Front Page and Declaration
Acknowledgement
Abstract

Table of Contents

1. Introduction

This work focuses on extracting demographic information such as age group, gender, and race from facial images using deep convolutional neural networks with transfer learning. Extracting such demographic information from facial images is very useful in many applications. Various businesses can do online target marketing to promote their products to a specific group of people. For example, a company in the textile industry can advertise different types of clothing appropriate for particular age groups or genders. Similarly, companies that sell footwear can reach out to their target customer group such as young people if the product is designed for their use. Furthermore, this work can also be applied to security and surveillance. Here, a potential suspect can be tracked down by narrowing down the search space of the facial footage due to the classification of faces according to age group, gender, and race.

In this work, the RetinaFace algorithm (Deng et al., 2020) was used to detect the facial images from the UTKFace dataset (Zhang, Song and Hairong, 2017). RetinaFace is a deep learning framework for detecting faces accurately in the presence of varying poses, illumination, and other observed deformations in facial images. This algorithm extracted 97.7% of UTKFace data as the dataset to train deep learning models to extract demographic information from facial images.

Once the faces were detected, they were cropped to a common resolution of 100 x 100. The deep learning frameworks like VGG16 (Simonyan and Zisserman, 2015), Resnet50 (He et al., 2016), and EfficientNet (Tan and Le, 2020) trained on ImageNet (Deng et al., 2009) were used as the base networks. Then, fully connected layers were added to these base networks to build the classifiers. Furthermore, similar models trained on VGGFace (Parkhi, Vedaldi, and Zisserman, 2015) and VGGFace2 (Cao et al., 2018) datasets to detect faces were deployed retraining the latter layers on the UTKFace data to classify age groups, gender, and race from facial images. Finally, the test results of all classifications of the deep models were compared.

1.1 Scope of the Work

This work is based on UTKFace dataset (Zhang, Song and Hairong, 2017) which consists of 23k facial images labelled with age, gender, and race. Since age ranges from 1-116 years in the UTKFace data, five age groups namely 0-14, 14-25, 25-40, 40-60, and 60+ year groups are identified for age group classification. The gender of the facial images in the entire dataset is binary and therefore gender classification focuses on classifying facial images into either male or female. The dataset also contains 5 types of races: White, Black, Asian, Indian, and Others. Thus, race classification focuses on classifying facial images into the above labelled races.

1.2 Research Question

How accurately can deep CNNs with transfer learning classify age, gender, and race from facial images?

This study focuses on how transfer learning could be used to modify existing face recognition deep models to classify facial images into age group, gender and race. Moreover, it also investigates on how age feature embeddings could be further utilised to achieve improved performance in age group and race classifications.

1.3 Project Objectives

- To classify facial images into age groups, gender, and race using deep CNNs with transfer learning.
- To investigate if gender could be used as a prior in age classification for improved performance.
- To investigate if gender could be used as a prior in race classification for improved performance.

1.4 Organisation of Chapters

1. Introduction: Provides a background for the study.
2. The UTKFace Dataset: Provides an exploratory data analysis.
3. Literature Review: Provides a review of the existing research on the topic.
4. Methodology: Provides the details of the methods used to extract demographic information from facial images.
5. Results: Shows the results obtained from running all deep learning models
6. Analysis and Discussion: Provides explanations and interpretation of results
7. Conclusion: Provides the conclusion of this research study and future research direction.
8. References: Shows the references used in this study.

2. Dataset

The UTKFace dataset originally contained 24,109 images of faces with different image size, illumination, and varying face poses (Zhang, Song and Hairong, 2017). Each face was labelled with age, gender, and race. The following table shows the attribute types of each feature.

Table 1: Attributes of each feature in UTKFace dataset. In this work, age is divided into 5 groups for age group classification.

Feature	Attribute
Age	Originally contained a range of faces in 1 – 116 years Divided into the following age groups: 0 – 14: Child 14 – 25: Youth 25 – 40: Adult 40 – 60: Middle Age 60+: Old
Gender	0 – Male, 1 – Female
Race	0 – White 1 – Black, 1 – Asian, 2 – Indian, 3 - Others

Initially, Viola and Jones's (2001) face detection algorithm was used to extract faces from the UTKFace dataset. This technique used a manual feature engineering-based face detection framework to detect the faces. It missed out on detecting more faces from the UTKFace dataset as its facial images had varying poses, illumination levels, and other deformations. Therefore, a deep learning-based face detection algorithm was sought to automate accurate face detection as a vital pre-processing step.

Wang and Deng (2020), in their study, provide a survey of face detection algorithms that used deep learning models to accurately detect faces. This emphasises how deep learning-based face recognition models can easily achieve state-of-the-art performance. Therefore, a similar deep learning-based face detection algorithm called “RetinaFace” was used in this work to extract the faces and resize them to 100 x 100 fixed size (Deng et al., 2020). The table below shows the details of the UTKFace images processed by RetinaFace algorithm.

Table 2: Number of images processed by RetinaFace algorithm

RetinaFace Performance	UTKFace Images	Percentage
Face detected	23561	97.73%
Face not detected	521	2.16%
Unprocessed images	23	0.10%
Images with incomplete labels	4	0.02%
Total	24109	100%

Since this study is about extracting demographic information from facial images, a total of 23,561 facial images of UTKFace dataset processed by RetinaFace algorithm were used as the dataset for classification tasks.

