Tech Science Press

Semantic Human Face Analysis for Multi-level Age Estimation

Rawan Sulaiman Howyan[1,2,*] and Emad Sami Jaha[1]

[1]Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
[2]Department of Computer Science, Faculty of Computers and Information Technology, University of Tabuk, Tabuk, Saudi Arabia
*Corresponding Author: Rawan Sulaiman Howyan. Email: rhowian@stu.kau.edu.sa or rhowian@ut.edu.sa
Received: 16 April 2021; Accepted: 05 June 2021

Abstract: Human face is one of the most widely used biometrics based on computer-vision to derive various useful information such as gender, ethnicity, age, and even identity. Facial age estimation has received great attention during the last decades because of its influence in many applications, like face recognition and verification, which may be affected by aging changes and signs which appear on human face along with age progression. Thus, it becomes a prominent challenge for many researchers. One of the most influential factors on age estimation is the type of features used in the model training process. Computer-vision is characterized by its superior ability to extract traditional facial features such as shape, size, texture, and deep features. However, it is still difficult for computers to extract and deal with semantic features inferred by human-vision. Therefore, we need somehow to bridge the semantic gap between machines and humans to enable utilization of the human brain capabilities of perceiving and processing visual information in semantic space. Our research aims to exploit human-vision in semantic facial feature extraction and fusion with traditional computer-vision features to obtain integrated and more informative features as an initial study paving the way to further augment the outperforming state-of-the-art age estimation models. A hierarchical automatic age estimation is achieved upon two consecutive stages: classification to predict (high-level) age group, followed by regression to estimate (low-level) exact age. The results showed noticeable performance improvements, when fusing semantic-based features with traditional vision-based features, surpassing the performance of traditional features alone.

Keywords: Semantic features; semantic face analysis; feature-level fusion; computer-vision features; hierarchical age estimation; age group classification; facial aging

1 Introduction

Humans have a great inherent ability to perceive visual signals around them in the environment using their eyes and beyond to analyze those signals by their brains to make some inferred decisions or actions based on their background, knowledge, and experience [1]. Nowadays, the continuing advances in

computer science along with artificial intelligence technologies make it possible to imitate some human capabilities in such visual perception. Numerous techniques based on machine learning and computer vision have been proposed and elaborated aiming to successfully simulate desirable human behaviors and capabilities in receiving and processing visualized data and then making appropriate decisions or reactions.

In recent years, analyzing information that can be derived from human face has attracted great interest in several areas associated with pattern recognition, computer vision and image processing approaches. There is often further implicit and semantic information which can be deduced from a human face image such as gender, age, identity, race, and sentiment expressions. The problem of estimating the age of human *via* a face image is deemed as a very active research issue in the last decades, due to the need to provide and develop the property or ability to determine the age of individuals from their faces (see Fig. 1) in many real-world applications, for example forensics, crime detection and prevention, identification of persons missing for several years, surveillance systems for suspects discovery, facial recognition and verification, access control for web content based on age, age simulation, cosmetics surgery, biometrics, and many other fields [2,3].

Figure 1: Albert Einstein's photographs in different ages

1.1 Facial Aging

Facial age estimation can be defined as the process of assigning an age group or a specific age to a person's face image based on some appearance features [3]. With age progression, there may be many changes that occur in face appearance alongside aging progression. The characteristics of these changes can vary between people in their childhood and adulthood stages. During childhood stage, the most prominent changes are those which appear in the face and visibly affect its shape and size and its components, which occur because of the rapid growth of bones and skull, such as the chin becomes more salient as the cheeks extend. On the other hand, the changes after adulthood are highly associated with the soft tissue, skin texture, fine lines, wrinkles, sagging, skin elasticity loosing, and many other skin-related details, in addition to minor changes in face shape [4]. Many intrinsic and extrinsic factors may affect facial aging or appearance and make the person's face looks younger or older, such as genes, muscle tone, collagen ratio, sunlight, lifestyle, make-up, nutrition, cosmetic operations, weather conditions, and smoking [5]. Moreover, many research explorations have shown that facial aging may differ by population and learning facial age jointly with expression, ethnicity, and/or gender is more difficult than learning facial age disregarding these considerations [6]. All these factors and more make the age estimation process more difficult and complex. Thus, there is still an urgent need to design and develop robust algorithms integrating their strength capabilities to better confront such challenging factors and help determine the facial age as accurately as possible. Figs. 2 and 3 show the usual changes occurring to young and old people with age progression.
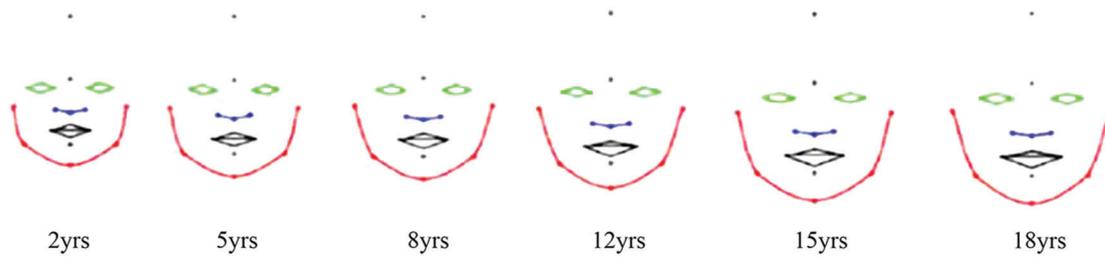
Figure 2: Facial growth in children and young people with age progression [7]
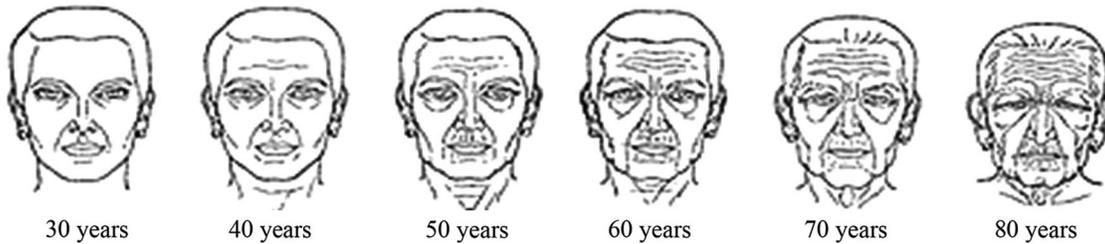
Figure 3: Sketch of changes occurring in the adult face with age progression [8]

