Insightful Identity Analysis: Detecting Age, Gender, and Ethnicity

Sarvagya Kaushik
IIITD
sarvagya21350@iiitd.ac.in

Kunal Sharma
IIITD
kunal21331@iiitd.ac.in

Vansh
IIITD
vansh21363@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. 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. 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. 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.

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.

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. GRA Net model 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.

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. 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.

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

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

Figure 1. Attention formula in GRA_Net.

classification accuracies achieved by the proposed GRA Net model for UTKFace datasets are found to be 99.2%.

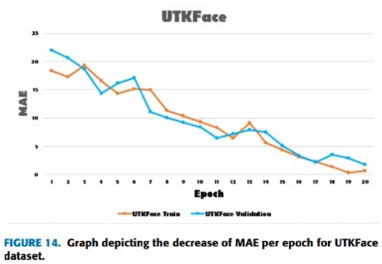
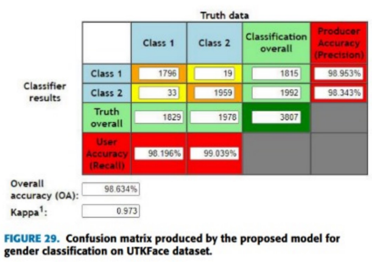


Figure 2. Loss vs Iteration graph for GRA_Net.



Truth data

Classifier results	Class 1	Class 2	Classification overall	Producer Accuracy (Precision)
Class 1	1796	19	1815	98.953%
Class 2	33	1959	1992	98.343%
Truth overall	1829	1978	3807	
User Accuracy (Recall)	98.196%	99.039%		

Overall accuracy (OA): 98.634%

Kappa[1]: 0.973

FIGURE 29. Confusion matrix produced by the proposed model for gender classification on UTKFace dataset.

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. Comparison of age and gender classification results for GRA_Net.

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.

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

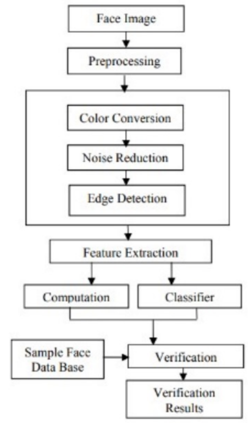


Figure 5. Structural flow of the model

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 gradients with respect to the x and y axis. And

$$(x,y) = \tan^{-1}\left(\frac{G_x}{G_y}\right)$$

where (x,y) is the 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.

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.

Training of the Model for Age Classification:

Model Used is Artificial Neural Network (ANN) with Back Propagation with these Training Phases: Feed-

$$Ratio1 = \frac{L\_eye\ to\ r\_eye\ distance}{eye\ to\ nose\ distance} \quad (7)$$

$$Ratio2 = \frac{L\_eye\ to\ r\_eye\ distance}{eye\ to\ lip\ distance} \quad (8)$$

$$Ratio3 = \frac{eye\ to\ nose\ distance}{eye\ to\ chin\ distance} \quad (9)$$

$$Ratio4 = \frac{eye\ to\ nose\ distance}{eye\ to\ lip\ distance} \quad (10)$$

Figure 6. Ratios to analyse facial information.

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

Figure 7. Naive Bayes Probability.

Forward Path Training, Feedback Path, Training of Feedback Path, Independent Training.

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

The UTKFace dataset consists of roughly 23k images of human faces (range from 0 to 116 years), annotated with age, gender and ethnicity with varying pose and illumination, making it a perfect fit for the age estimation task.. The images show 52.3 percent males and 47.7 percent females, which means that the gender distribution is almost balanced. Estimating age based on facial images alone is a difficult task. 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. The labels of each face image are embedded in the file name, formatted like [age][gender][race][date&time].jpg.

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

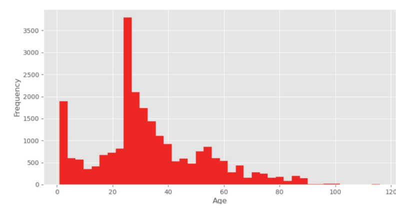


Figure 11. Age Distribution

Preprocessing performed: 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 onehot encoding were used to handle categorical factors such as gender and race, making them acceptable for a wide range of machine learning methodologies.

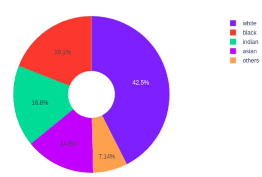


Figure 12. Race Distribution

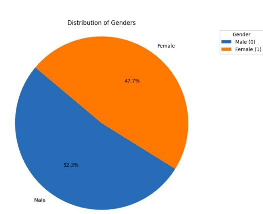


Figure 13. Age Distribution

5. Methodology

The following models were used to predict outcomes:

5.1. Logistic Regression

Logistic regression was used for binary classification tasks, predicting the probability of an event using the logistic function.

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

5.2. K-Nearest Neighbors

We also applied the k-Nearest Neighbors (k-NN) algorithm. The k value and distance metric were tuned to optimize model performance.

$$\hat{y} = \text{argmax}_j \sum_{i=1}^{k} I(y_i = j)$$

6. Results and Analysis

On further analysis of our data-set we found that few abnormalities in our data which required manual cleaning. Prepossessing steps has already been described on above sections.

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 while the age ranges from 0-116 years. Figure 13 shows that the percentage of male population is slightly greater than female. It's not capable of creating high bias. Figure 12 shows that our dataset majorly consists of images of white ethnicity with 42.5%. It is followed by black with 19.1%, Indian with 16.8% and Asian with 14.5%. Rest of the population are categorized by others. Figure 11 shows that the data is skewed to the left. Thus our dataset majorly consists of population less than 40 years. Figure 14 shows that the data is also normally distributed.

Now we describe about the performance of two models i.e. Logistic Regression and K-Nearest Neighbours for our classification problem. To classify image data into two categories: male and female using logistic regression, 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%.

Now using K-Nearest Neighbors, the model was trained by flattening the image dimensions into one dimension and setting the k parameter to 20. We obtained an accuracy of

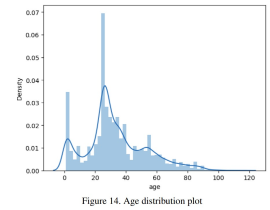


Figure 14. Age Distribution plot

0.7344 using k-NN model. 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.

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 confusion matrix visualising model's performance:

	Predicted: Male	Predicted: Female
Actual: Male	2095	373
Actual: Female	886	1385

The model demonstrates higher precision for females but better recall for males.

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. 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, models like CNNs, SVMs, and decision trees offer additional avenues for investigation to improve performance. 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.

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.