IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.13468

GENDER CLASSFICATION THROUGH FACIAL ANALYSIS

EDARA DEVENDRA SAI[1], JAGGUMAHANTHI PRASANTH [2], PALAPARTHI JOSEPH DINESH[3]

Koneru Lakshmaiah Education Foundation, Green fields, Vaddeswaram, Guntur, AP, India[1-3]

Abstract: Gender classification from facial photos is difficult due to the presence of a complex background, object occlusion, and varying lighting conditions. Face photos can be used for a variety of applications, including expression analysis, recognition, and tracking. This research investigates two deep learning-based approaches for gender classification using face photos. These approaches include CNN and Alex Net. Experiments were conducted to assess the effectiveness of both models in identifying male and female classes from facial photographs. The results indicate that both techniques were effective for gender classification. Additionally, a comparison study was carried out between these two models and a few well-known techniques for classifying gender.

Keywords: gender classification, gender recognition, CNN, Alex Net, Deep learning

I. INTRODUCTION

Gender is an essential component in interpersonal interactions in communities. With technological advancements, the use of smart devices has expanded, and social media has begun to draw everyone's attention. Daily studies on gender recognition have gained prominence and the number of applications using such techniques has expanded. Facial photos are commonly employed in such applications because they provide valuable information that can be used to extract human interaction.

The gender classification approach using facial photographs typically includes image manipulation, feature extraction, and classifying steps. These stages may differ based on the objective of the research and the characteristics of the methods to be utilized. As a result, the study's performance is significantly influenced by the classifier, method, and retrieved features.

Deep learning techniques are currently widely employed for a variety of applications, including classification, automatic feature extraction, object recognition, and so on, due to their high classification accuracy. Motivated by other domains, researchers have used deep learning approaches to predict and classify gender from facial photos. The next paragraphs provide a summary of several studies.

Jayaraman and Subramaniam (2019) set out to classify genders using various CNN architecture models. A dataset was generated using facial pictures taken from Malaysians and some Caucasians. Their results were 88% accurate with the VGG-16 model, 85% with the ResNet-50 model, and 49% with the Mobile Net model. Akbulut et al. (2014) used LRA-ELM and CNN architecture to recognize gender in facial pictures. The trials used around 11 thousand photos from the Audience dataset to identify age and gender. The proposed technique achieved 80% accuracy with LRA-ELM and 87.13% with CNN.

Gündüz and Decimole (2019) compared the suggested technique, Alex Net, and VGG16 models for gender classification, encompassing women, men, old, young, children, and newborns. The proposed CNN model achieved 72.20% accuracy, the VGG-16 model 99.41%, and the Alex Net model 65.63%.

Arora and Bhatia (2018) proposed a CNN model for gender classification using face pictures. In the experimental investigations, 1500 photos were used for training, and 1000 images from the CASIA database were confirmed. The experiments resulted in an accuracy of 98.5%.

Levi and Hassner (2015) suggested a simple convolutional network design to improve the performance of autonomous age and gender classification. This technique also produces optimal results with less training data. This model produced high classification accuracy on the Audience database (Eidinger et al., 2014).

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024
DOI: 10.17148/IJARCCE.2024.13468

Ranjan et al. (Ranjan et al., 2017) introduced Hyperface, a solution for simultaneous facial recognition, position prediction, and gender recognition based on CNNs. This approach merges the middle layers of a deep CNN with a separate CNN before applying a multitasking learning algorithm to the fused features. Hyperface is built on the ResNet-101 model to improve algorithm speed, and Fast Hyperface variations for zone suggestions have been presented.

Raza et al. (2018) suggested a deep learning algorithm for predicting the gender of pedestrians. A preprocessing step was employed to separate the pedestrian from the image. Then, stacked auto encoders with a SoftMax classifier were employed for classification. In the MIT dataset, accuracy rates of 82.9%, 81.8%, and 82.4% were attained in the anterior, posterior, and mixed views, respectively, while the PETA dataset achieved roughly 91.5% accuracy.

Abdulhadi and Aly (2020) presented an interchange of classical CNN models with the Canet model for gender classification. Furthermore, picante lowered the network design size in sophisticated CNN models. This approach achieved 89.65% accuracy in gender classification.

II. METHODOLOGY

a) DEEP LEARNING

Deep learning is an artificial intelligence (AI) technique that seeks to emulate the human brain by learning through experience. In other words, it is a strategy for realizing the learning process by attempting to uncover the hidden representation of the data (Goodfellow et al., 2016). These representations are learned through a training program. For example, to learn how to detect an object, we must train the computer using many object images labelled according to different classes. This instruction may take hours, days, or even weeks.

