The changes are made by showing the images to the model at different angles and at different distances (Wong et al., 2016). In the model described in this section, a new image with a different structure was obtained from each image used to train the model in the data set by applying the data augmentation technique. Therefore, the number of images used in training the model has been doubled by using this technique. In addition, because the images are produced randomly and differently, the model learns the feature attributes in the images more comprehensively and more

efficiently. Thanks to the re-evaluation of the same images from different angles and different distances, it is provided to identify the features that will give better results in classification. In this way, the model can perform class separation of the data more successfully and accordingly, the classification accuracy percentage of the model increases. The model architecture described in this section is given in Figure 3.

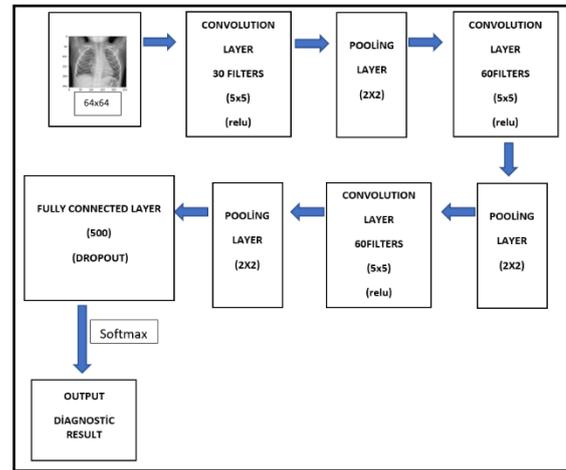


Figure 3. CNN architecture described in this section.

3.2. Convolutional Neural Network with Random Forest Classifier

Secondly, a hybrid convolutional neural network model was used to diagnose two different diseases based on image data. Keras and scikit-learn libraries belonging to the python programming language were used in the design of the classification model used. In the design of the hybrid model created in this section, the deep convolutional neural network model described in the previous section was used. The low, medium and high-level features of the image data were obtained using the deep convolutional neural network model described in the previous stage. After obtaining these features, random forest algorithm, one of the machine learning techniques, was used to classify these features in the model.

Random forest is a collective classification and regression algorithm that uses decision trees as a weak classifier (Breiman, 2001). Each of the decision trees used as weak learners in the algorithm is trained using a random data set derived from the original data set.

While classifying a sample with this algorithm, the classification of the sample is determined according to the majority decision by taking the decisions of weak classifiers for the sample in question. The sample evaluated according to the votes of weak classifiers is considered to belong to the class with the highest number of votes. Therefore, the most important difference of the

model described in this section from a conventional convolutional neural network is that after the features of the model are obtained, the random forest classifier is used as a different classifier instead of an artificial neural network. While classification operations are carried out with the random forest algorithm, the features that affect the classes the most are found in the tree's root and root nodes in decision trees, which are weak classifiers. Therefore, these features will be more effective when deciding on the class of an image instance. In this way, it will increase the classification performance of the model, since the features that have the most impact on determining classes will be more effective when deciding on the class of an image instance. Therefore, the strength of the model created in terms of performance is that the random forest algorithm is successful at the classification point. The hybrid model architecture described in this section is given in Figure 4.

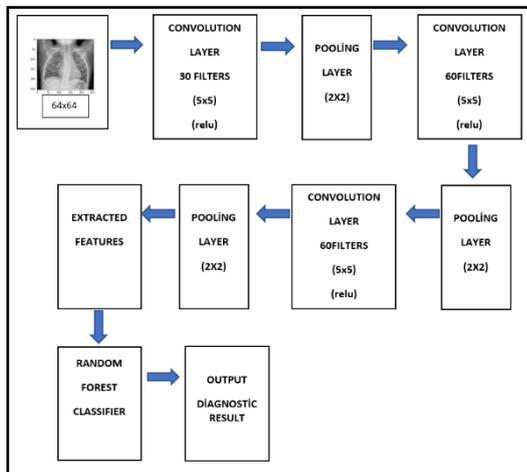


Figure 4. CNN with random forest classifier architecture described in this section.

