DOI: 10.6977/IJoSI.202412_8(4).0003
R. S. Tumpal, J. Hoque, etc /Int. J. Systematic Innovation, 8(4), 27-45 (2024)

Revolutionizing age and gender recognition: an enhanced CNN architecture

Rashna Sharmin Tumpal, Md. Khaliluzzaman[1]*, MD Jiabul Hoque[2], Roshni Tasnim[1]

[1] Department of Computer Science and Engineering, International Islamic University Chittagong, Chattogram-4318, Bangladesh

[2] Department of Computer and Communication Engineering, International Islamic University Chittagong, Chattogram-4318, Bangladesh

* Corresponding author E-mail: khalil@iiuc.ac.bd

(Received 27 November 2023; Final version received 22 May 2024; Accepted 31 July 2024)

Abstract

The recognition of age and gender in images has had a significant impact on computer vision, particularly with the increasing application of digital platforms. Deep Convolutional Neural Networks (DCNNs) show promising performance. However, they demand substantial computational resources, limiting their deployment in real systems, especially those with constraints on resources or cost. This study performs a sensitivity analysis in order to show how some changes in the architecture of the network can influence the tradeoff between accuracy and performance. For that, in this work, we have investigated various CNN architectures and introduced an effective convolutional neural network (CNN) model to precisely predict gender and age attributes using the Adience dataset. Amidst unfiltered and diverse image sources from various devices, our model exhibits an impressive 92.24% accuracy across eight distinct age groups and two gender categories. The model's strength lies in its adeptness at handling intricate image data, allowing comprehensive adjustments to age and gender parameters. By employing advanced deep learning techniques and comparing with MiniVGGNet, our model showcases exceptional performance.

Keywords: Adience dataset, Age and Gender recognition, Computer Vision, Deep Convolutional Neural Network (DCNN), MiniVGGNet.

1. Introduction

Facial attribute recognition, particularly age and gender classification, has garnered considerable attention in recent years due to its multifaceted applications in various domains such as security, marketing, computer vision, and human-computer interaction (Smith et al, 2021; Parkhi etc., 2015). The ability to accurately infer age and gender from facial features has been a longstanding challenge. However, advancements in deep learning, especially with convolutional neural networks (CNNs), have revolutionized this field (Liu and Chen, 2020).

Researchers have thoroughly investigated gender and age identification using a variety of handcrafted features and machine learning techniques (Smith et al, 2021). The pursuit described encompasses a wide range of implications, which include applications in the field of robotic programming that necessitate the use of demographic information. Additionally, it extends to surveillance systems that engage in human-computer interaction, human pose estimation, automated analysis of gender-related behaviors, and biometrics, as well as access management in secure environments (Parkhi etc., 2015). These various applications highlight the significance and potential impact of this pursuit. The customized Convolutional Neural Network (CNN) model designed for age and gender classification holds substantial practical consequences and has the potential for diverse applications in real-world situations. On digital platforms, it can improve user experiences by offering customized content, recommendations, and advertisements that are specifically designed for the user's demographic profile. The model in surveillance systems assists in the identification of demographic patterns, enhancing crowd management and enabling focused monitoring in public settings (Toshev and Szegedy, 2015). Rapid patient classification in healthcare applications can enhance the development of tailored treatment programs. Businesses can enhance their marketing and advertising efforts by doing a more precise analysis of customer demographics. Smart devices can utilize this framework to provide customized interactions, such as modifying the responses of voice assistants according to the user's age and gender. In addition, educational platforms have the ability to tailor learning materials to better align with the age and gender of students, hence improving their learning experience. These applications highlight the broader importance of the model, demonstrating its potential influence in other fields by offering accurate age and gender categorizations (Smith et al, 2021).

In recent years, there has been a notable demonstration of the impressive capabilities of deep learning and convolutional neural networks (CNNs) (LeCun at al., 2015). These developments have demonstrated outstanding performance in the field of computer vision, namely in tasks like image classification and recognition. The ability of convolutional neural networks (CNNs) is to iteratively acquire complex information from the first to the final layers. That resulting in the creation of more abstract and advanced representations, which signifies a significant paradigm shift in the domain (Liu and Chen, 2020).

The contemporary state of age and gender recognition pose complex issues that persistently influence research endeavors within the field of computer vision. The domain continues to encounter persistent challenges in the areas of apparent age estimate, accurate age prediction, robust face recognition, and combined identification and verification (Parkhi etc., 2015; Yang etc., 2018). It has wide-ranging societal and technological consequences. Precise identification improves user experiences on digital platforms by providing tailored content and services. It enhances public safety by supporting surveillance systems in doing demographic analysis. Within the healthcare field, the use of patient profiling enables rapid assessment of patients, hence enhancing the development of treatment strategies. Marketing businesses derive advantages by employing focused methods that are tailored to specific demographic profiles. Educational technology modifies content to more effectively align with students' age and gender. Smart devices enhance their intuitiveness by analyzing and comprehending user demographics. This research fulfills the need for sophisticated and dependable techniques of classifying age and gender. It fosters advancements in computer vision and artificial intelligence, facilitating technological growth (Rothe at el., 2016; Sun at el., 2014).

The utilization of convolutional neural networks (CNNs) and deep learning methods has been propelled as transformational instruments in the discipline in response to these issues. These technologies present intriguing opportunities for improving accuracy, reducing biases, and achieving resilience against fluctuations in illumination, posture, facial expression, and biases within datasets. The convergence of convolutional neural networks (CNNs) and deep learning is positioned at the forefront of technological progress, enabling breakthroughs in addressing the complexities associated with age and gender recognition tasks.

This paper introduces a customized developed convolutional neural network (CNN) model. Its design is customized for the accurate classification of gender and age. The model underwent training and evaluation on the Adience dataset. The model showcased exhibits a high level of precision, achieving a remarkable accuracy level of 92.24%. The analysis was conducted on a varied dataset comprising of eight distinct age groups and two gender divisions. We improved the performance of our model by utilizing advanced deep learning techniques. We conducted rigorous assessments of MiniVGGNet. This approach required making incremental adjustments to parameters in a total of ten models. Consequently, we achieved an exceptional level of accuracy. Our approach focuses performing a comprehensive evaluation utilizing the Adience standard. This highlights the model's capacity to efficiently process a diverse array of raw image data acquired from various sources, such as smartphones. This approach enables significant adjustments to age and gender variables. Therefore, it enhances the complexity of the dataset and establishes a robust framework for evaluation. Furthermore, we have incorporated a distant-view Convolutional Neural Network (CNN) model for the sake of comparison. This further strengthens the exceptional performance of our proprietary approach. The importance of our model in the field of gender and age prediction is reinforced by doing comparative comparisons with other prominent methods. This substantiates the superiority of our model in these areas. The contribution of the paper is provided below:

- Investigates various CNN architectures and introduced an effective convolutional neural network (CNN) model to precisely predict gender and age attributes using the Adience dataset.
- Model's performance on the Adience benchmark dataset was subjected to a rigorous examination.
- Implementation of a baseline Convolutional Neural Network (CNN) model is employed to establish a comparative benchmark for distant view images.
-

- Model's efficacy is demonstrated by its exceptional achievement of 92.24% accuracy over eight age and two gender groups.
- Our suggested model has exhibited superiority over state-of-the-art approaches, hence confirming its robustness and excellence in predicting gender and age.

The rest of the paper is summarized as follows. The related work is detailed at section 2. The proposed method is presented at section 3. The experimental analysis is explained at section 4. The paper is concluded at section 5.

2. Related works

Deep convolutional neural networks (CNNs) have played a key part in advancing age and gender classification, which are critical components of facial recognition technology. This comprehensive literature review examines several approaches, systems, and tactics. Those have been proposed in recent decay with the goal of improving accuracy in recognition tasks. Many authors in various articles examine innovative techniques. The strategies include shallow convolutional neural network (CNN) designs, ensemble learning, and attribute identification algorithms. The primary objective is to improve the precision of age and gender predictions. This thorough analysis investigates the development of these approaches. It emphasizes their noteworthy findings. Furthermore, it examines the diverse applications in the domain of face attribute recognition. Such as, Patil at el. (2021) show that performance can be significantly improved by learning representations while using deep convolutional neural networks (CNN) and Extreme Learning Machines (ELMs). The methodology employed in this study utilizes CNN and ELM to acquire fundamental representations. The CNN module is responsible for extracting features from input images, while the ELM module is responsible for classifying the intermediate results. This research advised to use CNN for giving more accurate results of age and gender classification.

