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.

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

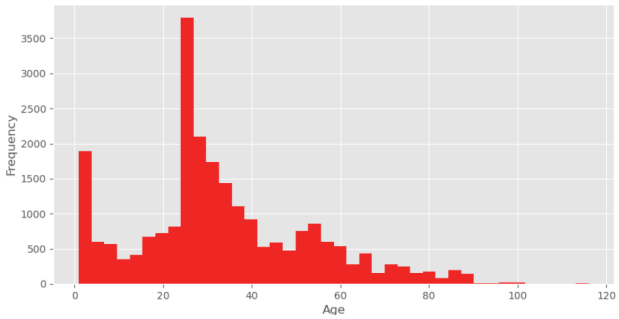

Figure 1. Age Distribution

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.

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.

4. Dataset

UTKFace dataset is a large-scale face dataset with long age span (range from 0 to 116 years old). The dataset 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.Estimating age based on facial images alone is a difficult task, even with advanced Deep Learning methods. 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.

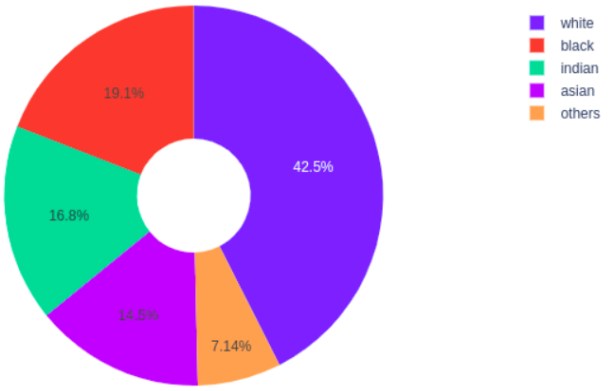

Figure 2. Race Distribution

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

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 yyyymmddHHMMSSFFF, 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. Naive Bayes Classification

Naive Bayes is a probabilistic machine learning algorithm commonly employed for classification tasks, particularly in the context of [mention the specific domain or application, e.g., text classification, spam detection, etc.]. It is based on Bayes' theorem and assumes independence between features, making it computationally efficient and well-suited for high-dimensional datasets. that the features used for classification are conditionally independent given the class label. This simplifying assumption facilitates efficient model training and inference.

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

Here Xs are the features while Y is label.

5.3. Random Forest Classification

Random Forest is a versatile ensemble learning algorithm widely employed for both classification and regression tasks. It operates by constructing a multitude of decision trees during training and outputs the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. This ensemble approach enhances the robustness and generalization of the model. Random Forest provides a measure of feature importance based on the information gain or Gini impurity reduction. This information is valuable for understanding the contribution of each feature to the model's decision-making process. The entropy ($H(X)$) of the random variable X is calculated

using the formula:

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

Here,

- $H(X)$ is the entropy of the random variable X
- n is the upper limit of the summation, representing the total number of possible outcomes of X.
- $P(x_i)$ is the probability of the ith outcome of X.

5.4. 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.

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

$\hat{y} =$ 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

5.5. Support Vector Machines (SVM)

Support Vector Machines (SVM) are powerful machine learning models used for classification and regression tasks. The SVM algorithm aims to find a hyperplane that best separates the data into different classes while maximizing the margin between the classes.

The data points that lie closest to the hyperplane and influence its position are called support vectors. These vectors play a crucial role in determining the optimal hyperplane and, hence, the SVM decision boundary. The kernel trick helps in classification of high feature data.

Given a set of 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 corresponding class label, SVM seeks to find the 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.

SVM aims to minimize $\frac{1}{2}\|w\|^2$ subject to the constraint $y_i(w \cdot x_i + b) \geq 1$.

5.6. Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN) are deep learning models designed for processing structured grid data, such as images. They are particularly effective in image classification, object detection, and image segmentation tasks. CNNs use convolutional layers to detect spatial hierarchies of features in the input data. Convolutional operations involve sliding small filters (kernels) across the input to extract local patterns. Pooling layers reduce the spatial dimensions of the input, emphasizing the most essential features while preserving spatial hierarchies. Common pooling operations include max pooling and average pooling. fter feature extraction, fully connected layers are employed for high-level reasoning. They connect every neuron in one layer to every neuron in the next layer, enabling complex relationships to be learned. Non-linear activation functions, such as ReLU (Rectified Linear Unit), introduce non-linearity to the model, allowing it to capture complex patterns and relationships in the data.

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 5 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 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 and ethnicity 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 on Gender:** 0.7344
- **Accuracy on Ethnicity:** 0.56

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:

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

6.2.3 Support Vector Machine (SVM)

In this section, we present the results and analysis of support vector machine model for gender, ethnicity and age prediction using image data. The model was trained by flattening the image dimensions and we used sklearn library to train our model. We used RBF kernel for our model. It performed better than the linear kernel.

The performance of our SVM model is summarized below:

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

The confusion matrix for gender prediction :

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

The mean absolute error on age prediction was around 110. Classic ML models failed to provide good prediction for age. Although, SVM performed quite well than other previous models.

6.2.4 Random Forest

In this section, we present the results and analysis of Random Forest model for gender, ethnicity and age prediction using image data. The model was trained by flattening the image dimensions and we used sklearn library to train our model. We normalized the features before feeding it to our model for training. Following this prepossessing helped us gain maximum accuracy possible with random forest model. We used 100 estimators for our ensemble learning process and used gini impurity as criterion.

The performance of our Random Forest model is summarized below:

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

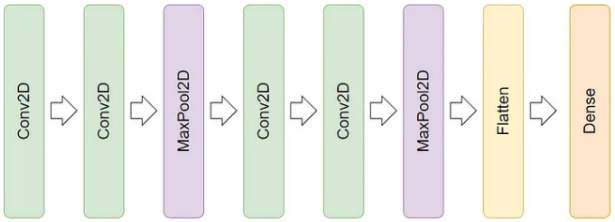

Figure 3. CNN model architecture

The confusion matrix for gender prediction :

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

Age prediction was not worth noting for this model. It performed quite well for ethnicity prediction in comparison with logistic regression and naive bayes model.

6.2.5 Convolution Neural Networks

In this section, we present the results and analysis of our Convolution Neural Network model for gender, age and ethnicity prediction using image data. Since we are using CNN architecture we weren't required to flatten the image data and we also didn't require to change image to gray scale. Keeping the input channels to 3 helped us get the best result in comparision with other models. We classified age by dividing it into 12 buckets.

The performance of our CNN model (Tiny VGGNet architecture) is summarized below:

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

We used different prepossessing step before training this architecture. Instead of resizing the model shape to 28x28 we resized it to 32x32 so that the CNN. We didn't gray scale our image and we trained our model in 3 channels i.e RGB. We also normalised there RGB values. We divided our training data into batch sizes of 32. 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

In this section, we present the results and analysis of our Naive Bayes model for gender, age and ethnicity prediction using image data. Since we are using Gaussian Naive Bayes architecture we had to change the RGB scale to Grey scale and apply the image transformation and we also flatten the image. Then we applied the PCA on the flatten image to reduce the dimension of the image.

The performance of our Naive Bayes model is summarized below:

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

We used different number of components in PCA and was able to achieve the best accuracy on 100 components. For age predictions we digitized age into 10 groups, which helped in restricting the continuous value of age to few classes.

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 same 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.