Facial Gender Classification and Age Prediction

Loukya Malapati
Department of Computer Science
Georgia State University
Atlanta, GA 30324
lmalapati1@student.gsu.edu

Ashritha Peddi
Department of Computer Science
Georgia State University
Atlanta, GA 30324
Apeddi3@student.gsu.edu

Abstract

This project focuses on the development of a facial classification system using deep learning techniques, aiming to classify individuals by gender and predict their age based on facial attributes. Leveraging the UTKFace dataset, which contains over 20,000 annotated face images spanning a wide age range (0 to 116 years) and encompassing diverse poses, expressions, and lighting conditions, we employ deep learning models for accurate analysis. Our primary objectives include binary gender classification and age prediction using facial features such as structure, hair color, and other distinguishing characteristics. The dataset's rich metadata, including precise age, gender, and ethnicity labels, aids in training and evaluation. The applications of this research extend to various fields, including Human-Computer Interaction, content control, employment, security, surveillance, forensics, banking, Identity database management, criminal investigations, and surveillance monitoring. To achieve our goals, we explore several deep learning models, including baseline CNN, ResNet, and VGGNet assessing their effectiveness in gender and age classification. This project contributes to advancing deep learning techniques with practical applications across numerous domains.

1 Introduction

1.1 Problem Statement and Objective

Facial classification, encompassing gender identification and age prediction, plays a pivotal role in diverse applications, from security to human-computer interaction. However, current methods face challenges in achieving high accuracy and adaptability across varied datasets. Existing systems often struggle with the intricate task of deciphering facial features, especially in the context of diverse poses, expressions, and lighting conditions. Additionally, the limitations of available datasets hinder the development of robust models for accurate gender and age classification.

Recognizing the critical importance of precise facial classification in applications such as content control, employment, forensics, and identity database management, there is a clear need to overcome the limitations of current methodologies. To address these challenges, we are developing a facial classification system using deep learning techniques.

The objective is to develop deep learning models that accurately classify an individual's gender and predict their age based on facial features.

Gender Classification: Develop deep learning models that classify individuals into two categories: male and female, based on facial features.

Age Prediction: Create deep learning models capable of predicting the age of an individual by analyzing facial attributes, including age, gender, and ethnicity annotations.

1.2 Methodology

We will follow a supervised deep learning approach for this project, using the UTKFace dataset as our primary data source. The methodology involves the following steps:

Dataset Description:

UTKFace dataset is a large-scale face dataset with long age span (range from 0 to 116 years old). It has a total of 23702 images. The images cover large variations in facial expression, illumination, pose, resolution and occlusion.

We chose this dataset because of its relatively more uniform distributions, the diversity it has in image characteristics such as brightness, occlusion and position and also it involves images of the public data. Each image is labeled with a 3-element tuple, with age (in years), gender (Male-0, Female-1) and races (White-0, Black-1, Asian-2, Indian-3 and Others-4) respectively.

Table 1. Composition of sets by gender

Gender	Training	Validation	Test	Total
Male	9900	1255	1234	12389
Female	9061	1115	1137	11313
Total	18961	2370	2371	23702

Table 2. Composition of sets by age

Age Group	Training	Validation	Test	Total
0-10	2481	303	278	3062
11-20	1222	150	158	1530
21-30	5826	765	753	7344
31-40	3618	462	456	4536
41-50	1767	223	254	2244
51-60	1858	214	226	2298
61-70	1057	137	122	1316
71-80	577	57	65	699
81-90	413	45	46	504
91-100	114	11	12	137
101-116	28	3	1	32
Total	18961	2370	2371	23702

Data Preparation and Exploration:

In this pivotal phase of our project, we undertook a comprehensive exploratory data analysis (EDA) on the dataset to lay the foundation for our facial classification system. This involved meticulous organization of the dataset into a structured Data Frame, encompassing image paths, age labels, and gender labels. To enhance interpretability, we strategically mapped gender labels, assigning '0' to 'Male' and '1' to 'Female.' A deeper understanding of the dataset's characteristics was gained through the visualization of distribution curves for age and gender.