In the study, Abir etc. (2023) put out a comprehensive facial recognition framework that integrates identification, age, and gender recognition. Their research showcases the capacity of machine learning to make complex predictions. Srivastava and colleagues (2023) provide research indicating a notable rise in the

They demonstrated a system for real-time gender recognition on an embedded device that achieves very high accuracy (up to 98.73% in the field).

utilization of deep neural networks, namely convolutional neural networks (CNNs), to effectively demonstrate the accuracy of age and gender categorization. The observed accuracy surpasses that of other previously established claims. The model consists of three convolutional layers, two max pooling layers, and two dense layers. The study revealed that the mean accuracy for age recognition is 82.2%, while the mean accuracy for gender recognition is 94.10%. The authors Balan etc. (2022) have introduced the concept of a Pike neuron based convolutional neural network (SN-CNN). The accuracy of age and gender recognition was assessed by the evaluation of the orthopantomogram dataset, yielding a significant degree of performance. Abood etc. (2023) introduced a fresh methodology in their research, wherein they employed the AlexNet model to categorise age and gender. The performance evaluations were conducted by utilising the UKTFace dataset. The current investigation successfully attained a reasonable degree of precision in the classification of age and gender. The introduction of a novel convolutional neural network (CNN) classification technique was presented by Mamatkulovich etc. (2023) in their paper. This algorithm demonstrates notable advantages over current approaches, as it exhibits much reduced training parameters and training time. In spite of its lower complexity, our model exhibited superior accuracy in classifying age and gender on the UTK-Face dataset. In a previous study, Sharma etc. (2022) put out an enhanced convolutional neural network (CNN) model designed for the assessment of age and gender. The authors conducted a comparative analysis of their method with the current state-of-the-art approach using the UKTFace dataset, and achieved a respectable level of accuracy.

The authors Agbo-Ajala at el. (2020) explore techniques for automatically determining a person's age and gender from raw images of their faces. They framed the challenge as one of multiple classes and suggested using a loss function based on categorization to guide model training. They studied the precision with which this model could classify data using the original dataset. Proposed model obtains the state-of-the-art performance in both age group and gender classifications. Shallow deep convolutional neural network architecture for Gender Recognition from face images was proposed by Greco at el. (2020) research. Author also conducted sensitivity analysis to demonstrate how varying the network's architecture can modify the efficiency-accuracy tradeoff.

A deep learning based network called GRA_Net is presented for estimating a person's age and gender based on their appearance by Garain at el. (2021). The author views the issue of age prediction as a hybrid of categorization and regression. The authors Nada at el. (2020) propose a new method for verifying a user's gender and age range as determined by their image. Their Deep Learning-based inputs now include a layer validator that checks inputs like user image, gender, and birth date. Using a Convolutional Neural Network, they were able to determine a person's gender and approximate age from just one image. After analysis, it was shown that this approach accurately predicted both gender and age. Hassan at el. (2020) propos a different model for the gender and age classification problem employing multiple sub-CNN is proposed. Here, each sub-CNN was evaluated separately, in addition to the final model evaluation that applied the voting ensemble. The idea behind employing several smaller CNNs and combining their predictions into one larger one is to increase accuracy. They receive a diverse representation for the photographs Feature. Therefore, improving classification accuracy necessitates developing a more accurate model of the age estimation problem. Compared to employing a single CNN model, the error rate in all age group classifications is lower when the voting ensemble method is used.

Benkaddour etc. (2021a) introduce three CNN network models with various architectures were tested using the number of filters and number of convolution layers. Authors showed that CNN networks significantly increase the system's usability and recognition accuracy. Later, Benkaddour etc. (2021b) proposed another CNN method for age and gender recognition to reduce the complexity of the proposed model. The model is evaluated based on the Adience dataset and achieved the accuracy for age is 91.75% and for gender is 95.6% respectively. The authors of Levi at el. (2015) explored and proposed a novel approach by exploiting the neighborhood information among image samples, which showed that accurate attribute detection is possible by utilizing the automatically generated neighborhood graph topology. A DNN regression is shown in Human Posture Estimation through Deep Neural Networks by Toshev at el. (2014). This research achieves high precision pose estimations and has completed their work on real-world images on four academic benchmarks. Author propose a DNN- based pose predictors, which allowed to increased precision of joint localization, where they focuses on the relevant region of the image, which is then cropped, and the pose displacement regression is combined with the sub-image. During training, simulated predictions are generated.

Ozbulak at el. (2014) introduce a method, where, a collection of Deep hidden Identity features (DeepID) are learned using deep learning and are derived from the final hidden layer neuron activations of deep Convolutional networks (ConvNets). It has a 160-dimensional DeepID system that can accurately predict 10,000 classes at the end of a cascade. DeepID reduced the number of neurons in the top layer and extracted features as it learned to recognize classes in the training set. Additionally, it eliminates a few neurons from the hidden layers. They combined the Bayesian strategy with DeepID-based neural network training for face verification. In the paper Sun at el. (2016), transferring cutting-edge deep Convolutional Neural Network models is investigated for autonomous age and gender prediction with two widely used soft biometric features. Two different face recognition models—one general and one specific to the face recognition domain—were chosen to explore whether these models can be used to classify people by age and gender. For age and gender classification tasks, few findings were made by comparing generic AlexNet-like and domain-specific VGG-Face CNN models against task-specific GilNet CNN models on the difficult Adience benchmark. According to experimental findings, transferred models perform 7% and 4.5% better for age and gender categorization tasks than the most recent GilNet model respectively.

A CNN for Pedestrian Gender Recognition is suggested by Ng at el. (2013) using a convolutional neural network that had been trained to discriminate between both genders of pedestrians. They achieved an accuracy of 80.4% on a dataset consisting of frontal and rear images of pedestrians' whole bodies using a relatively simple architecture and minimal picture preprocessing. A deep CNN was proposed by Raza at el. (2017) to determine the gender of a pedestrian. The method employs a pre-existing deep decomposition neural network to examine the pedestrian's images. Next, the backdrop is removed from the parsed photos, and whole-body and head-and-shoulders images are created. Later, we feed these two categories of images into the recommended fine-tuned CNN model. Input images of the full body are sorted by gender based on whether they were taken from the front, the back, or a combination of the two. Images are split into eight categories based on the attire on their upper bodies. The proposed method was found to be more effective in making predictions across a wide range of sub-classifications.

According to author knowledge, this paper (Cao at el., 2008) is the first to investigate gender recognition using still images of the human body. Their PBGR system, which combines part-based representation with ensemble learning, can identify the gender from a single frontal or back view image with an accuracy of 75.0% and also act robustly in the presence of slight misalignment. In the paper by Deng at el. (2014), a brand-new, massive dataset (PETA) containing 19000 images and 61 annotated attributes is presented. By utilizing the neighborhood information between image samples, they investigated and presented a novel strategy to deal with such big and heterogeneous data in the context of attribute classification. Author demonstrated that using the automatically inferred neighborhood graph topology, accurate attribute detection is possible. Ranjan at el. (2017) present a multi-task CNN-based technique for simultaneous face detection, face alignment, posture estimate, gender and smile is classification, and age estimation. Although both their method and Hyper Face use the MTL framework, their way performs far better. This work demonstrates how domain-based regularization and network initialization from face recognition tasks help subject-independent tasks.

Author provided a complete method for recognizing age, gender, and emotions by Dehghan at el. (2017). They demonstrate that their innovative deep architecture can outperform competing commercial and academic algorithms on a number of benchmarks when combined with their substantial, internally acquired data. In order to conduct attribute identification of a set of pedestrian images, straightforward deep network architecture was presented by Kurnianggoro at el. (2017). Using a public dataset, experiments were carried out to evaluate the effectiveness of this network. The results show that the suggested network performs better with fewer parameters used. Additionally, it is discovered that several characteristics, such as the presence of a backpack and clothing color, cannot be accurately identified because of self-occlusion or color confusion. As a result, it can be concluded from the experiments in this work that the particularly designed network performs better than the general purpose network. Hence, it is suggested to use a specially designed network for person re-identification research while considering human attributes as one of its components. The RoR-152+IMDB-WIKI-101 model by Zhang at el. (2017), while combined with age-group mechanisms, demonstrates superior performance compared to traditional CNN architectures. This highlights its ability to effectively estimate age and gender in complex image datasets encountered in real-world scenarios. The fine-tuning method used to Adience dataset enhances and reinforces its already exceptional performance, establishing it as a state-of-the-art model.

The literature investigate incorporates a comprehensive examination of various methodologies in age and gender recognition. The authors argue in favor of utilizing deep convolutional neural networks (CNNs) as a means to improve accuracy. They propose various techniques, including as shallow architectures, multi-CNN ensembles, and compact models, to achieve this objective. The researchers investigate many tasks, including the identification of gender in pedestrians, estimation of posture, and detection of attributes. They employ innovative approaches that prioritize domain-specific learning, dataset size, and specialized network designs to enhance performance across diverse recognition tasks. Presented here is a tabular representation that concisely summarizes the data extracted from the present state of the art. The summary of the related work is shown in Table 1.

