Journal of Information Systems Engineering and Management
2025, 10(31s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Accurate Age and Gender Prediction Using DNN Model from Real World Camera Feeds

P. Jayabharathi[1], K. Rohini[2]
[1,2] Department of Computer Science,
Vel's Institute of Science Technology & Advanced Studies, Chennai, Tamil Nadu, India

ARTICLE INFO

Received: 24 Dec 2024

Revised: 12 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

Introduction: The enormous rise of video and image data's has created a tremendous need for intelligent systems to autonomously understand and analyse information, while human interpretation is difficult. In social interactions, a person's face is crucial for expressing their identity and emotions. Compared to machines, humans are not very good at distinguishing between distinct faces.Predicting age and gender from live video broadcasts is a difficult problem with many potential uses. Traditional approaches are unable to detect faces with any degree of accuracy due to the wide variation in facial images.

Objectives: To develop a humanoid systems employing Deep Neural Network (DNN) with improved preprocessing and overfitting prevention for predicting the age and gender accurately..

Methods: The study uses a TensorFlow DNN for estimating the the age and gender of a person. The data processing including cleaning and data augmentation ensures the model's optimal performance. A Dropout Layer (DL) has been added to the feature extraction to prevent overfitting. The developed DNN model is evaluated using the face images in the Kaggle repository. Deep CNN (DCNN), Support Vector Machine (SVM), and K-Means clustering are compared with the suggested approach using the performance measures to determine the efficiency of the model.

Results: The findings show that, the TensorFlow DNN model outperforms the contrasted method with an accuracy, precision, recall, and F1-Scores of about 98%, 96%, 95%, and 75%, respectively, showing its efficiency in age and gender predictions. Moreover, the face detection capability of the system allows for the development of a central suggestion system for humanoid robots, thus making them more functional and coordinated according to categorized age groups.

Conclusions: The research presents a TensorFlow DNN model optimized for real-time age and gender prediction from camera feeds. The suggested model outperforms the algorithms such as DCNN, SVM, and K-Means clustering by embracing tedious data preprocessing methods, such as cleaning and augmentation, and leveraging benchmark datasets like IMDB and Adience.

Keywords: TensorFlow Deep Neural Network (DNN), Long Short term Memory(LSTM), Humanoid System, Suport Vector Machines(SVM), K-Means

INTRODUCTION

Advances in technology have made it easy and convenient to ascertain the gender and age of an individual using search trends and computer vision algorithms. These are above the normal applications of technology, and room is provided to include in human-like platforms for the aim of facilitating meaningful interaction and relationship with humans by machines. Humanoid models that are capable of identifying the gender and age of a human can reasonably adjust responses and behavior according to the human factor (Ounis et al., 2023).

In addition, a single system for recommendations on humanoid platforms enables enhanced functionality and synchronization in isolated age groups. This enables humanoid robots to provide responses and actions appropriate to the particular need and preference of a person by age groups, thus enabling more interactive and customized interactions (Devi et al., 2025)

With the addition of the age prediction and gender-prediction ability, we can create even more interactive and personal human-computer interactions. It not only further improves the experience of a user, but relations and ties of all sorts that can be constructed in applications covering everything from teaching and health care to customer support and entertainment become more substantial as well. As we continue to push the boundaries of what can be done with robotics and AI, the convergence of predictive capabilities is going to have enormous potential to alter how people engage with robots in the future((Esposito et al., 2024).

Chen et al., (2022) presented powerful and efficient face detection and recognition algorithms that and can be applied to video monitoring in real time. A system for managing medical data that incorporates video capture, image preprocessing, and facial recognition and identification is suggested. The most significant problems with face detection are complex backdrops, occlusion, and illumination. A wide variety of algorithms have been proposed to reduce these problems.

Rajendra et al., (2021) proposed a method based on basic image operations to implement the Data Augmentation Technique to expand the dataset's size. The gender of the user is then classified through the convolution neural network. The model has proven practically that the classifier's learning performance is improved by increasing the dataset and reducing the features.

