Foundation University Journal of Engineering and Applied Sciences
Vol. 3, Issue 1, 2022.
DOI: 10.33897/fujeas.v3i1.506

Research Article

Facial Based Gender Classification for Real Time Applications

Anmol Haider[1], Muhammad Imran[1*]

[1]Shaheed Zulfiqar Ali Bhutto Institute of Science and Information Technology, Islamabad, Pakistan

[*]Corresponding Author: Muhammad Imran. Email: dr.imran@szabist-isb.edu.pk

Abstract:

Appearance and facial features play an important role in gender recognition through images. For gender classification, multiple techniques were presented to acquire better results in which preprocessing part is one of the major and very important for gender classification as it removes noise, enhances, images, and eliminates any unnatural colors from an image. Another major aspect is the efficient feature extraction method. If features extracted accurately then the result of classification will improve. Over the past few years, gender classification techniques work perfectly for a controlled environment. However, challenges occurred for real-time applications due to low resolution, off-angle poses, faces with occlusion, and various expressions. The main focus of this study is to overcome existing challenges and propose a method that can be implemented in real-time applications. This research work proposed a novel method in which CNN has been used for classification of gender for real-time application. To assess the performance of proposed method experiments were conducted on static images and video data sets. The proposed research work achieved 98% of accuracy during the experiments.

Keywords: Gender classification; Recognition; Feature extraction; real-time application

1. Introduction

The human face conveys a lot of information that is easy to identify by humans, however, it is difficult to identify by machines. In recent years, gender classification attracts the attention of various researchers (Lin, Wu, Zhuang, Long, & Xu, 2016; Ng, Tay, & Goi, 2015). Different gender classification techniques have been proposed by researchers in the last few decades (Abbas et al., 2021; Cartwright & Nancarrow, 2022; Lin et al., 2016; Moeini & Mozaffari, 2017; Ng et al., 2015).

Now a day's deep learning is a latest technology just likes driverless cars. It is also key of voice control in consumer's devices such as tablets, TVs, Phones speakers and hand frees. There are a number of deep learning networks used for different purpose such as convolutional neural network, Generative adversarial network, Wasserstein convolutional neural network, style transfer network and so on. For deep learning we first learn the models that perform specific task including understanding of forms texts, images, digital images, black images and sound. Deep learning model can give good high accuracy while some times its performance is exceeding the human levels. In deep learning the model is trained by using labeled data, images, and architecture of the neural network contain numbers of CNN layers. Deep learning is used for several computer vision applications such as image classification, object detection, object segmentation, image reconstruction, image super resolution and image style transfer. It is considered most important in the fields of the Surveillance system, Security purposes, Mobile applications, Advertisements, and many more. With so many applications there are still various challenges present such as analysis of automatic video data is a very difficult task, face detection through a live stream is also a difficult task due to various expressions, different poses, face alignment, illumination conditions. Various researchers presented their solution to overcome these problems.

In recent years, gender classification attracts various researchers due to its usage in almost every field such as attendance system, surveillance system, security purposes, mobile applications, advertisements, and many more.

However, users' need a system that can be applied in real-time applications because of increasing development in the field of the internet nowadays, live video has much attraction with many users that can share their videos. These shared videos from real-world record human faces that's why analysis of face is very important in real-time applications.

Many methods have already been proposed for gender classification in both controlled and uncontrolled situations. However, problems occurred in an uncontrolled situation when there are a high rate of noises, various illumination conditions, and occluded faces or covered faces in real-time applications (Lin et al., 2016; Ng et al., 2015). To mitigate these problems, this research work proposed a new method that will work on uncontrolled conditions and enhances the performance of gender classification techniques in real-time application.

The remaining sections of the paper are as follows. Section 2 presents the literature review, and Section 3 Section 3 discusses the proposed framework. Results and discussed in Section 4. Conclusion is presented in Section 5.

2. Literature Review

Feature extraction is one of the most important steps in gender classification. Bukar et al. (Bukar, Ugail, & Connah, 2016) proposed a method SAM (supervised appearance model). The proposed method addressed the problem of feature extraction for gender classification. In the recent past, the Active Appearance Model (AAM) was used to capture the shape and texture variation for feature extraction.

