AGE, GENDER AND ETHNICITY PREDICTION

Priyash Shah
IIITD
priyash21553@iiitd.ac.in

Aditya Arya
IIITD
aditya21510@iiitd.ac.in

Divyansh Mishra
IIITD
divyansh21420@iiitd.ac.in

Aditya Mishra
IIITD
aditya21125@iiitd.ac.in

Abstract

Age, gender, and ethnicity are essential demographic factors relevant in various fields such as marketing, healthcare, and social sciences. By developing a machine learning model that can accurately detect these attributes, we can contribute to research and practical applications in these areas. It can assist in patient diagnosis and treatment planning in the healthcare domain. By working on such a project, we can develop skills directly applicable to real-world scenarios.Developing an age, gender, and ethnicity detection model involves various aspects of machine learning, including data prepossessing, feature extraction, model training, and evaluation. By working on this project, we can enhance our skills in these areas and gain a deeper understanding of machine learning algorithms and techniques.

1. Introduction

This research goes into the development of a strong machine learning model in our drive to capture the useful demographic insights of age, gender, and ethnicity. The goal is straightforward: to create a model that can reliably and precisely recognize these crucial properties. By attaining this objective, we hope to support a variety of applications in industries like marketing, medicine, and social sciences.

This project extends beyond purely theoretical model creation. It calls for a thorough understanding of machine learning, covering essential elements like feature extraction, data preparation, model training, and evaluation. The journey we take is expected to be informative and will provide us a chance to practice applying machine learning algorithms in real-world settings. Our inspiration comes from the UTKFace dataset's potential as well as its distinctive features. This dataset enables us to undertake inclusive research that addresses a diverse demographic spectrum because it covers a wide age range and includes gender and ethnicity annotations. It tests us with complications from the actual world, such as changes in facial expression, emotions, illumination, occlusion, and resolution. The collection also promotes multi-task learning by mixing data on age, gender, and ethnicity—a feature that helps push the limits of facial analysis methods.

By utilizing the amount of information offered by the UTKFace dataset, we will explore the intricate details of creating this age, gender, and ethnicity detection model in the parts that follow. We seek to uncover the potential of this model in diverse real-world scenarios and develop machine learning applications through thorough data processing, feature engineering, model architecture design, and rigorous evaluation.

2. Motivation

For those working in the fields of computer vision and facial recognition, the UTKFace dataset is a helpful resource. The unique characteristics of this dataset and its potential applications across numerous domains serve as the inspiration for training on it. The dataset is revolutionary in the field of age diversity as it has members ranging from 0 to 116, annotations for gender and ethnicity as the annotation chances for inclusive research and one can perform well across a range of demographics. Aspects of position, facial emotions, illumination, occlusion, and resolution are all covered by UTKFace. One can use the dataset for multi-task learning by mixing information on age, gender, and ethnicity. The dataset also contributes to the advancement of state-of-the-art techniques in facial analysis.

3. Survey

One of the model proposed earlier was ”Two Staged CNN”, which predicts age and gender and also extracts facial representations suitable for face identification by using a modified MobileNet, at second stage the extracted facial representations are grouped using hierarchical agglomerative clustering, achieving 94.1% accuracy and 5.04 MAE on gender recognition. Other model used Multi-Task CNN based on joint dynamic loss weight adjustment, having 98.23% accuracy on gender classification and 70.1% accuracy on age classification. Clear that previous methods have a common shortcoming of higher MAE and low accuracy mainly for the task of age estimation.Keeping in mind the strengths and weaknesses, GRA_Net model have introduced following contributions in the previous works. Introduced Gates for Residual Attention Network used as a backbone of the architecture, handled the poor performance caused by minor changes in facial orientation by applying attention masks through various channels covering as many combinations as possible. Other work which tried to resolve the issue of the poor performance was Feature Extraction based Face Recognition, Gender and Age Classification (FEBFRGAC) algorithm. The algorithm yields good results with small training data, even with one image per person. The model involved three stages for training, basically pre-processing, feature extraction and classification.