3. Research method

This section of our research work represents the detailed concept on recognizing the gender and age estimation by using deep neural network. The method proposed in this paper represents a significant improvement in the field of facial attribute recognition, namely in the area of gender and age classification. By utilizing convolutional neural networks (CNNs), our methodology seeks to tackle the complex task of precisely classifying gender and age features in facial images. By utilizing the distinct characteristics of the Adience dataset as described by Eidinger etc. (2017)

Table 1. Summary of the related works

Reference	Focus	Key Findings
Abir etc. (2023)	Integration of age and gender recognition inside a facial recognition framework.	A comprehensive framework for facial recognition has been developed, which integrates identification, age, and gender recognition. This framework places particular emphasis on the capacity of machine learning to make intricate predictions.
Srivastava etc. (2023)	Use of convolutional neural networks for reliable age and gender classification.	Demonstrated growing interest in using CNNs for accurate age and gender categorization; using a model with three convolutional layers, achieved mean accuracy rates of 82.2% for age and 94.10% for gender, both improvements above prior work.
Balan etc. (2022)	Recognition of both age and gender is remarkably accurate.	Presented a convolutional neural network (CNN) based on Spiking neurons (SN-CNN), with impressive results on the orthopantomogram dataset for identifying age and gender.
Abood etc. (2023)	Accuracy in determining age and gender is satisfactory.	Developed a novel method for age and gender categorization using the AlexNet model, which showed promising results on the UKTFace dataset.
Mamatkulovich etc. (2023)	Creation of a new convolutional neural network method for age and gender discrimination	has presented a unique CNN classification technique that, while attaining better accuracy on the UTKFace dataset, requires fewer training parameters and less time overall to implement.
Sharma etc. (2022)	Fairly precise gender and age determinations	Built a better CNN model for determining age and gender, with results that are competitive with state-of-the-art methods on the UKTFace dataset.
Patil at el. (2021)	Age and Gender recognition using CNN and ELM	Convolutional Neural Networks (CNNs) and Extreme Learning Machines (ELMs) utilized to acquire fundamental representations for better outcomes
Agbo-Ajala at el. (2020)	Age and Gender Multiclass Classification	Using unedited frontal imagery and multiclass classification algorithms, achieved state-of-the-art performance in age and gender classification.
Greco at el. (2020)	Condensed CNN for Identification of Gender	Suggested a small CNN architecture that can be used in embedded devices and achieves high real-time gender recognition accuracy of up to 98.73%.
Garain at el. (2021)	GRA_Net for Age and Gender prediction	Presents GRA_Net, a network that uses facial image classification and regression to handle age prediction.
Nada at el. (2020)	Verification of Gender and Age with Images	Using CNN, a double-check layer validator was proposed to reliably determine gender and age from single-person images with remarkable results.
Hassan at el. (2020)	Group of Sub-CNNs for Gender and Age	showed increased accuracy in age estimate and classification by the use of a voting ensemble to combine predictions from several sub-CNNs.
Benkaddour etc. (2021a)	CNN Topologies for Enhanced Precision	Investigated several CNN architectures, showing notable advances in recognition systems' usability and accuracy.
Benkaddour etc. (2021b)	Simple CNN architecture to reduce the complexity	A simple CNN model is formed to estimate the age and gender based on the Adience dataset.
Levi at el. (2015)	Identifying attributes through neighborhoods	Demonstrated precise attribute identification using the neighborhood graph topology among picture samples that was automatically created.
Toshev at el. (2014)	Human Posture using DNN Regression	Using DNN regression and concentrating on joint localization and simulated predictions during training, the model were able to estimate human posture with great precision.

Ozbulak at el. (2014)	Features of DeepID for Face Verification	Created a 160-dimensional DeepID system by combining Bayesian method with deep learning and neuron reduction for facial verification.
Sun at el. (2016)	Transferring CNN Models according to Gender and Age	Examined the transfer of cutting-edge CNN models for gender and age prediction; on tasks requiring gender and age classification, the models outperformed current ones by 5% and 7%, respectively.
Ng at el. (2013)	Gender Recognition in Pedestrians	Suggested a simple CNN architecture that achieved 80.4% accuracy in gender classification of pedestrians with little preprocessing.
Raza at el. (2017)	Gender Analysis Using Pedestrian pictures	A deep CNN model was used to evaluate gender utilizing different viewpoints and categorize apparel, resulting in improved prediction performance.
Cao at el., (2008)	Gender Identity in the Human Body	Designed PBGR system that demonstrates robustness to misalignment and 75.0% accuracy in gender recognition from human body photos.
Deng at el. (2014)	Finding attributes in Large-Sample Datasets	Demonstrated efficient attribute classification by presenting a method for neighborhood information-based correct attribute detection in a big dataset (PETA).
Ranjan at el. (2017)	Multi-Task CNN for Diverse Features	Presented a multi-task CNN method that outperformed Hyper Face utilizing domain-based regularization for face identification, alignment, posture, gender, smile categorization, and age estimation.
Dehghan at el. (2017	Robust Framework for Age, Gender, and Emotions	Outperformed current algorithms in identifying emotions, age, and gender by utilizing a deep architecture that was novel and heavily data-driven.
Kurnianggoro at el. (2017)	Identification of Attributes in Images of Pedestrians	Improved attribute identification efficiency with fewer parameters by proposing a specialized network for person re-identification study, which is especially helpful in self-occlusion settings.
Zhang at el. (2017)	Introduced two mechanisms to enhance age estimation	Proposed RoR-152+IMDB-WIKI-101, a novel CNN-based method for age and gender estimation.

and implementing advanced deep learning methods, we have developed a tailored model. This model explores the complex correlation between classes, hence enabling precise classification.

The aim of this methodological innovation is to surpass conventional methods. Our objective is to improve the accuracy and reliability of our predictions by developing a more extensive and simple framework for predicting age and gender. This innovative approach aims to make a significant contribution to the field of facial attribute recognition by introducing a new paradigm. By doing so, we want to improve the precision and effectiveness of convolutional neural network (CNN) models in handling intricate classification tasks. Fig 1 presents the workflow diagram of this research; which is explained in following with a sequential way.

Here, the dataset we are going to use in our research work is Adience dataset. It contains 19322 images which resolution is 816 by 816 with the horizontal and vertical resolution of 96 dpi and 24 bits. Fig 2 represents some sample image of different attribute class from the Adience dataset. Input images are close-view image.

DOI: 10.6977/IJoSI.202412_8(4).0003
R. S. Tumpal, J. Hoque, etc /Int. J. Systematic Innovation, 8(4), 27-45 (2024)

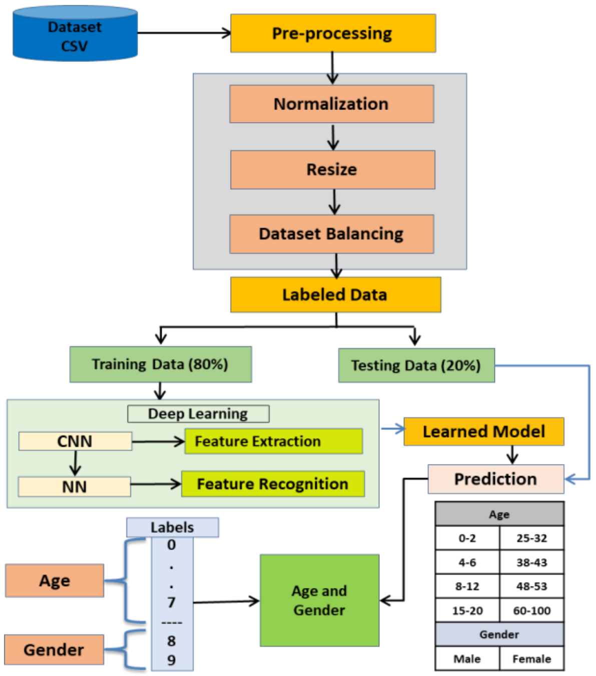


Fig 1. Workflow diagram of the age and gender recognition framework with data preprocessing.

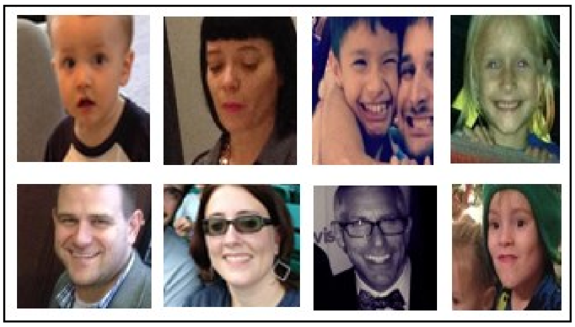


Fig 2. Sample image of different attribute class from Adience dataset

According to the steps to get preprocessed data, images are rearranged in the class label first. Within 30 age classes, we create 8 age classes, and from 3 gender classes, we take 2. We include all the age classes in those eight classes. Our age classes are 0-2, 4-6, 8-12, 15-20, 25-32, 38-43, 48-53, and 60-100, and our gender classes are male and female. Now, after reading the entire csv file, the image has been converted into channel 1, i.e., gray scale image. After that, the images are divided by 255.0, minus 0.5, and multiplied by 2.0 for scaling. After scaling, resize the original image into 75 by 75 images and give the image ID from 1 to increase, plus 1 for the next image. Then create five pickle files for each of the five CSV files. Now read the pickle files and make a directory. By reading the csv file, the image is divided into different folders, which are divided into age class and then male and female class, respectively. When reading every label line from the final label text file, the newline characters should be stripped from the end of every line to prevent extra labels from showing up. Then rename the image according to the label of age, class, and gender, and then add a random number. All the images are in jpg format. The image file name is like (0-2)_m_503.jpg, which means age is 0 to 2 and it is male. Some preprocessed image data is shown in Fig 3. Finally, labeled data are divided into train and test data, following the partition of 80% and 20%.