Foggia et al., (2024) outlines a work on the application of multi-task neural networks (MTNs) to voice-based soft biometrics recognition in social robots, such as gender, age, and mood. MTNs remove the requirement for cloud-based solutions that add network latency by enabling effective audio signal analysis for a variety of functions on low-power embedded devices.

Abdolrashidi, et al., (2020) used CNNs to extract information that improved the accuracy of gender prediction significantly. CNN coupled with deep learning approaches have been used to reach cutting-edge performance. Extensive trials on the IMDB- WIKI dataset, the most readily accessible compendium of facial images with Gender labels, were used to determine the image- based Gender estimation.

Kumar et al., (2020) suggested a technique employing CNN (ConvNet), a DL algorithm, to extract features. The suggested ConvNet method needs much less preprocessing compared to other state of art methods. Despite the fact that the filters are constructed by manually, ConvNets may be trained to identify and use these features.

Bhat et al., (2019) designed a framework to determine the gender and age group of facial images using a DL framework precisely. This technique focuses on the most significant and relevant regions of the face with an attention mechanism to produce more accurate predictions. The method substantially improves the accuracy of their age prediction model by including gender predictions into their feature embedding with a use of multi-learning approach.

Elkarazle et al., (2022) presents a hybrid model employing CNN and an Extreme Learning Machine (ELM). The strengths of both the techniques are utilized in hybrid architecture. ELM classifies the intermediate results after CNN uses the input images to acquire the features and the ELM classifies it. This approach is tested with the use of two well-known datasets, MORPH-II and Adience Benchma.

OBJECTIVES

- To develop a technique for the prediction of age and gender using TensorFlow and a DNN.
- To evaluate the proposed framework's efficiency with the other well known methods SVM, and K-Means clustering, and to reduce overfitting in an attempt to enhance the generalization of the model.
- To enhance the performance of humanoid systems that leverages the model's predictive ability to allow humanoid robots to adjust their interactions according to user demographics.

METHODS

The study presents a TensorFlow DNN age and gender prediction framework from live camera streams, particularly in humanoid systems. Aside from the preprocessing operations of data cleaning and augmentation, the model utilizes advanced DL for precise predictions from the images acquired by the camera. To improve robustness and avoid overfitting, a Dropout Layer (DRL) is incorporated in the model structure strategically. Not only does this novel technique ensure accurate predictions but also facilitates effortless integration with humanoid platforms, wherein age and gender prediction abilities serve as the pivotal feature for customizing interactions according to individual users.

Data Balancing and Cleaning: Random Sampling: Randomly sampled enough number of samples from classes with sufficient observations so that there was a balanced ratio among the classes. This provides the model with sufficient representation by each age group for training.

Manual Filtering: Filtered out wrong annotated samples by hand from both classes to remove any inaccuracies and unreliability in the training data.

Data Augmentation: To fight the sparsity of the minor sample classes, we performed data augmentation by including extra samples from a benchmarked dataset (Adience dataset). The augmentation increased training data diversity and model generalizability to new samples.

Offline Data Augmentation: Applied some offline data augmentation techniques, such as right flipping, rotation, scaling, and noise injection, to further diversify the dataset.Each of these procedures exposes the model to a wider variation of face images and makes the model stronger consequently.

Online Data Augmentation: Augmenting data online at training time is performed by resizing the input images to 256x256 pixels and then center cropping them to 224x224 pixels. Pre-processing assists in obtaining uniformity of the input size and enables the model to learn invariant features regardless of age. Through these data cleaning and balancing methods, we tried to reduce the impact of sparsity in data and enhance the performance of the model in precise prediction of age by age group.

TensorFlow DNN Model

The DRLs are separated by MaxPooling (MP) and Rectified Linear Unit (ReLU) layers. DRLs can identify low-level features in input data and produce feature maps with a variety of image features.

Dropout Layer (DRL): Added to the model design, the dropout layer aims to combat the issue of overfitting. The layers probabilistically lose neurones at random after training, which lessens the model's dependence on specific features and increases its generality.

