INTERNATIONAL JOURNAL OF SOFTWARE ENGINEERING & COMPUTER SYSTEMS (IJSECS)
ISSN: 2289-8522 e-ISSN: 2180-0650
VOL. 9, ISSUE 1, 59 – 67
DOI: https://doi.org/10.15282/ijsecs.9.1.2023.6.0110

ORIGINAL ARTICLE

Twofold Face Detection Approach in Gender Classification using Deep Learning

Muhammad Firdaus Mustapha[1], Nur Maisarah Mohamad[1,*] and Siti Haslini Ab Hamid[2]

[1]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Kelantan, 18500 Machang, Kelantan, Malaysia.
[2]FH Training Center, 16800 Pasir Puteh, Kelantan, Malaysia.

ABSTRACT – Face classification is a challenging task that is crucial to numerous applications. There are many algorithms for classifying gender, but their ability to evaluate their effectiveness regarding scientific data is constrained. Deep learning is popular among researchers in face classification problems. The detection of many faces is complicated and becomes a necessity in real problems. The proposed research aims to examine the effect of twofold face detection approach on the accuracy of gender classification, as well as the effect of using small datasets on accuracy. In this study, we use a small dataset to classify facial images based on their gender. The following phases involve deep learning methods along with the OpenCV library version 3.4.2 which is recommended to serve as a twofold face detection approach. In the experiments conducted, Phase 1 is the designated training phase, and Phase 2 serves as a testing phase. Two different algorithms are used in the testing phase to detect one face in the image (Experiment 1), while the remaining algorithm detects multiple faces in the image (Experiment 2). The FEI dataset is used to evaluate the accuracy of the proposed research, which results in 84% accuracy for Experiment 2 and 74% for Experiment 1, respectively.

ARTICLE HISTORY
Received: 27 October 2022
Revised: 22 March 2023
Accepted: 2 May 2023
Published: 22 May 2023

KEYWORDS
Gender Classification
Deep Learning
CNN
SVM

INTRODUCTION

With the continuous advancement of computer science and information technology, the capabilities and intelligence of computers have increased in recent years. Gender classification has attracted much attention due to its potential applications in identity verification, video capture, human-machine interface, and intelligent robots. Gender classification has since gained more attention because of advances in face detection and increased demand for intelligent visual surveillance.

Decades ago, researchers have looked at a broad field of research on gender classification in computer vision, subsequently producing some recommendations in the literature. The proposal includes implementing new models named Advanced Convolutional Neural Network (A-CNN) [1] to classify gender-specific faces. Abdalrady and Aly [2] proposed a simple feature fusion method using two simple Principal Component Analysis (PCANet) that trained on real-world images. In [3], they proposed gender classification with multi-feature methods that used Scale-Invariant Feature Transform (SFT) and Support Vector Machine (SVM) for feature extraction and classification. Other works include using Spatially-enhanced Local Binary Pattern (SLBP) and Histogram of Oriented Gradients (HOG) filters [4], Cycle-Consistent Adversarial Networks (CycleGANs) [5], Multipatch gender classification methods [6] and other strategies such as clustering and transfer learning [7].

The task of gender classification poses various challenges for machines, especially when dealing with occlusions, low quality and inconsistencies in images or even small amounts of data. In this research, the trained OpenCV and Single Shot Detector – Residual Network (SSD-ResNet) face detection algorithms were selected. Then, feature extraction and embedding are performed using trained FaceNet. Finally, gender is classified as either male or female using the SVM structure. Instead of an intermediate congestion layer as in earlier deep learning techniques, FaceNet uses a deep convolutional neural network (DCNN) that has been trained to directly improve the embedding itself. The training and testing phases form the two parts of this research pipeline. In the testing phase, there are twofold face detection approaches: an algorithm that can detect a single face in an image for Experiment 1 and an algorithm that can detect multiple faces in an image for Experiment 2. In conclusion, there are two main contributions in this paper: 1) to study the effect on gender classification accuracy of twofold face detection approaches, specifically one that detects one face in an image and another one that detects multiple faces in an image and 2) the effect of using small datasets on accuracy, as typically, machine learning projects require large amounts of data to correctly predict classes.

The following describes the structure of this paper: Section 2 provides a summary of related work. The setup of the experimental data, the suggested strategy and architecture for tackling the issue, as well as the proposed evaluation matrix are presented in Section 3. A discussion of the findings is presented in Section 4. Section 5 ends with a summary and potential areas for further research.

*CORRESPONDING AUTHOR | Nur Maisarah Mohamad | ✉ maisarah97.mohamad@gmail.com

RELATED WORK

OpenCV and Deep Leaning Face Detector

To be included in the last step of the pipeline, it is obviously important to identify the faces in the prepared image first. There are many methods for face detection, including statistical techniques [8], round or oval shape detection [9], and skin texture detection [10]. Convolutional Neural Network (CNN) [11] is the technique used. Although OpenCV [12] does not have the ability to recognize or categorize faces, it can help with face detection. Face Detector with Deep Learning is based on Single Shot Detector (SSD) [13] and has a small Residual Network (ResNet) backbone [14]. The use of ResNet will further improve the accuracy of the network. In this study, we bound faces in images using a CNN model with SSD-ResNet architecture for face detection. To create a feature map for face detection, the last three layers in Figure 1 use the original SSD layer.