3.1. GRA_Net

The model consists of multiple layer, each containing an attention block. Each attention block combines features from the previous layer with attention weights to produce refined feature representation. Gating mechanism dynamically controls the influence of attention on the feature at each layer(how much attention to be applied). The formula derived for the attention is:

$$O_{i,c}(X) = K_{i,c}(X) \cdot P_{i,c}(X)$$

It is trained using standard deep learning techniques, such as backpropagation and gradient descent.

$$\frac{\partial \mathbb{K}(X,\theta)}{\partial \phi}\mathbb{P}(X,\phi) = \mathbb{K}(X,\theta)\frac{\partial \mathbb{P}(X,\phi)}{\partial \phi}$$

Figure 1. A small image.

Loss achieved was 1.07 which is minimal till now comparing from the MAE of other models,metric used was MAE. The graph of Loss vs Iteration shows fluctuations, thus indicating a presence of high noise in the dataset. The classification accuracies achieved by the proposed GRA_Net model for UTKFace datasets are found to be 99.2%.

3.2. FEBFRGAC

In the model geometric features of facial images like eyes, nose, mouth etc. are located by using Canny edge operator and the face recognition is performed.

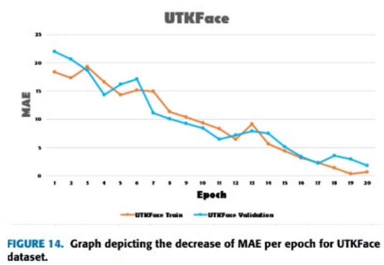

Figure 2. loss v/s iteration graph for GRA_Net

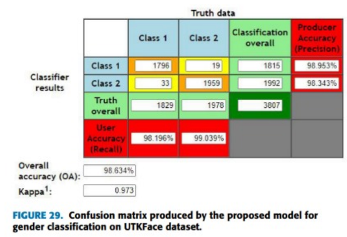

Figure 3. confusion matrix for GRA_Net

Model	Gender(%)	Age(%)
Facenet	91.2	56.9
Finetuned Facanet (FFNet)	96.1	64
MTCNN	98.23	70.1
RAN (Wang et al. (2017))	97.5	85.4
Proposed model	**99.2**	**93.7**

Figure 4. table:Comparison of both age and gender classification results with some past methods for GRA_Net

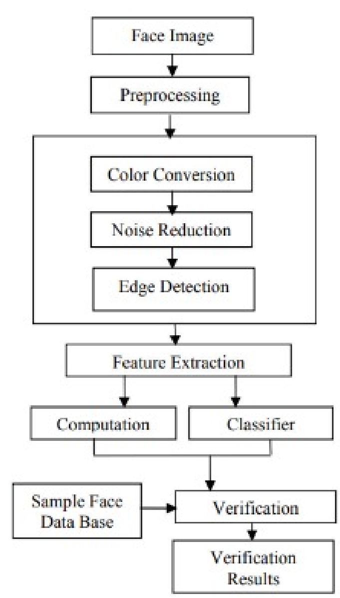

Figure 5. Structural flow of the model

In the preprocessing, first we perform color conversion in which an An RGB color image is an MxNx3 array of color pixels is a triplet corresponding to the red, green and blue components of an RGB image at a specific spatial location. Three dimensional RGB is converted into two dimensional gray scale images for easy processing of face image. After that followed by the Noise reduction, the filter for the reduction is applied to the binary image for eliminating single black pixels on white background. 8-neighbors of chosen pixels are examined if the number of black pixels are greater than white pixels then it is considered as black otherwise white. The last step in the pre processing is Edge detection, in which Canny edge detection finds edges by looking for local maxima of the gradient of f(x, y). The gradient is calculated using the derivatives of the Gaussian filter. The method uses two thresholds to detect strong and weak edges and includes the weak edges in the output only if they are connected to strong edges, i.e., to detect true weak edges.

$$G(x,y) = \sqrt{G_x + G_y}$$

where Gx and Gy are the gradient wrt x and y axis. And $(x,y) = \tan^{-1}\left(\frac{Gx}{Gy}\right)$ where (x,y) is edge direction.

