ABSTRACT

Gender classification is a critical task in computer vision. This task holds substantial importance in various domains, including surveillance, marketing, and human-computer interaction. In this work, the face gender classification model proposed consists of three main phases: the first phase involves applying the Viola-Jones algorithm to detect facial images, which includes four steps: 1) Haar-like features, 2) Integral Image, 3) Adaboost Learning, and 4) Cascade Classifier. In the second phase, four pre-processing operations are employed, namely cropping, resizing, converting the image from (RGB) Color Space to (LAB) color space, and enhancing the images using (HE, CLAHE). The final phase involves utilizing Transfer learning, a powerful deep learning technique that can be effectively employed to Face gender classification using the Alex-Net architecture. The performance evaluation of the proposed gender classification model encompassed three datasets: the LFW dataset, which contained 1,200 facial images. The Faces94 dataset contained 400 facial images, and the family dataset had 400. The Transfer Learning with the Alex-Net model achieved an accuracy of 98.77% on the LFW dataset. Furthermore, the model attained an accuracy rate of 100% on both the Faces94 and family datasets. Thus, the proposed system emphasizes the significance of employing pre-processing techniques and transfer learning with the Alex-Net model. These methods contribute to more accurate results in gender classification. Where, the results achieved by applying image contrast enhancement techniques, such as HE and CLAHE, were compared. CLAHE achieved the best facial classification accuracy compared to HE.

Keywords: Alex-Net, CLAHE, Deep learning, Gender Classification.
تصنيف الجنس بالاعتماد على الوجه باستخدام نموذج التعلم العميق

براق عبد الرضا حسن، فاتن عبد علي داود

قسم علوم الحاسبات، كلية العلوم، جامعة بغداد، بغداد، العراق

الخلاصة

تصنيف الجنس هو مهمة حرجة في رؤية الحاسوب. تحمل هذه المهمة أهمية كبيرة في مجموعة متنوعة من المجالات، بما في ذلك المراقبة والتسويق والتفاعل بين الإنسان والحاسوب. في هذه العمل نموذج تصنيف جنس الوجه المقترح من ثلاث مراحل رئيسية: المرحلة الأولى تطبيق خوارزمية فيولا-جونز لاكتشاف صور الوجه، والتي تشمل أربع خطوات: (1) Haar-Like Features، (2) Integral Image، (3) Adaboost Learning، (4) Cascade Classifier. المرحلة الثانية تشمل أربع عمليات ما قبل المعالجة، وهي القص، وتغيير الحجم، وتغيير الصورة من (RGB) إلى (LAB)، وتحسين الصورة باستخدام (HE,CLAHE). المرحلة النهائية تستخدم نقل التعلم، وهو تقنية قوية في التعلم العميق، يمكن استخدامها في تصنيف جنس الوجه باستخدام نموذج (Alex-Net).

تتضمن تقييم أداء النموذج المقترح لتصنيف الجنس ثلاث مجموعات بيانات: مجموعة بيانات LFW التي تحتوي على 1200 صورة وجه، ومجموعة بيانات Faces94 التي تحتوي على 400 صورة وجه. حقق نموذج Alex-Net الدقة بلغت 98.77% على مجموعة بيانات LFW. وعلاوة على ذلك، حقق النموذج معدل دقة بلغ 100% على كل من مجموعة بيانات LFW ومجموعة بيانات العائلة. وتباين النتائج المفترض في النتائج المحققة، يؤكِّد النظام المقترح على أهمية استخدام تقنيات التعلم العميق. النتائج التي تم الحصول عليها من خلال تطبيق تقنيات تحسين تباين الصور، مثل HE وCLAHE. حقق نموذج Alex-Net أفضل دقة في تصنيف الوجه مقارنةً بـ HE.

الكلمات المفتاحية: Alex-Net، التعلم العميق، تصنيف الجنس.

1. INTRODUCTION

The face is the essential focus of attention in society and is usually used as identification. The facial image is considered to be a reliable indicator for gender detection. Every human has a distinctive face that may be used to identify them from other people (Kanget al., 2017; Joodi et al., 2023).