AAM utilized the Principle Component Analysis (PCA) for a dimensional reduction in an unsupervised manner but cannot handle how the predictor variables related class labels mean it's only used to detect the texture of face image. To overcome this problem authors proposed a model named as Supervised Appearance Model (SAM) which replace PCA with PLS (Partial Least Square).

The proposed method is performed by forming a parameterized model using PLS dimensionality reduction to capture the variations as well as combine them in a single model. PLS can do both dimensionality reduction and regression simultaneously. The results of the experiment were compared with previous well know techniques.

Antipov et al. (Antipov, Berrani, & Dugelay, 2016) presented an approach named DCNN (Deep Conventional neural network) which is an advanced form of convolution neural network (CNN) that is a very powerful recognition technique having different layers in its architecture. Face detection was performed by voila Jones

Experiments were performed on LFW and CASIA web face databases Images in both databases were centered faces having resolution 250*250.

Zhang et al. (Zhang & Xu, 2018) proposed the Local deep neural network (LDNN) technique. In this proposed model, local image patches were selected based on the detected facial landmarks. The detected patches then used for network training which reduced the cost of training Author proved that local deep neural network (LDNN) was cheaper in terms of computational time than (Conventional neural network) CNN. The Face detection was obtained by viola Jones. The experiments were performed on the Audience dataset which contained 26580 images of the face. The proposed framework achieved 80.64% accuracy.

Briones et al. (González-Briones, Villarrubia, De Paz, & Corchado, 2018) presented a method named a multi-agent system. In the proposed method multiple classifiers were applied for the experiment and compared each other to obtain the best classifier. The proposed method was composed of well-known techniques such as Fisher Faces, and LBP (Local Binary Pattern) for face recognition. Face recognition was performed through a combination of fisher faces and ANN (Artificial neural network). LPB (local binary pattern) was used as a feature extraction technique. A bilateral filter was used to extract edges of faces. Experiments were performed on the FERET database having 14051 images with different angles. The result of experiments achieved 85% accuracy.

Santana et al. (Castrillón-Santana, Lorenzo-Navarro, & Ramón-Balmaseda, 2016) presented a multi-scale approach where features were extracted from the face, head, and shoulders areas. This method used various feature extraction techniques such as HOG (Histogram of Oriented Gradients), LBP (Local Binary Patterns), LTP (Local Ternary Patterns), and WLD (Weber Local Descriptor) LOSIB (Local Oriented Statistics Information Booster).

Mansanet et al. (Mansanet, Albiol, & Paredes, 2016) presented a method named Local Deep Neural Network (LDNN). In the proposed method Local-DNN was responsible to obtain local features. This proposed method was the general framework that applies only in the local feature. SVM (Support Vector Machine) was adopted as a classifier. Experiments were performed on two databases such as (LFW) Labeled Faces in the Wild contained 13233 face images and the Gallagher database contained 28231 face images. Proposed framework achieved 96.25% on LFW database whereas 90.50% on Gallagher's Database.

Huang et al. (Huang et al., 2014) proposed a method named Local circular pattern (LCP). In the past Local binary patterns (LBP) were used to extract the features. However, LBP still has various limitations such as it cannot deal with noisy images because it worked on image pixel values. To overcome this problem LCP was proposed, that worked on Cluster-based quantization rather than binary quantization hence this easily remove noise. Experiments were performed on the FRGC database having 1876 images. And achieved 95.65 accuracy. Off angle face images are still a very challenging field for recognition of a face.

Alomar et al. (Alomar et al., 2013) proposed a multi-scale Band let and local binary pattern (LBP) method for gender classification from face images. The band let is one of the multi-resolution techniques that can adapt to the orientation of edges, and also it enabled better capturing of texture from face images. In the proposed method LBP and Band let were used to extract the features and minimum distance classifier (MDC) was utilized to classify the gender. The first Band let transformation was performed to detect the geometric shape of the image. After that LBP was applied to extract the features of the face image. The experiments were performed using FERET grayscale face database having 994 images and achieved 95.8%accuracy.

Alignment of the face is a very common problem and to solve this problem Eidinger et al. (Eidinger, Enbar, & Hassner, 2014) proposed a method in which three steps were performed. The first step was Detection of the face, which was performed by voila & Jones. The second step was feature extraction and LBP was used for it. After that gender classification was performed by dropout SVM that able to remove the over fitting problem. The experiment was performed on the Gallagher database which contained 28231 images and achieved 88.6% accuracy.