Feature extraction in this context involves both global features such as interocular distance and facial proportions, and grid-based features encompassing skin color, facial regions like lips and eyes, and facial attributes like mustache and nose wing for characterizing face images. These combined features are used to represent and analyze facial information. Computation: Ratios are computed. Ratio1, ratio2, ratio3 and ratio4 are computed using the Equations (7), (8), (9) and (10) respectively.

$$Ratio1 = \frac{L\_eye\ to\ r\_eye\ distance}{eye\ to\ nose\ distance} \tag{7}$$

$$Ratio2 = \frac{L\_eye\ to\ r\_eye\ distance}{eye\ to\ lip\ distance} \tag{8}$$

$$Ratio3 = \frac{eye\ to\ nose\ distance}{eye\ to\ chin\ distance} \tag{9}$$

$$Ratio4 = \frac{eye\ to\ nose\ distance}{eye\ to\ lip\ distance} \tag{10}$$

Figure 6. Ratios to analyse facial information

For gender classification, a Naive Bayes approach is used to calculate the gender given features using the posterior probability of gender, where $P(C_i) = 0.5$, and we assume that the distribution of gender is Gaussian with mean μ_i and covariance σ_i.

$$P(C_i/x) = \frac{P(x/C_i)\,P(C_i)}{\sum_{i=female,\,male} P(x/C_i)}$$

Figure 7. Naive Bayes Probability

Training of the Model for Age Classification:

Model Used:

Artificial Neural Network (ANN) with Back Propagation. Training Phases:

Feed-Forward Path Training:

In the initial phase, the ANN trains its feed-forward path with face image inputs using backpropagation. It adjusts internal weights and biases to minimize prediction errors. This phase generates an initial output based on the input data.

Feedback Path: Activated after the initial output from the feed-forward path. Adjusts how various sections of the input impact the final result, modulating network information processing. Learns to produce varying signals contingent on the feed-forward algorithm's output, enhancing adaptability.

Training of Feedback Path: In the second phase, the feedback path is fine-tuned using pairs of face images. This helps the network differentiate between images and adapt feedback signals, potentially varying based on input stability.

Independent Training: The feed-forward and feedback paths are trained separately. They do not influence each other during the training phase, optimizing each part for its specific role within the network.

3.2.1 PERFORMANCE ANALYSIS

Gender	Sample size	Correctly Labeled(CL)	Correct Rate(CR)	Total CR
Male	40	38	95%	94.82%
Female	18	17	94.44%	

Figure 8. Table: Gender Recognition

Subject	Algorithm	Ratio1	Ratio2	Ratio3	Ratio4
Mean	FEBFRGAC	1.4384	1.4384	0.6789	1.3773
	ACFI	1.3697	1.3697	0.5574	0.5602
Var	FEBFRGAC	0.0456	0.0225	0.0253	0.3142
	ACFI	0.0227	0.0032	0.0012	0.0072
S.D	FEBFRGAC	0.2135	0.1501	0.1591	0.5605
	ACFI	0.1507	0.0567	0.3475	0.0268

Figure 9. Table: ACFI (Age Classification using Facial Image) vs FEBFRGAC

The values of mean, variance and standard deviation using FEBFRGAC are much higher than ACFI, which gives better results for a smaller number of facial image database.

4. Dataset

UTKFace dataset is a large-scale face dataset with long age span (range from 0 to 116 years old). The dataset

Algorithm	AG	Sample size	CL	CR	Total CR
FEBFRGAC	Y	28	25	89.3%	89.65%
	M	20	18	90%	
	O	10	09	90%	
CAGBFF	Y	44	37	84.4%	78.49%
	M	32	25	78.1%	
	O	17	11	64.7%	

Figure 10. Table: CAGBFF (Classification of Age Group based on Facial Features) vs FEBFRGAC

