Human Gender Prediction on Facial Images Taken by Mobile Phone using Convolutional Neural Networks

Mehmet HACIBEYOGLU*[1], Mohammed Hussein IBRAHIM[1]

Abstract: The interest in automatic gender classification has increased rapidly, especially with the growth of online social networking platforms, social media applications, and commercial applications. Most of the images shared on these platforms are taken by mobile phone with different expressions, different angles and low resolution. In recent years, convolutional neural networks have become the most powerful method for image classification. Many researchers have shown that convolutional neural networks can achieve better performance by modifying different network layers of network architecture. Moreover, the selection of the appropriate activation function of neurons, optimizer and the loss function directly affects the performance of the convolutional neural networks. In this study, we propose a gender classification system from facial images taken by mobile phone using convolutional neural networks. The proposed convolutional neural networks have a simple network architecture with appropriate parameters can be used when rapid training is needed with the amount of limited training data. In the experimental study, the Adience benchmark dataset was used with 17492 different images with different gender and ages. The classification process was carried out by 10-fold cross validation. According the experimental results, the proposed convolutional neural networks predicted the gender of the images 98.87% correctly for training and 89.13% for testing.

1. Introduction

This study addresses the problem of human gender classification from face photographs taken with mobile phones. In image processing, automatic human gender classification is an interesting subject, because gender contains very important and rich information about people's social activities [1]. Gender is the one of the most obvious and basic knowledge of human, but this knowledge opens the door to gather information estimates in different practical applications. In the gender classification process, gender of the individual is estimated by determining the distinct characteristics of the femininity and the masculinity [2]. Gender classification applications have great potential, especially in areas where artificial intelligence and image processing are used. For example, a computer system with gender classification function can be used in areas including human-computer interaction (HCI) [3] like mobile applications and video games [4], demographic surveys [5], security and surveillance industry [6], and social media applications [7].

A variety of studies on gender classification have been made in the literature up to the present. [8] trained a fully connected back-propagating neural network to differentiate gender in human faces and tested on a data set of 90 samples. The trained artificial neural network performed as well as the people. [9] applied principal component analysis to the three dimensional structure and gray level image data from laser-scanned human heads separately. They compared the quality of the information in the three-dimensional head and gray level image data. According to experimental results, they said that three-dimensional head data provide better gender classification than gray level images. [10] used a linear neural network to make human gender classification from the face

photographs. They trained the linear neural network on 80 male and 80 female full face images with the hair cropped. They tested the proposed method with the portion of the face images and they found a surprising interaction between the gender of the face and the partial image conditions. [11] used nonlinear support vector machines (SVMs) appearance based gender classification with low-resolution faces from the FERET faces database on 1,755 images. The performance of the SVMs was compared with well-known traditional pattern classifiers and large ensemble-radial basis function. According to the experimental results SVMs showed better performance than other methods. [12] proposed AdaBoost based gender classifier from a low resolution grayscale picture of their face. They achieved 80% accuracy in identification with less than 10 pixel comparisons and 90% accuracy with less than 50 pixel comparisons. [13] presented comprehensive comparison of the state-of-art gender recognition techniques. They performed critical analysis of different face-based gender classification techniques. They determined significant problems of gender classification like expression changes illumination effects, computational time and high data dimensions. They proposed a technique which includes preprocessing, feature extraction and classification steps. According to the experimental results, proposed technique is quite good and comparative more efficient to existing techniques.

[14] applied convolutional neural networks (CNNs) for age and gender classification. They proposed a simple convolutional net architecture on the Adience benchmark for age and gender estimation. They carried out CNNs dramatically outperformed current state-of-the-art methods. [15] utilized deep-convolutional neural networks (D-CNN) based on VGGNet architecture for gender recognition. They examined their study on the unfiltered image of the face for gender recognition. According to their experimental results, proposed method outplayed the current