Rai et al. (Rai & Khanna, 2014) worked on the front face images by the use of feature selection which is based on combine information and fusion of extracted features from shape and texture of face images for classification of gender.

Experiments were performed on the FERET database containing more than 14 thousand images and only

380 images used for testing. The result of the experiment achieved 78.71% accuracy.

Perez et al. (Perez, Tapia, Estévez, & Held, 2012) presented a method for gender classification named a local binary pattern (LBP) based classifier. In the proposed method LBP was used to for feature extraction named MCT (Modified Census Transform) which was responsible for extracting the feature of the face twice to enhance the performance of the face extraction process. The proposed method was used to extract the local feature therefore face alignment was also an important achievement for this approach. Face alignment used the location of fiducial points such as eyes, mouth, and nose so on. Experiments were performed on LWF (Labeled wild faces) database containing 13,233 images. Adaboost was used as a classifier which gives 87%accuracy which was higher as compared to another classifier such as Gabor, jets.

3. Proposed Framework

3.1 Overview

The flowchart of the proposed method for gender classification is shown in Figure 1. The first step is the face detection step, Cascade face detection model is used for face detection in live video. The next step is preprocessing, in this step, images are feed into BLOB which is used to enhance the quality of an image. Finally, classification is performed by CNN.

The details of these steps are mentioned below. In live video streaming gender, classification is a challenging field, and it is important for real-time applications but due to occluded faces, blur motion, various illumination conditions these mentioned problems still exist. This research is focused on the above-mentioned challenges that occurred during the face detection phase for gender classification in real time application.

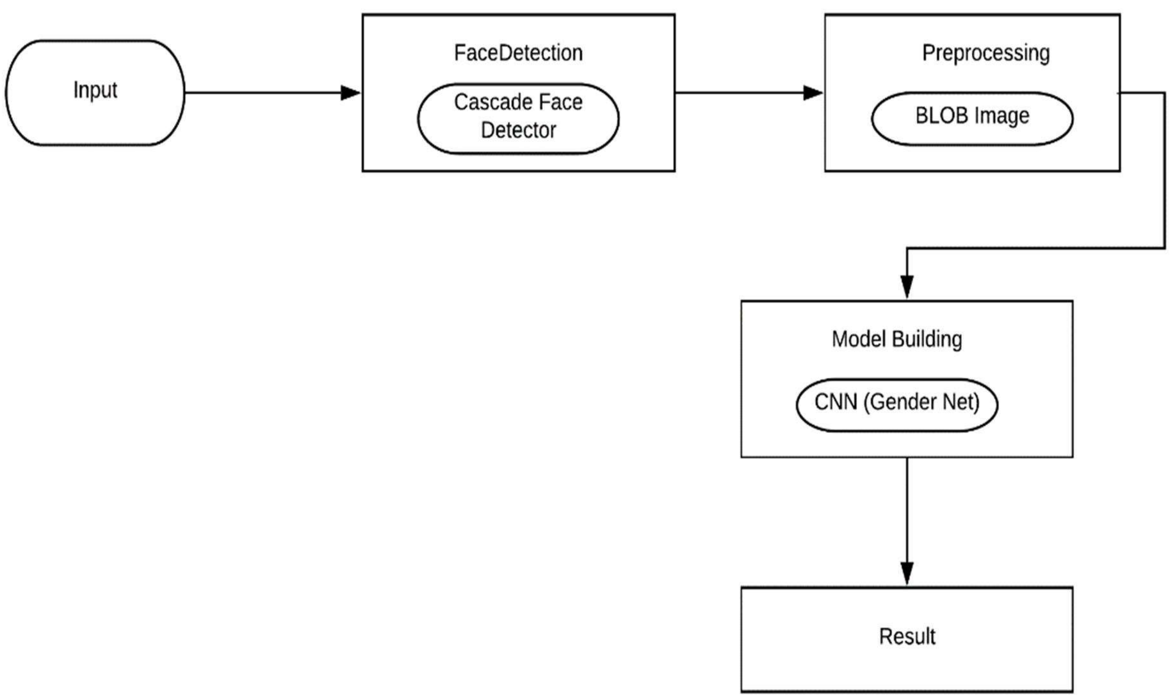

Figure 1: Proposed framework

3.2 Input

Classification of gender is performed by three type of input such as multiple faces are detected from video. Static images are also detected by this proposed method. Real time gender classification is also performed by given method.

