Automated Gender Classification Through Face Detection: A Deep Learning Approach

Atanu Kumar Das [1], Payel Sengupta [1], Chiradeep Mukherjee [2], Arup Mallick [3]

[1]Department of Computer Science and Engineering, Brainware University , Barasat, Kolkata , 700125, West Bengal, India.
[2]Department of Computer Science Technology, University of Engineering & Management, Kolkata ,West Bengal, India.
[3]Department of Computer Science & Engineering , Hooghly Engineering & Technology College, Hooghly, West Bengal, India.

Abstract

Gender identification using deep learning techniques has received much attention in recent research and development. This study investigates the application of convolutional neural networks (CNN) for gender classification on human faces. This model leverages CNN's powerful feature extraction capability to analyze facial features from images. Identify and classify gender accurately. In addition to remembering gender, this model also aims to provide insights into mental health by interpreting facial expressions, body language, and other visual signals. The training process uses a large labeled dataset of facial images, allowing CNN to detect and recognize them. Expressions are explicitly related to emotions such as sadness, disgust, happiness, neutrality, or surprise. The study took advantage of a publicly available facial database. To ensure effective training, testing, and evaluation, the proposed model has used the CNN-based gender classification method, and this research advances the field of computer vision. This model gives 98.11% on women and 90.60% on men face detection. It has hypothetical applications in multiple fields, such as the interaction between humans and computers like demographic studies and mental health assessment.

Keywords: CNN, face detection, augmentation, CelebA dataset, deep learning

1 Introduction

Gender recognition is a core aspect of computer vision, as well as a variety of other disciplines. As a case in point, security systems can use gender detection to enrich monitoring and surveillance through more detailed identification of individuals. In human-computer interaction, altering interfaces responsive to recognized gender can improve the personalization of user experiences. As with other demographic characteristics, gender identification allows companies to customize marketing efforts for particular audience segments with precise ads and demographic analysis. Modeling must be critical to maintain impartiality, fairness, and protection of individual privacy. Misclassifications or biased predictions can trigger significant social and ethical problems. The transformation of data processing and interpretation across multiple industries by machine learning (ML) [1] is now responsible for promoting automated decision-making and predictive analytics. ML algorithms let computers glean knowledge from data to make predictions or decisions without any explicit programming needed for each job. The adoption of new systems has caused progress in healthcare, finance, marketing, and security. An important application involves gender identification, designed to interpret an individual's gender from input data, including sound or visual recordings [2]. Accurate attendance tracking in educational institutions is crucial, particularly for managing large student populations. In this paper, related work of face detections is described in Section 2, details about the dataset are discussed in Section 3, and the method with experimental results and conclusion are discussed in Section 4 and Section 5, respectively.

2 Related Work

Traditional manual methods are often inefficient, error-prone, and labor-intensive. To overcome these challenges, a face recognition-based attendance system that takes advantage of deep learning and computer vision has been proposed [3]. This system simplifies the attendance process, reduces the possibility of fraud, and improves accuracy. It employs robust face detection algorithms and histograms of oriented gradients (HOGs) for feature extraction, building a comprehensive database of enrolled students. The integration of deep learning-based face detection with Principal Component Analysis (PCA), Support Vector Machine (SVM) [4], and K-Nearest-Neighbor (KNN) classification enhances the system's performance. Each student is assigned a unique identifier, and the central dataset is updated regularly to ensure efficient and accurate attendance tracking. The proposed system demonstrates a high accuracy rate of 96.8%, highlighting its potential to significantly improve attendance management in educational settings. Gender detection forms one of the parts of face recognition systems. The author proposed gender classification in [5] in a novel way using hybrid classical-quantum neural networks. A quantum variational circuit and a deep neural network have already been trained to develop a strong binary classifier. The quantum circuits with convolutional neural networks achieved high gender classification accuracy with publicly available facial image databases compared to the domestically more prominent models.