2.1 Exploratory Data Analysis

Figure 1 shows a sample of faces in the dataset used in this work. Notice that these faces have not been aligned with the horizontal axis. Figure 2 shows the age group distribution in the UTKFace data used in this study.

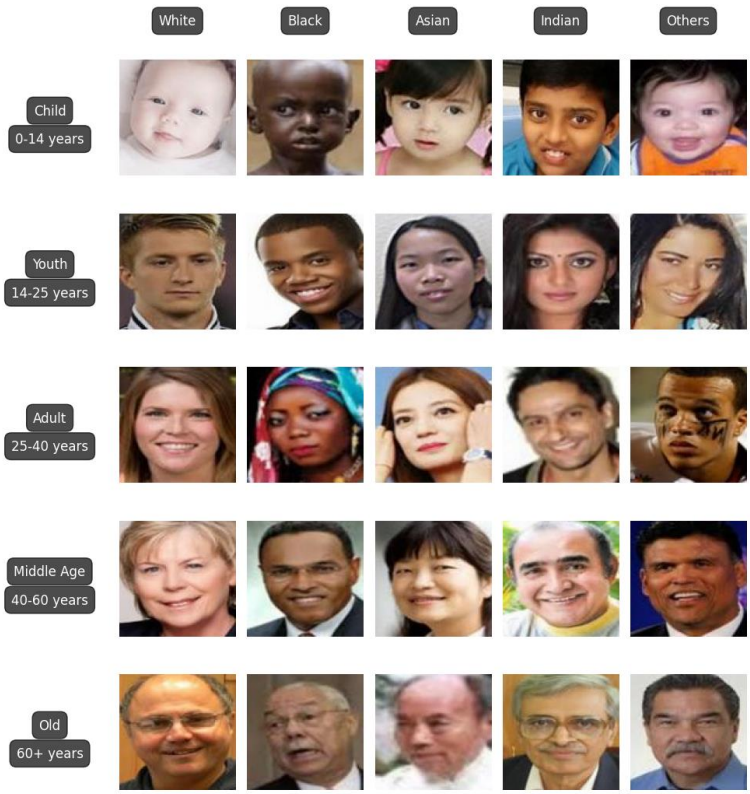

Figure 1: Sample facial images from UTKFace dataset categorised according to age groups and race. Different age groups such as Child, Youth, Adult, Middle age, and Old are shown in each row and different races such as White, Black, Asian, Indian, and Others are shown in columns.

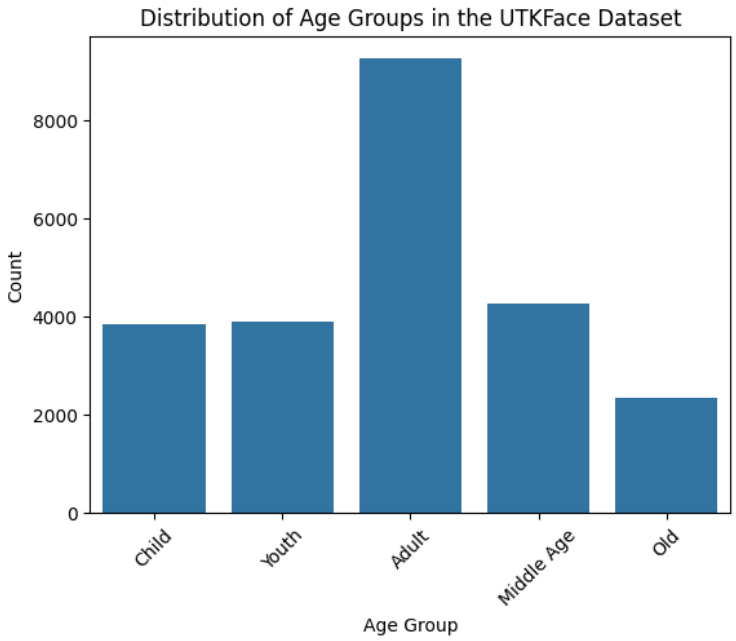

Figure 2: Age group distribution of UTKFace data used in this study. Child: 0-14 years, Youth: 14-25 years, Adult: 25-40 years, Middle Age: 40-60 years, and Old: 60+ years.

Figure 3 shows the gender distribution in the UTKFace data used in this work. This distribution is balanced between the two genders. Figure 4 shows the distribution of age groups across genders in the dataset. According to this distribution, more adult images are present in the dataset compared to other categories.