3.3 Face Detection

As input data can have multiple objects so first step is face detection from input data. In the proposed technique Face is detected using Haar Cascade. The Haar cascade extract various feature such as lines, edges, and rectangles. An edge is explained as a sharp change in contrast from bright to dark in either vertical or horizontal direction so whenever the Haar cascade algorithm is applied on the square of pixels it follows that property, it marks them as an edge feature. The whole region of the image is extracted for lines, edges, and rectangles. We keep the rectangle on the face for not losing of the face because in cascade classifier one single blink of an eye is also responsible to lose a face, so we adjust the rectangle according to the face position.

3.4 Preprocessing

After detecting the face next step is preprocessing. In this step detected face is converted into the BLOB (Binary Large Objects). BLOB is used to enhance the quality of image such as illumination condition and normalization by using of mean subtraction and scale factor. It has four steps such as resizes and crop images from the middle of subtracting mean values, scales values by scale vector, and swap Blue (B) and Red (R) channels.

3.4.1 Mean Subtraction

Illumination is the amount of source light that destroy the pixel value Mean subtraction is applied when there were illumination changes in the input images. Therefore, it is used to aid in Convolution Neural Networks as a technique.

Typically, the results of these three topple consist of the mean value of the Red, Green, and Blue channels respectively. In other cases, the mean values of Red, Blue, and Green is computed channel wise other than pixel-wise. Both these methods are perfectly valid forms of mean subtraction. When the image is ready to pass our proposed system, we subtract the mean from every input channel of the input image.

$$R = R - U_R \tag{1}$$

$$G = G - U_G \tag{2}$$

$$B = B - U_B \tag{3}$$

3.4.2 Scale Factor

Image scaling is used to resize the images. The scaling factor is used in the proposed method which adds in the normalization of the image. As mentioned above the scaling is used to normalize the image so that all images should be in same size. Scaling is done for Red, Green and Blue component of the image.

$$R = (R - U_R)/\sigma \tag{4}$$

$$G = (G - U_G)/\sigma \tag{5}$$

$$B = (B - U_B)/\sigma \tag{6}$$

We also manually set the scale factor to scale the input image space into a particular range.

3.5. Model Building

The CNN performs both feature extraction and classification within a single network structure through learning on data samples. CNN is specifically designed to cope with shortcomings of the traditional feature extractor that is characterized by being static, is designed independently of the trainable classifier, and is not part of the training procedure. A final benefit of CNNs is that they are relatively easier to train since they have fewer parameters than fully connected MLP neural networks with the same number of hidden layers. Filters are used to analyze the value of nearby pixels. By the rule of thumb take 5 * 5 filter size then move it on the image from upper left to lower right. For every point on the image, the filter value is calculated by using the convolution operation. Filters reduce the number of weights in the neural network when the location of features such as eyes, nose, lips, and cheeks changes then classification is not performed by a neural network. When a model building starts, we manually set the values of filters after that is continuously updated throughout the training process.

Convolution is composed of independent filters. Every filter is independently convolved with image and ends with six feature maps. The pooling layer is used to reduce the spatial size of the image and it operates on each feature map separately. Block diagram of CNN is shown in Figure 2.

4. Experimental Results

4.1 Performance Parameters

To assess the performance of proposed solution following established performance parameters are used in which precision determine the number of positive class means how many times proposed method gives accurate result.