[1] *Computer Eng., Necmettin Erbakan Uni., Konya – 42090, TURKEY*
* Corresponding Author: Email: hacibeyoglu@konya.edu.tr

advance updated methods. [16] proposed a hybrid system which includes CNNs and extreme machine learning for age and gender classification. They extracted the features from the input image using CNNs and they classified the obtained features with extreme machine learning. In the experimental study, they used two well-known benchmark datasets MORPH-II and Adience. According to experimental results, the proposed hybrid architecture outperformed other studies on the same datasets. For improving gender and age classification, [17] proposed a novel feedforward attention mechanism to discover the most informative and reliable parts of a given face. They used Adience, Images of Groups, and MORPH II benchmarks in the experimental study. According to the experimental results, including attention mechanisms enhances the performance of CNNs in terms of robustness and accuracy. [18] proposed a novel approach for gender prediction in mobile/smart phones using gait information. They extracted features from gait cycles using histogram of gradient descriptor. In the experiments they used a dataset consists of 654 gait data obtained from 109 instances. They argue that the proposed method achieves higher success than the existing methods. [19] developed a new cost function from the well-known softmax and sigmoid functions for concurrent and jointly fine-tuned deep neural networks to classify age and gender from speech and face images. They proposed a new deep neural network architecture. They argue that the framework of the proposed work can be applied to any classification/prediction problem such as image classification and language identification. [20] proposed a multi-agent system for gender and age classification from images. The multi-agent system carried out different techniques (Fisherfaces, Eigenfaces, LBP, ANN) and their combination with filters (Gabor, Sobel) for the acquisition and pre-processing of images. In the experiments, the proposed multi agent system operated in environments with hard conditions for image processing like sharpness, brightness, saturation or contrast.

In this paper, gender classification is performed from face photographs taken by mobile/smart phone cameras using convolutional neural networks. The rest of the paper is organized as follows. In Section 2, the structure of CNNs is given. In Section 3, the proposed CNNs for gender classification is explained. In Section 4, the experimental study is given and the paper is finalized with conclusions.

2. Convolution Neural Networks

CNNs, the most important architecture of deep learning networks, has been successfully used in many image classification and prediction problems [21,22]. [23] used deep CNNs to represent face space which illustrates the progress of automated face recognition. [24] proposed a novel hybrid CNNs and recurrent neural networks for facial emotion recognition. [25] used deep learning neural networks to classify brain tumors into for classes normal, glioblastoma, sarcoma and metastatic bronchogenic carcinoma tumors by using a dataset consists of 66 brain magnetic resonance images. [26] used deep CNNs for vehicle and face recognition. They used a vehicle dataset which consist of images from multiple perspectives and the proposed algorithm verified with deep learning framework Caffe. The proposed model which uses deep CNNs performed well in the experiments. The CNNs architecture consists of layers such as convolution and pooling, and these layers are obtained specific feature from the classified image. The fully connected layers and the classification layers are placed after the convolution and pooling layers. Figure 1 shows the basic structure of CNNs [22].
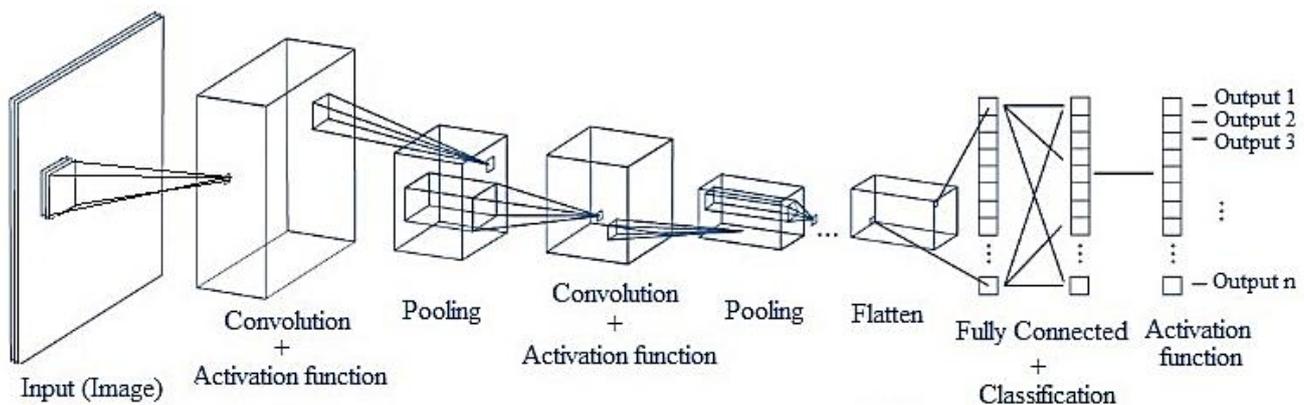

Fig. 1. The basic structure of convolution neural networks

2.1. Input Layer

Input layer is the first layer of CNNs. This layer is given at a certain size. The image size in this layer is very important for the success of the designed CNNs. Increasing the size of the incoming image data may increase the success of the system as well as the training process [27]. When the size of the image data is chosen to be low, the training process will be reduced but the success of the system may be reduced. When selecting the size of the incoming image data big, will increase the training process but may increase the success of the designed system. Therefore, it is useful if the image data is selected in a size suitable for the system being designed.