1.2 Traditional Computer Vision-based Age Estimation

Several different computer vision methods have been investigated and used for facial age estimation implicating either extracting hand-crafted features or learning deep features *via* Conventional Neural Network (CNN/ConvNet) architectures. The anthropometric model is focused on some measurements between the main facial components. A set of landmark points is identified on the eyes, lips, nose, eyebrows, ears, chin, and forehead. Then, a number of measurements, like axial distance, the angle between components, shortest distance, tangential distance, angle of inclination, and ratios of distances can be calculated among face components [9]. In [10,11] the authors relied on anthropometrical measurements in the process of age group classification using Artificial Neural Networks (ANNs), while in Thukral et al. [12], a hierarchical approach was applied in the process of human age estimation using 2D landmarks as shape features to train Support Vector Machine (SVM) and Support Vector Regression (SVR).

Active Shape Model (ASM) is a statistical model used to obtain descriptive information about face shape *via* a collection of facial landmark points [13]. In the age estimation process, other models are often combined with ASM in order to obtain more accurate description for a human face. The method used in [14], fuses ASM, interior angle formulation, anthropometric model, carnio-facial development, heat maps, and wrinkles together to extract aging features. Then, a CNN model is used to classify images among different age groups. Another enhanced prediction model was developed in [15] by combining both texture and shape features, and applying Partial Least-Square (PLS) for dimensionality reduction.

Age determination can be successfully performed using a poly regression model. Active Appearance Model (AAM) is used to extract features in many research explorations related to face images and was exploited by different researchers like Shejul et al. [16] to extract facial features for human age estimation task. It is worth noting that AAM is a statistical model distinguished from the ASM, such that it combines two different models for modeling shape and gray-level appearance. A wrinkle model was combined with AAM in [17] in order to enhance the results of basic AAM, and a CNN regressor was trained for exact age prediction. In [18], AAM, Local Binary Patterns (LBP), and Gabor Wavelets (GW)

algorithms were fused together for features extraction and a novel hierarchical approach was implemented with two consecutive regressors. A skin spots feature was added in [19] using Local Phase Quantization (LPQ) algorithm. However, this feature gave a slightly negative effect on the results due to the variation of lighting among images in the used datasets.

AGing pattErn Subspace (AGES) was proposed by Geng et al. to be used in automatic age estimation [20]. They defined the aging pattern as a sequence of facial images of a specific person, and those images must be ordered according to the age series. Hence, rather than using independent face images, the AGES approach considered each aging pattern as one sample. Also in AGES, each image is in one specific location on the axis t which refers to the time. If a person's picture is available at a certain age, then it is placed in the appropriate place in the aging pattern. Otherwise, if the image is not available, its location remains empty. Each image is transformed using AAM into a vector of features, then all vectors are combined in one long vector with an empty value for each missing image. For unavailable images, the Principal Component Analysis (PCA) can be performed frequently to reconstruct the vectors of these lost images. In [21], a three-level hierarchical approach was performed using SVM and SVR with an AGES representation model, where shape, texture, and wrinkles features were fused together in one integrated model.

CNN is one of the commonly recently used deep learning models to analyze images, videos, and other 2D and 3D data. Images can be interpreted by CNNs as three-dimensional volumes, then, after each layer of the CNN model, the input and output volumes are represented mathematically in terms of multi-dimensional matrices. After that, the dimensions of these matrices are transformed as the image goes deeper into a CNN model. CNN structures are composed of many layers, such as convolutional layers, sub-sampling layers, and fully connected layers [22].

A CNN model proposed in [23] combines handcrafted features and multistage learned features of the facial images. This model includes two approaches: the first approach is based on a feature-level fusion of several local handcrafted features of wrinkles, skin with some other Biologically Inspired Features (BIFs), and the second approach is score-level fusion of feature vectors learned from a CNN with multiple layers. In [24], a new CNN architecture is introduced as Directed Acyclic Graph Convolutional Neural Networks (DAG-CNN) for estimating human age, which automatically learns discriminative features obtained from different layers of a GoogLeNet CNN [25] and VGG16 CNN [26] models and combines them together. As such, in [24], they built two variant architectures. The first is DAG-GoogLeNet based on GoogLeNet CNN and the second is DAG-VGG16 based on VGG-16. Finally, the task of estimating human age was implemented in the decision layer. In [27], a transfer learning was used to solve the problem of gender and age recognition from an image using both VGG19 and VGGFace pre-trained models. A hierarchy of deep CNNs was evaluated, which initially classifies persons by gender then predicts their age using separate male and female age prediction models.