The problem statement of gender classification from face image can be summarized into Pose Variations, Varying Illumination Conditions, and Background Variation. Gender classification plays a crucial role in demographic analysis and behavior tracking. Gender-aware systems also find application in social robotics and Human-Computer Interaction (HCI) (Lahariya et al., 2021). They worked to design a framework based on deep learning with transfer learning and fine-tuning for gender classification from face images, remove the background from the face image, and use it in the proposed model by applying the Viola-Jones algorithm, enhance the face image and use it in the proposed model by applying several preprocessing techniques, which are cropping, resizing, Images Convert from RGB to LAB, and Contrast enhancement using Contrast Limited Adaptive Histogram Equalization (CLAHE). Finally, using transfer learning Alex-Net to train the classification model for face images and evaluate the best classification result using Accuracy. In the literature review, several methods for gender classification from facial images have been proposed.

(Dhorne et al., 2018) proposed an efficient convolutional network architecture called a VGG-Net that may be applied in severe situations with a shortage of training data for learning a D-CNN based on this architecture. Particularly for gender identification, this model is
learned. Several face images from the celebrity and LFW datasets were used in the experiments. The test results have shown a gender classification accuracy of 95%.

(VenkateswarLal et al., 2019) suggested building a method for face recognition by feature descriptors and preprocessing face image techniques, which works well in natural settings. The mined color descriptors, texture, and preprocessed face images were copied and categorized using the support vector machine technique. The experimental consequences of the proposed method are presented using various data samples, such as LFW and FERET datasets. Due to using additional pre-processing and derived feature descriptors, the results indicate that the recommended strategy has admirable classification accuracy and a blend of pre-processing approaches. The average classification accuracy for the manipulated two data samples is 99% and 94%, respectively. (Galla et al., 2020) converted the visualized data to audio data to submit better sensation for blind persons by modeled feature extraction and classification. The proposed Feature extraction used the Multiscale-Invariant Feature Transform (MSIFT), while the support vector machine technique and the LASSO classifier fare used to perform feature optimization and type. The obtained classification results are illustrated as EN with an accuracy of 93.5%. LR with an accuracy of 93.2% and RR with a precision of 89.6%. (Zaman, 2020) propose for gender classification a custom CNN architecture based on face images into female and male. This CNN contains seven convolutional layers two fully connected layers, and between the convolutional layers are used batch normalization layers. Compared to Google-Net and Alex-Net, the number of parameters in bespoke CNN is three times and 30 times less, respectively. The proposed custom CNN also produced results on par with cutting-edge techniques, producing the top result with 96% accuracy on the CelebA dataset. Custom CNN trained on the CelebA dataset correctly classified Gender for the IMDB and WIKI datasets with 97% and 96% accuracy, respectively. The proposed method obtained an accuracy of 95% in LFW datasets.

(Azimi and Meghdadi, 2021) proposed a real-time deep neural network model that performs gender classification more quickly and with less computation by decreasing the model parameters and computation process. The proposed model is a combination of multifold filters and is relatively light. With an accuracy of 91.4%, 95%, 95%, and 95%, respectively, the performance of this proposed method was assessed in Adience, LFW, FEI, and CAS datasets. (Alamri et al., 2022) proposed using machine learning techniques in the LFW dataset for face detection and gender identification. The SVM classifier, the LBPH technique, and the SIFT feature extraction method were trained and evaluated utilizing the LFW dataset. The SIFT with an SVM classifier with an accuracy of 65% is reported, while face recognition with LBPH with a precision of 88% is addressed, exceeding earlier research. Moreover, the LBPH provided a gender detection accuracy of 91%.

This work aims to successfully develop an accurate and reliable gender classification system that can have numerous practical implications. It can contribute to Demographic analysis and crowd monitoring in surveillance applications. Moreover, accurate gender Classification can enable personalized interactions and tailored experiences in the human-computer interaction field. This work serves to remove the background from the face image. Enabling its utilization in the proposed model by applying the Viola-Jones Algorithm is also a target for this work. It also aims to address challenges related to lighting conditions and expression variations using preprocessing techniques.
2. MATERIALS AND METHODS

2.1 Proposed Model for Gender Classification

The proposed model aims to classify Gender from face images and consists of three main phases. In the first phase, the Viola-Jones algorithm detects the face image. In the second phase, four steps are used as pre-processing techniques, including cropping, resizing, Image Convert from RGB to LAB, and Images Enhancement. The final phase is the proposed model of gender classification deep learning using Transfer Learning Alex-Net. The detailed overall framework for Classification according to Gender is shown in Fig. 1.