2.2. Convolution Layers

This layer is known as the characteristic extraction layer of CNNs, and the extraction process is performed by filtering the images at a certain size. Depending on the size of the images, the filters may be in different sizes such as 2x2, 3x3, 5x5, the output of this layer produces an output value after the activation functions Sigmoid, Tanh, and Rectified Linear Unit (ReLU) are applied. An example of convolution layers is shown in Fig. 2 below [28].
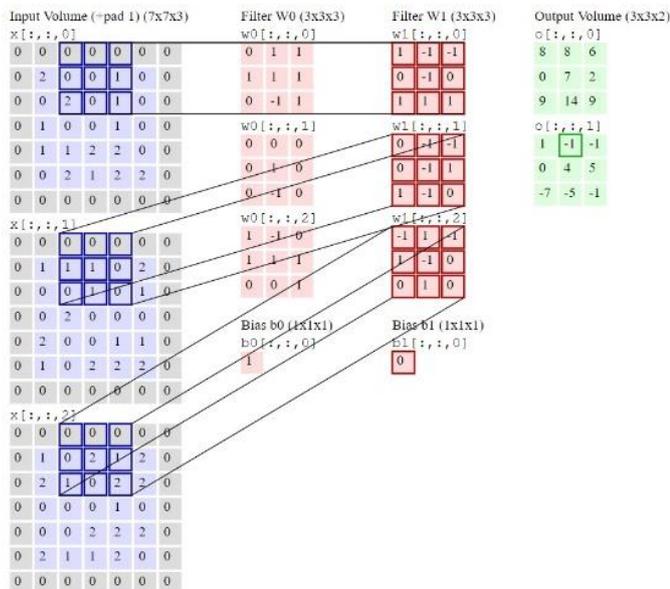
Fig. 2. An example of convolution layers

Fig. 3. Maximum pooling and average pooling with 2x2 filter in 4x4 input image

2.3. Pooling Layers

In CNNs, the pooling layer is often placed after the convolution layer, and the main purpose of this layer is to reduce the size of the input matrix for the convolution layer. As in the convolution layer, certain filters are defined in the pooling layer, these filters can be moved by a certain step on the image, and the values of the pixels in the image can be calculate in two different ways, average pooling and maximum pooling [27]. In maximum pooling, the maximum values of the pixels in the image are selected and in the average pooling, the average values of the pixels in the image are taken. There are two important parameters in the pooling layer, one of which is the filter size and the second is the stride parameter. Maximum pooling and average pooling is shown in Fig. 3 below [27].
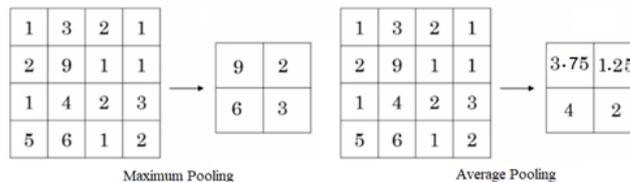
2.4. Fully Connected Layer

In the architecture of the CNNs, the fully connected layer is settled after the convolution and the pooling layers. This layer is connected by all nodes in the convolution or pooling layers, the number of layers can be various according to structure of the CNNs. The number of weights is equal to numbers of nodes in the convolution layers multiply the number of nodes in the connected layers [29].

2.5. Classification Layer

This layer of CNNs comes after the fully connected layer, the learning process of the CNNs takes place in this layer. This layer takes its inputs from the fully connected layer, and its outputs equal the number of objects to be classified [27].

3. Convolution Neural Networks for Gender Classification

In this work, a new CNNs architecture is designed for gender classification. Alexnet CNNs architecture [30] and similar works in the literature is taken as examples in the development of the new CNNs architecture. The number of layers and the number of nodes in each layer of the proposed CNNs architecture are determined according to the size of the input images. The used activation function, optimizer, and the dropout coefficients are determined empirically. The size of the person images is set to 100x100x3 before being given to the CNNs architecture. The proposed CNNs architecture for gender classification is shown in Table 1 below.