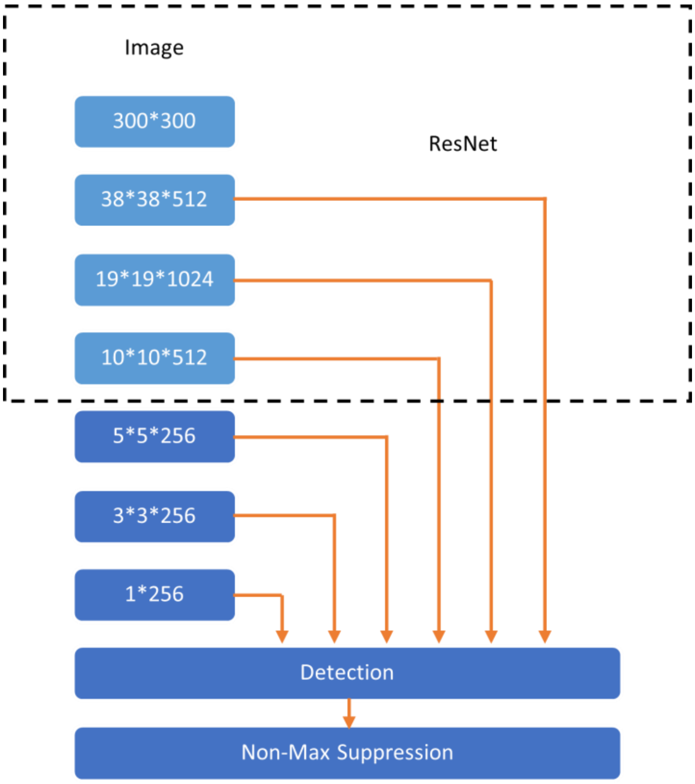

Figure 1. Face detection model based on SSD-ResNet [14].

FaceNet Facial Feature Extraction

A DCNN called FaceNet was used to identify features in images of human faces. A researcher from Google, Schroff et al, published it in 2015 [15]. For each detected face, a face embedding will be generated using the FaceNet model. High-dimensional vectors can be translated into low-dimensional spaces called embedding. By grouping inputs with similar semantic properties together in an embedding space, embeddings ideally capture some input semantics. Models can benefit from learning and using the same embedding. FaceNet model illustrates in Figure 2.

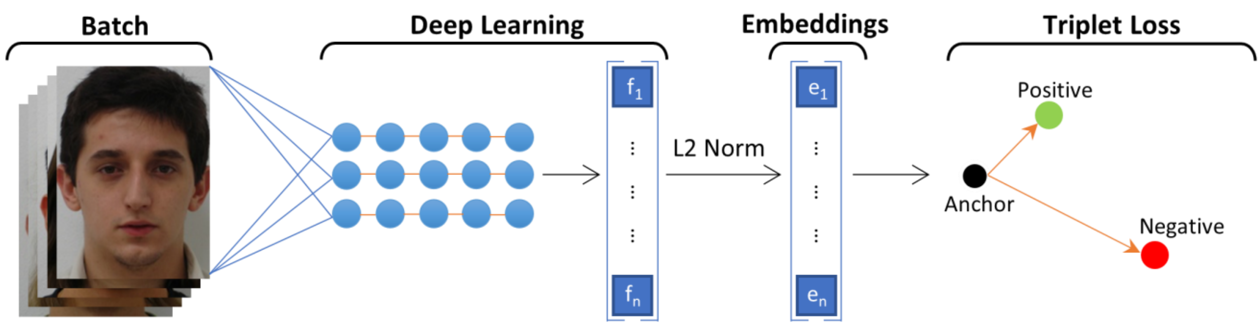

Figure 2. FaceNet model [15].

FaceNet uses face images as input and produces a vector of 128 numbers that represent the most salient facial features. These vectors are known as "embeddings" in machine learning [16]. The embedding process involves transferring all important data from an image to a vector. FaceNet basically takes a face and turns it into a vector of 128 numbers. Figure 3 shows the embeddings process.

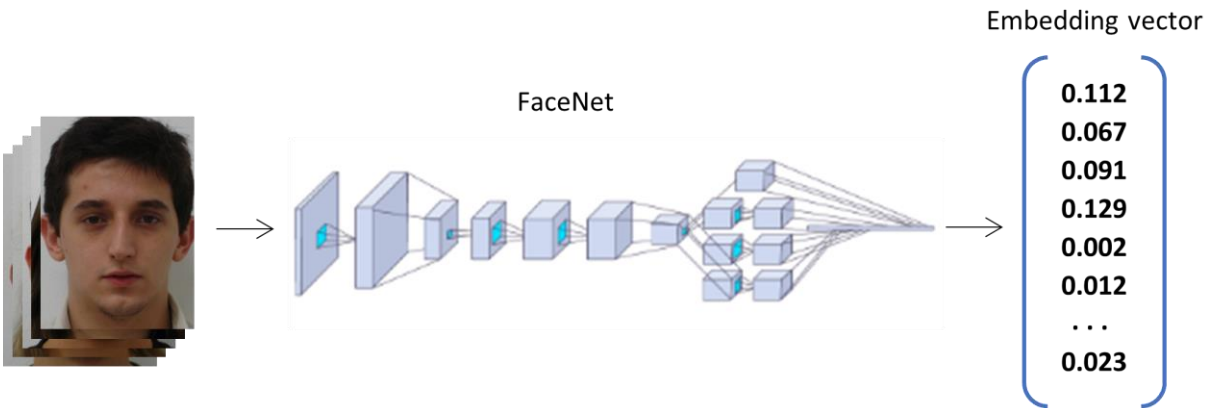

Figure 3. Embedding vector process [16].