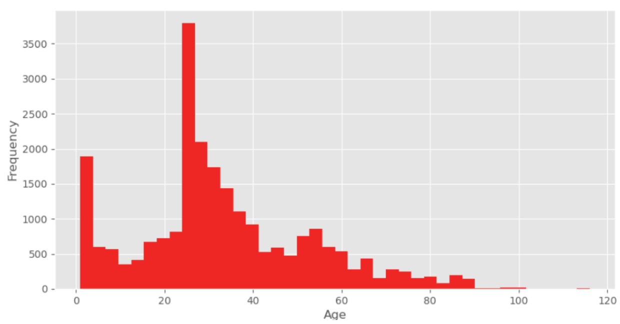

Figure 11. Age Distribution

consists of over 20,000 face images with annotations of age, gender, and ethnicity. The images cover large variation in pose, facial expression, illumination, occlusion, resolution, etc. This dataset could be used on a variety of tasks, e.g., face detection, age estimation, age progression/regression, landmark localization, etc.The UTKFace dataset consists of roughly 23k images of human faces with varying pose and illumination, making it a perfect fit for the age estimation task. Over the last several years, there has been extensive research in this area motivated by the increasing use of Deep Learning. Estimating age based on facial images alone is a difficult task, even with advanced Deep Learning methods. This is due to various external factors that influence age, such as overall health and skin care habits, as well as genetics. Additionally, the lack of high-quality labeled data has made it challenging to train deep models. However, this issue has been resolved with the availability of large labeled face datasets like VGGFace2. Moreover, the idea of Transfer Learning, which makes use of models, that were already (pre-)trained on very large datasets like ImageNet, is very appealing when working with small datasets. The dataset comprises of around roughly 24,000 images of individuals with 0-116 years of age, annotated with age, gender and ethnicity. The images show 52.3 percent males and 47.7 percent females, which means that the gender distribution is almost balanced.

The labels of each face image are embedded in the file name, formatted like [age][gender][race][date&time].jpg.

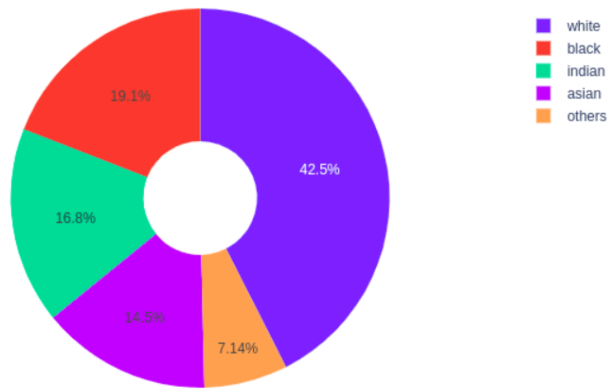

Figure 12. Race Distribution

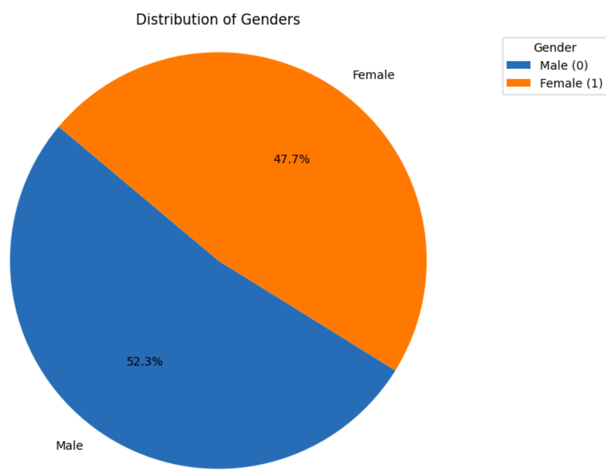

Figure 13. Age Distribution

a. Age: An integer from 0 to 116, indicating the age.

b. Gender: Either 0 (male) or 1 (female).

c. Race: An integer from 0 to 4, denoting White, Black, Asian, Indian, and Others (like Hispanic, Latino, Middle Eastern).

d. Date & Time: In the format of yyyymmddHH-MMSSFFF, showing the date and time an image was collected to UTKFace.

4.1. Prepossessing Techniques

Effective preprocessing was critical in preparing the UTK Face dataset for a variety of machine learning tasks. Because this dataset included facial photos labelled with age, gender, and race information, numerous preparation approaches were used to improve the dataset's overall quality and adaptability. One critical step was to resize all photos to a standard dimension, ensuring interoperability with multiple machine learning methods and simplifying data processing. When colour information was not required for the task, grayscale conversion was used to reduce data complexity and processing resources. Additionally, pixel values were normalised to a standard scale, frequently [0, 1], which improved model convergence during training.Encoding methods such as label encoding or one-hot encoding were used to handle categorical factors such as gender and race, making them acceptable for a wide range of machine learning methodologies. These preprocessing processes optimised the UTK Face dataset, ensuring its suitability for diverse facial recognition and classification applications across several machine learning paradigms.