1.3 Human Perception for Face Aging

With the increasing growth of real-world applications in the last decade, there has been also a growing interest in studying visual perception of human in facial age estimation tasks, and comparing the ability of machines versus humans in such tasks. Some researchers, such as [28], explored the ability of humans to estimate the age of subjects from their face images. Such that they showed a set of pictures to a group of participants and asked them to guess the age of each person shown in a picture based on their facial appearance. Simultaneously, they used the same images in building a separate machine learning model. After conducting performance evaluation and comparison, they concluded that the performance of human perception and machine learning attained close results in the age prediction task. Han et al. [29] proposed a framework for automatically estimating demographics from human face based on biologically inspired

features (BIF). Furthermore, they used crowdsourcing mechanism to estimate human perceptual performance in demographic prediction on a variety of aging datasets, including FG-NET, FERET, MORPH II, PCSO, and LFW. This enables comparison between computer and human abilities in predicting three demographics (age, gender, and race). They found that their proposed framework can closely match the performance of human in demographic prediction in all three demographic prediction tasks and it performs slightly better than human with PCSO and MORPH II datasets.

In [30], a study was conducted on the use of a CNN-based model for estimating women's face ages automatically, and further for comparing machine and human performance in terms of which face regions they are much focused on during performing the same task on the same image. The reported human prediction was almost as accurate as the machine prediction using VGG16 model (i.e., 60% accuracy for the CNN model against 61% for human prediction). Their experimental results of tracking human's eyes showed that, when a participant focused their gaze more on the eyes or the mouth regions, their accuracy increased in estimating a person's age; conversely, their accuracy degraded when their gaze concentrated more on other facial skin regions. This may provide some clues to the fact that, indeed, humans may be able to accurately estimate age of others based on some semantic facial features. It also indicated that human-vision capabilities are still competitive and cannot be neglected, even when compared with nowadays sophisticated computer-vision capabilities, and, thus, they can be rather utilized to be integrated with computer-vision capabilities or perhaps to be emulated for achieving augmented performance in such a difficult age estimation task. As such, a reasonable open question can be: What would be the performance of age estimation in the fusion of the capabilities of both humans and machines?

Although some researchers in the age estimation field have paid attention to studying human perception and interpretation of facial aging and comparing their abilities with machines, exploiting humans' vision capabilities in analyzing face images and deriving detailed semantic features for age estimation purposes is yet to be investigated, which may, in turn, help bridging the semantic gap and effectively improve the latest machine-based performance. However, several previous biometrics research studies concerned with recognizing human face [31], identifying subjects [32] and recognizing gait signature [33] have employed human-vision in analyzing and annotating images or videos to derive semantic biometric traits. Their results proved the powerful and effectiveness of such semantic traits in improving model-based performance.

The main contributions of this paper can be summarized as follows:

- Exploiting human-vision capabilities in the process of extracting a novel set of semantic features to be used for a proposed semantic-based age estimation approach in fusion with different combinations of traditional vision-based features.
- Analyzing our proposed semantic features using different statistical tests to investigate and select the most significant set of semantic features for high-level (age group classification) and low-level (regression for exact age prediction) stages.
- Utilizing these semantic features in augmenting and improving the performance of traditional computer vision models in predicting exact age and age groups/levels via face image analysis, allowing for exploring and proving that these two forms of facial features (i.e., traditional vision-based and semantic human-based) are differently informative and effectively integrative in human age estimation tasks.
- Providing an initial research study that may establish further promising research tracks to be undertaken by researchers in the future utilizing our obtained results and findings in how to supplement computer vision-based features with human-based features for augmented facial-based human age estimation, in a way, filling/bridging the semantic gap between humans and machines.

The remainder of this paper is organized as follows. Section 2 explains images dataset, proposed novel semantic features, computer vision-based features, hierarchical multi-level age estimation and evaluation metrics used to evaluate proposed approaches. Experiments and results are presented in Section 3. Section 4 provides a conclusion and future work.

2 Methods and Materials

In this section, we will explain the face images dataset, computer-vision features, semantic features, hierarchical multi-level estimation approach, and the metrics used for model evaluation.

2.1 Dataset

To evaluate our proposed approach, we used a ready-made and publicly available standard dataset called Face and Gesture Recognition Network (FG-NET) [34]. It is one of the most popular and frequently used datasets in the field of facial aging and age estimation [3,35,36]. It contains 1002 face images for 82 individuals, 48 males and 34 females. Individuals in FG-NET dataset have from 6 to 18 color or grayscale images at different ages of the same individuals, with an average of 12 images per individual. The range of ages in FG-NET dataset varies between 0 and 69 years old. There are 68 landmark points provided with each image in FG-NET that represent and localize the face boundaries, eyebrows, lips, eyes, and nose.

Most FG-NET images are collected by scanning the personal photos of individuals from multiple races. Therefore, the quality of images depends on the skills of photographers, camera used, lighting when taking photos, and the quality of photographic paper. This consequently causes a variation among images in quality, resolution, expression, pose, viewpoint, and illumination. In addition, some occlusions appear in a number of images, like makeup, spectacles, hats, beards, and mustaches that may cover or blur the aging signs in some areas of skin. All these conditions posed additional challenges in building an adaptable model capable of analyzing facial aging and achieving desired results. A sample of images for one individual in FG-NET at variant ages is shown in Fig. 4.