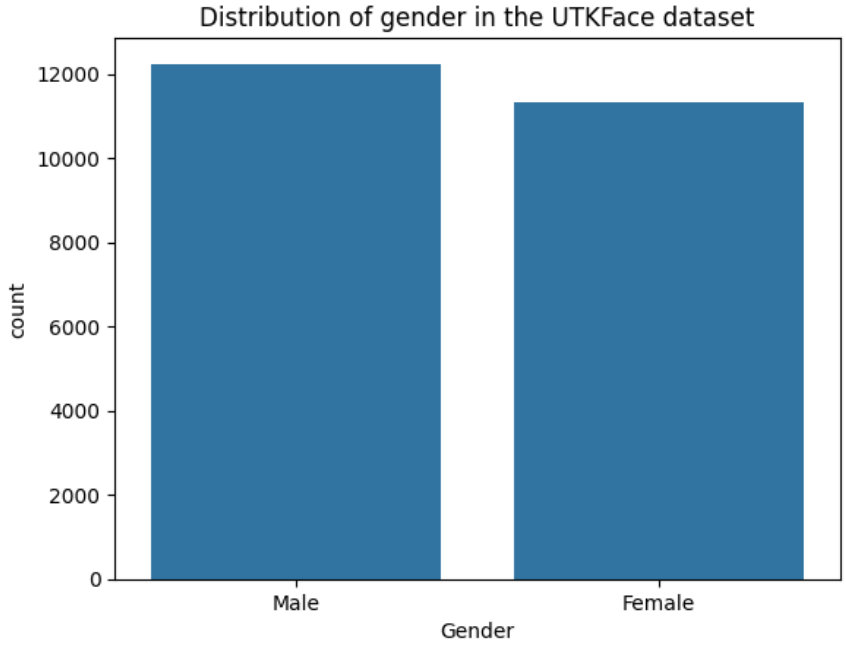

Figure 3: Gender distribution of the UTKFace dataset used in this study. UTKFace data only contains binary genders; therefore, the dataset used here has either males or females.

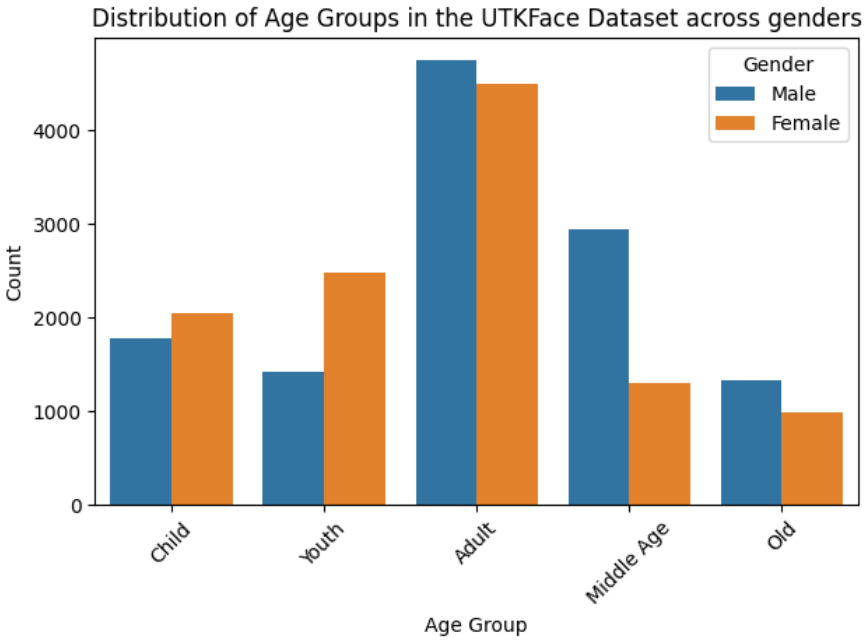

Figure 4: Age group distribution in the dataset according to gender. This shows more females in child and youth categories than males and more males in middle age and old categories than females. Child: 0-14 years, Youth: 14-25 years, Adult: 25-40 years, Middle Age: 40-60 years, and Old: 60+ years.

Figure 5 shows the race distribution in the UTKFace data used in this study. Similarly, Figure 6 shows the race distribution of the data across genders. Both these figures indicate that this dataset is White race dominant.

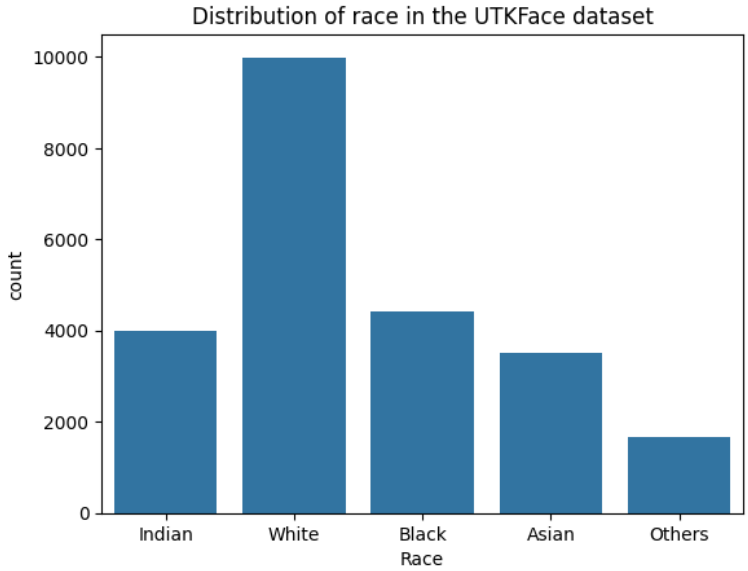

Figure 5: Race distribution of the UTKFace data used in this study. This shows that there are more white people than any other ethnic group.