3.3. Convolutional Neural Network with Support Vector Machines

Thirdly, a hybrid convolutional neural network model similar to the previous section was used to diagnose two different diseases based on image data. In the design of the classification model used, keras and scikit-learn libraries belonging to the python programming language were also used. The deep convolutional neural network model was used in the design of the hybrid model created in this section and the low, medium and high-level attributes of the image data were obtained using this model. Then, support vector machines algorithm, one of the machine learning techniques, was used to classify these attributes in the model.

Support vector machines are basically a classification algorithm that tries to find the plane that separates the classes with the widest margin in the sample space and in the most convenient way (Vapnik and Chapelle, 2000). Considering a linearly separable sample space, there are many plane

solutions, but the support vector machines algorithm calculates the widest margin of these solutions by derivative. Therefore, in the model created in this section, the support vector machines classifier was used to classify the features obtained from the images. The hybrid model architecture described in this section is given in Figure 5.

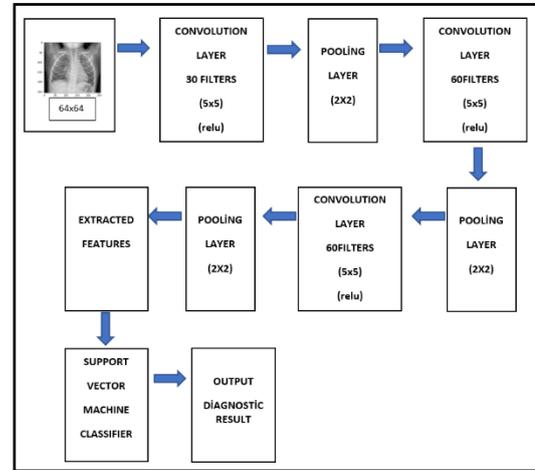


Figure 5. CNN with support vector machines classifier architecture described in this section.

4. RESULTS

In this section, the results obtained from three different classification models tested using two different medical image data sets are given and the performance of the models in the classification process is evaluated.

4.1. Results from The Deep Convolutional Neural Network Model

The deep convolutional neural network model was initially trained and tested using the pneumonia disease dataset. The number of images used in training of the model was increased by applying data augmentation technique to the part of the pneumonia disease data set reserved for education. The classification confusion matrix results obtained by classifying the test data of the model are given in Figure 6.

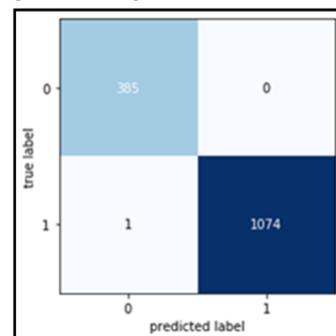


Figure 6. The result of the classification confusion matrix of the first model for the lung x-ray image data set.

When the classification confusion matrix is analyzed, it is seen that the model made 1 incorrect classification in total. The model incorrectly estimated the class of 1 person with pneumonia and included this person in the class of healthy people. The classification report results of the model are given in Figure 7.

Classification Report:				
	precision	recall	f1-score	support
0	1.00	1.00	1.00	385
1	1.00	1.00	1.00	1075
accuracy			1.00	1460
macro avg	1.00	1.00	1.00	1460
weighted avg	1.00	1.00	1.00	1460

Figure 7. The classification report results of the first model for the lung x-ray image data set.

Precision, recall, f1-score and accuracy values obtained from the classification process performed in the classification report results can be seen. The training of the model was carried out in a total of 25 iterations. The classification accuracy and loss graphs obtained during the training of the model are given in Figure 8 and Figure 9.