Sequence Learning Block : LSTM networks are arranged in three layers to form the sequence learning block (Telmem et al., 2024). LSTM's sequential processing makes it especially flexible for jobs like predicting age and gender from facial picture data. With 20 neurons per layer, LSTM has demonstrated remarkable accuracy in capturing patterns and long-distance correlations.

Output Layer: The size of the model's final output layer is determined by the prediction horizon. The number of output layer neurones in age and gender prediction issues is equal to the number of age and gender class neurones. The model can accurately predict from various age groups and gender types by dynamically varying the output layer's size.

Training Procedure: The model's weights are continuously updated during the training phase using stochastic gradient descent (SGD) (Upadhyay et al., 2024). The learning rate is continuously modified based on the validation loss to guarantee optimal convergence and keep the model from getting stuck in local minima. Additionally, early stopping is employed to avoid overfitting; training is stopped if, after a predefined number of epochs, the validation loss does not decrease. The model's performance is evaluated using test data after training.

Tensor Flow Model Architechture: Figure 1 displays the suggested architecture diagram. Using a camera, real-time face picture gathering is part of the data collection process. The relevant features are extracted from the photos using CNN. When learning compelx connections, non-linearity is introduced by the Rectified ReLU activation. By reducing the dimensionality, the pooling procedure maintains its key characteristics. The maps' contrast is further improved via normalisation. The CNN layer improved upon the features that the first layer had extracted. The flattened feature maps are processed by the fully connected layer in order to provide higher level representation. During training, the DL layer deactivates neurones to avoid overfitting. Then the FC layer computes the probability distribution over output classes using SoftMax. The final sotfmax probabilities determine age and gender predictions for each input facial image.

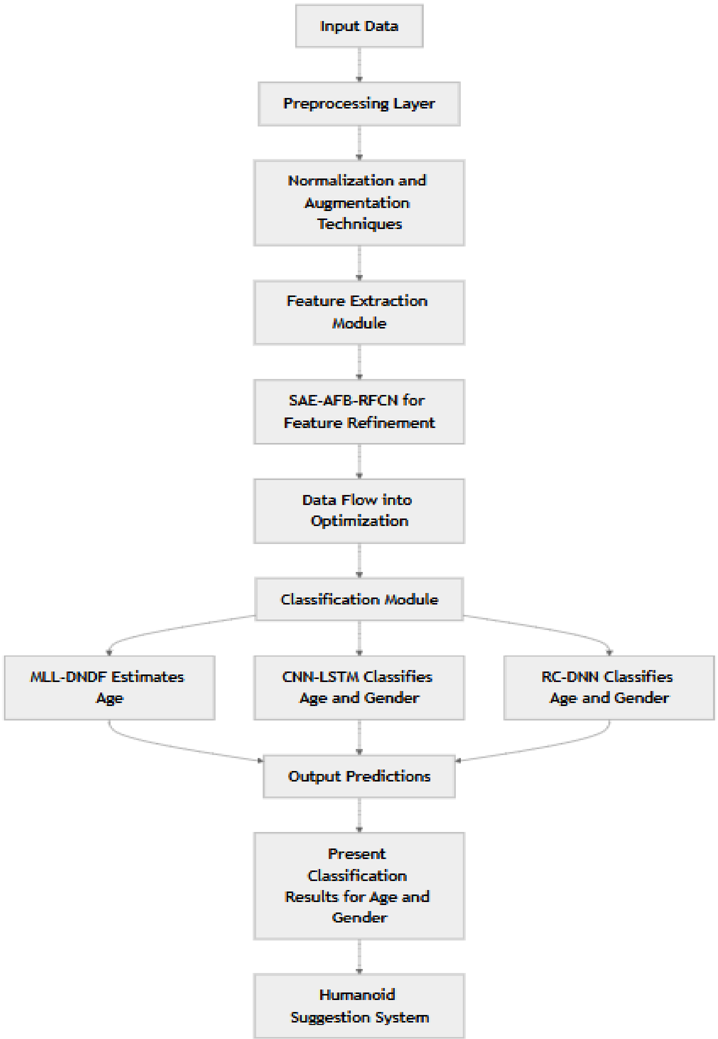

Figure 1. Proposed architecture

Training and Testing Overview

Initialization: Each network layer's weights are first set at random using a Gaussian distribution with a 0.01 standard deviation. Images and related labels from the benchmark dataset are utilised to train the network from the start rather than using pre-trained models.