5. Methodology

Following models were used to predict the outcome.

5.1. Logistic Regression

Logistic regression is a widely used statistical technique in data analysis and machine learning, particularly for binary classification tasks. It predicts the probability of an event based on independent variables, making it valuable for various applications, from medicine to marketing. It models this relationship using the logistic function, allowing us to estimate event likelihood. In research methodology, logistic regression assesses predictor impacts on outcomes, aiding data-driven decision-making.

$$P(Y=1) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p)}}$$

Where: - $P(Y=1)$ is the probability of the event occurring. - $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ are coefficients that represent the relationship between the independent variables $X_1, X_2, \ldots, X_p$ and the probability of the event. - e is the base of the natural logarithm.

5.2. K-Nearest Neighbours

To meet our research objectives, we tried the k-Nearest Neighbours (k-NN) algorithm in this work. The methodology starts with data collecting from appropriate sources, and then moves on to rigorous data preprocessing stages such as cleaning, feature selection/engineering, and data splitting. For the k-NN approach, we choose a suitable 'k' value and distance metric, and then proceed with model training and prediction. To ensure the model's robustness, performance is evaluated using appropriate metrics and cross-validation procedures. If necessary, hyperparameter adjustment is performed to improve model performance. The k-NN analysis's results and insights are provided and explored in the following sections, offering light on its application to our study subject.

$$\hat{y} = \arg\max_j \left(\sum_{i=1}^{k} I(y_i = j) \right) \quad (1)$$

$\hat{y}$ = Predicted class label
k = Number of nearest neighbors to consider
y_i = Class label of the i-th nearest neighbor
j = Class label for which we are calculating the majority vote
$I(\cdot)$ = Indicator function that equals 1 if the condition inside the parenthese

6. Results and Analysis

On further analysis of our data-set we found that few abnormalities in our data which required manual cleaning. After cleaning the data we had to process our image to ready to feed into machine learning models. Prepossessing steps has already been described on above sections, now we discuss about findings and analysis.

6.1. Data Insights

After prepossessing the data set we explored our data set further. We have 3 labels in total gender, ethnicity and age. Gender and Ethnicity are categorical while age is continuous. The data is categorized into 2 genders and 4 ethnicity while the age ranges from 0-116 years.

6.1.1 Gender Distribution

Figure 13 gives visualization for gender distribution. We can see that the percentage of male population is slightly greater than female but the difference is minor. It's not capable of creating high bias.

6.1.2 Ethnicity Breakdown

Figure 12 gives visualization for ethnicity distribution in our data set. Our data set majorly consists of images of white ethnicity with 42.5 percent. It is followed by black with 19.1 percent, Indian with 16.8 percent and Asian with 14.5 percent. Rest of the population are categorized by others.

6.1.3 Age Distribution

Figure 11 gives visualization for age distribution in our data set. From surface observation we can see that the data is skewed to the left. Thus our data set majorly consists of population less than 40 years. From figure 14 we can see that the data is also normally distributed.

6.2. Model Performance

In following section we describe about the performance of two models i.e Logistic Regression and K-Nearest Neighbours for our classification problem.

6.2.1 Logistic Regression

In this section, we present the results and analysis of our logistic regression model for classifying image data into two

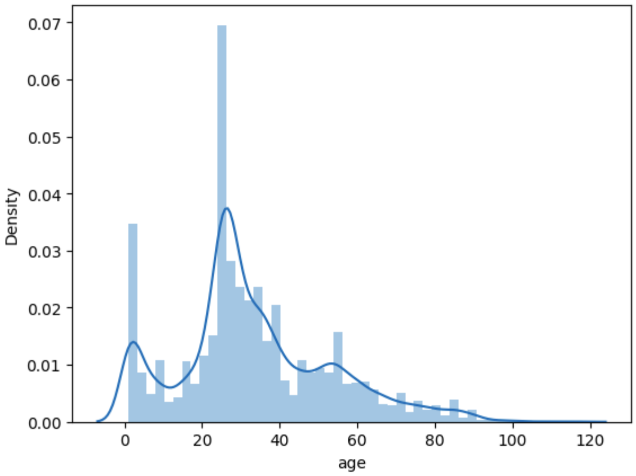

Figure 14. Age distribution plot

categories: male and female. The model was trained using a batch size of 32, binary cross-entropy as the loss function, and stochastic gradient descent (SGD) as the optimization algorithm. After 10 epochs of training, we achieved an accuracy of 80