Deep learning-based algorithms require a huge quantity of training data and take longer to train than traditional machine learning methods. Finding unique attributes for any object or character in an image is a time-consuming and challenging task because the object or character has so many qualities. At this stage, unlike traditional machine learning, where features are extracted manually, issues can be solved using deep learning approaches that automatically extract relevant features from data. Deep learning is a neural network with numerous hidden layers, which can number hundreds of millions. After being taught on the network, an image can be used to create complex notions from simple ones. When a picture is trained in a network, it can learn to recognize items like characters, faces, and cars by combining simple features like shape, edges, and corners. As the image progresses through the layers, each layer learns a simple feature before moving on to the next layer one by one. As the number of layers grows, the network can learn more and more complex features and eventually combine them to predict the image.

b) CONVOLUTIONAL NEURAL NETWORKS (CNN)

CNN, one of the deep learning approaches, is a very powerful neural network. It is commonly used to solve difficulties in computer vision and image processing. CNNs may replace input data with trainable parameters in each layer while simultaneously making accurate guesses about the nature of the pictures.

CNN architecture consists of five basic types of neural layers: convolution, activation, pooling, fully connected, and dropout. Each layer type serves a particular purpose. Each layer of CNN converts the input volume to an output volume of neuron activation, which is then delivered to fully connected layers. Simpler features, such as edge information, are collected in the initial layers, and more complex characteristics characterizing the image are produced in deeper layers. The following section describes in detail the operations done on CNN layers.

CNN's foundation is its convolutional layer. The transformation process is carried out by rotating a filter of various sizes on the image, such as 3 * 3, 2 * 2. Filters use a convolution procedure on the images from the preceding layer to generate the output data, and consequently, the activation map is created. The resulting activation map can be described as zones in which each filter has its unique attributes. During the training process, the coefficients of the filters are adjusted for each learning in the training set, determining which parts of the data are significant in determining the features. Simple picture features, such as edges.

Activation Layer: The convolution layer's mathematical operations give the network a linear structure. The activation functions utilized in the activation layer cause the network to become nonlinear. As a result, the network learns more quickly. It is critical to select activation functions in a neural network architecture. The most frequent of these are the sigmoid function, which is typically employed in classification issues, SoftMax, which is an extension of the sigmoid function for multiple categories, and the Rectified Linear Units Layer (Reule), which is used as the activation function in most research.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.13468

Pooling Layer: In CNN topologies, the pooling layer is responsible for size reduction. Although size reduction may result in information loss, these losses are advantageous to the network. Because the reduction in size reduces the computational effort for the network's subsequent layers, it also prevents network overfitting. The two most widely used forms are average pooling and maximum pooling.

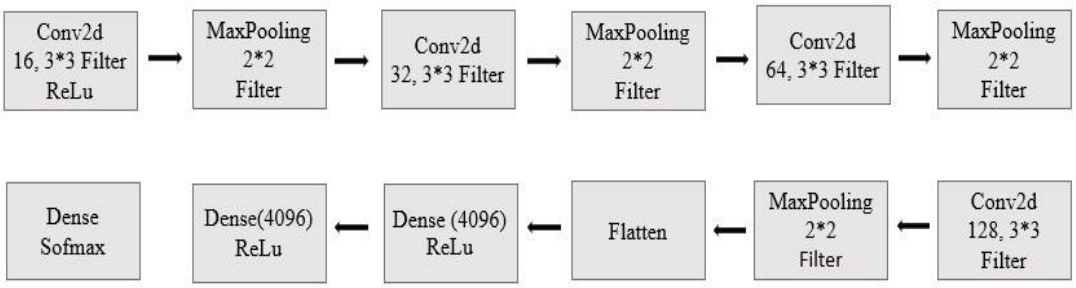
fully Connected Layer: Neurons in this layer are totally connected to all activations from the previous layer. As a result of these layers, two-dimensional feature maps are reduced to one-dimensional feature vectors. The generated vector can be classified into a set number of categories or utilized as a feature vector for additional processing.