Target Representation: The ground truth classes' sparse binary vectors are used to represent the training target values. Label vectors are made with the number of classes equal to the number of training images for tasks involving the categorisation of gender and age. For every training image, each label vector has a 1 in the ground truth class index and 0s elsewhere.

Network Training: In addition to utilising network architecture, two strategies are used to reduce overfitting. Dropout regularisation is used first, randomly setting the network neurones' output value to zero. The dropout ratio of one of the network's two DRL's is 0.5, which indicates that there is a 50% chance of a neuron's output value being zero. Second, each forward and backward training cycle includes a random cropping of 227x227 pixels from the input image. This method is similar to employing multiple mirrored and cropped versions of the image.

Prediction: Two techniques are applied to predict the age and gender of new faces:

- Center Crop: A centered cropped image of the face is fed to the network with dimensions of 227x227 pixels.
- Over-sampling: Five 227x227-pixel crop sections are extracted from the edges of the facial image, along with an additional crop region covering the center. Each of the five images, along with their horizontal reflections, is presented to the network. The final prediction is obtained by averaging the predictions from all these variations.

Resolving Image Quality Issues: Motion blurs, occulusions, and misalignment of the Adience photos affect the accuracy of the predictions. Instead of trying to enhance alignment quality, this is addressed by providing the network with numerous translations of the same face to accommodate for minor alignment issues.

RESULTS

Dataset: The images from IMDB and Adience dataset (Noor et al.,2024)were downloaded from Kaggel repository, which is a benchmark dataset for face images which has various real-world imaging constraints. It consisits of 26,580 images of 2,284 individuals under various age groups. For this prediction, pre-trained age and gender models have been used. The age deploys and gender_depoly prototxt files are used for training, which elucidate the network settings and the age net and gender net Caffe model files define the internal states of the parameters of the layers. Figure 2 shows some of the input images with actual age and the predicted results with age and gnder are shown in figure 3.