Fig 3. Some preprocessed image data from all class

3.1 Proposed CNN model

To achieve the best CNN model for predicting age and gender, we have applied different CNN layers and dense layers. From which we can estimate the best model for achieving our goal. For that, the model is trained with the various CNN and dense layers, namely NewNet1, NewNet2, … ,NewNet9. All the CNN layers are contained filters of size 3x3. The training and testing accuracy for different configurations are given in Table 2. From these entire networks, the NewNet3, NewNet8 and NewNet10 performed better. However, we select model 8 which is named here as NewNet8 because it gives higher training and testing accuracy with small number of parameter, where all other takes huge number of parameter and gives low training and testing accuracy. So finally, we choice the model NewNet8 is our proposed model. For train all the models Adam optimizer (Kingma at el., 2014) and categorical cross entropy (Liu and Chen, 2020) for loss function are used. For each model here used 50 epochs with the batch size of 128.

Based on the NewNet8 model, now we discuss the proposed convolutional network architecture given in Fig 4. It contains four convolutional layers and two fully connected layers that use the extracted features from the previous convolutional layers and acquire from them and thus perform the classification at the output layer.

Here, Close view RGB images of size 75 x 75 with three channels will be input to this convolutional neural network model. The first convolutional layer consists of 32 filters of size 3 x 3 and 3 channels each. The SAME padding is used here i.e. the input images are zero padded in such a way that the filters convolve over every pixel of the input image. For SAME padding the output image after convolution is the same as the input image. The 32 filters convolve the input images of size 75 x 75. Because of the SAME padding, the output feature maps after convolution of size 75 x 75 and 32 channels. To introduce non-linearity the output feature map is run through a ReLU activation function. This activation function turns any negative value into zero. Thus it makes every value non-negative. The max pooling filter is of size 2 x 2 and it is used to reduce the dimensionality of the feature maps. This filter moves by a stride of 2. And the output image size is 37x37x32.

The second convolutional layer contains 64 filters of size 3 x 3 and 64 channels each. The SAME padding is also used here. 64 filters are convolving the input

Table 2. Testing and training accuracy for the different model configurations

Model	CNN Layers	Dense Layers		Parameters	Train Accuracy (%)	Test Accuracy (%)
		Neurons	Dropout	Trainable		
NewNet1	64-128-256	1024 512	0.5 0.3	76,665,226 76,663,306	93.46	91.24
NewNet2	64-128-256	1024 512	0.5 0.4	76,665,226 76,663,306	93.39	89.19
NewNet3	32-64-128-256	1024 512	0.5 0.3	17,700,554 17,698,570	94.39	92.16
NewNet4	64-128-256	1024 256	0.5 0.3	76,399,242 76,397,834	93.54	91.84
NewNet5	64-128-256	1024 256	0.5 0.4	76,399,242 76,397,834	93.28	91.57
NewNet6	64-64-128-256	1024 256	0.5 0.4	17,454,026 17,452,490	94.50	92.15
NewNet7	32-64-128-256	1024 256	0.5 0.4	17,434,570 17,433,098	94.55	91.90
NewNet8	**32-64-128-256**	**1024** **256**	**0.5** **0.3**	**17,434,570** **17,433,098**	**94.60**	**92.24**
NewNet9	32-64-128-256	1024 512	0.4 0.3	17,700,554 17,698,570	94.56	91.85
NewNet10	32-64-128-256	1024 256	0.4 0.3	17,434,570 17,433,098	94.39	92.15
NewNet11	64-128-256	1024 1024	0.4 0.3	77,197,194 77,194,250	93.17	91.78
MiniVGGNet	32-32-64-64	512	0.5	10,690,858 10,689,450	91.01	91.11

DOI: 10.6977/IJoSI.202412_8(4).0003
R. S. Tumpal, J. Hoque, etc /Int. J. Systematic Innovation, 8(4), 27-45 (2024)

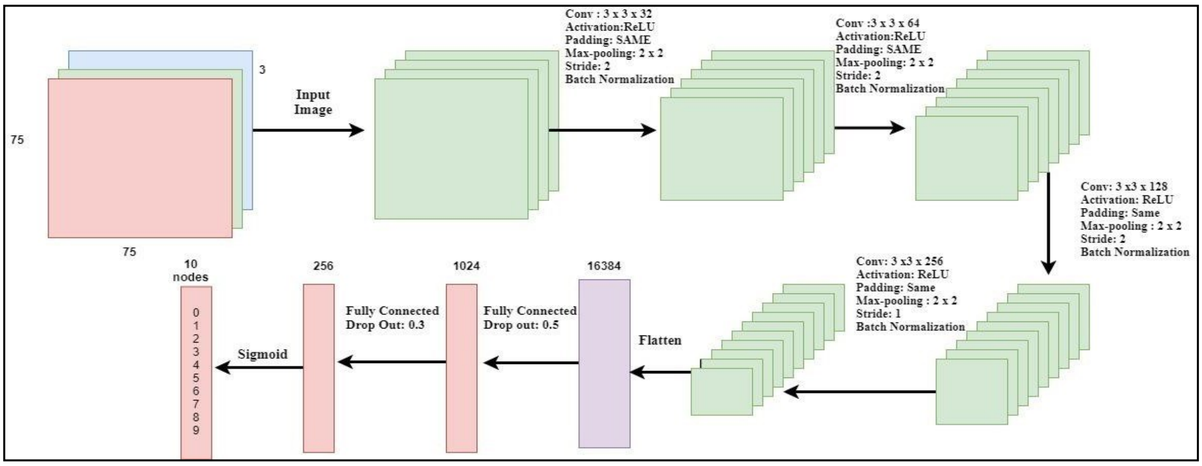


Fig 4. Proposed Convolutional Neural Network model architecture

images of size 37 x 37, and the output feature maps after convolution are of the size 37 x 37 and 64 channels. ReLU activation function is also applied to the feature maps. Max pooling filter of size 2 x2 which moves by a stride of 2, perform max pooling operation on the feature maps and output after this are images of size 18 x 18 and 64 channels.

The third convolutional layer consists of 128 filters of size 3 x 3 and 128 channels each. Here the SAME padding is also used. 128 filters convolve the input images of size 18 x 18 and the output feature maps after convolution is of size 18 x 18 and 128 channels. ReLU activation function is also applied to the feature maps. Max pooling filter of size 2 x2 which moves by a stride of 2, perform max pooling operation on the feature maps and output after this are images of size 9 x 9 and 128 channels.

In the similar way, the fourth convolutional layer is applied with the 256 filter size of 3x3. In this layer the Max pooling filter of size 2x2 that modes by a stride of 1. After the pooling operation the size of the image is 8x8 with 256 channels. The result of the last pool layer is flattened into a victories form which contained 16,384 elements (8x8x256) are followed by a fully connected (FC) layer of 1024 neurons. This vector's each element is fully connected with 1024 neurons of this layer. Fully connected layers are used to use the final features extracted from the previous convolution and pooling layers to classify the image into the respective classes in the output layer. The second FC layer is followed by the first FC layer with the neuron of 512. These FC layers learn from the features.

As the model is overfitting to the training data, to drop the neurons to prevent the model from overfitting the dropout layer is used. We have applied dropout on fully connected layers. In this model, we have used two dropouts. Here, Dropout 1= 0.5, Dropout 2= 0.3. To improve the model performance and solidity we have used Batch Normalization. It normalizes the input of every layer. Here, the batch normalization is 100. To improve the learning FC-256 layer is used. It is also used to increase the accuracy of the prediction. It contains 1024 neurons. FC-256 layer is fully connected to the output layer. It has 10 nodes representing the 10 class of the dataset. The attribute score for each of the attributes is then computed in the 10 nodes respectively. This is how the proposed CNN model shall work with the far view images we provide to this model.

3.2 Optimizer and loss function

The optimizer used in this work is Adam (Kingma at el., 2014) optimizer with the learning rate of 0.001. The loss function evaluates a classification model's performance or translates decisions to their related costs. Here we used categorical cross-entropy (Liu and Chen, 2020) loss for the models as our task is multiclass classification. The loss function is presented in equation (1).

$$Loss = -\sum_{i=i}^{N} log\hat{\boldsymbol{y}}_{\boldsymbol{i}} \qquad (1)$$

Here, the variable $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ represents the predicted value derived from the output of the model for the i[th] sample or class. The loss function evaluates the variance between what is expected and the true target values. The given expression corresponds to the categorical cross-entropy loss function, commonly utilized in classification tasks involving many classes (N) for prediction.