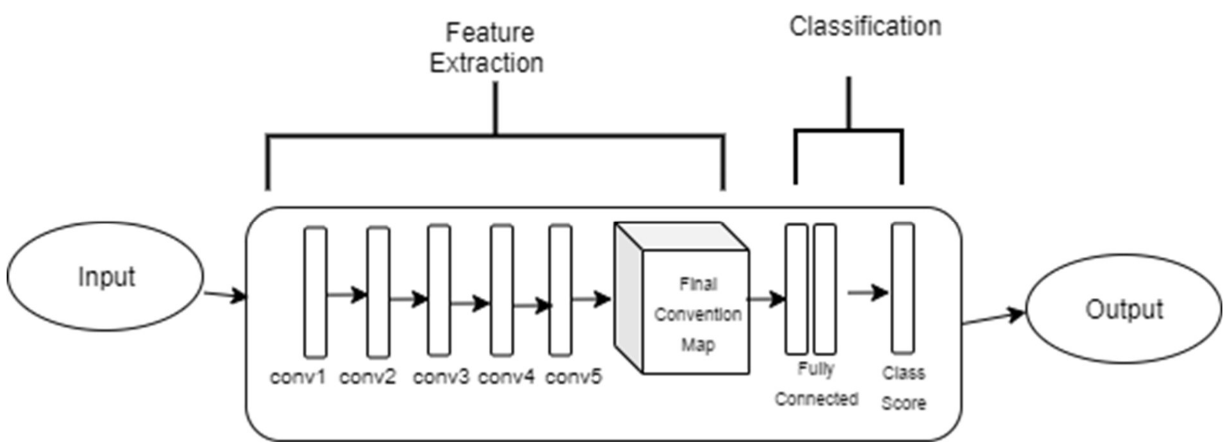

Figure 2: Block diagram of CNN

4.1.1 Precision

Precision is the percentage of the relevant result that is taken from the system during training or testing. In this research, Precision determines the number of positive class predictions that belong to the positive class only.

$$Precision = True\ Positive \div (ture\ positive + flase\ positive) \quad (7)$$

4.1.2 Recall

The recall is the percentage of relevant result that is classified by the proposed method in this research. The recall is determining the number of positive class predictions made from all positive.

$$Recall = True\ Positive \div (positive + Flase\ Negative) \quad (8)$$

4.2 Comparative Analysis of Experimental Study I (Static Images) Avg

To validate the performance of the proposed system, the results of the proposed systems are compared with various other well-known techniques. These techniques include Santana et al (Castrillón-Santana et al., 2016) , Endanger et al. (Eidinger et al., 2014), and Xu et al. (Zhang & Xu, 2018). The results are shown in Table 1 and Figure 3.

Table 1: Comparison table of static images

Methods	Accuracy	Precision	Recall	F-measure
Santana et al. (Castrillón-Santana et al., 2016), 2016	92.46	0.897	0.896	0.896
Endanger et al. (Eidinger et al., 2014), 2014	88.6	0.815	0.810	0.812
Zhang et al. (Zhang & Xu, 2018), 2018	93	0.901	0.902	0.901
Proposed Method	98	0.95	0.94	0.96

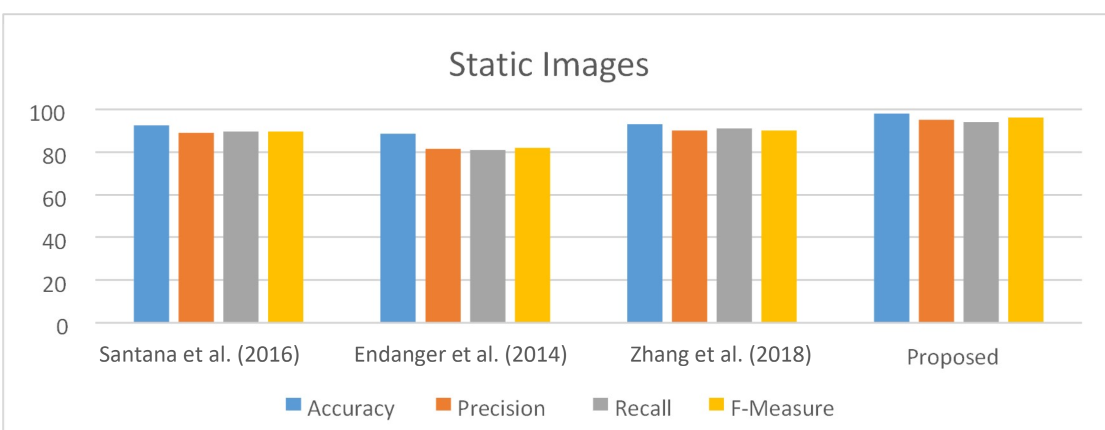

Figure 3: Comparison chart of static images

4.3 Comparative Analysis of Experimental Study 2 (Videos)

To validate the performance of the proposed system, the results of the proposed systems are compared with various other well-known techniques. These techniques include Bukar et al. (Bukar et al., 2016), Mansanet et al. (Mansanet et al., 2016), and Perez et al. (Perez et al., 2012). A comparative analysis is shown in Table 2 and Figure 4.