SVM Classifier

The SVM classification [17] method shown in Figure 4 aims to maximize the marginal distance between classes using the decision boundaries obtained by using various kernels.

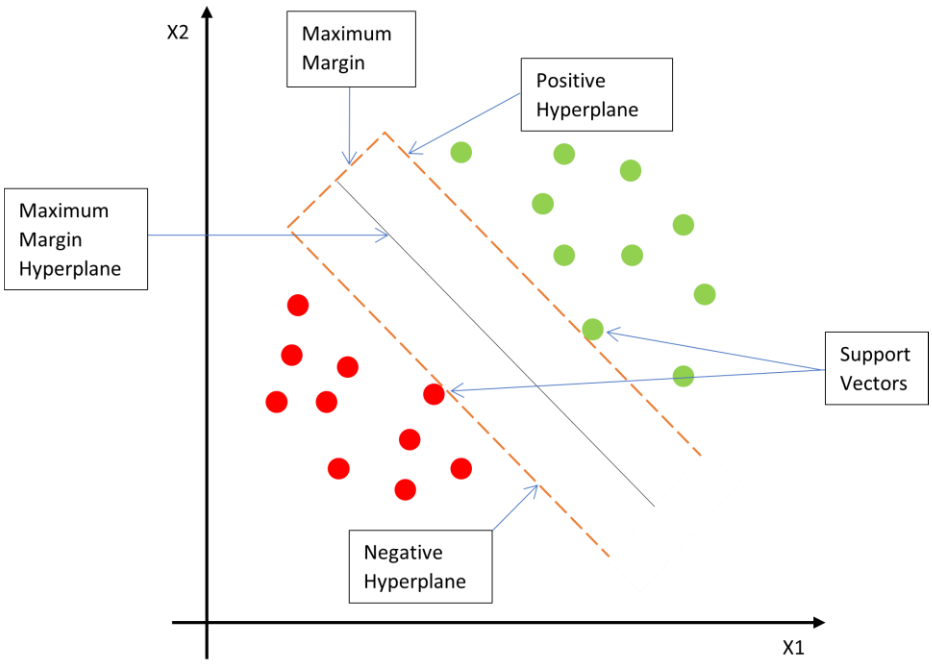

Figure 4. SVM structure [17].

SVM is designed to define a hyperplane that serves as a separator between two types of classes. This can be achieved by increasing the margin from the hyperplane into two classes. The support vector is the sample most closely related to the margin chosen to define the hyperplane. Moreover, it is a very versatile classifier for various pattern recognition issues, including age estimation [18], face recognition [19], and gender classification [20]. Mapping high-dimensional data (such as images) to low-dimensional representations (embeddings) has become a fairly common practice in modern machine learning.

METHODOLOGY

For this study, two experiments were carried out. Each experiment's goal is to evaluate the effectiveness of gender classification models using various Deep Learning approach on small amounts of data. Experiment 1 investigates the effectiveness of the algorithm that capable to detect a single face in an image, while Experiment 2 investigates the effectiveness of the algorithm that capable to detect multiple faces in an image.

Hardware and Software Requirements

In this experiment, a deep learning model used to classify gender. The used of personal computer (PC) has the following specifications: an Intel i7-3770 CPU, 4 GB of RAM, and Windows 10 Pro as its operating system. The library of OpenCV version 3.4.2 and the latest version of Python 3.9.0, are used to implement gender classification models.

Dataset

The experiments were performed on FEI face dataset [21]. The Brazilian face dataset forms the FEI face dataset. This dataset is a collection of facial images taken between June 2005 and March 2006 at the FEI Artificial Intelligence Laboratory in Sao Bernardo do Campo, Sao Paulo, Brazil. There are 2800 images in total, or 14 images for every 200 people. A uniform white background is used for all color images, which are all taken in an upright frontal position with the profile rotated up to approximately 180 degrees. Each image is 640x480 pixels in original size, although the scale may vary by about 10%. Most faces at FEI are students and workers, who range in age from 19 to 40 and have a variety of looks, hairstyles and accessories. There were 200 subjects in total, 100 of each gender (male and female). Figure 5 displays some examples of image variations from the FEI face dataset.

Figure 5. A sample FEI dataset with various image variations [21].

In this work, only 200 neutral faces of the FEI dataset are used for gender classification, as well as many papers in the literature, usually addressing neutral faces without differences in expression and pose. 100 faces are female images, and 100 faces are men. 75% of the sample is used as a training sample and the rest of the sample is used for testing. The number of female and male samples is in the same number in the training and testing phases as described in the Table1.

Table 1. Dataset distribution.