Security systems, including the use of Closed Circuit Television (CCTV) [6], have some issues, one of the significant issues being storage issues due to large volumes of data. The major disadvantage of continuous recording is the high storage capacity required for the recorded data, which becomes essential for long-term monitoring. The work of [7] also presents an efficient technique, as the surveillance camera captures the video only if it detects motion, consequently minimizing disc space. It is a system that can be installed in homes, corporate spaces, and other areas and makes surveillance better by monitoring any strange occurrence. Haar cascades are used to detect the face, and the Inception V3 model for gender classification has an accuracy of 97.4% on the IMDB dataset. Biometric attributes such as age and gender can assist in numerous applications, including security [8], marketing, and individual services; thus, it requires automatic identification. The effectiveness of different CNN architectures is compared as part of the study to determine which model is most effective: VGG16, ResNet50V2. The consequences can be any civil application that demands recognition of facial features, including forensic case studies, specific advertising, etc. The advancement of social media needs accurate age and gender classification. Now, [9] put forward a multi-phase method with multiple CNNs to improve the classification accuracy. These components involve face detection, background elimination, facial landmarking, and applying several CNN networks. A voting system where the voting is done to establish a more accurate model of the classification in the different individual networks is used. It suits the realization of the ideal design for many cases because it does not have the flaws of traditional systems.

Flask, deep learning, and CNNs have been used to develop a real-time system for detecting age, gender, and emotions from facial images, successfully done in [10]. This system employed cascaded classifiers for face detection and CNN all through the feature extraction and classification phase. It is ideal for memory-constrained devices to get precise real-time outputs across marketing, health, and security applications. Robust research indicates that the system performs accurately and can detect demography and emotion efficiently, even under certain environmental factors. Recent literature has shown that DL methods have a heightened application when operational transportation and modeling paradigms are questioned, especially in urban settings. [11] investigates the CNN-based methods [12] adoption for age and gender recognition and its future impacts on transportation systems. These benchmarks include AlexNet, among others, and the research analyses their performance in categorizing demographics from the images of faces. This work emphasizes the need for proper demographic scans in the transport systems, indicating that CNN-based approaches can significantly support advanced analytics and inform policies within urban development.

The current study seeks to develop a system for gender perception and classification based on critical input parameters. The framework discussed in [13] involves data collection of a sizeable, heterogeneous set, data preparation, feature identification, model selection, model training, validation, and then deployment of the resulting system. Therefore, constant monitoring and maintenance will help the model over its lifespan. The research also seeks to develop a highly efficient machine-learning model that accurately predicts the gender of a user as characterized by the image. This objective has been derived from a social issue about the gender split of government subsidies. Where necessary, subsidies are available to only males or females, leaving out the opposite gender. The rationale of the proposed gender classification model is to enhance the efficacy of the verification process further, thereby allowing only qualified persons to be registered and receive the correct subsidies.

3 Preparation of Dataset

Data variability is essential for adequately generalizing models among different populations. Openly available datasets, such as the CelebFaces Attributes Dataset (CelebA) shown in Fig.1 and number of samples are described in Table 1, frequently contain annotations and represent a diverse range of people, thereby serving well for training gender categorization models. Still, data preprocessing is essential for sustaining consistency and eliminating noise. The preprocessing [14] of visual gender recognition can include image scaling, expanding the dataset through rotation techniques, normalized pixel values, and rotating to diversify the collection. An essential part of preprocessing involves feature extraction.

Fig. 1 Sample images of men and women from CelebA dataset for face recognition.

Table 1 CelebA Dataset category-wise (men and women) with the number of images [15].