![Image of the overall framework for the Proposed Gender Classification]

**Figure 1.** The overall framework for the Proposed Gender Classification

2.2 Data Collection

This step is performed for face image loading using different types of datasets. Brief descriptions of these datasets are given in the following sub-sections.
2.2.1 LFW dataset

LFW dataset is the most often used benchmark for gender classification, even though it was initially developed for unrestricted face recognition such as facial expression, background, lighting, Race, Gender, age, color, saturation, camera quality, clothing, hairstyles, and others. LFW contains 13,233 images of 5749 people (10,256 males and 2977 females) (Chen et al., 2017; Bajrami et al., 2018). The dataset may be accessed and downloaded at the following website: http://vis-www.cs.umass.edu/lfw/. The dataset was created and maintained by researchers at the University of Massachusetts. Fig. 2 shows Samples of LFW datasets.

2.2.2 Faces94 dataset

Each person has 20 images in total. 152 people are depicted, with a 180 by 200 pixel image resolution (portrait format). The collection has 399 images of females and 2,660 images of males (Thepade et al., 2018; Berbar, 2022). There are many images in this dataset. The dataset can be observed and obtained at the following web http://cswww.essex.ac.uk/mv/allfaces/faces94.html. Samples of datasets are shown in Fig. 3.

Three datasets of facial images are used to test and evaluate the performance of the proposed model for gender classification. The descriptions of these images are illustrated in Table 1.

<table>
<thead>
<tr>
<th>Type of dataset</th>
<th>Images.</th>
<th>Female Images</th>
<th>Male Images</th>
<th>Image type &amp; Format</th>
<th>Resolution of each image</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFW</td>
<td>13,233</td>
<td>2977</td>
<td>10,256</td>
<td>RGB JPG</td>
<td>250 × 250</td>
</tr>
<tr>
<td>Faces94</td>
<td>3,059</td>
<td>399</td>
<td>2,660</td>
<td>RGB JPG</td>
<td>180 × 200</td>
</tr>
<tr>
<td>Real data</td>
<td>400</td>
<td>200</td>
<td>200</td>
<td>RGB JPG</td>
<td>250 × 250</td>
</tr>
</tbody>
</table>

2.2.3 Real dataset

The real dataset is collected from the family individual for unconstrained face classification such as facial expression, background, lighting, race, Gender, age, color, saturation, camera quality, clothing, hairstyles, and others. Real contains 400 images (200 males and 200 females). Face parts in the images were detected through the Viola-Jones face detector. Samples of the real datasets are shown in Fig. 4.

Figure 2. Samples of images from the dataset
2.3 Face Detection Using Viola-Jones Algorithm

Face detection aims to extract the proposed system's region of interest (ROI), a person's face. This detection is used for further processing by removing features and classifying them. The algorithm that is used for face detection is the Viola-Jones algorithm.

It is a face-detection technique that produces speedy, accurate, and effective results (Chaudhari et al., 2018; Suma and Raga, 2018). The Viola-jones algorithm consists of four main steps as follows:

- **Haar-Like Features**: The majority of the Features of human faces are identical. We can match their attributes using the Haar-like features that aid in extracting information from the input image (Heydarzadeh et al., 2010; Hassan and Dawood, 2022).
- **Integral Image**: An integral image calculates Haar-like features relatively quickly (Zhang et al., 2010; Hassan and Dawood, 2022).
- **Adaboost**: The main idea is to train many weak classifiers and then associate between these weak classifiers to acquire a strong classifier (Peng et al., 2018; Hassan and Dawood, 2022).
- **Cascade Classifier**: After choosing the greatest qualities in each window, now must choose which of these windows contains faces. Therefore, only the area that completes all steps is categorized as a face. The accuracy of detection increases with the number of steps in the cascade classifier (Kang et al., 2016; Hassan and Dawood, 2022).