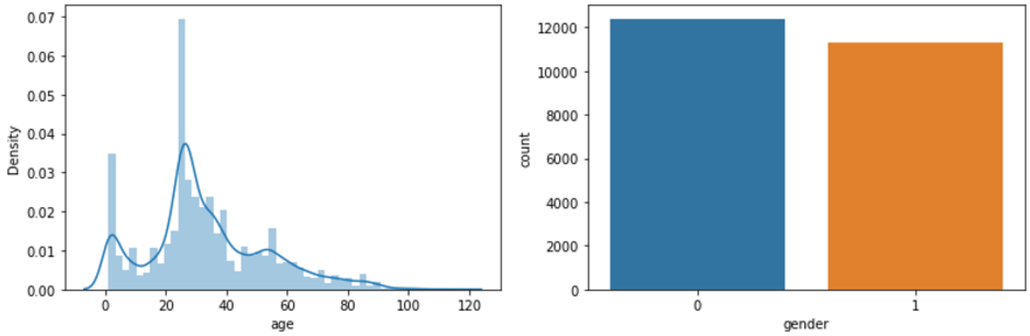

Figure 1: Age and Gender Distributions

Subsequently, we seamlessly transitioned into the data preparation stage, converting color images to grayscale for computational efficiency and resizing all images to a uniform dimension, essential for optimal model training. The processed images were then stored in a memory-efficient NumPy array to align with neural network requirements. Finally, we applied data normalization to standardize pixel values, ensuring consistent convergence during network training. This combined exploratory analysis and meticulous data preparation set the stage for the development of a robust facial classification model.

2 Deep Learning Algorithms

2.1 Convolutional Neural Network (CNN)

CNN is the most popular deep learning algorithm, used tremendously in computer vision. It is computationally efficient and instinctively spots prominent features without any supervision. CNN has the same layers compared with the traditional neural network, but the hidden layer consists of different kinds of internal layers, specifically the convolutional layer, pooling layer, fully connected layer, and normalization layer.

Model Architecture:

- The Convolutional Neural Network (CNN) architecture comprises several key components. It begins with an input layer of shape (64, 64, 1), representing the dimensions of the resized grayscale images.
- The model incorporates three convolutional layers (Conv2D) with increasing filter sizes (32, 64, 128) and uses rectified linear unit (ReLU) activation functions for feature extraction. MaxPooling2D layers follow each convolutional layer to reduce spatial dimensions.
- A Dropout layer with a dropout rate of 0.25 helps prevents overfitting. The flattened output is then connected to two branches: one for age prediction and another for gender prediction.
- Both branches consist of fully connected (Dense) layers with dropout to enhance generalization.
- The age prediction branch concludes with a single output neuron using ReLU activation, while the gender prediction branch ends with a sigmoid-activated neuron.