Category	Number of images	Dimension
Men	1173	48 x48
Women	1134	48 x48

In images, this may consist of marking facial landmarks and obtaining details, like the eyes' separation or the jawline's outline [16]. Pitch, tone, and spectral characteristics might be vital in identifying gender through voice. Improved model accuracy and resilience greatly depend on efficient feature extraction. Images in the CelebA dataset, mainly used for facial expression recognition, are divided into two gender categories: male and female. The data consists of automatically aligned 48x48 pixel facial photographs to center the face and fill the frame with equal dimensions in each image. Data pre-processing encrypts or modifies the data to make it easier for the machine learning algorithm to understand. The CelebA dataset is already pre-processed, so additional pre-processing methods are unnecessary. However, training a deep learning model typically demands a large volume of data. To address this, we employ a simple yet effective technique called Data Augmentation [17] [18]. Data Augmentation generates multiple realistic variations of each training sample, artificially increasing the size of the training dataset. This technique helps reduce overfitting by exposing the model to a broader variety of data during training. In our work, we applied several data augmentation methods, including rotation, width shift, height shift, shear, horizontal and vertical flips, and resizing images to different dimensions. These augmentations are added to the training set, making the model more robust to object orientation, position, and size variations within the image. Additional adjustments can also be made, such as altering contrast and lighting. By applying these transformations, we effectively expand the size of the training set, improving the model's generalization ability. Data Augmentation is a widely used technique to enhance model performance, mainly when the initial dataset is small, when overfitting is a concern, or when squeezing out better accuracy from the model is desired.

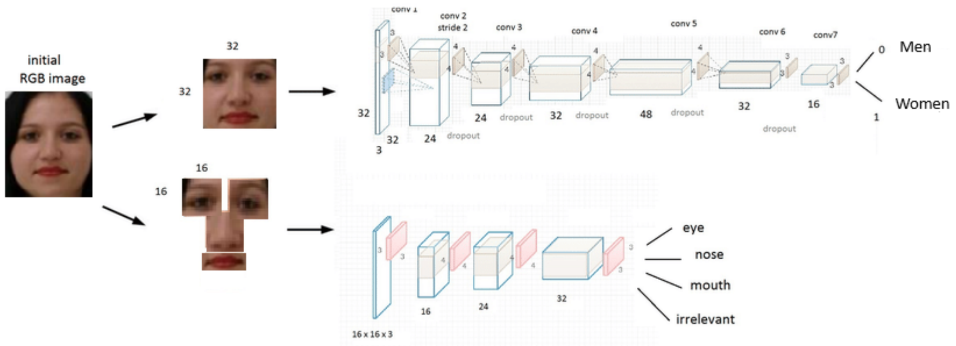

Fig. 2 Layerwise architecture of the proposed Convolutional Neural Network (CNN) model for face detection

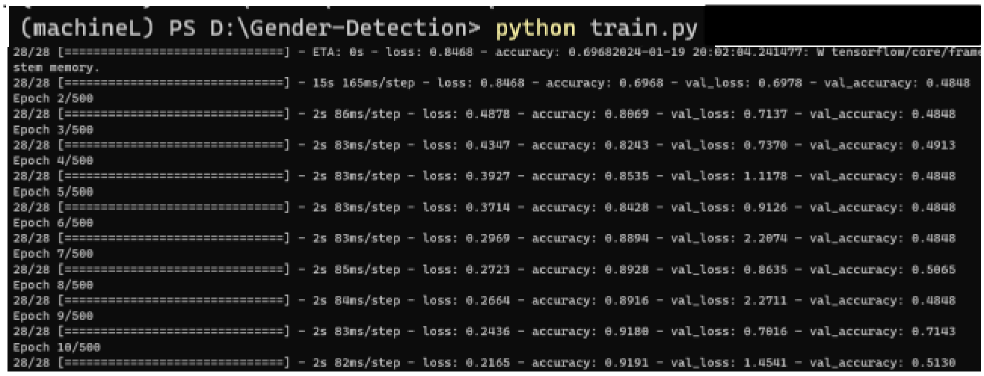

Fig. 3 Visualizations of the feature maps indicate the CNN model effectively captures relevant patterns in the input data