2.4 Face Image Pre-processing

The pre-processing step enhances the image quality to achieve the best results. A detailed explanation of the pre-processing steps applied to the input images follows.
2.4.1 Image Cropping

The image’s cropping is done to reduce the quantity and keep only the face region (Ciocca et al., 2007). The cropping process is done by taking the values from each side of the face bounding box that has been formed. The result of Image Cropping is illustrated in Fig. 5.

![Image Cropping Example](image)

**Figure 5.** *a*) Image before Cropping, *b*) Image after Cropping

2.4.2 Image resizing

In this step, the image size is resized before it enters the Alex-Net. The image must be the same size as the other images in the database. After the cropping, the images are of different sizes varying from 130x118 to 132x120 pixels (Unal et al., 2016). Therefore, all photos in the database were resized to a constant value (227x227) to reduce the calculation cost and complexity of the problem.

2.4.3 Image Converted from RGB to L*A*B*

The International Commission on Illumination first identified and defined this color space using CAE. One channel in this color space is designated for Luminance (Lightness), while channels a and b are chromaticity layers. Regarding the red-green and blue-yellow axes, the layer color is represented by the a* and b* layers, respectively (KaviNiranjana et al., 2015). The cube root of the relative luminance is used to determine the lightness correlation in CIELAB. Additionally, since uniform changes in L*A*B* color space’s Components are intended to match consistent differences in perceived color, it is possible to perceive every color as a dot in a three-dimensional space, allowing one to estimate the comparative perceptual variances between any two L*A*B* hues and calculating the Euclidean distance between them (Rathore et al., 2012). The process of converting images from RGB to L*A*B*, using Eq. (1).

\[
\begin{align*}
L^* &= 116 \left( \frac{Y}{Y_n} \right)^{\frac{1}{3}} - 16 \\
a^* &= 500\left[ f \left( \frac{X}{X_n} \right) \right. \\
b^* &= 200\left[ f \left( \frac{Y}{Y_n} \right) - f \left( \frac{Z}{Z_n} \right) \right]
\end{align*}
\]  

(1)

\[X, Y, Z, X_n, Y_n, \text{and } Z_n \text{ are the coordinates of CIEXYZ color space. For converting images to the CIEXYZ from the RGB, use Eq. (2).} \]
\[
\begin{bmatrix}
X \\ Y \\ Z
\end{bmatrix} =
\begin{bmatrix}
0.608 & 0.174 & 0.201 \\
0.299 & 0.587 & 0.114 \\
0.000 & 0.066 & 1.117
\end{bmatrix}
\begin{bmatrix}
R \\ G \\ B
\end{bmatrix}
\] (2)

Xn, Yn, and Zn correspond to the white value of the parameter, Eq. (3).

\[
f(x) = \begin{cases} 
    x^\frac{1}{3} & x > 0.008856 \\
    7.787x + \frac{16}{116} & x \leq 0.008856 
\end{cases}
\] (3)

Converting color reproduction from one basis to another is converting color space. It happens when transforming an image shown in a single color space to a different color space (Rathore et al., 2012).

2.4.4 Facial Images Enhancement Techniques

Image enhancement is the process that converts the image at the input into one that is more suited for the needed application. In this proposal, techniques are used, such as HE and CLAHE for image contrast enhancement.

- **Histogram equalization (HE)**

  Histogram equalization is the method that re-distributes all pixel values to be as close to the intended histogram as possible (Benitez-Garcia et al., 2011). Equalization of the histogram allows for increased contrast in areas of low local contrast. It automatically produces a transformation function to provide an image with a consistent histogram at the output. The equalization of the histogram is a technique for adjusting the contrast of an image by utilizing the histogram of the image. Usually, this method is applied to enhance the overall contrast of many images, primarily when near-contrast values represent the image's valid data (Musa et al., 2018). As a result, the intensities can be more evenly dispersed across the histogram. To accomplish this, histogram equalization is used to efficiently spread out the most common intensity values in the data. (Fadhil and Dawood, 2021). The color image and its histogram are shown in Fig. 6, followed by the equalized image and its histogram.

