Arab Face Recognition and Identification Based on Ethnicity and Gender Using Machine Learning

Mohammed Abual-Rub
Department of Management Information Systems, King Faisal University, Al Ahsa, Saudi Arabia
mabualrub@kfu.edu.sa

Khalid Nahar
Department of Computer Science, Yarmouk University Jordan
khalids@yu.edu.jo

Ammar Almomani
Department of Information Technology, Al-Balqa Applied University, Jordan
ammarnav6@bau.edu.jo

Firas Alzobi
Department of Information System and Networks, The World Islamic Sciences and Education University, Jordan
firas.alzobi@wise.edu.jo

Abstract: *Researchers are highly interested in the classification of ethnicity using the human face since every individual has features that distinguish him from others, and every group of people shares some features that set them apart. These features are called ethnicity. A shortage of academic inquiry into the Arab world is well acknowledged. To achieve this, this research seeks to generate an Arab dataset by first grouping all Arab countries into similar categories and then classifying these labels using machine learning methods. The Arab face dataset created consists of five labels: Arab Gulf States, Egypt, Levant, Maghreb, and North and East Arab African Countries. This paper uses six types of Machine Learning to classify gender and ethnicity: Artificial Neural Network (ANN), logistic regression, Support Vector Machine (SVM), naïve bayes, K-Nearest Neighbors (KNNs), and random forest. SVM model has recorded the best result to classify gender and ethnicity with 92.7% Area Under the Curve (AUC) and 57.6% accuracy, and ANN model has recorded the best result to classify ethnicity with 92.2% AUC and 72.2% accuracy.*

Keywords: *Ethnicity, facial recognition, neural network, multi-class classification, machine learning, facial features.*

Received January 28, 2025; accepted May 5, 2025
https://doi.org/10.34028/iajit/22/4/5

1. Introduction

Arab is a term that refers to the people who speak Arabic as a first language, it is a cultural and linguistic term. The term "Arabs" does not refer to race; you can find Arabs with blue eyes and light-colored hair, while others might have dark skin [5]. Arab countries have a rich diversity of ethnicities [3]. Arabs of the Levant (Lebanese, Syrians, Jordanians, and Palestinians) are different from Arabs of the Gulf (Saudi Arabia, Emirates, Qatar...) regarding skin, eye, and hair colors, colored or brown eyes, and brown or dark hair. Gulf Arabs, on the other hand, have darker skin colors-few might have light skin, but the majority have brown skin and dark hair, and eye colors. Whereas Arabs in North African countries have dark skin. An ethnicity is a group of people with attributes distinguishing them from others. These attributes include culture, nation, religion, history, society, and language [16]. Ethnicity is, sometimes, based on inherited status, the society within which one lives, or the genetic ancestry of a person's traits. Many researchers and scientists in artificial intelligence have been interested in classifying ethnicity using the face; some classified it as Black and White, and some added Asian to the list [4, 12, 20]. Others have classified it as Chinese, Japanese, and Korean [6]. In addition, some researchers have expanded their research scope to include East Asian countries, such as Malaysia, Thailand, Vietnam, Burma, and Indonesia [10]. Other researchers, however, have classified the Arab countries into three categories: the Gulf, Levant, and Egypt [3].

Many important constraints limit the generalisability and effectiveness of present Arab ethnicity recognition techniques and datasets. First, most publicly accessible facial recognition datasets do not fairly represent Arab populations and ignore the range of ethnic groups found throughout the Arab world. Lack of diversity means that models are not adequately trained to distinguish between minute phenotypic variations unique to distinct Arab ethnic groups. Many of the ethnicity detection methods in use today were also trained on datasets mostly composed of East Asian or Western facial features, which led to prejudices and inadequate generalisation when applied on Arab individuals. Lack of a defined benchmark dataset especially intended for Arab ethnicity categorisation aggravates this issue by making rigorous assessment and comparison of model performance challenging. Moreover, many times neglected in current datasets are real-world elements necessary for building dependable and usable ethnicity recognition models: variations in illumination, occlusions (such as veils and head coverings), ageing effects, and expression diversity. Moreover, ethical and cultural aspects are generally overlooked; some databases feature images obtained without the necessary consent, which begs for issues of data privacy and fairness. This work proposes a more representative dataset and an optimal recognition technique to improve accuracy, fairness, and cross-ethnic generalisation for Arab ethnicity categorisation.

This research focuses on Arab ethnicity. As mentioned, “Arabs” refers to people who speak Arabic as a first language. The Arab World consists of 22 countries in the Middle East and North Africa. Our Arab face dataset includes all Arab countries. It sorts them into five categories: Arab Gulf States countries, Levant countries, Maghreb countries, North and East Arab African countries, and Egypt. The research seeks to enhance recognition accuracy and generalization across real-world conditions by incorporating variations in lighting, occlusions, age diversity, and expression. Through these efforts, this study aims to improve the reliability and fairness of ethnicity recognition for Arab populations. The significance of this research mainly lies in its contribution to improving the accuracy, fairness, and applicability of ethnicity recognition systems for Arab populations. Beyond technical advancements, this work has practical implications in various domains, including security, identity verification, and social applications, ensuring that AI-driven ethnicity recognition systems are more inclusive and culturally sensitive. Ultimately, this research contributes to the broader goal of reducing algorithmic bias in facial recognition and promoting ethical AI development tailored to underrepresented populations. The remaining sections of this paper discuss the following: literature review, methodology, experimental results, and conclusion.

2. Literature Review

Recent studies have utilized machine learning and deep learning techniques for ethnicity classification based on facial images. This section reviews key research conducted on this topic from 2016 to 2025.

Lakshmiprabha [15] analyzed facial images using feature extraction methods, including Active Appearance Model (AAM), Gabor wavelets, Local Binary Pattern (LBP), and Wavelet Decomposition (WD). The aim was to study gender, age, expression and ethnicity classification. AAM provided the highest accuracy of 93.83% for ethnicity recognition. The datasets used were FG-NET, Cohn-Kanade, and Productive Aging Laboratory (PAL) facial database. Chen *et al*. [6] proposed models for Chinese, Japanese and Korean ethnicity classification using K-Nearest Neighbor (KNN), Support Vector Machine (SVM), two-layer neural network, and Convolutional Neural Network (CNN). CNN achieved the best accuracy of 89.2%. Wang *et al*. [27] extracted features using Deep Convolutional Neural Networks (DCNN) to classify between White/Black, Chinese/Non-Chinese and Han/Uyghurs groups. Accuracy of 99-100% was obtained across groups. Narang and Bourlai [19] utilized deep CNN for classifying gender and Asian/Caucasian ethnicity from Nighttime Near-Infrared (NIR) facial images. An accuracy of 78.98% was reported. Gudi [9] presented a deep CNN approach to generate 3D AAMs from 2D real-world facial images for ethnicity classification into Caucasian, East Asian, South Asian and African. An overall accuracy of 92.24% was achieved.

Anwar and Islam [4] extracted features using pre-trained CNN and classified ethnicity using an SVM on 10 datasets. High average Accuracy of 98-99% was attained. Trivedi and Amali [25] compared SVM and logistic regression for classifying Chinese, White and Hispanic ethnicity images. Logistic regression performed better with smaller training data. Srinivas *et al*. [23] introduced the Wild East Asian Face Dataset (WEAFD) containing labeled images of East Asians. Two CNN models were presented for classifying gender, age and ethnicity. Gender accuracy was 88% but ethnicity accuracy was only 33% due to limited training data. Masood *et al*. [17] classified Mongolian, Caucasian and Negro ethnicity in Face Recognition Technology (FERET) dataset images using Artificial Neural Network (ANN) and pre-trained 16-layer CNN. CNN achieved 98.6% accuracy compared to 82.4% for ANN. Heng *et al*. [10] proposed a hybrid CNN-SVM approach for classifying Bangladeshi, Chinese and Indian ethnicity, achieving 95.2% accuracy. Das *et al*. [8] presented Multi-Task CNN (MTCNN) with joint dynamic weight loss for classifying gender, age and race in University of Tennessee, Knoxville Face dataset (UTKFace) and Bias Estimation in Face Analytics dataset (BEFA). ethnicity accuracy ranged from 84-90% for the datasets. Karkkainen and Joo [12] introduced the Fairface dataset containing over 100,000 labeled face images across 7 ethnicities. Their model achieved average Classification Accuracy (CA) of 94.4% across ethnicity, gender and age. Acien *et al*. [1] compared ResNet50 and Visual Geometry Group Face (VGGFace) CNN models for gender and 3-class ethnicity classification on the Labeled Faces in the Wild (LFW) dataset. Accuracy ranged from 80-94%.