Set Type	Split Percentage	Number of Images
Training	75%	150 (75 male, 75 female)
Test	25%	50 (25 male, 25 female)
Total	100%	200 (100 male, 100 female)

Proposed Research Pipeline

The model pipeline of the proposed research is illustrated in Figure 6. The facial detection procedure must exist as part of a method for analyzing the human face. Therefore, Figure 6 shows the facial detection procedure in Phase 1 and Phase 2. First, the input of the dataset in the proposed experiment is a different dataset, i.e., Phase 1 has a 75% training dataset while the remaining dataset is a test dataset that has 25% of the total dataset. Next, these two phases carry out the face detection procedure using the OpenCV library which is combined with a pretrained face detection deep learning model. In this face detection procedure, the deep learning model acts as a preprocess as well as aligning the face using an existing pre-training model that belongs to the SSD-ResNet model that will count the face landmarks. Next, FaceNet acts as a Face alignment which is an additional step that has proven its use to improve facial detection performance in some previous research [22]–[24]. Phase 1 output is the result of embedded faces of training face datasets. However, the existing algorithm has been modified in Phase 2, which is a test phase, to allow detection to occur in two types of algorithms in different experiments. Next, the embedded face taken from Phase 1 is used in Phase 2 with the aim of comparing faces using the SVM method. Finally, the Phase 2 output will evaluate the overall results of the proposed gender classification. The objective of both experiments is to detect only one face in an image by using two different algorithms as depicted in Figure 7.

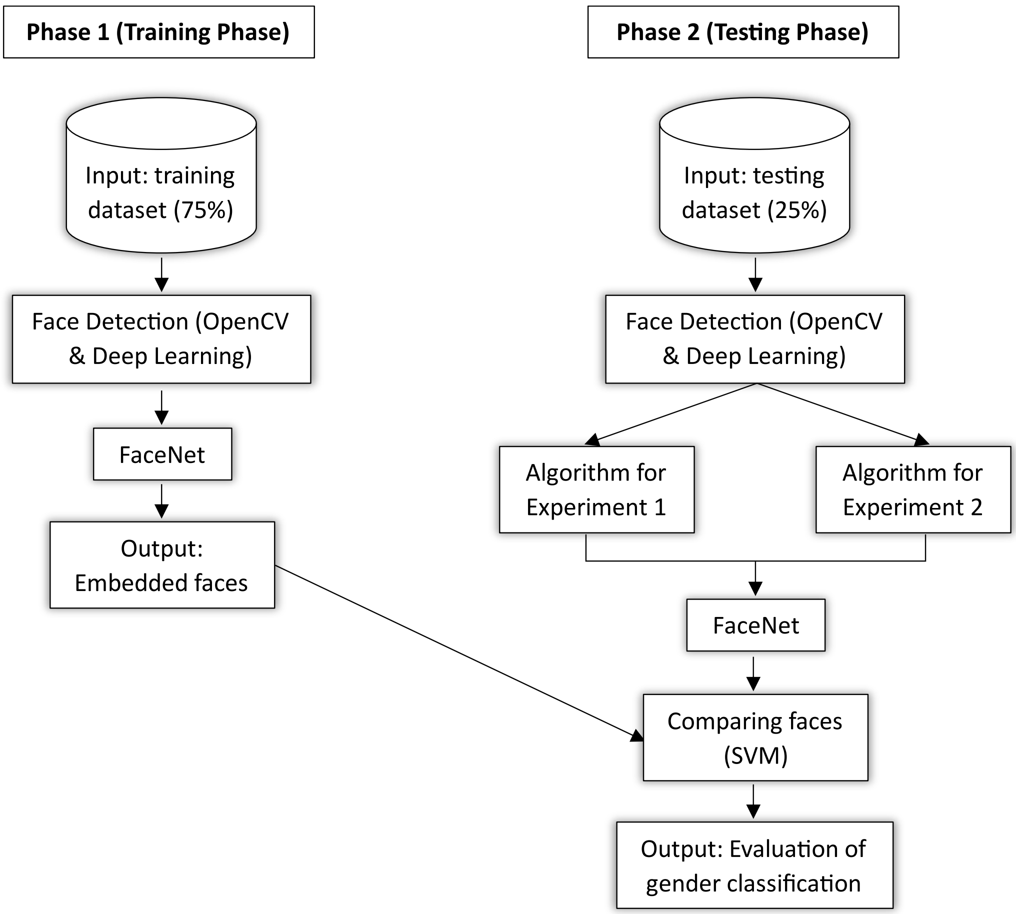

Figure 6. Proposed model pipeline.

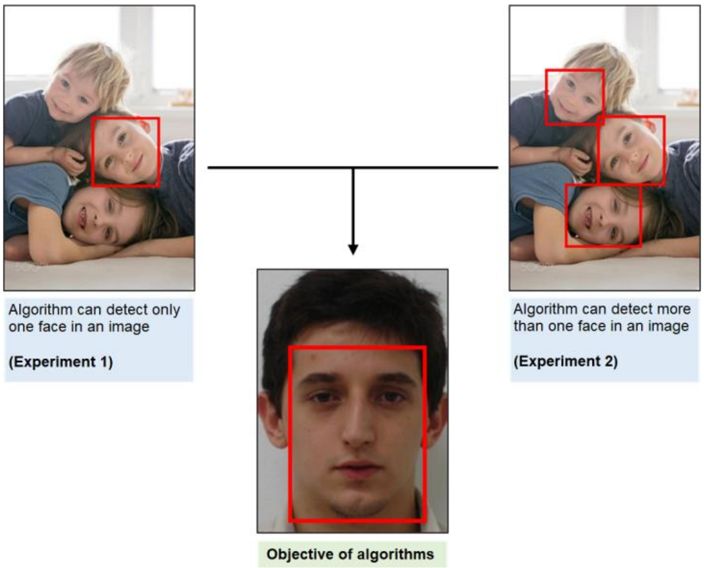

Figure 7. Example of modified algorithms.