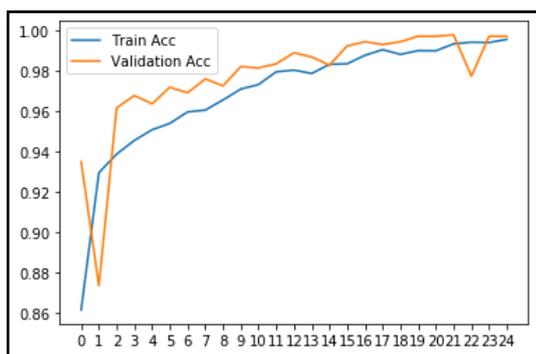


Figure 8. The graph of the classification accuracy values obtained by applying the first model for the lung x-ray image data set.

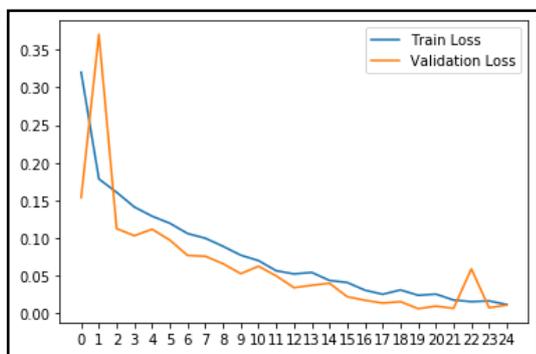


Figure 9. The graph of loss values obtained by applying the first model for the lung x-ray image data set.

According to the results, the deep convolutional neural network model, trained using pneumonia disease data, managed to accurately estimate the classes of 1460 test data by 99.8% in the post-training test process. The model has managed to classify 1459 of these images correctly, and it has incorrectly estimated the class of 1 test image.

The deep convolutional neural network model was evaluated in the second stage using the second dataset for brain tumor type detection. The number of images used in training the model was increased by applying data augmentation technique to the part of the brain tumor data set reserved for education, similar to the first data set. The classification confusion matrix results obtained by classifying the test data of the model are given in Figure 10.

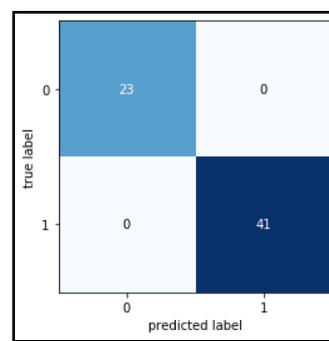


Figure 10. The result of the classification confusion matrix of the first model for the brain MRI image data set.

When the classification complexity matrix is analyzed, it is seen that the model correctly predicts the class of all samples and does not make any incorrect classification. The classification report results of the model are given in Figure 11.

Classification Report:				
	precision	recall	f1-score	support
0	1.00	1.00	1.00	23
1	1.00	1.00	1.00	41
accuracy			1.00	64
macro avg	1.00	1.00	1.00	64
weighted avg	1.00	1.00	1.00	64

Figure 11. The classification report results of the first model for the the brain MRI image data set.

Precision, recall, f1-score and accuracy values obtained from the classification process performed in the classification report results can be seen. The training of the model was carried out in a total of 25 iterations. The classification accuracy and loss graphs obtained during the training of the model are given in Figures 12 and Figure 13.

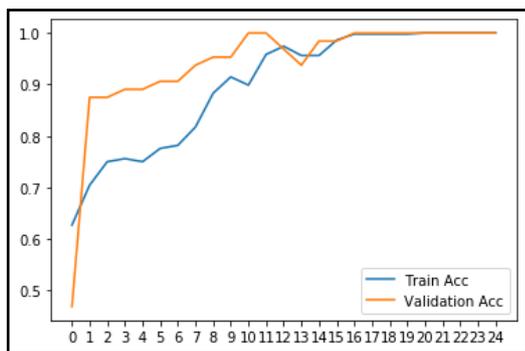


Figure 12. The graph of the classification accuracy values obtained by applying the first model for the brain MRI image data set.