![Images](image1.png)  

**Figure 6.** (a): Original Image, Histogram of the original image.  
(b) Image enhancement using the HE method, Histogram of HE image
Contrast Limited Adaptive Histogram Equalization (CLAHE)
CLAHE is a well-known technique for enhancing local contrast that has proven effective and practical in several applications (Dawood and Abood, 2018). The fundamental concept of CLAHE is to conduct histogram equalization on non-overlapping portions of the image, utilizing interpolation to fix border discrepancies (Musa et al., 2018). Moreover, CLAHE features two crucial hyper-parameters: the clip limit (CL) and the number of tiles (NT). The first one, called CL, is a numerical number that regulates noise augmentation. Once the histogram for each sub-area has been determined, it is redistributed so that the height does not exceed a chosen “clip limit." To execute the equalization, the cumulative histogram is then constructed (Wu et al., 2017). The second (NT) integer value determines the number of equal-sized non-overlapping sub-areas; depending on its value, the image is divided into several sub-areas of this kind. The proposed method applies CLAHE to LAB color space to improve face images. Fig.7 shows the color image and histogram, then indicates the CLAHE image, which is the histogram of the CLAHE image.

![Figure 7](image)

**Figure 7.** (a): Original Image, Histogram of the original image. (b): Image enhancement using the CLAHE method, Histogram of CLAHE image.

### 2.5 Gender Classification

Classification is the final phase in the proposed model. This phase requires enhancing the performance of gender classification from face images by using Transfer Learning: Alex-Net. In this section we will discuss the basic methods of data needed to train Alex-Net. Alex-Net is given the samples in the form of a fixed-size image. The input image is colored and sized $227 \times 227 \times 3$.

#### 2.5.1 Transfer Learning Alex-Net Model

A practical method for resolving a classification issue is transfer learning. In this work, we applied Transfer Learning Alex-Net for gender Classification based on the face image (Anusriet et al., 2021). As shown in Fig. 8, The Alex-Net architecture contains three pooling, five convolutional layers, and three fully connected layers (Rafique et al., 2021). We adopted a modified Alex-Net to fit our issue domain. The final layer fully connected layer is swapped out with another layer fully connected layer to classify two classes (female or male) instead of 1000 classes. The parameters of the layers are given in Table 2.
Convolution layer

The convolution layer (CL) applies to original images using a unique convolution kernel feature detector. The image is scaled to $227 \times 227 \times 3$ during training. To create new pixels via convolution, 96 numbers of $11 \times 11 \times 3$ convolution kernels are applied in the first convolution layer. The $27 \times 27 \times 96$ pixel layer produced from the first layer serves as the second layer's input data. Each pixel's edges are filled with two pixels to help with further processing. The computing formula for the second layer's convolution kernel is the same as that for the first layer, with a size of $5 \times 5$ and a stride of 1 pixel. A collection of $13 \times 13 \times 128$ pixel layers produced by the second layer serves as the third layer's input data.

Pooling layer

Max pooling is primarily used to select the maximum value in the array. The max pooling does not impact the judgment when several pixels translate the image. The main benefit of Max pooling is noise reduction. The first, second, and fifth convolution layers are followed by the maximum pooling layers (Lin et al., 2020). The convolution layer's calculation using Eq. (4):

$$ o = \frac{(1-KS + 2P)}{s} + 1 $$

(4)

Fully connected layer

Alex-Net, There are three completely interconnected levels. The first and second layers each contain 4096 convolution kernels, with a size of $6 \times 6 \times 256$. That is due to the exact match between the size of the convolution kernel and the processing feature map. Only one-pixel value from the feature map's size and a one-to-one correlation are multiplied by each coefficient in the convolution kernel. As a result, are 4096 neurons in the pixel layer after convolution, which is $4096 \times 1 \times 1$. This study has two classes; thus, two neurons in the third ultimately linked layer are tuned to one type each (Lin et al., 2020). The original Alex-Net requires an input RGB image of dimensionality $227 \times 227$, in line with the increased dimensionality of images. In addition, the Alex-Net has an output layer with thousands of neurons corresponding to each type of Image-Net internal object.