Molina *et al*. [18] proposed a two-model approach using CNN and image processing for classifying gender and ethnicity. Accuracy of 95-98% was attained for gender and ethnicity classification. Darabant *et al*. [7] trained CNN models on a 175,000-image dataset to classify African, Asian, Caucasian and Indian ethnicities. Accuracy of 95-96% was achieved. Khan *et al*. [14] used a segmentation-based DCNN model for multi-dataset ethnicity classification. Up to 100% accuracy was obtained. Al-Humaidan and Prince [3] collected an Arab ethnicity dataset and compared deep-learning classification techniques. Accuracy ranged from 56-74%, highlighting the challenge.

Sunitha *et al*. [24] developed an intelligent deep learning-based method for ethnicity recognition and classification using facial images to identify and classify ethnicity based on facial features. The proposed model, called Intelligent Deep Learning- Ethnicity Recognition and Classification using Facial Images (IDL-ERCFI), utilizes an eXtreme inception (Xception) network for feature extraction and applies Principal Component Analysis (PCA) for feature reduction. Furthermore, ethnicity classification is performed using the ideal Kernel Extreme Learning Machine (KELM) approach. After it is evaluated on the Beijing University of Posts and Telecommunications (BUPT-GlobalFace) dataset which contains 1.3 million images, the proposed method achieved an accuracy of 98.97%. Kanwar and Singh [11] developed an ethnicity prediction system based on a CNN model. They trained the model using a large-scale dataset of 23709 facial images of people from different ages, genders and ethnicities. The model achieved an accuracy of 86.69%. Al-Dabbas *et al*. [2] proposed two classification models using machine learning and deep learning, using Milborrow University of Cape Town (MUCT) database for training and evaluation, the achieved accuracy, in terms of classification reached 96.01% accuracy.

In summary, machine learning approaches, particularly CNN models, have shown promising accuracy for ethnicity classification from facial images. However, there is limited focus on comprehensive multi-class Arab ethnicity classification. This research aims to address this gap using diverse Arab datasets.

3. Methodology

This research presents the first comprehensive data set dedicated to the Arab world. It presents a gender-balanced data set for the Arab faces. The data has been collected for each country separately, then it has been classified. This research implements machine learning and deep learning models to ethnicity and gender classification and evaluates models. Figure 1 illustrates the proposed model.

Figure 1 illustrates the proposed machine learning workflow model, designed using the Orange data mining tool, for image classification. The pipeline begins with two datasets: one for training and one for testing, both of which are visualized using image viewers and converted into feature vectors using image embedding widgets. The embedded training data is then fed into multiple classification models, including a neural network, logistic regression, SVM, KNN, and naïve bayes. These models are evaluated using the “test and score” widget, which tests them on the embedded test data. The performance results are visualized through a confusion matrix and can be further examined using an image viewer. This workflow provides a comprehensive setup for comparing different machine learning models on image data.

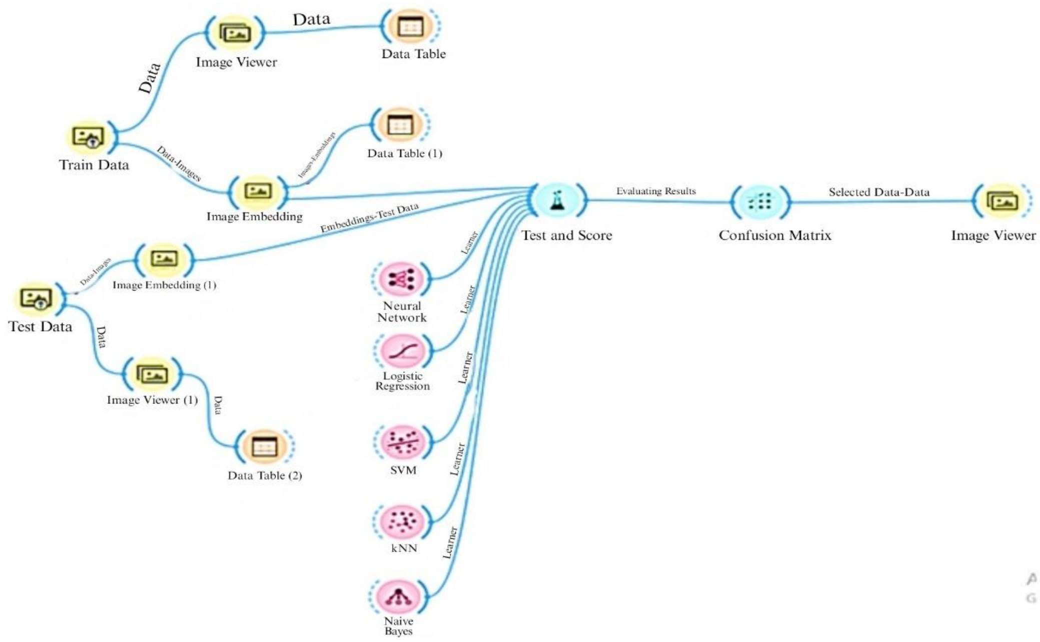

Figure 1. The proposed model.

3.1. Datasets

The proposed dataset provides face images collected from all Arab countries. The following steps have been followed in the of creating this dataset:

1. Download Google images of celebrities (actors, singers, broadcasters, athletes, businessmen /businesswomen, ministers). Most of the images collected, however, are of non-celebrities. The data collection focuses on collecting unmodified images.
2. Each of the twenty-two Arab countries has a separate dataset. Within each of these datasets, there are two main categories: males and females.
3. The data collected has been classified into five main

groups:

- Arab Gulf States: Saudi Arabia, Bahrain, Kuwait, Oman, Qatar, UAE., and Yemen.
- Levant countries: Jordan, Lebanon, Iraq, Palestine, and Syria.
- Maghreb countries: Morocco, Algeria, Libya, and Tunisia.
- North and East Arab African countries: Comoros, Djibouti, Mauritania, Somalia, and Sudan.
- Egypt.

Table 1. Dataset information.

Label	Countries in labels	No. of images for country	No. of images for label
Female/Arab Gulf States	Saudi Arabia	55	423
	Bahrain	60	
	Kuwait	60	
	Oman	53	
	Qatar	55	
	U.A.E.	70	
	Yemen	70	
Male/Arab Gulf States	Saudi Arabia	48	360
	Bahrain	56	
	Kuwait	45	
	Oman	53	
	Qatar	50	
	U.A.E.	53	
	Yemen	55	
Female/Levant	Jordan	51	264
	Lebanon	50	
	Iraq	70	
	Palestine	45	
	Syrian	48	
Male/Levant	Jordan	60	324
	Lebanon	57	
	Iraq	90	
	Palestine	53	
	Syrian	64	
Female/Maghreb	Morocco	63	256
	Algeria	60	
	Libya	82	
	Tunisia	51	
Male/Maghreb	Morocco	41	200
	Algeria	57	
	Libya	48	
	Tunisia	54	
Female/North and East Arab African countries	Comoros	45	240
	Djibouti	48	
	Mauritania	44	
	Somalia	50	
	Sudan	53	
Male/North and East Arab African countries	Comoros	48	252
	Djibouti	45	
	Mauritania	53	
	Somalia	47	
	Sudan	59	
Female/Egypt	Egypt	120	120
Male/Egypt	Egypt	180	180
Total	-	2619	2619

The grouping criteria were based on commonly referenced sociopolitical and cultural divisions in regional studies and existing literature on Arab identity classification. While we acknowledge that Iraq shares cultural and phenotypic traits with both AGS and Levant populations, it was grouped with the Levant in our framework to maintain balance across clusters and to align with classifications used in previous ethnicity recognition studies. Future work will consider dynamic or data-driven clustering strategies to better capture ethnic overlap and refine CA. Table 1 displays the details of the dataset used in this research.