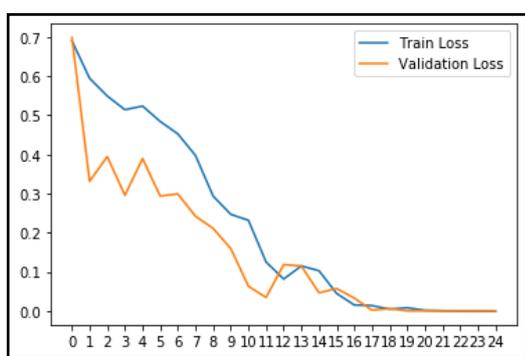


Figure 13. The graph of loss values obtained by applying the first model for the brain MRI image data set.

According to the results obtained, the deep convolutional neural network model, which was trained using brain MRI image data, succeeded to predict the classes of 64 test data 100% correctly in the post-training test process.

4.2. Results from The Convolutional Neural Network with Random Forest Classifier Model

The convolutional neural network model with a random forest classifier was initially trained and tested using the pneumonia disease dataset. The classification confusion matrix obtained by classifying the test data of the model is given in Figure 14.

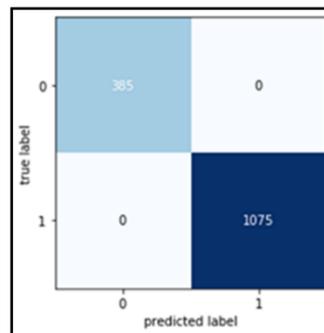


Figure 14. The result of the classification confusion matrix of the second model for the lung x-ray image data set.

When the classification confusion matrix result is examined, it is seen that the model correctly predicts the class of all the samples and does not make any incorrect classification. The classification report results of the model are given in Figure 15.

Classification Report:				
	precision	recall	f1-score	support
0	1.00	1.00	1.00	385
1	1.00	1.00	1.00	1075
accuracy			1.00	1460
macro avg	1.00	1.00	1.00	1460
weighted avg	1.00	1.00	1.00	1460

Figure 15. The classification report results of the second model for the the lung x-ray image data set.

Precision, recall, f1-score and accuracy values obtained from the classification process performed in the classification report results can be seen. According to the results obtained, the convolutional neural network model with random forest classifier, trained with the obtained features, managed to accurately predict the class of all 1460 test data at the test stage. Therefore, the model was able to achieve 100% accuracy in the testing phase. The calculated classification accuracy rate and total error value results of the model are given in Figure 16.

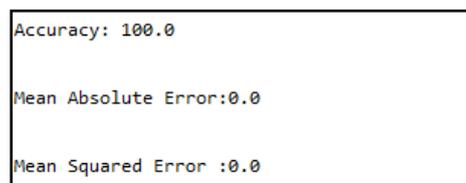


Figure 16. The classification accuracy rate and total error value result of the second model for the lung x-ray image data set.

The convolutional neural network model with a random forest classifier was evaluated in the second stage using the second dataset for brain tumor type detection. The classification confusion matrix results obtained by classifying the test data of the model are given in Figure 17.

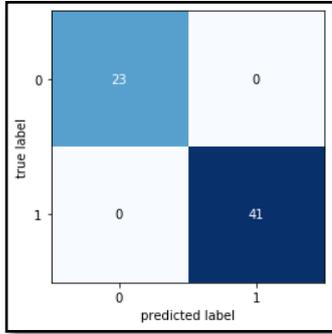


Figure 17. The result of the classification confusion matrix of the second model for the brain MRI image data set.

When the classification confusion matrix is analyzed, it is seen that the model correctly predicts the class of all samples and does not make any incorrect classification. The classification report results of the model are given in Figure 18.

Classification Report:				
	precision	recall	f1-score	support
0	1.00	1.00	1.00	23
1	1.00	1.00	1.00	41
accuracy			1.00	64
macro avg	1.00	1.00	1.00	64
weighted avg	1.00	1.00	1.00	64

Figure 18. The classification report results of the second model for the brain MRI image data set.

This model has been used in the deep convolutional neural network model explained in the first part in obtaining the features used in classification. The training of the deep convolutional neural network was carried out in 25 iterations. Then the attributes learned by the trained model were classified with the random forest classifier. It can be seen in the results that the model tested with the test data has reached 100% accuracy at this stage. The calculated classification accuracy rate and total error value results of the model are given in Figure 19.