Table 2 Comparison table of Videos

Methods	Accuracy	Precision	Recall	F-measur
Bukar et al. (Bukar et al., 2016) 2016	92.50	0.897	0.896	0.896
Mansanet et al. (Mansanet, Albiol, & Paredes, 2016)	96.25	0.95	0.91	0.92
Perez et al. (Perez et al., 2012)	87	0.85	0.86	0.82
Proposed Method	98.1	0.905	0.901	0.905

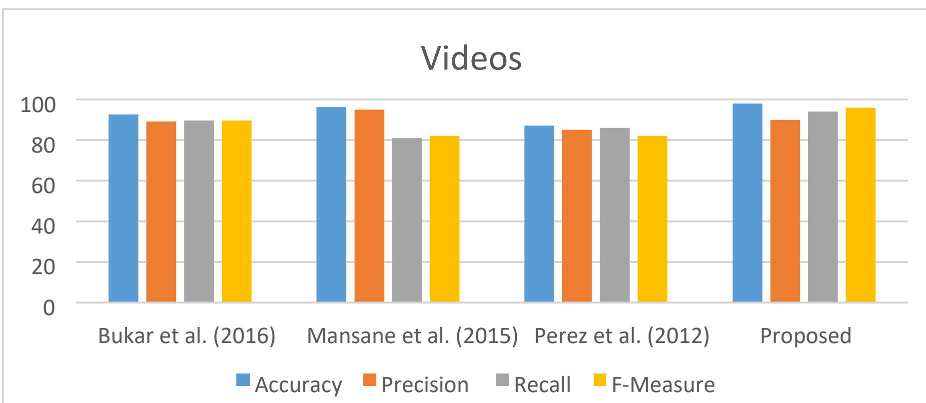

Figure 4: Comparison chart of videos

From Table 2 and Figure 4, one can observe that performance of the proposed approach is better than the previous approaches. As for accuracy, precision, and recall and F-measure of the proposed method is higher than the previous research.

4.3 Comparative Analysis of Experimental Study 3 (Real-Time)

To validate the performance of the proposed system in real-time, the results of the proposed systems are compared with various other well-known techniques. These techniques include Bukar et al. (Bukar et al., 2016), Mansanet et al. (Mansanet et al., 2016), Comparative analysis is shown in Table 3 and Figure 5.

From Table 3 and Figure 5, it is observed that performance parameter such as accuracy, precision-recall and F-measure of the proposed method is higher than the previous research.

Table 3: Comparison table of Real-Time

Methods	Accuracy	Precision	Recall	F-measure
Bukar et al., (Bukar et al., 2016), 2016	92.50	0.897	0.896	0.896
Mansanet et al., (Mansanet et al., 2016), 2016	96.25	0.95	0.91	0.92
Zhang et al., (Zhang & Xu, 2018), 2018	87	0.85	0.86	0.82
Proposed Method	98.1	0.905	0.901	0.905

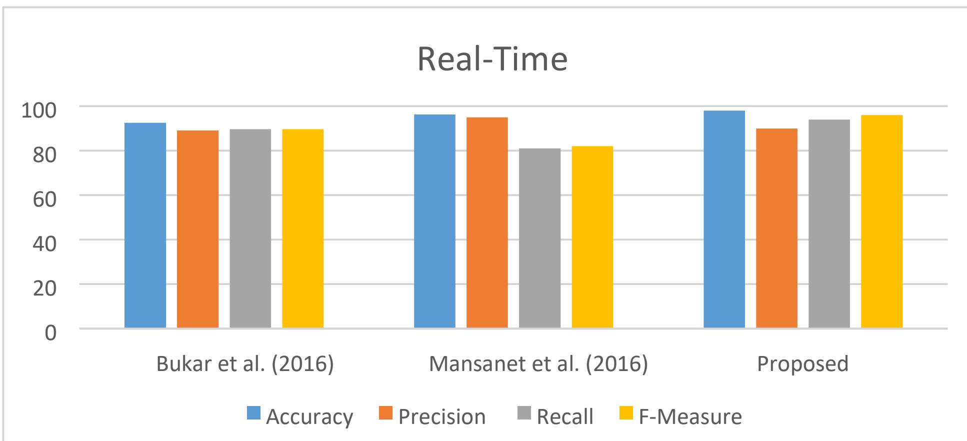

Figure 5: Comparison chart of real-time

5. Conclusion

In this study, the framework is proposed for the classification of gender in real time application. Classification of gender is very important in surveillance systems, security purposes, mobile applications, advertisements, and many more.