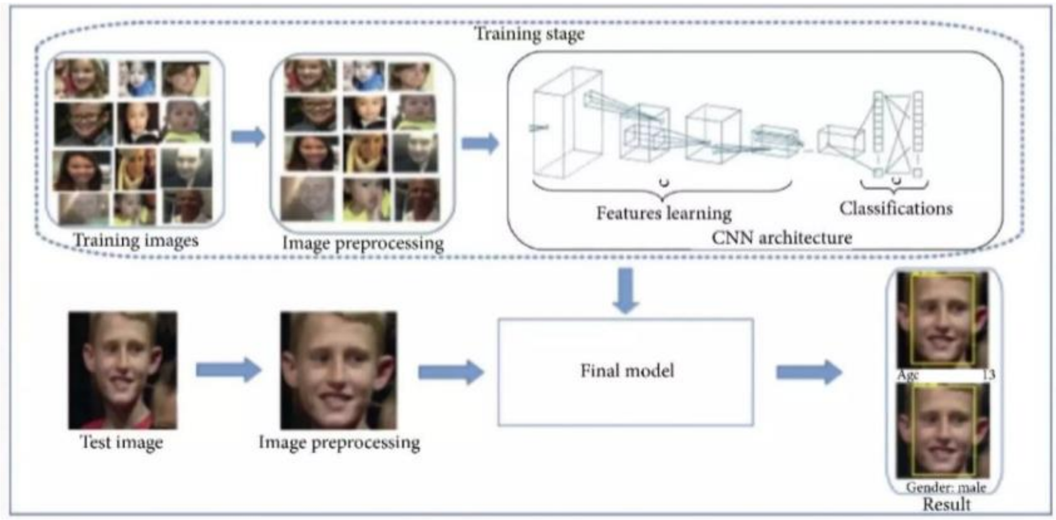

Figure 2: Architecture of CNN

2.2 ResNet50 (Residual Neural Network)

ResNet50 stands for "Residual Network with 50 layers." It is a convolutional neural network architecture that contains 50 layers, and it is a part of the ResNet family of models. ResNet was introduced to address the vanishing gradient problem in deep neural networks by introducing skip connections or residual connections. These connections allow the network to skip one or more layers, facilitating the flow of gradients during backpropagation. ResNet50 specifically refers to a variant with 50 layers, including convolutional, pooling, and fully connected layers, and it has demonstrated excellent performance in image recognition tasks.

Model Architecture:

- The ResNet50-based model for gender classification utilizes transfer learning, leveraging a pre-trained ResNet50 model on the ImageNet dataset.
- The ResNet50 layers are frozen, and additional dense layers are added for gender prediction.
- The architecture includes flattening the output, followed by dense layers with 512, 256, 128, 64, and 32 units, each activated by ReLU.
- For Gender: The final dense layer has 2 units for binary classification (Male or Female) with a sigmoid activation function.
- For Age: The final dense layer has 5 units for age classification into five age groups, with a softmax activation function.

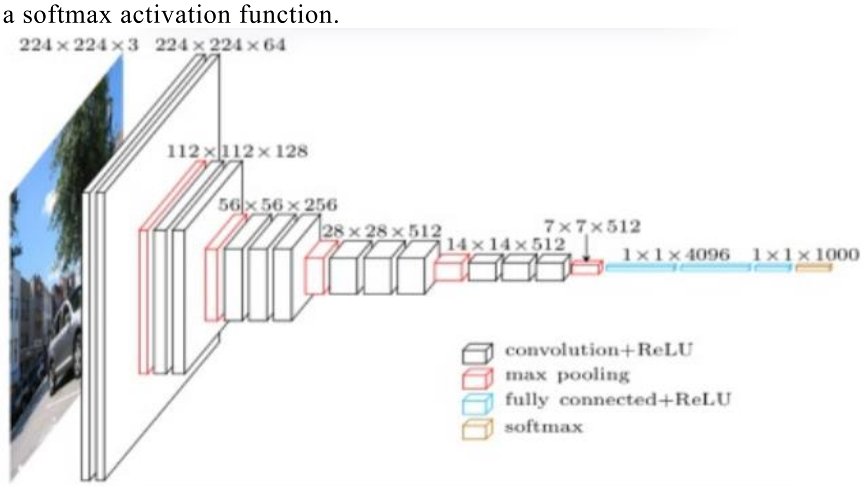

Figure 3: Architecture of ResNet

2.3 VGGNet (Visual Geometry Group)

VGG stands for Visual Geometry Group; it is a standard deep Convolutional Neural Network (CNN) architecture with multiple layers. The “deep” refers to the number of layers with VGG-16 or VGG-19 consisting of 16 and 19 convolutional layers. The VGG architecture is the basis of ground-breaking object recognition models. Developed as a deep neural network, the VGGNet also surpasses baselines on many tasks and datasets beyond ImageNet.