An additional experimental dataset was constructed to validate model generalization. This dataset includes 550 images, comprising 275 male and 275 female samples, selected proportionally from the five ethnicity classes defined earlier. All images were manually reviewed for clarity and face visibility, resized to 224×224 pixels, and labeled for both gender and ethnicity. The dataset serves as an independent benchmark for evaluating the trained classifiers under realistic and diverse conditions.

3.2. Image Analytics

The collected images varied in quality due to differences in source platforms. To ensure consistency, all images were manually curated to exclude low-resolution, obscured, or profile-view photos. Each image was converted to a standardized size of 224×224 pixels, matching the input specifications of the InceptionV3 embedding model. The original color depth and facial details were preserved to maintain realism during classification.

Although no automated image alignment or face landmark detection was applied, manual checks ensured that most images featured front-facing, unobstructed facial views. Additionally, no image augmentation techniques (e.g., rotation, flipping, or contrast adjustment) were employed in this version to avoid introducing bias from artificially generated data. These decisions were made to preserve dataset integrity and focus on evaluating model performance using raw, real-world images. In future work, augmentation and alignment will be integrated to improve model robustness and cross-domain generalization.

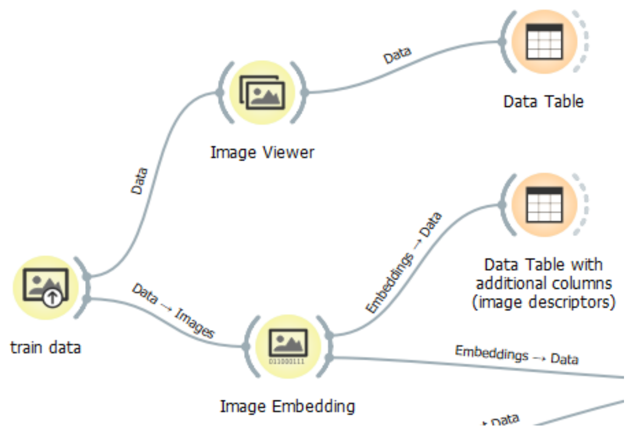

Figure 2. Image analytics for training the dataset.

As depicted in Figure 2, the following steps show the analysis of images for training the datasets:

- **Import Image**

In this step, train dataset images are imported and labelled with gender and ethnicity.

- **Image Embedding**

Image embeddings were generated using the InceptionV3 deep learning architecture, pretrained on the ImageNet dataset. Each image, after resizing to 224×224 pixels, was processed through the network up to the final global average pooling layer. This produced a 2048-dimensional feature vector for each image, encapsulating semantic and structural facial characteristics. These vectors were then used as the standardized input features for all machine learning classifiers in the study, allowing consistent model comparison based on the same representation space. The input for this neural network was a List of images while, output includes:

- **Embedding's**: images represented with a vector of numbers.
- **Skipped Images**: list of images whose embeddings were skipped during calculation.
  - Images can be read via embedding and sent to a server for analysis or evaluated locally. Each image's feature vector is determined using a deep learning algorithm. The table of data that is returned is improved by the addition of new columns (image descriptors).
  - Embedders: inception v3 Embedders is used in this model.

InceptionV3 is Google's image recognition deep neural network. It learns from the vast number of images in the ImageNet database.

- **Import Viewer:** to view the Imported dataset image.

3.3. Classification Models

This section introduces Machine Learning classification models that will be applied to classify ethnicity and gender. Six types of machine learning have been used separately as depicted in Figure 3 and explained in the upcoming subsections.

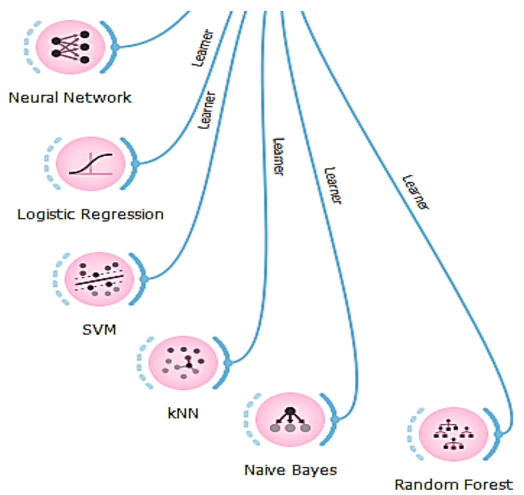

Figure 3. Classification models used.

3.3.1. Artificial Neural Network

With sklearn's multi-layer perceptron algorithm, the neural network widget can learn both linear and nonlinear models. Figure 4, is a screenshot of NN parameter setting screen in Orange tool.

The model parameters:

1. Neuron per hidden layer.
2. Activation function: Rectified Linear Unit Function (ReLu).
3. Adam (stochastic gradient-based optimizer).
4. Alpha: L2 penalty (regularization term) parameter.
5. Max iterations (maximum number of iterations).

- **Preprocessing**: when no other preprocessors are specified, neural network will employ its own default preprocessing. This is the sequence in which they are carried out:

1. It gets rid of any occurrences where the intended value is unknown.
2. Categorical variables are quantified (with one-hot-encoding).
3. Delete blank columns.
4. Mean values are used to replace missing data.

The data is normalized using the mean as a center and a standard deviation of 1 as a scaling factor.

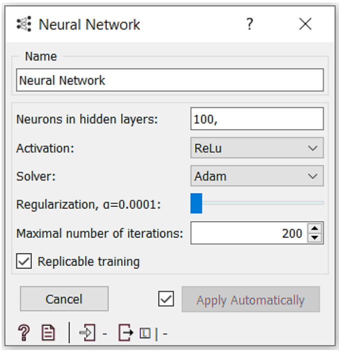

Figure 4. Neural network parameters.

3.3.2. Logistic Regression

Logistic regression classification algorithm with Least Absolute Shrinkage and Selection Operator (LASSO) (L1) or Ridge (L2) organization. Figure 5, is a screenshot of logistic regression parameter setting screen where parameters can be fine tunned.

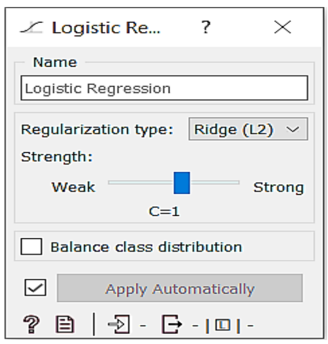

Figure 5. Logistic regression parameters.

The model parameters:

- Ridge (L2) as regularization type.

The preprocessing steps applied when using logistic regression are:

1. Delete instances with unknown target values.
2. Categorical variables are quantified (with one-hot-encoding).
3. Delete blank columns.
4. Mean values are used to replace missing data.

3.3.3. Support Vector Machine (SVM)

SVMs map inputs to higher-dimensional feature spaces. SVMs use a hyperplane to partition the attribute space to maximize the gap between instances of distinct classes or class values. The LIBrary for Support Vector Machines (LIBSVM) package, which contains a widely used implementation of SVM, is included in Orange.

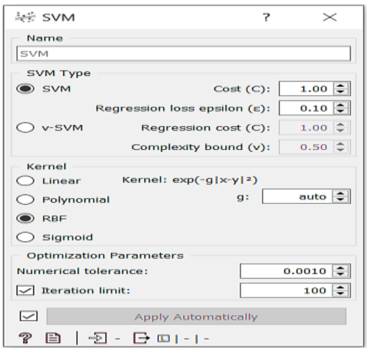

Figure 6. SVM parameters.

The model parameters as shown in Figure 6 include:

- Cost (C): a loss penalty metric relevant to classification and regression problems.
- Regression loss Epsilon (ε): a parameter to the epsilon-Support Vector Regression (epsilon-SVR) model, applies to regression tasks. Defines the distance from true values without penalty associated with predicted values.
- Kernel is a function that can generate a model with linear, polynomial, Radial Basis Function (RBF), or sigmoid kernels since each kernel is a function that maps one space of attributes to another space of features that fits the maximum-margin hyperplane. As our kernel, we employ the RBF.
- Numerical tolerance.
- Iteration limit.

When no preprocessors are specified, SVM will use its default preprocessing criteria. They are carried out in the following sequence:

- Instances with undefined destination values are purged.
- Categorical variables are quantified (with one-hot encoding).
- Deletes blank columns.
- Replaces missing data with the mean.