Table 1. The architecture of the CNNs for gender classification

Layer (type)	Output Shape	No. of Parameters
conv2d_1 (Conv2D)	(None, 32, 98, 98)	896
max_pooling2d_1 (MaxPooling2)	(None, 32, 49, 49)	0
conv2d_2 (Conv2D)	(None, 64, 49, 49)	18496
max_pooling2d_2 (MaxPooling2)	(None, 64, 24, 24)	0
conv2d_3 (Conv2D)	(None, 128, 24, 24)	73856
max_pooling2d_3 (MaxPooling2)	(None, 128, 12, 12)	0
conv2d_4 (Conv2D)	(None, 256, 12, 12)	295168
max_pooling2d_4 (MaxPooling2)	(None, 256, 6, 6)	0
flatten_1 (Flatten)	(None, 9216)	0
dense_1 (Dense)	(None, 256)	2359552
dense_2 (Dense)	(None, 256)	65792
dropout_1 (Dropout)	(None, 256)	0
dense_3 (Dense)	(None, 1)	257
activation_1 (Activation)	(None, 1)	0

When we look at to the table above, there are a total of 2,814,017 trainable parameters in CNNs architecture designed for gender classification.

All convolution layers and the first two fully connected layers are activated by an activation function named ReLU [31] which contributes a lot to the recent success of deep neural networks.

ReLU is a parameter-free activation function that can be defined as follows:

$$f(x) = \begin{cases} x, if\ x > 0 \\ 0, if\ x \le 0 \end{cases} \tag{1}$$

Sigmoid activation function is used for the last fully connected layer. Sigmoid activation function takes a real-value and outputs value in the range of 0 and 1 which can be defined as follows:

$$f(x) = \frac{1}{1+e^{-x}} \tag{2}$$

Binary crossentropy is used as loss function which can be defined as follows:

$$crossentropy(t, o) = -(t * \log(o) + (1 - t) * \log(1 - o)) \tag{3}$$

RMSProp with 0.0001 learning rate is used as an optimizer which works well in on-line and non-stationary settings [32]. RMSProp is gradient-based and adaptive learning optimization technique. It uses the average of moving gradients to normalize the gradient itself.

Dropout a technique is used to improve over-fit on neural networks, briefly dropout refers to ignoring neurons during the training phase of certain number of neurons which is chosen at random. Dropout layers is shown in Figure 4 below [22]. In this study, 0.25 rate dropout is used after the dense_2 layer of classification layer.
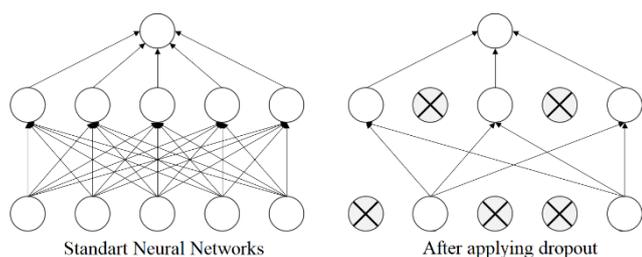

Fig. 4. Dropout layers

4. Experimental Study

We implemented our CNNs architecture using Phyton programming language on the Anaconda platform which is a free and open source distribution. Training and testing processes was performed on the Nvidia Geforce GT950mx graphics card. The experimental study was carried out with using the Adience benchmark dataset [33]. The Adience benchmark dataset is designed for age and gender classification which consist of face images automatically uploaded to Flickr website from smart phones. These images represent real-world/ social media images that have not been filtered or preprocessed. Many images in the Adience benchmark dataset have low resolution, extreme blur, expressions and poses in different variations for different age groups. Some example images of Adience benchmark dataset is shown in Figure 5.

Fig.5. The example images of Adience benchmark dataset

In the experimental study, a total of 17492 images including 8120 males and 9372 females were used. The classification process of CNNs was applied with 10-fold cross validation, which estimates the mean of the errors obtained on 10 different testing subsets. The training in each fold executed approximately 2650 seconds. The CNNs executed 50 epochs and measured classification accuracy and classification loss values in the training process are shown in the Figure 6.
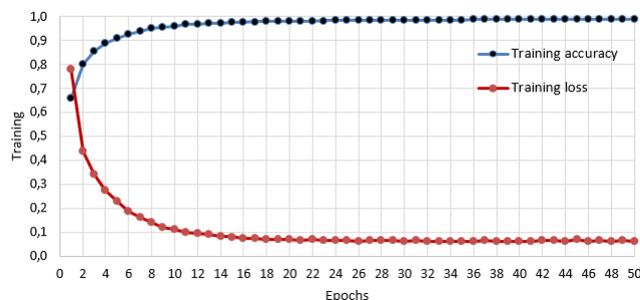