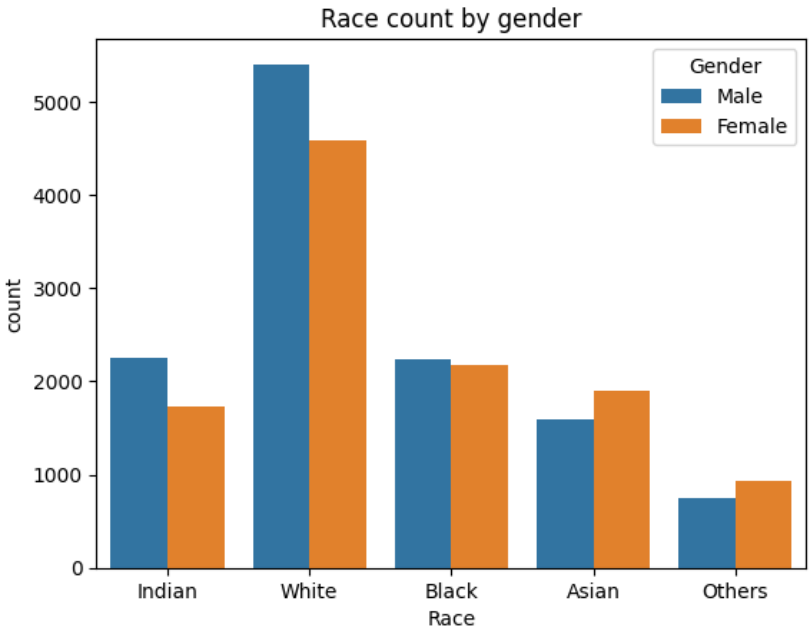

Figure 6: Race distribution of the dataset according to gender. This shows that both males and females in the White race are more than any other ethnicity.

Figure 7 shows the overall percentage summary of the gender and race of people in the UTKFace dataset used in this study. This donut chart shows that the dataset is balanced regarding gender and white-race dominant in terms of race.

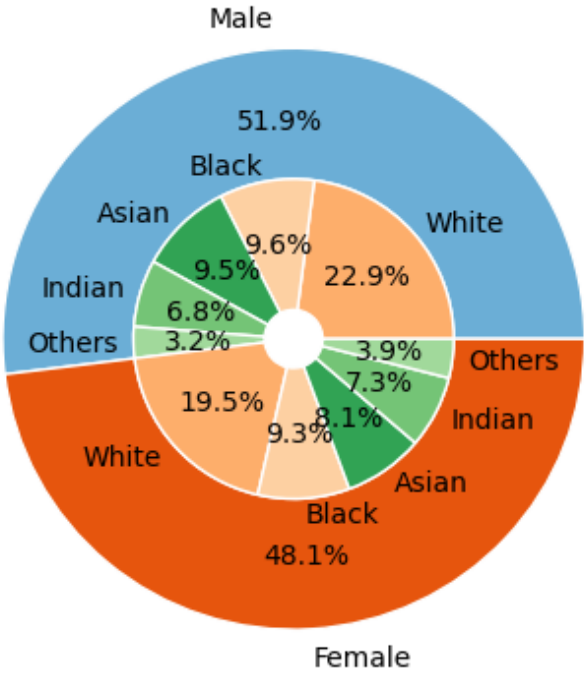

Figure 7: Donut chart on gender and race percentages in the UTKFace dataset used in this work. The outer pie chart shows the gender distribution as a percentage and the inner pie chart shows the race percentages associated with each gender.

2.2 Data Preprocessing

In this work, the RetinaFace algorithm was used to detect the faces from the UTKFace dataset. The details of the facial images used in this work are given in Table 1 in section 2.

The RetinaFace algorithm detects only the area of the faces. We as humans look at people's hair when estimating gender or race of a person. Therefore, knowing details about the hair of a person would increase the likelihood of accurate classifications of age group, gender, and race. Thus, a 20% margin of height and width of a detected face is considered when drawing the bounding box containing the face of a person for face extraction with hair.

ABC et al. (year), empirically showed that the best special resolution for faces for demographic feature classifications is 100 x 100. Therefore, the faces were then resized to a special resolution of 100 x 100.

In this work, the convolutional bases of trained deep convolutional models like VGG16, ResNet, and SeNet on very large facial datasets (2-3 million images) like VGGFace and VGGFace2 used for face detection were retrained along with the newly attached fully connected layers on UTKFace data for age group, gender, and race classifications. The preprocessing step used in the above VGG16 model is reused in this study. The mean values of R, G and B colour channels of VGGFace2 images were subtracted from UTKFace R, G, B images as a preprocessing step. Then, all the intensity values of the images are normalised to a range of 0 and 1.

Furthermore, data augmentation was also implemented and experimented. The following image augmentation techniques were deployed randomly in some experiments to increase the robustness of the trained models.

- Flip left-right
- Brightness
- Contrast
- Saturation
- Hue
- Rotation (up to 20 degrees)

2.3 Ethical concerns of the dataset

UTKFace dataset is owned by Zhifei Zhang, Yan Song, and Hairong Qi. This dataset was first used in their work on age progression at the University of Tennessee (ref). According to the author's website (Zhang, Song, and Hairong, 2017), this dataset is permitted to be used for non-commercial research purposes and thus could be used in this study.