4. Results and discussion

This section contains a detailed experimental analysis of our proposed system. As we discussed previously, the dataset we have used to train and test our model is the Adience dataset. It contains close-view images. The Adience dataset contains 19,322 images labeled with their gender and age. Faces are divided into five folds. Ages are divided into different groups, which are 0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, and 60+. Gender is divided into male and female classes. The original dataset has 30 age labels, but we cut the labels 'none' and blank. And include other age labels into our 8-year-old class. It captures variations in pose, light, appearance, and more. We use 19,322 images from the dataset for training and 3068 for testing. The original size of the images is 816 by 816, and we used them by resizing them to 75 by 75.

For doing our research work efficiently and correctly, from preprocessing our dataset to evaluating the performance of our model, we used some packages, tools, and a development environment. We need a Python language for easily converting our ideas into code. Python is a programming language that is interpreted at a high level. It has the largest collection of packages for implementing machine-learning algorithms. It is the most mature and well-supported programming language in the area of machine learning to achieve greater productivity with systematic efforts. For implementing computer vision, Python allows developers to automate tasks that involve visualization. We used many packages and libraries of Python, like Numpy, Scikit, OpenCV, TensorFlow, Keras, Pandas, and so on.

In our proposed model, we used the Adience dataset for evaluating our model using the Keras Python library. It contains all the methods for optimizing and losing functions. We used the Adam optimizer and categorical cross-entropy for the loss function. The same padding is used for every convolutional and max pooling layer. Batch size is 100, and the number of epochs is 50 for every experiment. Dropouts are 0.5 and 0.3 after two fully connected layers. Batch normalization is used for every convolutional layer. We preprocessed the Adience dataset and augmented the dataset. Then we use 19,322 of those images for training and 3068 for testing purposes.

Table 3. Comparison of the Classification report among the models

Network	Class	0-2	4-6	8-12	15-20	25-32	38-45	48-53	60-100	M	F
NweNet1	Precision(%)	73	79	74	56	62	57	52	83	92	90
	Recall(%)	91	62	72	67	78	18	18	54	88	93
	F1-score(%)	81	69	73	61	68	52	26	65	90	91
NewNet2	Precision(%)	77	70	79	88	63	38	32	50	83	92
	Recall(%)	72	69	51	22	61	73	26	61	92	82
	F1-score(%)	75	69	62	35	62	50	29	55	87	87
NewNet3	Precision(%)	84	86	66	70	64	66	47	81	89	96
	Recall(%)	85	66	82	52	83	52	45	54	96	89
	F1-score(%)	85	74	73	60	72	58	46	65	92	92
NewNet4	Precision(%)	85	78	72	65	63	62	49	67	89	94
	Recall(%)	81	73	75	57	79	47	44	63	94	89
	F1-score(%)	83	75	74	61	70	54	46	65	91	91
NewNet5	Precision(%)	78	85	69	51	67	60	58	73	89	93
	Recall(%)	89	62	76	73	71	56	26	49	94	90
	F1-score(%)	83	71	73	60	69	58	36	59	91	91
NewNet6	Precision(%)	91	79	77	72	62	59	49	64	94	92
	Recall(%)	76	72	69	48	79	62	42	71	91	95
	F1-score(%)	83	75	73	58	69	60	45	67	92	93
NewNet7	Precision(%)	88	83	71	71	64	68	40	62	87	95
	Recall(%)	84	71	77	50	84	44	38	68	95	87
	F1-score(%)	86	77	74	59	72	55	39	65	91	91
NewNet8	**Precision(%)**	**88**	**79**	**69**	**61**	**68**	**63**	**62**	**71**	**92**	**91**
	Recall(%)	**79**	**75**	**79**	**70**	**75**	**58**	**34**	**54**	**93**	**93**
	F1-score(%)	**83**	**77**	**74**	**65**	**71**	**60**	**44**	**61**	**92**	**93**
NewNet9	Precision(%)	73	81	79	73	66	68	51	51	95	88
	Recall(%)	90	67	70	60	79	53	47	77	86	95

	F1-score(%)	81	74	74	66	72	60	49	61	90	92
NewNet10	Precision(%)	93	74	82	71	60	65	46	67	90	95
	Recall(%)	70	80	71	49	83	49	49	62	95	90
	F1-score(%)	80	77	76	58	70	55	47	64	92	92
NewNet11	Precision(%)	74	79	60	59	72	58	64	61	92	92
	Recall(%)	84	65	81	71	68	54	17	70	91	93
	F1-score(%)	79	71	69	62	70	56	27	65	92	92
MiniVGG Net	Precision(%)	82	77	79	71	55	48	41	70	93	89
	Recall(%)	79	69	48	41	78	57	26	59	88	94
	F1-score(%)	80	73	60	52	64	52	32	64	90	92

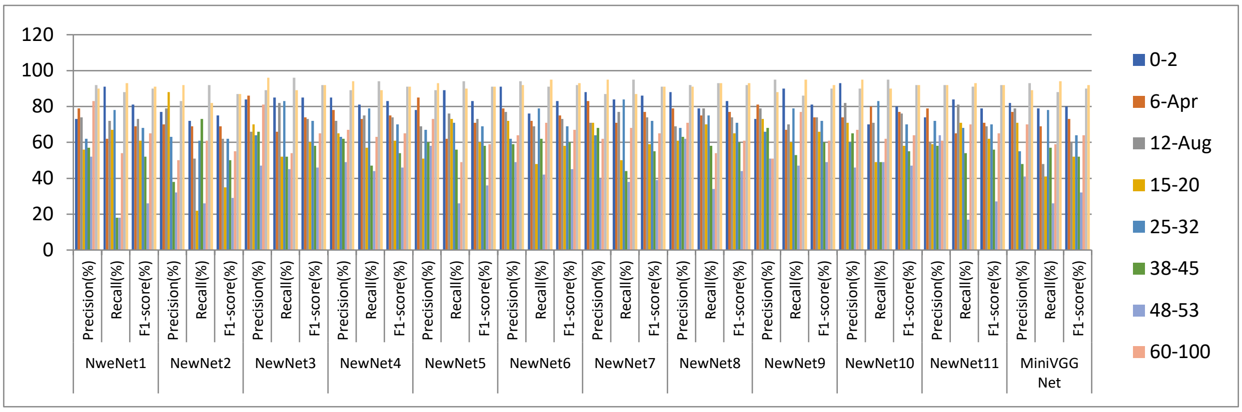


Fig 6. Comparison of the recall, precision and F1-Score among the models

4.1 Experimental results

As we discussed before we experiments 11 model with our dataset and we get different accuracy with different number of parameter. We build 10 models from our own with several numbers of convolutional layer and parameter. From all those, NewNet8 is selected as our proposed model, because of its high training and testing accuracy. Classification report comparison among 11 models is presented in Table 3. The table shows all the networks precision, recall, F1-Score and support in percentage. The values are also shown in Fig 6.

For NewNet8, the accuracy for the age labels (0-2),(4-6),(8-12),(15-20),(25-32),(38-43),(48-53),(60-100), and for the gender labels (female and male), respectively, is 92.24%. Recall for the age labels (0-2), (4-6) (8-12), (15-20), (25-32),(38-43),(48-53),(60-100) are 79%, 75%, 79%, 70%, 75%, 58%, 34%, 54% respectively and gender labels i.e., female and male are 93% and 93%, respectively. Age labels (0–2), (4-6), (8–12), (15–20), (25–32), (38–43), (48–53), and (60–100) have F1-scores of 83%, 77%, 74%, 65%, 71%, 60%, 44%, and 61%, respectively. Gender labels for women and men have F1-scores of 93% and 92%, respectively. These values are comparing with all other network which is better than other. That's why we choose NewNet8.

Fig 7 shows the validation accuracy and all the training for 11 networks and MiniVGGNet. The training accuracy for NewNet1, NewNe2, NewNet3, NewNet4, NewNet5, NewNet6, NewNet7, NewNet8, NewNet9, NewNet10, NewNet11, and MiniVGGNet is 93.46%, 93.39%, 94.39%, 93.54%, 93.28%, 94.50%, 94.55%, 94.60%, 94.56%, 94.39%, 93.17%, and 91.01% respectively. The testing accuracy for NewNet1, NewNe2, NewNet3, NewNet4, NewNet5, NewNet6, NewNet7, NewNet8, NewNet9, NewNet10, NewNet11, and MiniVGGNet is 91.24%, 89.19%, 92.16%, 91.84%, 91.57%, 92.15%, 91.90%, 92.24%, 91.85%, 92.15%, 91.78%, and 91.11% respectively. From these we see that accuracy is greater for NewNet8. Fig 8 show that, all the training and validation loss for 11 network and MiniVGGNet. From all this we see that NewNet8 gives less loss then other network.