Fig.6. The convergence of classification accuracy and classification loss values during the training phase

When we look at Figure 6, it seems that the proposed CNNs regularly reduces the classification loss during the training phase. The classification loss of the proposed CNNs initialized at 0.78, it decreases continuously from the first epoch and decreases to 0.06 in the last epochs. The classification accuracy varies inversely proportional with classification loss during the training phase. The classification accuracy of the proposed CNNs initialized at 65.6%, it increases continuously from the first epoch and reaches 98.8% in the last epochs. In general, CNN's convergence figure shows that the system designed for gender classification is trained well without over-fitting and under-fitting. The trained CNNs is tested using images without gender label in the testing phase. According to the testing experimental result the proposed CNNs obtained the confusion matrix which is given in Table 2 below.

Table 2. The confusion matrix of testing phase

		Predicted	
		Male	**Female**
Actual	**Male**	7033	1087
	Female	814	8558

For classification problems accuracy, sensitivity, and specificity are important evaluation measures which can be calculated as shown below by using the confusion matrix.

$$Accuracy = TP + TN/TP + TN + FP + FN \tag{4}$$

$$Precision = TP/\ TP + FP \tag{5}$$

$$Specificity = TN\ /\ TN + FP \tag{6}$$

where:
- True positive (TP) is the number of males correctly classified
- True negative (TN) is the number of females correctly classified
- False positive (FP) is the number of females incorrectly

classified as male

• False negative (FN) is the number of males incorrectly classified as female

According to the equations 4-6, accuracy, precision and specificity evaluation measures are calculated 89.13%, 89.14%, and 88.8%, respectively. When the accuracy measure is considered the proposed CNNs correctly predicted the majority of the genders from images taken by mobile phones. Also, the values of precision and specificity measures seem to be close to the value of accuracy. This is important for the robustness of the proposed CNNs because precision shows how the proposed CNNs correctly classifies male instances and specificity indicates how the proposed CNNs correctly classifies female instances. The precision and specificity results show that the proposed CNNs predicted the male instances a little better than female instances. As a conclusion of the experimental study, the proposed CNNs showed good performance for gender prediction on facial images taken by mobile phones.

5. Conclusions and Future Work

In this study, we have proposed a novel CNNs architecture for gender classification from facial images taken by mobile phones. In the literature, many studies on gender classification have used images with high resolution which are prepared in the laboratory environment. But in the real world, most of the social media or identity authentication applications usually use face photographs taken from mobile phones. Therefore, in order to achieve the desired success of the proposed CNNs, the Audience data set including face images taken by mobile phone with different angles and expressions belonging various ages and genders was used. The proposed CNNs was trained in a short time with a small number of images and achieved a good classification accuracy according to the experimental study. For future work, the proposed CNNs will be used in mobile applications that include, gender classification, and social media statistics. In addition, novel CNNs will be developed to predict age or identity using the knowledge and experience gained from this study.

References

[1]   J Richard Udry. The nature of gender. Demography, 31(4):561–573, 1994

[2]   Wu, Yingxiao, et al. "Human gender classification: A review." arXiv preprint arXiv:1507.05122 (2015).

[3]   Beckwith, L., Burnett, M., Wiedenbeck, S., & Grigoreanu, V. (2006, May). Gender hci: Results to date regarding issues in problem-solving software. In AVI 2006 Gender and Interaction: Real and Virtual Women in a Male World Workshop paper (pp. 1-4).

[4]   Yu, S., Tan, T., Huang, K., Jia, K., & Wu, X. (2009). A study on gait-based gender classification. IEEE Transactions on image processing, 18(8), 1905-1910.

[5]   Hoffmeyer-Zlotnik, J. H., Hoffmeyer-Zlotnik, J. H., & Wolf, C. (Eds.). (2003). Advances in cross-national comparison: A European working book for demographic and socio-economic variables. Springer Science & Business Media.

[6]   Demirkus, M., Garg, K., & Guler, S. (2010, April). Automated person categorization for video surveillance using soft biometrics. In Biometric Technology for Human Identification VII (Vol. 7667, p. 76670P). International Society for Optics and Photonics.

[7]   Marquardt, J., Farnadi, G., Vasudevan, G., Moens, M. F., Davalos, S., Teredesai, A., & De Cock, M. (2014, January). Age and gender identification in social media. In Proceedings of CLEF 2014 Evaluation Labs (pp. 1129-1136).