2.2 Extracting Semantic Features

In this research, the Figure-Eight crowdsourcing platform [37] was used to collect our novel semantic features. Whilst, most of these semantic aging features were inspired by different forensic concepts of face aging [5], we designed web-based annotation forms and used them to show each FG-NET image to several annotators on the platform and asked them to annotate the displayed image by choosing the most applicable label from a set of given descriptive labels for each facial feature, in such a way describing the degree of presence of each semantic feature on the displayed subject face. These labels are either ordinal such as (None, Minimal, Fair, Marked, or Prominent) and (No wrinkles, Wrinkles with expressions, Wrinkles with rest, or Prominent wrinkles) or binary nominal labels such as (male or female) and (absent or present). The annotator should choose only one most appropriate option/label per feature, where each face image was labeled by multiple annotators. Subsequently, the labels were rescaled using z-score standardization given as:

$$Z = \frac{(x - \mu)}{(\sigma)} \tag{1}$$

where x is a feature vector, μ is the mean and σ is the standard deviation. Then average-labels were calculated for every feature per face image. To facilitate the annotation task, we provided a few instructions with a visual example of an annotated image and grouped related features according to each targeted facial part. The designed annotation task form is displayed in Fig. 5.

This study exploited the human-vision capabilities of a set of people to analyze and extract changes occurring in texture and thickness of facial soft tissue during age progression and investigated their

effects on human facial morphology. Moreover, global aging signs like beard, mustache, gray hair, and pallor were included, in addition to other aging-related features introduced in [38–40]. Tab. 1 shows all proposed 32 semantic face features with their corresponding labels.

Figure 4: Sample of images for a single person in FG-NET dataset, at ages 2, 5, 8, 10, 14, 16, 18, 19, 22, 28, 29, 33, 40, 43, and 43 years old, shown from left to right, respectively

Figure 5: Designed annotation form for semantic feature acquisition

Table 1: Proposed semantic features with their labels grouped according to facial components
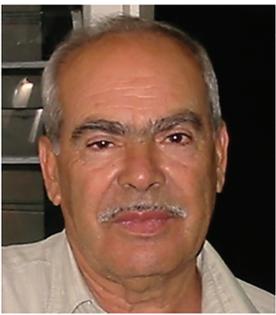
Face Component	Semantic Feature Description	Abbreviation	Semantic Labels
Forehead	Lines and wrinkles forming horizontally on forehead	F1	None, Minimal, Fair, Marked, or Prominent
	Elongation or extension of forehead height	F2	
Eyebrows	Vertical or slightly diagonal lines between eyebrows	F3	
	Sagging or dropping of eyebrows	F4	
Eyes	Lines and wrinkles around outer corners of eyes	F5	
	Eye bags above and below eyes	F6	
	Sagging or dropping of upper eyelids	F7	
	Sunken or hollowing eyes	F8	
	Sunken or hollowing lower eyelids	F9	
	Dark circles under eyes	F10	
Nose	Bulbous nose	F11	
	Lines and wrinkles on nose	F12	
Mouth	Vertical lines above and below lips	F13	
	Elongation or expansion of region between nose and mouth	F14	
	Smile lines or nasolabial folds	F15	
	Curved lines around mouth corners	F16	
	Thinning and flattening of lips	F17	
	Mustache	F18	No (absent) or Yes (present)
Cheeks	Sagging or dropping of cheeks	F19	None, Minimal, Fair, Marked, or Prominent
	Sunken or hollowing of cheeks	F20	
	Lines and wrinkles on cheeks	F21	
Jawlines	Jowls (small focal accumulation of fat in the lower cheek overlying jawbone)	F22	
Chin	Sagging or dropping of chin	F23	
	Marionette lines	F24	
	Lines and wrinkles on chin	F25	
	Beard	F26	No (absent) or Yes (present)
Whole face	Elasticity losing	F27	None, Minimal, Fair, Marked, or Prominent
	Age spots or liver spots	F28	
	Pallor of skin, lips and eyes color	F29	
	Gray hair in eyebrows and eyelashes (mustache and beard if any)	F30	
	Wrinkling degree of entire face	F31	No Wrinkles, Wrinkles with Expressions, Wrinkles with rest, or Prominent Wrinkles
	Gender	F32	Male or Female

Quality assurance services offered by the Figure-Eight platform were all set up to ensure data reliability, such as limiting tasks only for the highest accuracy and most experienced annotators, limiting the maximum number of judgments available for each annotator, pre-test questions for checking annotator understanding and excluding ineligible annotators, setting up detection and prevention of the extra-fast or random responses, and deploying our task for geographically unconstrained annotators to enable better reflection of average human-annotator perception. Tab. 2 provides a summary of collected semantic data.

Table 2: Summary of collected semantic data

Semantic Data Summary	Quantity/Value
Annotations per image	5
Total annotations for all images	5010
Total labels (32 labels per annotation)	160320
Maximum annotations allowed per annotator	5
Total number of annotators	1259

To investigate and select the most significant semantic features in classification and regression stages, we applied two different statistical analysis tests:

2.2.1 ANalysis Of VAriance (ANOVA)

Analysis of variance (ANOVA) is a statistical test used to check whether the means of at least two groups differ from each other or not [41]. We performed ANOVA on our semantic features to measure the significance of each in distinguishing between different age groups in classification stage with a 0.01 significance level.