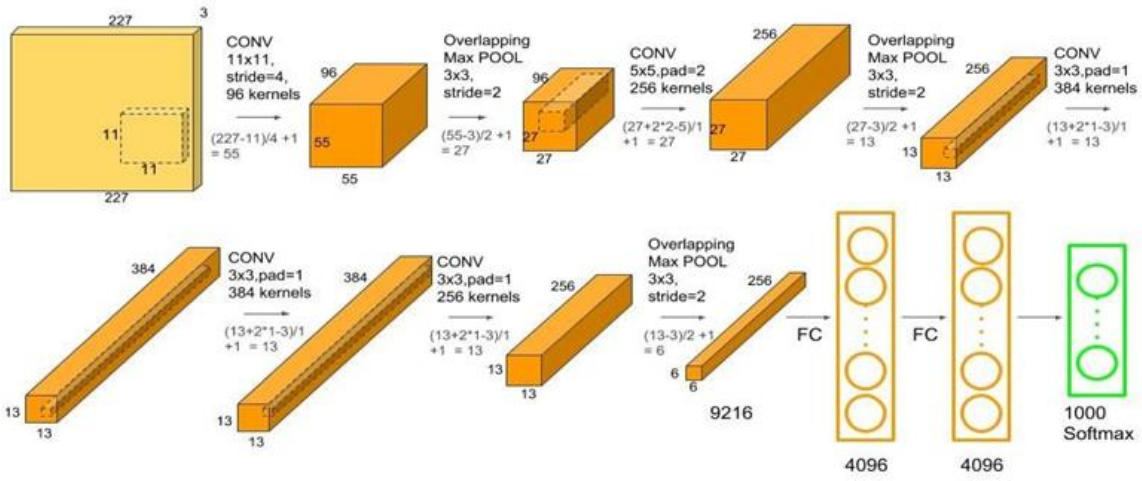
Dropout is one of the most utilized networking approaches in deep learning (Srivastava et al. 2014). When CNN is trained with large amounts of data, the network may become overfit. The basic logic, which involves eliminating some nodes from the network, avoids memorization.

The Image Net Big-Scale Visual Identification Competition (ILSVRC) is one of the world's largest competitions for object recognition. Alex Net, Le Net, ZF Net, VGG-16, Google Net, and Microsoft rest Net are the top-ranked CNN models. These models are critical for better understanding and improving deep learning and neural networks. As a result, it is the chosen method in many investigations.

Le Net: It is the model that produced the competition's first successful outcome and was released in 1998. Postal numerals were developed to make it easier to see the numbers on bank checks.

Alex Network: It is the winning model from the 2012 competition. Developed by Hrushevsky, Stukeley, and Hinton. Successive convolution layers are made up of completely connected layers using activation functions such maximal pooling, relu, and Sigmoid.





ResNet, the 2015 winning model, is the first to use Residual blocks and has 34 layers. This architecture stands out from others by focusing on a deeper design. Structure (He et al., 2016).

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.13468

III. EXPERIMENTS AND RESULT

This paper proposes a gender classification approach based on facial pictures and CNN models. The accuracy achieved using the CNN model and Alex Net architecture were: Compared to comparable research in the literature.

A. DATASET

KAGGLE (2018) conducted experiments with a dataset of 19,000 photos, comprising female, male, and child faces. The collection included unclear photos of children and babies with indeterminate gender, making it unsuitable for recognition. A total of 5000 photos were used for training, with 2500 female and 2500 male face images. A smaller dataset, the Audience dataset, was employed in this study. It consists of 26 thousand face photos of about 2,284 persons taken with smart phones and previously used in gender classification studies (Eidinger et al., 2014). Sample photos utilized in this study.

B. CNN MODELS

The study's data set was trained using two separate CNN models implemented in TensorFlow and Keres. The first of them is a CNN model with Conv2d, pooling, flatten, and thick layers, as well as RELU and Sigmoid activation functions. Figure 1 shows detailed information about the CNN model's layers. The other model was Alex Net. Figure 2 shows layer information for the Alex Net design. For both models, the dataset was split into 80% training and 20% testing. The proposed technique was assessed in terms of accuracy, precision, recall, and F1 score. The measures were derived from true positive (TP), true negative (TN), false positive (FP), and false negative (FN). These metrics are mathematically calculated as follows:

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

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

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

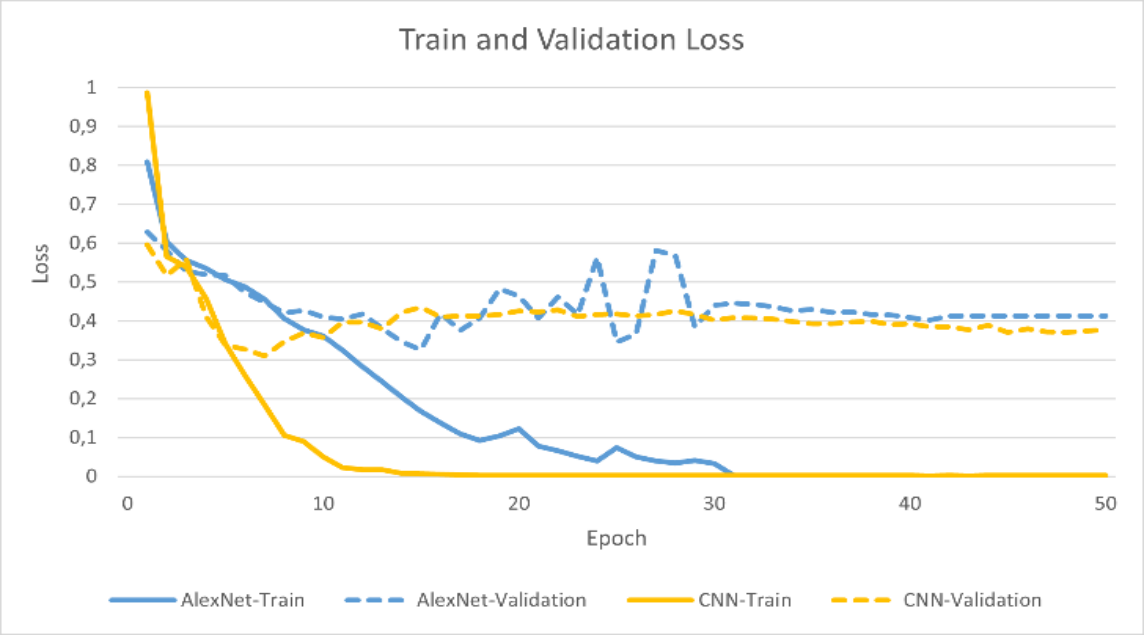
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.13468

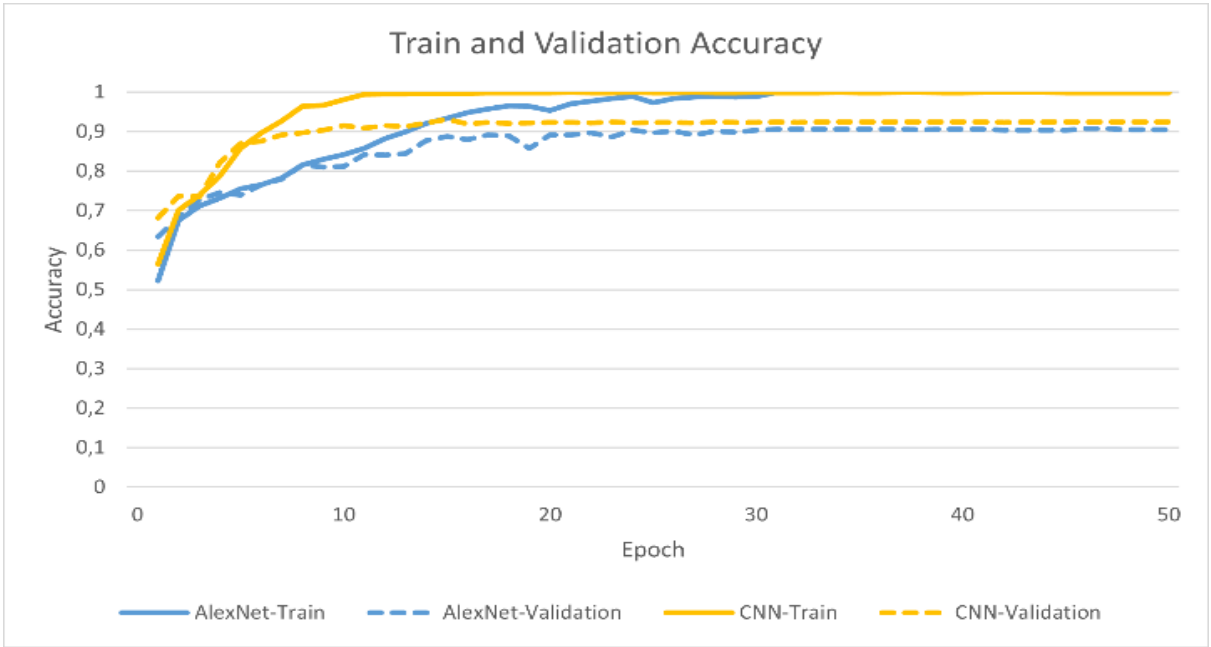
Accuracy is the ratio of accurately predicted data in the model to the entire dataset. Precision is the parameter that measures how much of the positively projected data is truly positive.

Remember: It is the metric that indicates how much of the data that should be predicted positively was forecasted favorably. F1Score: The harmonic mean of Precision and Recall values. As a result, it serves as an accurate predictor of success.