[8]   Golomb, B. A., Lawrence, D. T., & Sejnowski, T. J. (1990, October). Sexnet: A neural network identifies sex from human faces. In NIPS (Vol. 1, p. 2).

[9]   O'toole, A. J., Vetter, T., Troje, N. F., & Bülthoff, H. H. (1997). Sex classification is better with three-dimensional head structure than with image intensity information. Perception, 26(1), 75-84.

[10]   Edelman, B., Valentin, D., & Abdi, H. (1998). Sex classification of face areas: How well can a linear neural network predict human performance?. Journal of Biological Systems, 6(03), 241-263.

[11]   Moghaddam, B., & Yang, M. H. (2002). Learning gender with support faces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5), 707-711.

[12]   Baluja, S., & Rowley, H. A. (2007). Boosting sex identification performance. *International Journal of computer vision*, *71*(1), 111-119.

[13]   Khan, S. A., Ahmad, M., Nazir, M., & Riaz, N. (2014). A comparative analysis of gender classification techniques. Middle-East Journal of Scientific Research, 20(1), 1-13.

[14]   Levi, G., & Hassner, T. (2015). Age and gender classification using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 34-42).

[15]   Dhomne, A., Kumar, R., & Bhan, V. (2018). Gender Recognition Through Face Using Deep Learning. *Procedia Computer Science*, *132*, 2-10.

[16]   Duan, M., Li, K., Yang, C., & Li, K. (2018). A hybrid deep learning CNN–ELM for age and gender classification. Neurocomputing, 275, 448-461.

[17]   Rodríguez, P., Cucurull, G., Gonfaus, J. M., Roca, F. X., & Gonzalez, J. (2017). Age and gender recognition in the wild with deep attention. Pattern Recognition, 72, 563-571.

[18]   Jain, A., & Kanhangad, V. (2018). Gender classification in smartphones using gait information. Expert Systems with Applications, 93, 257-266.

[19]   Qawaqneh, Z., Mallouh, A. A., & Barkana, B. D. (2017). Age and gender classification from speech and face images by jointly fine-tuned deep neural networks. Expert Systems with Applications, 85, 76-86.

[20]   González-Briones, A., Villarrubia, G., De Paz, J. F., & Corchado, J. M. (2018). A multi-agent system for the classification of gender and age from images. *Computer Vision and Image Understanding*.

[21]   Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117.

[22]   Aloysius, N., & Geetha, M. (2017, April). A review on deep convolutional neural networks. In Communication and Signal Processing (ICCSP), 2017 International Conference on (pp. 0588-0592). IEEE.

[23]   O'Toole, A. J., Castillo, C. D., Parde, C. J., Hill, M. Q., & Chellappa, R. (2018). Face Space Representations in Deep Convolutional Neural Networks. Trends in cognitive sciences.

[24]   Jain, N., Kumar, S., Kumar, A., Shamsolmoali, P., & Zareapoor, M. (2018). Hybrid deep neural networks for face emotion recognition. Pattern Recognition Letters.

[25]   Mohsen, H., El-Dahshan, E. S. A., El-Horbaty, E. S. M., & Salem, A. B. M. (2018). Classification using deep learning neural networks for brain tumors. Future Computing and Informatics Journal, 3(1), 68-71.

[26]   Luo, X., Shen, R., Hu, J., Deng, J., Hu, L., & Guan, Q. (2017). A deep convolution neural network model for vehicle recognition and face recognition. *Procedia Computer Science*, *107*, 715-720.

[27]   Özkan İ. & Ülker, E. Derin Öğrenme ve Görüntü Analizinde Kullanılan Derin Öğrenme Modelleri. Gaziosmanpaşa Bilimsel Araştırma Dergisi, 6(3), 85-104.

[28]   https://skymind.ai/wiki/convolutional-network#edge-detection,

Accessed on: Agu. 1, 2018.

[29] Sermanet, P., Chintala, S., & LeCun, Y. (2012, November). Convolutional neural networks applied to house numbers digit classification. In Pattern Recognition (ICPR), 2012 21st International Conference on (pp. 3288-3291). IEEE.

[30] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[31] Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th international conference on machine learning (ICML-10) (pp. 807-814).

[32] Tieleman, T., & Hinton, G. (2012). Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, *4*(2), 26-31.

[33] Eidinger, E., Enbar, R., & Hassner, T. (2014). Age and gender estimation of unfiltered faces. IEEE Transactions on Information Forensics and Security, 9(12), 2170-2179.