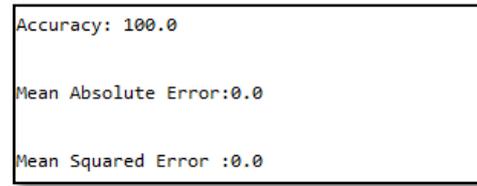


Figure 19. The classification accuracy rate and total error value result of the second model for the brain MRI image data set.

4.3. Results from The Convolutional Neural Network with Support Vector Machines Model

The convolutional neural network model with a support vector machine classifier was initially trained and tested using the pneumonia disease dataset. The number of images used in training of the model was increased by applying data augmentation technique to the part of the pneumonia disease data set reserved for education. The classification confusion matrix results obtained by classifying the test data of the model are given in Figure 20.

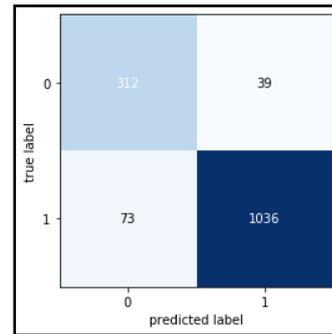


Figure 20. The result of the classification confusion matrix of the third model for the lung x-ray image data set.

When the classification confusion matrix result is analyzed, it is seen that the model correctly predicted the class of 1348 samples as well as 112 specimens incorrectly. The model estimated that 39 people who were actually patient were healthy and 73 people who were actually healthy were patients. The classification report results of the model are given in Figure 21.

Classification Report:				
	precision	recall	f1-score	support
0	0.81	0.89	0.85	351
1	0.96	0.93	0.95	1109
accuracy			0.92	1460
macro avg	0.89	0.91	0.90	1460
weighted avg	0.93	0.92	0.92	1460

Figure 21. The classification report results of the third model for the lung x-ray image data set.

Precision, recall, f1-score and accuracy values obtained from the classification process performed in the classification report results can be seen. According to the results obtained, the convolutional neural network model with support vector machine classifier succeeded in obtaining 92.3% accuracy rate by succeeding in correctly predicting the class of 1348 test data in the test phase. The calculated classification accuracy rate and total error value results of the model are given in Figure 22.

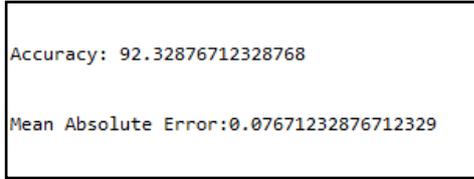


Figure 22. The classification accuracy rate and total error value result of the third model for the lung x-ray image data set.

The convolutional neural network model with a support vector machine classifier was evaluated in the second stage using the second dataset for brain tumor type detection. The classification confusion matrix results obtained by the classification of the test data of the model are given in Figure 23.

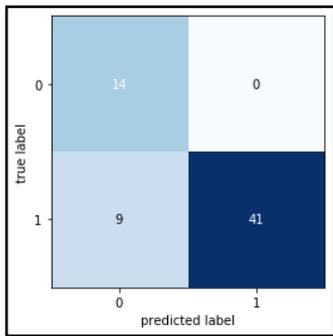


Figure 23. The result of the classification confusion matrix of the third model for the brain MRI image data set.

When the classification confusion matrix result is examined, it is seen that the model correctly predicted the class of 55 samples, but also incorrectly predicted the class of 9 samples. The model actually estimated the tumor type of 9 people with benign tumors incorrectly and identified the tumor type of these people as malignant. The classification report results of the model are given in Figure 24.

Classification Report:				
	precision	recall	f1-score	support
0	0.61	1.00	0.76	14
1	1.00	0.82	0.90	50
accuracy			0.86	64
macro avg	0.80	0.91	0.83	64
weighted avg	0.91	0.86	0.87	64

Figure 24. The classification report results of the third model for the brain MRI image data set.