Model Architecture:

- The VGGNet architecture is employed for both gender and age prediction in this project.
- The gender prediction model utilizes a pre-trained VGG16 base model, consisting of 13 convolutional layers and 3 dense layers.
- The VGG16 model, with weights pretrained on ImageNet, serves as a feature extractor, followed by flattening and multiple dense layers. For gender prediction, the final dense layer has 2 units with a sigmoid activation function, facilitating binary classification (Male or Female).
- Similarly, the age prediction model incorporates the VGG16 base model, consisting of the same 13 convolutional layers and 3 dense layers.
- The architecture includes flattened layers and additional dense layers for age classification into five groups. The final dense layer for age has 5 units with a softmax activation function, allowing classification into distinct age categories.
- Both models are compiled using appropriate loss functions and optimizers tailored for their respective tasks.

3 Model Training

3.1 Convolutional Neural Network (CNN)

The model is trained using the Adam optimizer, employing mean squared error (MSE) loss for age prediction and binary crossentropy loss for gender prediction. The training process spans 25 epochs with a batch size of 128, ensuring efficient updates of model parameters. Training and validation accuracy are monitored, and the trained model is saved for further use.

```
Epoch 22/25
149/149 [==============================] - 164s 1s/step - loss: 0.3074 - dense_3_loss: 0.0193 - dense_9_loss: 0.2881 - dens
e_3_accuracy: 0.0269 - dense_9_accuracy: 0.8901 - val_loss: 0.3356 - val_dense_3_loss: 0.0190 - val_dense_9_loss: 0.3167 -
val_dense_3_accuracy: 0.0337 - val_dense_9_accuracy: 0.8838
Epoch 23/25
149/149 [==============================] - 165s 1s/step - loss: 0.3021 - dense_3_loss: 0.0192 - dense_9_loss: 0.2829 - dens
e_3_accuracy: 0.0270 - dense_9_accuracy: 0.8924 - val_loss: 0.3498 - val_dense_3_loss: 0.0182 - val_dense_9_loss: 0.3316 -
val_dense_3_accuracy: 0.0335 - val_dense_9_accuracy: 0.8832
Epoch 24/25
149/149 [==============================] - 165s 1s/step - loss: 0.2927 - dense_3_loss: 0.0187 - dense_9_loss: 0.2740 - dens
e_3_accuracy: 0.0269 - dense_9_accuracy: 0.8981 - val_loss: 0.3251 - val_dense_3_loss: 0.0192 - val_dense_9_loss: 0.3059 -
val_dense_3_accuracy: 0.0337 - val_dense_9_accuracy: 0.8859
Epoch 25/25
149/149 [==============================] - 165s 1s/step - loss: 0.2983 - dense_3_loss: 0.0184 - dense_9_loss: 0.2799 - dens
e_3_accuracy: 0.0270 - dense_9_accuracy: 0.8949 - val_loss: 0.3291 - val_dense_3_loss: 0.0192 - val_dense_9_loss: 0.3100 -
val_dense_3_accuracy: 0.0337 - val_dense_9_accuracy: 0.8661
```

Figure 4: Training Accuracy Over 25 Epochs for CNN Model

3.2 ResNet50 (Residual Neural Network)

Gender Prediction:

The final dense layer has 2 units for binary classification (Male or Female) with a sigmoid activation function. The model is compiled using binary cross-entropy loss and the RMSprop optimizer. The ResNet50-based gender classification model is trained for 2 epochs with a batch size of 64. During training, the model learns to extract gender-related features from facial images. The training history, including loss and accuracy, is recorded, and used for visualization.

Age Prediction:

The final dense layer has 5 units for age classification into five age groups, with a softmax activation function. The model is compiled using categorical cross-entropy loss and the RMSprop optimizer. The age classification model is trained for 1 epoch with a batch size of 64. The training history is recorded to visualize the training and validation loss and accuracy. The trained ResNet50 age classification model is saved for future use.