It is a very challenging field of image processing and many researchers have presented their work to overcome limitations such as low-quality images, illumination conditions, and live video face detection in real-time. This research compares different classification methods that have been proposed by various researchers. After considering current literature there are still various problems such as only a few researchers considered enhancement techniques in their approach and real-time applicability is still a major problem in current methods. This research proposed a method that can overcome current literature challenges such as face alignment, illumination conditions, and real-time applicability. The proposed method is used for three types of input from images, videos, and real-time applications.

LWF, Groups, and GCLV three databases are used to apply proposed techniques to verify the accuracy and we achieved 99% accuracy in real-time gender detection which is higher accuracy as compared to previous research. In the preprocessing, step BLOB images are used to enhance the illumination condition and image quality. We apply the CNN model after the BLOB images to classify gender and feature extraction as well. In this research three types of input are taken which is a static image, multiple face detection in video, and Real-time gender detection. No researcher has done work on all three inputs.

References

Abbas, F., Yasmin, M., Fayyaz, M., Abd Elaziz, M., Lu, S., & El-Latif, A. A. A. (2021). Gender classification using proposed CNN-based model and ant colony optimization. *Mathematics, 9*(19), 2499.

Alomar, F. A., Muhammad, G., Aboalsamh, H., Hussain, M., Mirza, A. M., & Bebis, G. (2013). *Gender recognition from faces using bandlet and local binary patterns.* Paper presented at the 2013 20th International Conference on Systems, Signals and Image Processing (IWSSIP).

Antipov, G., Berrani, S.-A., & Dugelay, J.-L. (2016). Minimalistic CNN-based ensemble model for gender prediction from face images. *Pattern recognition letters, 70*, 59-65.

Bukar, A. M., Ugail, H., & Connah, D. (2016). Automatic age and gender classification using supervised appearance model. *Journal of Electronic Imaging, 25*(6), 061605-061605.

Cartwright, T., & Nancarrow, C. (2022). A Question of Gender: Gender classification in international research. *International Journal of Market Research, 64*(5), 575-593.

Castrillón-Santana, M., Lorenzo-Navarro, J., & Ramón-Balmaseda, E. (2016). On using periocular biometric for gender classification in the wild. *Pattern recognition letters, 82*, 181-189.

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.

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, 172*, 98-106.

Huang, D., Ding, H., Wang, C., Wang, Y., Zhang, G., & Chen, L. (2014). Local circular patterns for multi-modal facial gender and ethnicity classification. *Image and Vision Computing, 32*(12), 1181-1193.

Lin, F., Wu, Y., Zhuang, Y., Long, X., & Xu, W. (2016). Human gender classification: a review. *International Journal of Biometrics, 8*(3-4), 275-300.

Mansanet, J., Albiol, A., & Paredes, R. (2016). Local deep neural networks for gender recognition. *Pattern recognition letters, 70*, 80-86.

Moeini, H., & Mozaffari, S. (2017). Gender dictionary learning for gender classification. *Journal of Visual Communication and Image Representation, 42*, 1-13.

Ng, C.-B., Tay, Y.-H., & Goi, B.-M. (2015). A review of facial gender recognition. *Pattern Analysis and Applications, 18*, 739-755.

Perez, C., Tapia, J., Estévez, P., & Held, C. (2012). Gender classification from face images using mutual information and feature fusion. *International Journal of Optomechatronics, 6*(1), 92-119.

Rai, P., & Khanna, P. (2014). A gender classification system robust to occlusion using Gabor features based (2D) 2PCA. *Journal of Visual Communication and Image Representation, 25*(5), 1118-1129.

Zhang, Y., & Xu, T. (2018). Landmark-guided local deep neural networks for age and gender classification. *Journal of Sensors, 2018*.