The pictures shown in Experiment 1 and Experiment 2 in Figure 7 are examples of the differences between the two face detection algorithms in Experiment 1 and Experiment 2. The presence of these pictures further explains that the algorithm in Experiment 2 is more suitable to be used in the face detection module where there are many faces in an image.

The trained FaceNet model is then used to extract features in both the training and testing phases of the process. FaceNet uses end-to-end learning in its architecture (Figure 2). The basic architecture is Inception Network [25], which follows L2 norms [13] and produces face embedding. L2 normalisation is then used to calculate the vector's length, which represents how far the vector extends in Euclidean space. The norm used in feature extraction is the L2 Norm which is the most widely used Norm and is also called the Euclidean Norm. A triplet loss function is then presented (Figure 8).

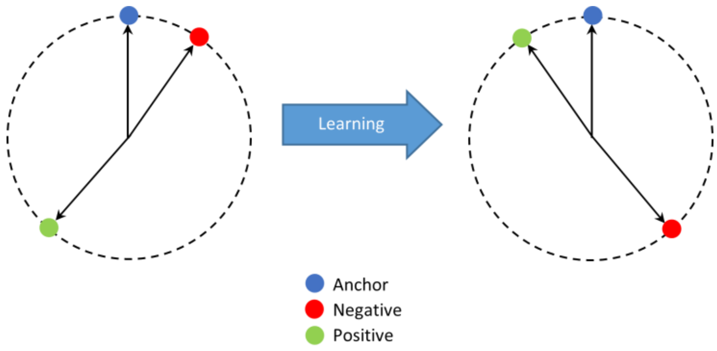

Figure 8. Triplet Loss in Cosine Similarity [15].

The triplet loss function [26] directly reflects the gender classification that is the goal of the proposed research. Triplet loss reduces the distance between positive and anchor. Both are of the same gender and maximize the separation between the opposite and negative gender.

The triplet loss function is the most suitable for determining gender. Triplet loss aims to create space between each face pair and every other face, allowing faces from one identity to co-exist with another face while still creating separation. Therefore, the identity of each person is different. In the testing phase, SVM is used to find the given test data against relevant subjects found in the database. Labels and features are used as input to the SVM classifier.

Evaluation

The proposed experiment is evaluated using a confusion matrix that will obtain the results of accuracy, precision, recall and F-1 scores [27]. Figure 9 provides an example of a confusion matrix.

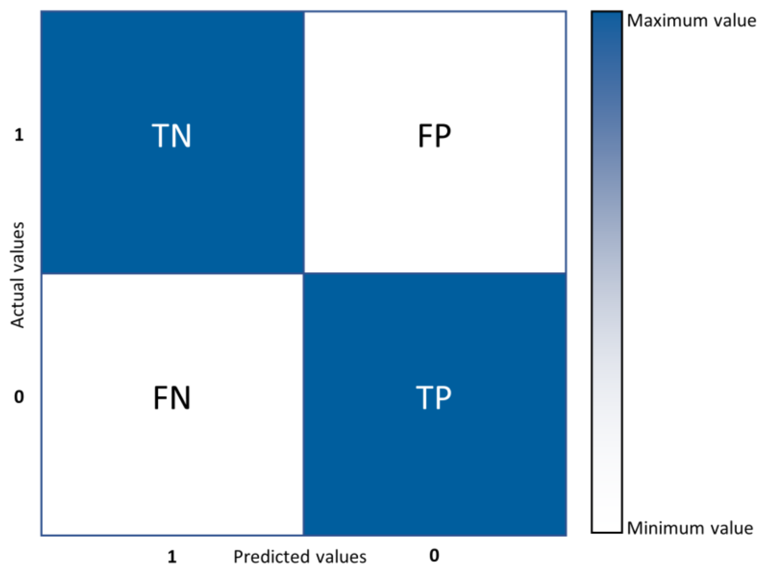

Figure 9. Binary classification in confusion matrix.

The model will achieve high accuracy if the number of True Negative (TN) and True Positive (TP) increases. Meanwhile, the accuracy of the model will decrease if the number of false positives (FP) and false negatives (FN) is high. The maximum value as well as the minimum value depends on the total amount of data in a class.

RESULT AND DISCUSSION

The results of both experiments are detailed in this section. Figure 10 illustrates the results of Experiment 1 and Experiment 2 in the order of the confusion matrix. Meanwhile, Table 2 shows the results of the comparison between the algorithms in Experiment 1 and Experiment 2.