3.3 VGGNet (Visual Geometry Group)

Gender Prediction:
The VGGNet-based gender prediction model is trained over 2 epochs using a batch size of 64 on the provided training dataset. Leveraging the pre-trained VGG16 architecture with weights from ImageNet, the model is fine-tuned to discern intricate patterns indicative of gender from facial images. Throughout training, the model adjusts its parameters to minimize the binary cross-entropy loss, optimizing its ability to accurately classify images as either male or female. The RMSprop optimizer with a learning rate of 1e-3 is employed to guide the optimization process. The training history, encompassing changes in loss and accuracy over epochs, is meticulously recorded for subsequent analysis and insight into the model's learning dynamics.

Age Prediction:
Similarly, the VGGNet-based age prediction model is trained for 2 epochs, utilizing a batch size of 64 on the designated training dataset. Employing the pre-trained VGG16 architecture with weights obtained from ImageNet, the model acts as a powerful feature extractor, capturing nuanced facial features indicative of age. Throughout the training process, the model refines its parameters to minimize categorical cross-entropy loss, enhancing its proficiency in categorizing images into predefined age groups. The RMSprop optimizer with a learning rate of 1e-3 is utilized to guide the optimization process. The training history, comprising fluctuations in both loss and accuracy across epochs, is meticulously logged for comprehensive analysis.

4 Experimental Results

4.1 Convolutional Neural Network (CNN)

The training history, represented by accuracy over epochs, is visualized through histograms. The model is finally tested with a test data and got the testing accuracy of 85.95%. Additionally, the model's predictive capabilities are demonstrated by displaying selected images, comparing actual gender and age with the model's predictions. Below, the visual representation displays the predicted gender and age categories for selected facial images using a trained convolutional neural network (CNN).

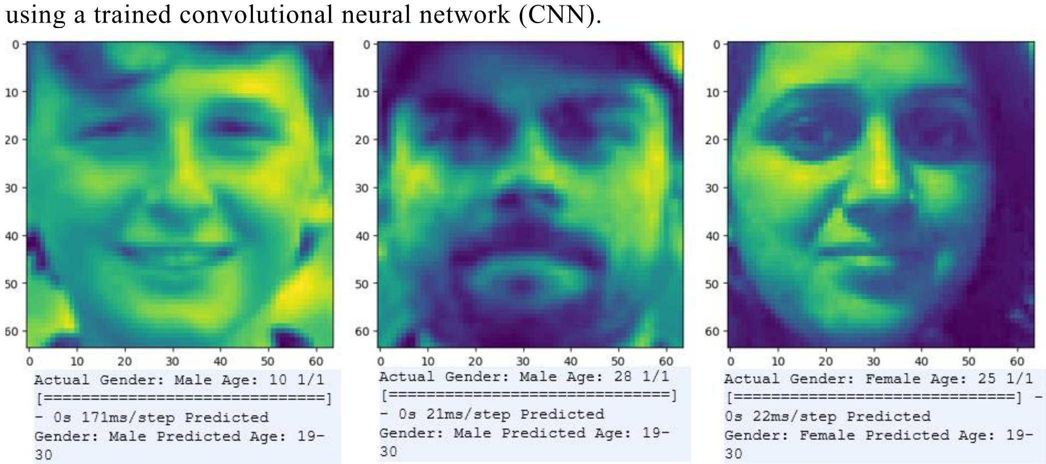

Figure 5: Predicted Gender and Age Categories for Selected Facial Images

4.2 ResNet50 (Residual Neural Network)

The trained model undergoes evaluation on the test set to assess its performance on unseen data. Key metrics such as loss and accuracy are calculated and reported. In the evaluation phase, the Age Model (ResNet) achieved a loss of 0.79 and an accuracy of 66.4% on the test set, while the Gender Model (ResNet) demonstrated a loss of 0.29 with an accuracy of 87.5%. These metrics signify the performance of the models on unseen data, indicating their effectiveness in predicting age groups and gender classifications, respectively.