The model's performance metrics provide valuable insights into its classification capabilities. The following statistics summarize the model's performance:

- **Training Loss:** 0.3654
- **Test Loss:** 0.3598
- **Test Accuracy:** 84.41

These results indicate that our logistic regression model performs well in classifying images into male and female categories. The relatively low training and test losses suggest that the model effectively minimized the classification error, and the test accuracy of 84.41 percent demonstrates its ability to correctly classify the gender of previously unseen images.

6.2.2 K-Nearest Neighbours

In this section, we present the results and analysis of our k-Nearest Neighbors (k-NN) model for gender classification using image data. The model was trained by flattening the image dimensions into one dimension and setting the k parameter to 20.

The performance of our k-NN model is summarized below:

- **Accuracy:** 0.7344

The classification report provides a detailed breakdown of the model's performance for each class (male and female):

Class	**Precision**	**Recall**	**F1-Score**	**Support**
0 (Male)	0.70	0.85	0.77	2468
1 (Female)	0.79	0.61	0.69	2273
Accuracy			0.73	4741
Macro Avg	0.75	0.73	0.73	4741
Weighted Avg	0.74	0.73	0.73	4741

The classification report highlights the precision, recall, and F1-score for both male and female classes, as well as the overall accuracy, macro average, and weighted average metrics.

The confusion matrix provides a visual representation of the model's performance:

$$\begin{bmatrix} 2095 & 373 \\ 886 & 1387 \end{bmatrix}$$

In the confusion matrix, the rows represent the true labels, while the columns represent the predicted labels. The diagonal elements represent correct predictions, while off-diagonal elements represent misclassifications.

Our k-NN model achieved an accuracy of 73.44 percent, indicating its ability to classify gender based on flattened image data. The classification report further reveals insights into the model's precision, recall, and F1-score for each class. Notably, the model demonstrates higher precision for females but better recall for males. The confusion matrix illustrates the distribution of correct and incorrect predictions.

7. Conclusion

In this study, we explored two distinct approaches for gender classification based on image data: logistic regression and k-Nearest Neighbors (k-NN). Each model offered unique insights into the task, shedding light on their respective strengths and limitations.

The logistic regression model achieved an accuracy of 80 percent after 10 epochs of training. Its relatively low training and test losses, along with an accuracy of 84.41 percent, demonstrated its effectiveness in classifying gender from images. However, further analysis and potential fine-tuning are recommended to explore the model's potential for enhancement.

On the other hand, the k-NN model, with 'k' set to 20, provided an accuracy of 73.44 percent. This model showed promise in gender classification, with distinct strengths in precision and recall for different gender categories. Future work may involve optimizing the choice of 'k' and exploring additional feature engineering techniques.

While both logistic regression and k-NN models have shown potential for gender classification from image data, it is important to acknowledge that there are numerous other machine learning and deep learning models that could be explored for further improvements in performance. Models like convolutional neural networks (CNNs), support vector machines (SVMs), and decision trees offer additional avenues for investigation. The choice of the most suitable model should be guided by the specific requirements of the task, dataset size, and computational resources.

In conclusion, this study has laid the foundation for gender classification from image data, demonstrating the capabilities of logistic regression and k-NN models. However, the journey towards optimal performance continues, and further research into a broader range of models and techniques is essential to achieve even better results. Additionally, addressing potential biases and ensuring fairness in classification models remains a critical concern for practical applications in diverse contexts.

References

[1] AVISHEK GARAIN,BISWARUP RAY, PAWAN KUMAR SINGH, ALI AHMADIAN, NORAZAK SENU and RAM SARKAR *GRA_Net: A Deep Learning Model for Classification of Age and Gender From Facial Images*, IEEE ACCESS, IEEE, June 3, 2021.

[2] Ramesha K,K B Raja, Venugopal K R and L M Patnaik *Feature Extraction based Face Recognition, Gender and Age Classification* , International Journal of Advanced Trends in Computer Science and Engineering, IJCSE, 2010.