The p-value for all semantic features is much less than the significance level, 0.01. Therefore, we rejected the null hypothesis for all features and accepted the alternative hypothesis states that there is a difference in the means of at least two age groups. Tab. 3 introduces selected semantic features in descending order according to their effectiveness by F-ratio and corresponding p-value.

Table 3: Semantic features arranged descendingly by F-ratio and P-value according to their effectiveness

Semantic Feature	F1	F27	F5	F31	F2	F4	F7	F9	F8	F25	F6	F24	F19	F3	F21	F22
P-value	1 e-169	1.20 e-168	6.59 e-161	1.30 e-152	1.66 e-149	2.94 e-149	7.15 e-149	1.23 e-144	4.49 e-143	3.95 e-142	1.31 e-139	4.01 e-136	3.93 e-134	1.90 e-133	2.89 e-130	2.27 e-129
F-ratio	397.2	393.6	368.1	341.7	332.1	331.3	330.2	317.3	312.6	309.8	302.4	292.3	286.5	284.6	275.6	273.1
Semantic Feature	F23	F13	F16	F20	F30	F10	F17	F15	F29	F28	F12	F14	F11	F18	F26	F32
P-value	1.41 e-128	6.46 e-125	7.61 e-125	8.93 e-125	1.69 e-120	8.54 e-114	8.73 e-114	3.34 e-108	2.21 e-107	8.16 e-103	9.81 e-101	6.74 e-99	1.36 e-94	8.64 e-28	2.67 e-09	2.24 e-06
F-ratio	270.8	260.7	260.5	260.3	248.7	231	230.97	216.6	214.5	203.1	197.98	193.5	183.1	45.9	14.58	9.8

2.2.2 Pearson Correlation Coefficient

Pearson's correlation coefficient is used in statistics to measures strength and direction of linear correlation between two continuous variables x and y [42,43]. Assuming $\bar{x}$ and $\bar{y}$ are the means of first and second variables, respectively, n is the number of pairs in x and y, the formula of Pearson's correlation coefficient r is defined as in Eq. (2). When the correlation is closer to +1, this indicates a stronger positive correlation and highly correlated variables, whereas when it is closer to -1, this means a stronger but negative correlation between variables. Correspondingly, the correlation becomes either way weaker as it closes to zero.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2}$$

As shown in Fig. 6, the correlation between age and each of our semantic features is strongly positive, except for three features, which are gender and presence of mustache or beard. This is due to that very few subjects appear in FG-NET dataset images with mustache or beard, such that only 36 images show subjects with a mustache and 31 with a beard. Furthermore, there is a lack of a direct or sensible relationship between human gender and age, due to the natural stability of gender over the lifetime. Nevertheless, extended analysis and exploration of different aging characteristics with respect to gender could enable more understanding of the similarities and differences between males and females (or average male and average female) in aging progression changes or signs from different perspectives and various aspects, such as what, when, where, and how similar or different those aging characteristics are expected to be, which could, in turn, be considered to help age estimation. Thus, these three semantic features were excluded from regression stage.

Tab. 4 provides semantic features in descending order based on their correlation strength with age in addition to corresponding p-value for each feature at 0.01 confidence level.

2.3 Extracting Computer-vision Features

To prepare the data images for computer vision-based feature extraction stage, all colored images were converted into grayscale. Next, since the FG-NET contains images from personal image collections of subjects, pose variation is likely observed in many subject images. Thus, images were rotated to a standard face pose. Image rotation was conducted based on the coordinates of the center of the two eyes to align all images vertically, where the rotation angle θ is computed as follows:

$$\theta = \tan^{-1}\frac{Y_L - Y_R}{X_L - X_R} \tag{3}$$

where (X_R, Y_R) and (X_L, Y_L) are the position of the right and left eyes, respectively. Fig. 7 provides an example of a face image before and after preprocessing.