4.3 VGGNet (Visual Geometry Group)

Following training, the VGG-based models are evaluated on dedicated test sets to assess their performance on unseen data. Evaluation metrics such as loss and accuracy are computed to gauge the models' effectiveness in predicting gender and age based on facial features. The age prediction model achieved an accuracy of approximately 67.2%. Simultaneously, the gender prediction model demonstrated an evaluation loss of 0.52 with an accuracy of around 77.7%. The results provide valuable insights into the generalization capabilities of the models and their potential for real-world applications.

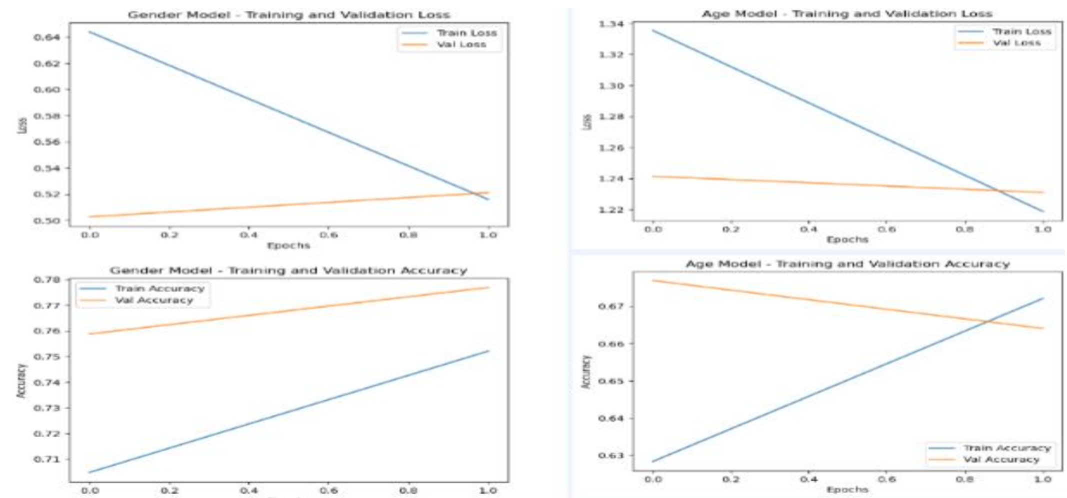

Figure 6: Training and Validation Loss/Accuracy for Age and Gender Prediction Model

5 Comparision of Models

Table 3. Comparision of Accuracy between Models

Model	Gender Prediction Test Accuracy	Age Prediction Test Accuracy
CNN	85.95%	
ResNet	87.45 %	66.41%
VGGNet	77.68%	67.23%

Table 4. Comparision between Models

Aspect	CNN	ResNet	VGGNet
Overall Architecture	Basic architecture with alternating convolutional and pooling layers	Deep architecture with skip connections (residual blocks)	Uniform architecture with 3x3 convolutional filters
Advantages	1.Simple architecture with fewer parameters. 2.Computationally less expensive. 3. Easier to interpret.	1. Effectively handles vanishing gradient problem. 2. Excellent performance in various tasks. 3. Captures intricate features.	1. Simple and easy to understand architecture. 2. Uniform structure simplifies interpretation.
Disadvantages	1. May struggle with capturing complex hierarchical features. 2. Limited depth might affect performance in some tasks.	1. Complex architecture can be computationally expensive. 2. May be harder to interpret. 3. Pre trained models can be large.	1. Depth can lead to computational cost. 2. Simplicity might limit representation of complex features.
Gender Classification	1. May require a larger dataset for effective feature learning. 2. Limited capacity for capturing intricate gender- related features.	1. Captures complex gender- related features effectively. 2. Pre trained model on ImageNet provides good feature extraction.	1. Simplicity is sufficient for gender classification. 2. Easier to fine-tune for specific tasks.
Age Classification	1. Limited depth may hinder learning hierarchical features relevant to age. 2. Performance highly dependent on dataset size.	1. Handles hierarchical features relevant to age well. 2. Pre trained model on ImageNet is beneficial for feature extraction.	1. Simplicity is advantageous for fine-tuning. 2. Easier to interpret and visualize features.