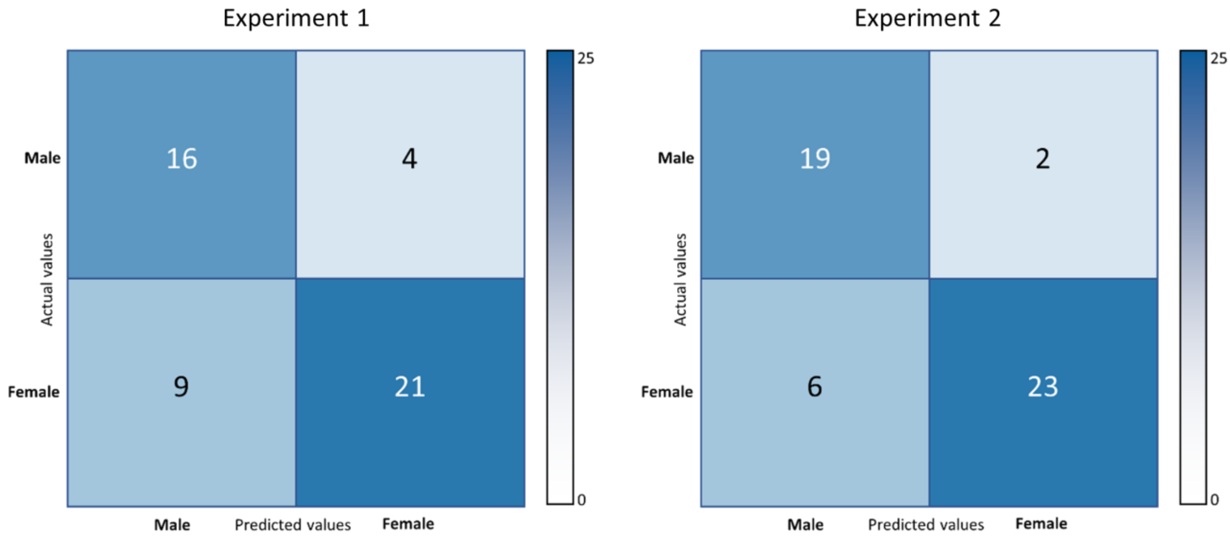

Figure 10. The confusion matrix indicates the proposed experimental findings.

Table 2. Comparison between the results of the algorithms used in Experiment 1 and Experiment 2.

Type of Dataset	Accuracy	Precision	Recall	F-1 Score
Experiment 1	74%	84%	70%	76.36%
Experiment 2	84%	92%	79.31%	85.19%

From the above results, the algorithm in Experiment 2 preceded Experiment 1 with 84% Accuracy and 92% Precision. Meanwhile, the results of Recall and F-1 Score in Experiment 2 are better than Experiment 1 which is 79.31% for Recall and 85.19% for F-1 Score. The comparison result with related research is shown in Table 3.

Table 3. Result comparison with related research.

Experiment	Dataset	Number of datasets	Dataset split (Train: Test)	Accuracy
Experiment 1	FEI	200	(75%: 25%)	74%
Experiment 2	FEI	200	(75%: 25%)	84%
Mustapha et al., [27]	AAF	3518	(70%: 30%)	85.12%

Furthermore, the results of this study have been compared to previous research [27] using the same model but using a different dataset of All-Age Face (AAF) datasets in different amounts of data. In addition, there are also similarities in terms of data splitting methods that use the hold out method but with different amounts on each section. Mustapha et al. [27] proposed an age group classification to address the issue of class imbalance in CNN-based classifications. Although there is a large difference in the number of datasets, the performance in terms of accuracy varies only marginally. Based on the comparison, the results of the proposed study show that the model can achieve competitive accuracy by using a small number of datasets. Table 4 shows the number of detected faces on the experiment.

Table 4. Number of faces detected.

Type of Dataset	Number of Face Images	Number of Face Detected
Training	150	135
Test (Experiment 1)	50 (25 male, 25 female)	44 (20 male, 24 female)
Test (Experiment 2)	50 (25 male, 25 female)	55 (28 male, 27 female)

There are only 135 faces out of 150 that can be extracted from the training dataset. Furthermore, only 44 faces out of 50 total faces in the test dataset for the algorithm that detects single face in one image are found. The algorithm used in Experiment 2, which can detect multiple faces, was able to detect 55 out of a total of 50 faces in the image. This shows that the algorithm used in Experiment 1 was unable to identify 6 additional faces in the dataset, including 5 male faces and 1 female face. The algorithm in Experiment 2 detects 55 faces out of 50 faces, there are 5 more faces detected which are 3 male images and 2 female images. Unfortunately, each dataset used contained only one face per image. Figure 11 illustrates the summary of Experiment 1 and Experiment 2.