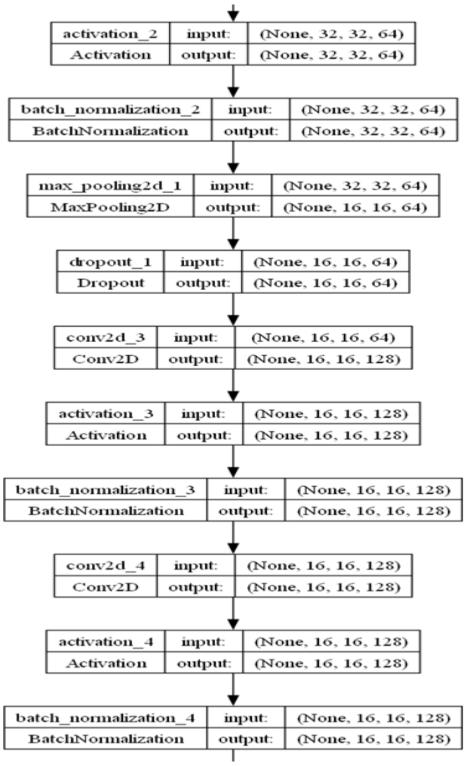

Fig. 4 Layer-wise pooling operations show a balance between retaining key features and downsampling unnecessary details for higher efficiency

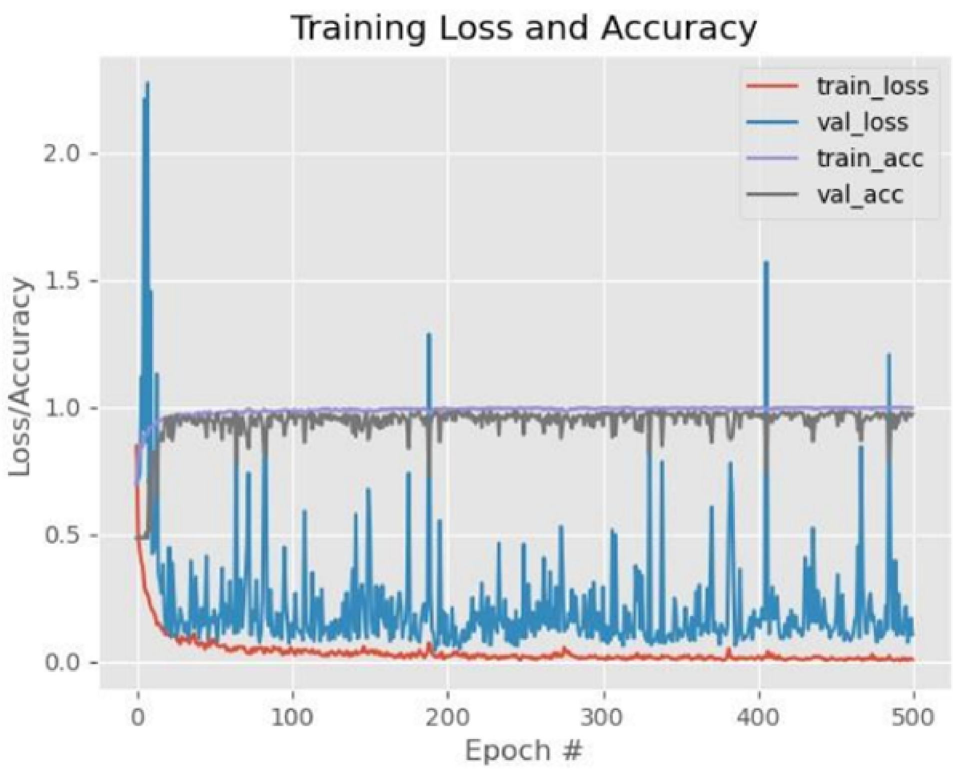

Fig. 5 Graphical representation of training loss accuracy with respect to epoch

4 Method and Experimental Results

CNN model allows for fine-tuning, with an option to make the last four layers trainable. Fine-tuning is a common practice in transfer learning, where the pre-trained layers are frozen to retain learned features, and only the top layers are trained on the target dataset shown in Fig. 2. Fine-tuning helps our model adapt to the specific characteristics of the CelebA dataset. The model includes dropout layers and kernel regularization (L2 regularization) in the fully connected layers. These techniques help prevent overfitting [19], essential when dealing with limited data [20]. The fully connected layers include batch normalization layers. Batch normalization helps stabilize and accelerate training by normalizing each layer's input. The custom loss function adaptive loss is implemented to handle imbalanced classes in the dataset. Using a weighted loss based on class weights addresses the potential issue of unequal representation of different facial expressions.