2.3.1 Active Appearance Model

Active Appearance Model (AAM) is a computer-vision model proposed by Cootes et al. [44]. It combines shape and gray-level appearance statistical models to represent and interpret objects in images. It was developed and used in many applications such as objects tracking, gait analysis, human eyes modeling, facial expression recognition, and medical image segmentation [45]. We have employed AAM [46] proposed for fast AAM fitting in-the-wild.
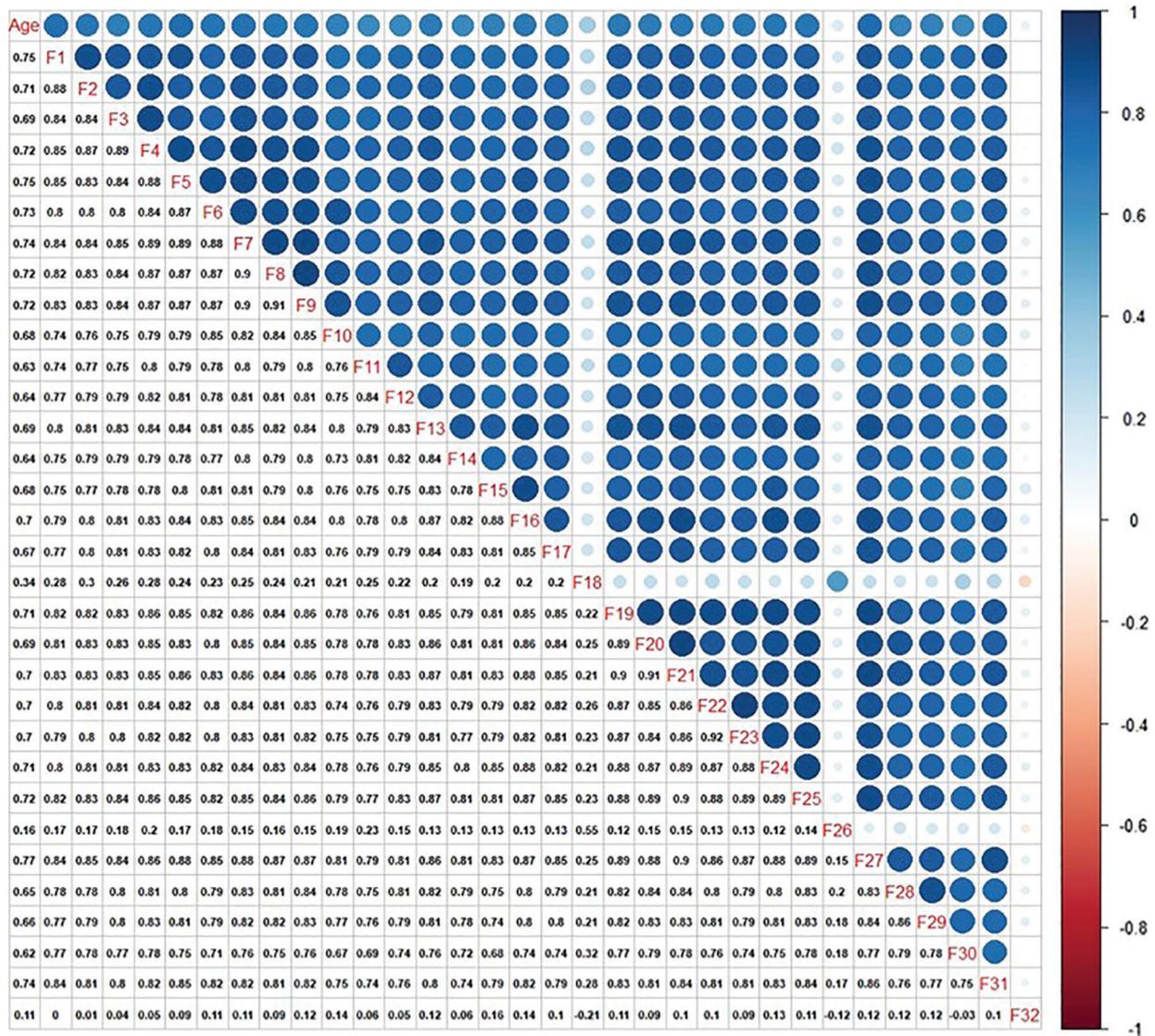
Figure 6: Correlation matrix between semantic features and age and between each other using Pearson's r correlation coefficient

In this work, to build the shape model, a set of 68 landmark points (X_1, Y_1, X_2, Y_2, …., X_{68}, Y_{68}) describing the shape of a human face is required across D training images. Procrustes analysis was applied to remove similarity transformation from original shapes and get D similarity-free shapes. Then, PCA was performed on these shapes to obtain the shape model $\hat{s}$, defined by a set of shape eigenvectors S and the mean shape s_0. The statistical shape model can be expressed as:

$$\hat{s} = s_0 + SP, \text{ where } P = S^T(s - s_0) \tag{4}$$

where s is a new similarity-free shape and p contains the shape parameters of a deformable model.

Table 4: Semantic features arranged in descending order according to their correlation strength with age

Semantic Feature	F27	F5	F1	F31	F7	F6	F4	F25	F8	F9	F24	F2	F19	F16	F22	F21
P-value	2.34 e-194	4.01 e-184	1.56 e-178	3.73 e-175	2.60 e-173	1.30 e-166	9.67 e-161	1.60 e-160	1.95 e-160	4 e-160	2.24 e-157	2.88 e-156	6.24 e-152	1.08 e-150	1.33 e-150	3.02 e-150
Pearson's r	0.77	0.75	0.75	0.74	0.74	0.73	0.72	0.72	0.72	0.72	0.71	0.71	0.71	0.7	0.7	0.7
Semantic Feature	F23	F13	F20	F3	F10	F15	F17	F29	F28	F12	F14	F11	F30	F18	F26	F32
P-value	3.79 e-149	5.10 e-145	2.12 e-144	3.40 e-140	3.21 e-136	6.11 e-136	1.56 e-132	1.06 e-126	4.66 e-122	7.20 e-116	5.96 e-115	1.54 e-110	1.10 e-108	5.02 e-29	5.37 e-07	0.656 e-3
Pearson's r	0.7	0.69	0.69	0.69	0.68	0.68	0.67	0.66	0.65	0.64	0.64	0.63	0.62	0.34	0.16	0.11