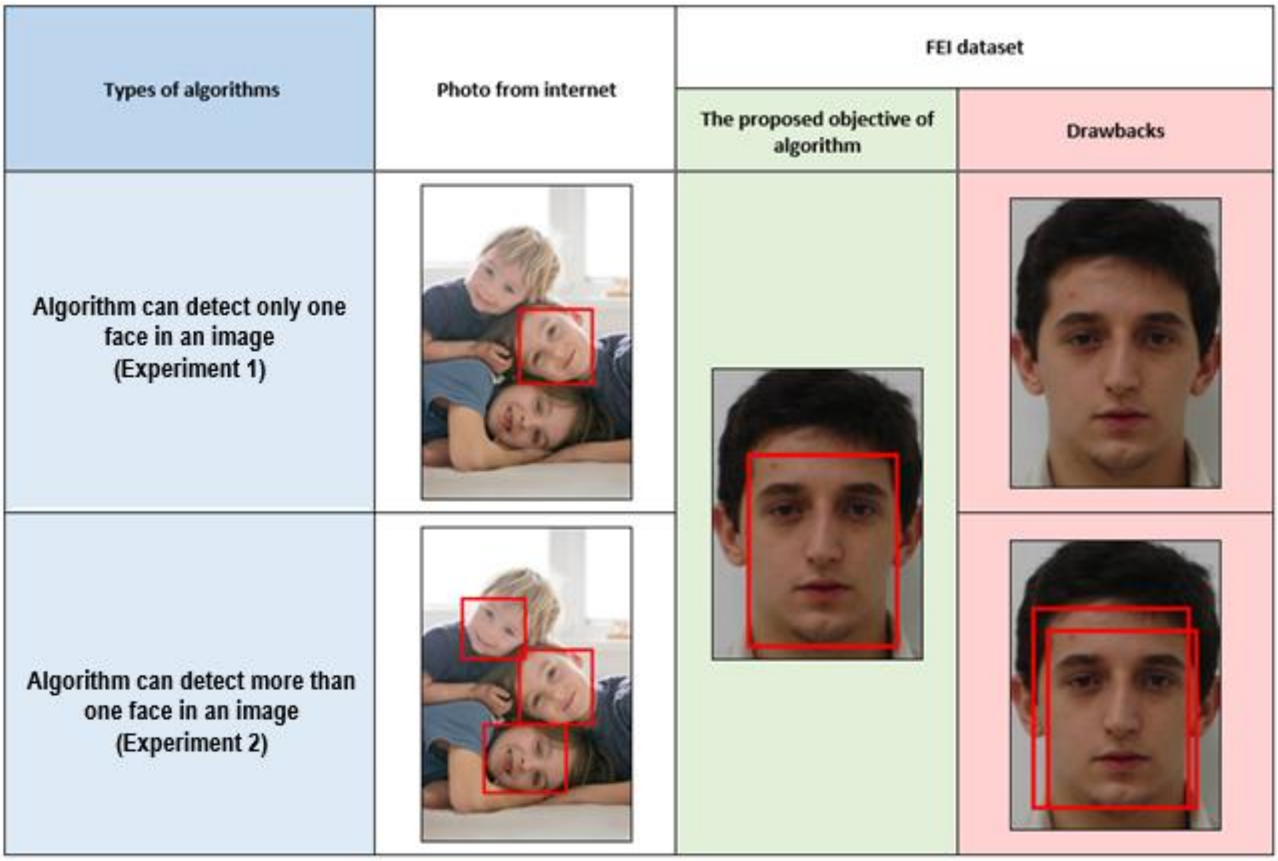

Figure 11. Summary of Experiment 1 and Experiment 2.

In conclusion, the algorithm for detecting multiple faces in an image as shown in Experiment 2 can be used if applied to a small number of datasets. However, when applied to many datasets, algorithms for detecting single faces in images are also important. Although all human faces were detected in Experiment 1, not all detected images in Experiment 2 contained images of human faces. There may also have misalignment face detection in the algorithm of Experiment 2. 6 additional faces that needed to be detected were also missed in Experiment 1. In both experiments, these algorithms were more likely to identify female faces than male faces based on gender.

CONCLUSION AND FUTURE WORK

It is possible to conclude that the FEI dataset is a balanced set of facial image data to be used as a gender classification in a small number of datasets. In most cases found in machine learning models, the accuracy of the model improves as the size of the training dataset increases. Meanwhile, the results of the experiment showed that a small number of datasets can achieve high classification accuracy if used for the appropriate type of problem. In other words, Experiment 1 is more suitable if applied to a dataset that has only one face in the image, as is the case with the proposed FEI dataset. Experiment 2, on the other hand, is more appropriate if the dataset used has multiple faces in one image as shown in Figure 11 under “Photos from the Internet”. However, both algorithms in the proposed experiment have their own drawbacks, which further confirm the fact that the greater the number of training datasets, the higher the accuracy of the model can be achieved.

In the future, existing deep learning algorithms especially for CNN algorithms in gender classification will be further enhanced by using the latest version of OpenCV that accompanies the use of Graphics Processing Units (GPUs) instead of CPUs with the same number of datasets. In addition, in line with the transformative changes in the Twelfth Malaysia Plan, this research will enhance digital connectivity for inclusive development.

ACKNOWLEDGEMENT

The Fundamental Research Grant Scheme (FRGS) of the Ministry of Education (MOE) supported this research with grant number 600-IRMI/FRGS 5/3 (234/2019).

REFERENCES

[1] H. Yousif and I. A. El Kader, "Gender face Recognition Using Advanced Convolutional Neural Network Model," *2021 Int. Conf. Digit. Soc. Intell. Syst. DSInS 2021*, pp. 255–259, 2021, doi: 10.1109/DSInS54396.2021.9670602.

[2] N. A. Abdalrady and S. Aly, "Fusion of Multiple Simple Convolutional Neural Networks for Gender Classification," Proc. 2020 Int. Conf. Innov. Trends Commun. Comput. Eng. ITCE 2020, pp. 251–256, 2020, doi: 10.1109/ITCE48509.2020.9047798.

[3] S. Kumar, S. Singh and J. Kumar, "Gender Classification Using Machine Learning with Multi-Feature Method," 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, pp. 0648-0653, 2019 doi: 10.1109/CCWC.2019.8666601

[4] M. Annalakshmi, S. M. M. Roomi, and A. S. Naveedh, "A hybrid technique for gender classification with SLBP and HOG features," *Cluster Comput.*, vol. 22, pp. 11–20, 2019, doi: 10.1007/s10586-017-1585-x.

[5] C. Yang and Z. Lv, "Gender based face aging with cycle-consistent adversarial networks," *Image Vis. Comput.*, vol. 100, p. 103945, 2020, doi: 10.1016/j.imavis.2020.103945.