The model architecture includes several dense layers with batch normalization and activation functions (ReLU) [21] to capture and learn hierarchical features from the input data. The model uses the softmax activation function in the output layer, which is suitable for multi-class classification problems. In this case, there are seven

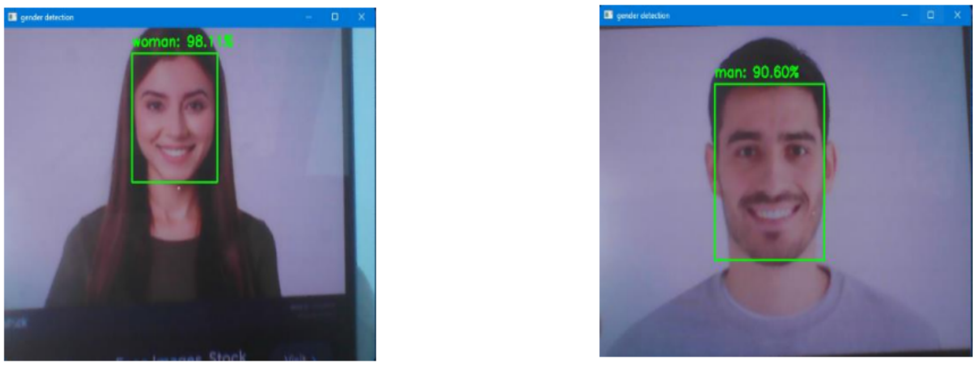

Fig. 6 Face detection accuracy result for women and men images on CelebA dataset

Table 2 CelebA Dataset [16] category-wise images accuracy using proposed model.

Category	Number of images	Training sample	Test sample	Accuracy(%)
Man	1173	938	234	90.60
Woman	1134	907	226	98.11

classes representing different facial expressions. This model is compiled using the Adam optimizer, a popular choice for training deep neural networks. Layer-wise pooling operations are shown in Fig. 4. This CNN model, which has fine-tuning, regularization, and adaptive loss handling, is designed to be effective for gender recognition on the CelebA dataset. The added layers act as a classifier on top of the features extracted by the pre-trained CNN model [22]. This approach is common in machine learning applications, where pre-trained models are powerful feature extractors that can be fine-tuned for new tasks. After training our proposed model and creating the history, we evaluated our model's performance and accuracy by testing it against a test dataset, and we have achieved an accuracy of around 90.60% for men and 98.11% for women reported in Table 2 and shown in Fig.6. Further, we tried to plot a graph for accuracy and loss using the history variable, which keeps track of each epoch in the form of a list shown in Fig. 5.

5 Conclusion

The gender detection system, created using the CelebA dataset using CNN, has shown excellent performance. Advanced data augmentation techniques such as rotation, flipping, and color modifications have increased the model's ability to generalize to real-world scenarios. Essential training techniques, such as batch normalization, regularisation, and hyperparameter tweaking, have reduced overfitting and increased robustness. The model's overall accuracy has been improved by addressing imbalanced class concerns through an adaptive loss function. This method has resulted in a highly accurate and dependable system for determining gender from photographs. The training procedure was further improved by applying optimization methods like Adam and the learning rate schedule. To maintain long-term performance, we will continue to monitor and improve the system going forward. Applications will include security, demographic analysis, and human-computer interaction. With policies in place to protect privacy, data security, and fairness, ethical issues continue to be crucial. To uphold our moral standards, we will communicate with stakeholders in the future. Overall, our system is a technically sound and socially responsible solution with the potential to impact various industries substantially.

References

[1] Mahesh, B.: Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet] **9**(1), 381–386 (2020)

[2] Francesca, P., Weldon, S.M., Lomi, A.: Lost in translation: Collecting and coding data on social relations from audio-visual recordings. Social Networks **69**, 102–112 (2022)

[3] G, A.S., Geerthik, S., Karthikeyan, R., Keerthana, G.: Face recognition based automated smart attendance using hybrid machine learning algorithms and computer vision. 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) (2024) https://doi.org/10.1109/icaaic60222.2024.10574896

[4] Sengupta, P., Mollah, A.F.: Dissected scene character recognition using hog descriptors. In: Internet of Things and Its Applications: Select Proceedings of ICIA 2020, pp. 199–209. Springer, ??? (2022)