The UTKFace dataset is anonymised and therefore no facial image could be identified with the actual name. However, a person's facial image is labelled with the actual age, gender, and race.

Moreover, this study focuses on estimating the age group of a person rather than estimating the actual age from a facial image. Thus, a person's exact age is not estimated. In addition to that, this study extracts a person's race and gender from a facial image. According to the

GDPR, it is a legal obligation to conduct a Data Protection Impact Assessment (DPIA) if there is a risk to people's rights when personal information is processed using new technologies (ref). Therefore, a DPIA is carried out and shown in section 2.3.

2.4 Data Protection Impact Assessment

Purpose of study:

To develop deep convolutional neural network models for age group, gender, and race classifications. This extraction of demographic features from facial images could be used in target marketing, automated systems for targeted services, or personalisation.

Legitimate interest:

This work is essential for research and development in Deep Learning for estimation of demographic attributes from facial images. Therefore, a publicly available facial dataset is used adhering to its usage licence.

Description of data processed:

UTKFace, a publicly available dataset with facial images of people labelled with age, gender, race, and date and time. Furthermore, this dataset does not contain any names.

Special category data: Includes sensitive attributes like age, gender, and race.

Processing steps:

- Preprocessing by detecting faces, resizing faces to 100 x 100 resolution, and normalisation.
- Use transfer learning to reuse previously trained convolutional bases of deep convolutional neural networks on face detection.
- Retrain deep convolutional nets using UTKFace data for demographic feature classifications.
- Validation and testing for accuracy and robustness.

Technology used:

- Deep learning frameworks such as Keras and Tensorflow
- Python programming language
- Google Colab for model training, validation and testing
- Python-opencv for image preprocessing
- GitHub, University drive and Google drive for storage

Minimisation:

Only attributes necessary for this study, such as age, gender, and race, are used. The date and time attributes are not used.

Data Accuracy:

- Data preprocessing techniques are used to prepare suitable input for the models.

- Data augmentation techniques are used to increase the robustness of the trained models.
- Use metrics like f1-score to evaluate model performance when data is imbalanced.

Retention:

- Only retain the data and processed models during the project duration.
- Dispose of intermediate and final datasets responsibly upon project completion.

Assessment of risks to data subjects:

- Inherent biases in the dataset may lead to unfair model predictions.
- The trained model could be deployed to extract information from faces to harm people.

Mitigation measures:

- Use the f1-score when comparing models when the dataset contains a class imbalance.
- Also use data resampling and augmentation to reduce inherence biases.
- The usage of trained models could be restricted for academic and research purposes and explicitly prohibit applying these models in general surveillance or in any intentional usage that may cause harm to people.
- Even though the faces are detected, they are not reidentified.
- Processing stages are in line with UK GDPR principles such as purpose limitation, data minimisation, and accountability.

3. Literature Review

Extracting demographic features like age group, gender and race from facial images has many applications like online target marketing, validating users according to age group for vending machines that sell age-restricted products, and reducing the search space in facial surveillance footage to track down a potential suspect with known facial demographics. Thus, it requires detecting and extracting faces before deploying methods to classify them according to these demographic features. Wang and Deng (2020), in their work, showed that deep learning-based face detection methods are more accurate and robust than hand-engineered feature-based classifiers (Viola and Jones, 2001). Therefore, RetinaFace (Deng et al., 2020) deep learning-based method is used in this work to detect faces from the UTKFace dataset (Zhang, Song and Hairong, 2017).

There are two main ways of training deep models depending on the tasks they are intended to solve. They are single task learning (STL) and multi-task learning (MTL) based models. In STL based methods, deep models are trained on the dataset per task. For example, here, the age group, gender, and race classifications are treated as three separate classification problems. Hence, a deep model is trained separately for each classification task (Dey et al., 2024; Lungin et al., 2023). In contrast, MTL based methods (Ito et al., 2018; Iqbal, Rukhsar and Baliarsingh, 2023; Foggia et al., 2023) generally have a common convolutional base and separate fully connected layers that form the classifier head per task. Figures 8 and 9 below pictorially depict the learning approaches.

