Dicle University
Journal of Engineering
https://dergipark.org.tr/tr/pub/dumf
duje.dicle.edu.tr

Research Article

Gender, age, and ethnicity estimation by image processing

Mesut UYSAL[1*], Mehmet Fatih DEMİRAL[2]

[1] Burdur Mehmet Akif Ersoy University, Institute of Science, mesuttt_4440@hotmail.com, Orcid No: 0009-0002-1650-8880
[2] Burdur Mehmet Akif Ersoy University, Faculty of Engineering and Architecture Department of Industrial Engineering, mfdemiral@mehmetakif.edu.tr, Orcid No: 0000-0003-0742-0633

ARTICLE INFO

Article history:

Received 24 October 2023
Received in revised form 9 February 2024
Accepted 10 February 2024
Available online 29 March 2024

Keywords:

CNN, image processing, python, open CV, classification

Doi: 10.24012/dumf.1380485

* Corresponding author

ABSTRACT

Today, with the increasing interest in technology, very useful studies are carried out in the field of image processing. Image technologies are also used in many fields such as security, defense, medicine, and industry. In this study, age, gender, and ethnicity were determined in the images by employing various deep-learning techniques and constructing a custom model using Convolutional Neural Networks (CNN). The dataset, consisting of 23,705 images obtained from the Kaggle dataset named "Face Data," was utilized for the analysis. The images were categorized based on gender, race, and age within the application, and the accuracy and losses of the results were visualized through graphs. Moreover, an interface was created using the Python Flask library, enabling real-time analysis of images captured from the camera to determine age, gender, and race. Among the 23,705 images, approximately 12,000 were male profiles and 11,000 were female profiles. These profiles were further classified into 5 distinct ethnicities as specified in the dataset. The ethnicities in the application were represented as follows: 0 for White, 1 for Black, 2 for Asian, 3 for Indian, and 4 for others. The most challenging aspect of this study is the variability of images due to factors such as posture, pose angle, brightness, and resolution at the time of shooting. Despite these challenges, the developed models showcased promising results, as evidenced by the accuracy metrics and visual representations provided in the study. The integration of real-time image analysis through the Python Flask interface enhances the practical applicability of the proposed techniques in various scenarios.

Introduction

Although people are by nature creatures prone to social interaction, various factors such as the environment, ethnicity, age, and gender influence their modes of interaction, speech patterns, sincerity, and communication dynamics.

Today, it is seen that the field of informatics is a rapidly developing field. Technologies such as blockchain, metaverse, and virtual printers are just a few of them. The image processing problem, which has been studied for a long time, can be summarized as follows:

It aims to make the image perceivable by computers and to benefit from the newly obtained image by processing it. Here, the input can be a video or a photograph, while the output is the section obtained from the image. To clarify the subject by giving an example through a human and a robot; A person driving a vehicle sends the data he receives by perceiving the traffic lights with his eyes to the brain and processes them there. If the light information sent is green, pass, if the information sent is red, stop, and the vehicle user acts according to the output is an example of this process. If the same problem is considered in the robot, the robot eye evaluates the data defined in the microprocessors defined by the received data and takes action accordingly, which falls within the scope of image processing.

Image processing is the operations performed on matrices. A matrix is a two-dimensional array of numbers that represents the pixels in an image. When the pictures are examined, it can be understood that each frame consists of various colors. In image processing, each element of these matrices is called a pixel. Each element of the matrix corresponds to a single pixel in the image, and the value of the element represents the intensity or color of the pixel [1]. Image processing consists of pixel cells that form these matrices. Each pixel contains a numerical value between 0-255. In the context of this study, aims to enter the field of image processing, especially targeting gender, ethnicity, and age estimation. This endeavor capitalizes on the structural features of human beings and the transformative capabilities of artificial intelligence within the rapidly evolving landscape of the informatics world. In this study, the dataset employed in this study sourced from the 'Age, Gender, and Ethnicity Face Data' dataset on Kaggle [Nipun Arora, Kaggle] [2]. 23705 image sources taken from Kaggle were integrated into the project. The information in the data set is listed as a Data frame. The reason for this is that there is more than one data type and the information is made more understandable with data frame. According to the listed data, the average age of the people belonging to the images is 33, the

youngest individual is 1, and the oldest individual is 116 years old. The age range given is mostly concentrated in the 20-40 age range. When we look at the distribution of gender and ethnic values in the application, approximately 12000 people are male and 11000 people are female. Again, in terms of ethnic distinction, approximately 10,000 people are distributed as 0 (White), 4000 people 1 (Black), 3000 people 2 (Asian), 3800 people 3 (Indian), and 1800 people 4 (Other) ethnic origin. In the dataset, 18964 people were allocated for training and 4741 for testing.

Theoretical basis

The contemporary landscape of technological advancements has sparked a surge of interest in image processing, propelling numerous impactful studies. The application of image technologies extends across diverse sectors such as security, defense, medicine, and industry. Within this dynamic context, this study delves into the intricate realms of age, gender, and ethnicity prediction using various deep learning techniques, prominently featuring the development of a bespoke CNN model.

Deep learning in image processing

Deep Learning in Image Processing Deep learning, a subfield of machine learning, involves the utilization of artificial neural networks to enable the learning of intricate patterns and representations. In the context of image processing, deep learning techniques, particularly Convolutional Neural Networks (CNN), have demonstrated remarkable efficacy in tasks such as feature extraction and pattern recognition [3].

CNN model setup for age prediction

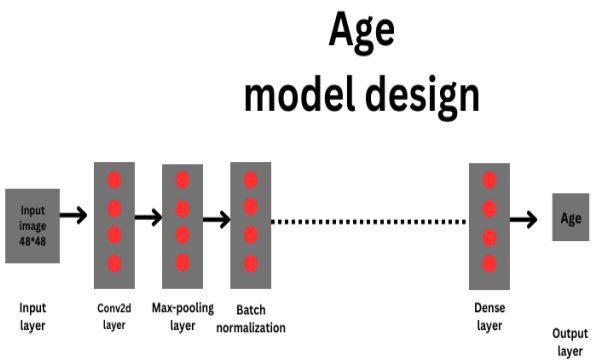

Figure 1. Age model design.

To accomplish deep learning, the integration of 4 CNN access neural networks into the model is performed respectively. These are four neural networks with 64, 128, 256, and 512 filters respectively. The first layer has (48, 48) input shape and the Relu activation function is used. In the second layer (3, 3) kernel size is defined. The third layer has a kernel size of (3, 3), and the same activation function is used. In the fourth layer, we again have a matrix dimension of (3, 3) and we set up the Relu activation function. The model also has a pool sizing layer. This model has 3 maximum pooling layers with pool size (2, 2). We also construct a model with a dropout layer with a ratio of 0.3 and 0.5 respectively. Finally, 2 dense layers with 128 and 1 neuron are added respectively. In the first dense layer, the Relu function is used, while in the second dense layer, no activation function is used since there is a regression problem. As a result, in this model, a convolutional neural network with 6 convolutions, 4 maximum layers, 4 batch normalization, 4 dropout layers, and 2 dense layers was constructed. In addition, there are 2,553,729 parameters in the model, of which 2,551,809 are trainable and 1920 are untrainable.

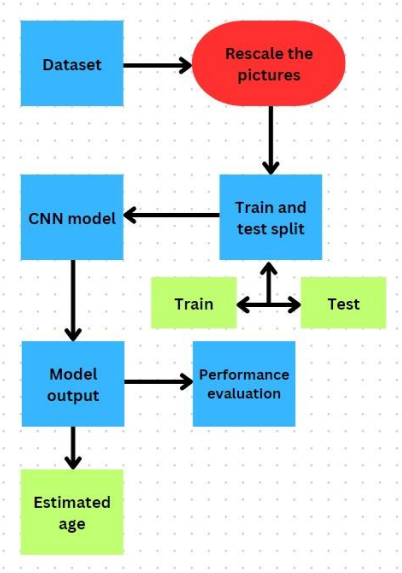

Figure 2. System architecture of the age model.

CNN model setup for gender prediction

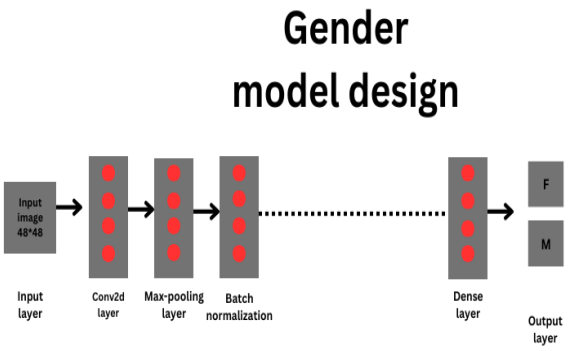

Figure 3. Gender model design.

For gender prediction, 20% of the data is allocated for testing and 80% for training. Random state 42 was chosen so that the data could be mixed and repeat more than one function call. As we created in the grief model, firstly the dimensions in the conv2D layer 48x48x1 and core images are defined. 'Relu' was used as an activation function. In the MaxPooling2D layer, the dimensions are reduced and the pool size is defined as (2, 2). In this way, the size of the attribute maps can be halved and the computational cost can be reduced. The normalization layer normalizes the mean and standard deviation of the attribute maps. Thus, the model runs faster. In the Dropout layer, it randomly resets 40% of the attribute maps by setting the value as 0.4 and prevents overfitting. In the flattened layer, multidimensional arrays help the model to learn faster and

better. In the Dense layer, we use sigmoid as the activation function, which relates each unit we give as input to each other as output. If the function is below 0.5, it gives 0, if it is above 0.5, it gives 1 result. In summary, this model is a neural network with 6 access layers, 3 maximum pooling, 3 normalization, and 2 dense layers. In total, in the model, there are 1,109,121 parameters, of which 1,108,225 are trained and 896 are untrained.

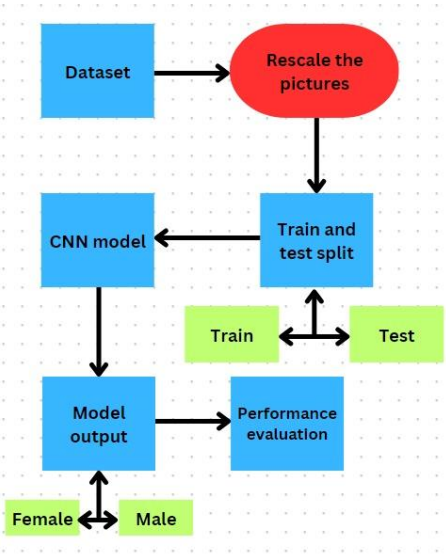

Figure 4. The system architecture of the gender model.

CNN model setup for ethnicity prediction

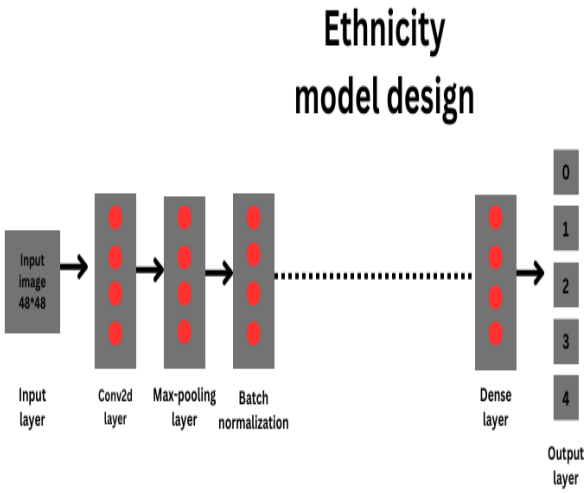

Figure 5. Ethnicity model design.

For ethnicity estimation, 20% of the data is allocated for testing and 80% for training. For the random state, the desired value can be given, but the value of 42 was chosen. Conv2D layer applies a 2D convolution layer on the input image using 64 filters of 3x3 size and RELU activation function. This layer is used to convolve the image features by sliding a filter over it. In the pooling layer, a 2x2 filter is selected to extract the feature map. Thanks to this layer, the number of parameters is reduced to prevent overfitting. In the Batch Normalization layer, the images are normalized to improve the performance of the model. After this stage, a conv2D layer with a 3x3 size with 128 filters and RELU activation is applied to the output of the previous layer with a 2D convolution process. With this layer, more features are extracted from the image by applying more filters. In the dropout layer, the data is randomly distributed with a probability of 40% by giving a rate of 0.4 and overfitting is prevented. In the smoothing layer, the data is reduced to one dimension and prepared as input for dense layers. The dense layer is a dense layer that applies a fully connected process using 5 units and a sigmoid activation function over the input vector. Here, each of the 5 classes (white, black, Asian, Indian, and others) gives us a probability output.

This model has a total of 1.109.381 parameters, of which 1.108.485 are trainable and 896 are non-trainable. The accuracy loss for ethnicity was found to be 0.7979, which is not a bad rate, but it is open to improvement.

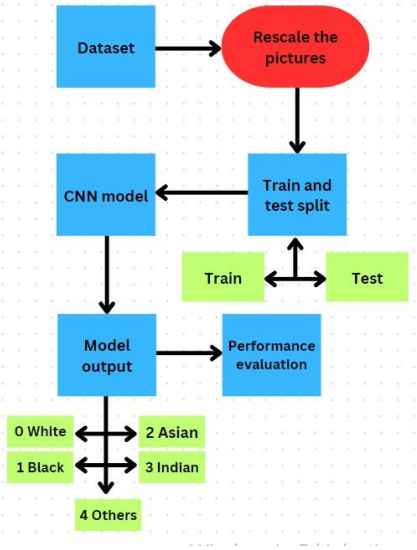

Figure 6. The system architecture of the ethnicity model.

Library, methods, and algorithms used

OpenCV

It is an open-source library that supports applications such as machine learning, image processing, and data analysis. It is used by many software languages such as Java, C++, Python, and MATLAB.

With the increase in R&D studies, the concept of computerized vision is gaining importance. With computer vision, we can understand and interpret images. For example, when uploading and labeling images on Facebook, the people who will be labeled come as a frame, and face detection algorithms are embedded in Facebook. In this way, faces in the picture can be found. Again, interpreting the state and movement images of someone who makes suspicious movements in terminals or crowded places and reporting them to security units, again taking into account the facial expression of the audience in the cinema and improving the films, computer vision is becoming valuable in parallel open cv. In this study, all codes are written in Python language. The OpenCV library contains hundreds of functions that support the capture, analysis, and manipulation of visual information connected to a computer by webcams, video files, and other types of devices. While simple functions

can be used to draw a line on a screen, more advanced parts of the library can include algorithms to detect faces, track motion and analyze shapes. In this application, OpenCV was used to import and resize images.

Pandas

One of the most important libraries used in data-related processing is Pandas. It is used in stages such as integrating the data set, reading, and processing the data. To install pandas on our computer, the "pip install pandas" command can be written from cmd. However, since Anaconda is used in this project, libraries such as NumPy, matplotlib, and skit-learn will be installed. This library was used in data processing by importing it into the project. Pandas are divided into 2 series and data frames. While a series consists of a single column; a data frame consists of multiple columns. In addition, the data frame contains more than one different data type such as int, string, bool is also an advantage for us. Since each of the values of age, gender, image name, and pixel values are of different types in the data set, the data set is defined in the data frame type.

	age	ethnicity	gender	img_name	pixels
0	1	2	0	20161219203650636.jpg.chip.jpg	129 128 128 126 127 130 133 135 139 142 145 14...
1	1	2	0	20161219222752047.jpg.chip.jpg	164 74 111 168 169 171 175 182 184 188 193 199...
2	1	2	0	20161219222832191.jpg.chip.jpg	67 70 71 70 69 67 70 79 90 103 116 132 145 155...
3	1	2	0	20161220144911423.jpg.chip.jpg	193 197 198 200 199 200 202 203 204 205 208 21...
4	1	2	0	20161220144914327.jpg.chip.jpg	202 205 209 210 209 209 210 211 212 214 218 21...

Figure 7. Information about the individuals in the data.

In Figure 7, the age, ethnicity, gender, image name, and pixel values of the first 5 people in the data set are given as Data frames. In projects with large data sets, such operations are performed to see the data.

	count	mean	std	min	25%	50%	75%	max
age	23705.0	33.300907	19.885708	1.0	23.0	29.0	45.0	116.0
ethnicity	23705.0	1.269226	1.345638	0.0	0.0	1.0	2.0	4.0
gender	23705.0	0.477283	0.499494	0.0	0.0	0.0	1.0	1.0

Figure 8. Structural content of the data set.

Figure 8 shows the structure of the dataset with the describe command to recognize the content of the data. In the dataset; there are 23705 people, the average age value is 33,300, and the standard deviation is 19,88.

Seaborn

The Seaborn library is built like the Matplotlib library and is used for data visualization. Visualizing data is extremely important. To understand, interpret, and draw conclusions from large data sets, we need to have objective visible values in front of us. Visualizing the data is known as expressing the data with a graph or line [4]. In the project, the distribution of the age variable is expressed by the plot method.

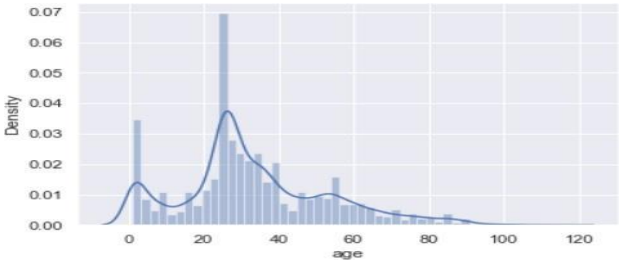

Figure 9. Age distribution of individuals in the data set.

According to Figure 9, it is seen that the people in the data set are predominantly between the ages of 20-40.

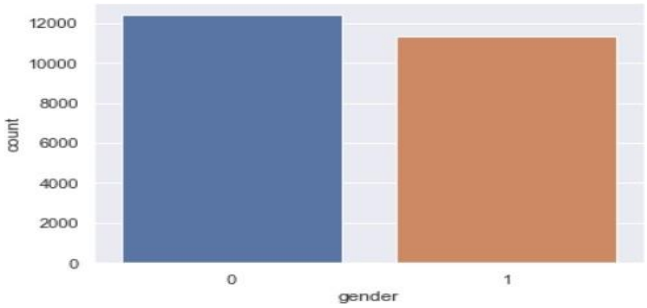

Figure 10. Gender distribution of individuals in the data set.

In Figure 10, the attributes of the gender variable are visualized with the count plot method. There are approximately 12 thousand men and 11 thousand women in the project.

NumPy

NumPy is a Python library that allows to work with matrix and array operations in mathematical operations. It is one of the most preferred libraries by data scientists. NumPy has a fixed size when creating arrays. But Python is an array whose size can change when creating arrays [5]. In the code, the pixel stack is transformed into a NumPy array, where the images are structured as 48x48.

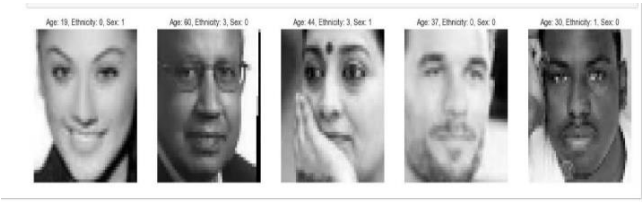

Image 1. Display of age, gender, and ethnicity values of the pictures in the data set.

As shown in Image 1, 5 pictures with a size of 20x10 were randomly drawn. The age, ethnicity, and gender of the people are shown in the picture. In the code, the pictures were sized 48x48 and entered into the CNN model. Then, since the pixel values of each image are between 0 and 255, the normalization process was performed by dividing the pixel value by 255 and compressing all values between 0 and 1. The normalization process provides better and faster training when training the model. Using the Batch hyperparameter, it is specified how many times it is processed.

Scikit-learn

The scikit learn library is one of the most basic libraries used in machine learning problems. Scikit-learn includes pre-implemented algorithms and metrics for a range of machine learning tasks such as classification and regression. This library enables data scientists and researchers to quickly apply various machine learning models and assess their performance efficiently.

This library was used to train-test the dataset and to measure the performance of the regression model for age. Of the data set used, 80% was used for training and 20% was used as test data.

Classification metrics

The process of obtaining a model using a method based on training data and using this model in prediction is called classification. In other words, classification is the process of predicting new incoming data with the experience gained from existing data. Features and result information about these features are kept in the training data. Categorical information is produced as a result of the result information and the prediction made. There are many methods in the literature. In these articles, preliminary information about the basic methods will be given [6].

Accuracy: The number of correct predictions / Number of all predictions is found by the formula. The overall performance of the model is found.

The accuracy value of our model: is 90.66.

Precision: It deals with how many predictions are correct.

This one as formulated is. True Positive / True Positive+ False

the precision value of our Positive Model;

Sensitivity= 0.91.

Recall (sensitivity): True Positive / True Positive + False Positive.

Recall values in this model; Sensitivity=0.91.

F1 score: The F1 score is used to balance precision and recall. It is formulated as F1=2 x precision x recall/precision + recall. The f1 score value of our model: is 0.91 [7].

Confusion matrix

The confusion matrix is the agreement table showing the agreement between the actual labels and the prediction of the model. In this matrix, the row contains the number of predicted positive and negative values, while the column contains the actual positive and negative values.

Figure 11 shows a confusion matrix showing the performance of the test set with ethnicity data.

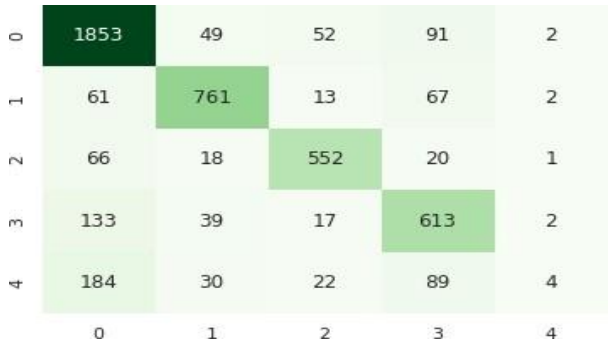

Figure 11. Performance results of test set with confusion matrix for ethnicity data.

- The ethnicity in the (0,0) position (first row) is generally well-predicted, as evidenced by the high value.
- The ethnicity in the (1,1) position (second row) also has good prediction results, with the highest value.
- The ethnicity in the (2,2) position (third row) is reasonably well-predicted, as shown by the high value.
- The ethnicity in the (3,3) position (fourth row) demonstrates a decent prediction, with the highest value.
- The ethnicity in the (4,4) position (fifth row) faces challenges, especially indicated by the relatively lower values.

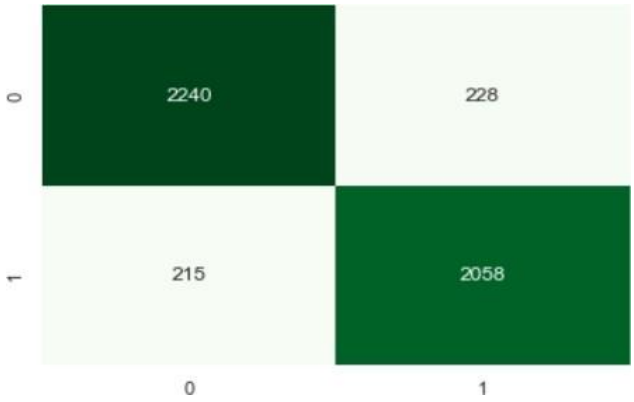

Figure 12. Performance results of test set with confusion matrix for gender data.

In Figure 12, there are 4741 data reserved for the test data in the confusion matrix. The predictions are as follows:

- True positives: 2240
- False positives: 228
- True negatives: 2058
- False negatives: 215

If the confusion matrix metric is interpreted here, the model predicts 2240 data as male and knows 2240 data correctly. However, it made a wrong prediction by predicting 215 data as female. In the same way, it predicted 2058 data as 1 (female)

and made the correct prediction. However, for 228 data, it is predicted as 0 (male).

Regression metrics

MAE (Mean Absolute Error): It is one of the regression metrics. It is referred to as mean absolute error. It is a more direct representation of the sum of the error terms by taking the sum of the absolute error values. The MAE of our model is approx. MAE 5.896764. This means that the error rate of our age regression model is approximately this value. The model can calculate the age of a person who is 65 as 60 or 70 and give us the result as a result.

Errors Graph: The graph shows the visual result of the errors graph of the model. According to the graph, errors can be seen. If it were a straight line, it would be assumed to be an error-free model. This graph expresses the relationship between the actual value and the predicted values. The width of the graph is 8 inches and the height is 6 inches. Some scatter points on the graph indicate error values (Figure 13).

Mean Square Error (MSE) is a measure used to represent regression models. It gives an absolute number of how much the predicted results differ from the actual values. In other words, it measures how close the model's predictions are to the real values.

A lower MSE value indicates better model performance. The MSE value of the Age model was obtained as 68.92.

Root Mean Square Error (RMSE): It gives an absolute number of how much the predicted results differ from the actual values. That is, it measures how close the model's predictions are to the true values.

RMSE is the square root of MSE (Mean Square Error). MSE is the mean value of the squared error. The error of the prediction is squared, these squares are taken and the result is a number.

A lower RMSE indicates a better model demonstration

The RMSE value of the Age model was found to be 8.302.

R^2 (Coefficient of Determination): is how close a part is to the recorded regression line.

R^2 indicates how well the model's predictions fit the observed pattern. The highest value can be 1. The closer R^2 is to 1 for the data set, the higher the performance of the model on the data. R^2 value in the age model was calculated as 0.821.

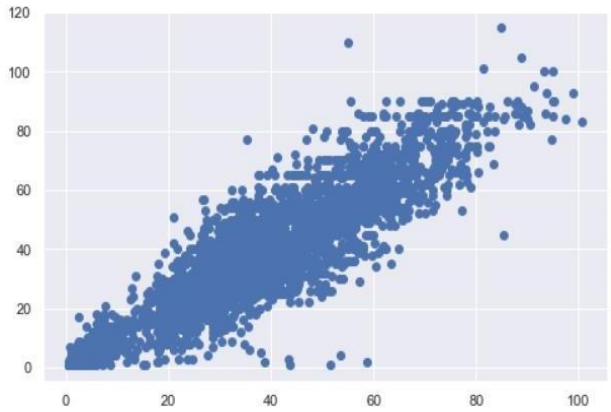

Figure 13. Errors graph of the model.

Keras

Keras library is a machine learning library that can work on tensor flow used for deep learning and is used in this project. Since deep learning techniques are used in practice;

- Defining training data,
- Layer and model formation,
- This library provides services such as epoch generation and optimization, definition of hyperparameters, etc. using for.

Keras has many usage advantages;

The created model can run smoothly on the CPU and also supports this structure when CNN models are preferred. For the age model in the application, the layers to be added are first defined. In the first hidden layer, a filter with 64 neurons and a 3x3 size is defined. The defined neuron may vary depending on the size of the application. Networks that usually start with 32 or 64 neurons can also be defined as 256, 512, or 1024 if there is a very complex dataset. Filter values should be increased gradually from smallest to largest. The structure specified as the kernel size is known as the kernel size of the cov2d parameter. The specified heap size can be (1,1) (3,3) (5,5) and (7,7). If the input images are large, such as 128x128, a kernel size greater than 3 can be chosen to aid learning. Since the images are 48x48 in size, a 3x3 kernel size was chosen. Layers of the CNN model were used here.

Firstly, the convolutional layer (CNN) layer is used to detect images whose data set is image and video. In this layer, the features of the image are taken and the feature table is created by multiplying the filter values and pixel values of the image. While we could define 2 parameters for the cov2D class in the convolutional layer, the value "same" was assigned to the padding value of the output volume to match the input size volume and maintain its size. The Max Pooling layer, used as an intermediate layer, is used to reduce the size of the image. The Max Pooling layer is a section of the pool layer in the Max Pooling layer. It is used to prevent loss of value from the image. Normalization operations are performed in the created layer and the values obtained are compressed between 0 and 1. In the flattened layer, the image is flattened and decomposed as a single vector. It is necessary to create as many layers as the output to be obtained.

When using artificial neural networks in the defined layers, activation functions are used. Because these functions are useful in expressing how neurons change. Generally, non-linear activation functions are preferred. This is because they increase the complexity of the model and encourage better model learning. These functions compress the output within a certain range. Relu function is used. In Relu, values less than zero take the value 0, and values greater than 0 take the value they are in. The value of Relu is already from 0 to plus infinity. The reason for choosing the "Adam" parameter from the optimizer function during the compilation phase after layers are created in the model is to control the learning, and generally, the "Adam" parameter provides a good optimization. Again, the "accuracy" metric can be used to see how the model will perform during training and to see the loss value.

Another important problem encountered in training artificial neural networks is Early Stopping. This event is briefly defined

as the process of memorizing the model and repeating itself continuously. This is caused by the increasing difference between the test validation and the loss value of the train. The defined callbacks parameter allows monitoring and triggering the performance and stopping the training process according to the situation. "Monitor" allows the training terminator to be measured according to the performance to be monitored. The behavior of the metric can be determined with the mode argument. It was stated that there is a tendency to increase by selecting max. In this way, if the verbose value is selected as 1, instantly updated results can be seen in the application.

Python flask

Flask is a framework used to create web services in Python language. It is a framework that can be learned quickly and has high performance. To install Flask, we install it with the pip install flask command in the terminal command of our project. We make use of the flask library as follows for the establishment of the camera connection and reading the image from the camera.

First of all, the flask library is integrated into the project for the process of creating a web interface by taking images, the function we write called video-stream generates the received images as bytes, and the received image frames are read and encoded in JPEG format. The byte stream is then responded to as an HTTP with the content type set to image/jpeg.

Framing and image framing in the flask library

The image retrieval process starts by sending a post request to HTTP. The received image is resized to 48x48 pixels by cropping only the pixels between 60-470 rows and 150-500 columns. Each pixel value in the image is divided by 255 and the normalization process is performed. And reshaping is done by making it 4-dimensional. After the frame editing process is finished, we transfer the 3 models we have previously trained to predict the age, gender, and ethnicity of the person to our application and perform the prediction operation for each model. For example, to predict gender, we need to integrate the previously trained gender. model5 model into the system, then the uploaded image is compared with the images in the trained model and sends us a prediction result between 0-1. Here 0 means male and 1 means female. If the prediction value is less than 0.5, the process of finding gender is performed by assigning it to the male group, otherwise to the female group. When we do the same process in age prediction, the person to be predicted is given to the previously trained age model. The results are stored in the age variable we define in the code. Since the age parameter is a numeric value, the output value is rounded to the nearest integer with the help of the round function in Python in order not to obtain a fractional number. (For example, when the age value of the person is 56,59, the output will give us the value 57).

Other parameters used

Epochs: Determines how many times the training data will be given to the model. In this code, 15 epochs are specified, which means the model will see each image 15 times.

Batch_size: Determines how many images the model will process in each step. In this code, batch size is set to = 64. At each step, the model will take 64 images and their losses and gradients will be calculated.

Verbose: Determines whether the model will show a message during its run. It can take values of 0,1 or 2. In this code, when verbose = 1 is defined in this code, a progress bar and loss metrics will be seen at the end of each epoch.

Literature review on image processing

The literature aims to investigate age, ethnicity, and gender determination according to facial features in our country and the world based on image processing-based machine learning techniques. Furthermore, by following the data of the person, mourning can be determined according to height, weight, and facial features. Dealing with CNN models involves handling more irregular data compared to traditional machine learning methods.

Uçar conducted a study on distraction, focusing on angular changes in head movements within a classroom environment. The study, which was carried out with C ++ software, utilized the libraries required in image processing such as Python Open Cv. In the classroom, the student's face and facial expressions are recorded every 5 seconds, and a training set is created. Students' head structures are recorded and stored in 3 dimensions. It is determined that the students who are in the direction of the teacher are more attentive. Throughout the study, the UPNA database was used for the face recognition of the students. Each of the available data was divided into 2 classes (careful, and inattentive) by 5 different people. The datasets were tested with different algorithms and the highest performance was obtained with the SVM algorithm [8].

Günay and Nabiyev estimated age using facial data, employing the LBP histogram. The LBP data, segmented by an attributive vector, underwent classification using methods like K-NN and minimum distance. Euclidean distances were calculated for all samples. The best experimental performance was calculated at 89% [9].

Ayata and Çavuş addressed the issue of person recognition and differentiation in images for security and criminology purposes. This study is based on artificial neural networks, a sub-branch of machine learning. Support vector machines and DSA methods were used to process all the data in the FEI dataset, Celeb A dataset, and Family dataset. The study achieved success rates ranging from 95% to 99%. For robust face recognition, artificial neural network-based classification was favored [10].

Eldem et al. conducted scientific research in the field of image processing with a study on the facial features of the person in the scientific field of image processing and the extent to which it resembles other individuals. The images processed in this study were obtained from cameras and external image recording devices. The images were digitized using the OpenCV library, known for its compatibility with C, C++, and Python and its versatility on platforms such as Android and Linux. Cameras were set at a distance of approximately 55-60 cm during data acquisition, and the dataset included information such as the person's name, surname, and a given ID [11].

In the studies of Gündüz and Cedimoğlu, the aim was to predict the gender of individuals using deep learning algorithms, regardless of age. Approximately 10% of the 6508 data were allocated for testing, while 90% were used for training. Gender

prediction was accomplished through data augmentation with a deep learning network. The study utilized Python programming language, Ubuntu, and Linux operating systems. CNN was employed, with Keras and Tensor flow libraries. VGG-16 achieved the best performance among the models used [12].

Kaya alp and Metlek predicted the gender of the person by taking 63228 images from the wiki database. A support vector machine was used in this application. Attention was given to creating a hyperplane, ensuring that the features of the classes consisted of the two most distant lines. In deep learning, speed plays a crucial role as it concurrently executes image recognition and classification processes. While CUDA is commonly employed in such projects based on the power of the computer graphics card, this study utilized MATLAB-based MatConvNet-1.0-beta15. At the point of success, it was classified according to SVM, and 80% of it was divided into training and 20% as tests. According to the results of the complexity matrix, the accuracy rate of the developed system is 94.48% [13].

In the scientific research of Toprak, the problem of age estimation with image processing techniques was addressed. LBP and HOGC histogram methods were used throughout the study. The data were classified according to the K-NN algorithm. When conducting such studies, models typically consider examples such as Anthropometric, Active, Appearance, Leaning Pattern, and Mourning Manifold. For instance, estimating the age range is treated as a classification problem, while precise age estimation is considered a regression problem. The IMDB-WIKI model was used for the database, consisting of 523051 images divided into a 90% training set and a 10% test set. For the performance of the system, mean absolute error (MAE), one of the error measurement techniques, was used [14].

Literature review: comparative analysis of deep learning approaches in image analysis studies

Table 1. Literature review comparison.

Features	Gender, Age, and Ethnicity Estimation by Image Processing	Recognition Of Students in The Classroom Environment and Detection of Distractions with Real-Time Image Processing [3].	Investigating The Effects of Facial Regions on Age Estimation [4].	Face Detection System Development with Image Processing Techniques [6].	Gender Estimation with Image by Using Deep Learning Algorithms [7].
Focus	Age, gender, race prediction	Age, gender, and race prediction	Age estimation based on facial regions	Image Processing Application	Gender Prediction
Model Type	CNN	CNN	LPQ and regional analysis	OpenCV and OpenCV Sharp	Alex Net and VGG-16
Dataset	A dataset comprising data from approximately 23,000 individuals.	UPNA Head Pose Database	FGNET and PAL Databases	An original dataset was curated specifically for this study, consisting of images captured through a computer camera.	A dataset generated from Wikipedia images
Performance Analysis	Emphasized performance differences in age and gender predictions	Highlighted lower performance in the genetic class compared to the gender class	Found eye region to be more effective in age prediction than other regions	Achieved 79% success in face recognition application	Mentioned 99.41% success of VGG-16 in gender prediction
Future Recommendations	Suggested increasing the number of training iterations, adding more filters and layers, and working with higher-resolution photos	Recommended adding more features (eye direction, emotional state, etc.), conducting in-class tests with professional cameras, and expanding the dataset for attention distribution	Proposed working on determining regional weights in addition to regional analysis for better prediction accuracy	Recommended enhancing the success and performance of the developed OpenCV-based application by adding new features and training with more extensive datasets	Suggested experimenting with data diversity, data augmentation, minibatch augmentation, optimizer differentiation, and layer augmentation for better results.

Contributions to the literature

This study makes significant contributions to the application of deep learning techniques in predicting age, gender, and race from images and videos, addressing gaps in the existing literature and enhancing current knowledge:

In-depth analysis in multi-class prediction: Particularly in scenarios where the genetic class is subdivided into five subclasses, the study provides a detailed analysis of the performance of age, ethnicity, and gender prediction models, offering a valuable perspective to the literature. Emphasizing the challenges of class imbalance and multi-class scenarios can guide future similar studies.

Challenges and solutions in real-time applications: The study thoroughly explains the challenges associated with capturing real-time images from web cameras and details the solutions developed to overcome these challenges. This can serve as a

crucial reference for the development of similar applications or the enhancement of existing ones.

Recommendations for future research: By providing suggestions on how the study can be further improved through future training with higher-resolution images and the utilization of additional filters and layers in convolutional neural networks, the research offers a guiding framework for researchers and practitioners.

Ethnic diversity and deep learning: While most deep learning-based studies predominantly focus on gender and age, there is a pressing need for more research addressing ethnic diversity. This would enable a more effective evaluation of model performance across a broad demographic spectrum.

These contributions provide a valuable perspective on image analytics based on deep learning, filling gaps in the literature and inspiring future research endeavors.

Results and outputs

Printouts of the image received in the web environment

The photo in the picture belongs to me and I am 26 years old as of the date of uploading the photo (Image 2). The format of the photo is jpeg.

Image 2. Reshaped photograph.

When the image is uploaded to the application and reshaped, age, gender, and genetic prediction values are given as output in Figure 11.

```
In [6]: ethincyt[0]
Out[6]: 1

In [7]: gender[0]
Out[7]: 'Erkek'

In [8]: age[0][0]
Out[8]: 26.563057
```

Figure 11. Age, gender, and ethnicity prediction values of the picture.

While estimating age and gender in the photograph, the regression method was used for age, and the classification method was used for gender. The photographs in the data set are in pixel form. Since the pictures are 48x48 in size, they are in grey (black and white) form. The small size of the image makes the training easier and facilitates our work in training the model.

Performance results of the ethnicity model

Table 2 shows the results of the test set of the ethnicity class. For each class, 4 metrics are given. Precision is the rate at which that model predicts that class correctly. Recall shows how many of the images belonging to that class are correctly predicted. F1-score, precision, and recall show how many of the images belonging to that class are correctly predicted. According to this model, the best performance of the model is 0 (0.85 f1-score) and the worst performance is 4. (0.04 f1-score.) The overall accuracy of the model is 0.80.

Table 2. Performance results of the ethnicity class of the model.

	Precision	Recall	F1-score	Support
0	0.82	0.89	0.85	2047
1	0.83	0.84	0.83	904
2	0.84	0.86	0.85	657
3	0.69	0.78	0.74	804
4	0.70	0.02	0.04	329
Accuracy			0.80	4741
Macro avg	0.78	0.68	0.66	4741
Weighted avg	0.79	0.80	0.77	4741

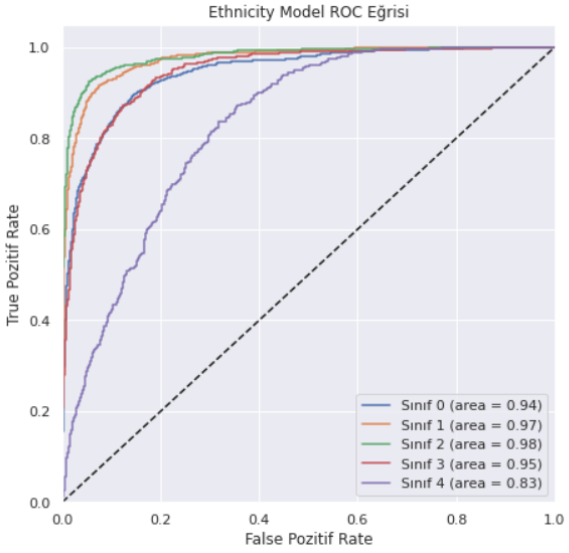

Figure 14. Ethnicity model roc curve.

Looking at the metrics of the model, we can say that it generally performs well. However, we see that some classes perform lower than others.

Precision, recall, f1-score values for classes 0, 1, and 2 are quite high. This indicates that our model is capable of predicting these classes accurately.

Precision and recall values for Class 3 are slightly lower but still acceptable. However, for this class, your model's performance is slightly lower than other classes.

Although the precision value for Class 4 is at an acceptable level, the recall and f1-score values are very low. This indicates that your model's ability to accurately predict this class is quite low. This means that our model often predicts this class.

The overall accuracy value is 80%. This means your model made 80% of all predictions correctly.

Performance results of the gender model

According to the data given in Table 3, a 91% accuracy value was obtained for both classes according to the gender model.

Table 3. Performance results of the gender model.

	Precision	Recall	F1-score	Support
0	0.92	0.90	0.91	2468
1	0.90	0.91	0.90	2273
Accuracy			0.91	4741
Macro avg	0.91	0.91	0.91	4741
Weighted avg	0.91	0.91	0.91	4741

Sensitivity: We see that the model has a precision of 92% for class 0 and 90% for class 1. This shows that most of the positive prediction examples provided by the model are positive.

Recall (Sensitivity): We see that the model has a sensitivity of 90% for class 0 and 91% for class 1. This shows that most of the true positive instances of your model are correctly detected.

F1-Score: The F1 score is a harmonic storage of precision and improvement and indicates that your model has an overall balanced performance. We see that your model has an F1 score of around 91% for both classes.

Accuracy (Accuracy): Your model is 91% overall. This shows that your model made 91% of all its predictions correctly.

Figure 15. Gender model roc curve.

Figure 15 shows a curve named “Gender Model 1”. The model's AUC (Area Under the Curve) score was calculated to be approximately 0.906. The blue line on the chart represents ROC curves. This curve initially rises quickly and then flattens out, indicating that the model is performing well. The X-Axis is labeled “False Positive Rate” and has a value between 0 and 1. The Y axis is labeled “True Positive Rate” and again takes a value between 0 and 1.

Performance results of the age model

Table 4. Result of the age model.

	MAE	**MSE**	**RMSE**	**R2 Score**
Age Model Metrics	5.892996	68.927456	8.302256	0.821662

Mean Absolute Error (Mean Absolute Error): This metric measures how much your model's predictions deviate from the true values. The average mean Absolute Error of your model is approximately 5.89. This shows that your model's predictions differ from the actual values by 5.89 units on average.

Mean Square Error (Mean Square Error): This metric measures how much your model's predictions deviate from the actual values, taking the squared errors. However, it causes major mistakes to be punished more. The average speed Square Error of your model is approximately 68.93.

Square Root Mean Square Error (Square Root Mean Square Error): This metric takes the square root of the distance Square Error and measures how much your model's predictions deviate from the actual values. The average speed of your model is approximately 8.30 Square Root of Square Error.

R2 Score (R Squared Score): This metric measures how much better your model's predictions are than predictions made using the mean of the target variable. Your model's R Square Score is 0.82, indicating that your model explains 82% of its variance. This shows that your model is pretty good.

As a result, the overall performance of your model is quite good.

Conclusion

In conclusion, this study applied deep learning techniques to predict age, gender, and race from images and videos. Utilizing a dataset of approximately 23 thousand individuals, Convolutional Neural Network (CNN) models were trained and evaluated. The integration of the model into a web environment for real-time image data input was achieved using the Python Flask library.

The performance analysis of the CNN models revealed nuances between age and gender predictions. The genetic class, encompassing multiple classes (5), exhibited lower precision, recall, F1-score, and support values compared to the gender class, which consisted of only 2 classes. This discrepancy can be attributed to the inherent complexity of predicting multiple genetic classes.

The application's image size, set at 48x48 pixels, influenced accuracy and precision. Larger and higher-resolution images are anticipated to yield more successful results, albeit with potential delays in processing time. Addressing real-time challenges, such as instantaneous changes in predictions from webcam-captured images, required meticulous attention to environmental factors and user-related variables.

Notably, the implementation of a cropping process significantly improved the model's ability to interpret pixel values, mitigating deviations and enhancing accuracy.

To advance these findings, future studies may consider increasing the number of training iterations, incorporating more filters and layers into CNN models, and experimenting with higher-resolution photos. These enhancements could contribute to the refinement and robustness of the predictive models.

In summary, while the study sheds light on the potential of deep learning in image-based predictions, continuous refinement, and adaptation are crucial to addressing challenges and unlocking the full capabilities of these predictive models.

Conflicts of interest

The authors declare that there is no conflict of interest in this study.

Declaration of ethical code

In this study, the authors undertake that they comply with all the rules within the scope of the "Higher Education Institutions Scientific Research and Publication Ethics Directive" and that they do not take any of the actions under the heading "Actions Contrary to Scientific Research and Publication Ethics" of the relevant directive.

Authors' contributions

MU realized the idea and implementation. MFD found and evaluated the performance analyses of the application. He analyzed whether the manuscript conformed to the template. MFD and MU revised the typos and logic errors together and translated the manuscript into English. MFD measured the plagiarism rate. Both authors reviewed and finalized the final version of the manuscript.

Sample Statement:

Author 1: app idea, literature review, coding, analysis result, planning

Author 2: application result, creation of draft text, reducing the similarity ratio

Acknowledgments

I would like to thank my advisor, Assoc. Prof., for his efforts and contributions in the process from the idea phase of this study to the implementation phase and the article writing template. Dr. We would like to thank Mehmet Fatih Demiral.

References

[1] Ç. Kılınç, "Why Do We Use Matrices for Image Processing?" [Online]. Available: https://medium.com/@cgtyklnc/why-do-we-use-matrices-for-image-processing-3b24a59abe4f/, Accessed on: Jan. 6, 2023.

[2] N. Arora, "Age, Gender, and Ethnicity Face Data." [Online]. Available: https://www.kaggle.com/datasets/nipunarora8/age-gender-and-ethnicity-face-data-csv/, Accessed on: Sep. 2, 2023.

[3] Protopars, "Derin öğrenme (deep learning) nedir?" [Online]. Available: https://www.protopars.com/derin-ogrenme-deep-learning-nedir/, Accessed on: May. 13, 2023.

[4] Statology, "The easiest way to use seaborn: import seaborn as sns". [Online]. Available: https://www.statology.org/import-seaborn-as-sns/, Accessed on: May. 19, 2023.

[5] T. Ergin, "Keras ile derin öğrenme model oluşturma" [Online]. Available: https://medium.com/@tuncerergin-/keras-ile-derin-ogrenme-model-olusturma-4b4ffdc35323, Oct 2, 2018. Accessed on: June 11, 2023.

[6] E. Uzun,"Makine öğrenmesi" [Online]. Available: https://erdincuzun.com/makine_ogrenmesi/makine-ogrenmesi-metotlari/, Accessed on: Jan. 10, 2023.

[7] M. F. Akca, "Sınıflandırma problemlerindeki metrikler". [Online]. Available: https://medium.com/deep-learning-turkiye/s%C4%B1n%C4%B1fland%C4%B1rma-problemlerindeki-metrikler-33ee5f30f8eb, Accessed on: April. 15, 2023.

[8] M. U. Uçar, "Recognition of students in the classroom environment and detection of distractions with real-time image processing." Master's thesis, Department of Electrical and Electronics Engineering, İskenderun Technical University, Hatay, 2019.

[9] A. Günay, and V. Nabiyev, "Investigating the effects of facial regions to age estimation." *Türkiye Bilişim Vakfı Journal of Computer Science and Engineering*, vol. 9, no. 2, pp.1-10, 2016.

[10] F. Ayata, and H. Çavuş, "Performance tests of ESA, YGH-DVM and DSA algorithms used in face recognition systems." *Fırat University Journal of Science and Technology*, vol. 34, no.1, pp. 39-48, 2022.

[11] A. Eldem, H. Eldem, and A. Palalı, "Face detection system development with image processing techniques". *Bitlis Eren University Journal of Science and Technology*, vol.6, no.2, pp. 44-48, 2017.

[12] G. Gündüz, and İ. H. Cedimoğlu, "Gender estimation with image by using deep learning algorithms." *Sakarya University Journal of Computer and Information Sciences*, vol.2, no.1, pp. 9-17, 2019.

[13] K. Kayaalp, and S. Metlek, "Detection of fish species with deep learning." *International Journal of 3D Printing Technologies and Digital Industry*, vol.5, no.3, pp. 569-576. 2021.

[14] Ö. Toprak, "Age estimation with image processing techniques," Master's thesis, Institute of Science and Technology, Maltepe University, Istanbul, 2019.