[5] Praveena, M., Praveena, M., Srinija, K.G., Srinija, K.G., Meghana, A., Meghana, A.: Perception of age and gender detection by using hierarchical deep learning architecture through vision. 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) (2022) https://doi.org/10.1109/icscds53736.2022.9760865

[6] Zeba, S., Amjad, M.: Identification and location monitoring through live video streaming by using blockchain. Multimedia Tools and Applications, 1–28 (2024)

[7] Thangaraj, R., Thangaraj, R., Pandiyan, P., Pandiyan, P., Pavithra, T., Pavithra, T., Manavalasundaram, V.K., Manavalasundaram, V.K., Sivaramakrishnan, R., Sivaramakrishnan, R., Kaliappan, V.K., Kaliappan, V.K.: Deep learning based real-time face detection and gender classification using opencv and inception v3. 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA) (2021) https://doi.org/10.1109/icaeca52838.2021.9675635

[8] Naaz, S., Pandey, H., Lakshmi, D.C.: Deep learning based age and gender

detection using facial images. 2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) (2024) https://doi.org/10.1109/accai61061.2024.10601975

[9] S, S., Kumar, N.N.: Age and gender voice recognition using deep learning. International Journal of Scientific Research in Computer Science Engineering and Information Technology (2024) https://doi.org/10.32628/cseit2410336

[10] Kotadia, A., Kalavadiya, E., Pathak, L., Patel, T.: Real-time human profiling: Unveiling age, gender and emotions using deep learning. International Journal of Innovative Science and Research Technology (2024) https://doi.org/10.38124/ijisrt/ijisrt24aug894

[11] Sonthi, V.K., Sikka, R., Anusha, R., Devi, P.B.S., Kamble, S.S., Pant, B.: A deep learning technique for smart gender classification system. 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (2023) https://doi.org/10.1109/icacite57410.2023.10182683

[12] Wu, W., Yin, Y., Wang, X., Xu, D.: Face detection with different scales based on faster r-cnn. IEEE transactions on cybernetics **49**(11), 4017–4028 (2018)

[13] Alonazi, M., Alshahrani, H.J., Alotaibi, F.A., Maray, M., Alghamdi, M., Sayed, A.E.: Automated facial emotion recognition using the pelican optimization algorithm with a deep convolutional neural network. Electronics (2023) https://doi.org/10.3390/electronics12224608

[14] Mishra, P., Biancolillo, A., Roger, J.M., Marini, F., Rutledge, D.N.: New data preprocessing trends based on ensemble of multiple preprocessing techniques. TrAC Trends in Analytical Chemistry **132**, 116045 (2020)

[15] Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild

[16] Liu, Z., Luo, P., Wang, X., Tang, X.: Large-scale celebfaces attributes (celeba) dataset. Retrieved August **15**(2018), 11 (2018)

[17] Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., Shen, F.: Image data augmentation for deep learning: A survey. arXiv preprint arXiv:2204.08610 (2022)

[18] Shorten, C., Khoshgoftaar, T.M., Furht, B.: Text data augmentation for deep learning. Journal of big Data **8**(1), 101 (2021)

[19] Dasgupta, R., Chowdhury, Y.S., Nanda, S.: Performance comparison of benchmark activation function relu, swish and mish for facial mask detection using convolutional neural network. In: Intelligent Systems: Proceedings of SCIS 2021, pp. 355–367 (2021). Springer

[20] Zamir, M., Ali, N., Naseem, A., Ahmed Frasteen, A., Zafar, B., Assam, M., Othman, M., Attia, E.-A.: Face detection & recognition from images & videos based on cnn & raspberry pi. Computation **10**(9), 148 (2022)

[21] Dubey, S.R., Chakraborty, S.: Average biased relu based cnn descriptor for improved face retrieval. Multimedia Tools and Applications **80**(15), 23181–23206 (2021)

[22] Guo, G., Wang, H., Yan, Y., Zheng, J., Li, B.: A fast face detection method via convolutional neural network. Neurocomputing **395**, 128–137 (2020)