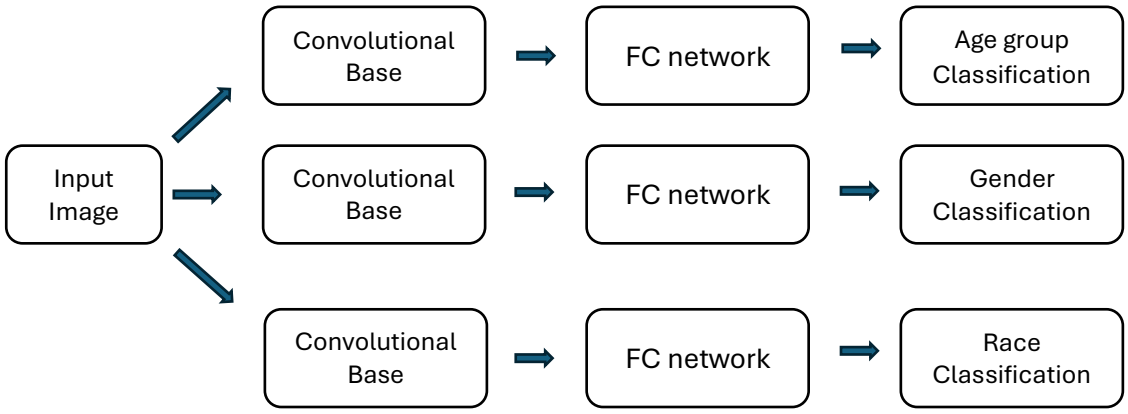

Figure 8: Single task learning model. Each classification task requires the input image to be fed to a separate convolutional base. Each convolutional base is connected to a fully connected network (FC) with two or three layers to output classifications such as age group, gender, and race. Each classification task requires a separate trained model.

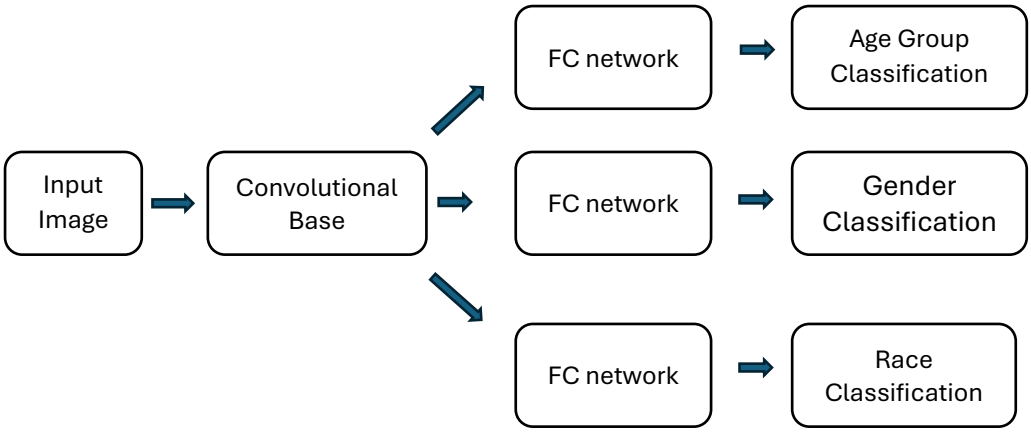

Figure 9: Multi task learning model. All classification tasks require the input image to be fed to a common convolutional base which is connected to three separate fully connected networks with two or three layers to build a classifier for each classification task such as age group, gender, and race classification. Here, each classification task does not require a separate trained model.

The main advantage of MTL based methods is they reduce the number of training parameters drastically by only training one convolutional base compared to STL based methods that train a convolutional base for each task. The loss function of MTL based methods usually consists of a weighted sum of losses defined for each classification task. These weights can be determined according to the level of difficulty of each classification task. In contrast, STL based methods generally perform better than MTL based methods as the convolutional base is trained to learn the features of data relating to each task separately.

Vallerie et al. (2023) proposed an STL-based deep convolutional neural network for age group classification. Their model architecture included two convolutional layers followed by three fully connected layers, achieving a test accuracy of 71.4% across five age groups: child, youth, adult, middle-aged, and elderly, using the UTKFace dataset. Through experimentation, they determined that a batch size of 64 yielded the best performance. Notably, their approach did not employ transfer learning; instead, the model was trained entirely on the UTKFace dataset.

Similarly, Puja et al. (2024) demonstrated the effectiveness of pre-trained models, including VGG16, VGG19, and ResNet50, for age group and gender classification using an STL-based approach. In addition to these models, they developed a custom CNN, featuring an initial convolutional layer with 32 filters, ReLU activation, batch normalization, and max pooling, followed by subsequent layers with a similar structure but an increased number of filters. Their experiments on the UTKFace dataset showed that VGG16, VGG19, ResNet50, and the custom CNN achieved test accuracies of 86.8%, 85.8%, 84.3%, and 97.7%, respectively, for gender classification.

Transfer learning refers to reusing a deep learning model trained on one task in another related task. For example, standard deep learning models such as VGG16 (Simonyan and Zisserman, 2015), Resnet50 (He et al., 2016), and EfficientNet (Tan and Le, 2020) trained on ImageNet (ref) can be used to classify facial images according to gender. ImageNet comprises 14 million images of 1000 different objects. Thus, deep models trained on ImageNet contain

good low-level parameters which are usually retained. In transfer learning, only higher layers and the classier head are retrained on another dataset to achieve another related classification task. As an example, retraining the higher layers and classifier head on another facial dataset enables the deep model to learn high-level features of the second classification task such as gender classification.

Koichi et al (2018) in their work used an MLT-based approach to predict age and gender using CNNs. They used about 210k facial images from IMDB-WIKI dataset for training and testing their models. Their models include AlexNet, VGG-16, ResNet-152, and WideResNet-16-8 pre-trained on ImageNet and achieved 91.5%, 93.4%, 92%, and 93.6% test accuracies respectively.

