# 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.*

*The project code and implementation details are available on GitHub for further reference.*

## 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 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.

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 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.

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 $\mu_i$ and covariance $\sigma_i$.

# 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.

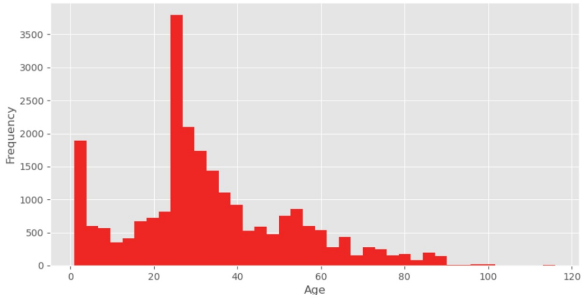

Figure 1. Age Distribution

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.

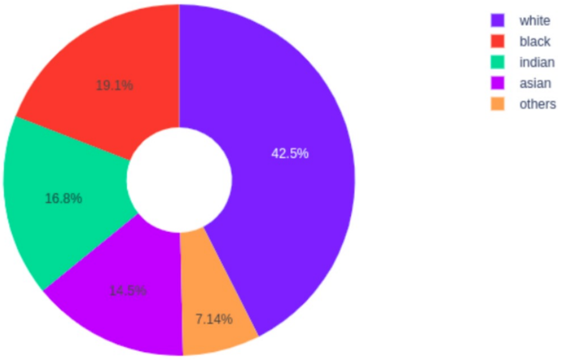

Figure 2. Race Distribution

# 5. Methodology

The following models were used to predict the outcome.

## 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. Naive Bayes Classification

Naive Bayes is a probabilistic algorithm for classification tasks, leveraging Bayes' theorem and assuming feature independence. This assumption simplifies computations and makes it suitable for high-dimensional data.

$$P(Y|X_1, X_2, \ldots, X_n) = \frac{P(Y)\cdot P(X_1|Y)\cdot \ldots \cdot P(X_n|Y)}{P(X_1)\cdot \ldots \cdot P(X_n)}$$

Here, $X$ are the features, and $Y$ is the label.

### 5.3. Random Forest Classification

Random Forest is an ensemble learning algorithm for classification and regression. It builds multiple decision trees during training and outputs the class mode (classification) or mean prediction (regression), enhancing robustness and generalizability.

Feature importance is assessed using information gain or Gini impurity. The entropy $(H(X))$ of a random variable $X$ is computed as:

$$H(X) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i)$$

Here,

- $n$ is the total number of possible outcomes of $X$.
- $P(x_i)$ is the probability of the $i$th outcome.

### 5.4. 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.

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

### 5.5. Support Vector Machines (SVM)

Support Vector Machines (SVM) are models used for classification and regression, finding a hyperplane that separates data into classes while maximizing the margin. Support vectors, the data points closest to the hyperplane, determine the decision boundary. The kernel trick enables SVM to handle high-dimensional data effectively.

Given training data:

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\},$$

where $x_i$ is the feature vector and $y_i$ is the class label, SVM seeks a hyperplane:

$$w \cdot x + b = 0,$$

such that:

$$y_i(w \cdot x_i + b) \geq 1, \quad \text{for } i = 1, 2, \ldots, n.$$

Here, $w$ is the weight vector, $x$ is the input feature vector, and $b$ is the bias term.

### 5.6. Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN) are deep learning models designed for structured grid data, such as images. They detect spatial hierarchies of features using convolutional layers, with pooling layers reducing spatial dimensions while preserving essential features. Fully connected layers and non-linear activation functions like ReLU enable complex pattern recognition. CNNs are widely used in tasks such as image classification, object detection, and segmentation.

## 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. Preprocessing steps has already been described on above sections, now we discuss about findings and analysis.

### 6.1. Data Insights

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. The percentage of male population is slightly greater than female. It's not capable of creating high bias. 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 1 shows that the data is skewed to the left. Thus our dataset majorly consists of population less than 40 years.

### 6.2. Model Performance

In the following section we describe about the performance of all the models for our classification problem.

#### 6.2.1 Logistic Regression

We trained a logistic regression model to classify images into male and female categories using a batch size of 32, binary cross-entropy as the loss function, and stochastic gradient descent (SGD). After 10 epochs, the model achieved 84.41% accuracy. Key metrics:

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

### 6.2.2 k-Nearest Neighbours (k-NN)

We used k-NN for gender and ethnicity classification, setting $k = 20$ and flattening image dimensions.

Performance:

- **Accuracy on Gender:** 0.7344
- **Accuracy on Ethnicity:** 0.56

The confusion matrix:

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

### 6.2.3 Support Vector Machine (SVM)

We trained an SVM model using the RBF kernel for gender, ethnicity, and age prediction.

Performance:

- **Accuracy on Gender:** 0.83
- **Accuracy on Ethnicity:** 0.69

The confusion matrix for gender prediction:

$$\begin{bmatrix} 714 & 118 \\ 122 & 646 \end{bmatrix}$$

### 6.2.4 Random Forest

Random Forest was used for gender, ethnicity, and age predictions with 100 estimators and Gini impurity.

Performance:

- **Accuracy on Gender:** 0.80
- **Accuracy on Ethnicity:** 0.61

The confusion matrix for gender prediction:

$$\begin{bmatrix} 684 & 148 \\ 172 & 596 \end{bmatrix}$$

### 6.2.5 Convolution Neural Networks (CNN)

We used a Tiny VGGNet CNN for gender, age, and ethnicity classification without flattening or grayscale conversion. Preprocessing included resizing to $32 \times 32$, normalizing RGB channels, and using batch size 32.

Performance:

- **Accuracy on Gender:** 0.87
- **Accuracy on Ethnicity:** 0.72
- **Accuracy on Age:** 0.47

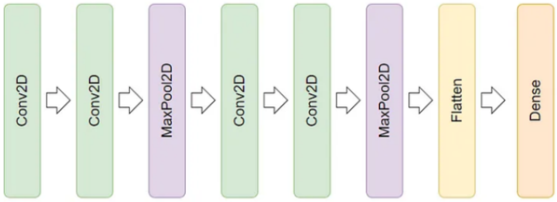

Figure 3. CNN model architecture

Given the complexity of our architecture, it converged in 6 iterations only. This wasn't possible with other models since they weren't as complex as CNN architecture.

### 6.2.6 Naive Bayes

Gaussian Naive Bayes was used for gender, age, and ethnicity classification after grayscale conversion, data flattening, and PCA. Age was digitized into 10 groups.

Performance:

- **Accuracy on Gender:** 0.79
- **Accuracy on Ethnicity:** 0.56
- **Accuracy on Age:** 0.37

## 7. Conclusion

We used 6 models to predict age, ethnicity, and gender for a given image data set. Many models performed quite well for binary classification such as gender classification but failed miserably for multi-class classification such as gender and regression problem such as age classification. Deep learning architecture like CNN architecture was able to learn the data very quickly and also performed quite well on all the prediction labels. It gave the highest accuracy on all the columns and excelled by flying colors in multi-class classification such as ethnicity classification. SVM performed quite well for both gender and ethnicity classification but was computationally expensive. It took more compute power. Random forest gave comparable result but wasn't as impressive as SVM. Both performed poorly on age prediction. The case was similar to KNN and logistic regression as well. We used several custom architectures but Tiny VGG net architecture gave the best result among all other architectures. The computation speed and complexity of CNN based architecture was also very less and was able to converge in only 5 iterations whereas other models required 20 iterations to converge given batch size and optimizer function.

## 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.