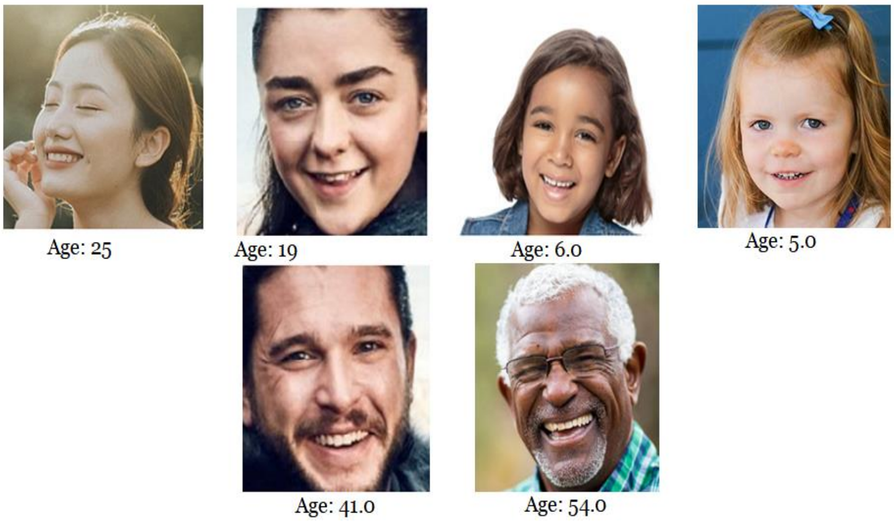

Figure. 2. Input Images with actual age

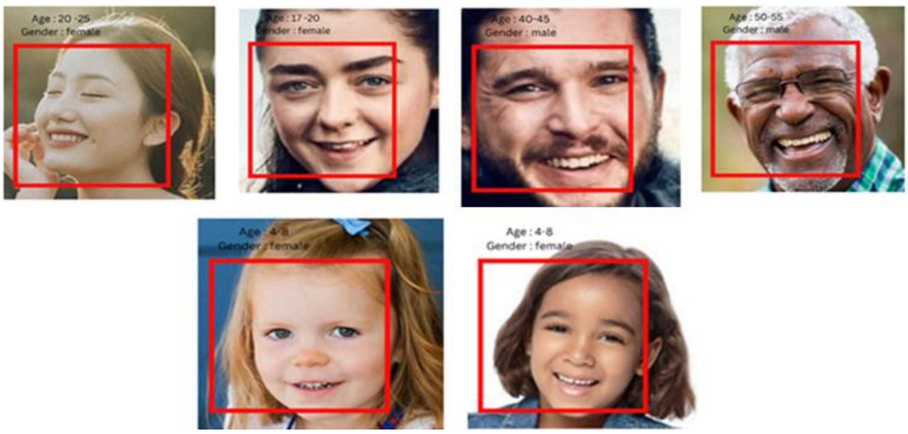

Figure. 3 Predicted Age and Gender for Input Image

Comparison Results: The suggested model's performance is contrasted with that of the LSTM, SVM, and K-Means methods with the performance metrics to evaluate the performance. The confusion matrix is shown in figure 4 and the Table 1 lists the performance metrics values.

Predicted	0-2	3-5	6-8	9-12	13-15	16-18	19-24	25-30	31-40	41-50	60+
Actual: 0-2	88	5	1	0	0	0	0	0	0	0	1
Actual: 3-5	4	90	3	1	0	0	0	0	0	0	0
Actual: 6-8	0	2	85	8	2	0	0	0	0	0	0
Actual: 9-12	0	1	5	88	4	2	0	0	0	0	0
Actual: 13-15	0	0	2	4	90	3	1	0	0	0	0
Actual: 16-18	0	0	0	2	4	92	3	1	0	0	0
Actual: 19-24	0	0	0	1	2	5	89	3	0	0	0
Actual: 25-30	0	0	0	0	1	3	4	88	4	0	0
Actual: 31-40	0	0	0	0	0	2	3	5	87	3	0
Actual: 41-50	0	0	0	0	0	0	2	4	6	85	3
Actual: 60+	1	0	0	0	0	0	1	2	2	5	89

Figure 4 : Confusion Matrix- Gender Classification (Binary Male/Female)

Table 1. Comparison Results for Various Algorithms

Metrics	DNN (%)	LSTM (%)	SVM(%)	K-Means(%)
Accuracy.	98	95	92	93
Precision.	96	94	90	91
Recall.	95	90	90	91
F1-Score.	95	92	90	90

DNN has the highest accuracy of 98%, properly classifying approximately 98% of occurrences. LSTM follows with 95%, which is still strong but lower than DNN. SVM and K-Means are marginally lower, at around 92% and 93% respectively. DNN leads again with a precision of 96%, indicating that when DNN predicts positive, it is correct 96% of the time followed by LSTM. SVM and K-Means lag behind but are close together, at 0.90 to 0.91. DNN has a recall of 95%, indicating that it accurately recognizes approximately 95% of the actual positives. LSTM is at 92%, a little decrease from DNN. SVM and K-Means perform equally, with values of 90% and 91%, respectively. DNN outperforms with a F1-score of 95%, showing its superior performance than other models. SVM and K-Means are marginally lower at 90% and 90%, respectively. DNN is the best-performing model across all metrics, implying that it has learned the data patterns most efficiently. LSTM performs well, particularly with sequential or time-series data, but falls somewhat below DNN. SVM and K-Means perform rather well but trail behind deep learning models, which are expected given that SVM is a classical classifier and K-Means is unsupervised.

CONCLUSION

The study presents a TensorFlow DNN model tailored for real-time age and gender prediction from camera feeds. By employing meticulous data preprocessing techniques, including cleaning and augmentation, and utilizing benchmark datasets like IMDB and Adience, the suggested model outperforms the performance of competing algorithms like DCNN, SVM, and K-Means clustering. With an outstanding accuracy of around 98%, precision of 96%, recall of 95%, and F1-Score of 75%, the TensorFlow DNN model showcases its robustness in accurately predicting age and gender. This achievement unlocks exciting possibilities for applications in various domains, including human-robot interactions and personalized services. For instance, imagine a humanoid companion adapting its behavior based on the age and gender of the individual it interacts with, offering tailored assistance and support. Such advancements herald a future where technology seamlessly integrates with human experiences, enhancing engagement and fostering meaningful connections.