6 Conclusion & Future Steps

In summary, our investigation highlighted the remarkable accuracy achieved by the baseline CNN model, indicating its efficacy in simultaneous gender classification and age prediction. Notably, the ResNet50 model surpassed the CNN in gender prediction accuracy, showcasing its robust performance in discerning intricate features. However, the VGGNet model demonstrated comparatively lower accuracy levels in both gender and age prediction tasks. Additionally, we observed that age prediction accuracy lagged behind gender prediction, emphasizing the challenges posed by unpredictable variables such as plastic surgery and makeup. Despite these challenges, our project underscores the potential of deep learning in facial classification, offering valuable insights into model strengths and areas for improvement.

Incorporating insights from hyperparameter tuning, we determined a learning rate of 10^(-3) to be optimal, as alternative values yielded suboptimal accuracy. The chosen batch size of 64 demonstrated superior performance, emphasizing its pivotal role in achieving high accuracy.

Looking forward, our future steps involve the exploration of advanced models such as Inceptionv3. By delving into these sophisticated architectures, we aim to unlock additional layers of feature representation, potentially leading to enhanced accuracy in gender and age prediction.
The relentless pursuit of improved age prediction accuracy remains a focal point, with dedicated efforts planned for further model refinement. As we continue our journey, the strategic adjustment

of hyperparameters and the introduction of state-of-the-art models are anticipated to contribute significantly to the project's overall success.

7 Team Contribution

Table 5. Contribution of Team Members

Task	Contributor
Data Selection	Loukya, Ashritha
EDA and Data Preparation	Loukya
CNN, ResNet	Ashritha
VGGNet	Loukya
Result Analysis	Loukya, Ashritha
Documentation	Loukya, Ashritha

8 Acknowledgement

We would like to extend our sincere appreciation to **Jonathan Shihao Ji** for his invaluable support and guidance throughout the duration of this project.

We are also grateful to peers for their insightful discussions during the project presentation. Their expertise and input greatly contributed to the refinement and success of our deep learning models.

Lastly, we want to express our thanks to each member of our team. The collaborative effort and dedication of every team member significantly enhanced the overall project.

9 References

[1] Felix Anda, Evaluating Automated Facial Age Estimation Techniques for Digital Forensics, 2018 IEEE Security and Privacy Workshops https://doi.org/10.1109/SPW.2018.00028.

[2] M. H. Rahman, M. A. Bashar, F. H. M. Rafi, T. Rahman, and A. F. Mitul, "An automatic face detection and gender identification from color images using logistic regression," 2013 International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, 2013, pp. 1-6. https://doi.org/10.1109/ICIEV.2013.6572576

[3] S.Sudhakar, V.Vijayakumar, C.SathiyaKumar, V.Priya, LogeshRavi, V.Subramaniyaswamy, Unmanned Aerial Vehicle (UAV) based Forest Fire Detection and monitoring for reducing false alarms in forest-fires, Elsevier- Computer Communications 149 (2020) 1–16, https://doi.org/10.1016/j.comcom.2019.10.007

[4] R.Vasanthi, R.Jayavadivel, K.Prasadh, J.Vellingiri, G.A kilarasu, S.Sudhakar, P.M.Balasubramaniam, A novel user interaction middleware component system for ubiquitous soft computing environment by using the fuzzy agent computing system, Journal of Ambient Intelligence and Humanized Computing (2020), Springer, doi.org/10.1007/s12652-020-01893-4.

[5] C. Huang, Y. Chen, R. Lin, and C. -. J. Kuo, "Age/gender classification with whole-component convolutional neural networks (WC-CNN)," 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Kuala Lumpur, 2017, pp. 1282-1285. https://doi.org/10.1109/APSIPA.2017.8282221