3.3.4. Naïve Bayes

Naïve bayes is a probabilistic classifier grounded in bayes' theorem, which assumes that features are conditionally independent given the class label. It operates by calculating prior probabilities for each class and then computing the likelihood of each feature under each class. These values are combined to determine the posterior probability of each class for a given input. The model predicts the class with the highest posterior probability. Due to its simplicity, computational efficiency, and effectiveness in high-dimensional spaces, naïve bayes is frequently used in text classification, facial recognition, and other domains with structured categorical or probabilistic data, Equation (1) represents the conditional probability of an instance x while Equation (2) is used when a vector of instances exists.

$$P(c|x) = \frac{P(x|c)P(c)}{p(x)} \tag{1}$$

Where:

$P(c|x)$ is the posterior probability
$P(x|c)$ is the likelihood probability
$P(c)$ is the class prior probability
$P(x)$ is the predictor prior probability

For X as a vector of instances, Equation (2) is applied.

$$P(c|X) = P(x1|c).P(x2|c) \dots\dots..P(xn|c).P(c) \tag{2}$$

3.3.5. K-Nearest Neighbor (KNN)

Apply the closest examples from training to your predictions. In order to make a prediction, the KNN widget employs the KNN algorithm, which takes the average of the k nearest training examples in feature space. KNN parameter setting screen is shown in Figure 7.

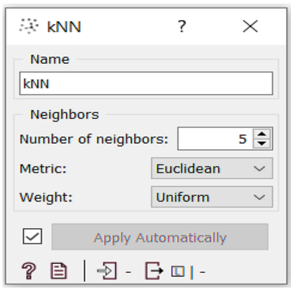

Figure 7. KNN parameters.

The model parameters:

- Number of neighbors.
- Euclidean ("straight line," distance between two points) as the distance parameter (metric) and weights as model criteria.
- Uniform (all points in each neighborhood are weighted equally) as weight.

3.3.6. Random Forest

Decision trees are created using the random forest method. A training data boot-strap sample is used to create each tree. When creating individual trees, the optimal characteristic for a split is chosen from a group of attributes chosen at random (thus the name "random"). Each tree in the forest is produced independently, and the final Model is determined by popular voting. Figure 8 displays the parameter settings of random forest method.

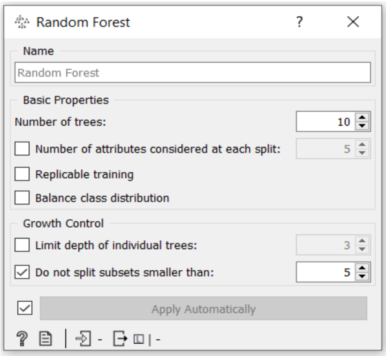

Figure 8. Random forest parameters.

The Model Parameters:

- Number of trees (10): to specify the total number of trees in the forest of potential decisions.

When no preprocessors are specified, random forest employs its own. This system carries them out in the following sequence:

- Instances whose destination values are unknown are deleted.
- Categorical variables are quantified (with one-hot encoding).
- Delete blank columns.
- Mean values are used to replace missing data. Imputes missing values with mean values.

4. Results and Discussion

Experiments have been conducted using orange data mining.

4.1. Classification Results for Gender and Ethnicity

The dataset is divided into 85% training sets and 15% testing sets without overlapping. The images used in train sets are not used in the test sets. Based on the provided evaluation metrics (AUC, CA, F1, precision, and recall), shown in Table 2, the SVM model achieved the best performance for predicting gender with an Area Under the Curve (AUC) of 0.927, CA of 0.576, and F1-score of 0.548. This indicates excellent discrimination ability and high accuracy. The neural network and logistic regression models also performed well, with AUC scores above 0.919, CA scores above 0.561, and F1-scores above 0.54. Naïve bayes and random forest models had comparatively lower metrics. KNN model performed the worst with an AUC of 0.805, CA of 0.374, and F1-score of 0.349. In summary, SVM, neural network and logistic regression were the top-performing models for gender classification in this study.

Table 2. Classification results for gender and ethnicity.

Model	AUC	CA	F1	Precision	Recall
SVM	0.927	0.576	0.548	0.581	0.576
Neural network	0.920	0.561	0.544	0.556	0.561
Logistic regression	0.919	0.583	0.559	0.567	0.583
Naïve bayes	0.882	0.486	0.450	0.455	0.486
Random forest	0.860	0.477	0.450	0.472	0.477
KNN	0.805	0.374	0.349	0.394	0.374

Figures 9 to 14 show the confusion matrices for all six models.

Predicted

Actual	Female/Arab Gulf States	Female/Egypt	Female/ Levant	Female/ Maghreb	Female/North and East Arab African Countries	Male/Arab Gulf States	Male/Egypt	Male/Levant	Male/ Maghreb	Male\North and East Arab African Countries	Σ
Female/Arab Gulf States	36	0	0	0	0	0	0	0	0	0	36
Female/Egypt	3	7	11	3	4	0	0	0	0	2	30
Female/Levant	10	4	8	10	2	0	0	2	0	0	36
Female/Maghreb	5	1	4	21	1	0	0	0	0	0	32
Female/North and East Arab African Countries	6	3	1	2	22	0	0	0	1	1	36
Male/Arab Gulf States	0	0	0	1	0	18	4	6	2	1	32
Male/Egypt	0	0	0	0	0	1	17	9	1	2	30
Male/Levant	0	0	0	1	0	2	8	17	3	0	31
Male/Maghreb	0	0	0	1	1	2	0	2	20	2	28
Male/North and East Arab African Countries	0	0	0	1	0	5	2	5	3	14	30
Σ	60	15	24	40	30	28	31	41	30	22	321

Figure 9. Confusion matrix for neural network model.

In Figures 9, the neural network shows strong performance in correctly classifying most ethnic groups. However, misclassifications are most prominent between Levant and AGS, suggesting that shared facial features (e.g., skin tone, structure) between these regions challenge the model's discriminative ability. Egypt is largely classified correctly, but a few instances are confused with African Arab countries, possibly due to geographical proximity and overlapping features.

Predicted

Actual	Female/Arab Gulf States	Female/ Egypt	Female/ Levant	Female/ Maghreb	Female/North and East Arab African Countries	Male/Arab Gulf States	Male/Egypt	Male/Levant	Male/ Maghreb	Male\North and East Arab African Countries	Σ
Female/Arab Gulf States	36	0	0	0	0	0	0	0	0	0	36
Female/Egypt	3	5	8	5	9	0	0	0	0	0	30
Female/Levant	9	2	9	8	5	1	0	1	1	0	36
Female/Maghreb	6	2	2	22	0	0	0	0	0	0	32
Female/North and East Arab African Countries	6	2	2	2	22	0	0	0	0	2	36
Male/Arab Gulf States	1	0	0	0	0	18	3	4	4	2	32
Male/Egypt	0	0	0	0	0	0	20	4	1	5	30
Male/Levant	0	0	0	1	0	3	4	19	2	2	31
Male/Maghreb	0	0	0	0	1	2	0	2	21	2	28
Male/North and East Arab African Countries	0	0	0	0	1	5	2	5	2	15	30
Σ	61	11	21	38	38	29	29	35	31	28	321

Figure 10. Confusion matrix for logistic regression model.

As can be seen from Figure 10, logistic regression shows competitive classification performance, especially in separating Maghreb and Egypt. Some confusion persists between Levant and AGS, echoing the neural network's trend. Minor overlaps are observed between Egypt and African Arab categories, likely due to facial similarity and dataset diversity.

Predicted

Actual	Female/Arab Gulf States	Female/ Egypt	Female/ Levant	Female/ Maghreb	Female/ North and East Arab African	Male/Arab Gulf States	Male/Egypt	Male/Levant	Male/ Maghreb	Male\North and East Arab African Countries	Σ
Female/Arab Gulf States	34	0	1	0	0	1	0	0	0	0	36
Female/Egypt	6	7	8	2	5	0	0	0	0	2	30
Female/Levant	12	2	5	11	2	1	0	2	1	0	36
Female/Maghreb	2	0	1	29	0	0	0	0	0	0	32
Female/North and East Arab African Countries	6	0	0	4	22	1	0	1	1	1	36
Male/Arab Gulf States	0	0	0	0	1	18	1	6	4	2	32
Male/Egypt	0	0	0	0	0	0	20	9	0	1	30
Male/Levant	0	0	0	0	0	4	3	22	0	2	31
Male/Maghreb	0	0	0	1	1	2	2	3	11	8	28
Male/North and East Arab African Countries	0	0	0	1	0	4	4	3	1	17	30
Σ	60	9	15	48	31	31	30	46	18	33	321

Figure 11. Confusion matrix for SVM model.

The SVM classifier in Figure 11, demonstrates moderate accuracy but suffers from greater confusion between AGS and Levant, and Maghreb and African Arab categories. This could be due to SVM's reliance on kernel-based separation, which may not capture subtle non-linear facial feature variations as effectively as neural networks.

Predicted

Actual	Female/Arab Gulf States	Female/ Egypt	Female/ Levant	Female/ Maghreb	Female/North and East Arab African Countries	Male/Arab Gulf States	Male/Egypt	Male/Levant	Male/ Maghreb	Male\North and East Arab African Countries	Σ
Female/Arab Gulf States	22	5	4	2	2	1	0	0	0	0	36
Female/Egypt	3	13	6	1	6	0	0	0	0	1	30
Female/Levant	8	10	3	10	3	0	2	0	0	0	36
Female/Maghreb	1	0	0	29	0	0	1	0	1	0	32
Female/North and East Arab African Countries	7	5	0	3	19	0	1	0	0	1	36
Male/Arab Gulf States	0	0	0	1	0	18	8	1	3	1	32
Male/Egypt	0	0	0	0	0	0	28	0	0	2	30
Male/Levant	0	0	1	1	0	5	10	5	8	1	31
Male/Maghreb	0	0	0	1	1	1	4	4	13	4	28
Male/North and East Arab African Countries	0	0	0	0	0	8	8	4	4	6	30
Σ	41	33	14	48	31	33	62	14	29	16	321

Figure 12. Confusion matrix for naïve bayes model.

In Figure 12, naïve bayes exhibits noticeable misclassifications across nearly all class boundaries. The model assumes feature independence, which may not hold for facial embeddings. This results in increased overlap between closely related ethnic groups, especially between Maghreb and Africa, and Levant and AGS.

KNN in Figure 13, struggles with high-dimensional facial embeddings, resulting in frequent misclassifications, particularly among Maghreb, African Arab, and Egypt classes. The curse of dimensionality reduces KNN’s ability to find meaningful nearest neighbors, leading to lower overall accuracy.

Predicted

Actual	Female/Arab Gulf States	Female/ Egypt	Female/ Levant	Female/ Maghreb	Female/ North and East Arab African	Male/Arab Gulf States	Male/Egypt	Male/Levant	Male/ Maghreb	Male\North and East Arab African Countries	Σ
Female/Arab Gulf States	32	0	3	0	0	1	0	0	0	0	36
Female/Egypt	11	6	6	2	3	0	0	0	0	2	30
Female/Levant	17	3	7	9	0	0	0	0	0	0	36
Female/Maghreb	8	1	4	19	0	0	0	0	0	0	32
Female/North and East Arab African Countries	15	4	4	2	9	0	1	1	0	0	36
Male/Arab Gulf States	1	0	1	1	0	18	4	4	3	0	32
Male/Egypt	0	0	0	1	0	2	11	13	2	1	30
Male/Levant	0	0	0	1	0	4	14	9	2	1	31
Male/Maghreb	1	0	1	0	1	4	4	9	5	3	28
Male/North and East Arab African Countries	0	0	0	0	1	11	5	6	3	4	30
Σ	58	14	26	35	14	40	39	42	15	11	321

Figure 13. Confusion matrix for KNN model.

In Figure 14, the random forest classifier provides a slightly better separation than KNN and naïve bayes, but still exhibits confusion between Levant and AGS, and Egypt vs. African Arab groups. This may reflect the model’s sensitivity to imbalanced data and the need for feature selection tuning.

Predicted

Actual	Female/Arab Gulf States	Female/Egypt	Female/ Levant	Female/ Maghreb	Female/North and East Arab African Countries	Male/Arab Gulf States	Male/Egypt	Male/Levant	Male/ Maghreb	Male\North and East Arab African Countries	Σ
Female/Arab Gulf States	36	0	0	0	0	0	0	0	0	0	36
Female/Egypt	9	2	10	1	7	0	1	0	0	0	30
Female/Levant	15	2	9	6	2	0	1	1	0	0	36
Female/Maghreb	4	1	3	21	0	1	0	2	0	0	32
Female/North and East Arab African Countries	5	0	5	3	20	1	0	0	0	2	36
Male/Arab Gulf States	0	0	0	1	0	19	3	4	1	4	32
Male/Egypt	0	0	0	0	0	0	11	13	4	2	30
Male/Levant	0	0	0	0	0	4	5	15	4	3	31
Male/Maghreb	0	0	0	1	0	3	2	6	10	6	28
Male/North and East Arab African Countries	0	0	0	0	2	7	0	8	3	10	30
Σ	69	5	27	33	31	35	23	49	22	27	321

Figure 14. Confusion matrix for random forest model.

4.2. Classification Results for Ethnicity

Eleven models have been generated for ethnicity Classification, the first Model is used to classify all labels, and the other models are used to classify between each two labels, i.e., Arab Golf States vs. Levant.

All experiments in this study were conducted using Orange data mining, a visual programming platform for data analysis and machine learning. The tool provided a flexible and transparent environment for constructing and evaluating the experimental pipeline. Through its widget-based interface, Orange enabled the integration of image embedding (via InceptionV3), model training (e.g., naïve bayes, SVM, logistic regression), and performance evaluation using standard metrics such as accuracy, AUC, F1-score, and confusion matrices. The visual workflow facilitated traceability and reproducibility of the experimental design, ensuring consistency across multiple classifiers. Furthermore, Orange’s compatibility with Python allowed advanced configuration of model parameters while preserving interpretability for comparative analysis.

We have added a dedicated subsection that defines all evaluation metrics used in the study AUC, CA, F1-score, precision, and recall along with their mathematical formulations and relevant citations. These metrics were selected to provide a comprehensive assessment of classification performance, especially in the context of multi-class and imbalanced datasets, which are common in ethnicity and gender classification tasks. Table 3 illustrates those evaluation metrics used.

Table 3. Evaluation metrics.

Metric	Equation	What It Indicates
Accuracy (CA)	(TP+TN)/(TP+TN+FP+FN)	Overall correctness of predictions
Precision	TP/(TP+FP)	Correctness of positive predictions
Recall	TP/(TP+FN)	Coverage of actual positives
F1-score	2×(Precision×Recall)/(Precision+Recall)	Balance between precision and recall
AUC	Area under the ROC curve	Ability to distinguish between classes

4.2.1. Model for Classifying All Labels

In Table 4, the neural network model achieved the highest CA with a CA score of 0.722 for predicting ethnicity labels. Logistic regression and SVM models also performed well, with CA scores of 0.632 and 0.519, respectively. The neural network had the best F1-score of 0.722, showing strong precision and recall. While random forest, naïve bayes, and KNN models had comparatively lower performance. The random forest model had the lowest metrics with AUC of 0.755, CA of 0.466, and F1 of 0.447. In summary, neural network, logistic regression and SVM models were the most effective for ethnicity classification in this study.

Table 4. Evaluation results for all labels.

Model	AUC	CA	F1	Precision	Recall
Neural network	0.922	0.722	0.722	0.722	0.722
Logistic regression	0.886	0.632	0.630	0.635	0.632
SVM	0.836	0.519	0.512	0.520	0.519
Naïve bayes	0.790	0.511	0.488	0.517	0.511
KNN	0.763	0.474	0.460	0.484	0.474
Random forest	0.755	0.466	0.447	0.457	0.466

4.2.2. Model for Classifying Arab Gulf States vs. Levant

Table 5 displays the evaluation results of the classification of Arab Gulf States vs. Levant, the neural network and logistic regression models perform the best overall, with accuracy scores of 0.827 and 0.823, respectively, and relatively high AUC, F1, precision, and recall scores. The KNN, random forest, SVM, and naïve bayes models appear to have lower overall performance than neural network and logistic regression models.

Table 5. Evaluation results for A.G.S. vs. Levant.

Model	AUC	CA	F1	Precision	Recall
Neural network	0.906	0.827	0.827	0.828	0.827
Logistic regression	0.900	0.823	0.823	0.823	0.823
KNN	0.847	0.762	0.761	0.762	0.762
Random forest	0.835	0.751	0.747	0.752	0.751
SVM	0.821	0.742	0.743	0.744	0.742
Naïve bayes	0.785	0.736	0.733	0.736	0.736

4.2.3. Model for classifying Arab Gulf States vs. Maghreb

Table 6 displays the evaluation results of classification of Arab Gulf States vs. Maghreb, it is clear that the neural network and logistic regression models again perform the best overall, with AUC scores of 0.887 and 0.872, and relatively high accuracy scores of 0.808 and 0.796, respectively. F1, precision, and recall also show high scores. The KNN and naïve bayes models also perform well, but with slightly lower scores than neural network and logistic regression models. The random forest and SVM models seem to have lower overall performance than others, with lower scores in most metrics.

Table 6. Evaluation results for A.G.S. vs. Maghreb.

Model	AUC	CA	F1	Precision	Recall
Neural network	0.887	0.808	0.808	0.809	0.808
Logistic regression	0.872	0.796	0.796	0.797	0.796
KNN	0.846	0.782	0.775	0.780	0.782
Naïve bayes	0.811	0.779	0.781	0.785	0.779
Random forest	0.826	0.759	0.754	0.754	0.759
SVM	0.815	0.753	0.754	0.756	0.753

4.2.4. Model for Classifying Arab Gulf States vs. Egypt

In Table 7 logistic regression and neural network models performed the best overall for classifying Arab Gulf vs Egypt ethnicity, with accuracy scores of 0.860 and 0.859 respectively. SVM, KNN and random forest models also performed well but had slightly lower scores. The naïve bayes model performed the worst with lower scores across all metrics than the other models.

Table 7. Evaluation results for A.G.S. vs. Egypt.

Model	AUC	CA	F1	Precision	Recall
Logistic regression	0.917	0.860	0.858	0.858	0.860
Neural network	0.924	0.859	0.858	0.857	0.859
SVM	0.902	0.845	0.845	0.845	0.845
KNN	0.877	0.822	0.820	0.819	0.822
Random forest	0.859	0.808	0.794	0.800	0.808
Naïve bayes	0.816	0.772	0.778	0.792	0.772

4.2.5. Model for Classifying Arab Gulf States vs. Arab African Countries

Table 8 shows the evaluation results of the classification of Arab Gulf States vs. Arab African Countries, again, neural network and logistic regression models perform the best overall results with accuracy scores of 0.858 and 0.851, respectively. The naïve bayes model records the lowest accuracy result of 0.748.

Table 8. Evaluation results for A.G.S. vs. Arab African countries.

Model	AUC	CA	F1	Precision	Recall
Neural network	0.931	0.858	0.858	0.858	0.858
Logistic regression	0.922	0.851	0.850	0.850	0.851
KNN	0.877	0.804	0.799	0.805	0.804
SVM	0.877	0.799	0.798	0.798	0.779
Random forest	0.855	0.777	0.772	0.775	0.777
Naïve bayes	0.802	0.748	0.751	0.761	0.748

4.2.6. Model for Classifying Levant vs. Egypt

Table 9 displays the evaluation results of the classification of Levant vs. Egypt; although neural network and logistic regression models perform the best overall results with accuracy scores of 0.724 and 0.710, respectively, but their results are less than the previous labels. The naïve bayes model records the lowest accuracy result of 0.573.

Table 9. Evaluation results for Levant vs. Egypt

Model	AUC	CA	F1	Precision	Recall
Neural network	0.755	0.724	0.716	0.714	0.724
Logistic regression	0.732	0.710	0.705	0.702	0.710
SVM	0.665	0.656	0.651	0.648	0.656
KNN	0.625	0.646	0.625	0.619	0.646
Random forest	0.578	0.629	0.606	0.599	0.629
Naïve bayes	0.657	0.573	0.583	0.648	0.573

4.2.7. Model for Classifying Levant vs. Maghreb

Table 10 displays the evaluation results of the classification of Levant vs. Maghreb; it seems that the neural network and logistic regression models have the highest AUC scores, with 0.748 and 0.736, and highest accuracy scores of 0.699 and 0.687, respectively. However, all models' overall performance seems relatively low, with accuracy scores ranging from 0.620 to 0.699 and F1-scores ranging from 0.608 to 0.698. The KNN and naïve bayes models perform slightly better than SVM and random forest models.

Table 10. Evaluation results for Levant vs. Maghreb.

Model	AUC	CA	F1	Precision	Recall
Neural network	0.748	0.699	0.698	0.698	0.699
Logistic regression	0.736	0.687	0.686	0.685	0.687
KNN	0.668	0.644	0.630	0.641	0.644
Naïve bayes	0.670	0.635	0.632	0.632	0.635
SVM	0.660	0.621	0.622	0.624	0.621
Random forest	0.645	0.620	0.608	0.613	0.620

4.2.8. Model for Classifying Levant vs. Arab African Countries

Table 11 displays the evaluation results of the classification of the Levant vs. Arab African countries, it seems that the neural network and logistic regression models have the highest accuracy scores, with 0.830 and 0.828, respectively. The overall performance of all models seems to be good, with CA scores ranging from 0.675 to 0.830 and F1-scores ranging from 0.670 to 0.830. The SVM and KNN models perform slightly less than the neural network and logistic regression models. The naïve bayes and random forest models show the lowest performance, with accuracy scores of 0.681 and 0.675, respectively.

Table 11. Evaluation results for Levant vs. Arab African countries.

Model	AUC	CA	F1	Precision	Recall
Neural network	0.910	0.830	0.830	0.830	0.830
Logistic regression	0.907	0.828	0.827	0.828	0.828
SVM	0.802	0.726	0.725	0.725	0.726
KNN	0.792	0.721	0.714	0.728	0.721
Naïve bayes	0.734	0.681	0.681	0.681	0.681
Random forest	0.737	0.675	0.670	0.675	0.675

4.2.9. Model for Classifying Maghreb vs. Egypt

Table 12 displays the evaluation results of classification of Maghreb vs. Egypt, for accuracy, the logistic regression model records a little bit higher score than neural network with 0.737. Naïve bayes model records the lowest accuracy of 0.615.

Table 12. Evaluation results for Maghreb vs. Egypt.

Model	AUC	CA	F1	Precision	Recall
Logistic regression	0.795	0.737	0.737	0.736	0.737
Neural network	0.805	0.736	0.736	0.735	0.736
SVM	0.744	0.691	0.688	0.687	0.691
KNN	0.740	0.671	0.675	0.688	0.671
Random forest	0.687	0.655	0.654	0.654	0.655
Naïve bayes	0.689	0.615	0.615	0.666	0.615

4.2.10. Model for Classifying Maghreb vs. Arab African Countries

Table 13 displays the evaluation results of the classification of Maghreb vs. Arab African Countries; from the table, we can see that logistic regression and neural networks have the highest AUC, CA, F1, precision, and recall scores. Overall, the performance of the models seems to be relatively good, with AUC scores ranging from 0.729 to 0.860, CA scores ranging from 0.666 to 0.778, F1-scores ranging from 0.666 to 0.778, precision scores ranging from 0.666 to 0.778, and recall scores ranging from 0.666 to 0.778.

Table 13. Evaluation results for Maghreb vs. Arab African countries.

Model	AUC	CA	F1	Precision	Recall
Logistic regression	0.860	0.778	0.778	0.778	0.778
Neural network	0.855	0.771	0.770	0.771	0.771
SVM	0.743	0.678	0.677	0.678	0.678
Random forest	0.742	0.677	0.677	0.677	0.677
Naïve bayes	0.729	0.676	0.678	0.678	0.676
KNN	0.727	0.666	0.666	0.666	0.666

4.2.11. Model for Classifying Arab African Countries vs. Egypt

Table 14 displays the evaluation results of the classification of Arab African Countries vs. Egypt, in terms of CA, logistic regression has the highest score of 0.817, followed closely by the neural network at 0.810. SVM, random forest, KNN, and naïve bayes have accuracy scores ranging from 0.768 to 0.663. Overall, we can see that logistic regression and neural networks perform the best among the six models in most evaluation measures.

Table 14. Evaluation results for Arab African countries vs. Egypt.

Model	AUC	CA	F1	Precision	Recall
Logistic regression	0.895	0.817	0.816	0.816	0.817
Neural network	0.888	0.810	0.807	0.809	0.810
SVM	0.805	0.768	0.765	0.765	0.768
Random forest	0.717	0.691	0.685	0.684	0.691
KNN	0.729	0.689	0.691	0.694	0.689
Naïve bayes	0.712	0.663	0.667	0.676	0.663

5. Discussion

The superior performance of the neural network model across most classification tasks can be attributed to its ability to capture non-linear relationships and high-dimensional patterns within the deep image embeddings. These embeddings, extracted using InceptionV3, contain complex facial features that require a model capable of learning deep representations. The logistic regression model also demonstrated strong performance, particularly in binary classification settings, due to its simplicity, robustness, and efficiency in handling linearly separable feature spaces. In contrast, models such as naïve bayes underperformed, largely because of their assumption of feature independence, which is not valid for the correlated features derived from deep learning embeddings. Similarly, KNN struggled due to the curse of dimensionality, which affects distance-based models when operating in high-dimensional spaces. These justifications support empirical results and are consistent with findings in prior literature.

Moreover, in addition to multi-class classification, we conducted pairwise label classification to evaluate how well the models can distinguish between specific ethnic groups. This approach serves two primary purposes: First, it helps isolate the classification boundaries between individual group pairs such as Arab Gulf States vs. Levant or Maghreb vs. African Arab countries where visual and phenotypic overlap is known to be high. Second, it provides granular diagnostic insights where classification confusion is most likely to occur, which can guide both dataset refinement and model tuning in future work. For example, while overall ethnicity classification may perform well, the pairwise comparison between Levant and AGS reveals notable confusion, reinforcing the need for more localized features or hierarchical classification strategies. Thus, pairwise classification serves as a complementary tool to multi-class analysis, enhancing the interpretability and practical application of the model outcomes.

It is clear also that, while the proposed model achieved classification accuracies of 57.6% for gender and 72.2% for ethnicity, we acknowledge that these results are relatively modest compared to other face recognition tasks. Several factors likely contributed to these outcomes. First, the dataset used though diverse contains variability in image quality, resolution, lighting conditions, and facial expressions, which introduces noise and may reduce model generalization. Second, ethnic and gender boundaries in the Arab world are often visually subtle and culturally overlapping, making the classification task inherently challenging. Third, the models were trained using default or lightly tuned hyperparameters, and no data augmentation or facial alignment was employed in this study, which could have enhanced robustness.

Although some studies have looked at how facial features can be used to identify ethnicity, most of them focus on more broad racial categories as Asian, Caucasian, and African and they utilise meticulously chosen datasets such UTKFace [28], FairFace [13], or MORPH [21]. Usually between 80% and 95%, these studies report great CA, which is largely due to high-quality images, low inter-class ambiguity, and balanced datasets. Conversely, our work seeks to solve a more complex problem: fine-grained intra-ethnic classification in the Arab world, where phenotypic overlap is greater, and face variation is more subtle due to common ancestry and geographic proximity.

Wang and Deng *et al*. [26], for example, classified broad ethnicity with over 90% accuracy using deep CNNs on the FairFace dataset. Their models, which operated on racially different groupings, ignored the slight variation observed in Arab subgroups. Rothe *et al*. [22] reported an accuracy of 88.2% for ethnicity using the MORPH dataset; but they also focused on high-contrast categories. Our work, which employed real-world, publicly available Arab facial images and attained 72.2% accuracy for ethnicity classification and 57.6% accuracy for gender analysis, reflects the greater difficulty of the task and absence of curated or standardised datasets for this group.

Though challenging, our approach offers a reasonable starting point for more research in Arab ethnicity categorisation by assessing various traditional classifiers, such as SVM, naïve bayes, and neural network, and presenting a repeatable pipeline using deep embeddings (InceptionV3). Future advancements could involve dataset expansion, picture alignment, and ensemble model methods.

6. Conclusions and Future Work

This study introduces the first dataset that focuses on Arab countries. Each of the twenty-two Arab countries has a separate dataset. Within these datasets, there are two main categories: males and females. The data set is gender-balanced, meaning that the images of males are equal to those of females. The data is examined using six machine-learning models. The SVM model has recorded the best result in classifying gender and Ethnicity with 92.7% AUC and 57.6% accuracy. In comparison, the ANN model recorded the best result in classifying ethnicity with 92.2% AUC and 72.2% accuracy. Other classifications have been conducted between each pair of labels to evaluate the results of classifying ethnicity; overall, logistic regression and neural network perform the best among the six models in terms of the majority of the evaluation measures. This study, however, has some limitations; the dataset does not cover all ages in a balanced manner. Increasing the number of images per country to cover all age groups is recommended for future studies.

To improve performance, we propose multiple future directions:

1. Applying data augmentation techniques (e.g., rotation, cropping, illumination normalization) to increase training data diversity.
2. Integrating automated facial alignment and landmark detection to reduce spatial variability.
3. Employing hyperparameter optimization (e.g., grid search or Bayesian tuning) for each classifier.
4. Experimenting with ensemble or hierarchical models that can better capture subtle inter-class distinctions.

These strategies are expected to significantly enhance CA in follow-up studies.

Acknowledgment

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU250411].

References

[1] Acien A., Morales A., Vera-Rodriguez R., Bartolome I., and Fierrez J., "Measuring the Gender and Ethnicity Bias in Deep Models for Face Recognition," *in Proceedings of the 23rd Iberoamerican Congress on Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, Madrid, pp. 584-593, 2019. https://doi.org/10.1007/978-3-030-13469-3_68

[2] Al-Dabbas H., Azeez R., and Ali A., "Two Proposed Models for Face Recognition: Achieving High Accuracy and Speed with Artificial Intelligence," *Engineering, Technology and Applied Science Research*, vol. 14, no. 2, pp. 13706-13713, 2024. https://doi.org/10.48084/etasr.7002

[3] Al-Humaidan N. and Prince M., "A Classification of Arab Ethnicity Based on Face Image Using Deep Learning Approach," *IEEE Access*, vol. 9, pp. 50755-50766, 2021. DOI:10.1109/ACCESS.2021.3069022

[4] Anwar I. and Islam N., "Learned Features are Better for Ethnicity Classification," *Cybernetics and Information Technologies*, vol. 17, no. 3, pp. 152-164, 2017. DOI:10.1515/cait-2017-0036

[5] Awad G., Hashem H., and Nguyen H., "Identity and Ethnic/Racial Self-Labeling among Americans of Arab or Middle Eastern and North African Descent," *Identity*, vol. 21, no. 2, pp. 115-130, 2021. https://doi.org/10.1080/15283488.2021.1883277

[6] Chen H., Deng Y., and Zhang S., "Where Am I From? -East Asian Ethnicity Classification from Facial Recognition," *Project Study in Stanford University*, pp. 1-5, 2016. https://api.semanticscholar.org/CorpusID:29258943

[7] Darabant A., Borza D., and Danescu R., "Recognizing Human Races through Machine Learning-A Multi-Network, Multi-Features Study," *Mathematics*, vol. 9, no. 2, pp. 1-19, 2021. https://doi.org/10.3390/math9020195

[8] Das A., Dantcheva A., and Bremond F., "Mitigating Bias in Gender, Age and Ethnicity Classification: A Multi-Task Convolution Neural Network Approach," *in Proceedings of the European Conference of Computer Vision Workshops*, Munich, pp. 1-14, 2018. https://doi.org/10.1007/978-3-030-11009-3_35

[9] Gudi A., "Recognizing Semantic Features in Faces Using Deep Learning," *arXiv Preprint*, vol. arXiv:1512.00743, pp. 1-9, 2015. https://doi.org/10.48550/arXiv.1512.00743

[10] Heng Z., Dipu M., and Yap K., "Hybrid Supervised Deep Learning for Ethnicity Classification Using Face Images," *in Proceedings of the IEEE International Symposium on Circuits and Systems*, Florence, pp. 1-5, 2018. DOI:10.1109/ISCAS.2018.8351370

[11] Kanwar A. and Singh K., "Prediction of Age, Gender, and Ethnicity Using CNN and Facial Images in Real-Time," *in Proceedings of the IEEE World Conference on Applied Intelligence and Computing*, Sonbhadra, pp. 668-674, 2023. https://ieeexplore.ieee.org/document/10263824

[12] Karkkainen K. and Joo J., "Fairface: Face Attribute Dataset for Balanced Race, Gender, and Age," *arXiv Preprint*, vol. arXiv:1908.04913, pp. 1-11, 2019. https://doi.org/10.48550/arXiv.1908.04913

[13] Karkkainen K. and Joo J., "FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation," *in Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, Hawaii, pp. 1548-1558, 2021. DOI:10.1109/WACV48630.2021.00159

[14] Khan K., Khan R., Ali J., Uddin I., Khan S., and Roh B., "Race Classification Using Deep Learning," *Computers, Materials and Continua*, vol. 68, no. 3, pp. 3483-3498, 2021. DOI:10.32604/cmc.2021.016535

[15] Lakshmiprabha N., "Face Image Analysis Using AAM, GABOR, LBP and WD Features for Gender, Age, Expression and Ethnicity Classification," *arXiv Preprint*, vol. arXiv:1604.01684, pp. 1-16, 2016. https://doi.org/10.48550/arXiv.1604.01684

[16] Lu C., Ahmed R., Lamri A., and Anand S., "Use of Race, Ethnicity, and Ancestry Data in Health Research," *PLOS Global Public Health*, vol. 2, no. 9, pp. 1-15, 2022. DOI:10.1371/journal.pgph.0001060

[17] Masood S., Gupta S., Wajid A., Gupta S., and Ahmed M., "Prediction of Human Ethnicity from Facial Images Using Neural Networks," *in Proceedings of the Advances in Intelligent Systems and Computing*, Cairo, pp. 217-226, 2018. https://doi.org/10.1007/978-981-10-3223-3_20

[18] Molina D., Causa L., and Tapia J., "Reduction of Bias for Gender and Ethnicity from Face Images Using Automated Skin Tone Classification," *in Proceedings of the International Conference of the Biometrics Special Interest Group*, Darmstadt,

pp. 1-5, 2020. https://ieeexplore.ieee.org/document/9211042

[19] Narang N. and Bourlai T., "Gender and Ethnicity Classification Using Deep Learning in Heterogeneous Face Recognition," *in Proceedings of the International Conference on Biometrics*, Halmstad, pp. 1-8, 2016. DOI:10.1109/ICB.2016.7550082

[20] Reddy N., Rao M., and Satyanarayana C., "A Novel Face Recognition System by the Combination of Multiple Feature Descriptors," *The International Arab Journal of Information Technology*, vol. 16, no. 4, pp. 669-676, 2019. https://iajit.org/PDF/July%202019,%20No.%204/11890.pdf

[21] Ricanek K. and Tesafaye T., "MORPH: A Longitudinal Image Database of Normal Adult Age-Progression," *in Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, Southampton, pp. 341-345, 2006. DOI:10.1109/FGR.2006.78

[22] Rothe R., Timofte R., and Van Gool L., "Deep Expectation of Real and Apparent Age from a Single Image without Facial Landmarks," *International Journal of Computer Vision*, vol. 126, no. 2, pp. 144-157, 2018. https://doi.org/10.1007/s11263-016-0940-3

[23] Srinivas N., Atwal H., Rose D., Mahalingam G., Ricanek K., and Bolme D., "Age, Gender, and Fine-Grained Ethnicity Prediction Using Convolutional Neural Networks for the East Asian Face Dataset," *in Proceeding of the 12th IEEE International Conference on Automatic Face and Gesture Recognition*, Washington (DC), pp. 953-960, 2017. DOI:10.1109/FG.2017.118

[24] Sunitha G., Geetha K., Neelakandan S., Pundir A., Hemalatha S., and Kumar V., "Intelligent Deep Learning Based Ethnicity Recognition and Classification Using Facial Images," *Image and Vision Computing*, vol. 121, pp. 104404, 2022. https://doi.org/10.1016/j.imavis.2022.104404

[25] Trivedi A. and Amali D., "A Comparative Study of Machine Learning Models for Ethnicity Classification," *in Proceedings of the IOP Conference Series: Materials Science and Engineering*, Busan, pp. 1-8, 2017. DOI:10.1088/1757-899X/263/4/042091

[26] Wang M. and Deng W., "Deep Face Recognition: A Survey," *Neurocomputing*, vol. 429, pp. 215-244, 2021. https://doi.org/10.1016/j.neucom.2020.10.081

[27] Wang W., He F., and Zhao Q., "Facial Ethnicity Classification with Deep Convolutional Neural Networks," *in Proceedings of the 11th Biometric Recognition Chinese Conference*, Chengdu, pp. 176-185, 2016. https://doi.org/10.1007/978-3-319-46654-5_20

[28] Zhang Z., Song Y., and Qi H., "Age Progression/Regression by Conditional Adversarial Autoencoder," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, pp. 5810-5818, 2017. DOI:10.1109/CVPR.2017.463

Mohammed Abual-Rub is an Assistant Professor at the Department of Management Information Systems, Business School, King Faisal University, KSA, where she has been a faculty member since 2011. Abual-Rub graduated with degree in computer Sciences from Yarmouk University, Jordan, in 1996, and an M.Sc. in Computer Science from USM University, Malaysia in 2006. He then completed His Ph.D. Computer Science from USM University, in 2011. His research interests are primarily in Bioinformatics, Machine learning, and Artificial Intelligence, where he is the author/co-author of more than 12 research publications.

Khalid Nahar is a professor in the Department of Computer Sciences-Faculty of IT, he received his BS and MS degrees in computer sciences from Yarmouk University in Jordan, in 1992 and 2005 respectively. He was awarded a full scholarship to continue his PhD in Computer Sciences and Engineering from King Fahd University of Petroleum and Minerals (KFUPM), KSA. In 2013 he completed his PhD and started his job as an assistant professor at Tabuk University, KSA for 2 years. In 2015 he gets back to Yarmouk University, and became the assistant dean for quality control. From 2020 to 2022, he was assigned the chairman of the training department at the accreditation and quality assurance center. During his stay at Yarmouk University, he was promoted upgrade to associate professor, then to professor in AI and ML in the department of computer science. Research Interests include: Artificial Intelligence and Machine Learning, Speech Recognition, Arabic Computing, Natural Language Processing, Multimedia Computing, Content-Based Retrieval, IOT, and Data Science.

Ammar Almomani is a distinguished academician and cybersecurity expert currently serving as the Head of the Research and Innovation Department at Skyline University College (SUC) in Sharjah, UAE. With a Ph.D. in Computer Science from Universiti Sains Malaysia (USM), he specialises in internet infrastructure security and cybersecurity based on artificial intelligence. His illustrious academic career spans 20 years across four reputable universities, complemented by three years of industrial experience Throughout his career, he has taught more than 40 subjects in computer science and cybersecurity, supervised numerous student research projects, and holds an exceptional publication record with over 127 scholarly articles and ten patents to his credit. Recognized among the top 2% of global influential researchers by Stanford University, Professor Almomani’s contributions continue to shape the future of cybersecurity and academia.With extensive expertise in cybersecurity, his research interests include network security, cryptography, intrusion detection systems, IoT security, and cloud computing. He has authored over 85 articles in these fields, demonstrating his prolific research output. Professor Almomani also possesses a unique ability to bridge academia and industry, creating innovation opportunities. He has five years of management experience and has played a pivotal role in accreditation efforts while developing specialized programs in computing. His active presence on ResearchGate with nearly 100,000 readers and followers highlights his broad influence and dedication to advancing cybersecurity and education globally.

Firas Alzobi is an associate professor in Computer Networks Security, chair of Information System and Networks Department in the World Islamic Sciences and Education University W.I.S.E/Jordan since 2022. He received his Ph.D in Computer Networks from the Technical University of Sofia/Bulgaria in 2005. He then joined Jerash private university/Jordan in 2005. He became chair of Computer science department in 2006. He then moved to King Saud University/KSA in 2007 until 2009. He then joined Tabuk University/KSA in 2009 until 2017. Finally he joined the WISE University. Dr. Firas research interests include Networks, Sybersecurity, Cloud Computing, IoT, and Artificial Intelligence. He has published more than 20 journal papers and conference papers.