As a result, the last three layers were modified as part of the debugging phase to address the gender classification problem from the facial image to (male/female). These layers are replaced with a full connection (fc), Softmax, and output layers. By implementing this fine-tuning phase, the network can infer the necessary bias and specificity of the data, which
enhances the ability of the network to make strong discrimination. In addition, to counteract overfitting, Two dropout layers (50% randomly) were added to the grid due to the limited size of the data set. Fig. 9 shows the implementation details of the entire Alex-Net network layers after modifications.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Number of Kernels</th>
<th>Kernel Size</th>
<th>Stride</th>
<th>Padding</th>
<th>Output Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>227×227×3</td>
</tr>
<tr>
<td>Conv1</td>
<td>96</td>
<td>11×11×3</td>
<td>4</td>
<td>0</td>
<td>55×55×96</td>
</tr>
<tr>
<td>Relu1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>55×55×96</td>
</tr>
<tr>
<td>Norm1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>55×55×96</td>
</tr>
<tr>
<td>Max pool1</td>
<td>-</td>
<td>3×3</td>
<td>2</td>
<td>0</td>
<td>27×27×96</td>
</tr>
<tr>
<td>Conv2</td>
<td>256</td>
<td>5×5×48</td>
<td>1</td>
<td>2</td>
<td>27×27×256</td>
</tr>
<tr>
<td>Relu2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>27×27×256</td>
</tr>
<tr>
<td>Norm2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>27×27×256</td>
</tr>
<tr>
<td>Max pool2</td>
<td>-</td>
<td>3×3</td>
<td>2</td>
<td>0</td>
<td>13×13×256</td>
</tr>
<tr>
<td>Conv3</td>
<td>384</td>
<td>3×3×256</td>
<td>1</td>
<td>1</td>
<td>13×13×384</td>
</tr>
<tr>
<td>Relu3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>13×13×384</td>
</tr>
<tr>
<td>Conv4</td>
<td>384</td>
<td>3×3×192</td>
<td>1</td>
<td>1</td>
<td>13×13×384</td>
</tr>
<tr>
<td>Relu4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>13×13×384</td>
</tr>
<tr>
<td>Conv5</td>
<td>256</td>
<td>3×3×192</td>
<td>1</td>
<td>1</td>
<td>13×13×256</td>
</tr>
<tr>
<td>Relu5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>13×13×256</td>
</tr>
<tr>
<td>Max pool5</td>
<td>-</td>
<td>3×3</td>
<td>2</td>
<td>0</td>
<td>6×6×256</td>
</tr>
<tr>
<td>Fc6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1×1×4096</td>
</tr>
<tr>
<td>Relu6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1×1×4096</td>
</tr>
<tr>
<td>Dropout6 (50%)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1×1×4096</td>
</tr>
<tr>
<td>Fc7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1×1×4096</td>
</tr>
<tr>
<td>Relu7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1×1×4096</td>
</tr>
<tr>
<td>Dropout7 (50%)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1×1×4096</td>
</tr>
<tr>
<td>fc8</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1×1×2</td>
</tr>
<tr>
<td>Softmax</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1×1×2</td>
</tr>
<tr>
<td>Class output</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Cross entropy with 'male' and 'female' other class</td>
</tr>
</tbody>
</table>

3. EVALUATION METRICS

A confusion matrix can be used to investigate how well a classification algorithm can do gender recognition from a face image. Table 3 illustrates the confusion matrix. The confusion matrix’s TP and TN values represent the proportion of accurate positive and negative classifications (Goodfellow et al., 2016). At the same time, FP and FN stand for the balance of incorrectly classified negative and positive instances, respectively. Several commonly used metrics can be presented from the confusion matrix for evaluating the classifier performance with varying concentrations of evaluations (Vujović, 2021; Al Jibory et al., 2022).
**Figure 9.** Conceptual of Transfer Learning Alex-Net model.