DOI: 10.6977/IJoSI.202412_8(4).0003
R. S. Tumpal, J. Hoque, /Int. J. Systematic Innovation, 8(4), 27-45 (2024)

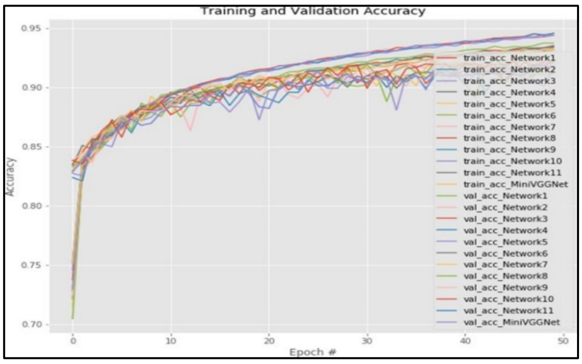


Fig 7. Validation accuracy and training curve for 11 model and MiniVGGNet

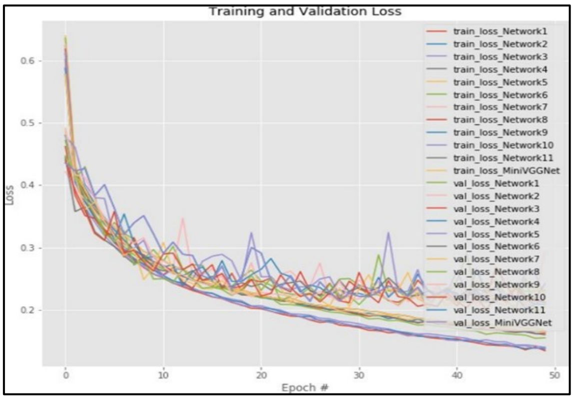
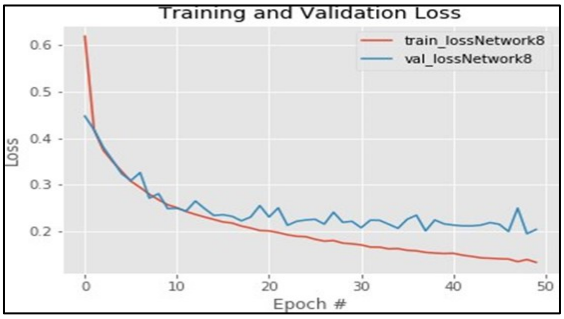
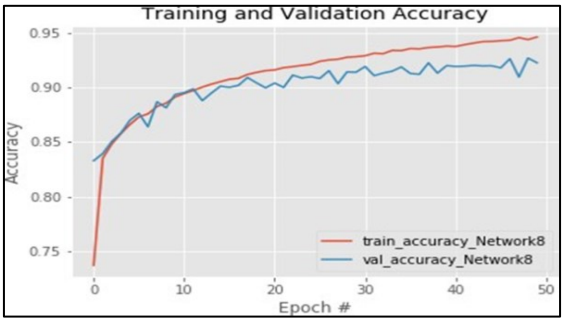


Fig 8. Training and validation loss curve for 11 models and for MiniVGGNet



(a)



(b)

Fig 9. Validation and training Loss and Accuracy Curve for proposed model: a) Loss Curve, and b) Accuracy Curve

The curve of Fig 9 shows the validation and training loss and accuracy for NewNet8 which is our proposed model.

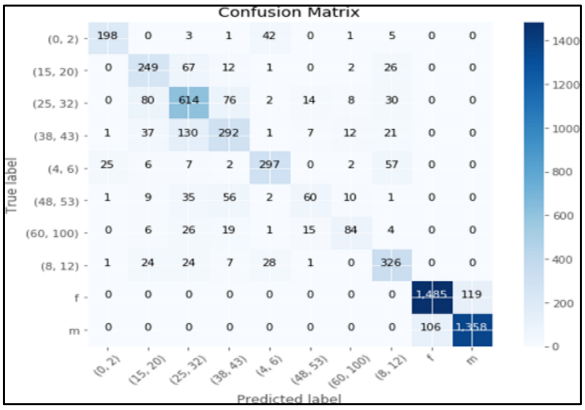


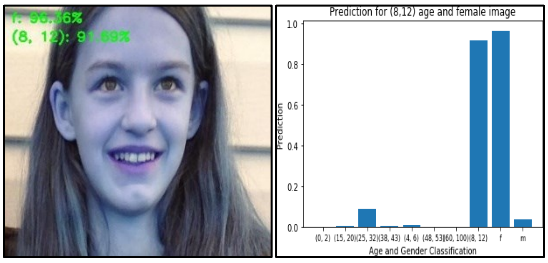
Fig 10. Confusion matrix of the proposed model

The model achieves a training accuracy of 94.60% and a testing accuracy of 92.24%. It consists of four convolutional neural network layers and two fully connected layers, with a very modest number of parameters. Notably, this model outperforms all 11 recommended models. Fig 10 presents the confusion matrix of the model employed in the present research. Based on the observation of Fig 10, it is revealed that the age category labeled as (8-12) demonstrates the highest level of accuracy in estimation, reaching 79.32%. The labels (48-53) and (60-100) have the lowest classification accuracies, specifically 34.48% and 54.19% respectively. The labels (48-53) and (60-100) exhibit a high degree of misclassification when compared to the labels (38-43) and (25-32), which correspond to values of 56 and 26, respectively. The accuracy of correctly estimating the age label (0-2) is 79.52%, whereas the accuracy of correctly estimating the age label (4-6) is 75%. The label (15-20) is incorrectly identified as label (25, 32) with an accuracy rate of 18.77%. The gender label "male" is classified with a higher accuracy rate of 92.76%, whereas the gender label "female" is classified with a little lower accuracy rate of 92.58%. There is a misclassification rate of 7% between male and female genders. The proposed model has a high level of accuracy in predicting the (0-2), (25-32) and (8-12) age groups. In the context of gender classification, the utilization of age labels within the ranges of 4-6, 15-20, 38-43, and 48-53 has been found to yield more accurate outcomes for the male and female classes, respectively.

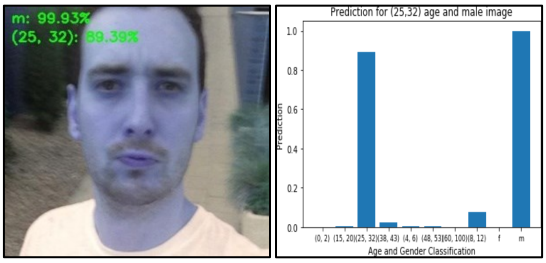
4.2 Experimental demonstration

In this section, we have presented some test samples that are classified the proposed model correctly. The images with different ages and genders are collected for testing is belonging to the separate classes whose image backgrounds are also different.

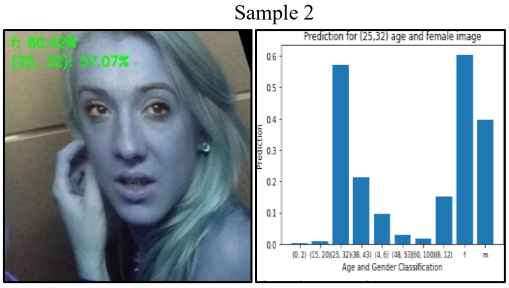
In Fig 11, the Sample 1 presents an image that is labeled with an age range of 8-12 and a gender label of Female. The Sample 2 presents an image that is annotated with an age range of 25-32 and a gender label of Male. The third sample presents an image that is distinguished by an age range of 25-32 and is labeled as Female in terms of gender. In a similar vein, Sample 4 exhibits an image that showcases an age range spanning from 38 to 43, accompanied by a gender label of Male. Likewise, Sample 5 portrays an image characterized by an age range of 48 to 53, along with a gender label of Male. Sample 6 presents an image that has been assigned an age classification of 15-20 and a gender classification of Male. Remarkably, all of these photos were accurately anticipated in relation to their age and gender characteristics.