REFERENCES

[1] Ounis HA, Ruhaiyem NIR, Badr AA, Abdul-Hassan AK, Alfadli IM, Binjumah WM, Altuwaijri EA, Nasser M. Multimodal Age and Gender Estimation for Adaptive Human-Robot Interaction: A Systematic Literature Review. *Processes*. 2023; 11(5):1488. https://doi.org/10.3390/pr11051488

[2] Devi, V. S., Ramisetty, U., Ramisetty, K., & Thimmareddy, A. (2025). Real-Time age, gender and emotion detection using YOLOv8. In ITM Web of Conferences (Vol. 74, p. 01015). EDP Sciences.

[3] Esposito A, Cuciniello M, Amorese T, Vinciarelli A, Cordasco G. Humanoid and android robots in the imaginary of adolescents, young adults and seniors. Journal of Ambient Intelligence and Humanized Computing. 2024 May;15(5):2699-718.

[4] Chen, J., Stantic, B., & Chen, J. (2022, December). Age prediction of social media users: case study on robots in hospitality. In International Conference on Robot Intelligence Technology and Applications (pp. 426-437). Cham: Springer International Publishing.DOI: https://doi.org/10.3390/pr11051488

[5] Rajendra, G., Sumanth, K., Anjali, C., Pardhasai, G., & Supraja, M. (2021, August). "Gender Prediction using Deep Learning Algorithms and Model on Images of an Individual". In Journal of Physics: Conference Series (Vol. 1998, No. 1, p. 012014). IOP Publishing.

[6] Foggia, P., Greco, A., Roberto, A., Saggese, A., & Vento, M. (2024). Identity, Gender, Age, and Emotion Recognition from Speaker Voice with Multi-task Deep Networks for Cognitive Robotics. Cognitive Computation, 1-11.

[7] Abdolrashidi, A., Minaei, M., Azimi, E., & Minaee, S. (2020). "Age and gender prediction from face images using attentional convolutional network". arXiv preprint arXiv:2010.03791.

[8] Kumar, Vijay. (2020). Age Prediction using Image Dataset using Machine Learning. International Journal of Innovative Technology and Exploring Engineering. 8. 107-113. 10.35940/ijitee.L1020.10812S319.

[9] S. F. Bhat, A. W. Lone and T. A. Dar, "Gender Prediction from Images Using Deep Learning Techniques," *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)*, Malatya, Turkey, 2019, pp. 1-6, doi: 10.1109/IDAP.2019.8875934.

[10] ELKarazle K, Raman V, Then P. Towards Accuracy Enhancement of Age Group Classification Using Generative Adversarial Networks. Journal of Integrated Design and Process Science. 2022 Apr 20;25(1):8-24.

[11] Tumpa, R. S., Khaliluzzaman, M., Hoque, M. J., & Tasnim, R. (2024). Revolutionizing Age and Gender Recognition: An Enhanced CNN Architecture. International Journal of Systematic Innovation, 8(4), 27-45.

[12] Telmem, M., Laaidi, N., Ghanou, Y., Hamiane, S., & Satori, H. (2024). Comparative study of CNN, LSTM and hybrid CNN-LSTM model in amazigh speech recognition using spectrogram feature extraction and different gender and age dataset. International Journal of Speech Technology, 27(4), 1121-1133.

[13] Upadhyay, S., BV, J. M. M., & Deepa, A. (2024, March). Human: Age and Gender Identification Using SFTA and Deep Learning Algorithm. In 2024 2nd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT) (pp. 01-05). IEEE

[14] Noor, A., Ahmad, W., & Shah, S. M. A. Deep Faces: Advancing Age and Gender Classification using Facial Images with Deep Features.