**Table 3.** Confusion Matrix

<table>
<thead>
<tr>
<th>Predicted condition</th>
<th>True condition</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total population</td>
<td>Condition Positive</td>
<td>Condition Negative</td>
</tr>
<tr>
<td>Predicted condition positive</td>
<td>True Positive</td>
<td>False Positive</td>
<td>PP</td>
</tr>
<tr>
<td>Predicted condition negative</td>
<td>False-negative</td>
<td>True negative</td>
<td>PN</td>
</tr>
</tbody>
</table>

**Accuracy:** This is seen in the ratio of real classifications to all sample tests used during training and testing. It is computed using Eq. (5) (Vujović, 2021; Alfarhany and Abdullah, 2023):

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

The volume of accurate predictions that have been tallied is used to determine how well the gender categorization from the facial image technique performs.
- **Recall or sensitivity ratio**: the ratio of true positive cases correctly predicted positive (Mateenet et al., 2020). This rate is calculated by dividing the images (male and female) correctly detected for a total of images (male and female) and false negative detected correctly, using Eq. (6)

\[
\text{Recall(Sensitivity) } TPR = \frac{TP}{TP+FN} = \frac{TP}{RP}
\]  

(6)

- **Precision or confidence ratio**: Precision represents the proportion of predicted positive cases correctly classified as positive. It is calculated by dividing the number of correctly detected images by the sum of false-positive and correctly detected images (Alkentar et al., 2021), using Eq. (7)

\[
\text{Precision (Confidence) } TPA = \frac{TP}{TP+FP} = \frac{TP}{PP}
\]  

(7)

- **F1-score**: it is the average of the weighted values for precision and recall. As a result, this score considers both false positives and false negatives, using Eq.(8) (Vujović, 2021; Ghadi and Salman, 2022):

\[
\text{F1score} = \frac{2 \times \text{Recall \times Precision}}{\text{Recall + Precision}}
\]  

(8)

4. RESULTS AND DISCUSSION

The proposed model was implemented using MATLAB (R2020a). The computer system specifications are as follows: Processor Intel Core i7-10510U CPU @ 1.80GHz, Memory size: 8 GB, Type of Operating system 64-bit. Before entering the data into the Transfer Learning Alex-Net model, it was divided into main parts: training constitutes 80%, and testing includes 20%.

Another set was taken from the validation set, consisting of 16% of the total training sets. This division is commonly employed to ensure proper evaluation and optimization of the model’s performance as given in Table 4. The training and validation sets were used during the model’s training, while the test set was used for the final model performance. In the proposed model, three datasets were used for training separately using the same metrics in training face image datasets with the Transfer Learning Alex-Net model.

The training was done for the model for 20 epochs with a batch size of 20, and the learning rate is set to 0.0003. Three cases will be illustrated.

Case 1: Results of training without Preprocessing for three datasets, as shown in Table 5.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>No. of Training Images</th>
<th>No. of Validation Images</th>
<th>No. of Testing Images</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFW</td>
<td>768</td>
<td>192</td>
<td>240</td>
<td>1200</td>
</tr>
<tr>
<td>Faces94</td>
<td>256</td>
<td>64</td>
<td>80</td>
<td>400</td>
</tr>
<tr>
<td>Real data</td>
<td>256</td>
<td>64</td>
<td>80</td>
<td>400</td>
</tr>
</tbody>
</table>
Table 5. Results of training without Preprocessing for three datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Training Accuracy</th>
<th>Validation Accuracy</th>
<th>Training Loss</th>
<th>Validation Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFW</td>
<td>100.00%</td>
<td>92.19%</td>
<td>0.0051</td>
<td>0.2481</td>
</tr>
<tr>
<td>Faces94</td>
<td>100.00%</td>
<td>95.00%</td>
<td>0.0515</td>
<td>0.0481</td>
</tr>
<tr>
<td>Real data</td>
<td>100.00%</td>
<td>95.31%</td>
<td>0.0088</td>
<td>0.0650</td>
</tr>
</tbody>
</table>

Case 2: Results of training with Preprocessing using Histogram equalization (HE) for three datasets, as shown in Table 6.

Table 6. Results of training with Preprocessing using (HE) for three datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Training Accuracy</th>
<th>Validation Accuracy</th>
<th>Training Loss</th>
<th>Validation Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFW</td>
<td>100.00%</td>
<td>91.15%</td>
<td>0.0037</td>
<td>0.2168</td>
</tr>
<tr>
<td>Faces94</td>
<td>100.00%</td>
<td>98.44%</td>
<td>0.0006</td>
<td>0.1413</td>
</tr>
<tr>
<td>Real data</td>
<td>100.00%</td>
<td>98.44%</td>
<td>0.0015</td>
<td>0.0766</td>
</tr>
</tbody>
</table>

Case 3: Results of training with Preprocessing using CLAHE for three datasets, as shown in Table 7.