Similarly, Mohammad et al (2023) in their work used an MTL-based approach to predict age, gender, and race on the UTKFace dataset. They experimented with MobileNet, ResNet-50, and SeNet-50 as the convolutional base and added a couple of fully connected layers to this base per task. Their models also used pre-trained ImageNet weights and achieved 89% and 78.7% test accuracies for gender and race classifications. Since they followed an MTL-based approach, they simultaneously trained their models to achieve high execution speed, and better memory utilisation compared to STL-based approaches.

Foggia et al (2023) also followed an MTL-based approach to gender, age, race and emotion recognition using MobileNet, ResNet and SENet models trained on ImageNet. According to their study, the method outperformed single-task CNNs by offering 2.5 to 4 times faster processing and 2 to 4 times reduced memory usage, all while maintaining comparable accuracy.

4. Methodology

In this work, standard deep learning frameworks such as VGG16 (Simonyan and Zisserman, 2015), Resnet50 (He et al., 2016), and EfficientNet (Tan and Le, 2020) trained on ImageNet are used as the base network. After that, a classifier is formed by attaching fully connected layers to this base for different classification tasks such as age group, gender, and race classifications. The attached fully connected layers and the closest layers in the base models are trained on the UTKFace data for the latter layers to learn task-specific dataset related high-level features. Finally, the test results of all classifications of the deep models are compared.

Moreover, a VGG16 model trained on VGGFace2 data for face recognition is further modified to classify age groups, gender and race on UTKFace data. Once again the classifier is formed by connecting fully connected layers for this VGG16 pre-trained convolutional base. Here, the final convolutional block in the VGG base is trained along with the fully connected layers to achieve the task-based classifications. Here single task learning and multi task learning approaches are used to train this VGG model.

VGG model that uses multi-task learning approach

This will be explained in detail form

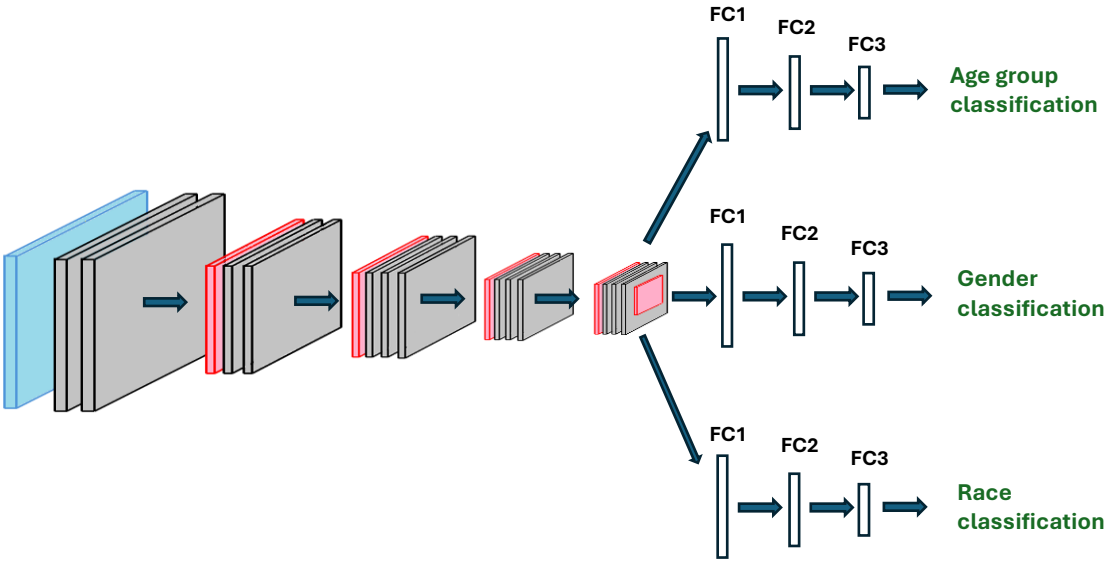

Figure 10 Multi task learning VGG model. Blue: Input image, Gray: 13 convolutional feature maps, Red: feature maps from pooling. There is a common convolutional base and separate classifiers with fully connected layers for each classification task.