(a)   (b)
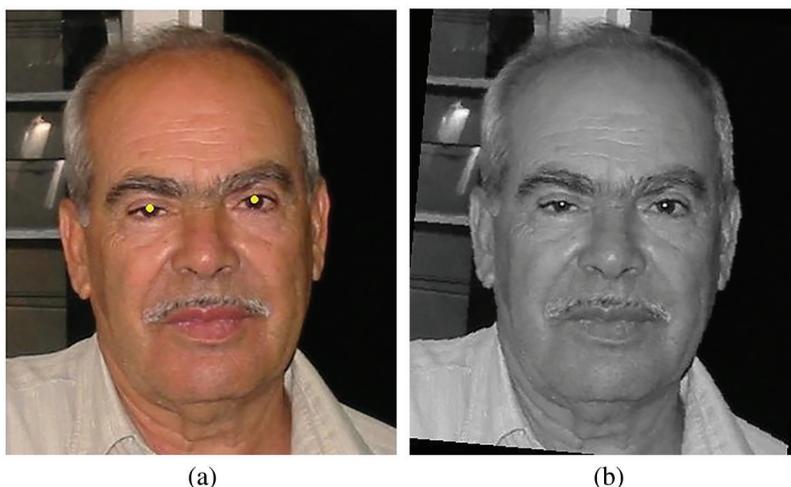
Figure 7: (a) is the original image showing eyes coordinates in yellow color. (b) is the same image after rotation and grayscale conversion

Building the appearance model requires removing variation of shapes from the texture. This was achieved by warping each texture I to the reference frame obtained using the mean shape. After that, PCA was performed on shape-free textures to obtain the appearance model $\hat{I}$ defined by a set of appearance eigenvectors A and the mean normalized appearance vector A_0. By assuming c represents the appearance parameters, the statistical appearance model can defined as:

$$\hat{I} = A_0 + Ac, \text{ where } c = A^T(I - A_0)$$ (5)

Fast-Forward algorithm was used to fit the AAM model to extract global facial shape and appearance from each image. Fig. 8 below illustrates detected landmark points with corresponding appearance.
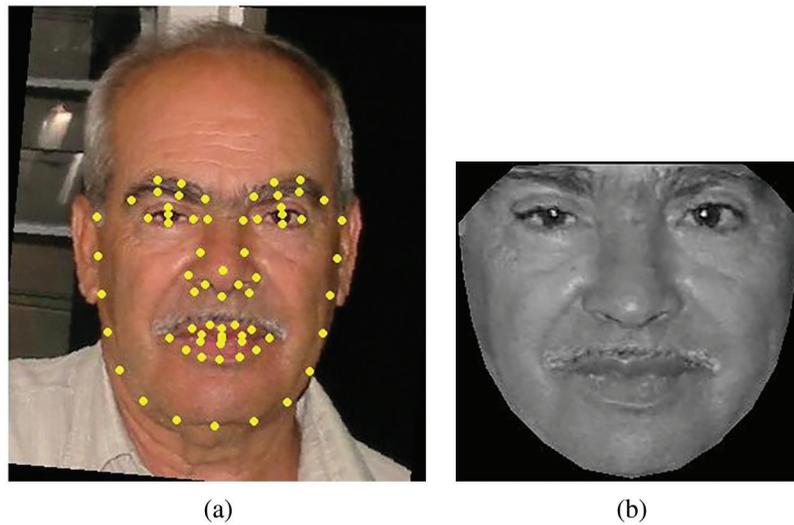
Figure 8: (a) landmark points define facial shape, (b) Shape-free patch represents face appearance

2.3.2 Geometric Ratios

Craniofacial changes occur in human face with age progression and convey much information about facial aging. They are useful in distinguishing children from adults and studying facial aging signs related to children, who are characterized by their rapid geometric growth [47].

Due to the fact that the facial shape may be affected by expressions and pose, this might influence on the localization of facial landmark points. Therefore, the ratios between distances were used to measure human craniofacial growth rather than solely using distances. We calculated 19 distances D *via* 23 landmark points and used them to form 18 different ratios R as geometric features as in [47,48]. All landmark points in this section were produced using AAM as demonstrated in the previous section, except for point number 23, which was manually localized due to hair covering the forehead or bald appearing for some individuals in the FG-NET dataset. The ratio R between two distances is calculated as:

$$R = \frac{D_1}{D_2}, \text{ where } D_i(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} \qquad (6)$$