[6] J. Cheng, Y. Li, J. Wang, L. Yu, and S. Wang, "Exploiting Effective Facial Patches for Robust Gender Recognition," *Tsinghua Science and Technology*, vol. 24, pp. 333–345, 2019, doi: 10.26599/TST.2018.9010090.

[7] C. H. Nga, K. -T. Nguyen, N. C. Tran and J. -C. Wang, "Transfer Learning for Gender and Age Prediction," IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan), Taoyuan, Taiwan, pp. 1-2, 2020, doi: 10.1109/ICCE-Taiwan49838.2020.9258347.

[8] T. A. Mohammed, S. O. Rafeeq, and O. Bayat, "The Statistical Learning Methods in image processing and Facial Recognition," ICEMIS'21: The 7th International Conference on Engineering & MIS, 2021, doi: 10.1145/3492547.3492656.

[9] A. Kumar, A. Kaur, and M. Kumar, "Face detection techniques: a review," *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 927–948, 2019, doi: 10.1007/s10462-018-9650-2.

[10] S. Garg, S. Mittal, and P. Kumar, "Performance Analysis of Human Face Recognition Techniques," *J. Comput. Theor. Nanosci.*, vol. 16, no. 9, pp. 3830–3834, 2019, doi: 10.1166/jctn.2019.8257.

[11] Y. Feng, X. An, and S. Li, "Research on Face Recognition Based on CNN," *Chinese Control Conf. CCC*, vol. 2018-Janua, pp. 9078–9082, 2018, doi: 10.23919/ChiCC.2018.8484211.

[12] M. Khan, S. Chakraborty, R. Astya and S. Khepra, "Face Detection and Recognition Using OpenCV," 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, pp. 116-119, 2019, doi: 10.1109/ICCCIS48478.2019.897449.

[13] R. Ranjan *et al.*, "A Fast and Accurate System for Face Detection, Identification, and Verification," *IEEE Trans. Biometrics, Behav. Identity Sci.*, vol. 1, no. 2, pp. 82–96, 2019, doi: 10.1109/tbiom.2019.2908436.

[14] X. Lu, X. Kang, S. Nishide and F. Ren, "Object detection based on SSD-ResNet," 2019 IEEE 6th International Conference on Cloud Computing and Intelligence Systems (CCIS), Singapore, pp. 89-92, 2019, doi: 10.1109/CCIS48116.2019.9073753.

[15] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. June, pp. 815–823, 2015, doi: 10.1109/CVPR.2015.7298682.

[16] S. Makinist, B. Ay, and G. Aydin, "Average Neural Face Embeddings for Gender Recognition," *European Journal of Science and Technology*, no. April, pp. 522–527, 2020, doi: 10.31590/ejosat.araconf67.

[17] Z. Rustam and A. A. Ruvita, "Application Support Vector Machine on Face Recognition for Gender Classification," *J. Phys. Conf. Ser.*, vol. 1108, no. 1, 2018, doi: 10.1088/1742-6596/1108/1/012067.

[18] C. Xiao, Z. Zhifeng, C. Jie and Z. Qian, "Combined Deep Learning With Directed Acyclic Graph SVM for Local Adjustment of Age Estimation," *IEEE Access*, vol. 9, pp. 370-379, 2021, doi: 10.1109/ACCESS.2020.3046661.

[19] L. Shi, X. Wang, and Y. Shen, "Research on 3D face recognition method based on LBP and SVM," *Opt. - Int. J. Light Electron Opt.*, vol. 220, no. June, p. 165157, 2020, doi: 10.1016/j.ijleo.2020.165157.

[20] A. Venugopal, O. V. Yadukrishnan and R. Nair T., "A SVM based Gender Classification from Children Facial Images using Local Binary and Non-Binary Descriptors," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, pp. 631-634, 2020, doi: 10.1109/ICCMC48092.2020.ICCMC-000117.

[21] C. E. Thomaz, "FEI Face Database," 2007.

[22] F. Cahyono, W. Wirawan and R. Fuad Rachmadi, "Face Recognition System using Facenet Algorithm for Employee Presence," 2020 4th International Conference on Vocational Education and Training (ICOVET), Malang, Indonesia, pp. 57-62, 2020, doi: 10.1109/ICOVET50258.2020.9229888.

[23] E. Jose, G. M., M. T. P. Haridas and M. H. Supriya, "Face Recognition based Surveillance System Using FaceNet and MTCNN on Jetson TX2," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, pp. 608-613, 2019, doi: 10.1109/ICACCS.2019.8728466.

[24] T. Nyein and A. N. Oo, "University Classroom Attendance System Using FaceNet and Support Vector Machine," 2019 International Conference on Advanced Information Technologies (ICAIT), Yangon, Myanmar, pp. 171-176, 2019, doi: 10.1109/AITC.2019.8921316.

[25] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 2818-2826, 2016, doi: 10.1109/CVPR.2016.308.

[26] F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp. 815-823, 2015, doi: 10.1109/CVPR.2015.7298682.

[27] M. F. Mustapha, N. M. Mohamad, G. Osman, and S. H. A. Hamid, "Age group classification using Convolutional Neural Network (CNN)," *J. Phys. Conf. Ser.*, vol. 2084, no. 1, 2021, doi: 10.1088/1742-6596/2084/1/012028.