VGG model that uses single-task learning approach

More details will be given here as well

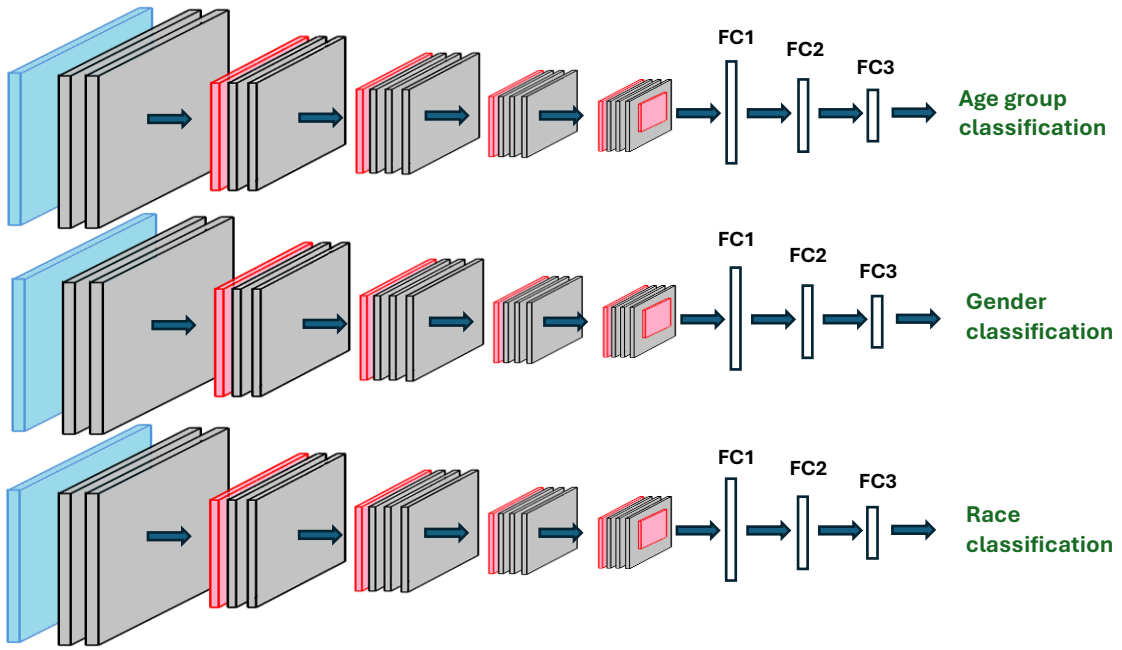

Figure 11: Single task learning based VGG model. Blue: Input image, Gray: 13 convolutional feature maps, Red: feature maps from pooling. There are separate convolutional bases and classifiers with fully connected layers for each classification task.

5. Results Analysis and Discussion

- Results of STL models

-Results of MTL models

- Their respective hyper-parameter tuning

- Results on how knowing gender affects age group and race classifications

- Model complexities STL vs. MTL and their accuracies

6. Conclusion and Future Work

7. References

Cao , Q., Shen, L., Xie, W., Parkhi, O.M. and Zisserman, A. (2018). VGGFace2: A dataset for recognising faces across pose and age. In: *Proceedings of the 13th IEEE international conference on automatic face & gesture recognition* . [online] IEEE, pp.67–74. Available at: https://ieeexplore.ieee.org/abstract/document/8373813.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In: *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [online] IEEE, pp.248–255. Available at: https://ieeexplore.ieee.org/abstract/document/5206848.

Deng, J., Guo, J., Ververas, E., Kotsia, I. and Zafeiriou, S. (2020). RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Available at: https://doi.org/10.1109/cvpr42600.2020.00525.

Dey, P., Mahmud, T., Chowdhury, M.S., Hossain, M.S. and Andersson, K. (2024). Human Age and Gender Prediction from Facial Images Using Deep Learning Methods. *Procedia Computer Science*, 238, pp.314–321. doi:https://doi.org/10.1016/j.procs.2024.06.030.

Foggia, P., Greco, A., Saggese, A. and Vento, M. (2023). Multi-task learning on the edge for effective gender, age, ethnicity and emotion recognition. *Engineering Applications of Artificial Intelligence*, 118, p.105651. doi:https://doi.org/10.1016/j.engappai.2022.105651.

He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp.770–778. Available at: https://doi.org/10.1109/cvpr.2016.90.

Iqbal, M.M., Rukhsar, A. and Baliarsingh, S.K. (2023). A CNN-based Prediction Model for Age, Gender, and Ethnicity Using Facial Images. In: *14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, pp.1–7. doi:https://doi.org/10.1109/icccnt56998.2023.10306826.

Ito, K., Kawai, H., Okano, T. and Aoki, T. (2018). Age and Gender Prediction from Face Images Using Convolutional Neural Network. In: *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. [online] IEEE, pp.7–11. doi:https://doi.org/10.23919/APSIPA.2018.8659655.

Lungin, V.J.J., Kamarudin, S.N.K., Mahmud, Y. and Shamsuddin, M.R. (2023). Age Group Classification Based on Facial Features Using Deep Learning Method. In: *4th International Conference on Artificial Intelligence and Data Sciences (AiDAS),*. [online] pp.298–302. doi:https://doi.org/10.1109/aidas60501.2023.10284587.

Parkhi, O.M., Vedaldi, A. and Zisserman, A. (2015). Deep Face Recognition. In: *BMVC - Proceedings of the British Machine Vision Conference*. [online] British Machine Vision Association, pp.1–12. Available at: https://ora.ox.ac.uk/objects/uuid:a5f2e93f-2768-45bb-8508-74747f85cad1.

Simonyan, K. and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. In: *Proceedings of the International Conference on Learning Representations (ICLR)*. ICLR. Available at: https://arxiv.org/abs/1409.1556.

Tan, M. and Le, Q. (2020). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. [online] Available at: https://arxiv.org/pdf/1905.11946.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition - CVPR.* [online] IEEE, pp.I–I. Available at: https://ieeexplore.ieee.org/document/990517.

Wang, M. and Deng, W. (2020). Deep Face Recognition: A Survey. *Neurocomputing*, 429, pp.215–244. doi:https://doi.org/10.1016/j.neucom.2020.10.081.

Zhang, Z., Song, Y. and Hairong, Q. (2017). [online] UTKFace: A Large Scale Face Dataset. Available at: https://susanqq.github.io/UTKFace/[Accessed 4 Oct. 2024].