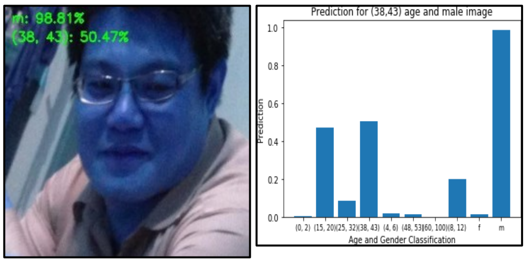
Sample 1



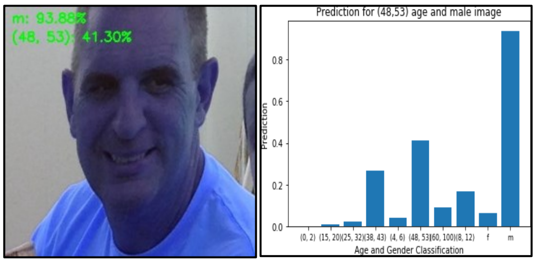
Sample 2



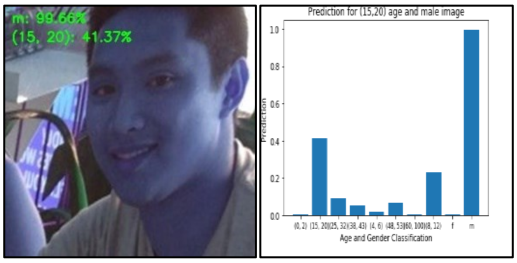
Sample 3



Sample 4



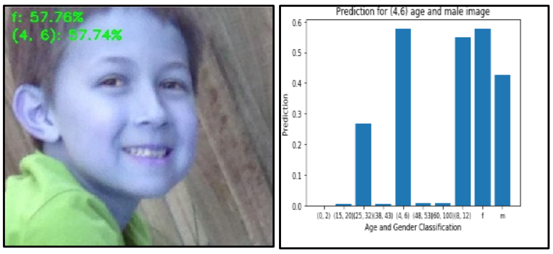
Sample 5



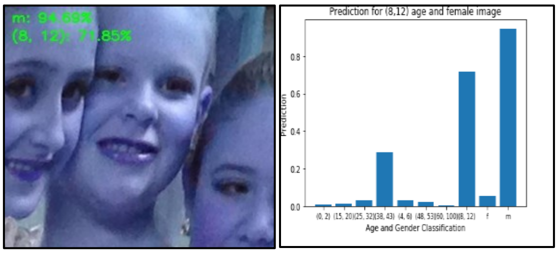
Sample 6

Fig 11. Samples of correctly identify images with proposed model: a) Image with prediction label, and b) histogram of the prediction

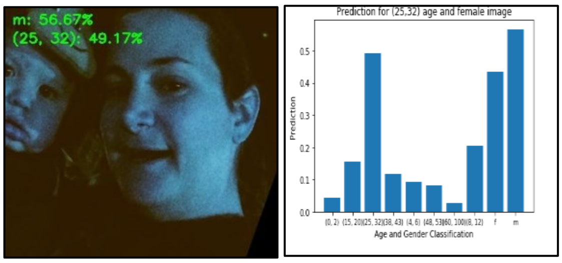
Fig 12 displays a collection of images representing different classes, showcasing situations where is classification has occurred. Significantly, it is obvious that both gender and age misclassifications are present in these situations. In regard to gender prediction, it is observed that Samples 7, 8, and 9 are misclassified. Conversely, misclassifications pertaining to age prediction are evident in Samples 10, 11, and 12. A notable observation arises from these results is that our model exhibits a greater level of accuracy in predicting gender labels in comparison to age labels. The misclassifications seen can be due to the presence of challenging conditions, like image blurriness, low resolution, and occlusions often encountered in the Adience benchmark images that were processed by our method.



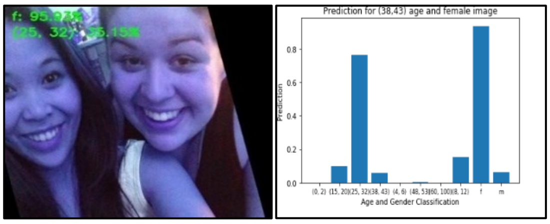
Sample 7



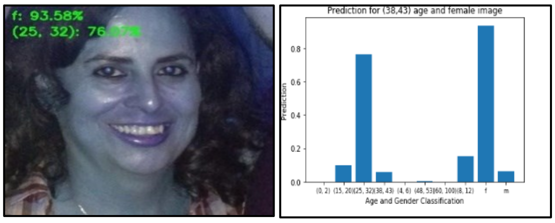
Sample 8



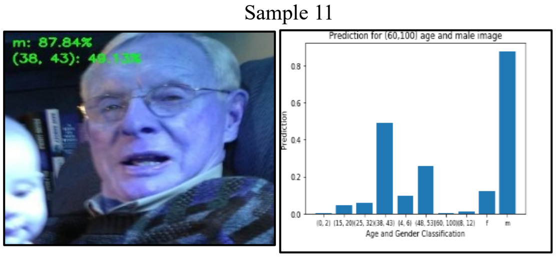
Sample 9



Sample 10



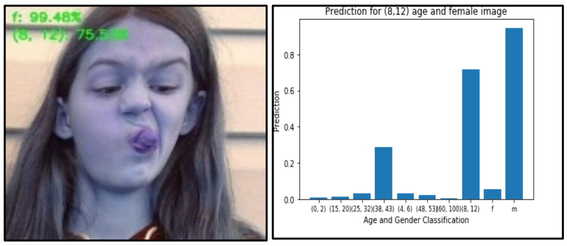
Sample 11



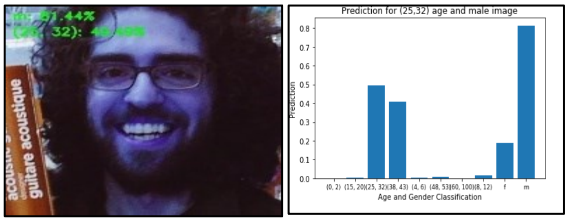
Sample 12

Fig 12. Samples of wrong prediction images with proposed model: a) Image with prediction label, and b) histogram of the prediction

In Fig 13, Sample 13 portrays a visually complex representation featuring a young female individual falling between the age ranges of 8 to 12 years. In a similar vein, Sample 14 presents a multifaceted depiction of a male individual within the age range of 25 to 32. Despite the inherent challenges associated with interpreting these images, our model demonstrates a remarkable ability to accurately predict both age and gender features.



Sample 13



Sample 14

Fig 13. Some challenging Samples that are correctly identified with proposed model: a) Image with prediction label, and b) histogram of the prediction

4.3 Discussions

In the method we have proposed, it has been observed that integrating both gender and age as features gives improved outcomes. This enhancement is achieved by using of straightforward model architecture, along with the utilization of Adience datasets, which consist of a substantial collection of around 19,322 images. Both classes have high confidence levels as well. Table 4 depicts the utilization of the ROR model (Zhang at el., 2017) in a sequential manner. Initially, the ROR model is pre-trained on the ImageNet dataset. Subsequently, it undergoes fine-tuning on the IMDB-WIKI101 dataset to enhance its learning capabilities. Finally, the model is further fine-tuned on the Adience dataset. At last, the ROR-152+IMDB-WIKI101 model, employing two mechanisms, has obtained state-of-the-art results on the Adience benchmark, with an accuracy of 79.99 ±2.23. The DAGER (Dehghan at el., 2017), and MiniVGGNet models achieve accuracies of 91.00% and 91.11%, respectively, when evaluated on the Adience Dataset.

GRA_Net (Nada at el., 2020) perform the operator both for age and gender separately and achieve the accuracy for age and gender 65.1± 2.1 and 81.4±0.6 percent respectively. In the study conducted by the authors Smith at el., (2021), design was employed on a Convolutional Neural Network (CNN) and Extreme Learning Machines (ELMs) utilized to highlight its efficacy in achieving improved results. The implementation of this architecture led to a notable enhancement in the accuracy of age and gender classification, with an average accuracy of 90.2±1.2. Ozbulak at el., (2016) proposed a method where the accuracy of age estimation was achieved at a rate of 92.00%. A model is developed using DCNN (Greco at el., 2020) for the recognition of gender. The model showed the accuracy of 84.45%. Srivastava and colleagues (2023) proposed a method that achieved the separate accuracy for age and gender was 82.2% and 94.10% respectively. However, the average accuracy was 88.15%. Previously, Benkaddour etc. (2021b) suggested a method for reducing the CNN model's complexity and achieved the separate accuracy for age and gender was 91.75% and 95.6% respectively. However, the average accuracy for age and gender was 93.68%.

In the present research, the MiniVGGNet and baseline Network, namely NewNet8, were employed for analysis on the Adience dataset. The MiniVGGNet and proposed model NewNet8 demonstrate competitive accuracies of 91.11% and 92.24% respectively, surpassing other networks and specific state-of-the-art models.

Table 4. Comparison with state of the art results of age and gender classification on Adience dataset

Method	Age	Gender	Accuracy (%)
CNN (Srivastava and colleagues, 2023)	Yes	Yes	Age: 82.2 Gender: 94.10
CNN-ELM (Smith at el., 2021)	Yes	Yes	90.2±1.2
Shallow CNN (Benkaddour etc., 2021b)	Yes	Yes	Age: 91.75 Gender: 95.6
DCNN (Greco at el., 2020)	No	Yes	84.45
GRA_Net (Nada at el., 2020)	Yes	Yes	Age: 65.1±2.1 Gender: 81.4±0.6
GilNet (Levi at el., 2015)	Yes	Yes	87.5±3.35
Ft- VGG-Face + SVM (Ozbulak) at el., 2016)	Yes	Yes	92.00
DAGER (Dehghan at el., 2017)	Yes	Yes	91.00
ROR-34 + IMDB-WIKI (Zhang at el., 2017)	Yes	Yes	79.99 ±2.23
MiniVGGNet	**Yes**	**Yes**	**91.11**
Propose Model (NewNet8)	**Yes**	**Yes**	**92.24**

5. Conclusions

This paper provides insights into the recognition of age and gender in images, which is essential for the advancement of computer vision. The sensitivity analysis performed in this paper reveals the impact of architectural changes on the tradeoffs between accuracy and performance. The model's exceptional adaptability and sophisticated parameter modifications are seen in its ability to achieve an accuracy of 92.24% across various age and gender categories using the Adience dataset. By utilizing sophisticated deep learning methods and conducting comparisons with MiniVGGNet, our model exhibits 1.13% higher performance. Our model demonstrates

exceptional performance in accurately predicting age categories such as (0-2), (8–12) and (25–32) with the accuracy value of 79.52%, 79.32%, and 74.51%, respectively. Additionally, it exhibits lower accuracy in predicting age ranges of (15–20), (38–43), and (48–53) with the accuracy of 69.79%, 58.17%, and 34.48%, respectively. For gender classification, the model performs 0.18% better on the male class with compare to the female class. In the future, our attention will shift towards boosting our Convolutional Neural Network (CNN) structure by employing unsupervised pre-training techniques in order to enhance its classification capabilities. Furthermore, our objective is to assemble an augmented dataset that includes a greater diversity of diverse classes in order to enhance the capabilities of our model.