Table 7. Results of training with Preprocessing using (CLAHE) for three datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Training Accuracy</th>
<th>Validation Accuracy</th>
<th>Training Loss</th>
<th>Validation Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFW</td>
<td>100.00%</td>
<td>98.96%</td>
<td>0.0326</td>
<td>0.0259</td>
</tr>
<tr>
<td>Faces94</td>
<td>100.00%</td>
<td>100.00%</td>
<td>0.0031</td>
<td>0.0015</td>
</tr>
<tr>
<td>Real data</td>
<td>100.00%</td>
<td>100.00%</td>
<td>0.0005</td>
<td>0.0011</td>
</tr>
</tbody>
</table>

5. CONCLUSIONS

The proposed method is used to classify the Gender from the face image, whether (male or female), before inputting the face image in Transfer Learning Alex-Net. We apply the Viola-Jones algorithm for face images to detect faces and exclude the background. After the face detection in images, a pre-processing step is used, such as cropping, resizing, Images Convert from RGB to LAB, and Contrast enhancement using (CLAHE). Deep learning techniques are the most common for gender detection, aiming to learn essential features from raw data automatically. Alex-Net-based Deep Learning model was proposed for gender classification tasks. A comparison has been made using Alex-Net with the dataset before and after preprocessing, where Alex-Net was applied to three datasets. The Alex-Net model delivered the best accuracy after preprocessing in LFW, Faces94, and Real datasets, where 98.77 %, 100%, and 100%, respectively.
NOMENCLATURE

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A*</td>
<td>A(green-red)</td>
<td>CNN</td>
<td>Convolutional Neural Networks</td>
</tr>
<tr>
<td>a*</td>
<td>The red-green color component of the LAB color space. It ranges from -128 to 127, where negative values indicate greenness and positive values indicate redness.</td>
<td>D-CNN</td>
<td>Deep Convolution neural networks</td>
</tr>
<tr>
<td>b*</td>
<td>The yellow-blue color component of the LAB color space. It also ranges from -128 to 127, where negative values indicate blueness and positive values indicate yellowness.</td>
<td>FC</td>
<td>Fully Connected Layer</td>
</tr>
<tr>
<td>B*</td>
<td>B(blue-yellow)</td>
<td>FN</td>
<td>False Negative</td>
</tr>
<tr>
<td>I</td>
<td>The size of the input image</td>
<td>FP</td>
<td>False Positive</td>
</tr>
<tr>
<td>KS</td>
<td>The dimensions of the convolutional kernel</td>
<td>HCI</td>
<td>Human-Computer Interaction</td>
</tr>
<tr>
<td>L*</td>
<td>The lightness component of the LAB color space. It ranges from 0 to 100, where 0 represents black and 100 represents white.</td>
<td>HE</td>
<td>Histogram Equalization</td>
</tr>
<tr>
<td>o</td>
<td>The output size after applying the convolution operation</td>
<td>LBPH</td>
<td>Local Binary Pattern Histogram</td>
</tr>
<tr>
<td>P</td>
<td>The number of pixels added around the input to control the spatial dimensions and preserve information during convolution.</td>
<td>LFW</td>
<td>Labeled Faces in the Wild</td>
</tr>
<tr>
<td>s</td>
<td>The step size or the number of pixels the kernel moves during the convolution operation</td>
<td>Relu</td>
<td>Rectified Linear Unit Function</td>
</tr>
<tr>
<td>X, Y, and Z</td>
<td>The amounts of the three primary colors (red, green, and blue) in a given color, while Xn, Yn, and Zn represent the reference white point.</td>
<td>RGB</td>
<td>Red, Green, Blue Color Space</td>
</tr>
</tbody>
</table>

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROI</td>
<td>Region Of Interest</td>
</tr>
<tr>
<td>TN</td>
<td>True Negative</td>
</tr>
<tr>
<td>TP</td>
<td>True Positive</td>
</tr>
</tbody>
</table>

REFERENCES


121