The deep convolutional neural network model described in the first part was used to obtain the features used in this classification. Then the features learned by the trained model were classified with the support vector machine classifier. It can be seen from the results that the model tested with the test data reached 85.9% accuracy at this stage. The calculated mean absolute error value of the model was determined as 0.14. The calculated classification accuracy rate and total error value results of the model are given in Figure 25.

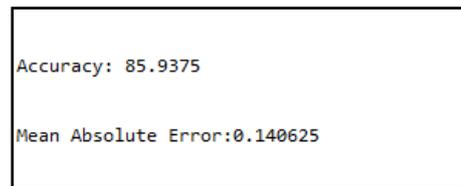


Figure 25. The classification accuracy rate and total error value result of the third model for the brain MRI image data set.

5. CONCLUSIONS AND SUGGESTIONS

When all model results are compared, it is seen that the most successful performance results for both data sets are obtained with the convolutional neural network with random forest classifier. The classification accuracy rates obtained for different data sets from the models are given in Table 1.

Table 1. Classification accuracy rates of all models.

Models	Accuracy rate for first dataset	Accuracy rate for second dataset
First Model (CNN)	%99,8	%100
Second Model (CNN+RF)	%100	%100
Third Model (CNN+SVM)	%92,3	%85,9

Since both data sets used in the study contain 2 classes, the models succeeded in classifying images with very high accuracy rates. Of the models used, the convolutional neural network model with a

random forest classifier has succeeded in accurately predicting all classes of test samples for both data sets. However, it is estimated that model performances may decrease below these levels if the number of classes increases. For example, although the hybrid model using the random forest classifier reached 100% accuracy in the study, it is estimated that this rate may decrease if the amount of data used increases. In addition, it is thought that the convolutional neural network with support vector machine classifier can exhibit higher performance in different types of problems with more class. As a result, it can be said that the hybrid model approaches used in the study give successful performance results in binary image classification problems.

Convolutional neural network-based models, which are used in the classification of medical images, require a high amount of data during the training phase. Even if the architectural classification used is good enough to perform, models cannot yield good and reliable results if there is not enough data. Therefore, in future studies to be carried out after this study, studies will be made to increase the samples in small size datasets with deep learning architectures such as generative adversarial networks or convolutional autoencoder. By using these methods, the number of data can be increased to make the study more reliable. Thus, evaluation of similar classification models can be studied by using sufficient and larger data for classification. The step of increasing data by producing synthetic image data, which is planned to be carried out in future studies, is expected to significantly increase the reliability of deep learning models.

REFERENCES

- Aghdam, H. H. and Heravi, E. J. (2017). Guide to Convolutional Neural Networks, NY: Springer, New York, USA.
- Afshar, P., Mohammadi, A. and Plataniotis, K. N. (2018). "Brain tumor type classification via capsule networks." Proc., 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, pp. 3129-3133. IEEE.
- Bejnordi, B. E., Lin, J., Glass, B., Mullooly, M., Gierach, G. L., Sherman, M. E., Karssemeijer, N., van der Laak, J. and Beck, A. H. (2017). "Deep learning-based assessment of tumor-associated stroma for diagnosing breast cancer in histopathology images." Proc., IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, Australia, pp. 929-932. IEEE.
- Breiman, Leo. (2001). "Random forests." Machine Learning, Vol. 45, No. 1, pp. 5-32.
- Chakrabarty, N. (2019). Brain MRI Images for Brain Tumor Detection.
- Chen, X., Xu, Y., Wong, D. W. K., Wong, T. Y., and Liu, J. (2015). "Glaucoma detection based on deep convolutional neural network." Proc., 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, pp. 715-718.
- Chollet, F. (2017). "Xception: Deep learning with depthwise separable convolutions." Proc., IEEE Conference on Computer Vision and Pattern Recognition, Hawai'i Convention Center, Honolulu, Hawaii, USA, pp. 1251-1258.
- Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., and Greenspan, H. (2018). "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification." Neurocomputing, Vol. 321, 321-331.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep learning, MIT press, Massachusetts, USA.
- He, K., Zhang, X., Ren, S. and Sun, J. (2016). "Deep residual learning for image recognition." Proc., IEEE Conference On Computer Vision And Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770-778. IEEE.
- Indraswari, R., Kurita, T., Arifin, A. Z., Suciati, N. and Astuti, E. R. (2019). "Multi-projection deep learning network for segmentation of 3D medical images." Pattern Recognition Letters, Vol. 125, pp. 791-797.
- Kermany, D. K. and Goldbaum, M. (2018). Labeled optical coherence tomography (OCT) and Chest X-Ray images for classification. Mendeley Data, 2.
- Khobragade, S., Tiwari, A., Patil, C. Y. and Narke, V. (2016). "Automatic detection of major lung diseases using Chest Radiographs and classification by feed-forward artificial neural network." Proc., IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), Delhi, India, pp. 1-5. IEEE.
- Korolev, S., Safiullin, A., Belyaev, M. and Dodonova, Y. (2017). "Residual and plain convolutional neural networks for 3D brain MRI classification." Proc., IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, Australia, pp. 835-838. IEEE.

- Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks." Proc., Advances in Neural Information Processing Systems (NIPS), Lake Tahoe, Nevada, USA, pp. 1097-1105.
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). "Backpropagation applied to handwritten zip code recognition." Neural computation, Vol. 1, No. 4, pp. 541-551.
- Martinez-Murcia, F. J., Ortiz, A., Gorriz, J. M., Ramirez, J. and Castillo-Barnes, D. (2019). "Studying the Manifold Structure of Alzheimer's Disease: A Deep Learning Approach Using Convolutional Autoencoders." IEEE Journal of Biomedical and Health Informatics.
- Mikołajczyk, A., and Grochowski, M. (2018). "Data augmentation for improving deep learning in image classification problem." Proc., International Interdisciplinary PhD Workshop (IIPhDW), Szczecin, Poland, pp. 117-122.
- Oh, S. L., Hagiwara, Y., Raghavendra, U., Yuvaraj, R., Arunkumar, N., Murugappan, M., and Acharya, U. R. (2018). "A deep learning approach for Parkinson's disease diagnosis from EEG signals." Neural Computing and Applications, 1-7.
- Qian, Y., Bi, M., Tan, T., and Yu, K. (2016). "Very deep convolutional neural networks for noise robust speech recognition." IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 24, No. 12, pp. 2263-2276.
- Shahzadi, I., Tang, T. B., Meriadeau, F. and Quyyum, A. (2018). "CNN-LSTM: Cascaded framework for brain Tumour classification." Proc., IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), Sarawak, Malaysia, pp. 633-637. IEEE.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014). "Dropout: a simple way to prevent neural networks from overfitting." The Journal of Machine Learning Research, Vol. 15, No. 1, pp. 1929-1958.
- URL-1: <https://www.kaggle.com/navoneel/brain-mri-images-for-brain-tumor-detection/metadata> [Date of Accessed: 09.09.2019]
- URL-2: <https://www.kaggle.com/paultimothymooney/chestxray-pneumonia> [Date of Accessed: 20.07.2019].
- Wong, S. C., Gatt, A., Stamatescu, V., and McDonnell, M. D. (2016). "Understanding data augmentation for classification: when to warp?" Proc., International conference on digital image computing: techniques and applications (DICTA), IEEE, Canberra, Australia, pp.1-6.
- Vapnik, V., and Chappelle, O. (2000). "Bounds on error expectation for support vector machines." Neural Computation, Vol. 12, No. 9, pp. 2013-2036.
- Varshni, D., Thakral, K., Agarwal, L., Nijhawan, R. and Mittal, A. (2019). "Pneumonia detection using cnn based feature extraction." Proc., IEEE Third International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, Tamil Nadu, India, pp. 1-7. IEEE.
- Zhang, J., Xie, Y., Wu, Q. and Xia, Y. (2019). "Medical image classification using synergic deep learning." *Medical Image Analysis*, Vol. 54, pp. 10-19