Table I summarizes the total results for both CNN and AlexNet. In terms of F1-score, both models yielded identical male and female classification results. For each class, CNN produced better results (92%).



Compared to Alex Net (90%). depicts the test and train accuracy obtained for the CNN and Alex Net models. The overall behaviour of both models appears comparable, as evidenced by the quantitative results. Similarly, depicts the training and testing losses for both models. The graphics show that overfitting did not occur during training due to the proximity of train and test accuracy scores. The loss values represent the sum of mistakes for each sample in the training and test sets. It begins with high values at the start of training and decreases as the model learns the data.



IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.13468

COMPARATIVE ANALYSIS OF PROPOSED METHOD WITH OTHER MODELS IN

Previous Works	CNN	Alex Net
Gündüz and Cedimoğlu [5]	72.20	65.63
Akbulut et al [4]	87.13	-
Levi and Hassner [6]	86.8	-
Yu at al [11]	91.50	-
Proposed	**92.40**	**90.50**

Additional experiments were carried out to see how the number of epochs affected the classifiers' validation accuracy. Table II summarizes the results obtained for various numbers of epochs.

The results show that the test accuracy for both the CNN and Alex Net models remained constant after a specific number of epochs (about 20). However, increasing the number of epochs beyond 20 increases processing time while not significantly improving accuracy.

ACCURACY OBTAINED FOR ALEXNET AND CNN (%)

Precision			Recall		F1-Score	
Classes	**Alex Net**	**CNN**	**Alex Net**	**CNN**	**Alex Net**	**CNN**
Male	89	91	92	93	90	92
Female	92	92	89	91	90	92

ACCURACY OBTAINED FOR DIFFERENT EPOCHS (%)

Epoch	***AlexNet***	***CNN***
10	81.10	90.40
20	90.20	92.20
30	89.80	92.30
40	90.50	92.40
50	**90.50**	**92.40**

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024
DOI: 10.17148/IJARCCE.2024.13468

IV. CONCLUSION

In this study, the Deep Learning models CNN and Alex Net are proposed for gender classification. Experiments were conducted on a dataset with fewer photos than those found in Audience, which had previously been used in gender recognition and classification experiments. Gender classification was achieved by combining the required models with the photos in the dataset. The accuracy rates acquired from experimental research were compared to the dataset's model performance.

The accuracy rates obtained using these approaches were compared to the accuracy rates of similar research in the literature, and improved results were found. Future studies will compare different CNN algorithms with a larger dataset of images.

REFERENCES

[1]. D. o. Heresbach, "A prospective multicenter study of back-to-back video colonoscopies revealed the miss rate for colorectal neoplastic polyps," Endoscopy, vol. 40, no. 04, pp. 284–290, 20.
[2]. D. Vazquez ´ et al., “A benchmark for endoluminal scene segmentation of colonoscopy images,” Journal of healthcare engineering, vol. 2017, 2017.
[3]. J. Bernal, F. J. Sanchez, G. Fern ´ andez-Esparrach, D. Gil, C. Rodr ´ ´ıguez, and F. Vilarino, “Wm-dova maps for accurate polyp highlighting in ˜ colonoscopy: Validation vs. saliency maps from physicians,” Computeri. Med. Imag. and Graph., vol. 43, pp. 99–111, 2015.
[4]. A. A. Pozdeev, N. A. Obukhova, and A. A. Motyko, “Automatic analysis of endoscopic images for polyps detection and segmentation,” in Proc. of EIConRus, 2019, pp. 1216–1220.
[5]. M. Akbari et al., “Polyp segmentation in colonoscopy images using fully convolutional network,” in Proc. of EMBC, 2018, pp. 69–72.
[6]. The article "Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy" was published in the National Biomedical Engineering Journal in 2018. P. Wang et al.
[7]. Y. B. Guo and B. Matuszewski, “Giana polyp segmentation with fully convolutional dilation neural networks,” in Proc. of VISIGRAPP, 2019, pp. 632–641.
[8]. M. Yamada et al., “Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy,” Scienti. repo., vol. 9, no. 1, pp. 1–9, 2019.
[9]. J. Poomeshwaran, K. S. Santhosh, K. Ram, J. Joseph, and M. Sivaprakasam, “Polyp segmentation using generative adversarial network,” in Proc. of EMBC, 2019, pp. 7201–7204
[10]. J. Kang and J. Gwak, “Ensemble of instance segmentation models for polyp segmentation in colonoscopy images,” IEEE Access, vol. 7, pp. 26 440–26 447, 2019.
[11]. O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical

