Research Article

Real-time Age and Gender Classification using VGG19

Muhammad Usman Tariq[1], Arslan Akram[2*], Sobia Yaqoob[3], Mehwish Rasheed[4] and Muhammad Salman Ali[5]

[1]*Mphil Scholar, Department of Computer Science, Superior University Pakistan.*

[2]*Ph.D Scholar, Department of Computer Science, Superior University Pakistan.*

[3]*Ph.D Scholar, Lecturer Department of Computer Science, University of Okara, Pakistan.*

[4]*Ph.D Scholar, Lecturer, Department of Computer Science and Information Technology, Superior University Pakistan.*

[5]*Mphil Scholar, Department of Computer Science, Superior University Pakistan.*

*Corresponding Author

Arslan Akram, Ph.D. Scholar, Department of Computer Science, Superior University Pakistan.

Submitted: 2023 Sep 28; **Accepted**: 2023 Oct 17; **Published**: 2023 Oct 20

Abstract

Unrestricted real-world facial photographs are arranged into specified age and gender groups using unprocessed face age and gender estimations. This explorer nation has now been prefabbed with earth-shattering enhancements due to its value in speedy real-world applications. However, conventional approaches utilizing unfiltered benchmarks show their incapacity to handle higher levels of variance in such unrestricted photographs. Convolutional Neural Networks (CNNs) enabled approaches have recently been widely used during categorization tasks due to their superior performance in facial psychotherapy. Dimension extraction and categorization are both components of the two-level CNN framework. The article extraction process extracts characteristics such as age and sexual identity, while the classification technique assigns the play photographs to the appropriate age and gender groups. We propose a ground-breaking end-to-end CNN swing in this implementation to achieve better and healthier age units and sexuality categorization of unfiltered real-world faces. We use a bulky person pretreatment approach to prepare and process the unfiltered real-world faces before they are input into the CNN poser in order to handle the significant discrepancies in those faces. When tested for sorting accuracy on the synoptical OIUAudience benchmark, an experimental result reveals that with us assistance achieves state-of-the-art achievement in both age gathering and gender arrangement. Our web is pretrained on an IMDb-WIKI with chanting labels, then fine-tuned on MORPH-II, and eventually on the OIUAudience (first) dataset's training set. In comparison to the best-reported results, the classification of age groups is improved by an excellent percentage (exact accuracy) and a very high percentage (validation accuracy), while the classification of genders is improved by an excellent percentage (exact correctness) and 93.42 percent (validation accuracy).

Keywords: Convolutional Neural Networks (CNNs), IMDb-WIKI, MORPH-II, OIU-Audience

1. Introduction

Human facial image analysis is a fundamental and difficult problem in computer vision. There are many possible applications for automated facial recognition, including security, marketing, animation, human-computer interaction, and surveillance. However, due to variations in visual perspective, facial look, and sophisticated facial expressions, it is difficult to identify the same person in different photos. This is where facial recognition technology comes in. Face recognition technology is a computer application that is used to identify or verify the identity of a person

in a digital image or video.

There are a few advantages to age and gender classification. One is that it can be used to help target marketing efforts. For example, a company might send different advertisements to men and women, or different age groups. Another advantage is that it can be used to help determine how much money to allocate to different. There are some limitations to facial recognition when it comes to age and gender classification [1]. For example, certain facial features may be more indicative of a certain age or gender than others, and these features may not be as easily identifiable in all individuals. Additionally, the accuracy of facial recognition software may vary depending on the quality of the images used for recognition. The authors provided a good approach for evaluating face pictures from start to finish using semantic face division. The maker's development combines several stack pieces for face recognition, including head present measurement, age confirmation, and direction gathering.

A genuinely stepped dataset is prepared using an unanticipated sporadic field (CRFs) based district, model. A CRF-based multi-class face division structure divides a facial image into six segments. The probabilistic gathering technique is used for each class, and probability maps are delivered. For each challenge, probability maps are employed as feature descriptors, and an RDF classifier is displayed [2]. The algorithm summarizes our recommended estimation. A division model is initially spread out among the CRFs. The proposed model MSF-CRFs yields the most plausible class for each super-pixel when it comes to face division. Each pixel within the super-pixel is subsequently assigned a comparable imprint. Age and direction makers employ the likelihood maps supplied during class division to show head posture. Probability maps were conveyed for each class. For every task (head stance, age, and direction) Author trains an RDF classifier with a component vector of the related probability maps. Feature descriptors are used by the probability maps.

Computational cost is an essential issue of MSF. To diminish the computational rate, the Author used a model (MSF-CRFs) which has a super-pixel-based investigation. In several instances, the taking care of period of the division was updated with The care of the division period was worked on several times using MSF-CRFs in conjunction with the MSF [3]. Makers included SEED calculation for super-pixel division in the proposed approach. Makers prefer SEEDS to SLIC and other methods because the SEED's speed is considerably superior to the numerous SOA strategies. SEED also provides far better superpixel division than traditional mix-up estimations. A probabilistic request technique is employed to create probability maps for each face class. Unpredictable Results Through the use of different probability maps, the forest classifier is prepared for any task (head stance, age, and direction.

The author recommends that the front view be used to extract the numerical and appearance components of a face. The all-out benchmark approach is utilized to isolate the component. Two major orders, guided and solo techniques, can be used for direction gathering. The author employed a coordinated AI method in this paper. For this concept, Maker employed three arranged classifiers: SVM, neural association, and adobos. The writer has prepared every one of the classifiers through an indistinguishable preparation dataset and comparable elements. creators chose to utilize a mix of mathematical and appearance highlights simultaneously. The distinguished issue of orientation acknowledgment needs consideration on strong element extraction as well as a production technique. Here, the accomplishment of the answer for this issue urgently relies upon the steadiness and versatility of the facial highlights, which utilize the qualities of elements vector. Likewise, creators inspected different designs and shading-based elements for recognizable proof of human face orientation. The authors of this work had a collection of images and used a dataset designed for the face recognition IITK right away.

Additionally, the author used the R language Rattle to examine the Local Invariant Feature-Based Gender Recognition facial parts detected from a selected image by utilizing skin shading division and morphological activities [4]. The extricated facial parts were also subjected to a proposed calculation. Creators encode the ostensible amount of 2-D spatial data in the record by removing qualities like surface, shape, and force appropriation and incorporating vectors with other crossover image highlights to further expand the separation exactness of shape ordering procedures [5]. This document shows how to extract elements based on their shape and shading In this exploration work creator proposed framework, spatially improved neighborhood double example (SLBP), and histogram of situated angles (HOG) are extricated to arrange the human orientation with the SVM classifier [6].

This half-and-half component choice has expanded the force of the proposed framework because of its portrayal of surface miniature examples and nearby shapes by catching the edge or inclination structure of the picture. The orientation arrangement precision is examined by utilizing the nearby element portrayal of the face pictures independently and these elements are connected to give a superior acknowledgment rate. The blend of two distinct nearby descriptors gives a great portrayal of the face picture and this is given to the SVM classifier which arranges as male or female. Likewise, the proposed work is contrasted and other two customary classifiers, for example, k-closest neighbor and meager portrayal classifier. The presentation was assessed on the FERET and LFW information base. to recognize the orientation of an individual as one or the other male or female from facial signals of the caught picture. The general stream graph of the proposed strategy is displayed in Fig. 1.

Figure 1: Audience Image Dataset Examples. Top Row: 6 Males of Various Ages and Bottom Row: 6 Females of Various Ages

This proposed technique involves four significant advances in particular face discovery, and preprocessing, including extraction and characterization. Face identification is the initial step to pick the area of interest to eliminate undesirable parts like neck, hand, and environmental elements. This cycle is refined through Viola-Jones calculations utilizing Harr-like highlights and AdaBoost. After the fruitful location of the face area, preprocessing like shading transformation and picture resizing is performed. Spatially upgraded nearby paired examples (SLBP) and HOG highlights are removed from the preprocessed picture. These highlights are melded to get half-and-half element vectors for characterization. In the preparing stage, the elements from preparing pictures are removed and the characterization model is worked with the assistance of their actual class marks. At the point when the question picture is taken care of as a contribution to the orientation arrangement framework, the SLBP and HOG highlights are separated and recognized as one or the other male or female.

Likewise, three unique classifiers are executed, and their presentation is contrasted with selecting the appropriate classifier. The proposed strategy is tried on standard controlled FERET and uncontrolled LFW information bases. The exhibition is investigated with every descriptor and their mixture mix with three unique classifiers However, when the scale and complexity of the convolutional neural network (CNN) becomes large and complex, the model parameters are too many, the scale is large, the training data set is large, and the training time is long, which becomes inconvenient to deploy on the mobile side, there are some shortcomings [7]. The result is a lighter network, but the age estimates must also be accurate. In view of the above difficulties, this study proposes an age estimation model based on ShuffleNet V2. Experiments have shown that the proposed network converges quickly during the training phase, is more accurate during the age estimation phase, and the size of the network model is smaller.

Since these algorithms are measured from a distance, facial recognition systems for gender recognition are invariant to many types of lighting and positional aberrations. Crack detection, noise detection, and sharpness are examples of other work in this area. Real-time moments, curvature, shape features, integral persistence, and symbolic representations are all examples of how firms can be represented. Region- and contour-based 2D shape maps and region- and contour-based feature vectors represent categories, respectively. In the path-based method, the shape

pixels in question are ignored in favor of path information, but pixels play an important role in creating shape descriptors in the area-based method. In recent years, there has been increasing interest in research on the automatic classification of invisible data, especially in the areas of object recognition, surveillance, medical photography (including image enhancement), and financial data analysis. The classification of invisible data is based on patterns found in a particular space. This approach was inspired by "deep learning," the latest version of an old computational technology called neural networks. Roughly inspired by the tightly connected neurons of the brain, these thriving systems mimic human learning by empirically changing the strength of simulated neural connections.

Human gender classification has been associated with various automatic algorithms since the advent of social media and the web. However, for advanced face recognition processes, the display and performance of algorithms that can be used with real original algorithms are not yet up to par. Deep neural networks can improve the accuracy of various applications when running on smartphone processors, but they cannot meet the needs of end-users due to their rigorous computational requirements [8]. However, this paper proposes a method that takes memory and battery limitations into account. Server architecture that implements the Deepgender model. This article's contributions are summarized as follows: Building a Lightweight Convolutional Neural Network Age Estimation Model Using Hybrid Attention Methods. The age estimation method combines classification and regression, which is easy to use and has high accuracy. Face detection and correction is to enhance the age-related feature information of faces by inputting face images and image enlargement. This is useful for network training.

2. Literature Review

The authors claim that age is a soft biometric that helps law enforcement agencies identify victims of the creation/distribution of child sexual exploitation material (CSEM). Accurately estimating the age of the subject can classify ownership of open content as illegal during the investigation. This automation of age classification speeds content discovery and allows you to prioritize evidence, including CSEM, and focus your investigation on digital evidence. The authors propose a deep convolutional neural network (DCNN) based on a 50-layer residual neural network (ResNet50) to estimate facial age in juveniles. Our proposed method is pre-

trained on the ImageNet dataset.5 The final fully connected (FC) Softmax layer with 1000 outputs has been replaced with an FC Softmax activation function layer with 20 outputs corresponding to the studied age category (1-20 years old) according to your needs. The convolutional layer parameters are then frozen during training. The ResNet50 architecture is used to estimate the age of the face and replace the last layer with 20 outputs.

Since the age estimation problem is treated as a classification problem, the category cross-entropy log loss function is used [9]. During the creation of the DeepUage model, an artistic method of processing DCA facial proportions was developed. Its preprocessing efficiency has been compared to that of other preprocessing approaches. It achieved an MAE of 2.73 years, which was the best of all the techniques tested. In comparison to previous facial age estimators for underage people, the DeepUAge model's accuracy in both validation and testing is superior to other techniques. Another author proposes extracting orientation-based components from human facial images using computerized expectations that are old enough. The authors demonstrate that learning representations with the use of profound convolutional neural networks improve these tasks significantly (CNN).

The feedforward neural organization technique used in this study improves the power of deep factor unconstrained acknowledgment projects that recognize orientation and age groups [10]. On both the Essex face dataset and the Audience benchmark, this examination was investigated and accepted for the orientation expectation and age evaluation. Convolutional neural organizations are coordinated in an arrangement of rotating two sorts of layers S-layers and C-layers, called convolution and sub-testing with at least one completely associated layer (FC) in the closures. Convolutional Neural Networks are an extraordinary sort of multi-faceted neural organization, they follow the way of its ancestor noncognition in its shape, design, and learning reasoning.

Generally, neural organizations convert input information into a one-layered vector, and they can perform both element extraction and arrangement. The info layer gets standardized pictures with similar sizes. Each info picture will go through a progression of convolution layers with channels (Kernels), Pooling, and completely associated layers (FC), and apply enactment capacity to characterize an item with probabilistic qualities. In the result layer, the initiation work, as a rule, relies upon the errand. For instance, on account of double or various twofold arrangements, the sigmoid actuation is utilized (which is ensured to have values in [0, 1]). For multi-way arrangement, the softmax actuation work is utilized to produce a worth between 0- 1 for every hub. authors assessed the use of profound neural organization for human age and orientation forecast utilizing CNN.

During this concentrate on the different plans produced for this assignment, age, and orientation order is one of the critical portions of exploration in the biometric as friendly applications with the objective of future conjecture and data divulgence about the

specific individual should be conceivable satisfactorily [11]. The method uses assembled approaches and computations by which profound learning is similarly the prime in use plans in the CNN for age and orientation order. A convolutional neural organization pipeline for programmed age and orientation acknowledgment has been proposed, The authors give a broad informational collection and benchmark for the investigation of old enough assessment and orientation order. In the research paper, the authors recommend another approach for managing assistance for the client's direction and age range, which is precisely reflected in his photo. Similarly, adopting the Deep Learning approach, adding a twofold check layer validator by linking between client photo, direction, and date of birth structure inputs, by identifying the direction and calculating the age from a lone person's photo utilizing a Convolutional Neural Network (CNN or ConvNets).

The main goal of this research is to provide a design endorsement work that can connect directly, DOB, and client photo inputs and support all of them while considering their relationship. The direction entered by the customer, for example, will be compared to the perceived direction from the client photo. Furthermore, the DOB submitted by the client will be authorized, as will the real age range has shown by the client's photos, and so on. Age, orientation, and the date of individual images have recently become crucial data for a long time and state-run administrations for commercial, ID, security, and other purposes. The design validators were proposed to reduce the client data area botches because this information was obtained from people through the effort framework. Makers seek to propose another response for supporting these data in this research by estimating age and direction from a single individual photo and distinguishing it from entered direction and date-of-birth.

The proposed method used three basic obstruction strategies: blackout, self-assertive splendor, and dark, each of which simulates a different type of challenge that would be familiar with real-world applications. A modified cross-entropy disaster is also presented, which lends less discipline to the age assumptions that land on the close-by classes of the ground truth class. The viability of our proposed technique is investigated by applying increment methodology and modified cross-entropy to two different convolution neural connections, the imperceptibly changed AudienceNet, and the significantly changed VGG16, to execute the age and direction course of action [12]. The recommended development structure addresses the Audience Net network's age and direction portrayal precision by 6.62 percent and 6.53 percent on the Audience illuminating assortment, respectively [13]. On the Audience enlightening record, the proposed increment system reduces the age and direction gathering accuracy of the relatively altered VGG16 network by 6.20 percent and 6.31 percent, respectively.

The suggested methodology is motivated by a desire to address the theory and power of convolutional neural networks by expanding facial images to replicate real life and unconstrained settings. These disorders are accompanied by obscured, obstructed, or absent

facial features. Wearing shades, for example, blocks the eyes and a portion of the nose, a facial shroud restricts the lips and nose, and over-the-top lighting causes components to evaporate quickly. To create real-life and unrestricted conditions, makers create hurdles on the three guiding aspects of human appearances, eyes, nose, and mouth. Face distinguishing proof is conducted on the data facial pictures to observe these components, and a face plan is also performed to chip away at the concept of data pictures. Makers hoped that by developing data, they would be able to lessen the issue of overfitting, which can occur when the arrangement data is limited. Makers planned to give less discipline to one-off age figures when cross-entropy adversity changed. The proposed data increment methodology with feature hindrance expands the data by simulating problems that might occur in real life.

Given an information facial image, the proposed information advancement with consolidated impediment will first discretionarily choose a facial part locality and then do one of the three-block philosophies erratically [14]. The obstruction methods mix selectable brilliance, power loss, and dim, which rehashes where the countenances are discouraged and low-level-headed under unfathomable lighting settings. The author can also conduct further arranging tests with the facial part regions that are being obliterated, enhancing the strength and hypothesis of the affiliation. The proposed altered cross-entropy difficulty improves the learning relationship by imposing less discipline on the all-out occurrence when incorrect age checks land in the ground truth age class's adjacent classes. The author proposes and shows that pre-prepared CNNs which were prepared on huge benchmarks for various purposes can be retrained and finetuned for age range arrangement from unconstrained face pictures. Additionally, the creator proposes to decrease the element of the result of the last convolutional layer in pre-prepared CNNs to work on the presentation of the planned CNNs models.

The author utilized a data set that has unconstrained face pictures mirroring the varieties of genuine subject pictures taken from the web vaults. The primary objective of the work is to invest entryway of the work of CNNs, which were recently prepared for contras tent errands on huge information bases, in the plan of a DNN for age range arrangement task. We have proposed and assessed a few CNN designs. A profound pre-prepared CNN model for face acknowledgment is utilized to extricate facial highlights. Then, at that point, these separated facial highlights are utilized to prepare a DNN for age range arrangement [11]. What's more, dimensionality decrease is performed on the last convolutional layer elements of already pre-prepared CNN models on various assignments other than age range grouping. The correspondingly diminished elements are then fused and prepared to assess the age by utilizing DNN engineering. The author gives broad test results in his research work, which exhibit the capacity of our proposed work to arrange the age from the facial pictures. The proposed work altogether outflanked detail-of-the-craftsmanship results by roughly 12% on the Audience benchmark.

Sr.	Year	Title	Methodology
1	2020	"Improving Underage Age Estimation Accuracy to Aid CSEM Investigation"	"Deep convolutional neural network (DCNN)"
2	2019	"Gender Classification in Human Face Images for Smart Phone Applications Based on Local Texture Information and Evaluated Kullback-Leibler Divergence"	"Improved local binary patterns (ILBP)".
3	2021	"CNN Based Features Extraction for Age Estimation and Gender Classification"	"Convolutional neural network(CNN)"
4	2019	"Exploiting Effective Facial Patches for Robust Gender Recognition"	"Convolutional neural network(CNN)"
5	2021	"Data augmentation with occluded facial features for age and gender estimation"	"Feature occlusion(VGG16) "
6	2020	"Age and Gender Prediction and Validation Through Single User Images Using CNN"	"DNN model"
7	2019	"Utilizing CNNs and transfer learning of pre-trained models for age range classification from the unconstrained face image"	"Convolutional neural network(CNN)"
8	2019	"Age estimation in facial images through transfer learning"	"local binary patterns (LBP) histogram of oriented gradients (HOG), binarized statistical image features (BSIF)"
9	2019	"A hybrid technique for gender classification with SLBP and HOG features"	"Spatially enhanced local binary pattern (SLBP) and HOG"

10	2019	"Deepgender: real-time gender classification using deep learning for smartphones"	"DNN"
11	2019	"Local Invariant Feature-Based Gender Recognition from Facial Images"	"SVM classifier"
12	2019	"Deep gender classification based on AdaBoost-based fusion of isolated facial features and foggy faces"	"CNN"
13	2021	"Face Attribute Dataset for Balanced Race, Gender, and Age"	"ADAM optimization"
14	2020	"Face Image Age Estimation Based on Data Augmentation and Lightweight Convolutional Neural Network"	"CNN"
15	2019	"A Unified Framework for Head Pose, Age and Gender Classification through End-to-End Face Segmentation"	"MSF-CRFs model"

Table 1: Using techniques in Others Research

3. Methodology

Our proposed methodology uses transfer learning for the classification of age and gender through facial images. Firstly, the method preprocesses an image using a face detector to crop a face image, and then a trained network is used to classify face images using transfer learning. For transfer, and learning purposes benchmarked trained network VGG 19 is used. VGG 19 includes different layers to process faces some are used to get features of face units and some are used for classification purposes. Section 3.1 illustrates benchmarked face image databases used to test the proposed method.

3.1 Datasets

There are various publicly lendable datasets worthy for age and gender assessment. IMDB-WIKI, UTK, and FG-Net Face datasets are used in this method to train the model The most frequently available dataset with age and gender labels is IMDB-WIKI. It was designed for the evaluation of Rothe et al DEX.'s (Deep Expectation) approach. The information for each image included the date it was taken (Dalal & Triggs, 2005). They snuck in 20,284 celebrities and obtained 460,723 tagged photos, with male faces accounting for 57% of the total. There are 62,328 photos in the Wikipedia subset, with 75 percent of them being male faces. The IMDB-WIKI collection contains a total of 523,051 photos

Figure 2: Some Samples in the IMDB and Wiki Datasets

The UTKFace collection contains 24,107 faces ranging in age from 0 to 116, with male subjects accounting for 52 percent of the photos. Each photograph is labeled with information such as age, gender, and ethnicity (Galla et al., 2019). The collection also includes 68 landmarks for each face The F G - N E T (Grappling and Communicate Recognition) dataset is a face-based approach senescence dataset. Human determination limit and age are noted on the photos. The dataset also includes 68 carefully labeled face landmarks for each presentation. At various ages, each person has sevenfold images. The sexuality hold is not included in the dataset (Gunay & Nabiyev, 2007).

3.2 Preprocessing

The techniques used for the detection are an open-source package called Dlib l to construct a classifier based on a histogram of directed gradients and The Multitask Cascaded Convolutional Neural Network (MTCNN) is used in the second approach. This technique is recommended since it is quicker and more accurate than HOG. Each face in a picture that the detector identifies is enlarged to the appropriate image size. The ideal size for our models is 64 by 64 pixels. In the original picture, the detector provides enlarged faces as well as the locations of their bounding boxes
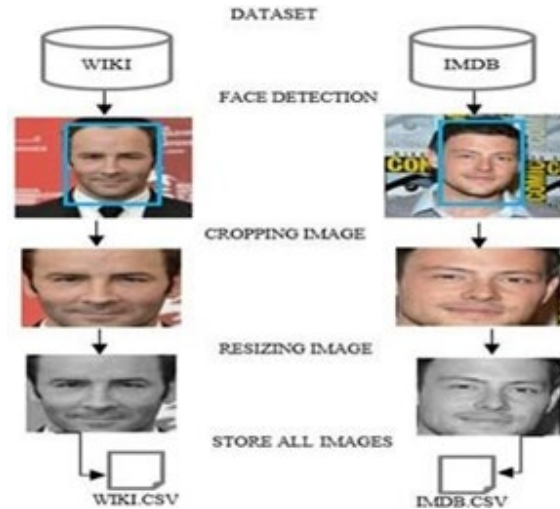
Figure 3: Preprocessing Data

A Face Alignment Network (FAN) is used for the detection of face landmarks. The facial landmark detector is used once the faces in an image have been detected. 68 facial landmarks, such as the shape of the eyes, nose, mouth, eyebrows, and jawline, are detected by the facial landmark detector. The center of each eye is calculated using the points on the contour of the eyes. After that, we drew a straight line between the nose's points. The distance between each eye and the nasal line is then calculated. Finally, we calculate the distance ratio. If the ratio is below a specific threshold, the face can be sent to an age and gender predictor.

3.3 Deep Learning

It is the category of machine learning, which is based on artificial neural networks, multiple layers of processing are used to extract progressively higher-level features from data. In transfer learning, a machine uses previous task knowledge to increase generalization about a new task. In our research, we used the VGG-19 deep convolutional neural network. Here's a quick rundown of the architecture.
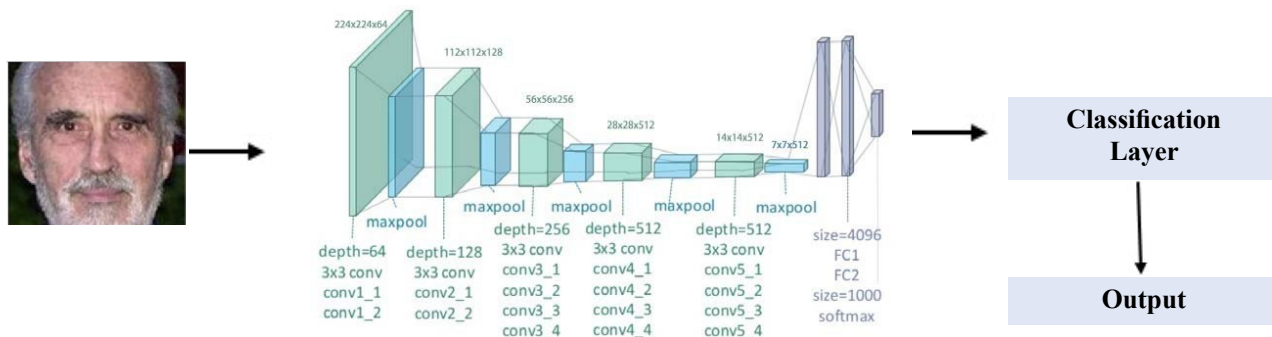

Figure 4: Methodology

3.4 VGG -19

VGG-19 is a hybrid of VGG-16 and VGG-16, with a few peak tweaks. It was created by the same group that created VGG-16 to improve ImageNet disagreement error evaluation. There are 19 levels in total. 128 strain convolutional bed x 2, 256 filters convolutional sheet x 4, 512 convolutional locations x 8, entire engaged bed x 2, and 1 output sheet. For ImageNet gainsay, it originally had 1000 product nodes, but we lowered the number to two to match our demands (Lee et al., 2019). A 0.5 dropout was employed with two completely neighboring layers.

3.5 Experimental Setup

The assembling set was arbitrarily divided into two sections: activity (80%) and investigation (20%). The dataset is split into these two pieces to ensure that the same patient who was used for the activity is not used for testing again. This ensures that the classifier isn't biased against specimens that haven't been seen. The CNN leader was trained on an NVIDIA Discoverer K80 (24 GB of GDDR5 memory among them, 11 GB GDDR5 storage each occurrence), Intel Xeon® CPU @2.3 GHz, RAM 14 GB per instance, and Keras API on top of TensorFlow (CUDA toolkit 9.0, cuDNN SDK v7, and Python 3.6). The upbringing

takes approximately 13 seconds per period, and it converges in only 20 epochs. Using mean absolute error as the loss function, we trained each combination of optimizer and learning rate for 60 epochs. During the training, we kept track of both the training and validation losses. The training was placed on Meta centrum VO l's infrastructure, which included four CPUs, 30GB of RAM, and one CPU. One combination required roughly 6 hours to train.

4. Results

We used a learning rate of 0.1 initially, however, the model diverged for each optimizer. The results in this paper are based on accuracy and the f1-score, which are described by the equations below:

The effectiveness of a classification algorithm where output seems to be a probabilistic number between 0 is measured by cross-entropy loss, also known as log loss. As the anticipated probability departs from the labeled probability, cross-entropy loss grows. As a result, making a prediction with a probability of.012 when the observation label is really 1 would be poor and have a high loss value. A log loss of 0 would indicate a flawless model.
$$C.E = -\sum c=1 M y o, c \log(p o, c)$$

4.1 Age Network

Adam optimizer's best learning rate was 0.001, resulting in a mean absolute error of 6.48. The best learning rate for the RMSprop optimizer is 0.001, with a mean absolute error of 6.65. With a learning rate of 0.01 and a stochastic gradient descent optimizer, we attained a mean absolute error of 6.88. Table 2 compares validation losses for each combination after 60 epochs. Based on validation loss, the Adam optimizer with a learning rate of 0.001 was the best optimizer, learning rate combination. On the test dataset, we tested this model. We tested the RMSprop optimizer with a learning rate of 0.001 and the gradient descent optimizer with a learning rate of 0.01 on the test dataset since they produced similar results. Table 2 shows the evaluation findings for the validation and test subgroups. Based on both findings, Adam with a 0.001 learning rate is the optimum combination of optimizer and learning rate.

Optimizer	Learning Rate	Val Loss	Test Loss
Adam	0.001	6.48	7.08
RMSprop SGD	0.001	6.65	7.23
	0.01	6.88	7.41

Table 2: Test Loss for Best Combinations for Age Network Training

4.2 Gender Network

We trained the gender estimation networks similarly to age estimation networks. The best learning rate for the Adam optimizer is 0.001 achieving 0.09 loss in validation The greatest result we got with the RMS prop optimizer was a validation loss of 0.09 for a learning rate of 0.001. Finally, we achieved a validation loss of 0.1 with the gradient descent optimizer with a learning rate of 0.01. It's worth noting that the learning rate of 0.0001 was far too low, and the model had a validation loss of 0.47. The validation loss after 60 epochs for each combination is shown in Table 5.5. RMSprop with a learning rate of 0.001 was the best optimizer and learning rate combo. However, Adam optimizers with learning rates of 0.001 and 0.0001, RMSprop optimizers with learning rates of 0.0001, and gradient descent optimizers with learning rates of 0.0001.

Optimizer	Learning Rate	Val Loss	Test Loss
Adam	0.001	0.09	0.10
Adam RMSprop	0.0001	0.10	0.12
RMSprop SCD	0.001	0.08	0.10
	0.0001	0.10	0.12
	0.01	0.10	0.12

Table 3: Test Loss for Best Combinations for Gender Network Training

The best-performing combination of optimizer and learning rate based on the evaluation of validation and test datasets was RMSprop with a learning rate of 0.001. The results of all evaluated combinations are displayed
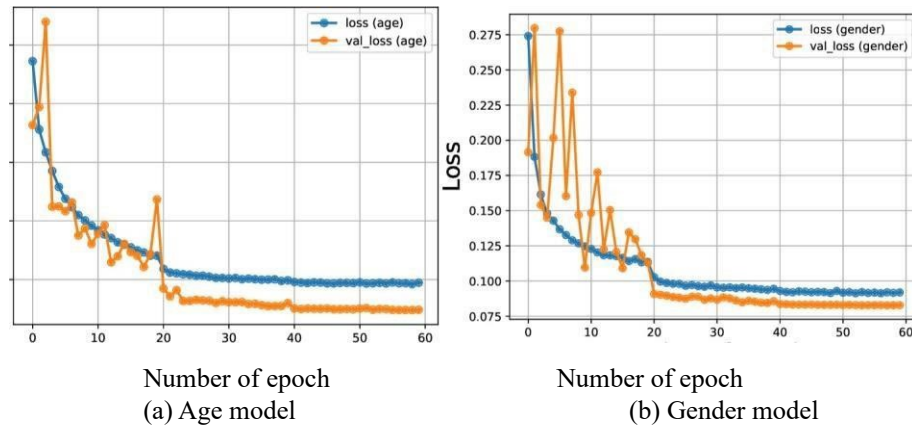
<div align="center">

Number of epoch
(a) Age model
      Number of epoch
(b) Gender model

Figure 5: Graphs of Training and Validation Loss for Best Age and Gender Models

</div>

4.3 Evaluation of Age Prediction

We utilized mean absolute error to train and evaluate the age prediction model (MAE). We obtained an MAE of 5.98 while evaluating the full dataset. Following that, we calculated the inaccuracy for each age label. Table 8 displays the results. we can see that the best MAE, 3.96, is obtained when the age label is 45. The oldest group, 40 years old, had an MAE of 4.73, which is excellent. This indicates that in the combined training dataset, the best MAE was attained at or near the apex of the age distribution. It's worth noting that the worst MAE of 10.63 corresponds to the age of 50. To clarify this, we delve further into the MAE on a gender-by-gender basis. The results displayed in Table 9 show that for males the MAE is like age labels 55 and 60. However, for females, the MAE is much higher but with fewer examples. There might be a few different reasons for this. From human annotator error to a particularly difficult set of images for the model.

4.4 Evaluation of Gender Prediction

We have additional variables to evaluate the success of our trained model because gender prediction is a binary classification. Precision and recall are two helpful measures. Precision refers to a model's ability to avoid labeling a positive example as a negative. The number of real positive instances is divided by the total number of positive examples to arrive at this figure. Recall, on the other hand, is a model's capacity to locate all positive cases. It is determined by dividing the number of genuine positives by the number of true negatives.

5. Conclusion

We developed a program that predicts age and gender There are two methods for detecting faces in the application. The first uses a histogram of directed gradients, while the second employs a cascade of convolutional networks. In terms of age estimate, the trained model outperformed the untrained model on all datasets. The FG-NET dataset had the greatest substantial reduction, with mean absolute error dropping from 12.85 to 6.61. On most datasets, the model outperformed the others in terms of gender estimation. On the UTKFace dataset, the error rate dropped from 20.45% to 10.6%. We got a 5.66 age mean absolute error. The mean absolute

error might be as low as 3.96 in rare situations. All male individuals are accurately classified by the model. In a real-world scenario, we proved that our system could estimate age and gender.

References

1. Annalakshmi, M., Roomi, S. M. M., & Naveedh, A. S. (2019). A hybrid technique for gender classification with SLBP and HOG features. Cluster Computing, 22, 11-20.
2. Khan, K., Attique, M., Syed, I., Sarwar, G., Irfan, M. A., & Khan, R. U. (2019). A unified framework for head pose, age and gender classification through end-to-end face segmentation. Entropy, 21(7), 647.
3. Afifi, M., & Abdelhamed, A. (2019). Afif4: Deep gender classification based on adaboost-based fusion of isolated facial features and foggy faces. Journal of Visual Communication and Image Representation, 62, 77-86.
4. Abu Nada, A. M., Alajrami, E., Al-Saqqa, A. A., & Abu-Naser, S. S. (2020). Age and gender prediction and validation through single user images using cnn.
5. Akram, A., Ramzan, S., Rasool, A., Jaffar, A., Furqan, U., & Javed, W. (2022). Image splicing detection using discriminative robust local binary pattern and support vector machine. World Journal of Engineering, 19(4), 459-466.
6. Benkaddour, M. K. (2021). CNN based features extraction for age estimation and gender classification. Informatica, 45(5).
7. Mallouh, A. A., Qawaqneh, Z., & Barkana, B. D. (2019). Utilizing CNNs and transfer learning of pre-trained models for age range classification from unconstrained face images. Image and Vision Computing, 88, 41-51.
8. Fekri-Ershad, S. (2019). Gender classification in human face images for smart phone applications based on local texture information and evaluated Kullback-Leibler divergence. Traitement du Signal, 36(6), 507-514.
9. Verma, V. K., Srivastava, S., Jain, T., & Jain, A. (2019). Local invariant feature-based gender recognition from facial images. In Soft Computing for Problem Solving: SocProS 2017, Volume 2 (pp. 869-878). Springer Singapore.
10. Kärkkäinen, K., & Joo, J. (2019). Fairface: Face attribute dataset for balanced race, gender, and age. arXiv preprint

arXiv:1908.04913.

11. Anda, F., Le-Khac, N. A., & Scanlon, M. (2020). DeepUAge: improving underage age estimation accuracy to aid CSEM investigation. Forensic Science International: Digital Investigation, 32, 300921.

12. Cheng, J., Li, Y., Wang, J., Yu, L., & Wang, S. (2019). Exploiting effective facial patches for robust gender recognition. Tsinghua Science and Technology, 24(3), 333-345.

13. Liu, X., Zou, Y., Kuang, H., & Ma, X. (2020). Face image age estimation based on data augmentation and lightweight convolutional neural network. Symmetry, 12(1), 146.

14. Dornaika, F., Arganda-Carreras, I., & Belver, C. (2019). Age estimation in facial images through transfer learning. Machine Vision and Applications, 30, 177-187.