References

Abir, I. M., Zaki, H. F. M., & Ibrahim, A. M. (2023). Evaluation of simultaneous identity, age and gender recognition for crowd face monitoring. ASEAN Engineering Journal, 13(1), 11-20. https://doi.org/10.11113/aej.v13.17612

Abood, Q. K. (2023). Predicting Age and Gender Using AlexNet. TEM Journal, 12(1).

Agbo-Ajala, O., & Viriri, S. (2020). Deeply learned classifiers for age and gender predictions of unfiltered faces. The Scientific World Journal, 2020.

Balan, H., Alrasheedi, A. F., Askar, S. S., & Abouhawwash, M. (2022). An Intelligent Human Age and Gender Forecasting Framework Using Deep Learning Algorithms. Applied Artificial Intelligence, 36(1), 2073724.

Benkaddour, M. K. (2021b). CNN based features extraction for age estimation and gender classification. Informatica, 45(5).

Benkaddour, M. K., Lahlali, S., & Trabelsi, M. (2021a). Human age and gender classification using convolutional neural network. In 2020 2nd international workshop on human-centric smart environments for health and well-being (IHSH) (pp. 215-220). IEEE.

Cao, L., Dikmen, M., Fu, Y., & Huang, T. S. (2008, October). Gender recognition from body. In Proceedings of the 16th ACM international conference on Multimedia (pp. 725-728).

Dehghan, A., Ortiz, E. G., Shu, G., & Masood, S. Z. (2017). Dager: Deep age, gender and emotion recognition using convolutional neural network. arXiv preprint arXiv:1702.04280.

Deng, Y., Luo, P., Loy, C. C., & Tang, X. (2014, November). Pedestrian attribute recognition at far distance. In Proceedings of the 22nd ACM international conference on Multimedia (pp. 789-792).

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. Journal of machine learning research, 12(7).

Eidinger, E., Enbar, R., & Hassner, T. (2014). Age and gender estimation of unfiltered faces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1681-1688).

Garain, A., Ray, B., Singh, P. K., Ahmadian, A., Senu, N., & Sarkar, R. (2021). GRA_Net: A deep learning model for classification of age and gender from facial images. IEEE Access, 9, 85672-85689.

Greco, A., Saggese, A., Vento, M., & Vigilante, V. (2020). A convolutional neural network for gender recognition optimizing the accuracy/speed tradeoff. IEEE Access, 8, 130771-130781.

Hassan, K. R., & Ali, I. H. (2020, November). Age and gender classification using multiple convolutional neural network. In IOP Conference Series: Materials Science and Engineering (Vol. 928, No. 3, p. 032039). IOP Publishing.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kurnianggoro, L., & Jo, K. H. (2017). Identification of pedestrian attributes using deep network. In IECON 2017-43rd Annual Conference of the IEEE Industrial Electronics Society (pp. 8503-8507). IEEE.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep learning." Nature, 521(7553), 436-444.

Levi, G., & Hassner, T. (2015). Age and gender classification using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 34-42).

Liu, Q., & Chen, L. (2020). "Advancements in age and gender classification using convolutional neural networks." IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(6), 789-802.

Mamatkulovich, B. B., & Alijon o'g'li, H. A. (2023). Facial Image-Based Gender and Age Estimation. Eurasian Scientific Herald, 18, 47-50.

Nada, A. A., Alajrami, E., Al-Saqqa, A. A., & Abu-Naser, S. S. (2020). Age and gender prediction and validation through single user images using CNN. Int. J. Acad. Eng. Res.(IJAER), 4, 21-24.

Ng, C. B., Tay, Y. H., & Goi, B. M. (2013). A convolutional neural network for pedestrian gender recognition. In Advances in Neural Networks–ISNN 2013: 10th International Symposium on Neural Networks, Dalian, China, July 4-6, 2013, Proceedings, Part I 10 (pp. 558-564). Springer Berlin Heidelberg.

Ozbulak, G., Aytar, Y., & Ekenel, H. K. (2016). How transferable are CNN-based features for age and gender classification?. In 2016 International Conference of the Biometrics Special Interest Group (BIOSIG) (pp. 1-6). IEEE.

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). "Deep Face Recognition." British Machine Vision Conference (BMVC).

Patil, J. (2021). Age and gender detection using CNN. International Journal of Scientific Research in Science and Technology, 8(3), 29-33.

Pranav, K. B., & Manikandan, J. (2020). Design and evaluation of a real-time face recognition system using convolutional neural networks. Procedia Computer Science, 171, 1651-1659.

Ranjan, R., Sankaranarayanan, S., Castillo, C. D., & Chellappa, R. (2017, May). An all-in-one convolutional neural network for face analysis. In 2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017) (pp. 17-24). IEEE.

Raza, M., Zonghai, C., Rehman, S. U., Zhenhua, G., Jikai, W., & Peng, B. Part-wise pedestrian gender recognition via deep convolutional neural networks. 2nd IET International Conference on Biomedical Image and Signal Processing (ICBISP 2017).

Rothe, R., Timofte, R., & Van Gool, L. (2016). "DEX: Deep EXpectation of apparent age from a single image." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 10-15.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

Sharma, N., Sharma, R., & Jindal, N. (2022). Face-based age and gender estimation using improved convolutional neural network approach. Wireless Personal Communications, 124(4), 3035-3054.

Smith, J., & Johnson, A. (2021). "Facial attribute recognition using deep convolutional neural networks: A comprehensive review." Journal of Artificial Intelligence, 15(3), 211-225.

Srivastava, D. K., Gupta, E., Shrivastav, S., & Sharma, R. (2023). Detection of Age and Gender from Facial Images Using CNN. In Proceedings of 3rd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications: ICMISC 2022 (pp. 481-491). Singapore: Springer Nature Singapore.

Sun, Y., Chen, Y., Wang, X., & Tang, X. (2014). Deep learning face representation by joint identification-verification. Advances in neural information processing systems, 27.

Sun, Y., Wang, X., & Tang, X. (2014). Deep learning face representation from predicting 10,000 classes. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1891-1898).

Toshev, A., & Szegedy, C. (2014). Deeppose: Human pose estimation via deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1653-1660).

Yang, J., etc. (2018). "Predicting Human Age by Combining Deep Learning and High-Order Functional Networks." IEEE Transactions on Image Processing, 27(9), 4572-4583.

Zhang, K., Gao, C., Guo, L., Sun, M., Yuan, X., Han, T. X., ... & Li, B. (2017). Age group and gender estimation in the wild with deep RoR architecture. IEEE Access, 5, 22492-22503.

AUTHOR BIOGRAPHIES

Md. Khaliluzzaman received his B.Sc. in CSE in 2007 from Khulna University of Engineering & Technology (KUET), Bangladesh. He completed his post-graduation in CSE from Chittagong University of Engineering & Technology (CUET), Bangladesh, in 2017. Currently, he is pursuing his PhD in CSE from Chittagong University of Engineering & Technology (CUET), Bangladesh, and working as an Associate Professor in the Department of Computer Science and Engineering at the International Islamic University Chittagong, Chattogram, Bangladesh. His research interests include image processing, computer vision, human-computer interaction, pattern recognition, and gait analysis.

MD Jiabul Hoque a prominent faculty member under the Faculty of Science & Engineering at International Islamic University Chittagong, is a distinguished PhD research fellow at Chittagong University of Engineering and Technology (CUET). He holds an M.Sc. in CSE from the University of Greenwich, London, UK, and a B.Sc. in CSE from CUET. With a strong focus on Machine Learning and IoT, he has authored over 15 research articles in prestigious international journals and conferences. Mr. Hoque actively participates in workshops, seminars, and symposiums in these fields, contributing significantly to academic and scientific progress. His dedication and expertise are driving advancements in these domains, making him a recognized Fig in the global research community. He can be contacted at email: jia99cse@yahoo.com.