where p_1, p_2, q_1 and q_2 are coordinates of p and q landmark points. Fig. 9 presents 23 landmark points used to calculate the ratios of geometric features and Fig. 10 illustrates 18 computed shape ratios of distances between different landmark points. The difference between an adult and a child in R_6 ratio between face width and height eyebrow center and chin points is shown in Fig. 11.
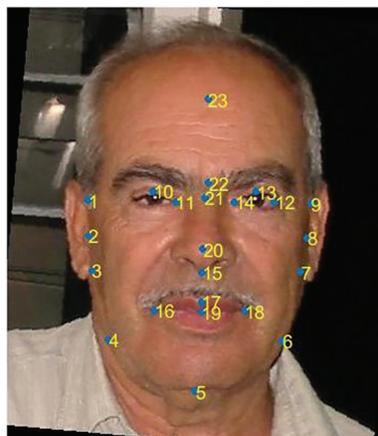
Figure 9: 23 landmark points for calculating ratios between distances
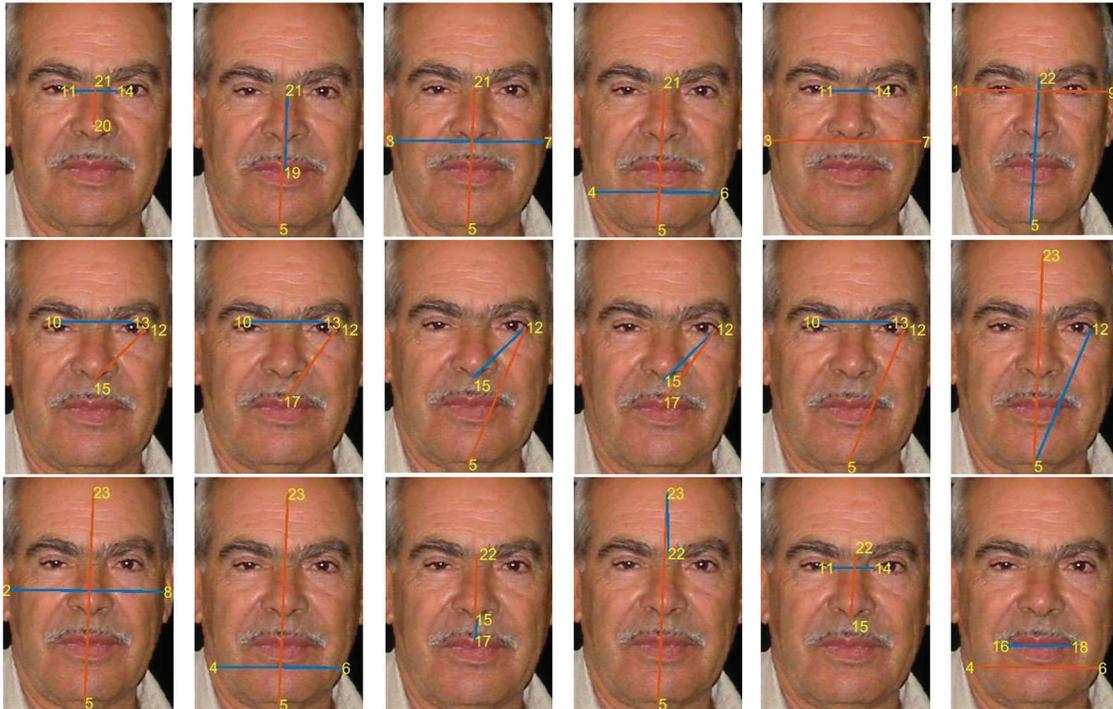
Figure 10: 18 Ratios between different landmark pair points distances
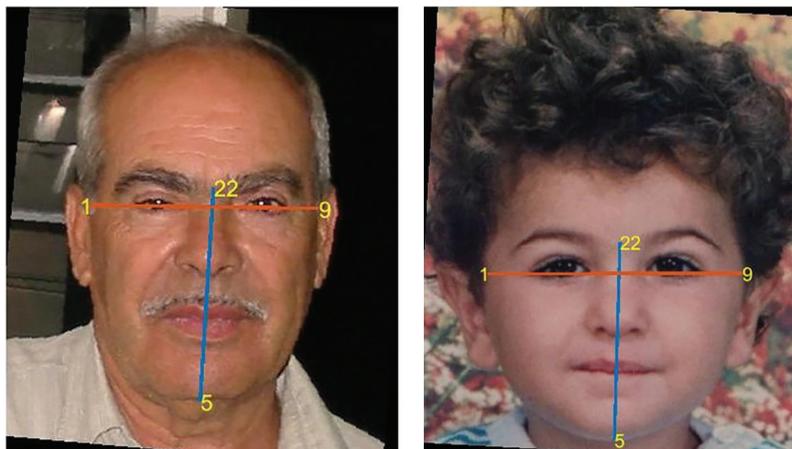
Figure 11: The difference between an adult and a child using R6 ratio

2.3.3 Local Binary Patterns

Lately, in human facial age estimation, local skin features analysis has achieved substantial efficiency [49] as it can represent the significant facial information related to soft tissue aging and remove noise such as hair, background, and non-skin areas. As shown in Fig. 12, eleven facial regions were cropped to be used in deriving local texture features using Local Binary Patterns (LBP).
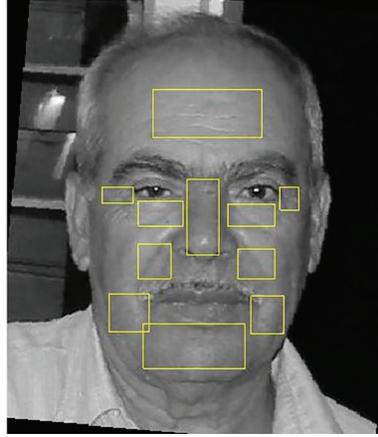
Figure 12: 11 skin regions used to extract LBP and Gabor features

LBP's success stems from the robust binary code that is highly sensitive to texture and soft tissue changes uninfluenced by the light, noise, and facial expressions [18]. LBP takes the value of each pixel as a threshold and calculates 8-bit binary code for their neighboring pixels. Then, a histogram is generated as a texture descriptor [50,51]. The LBP code is expressed as:

$$LBP_{P,R} = \sum_{P=0}^{P-1} s(g_P - g_C)2^P, \text{ where } s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \tag{7}$$

where g_C is the gray value of center pixel, g_P is the gray value of equally spaced pixels on the radius circle, s is a thresholding function, P refers to the number of neighboring pixels and R is the distance between center and neighboring pixels.

2.3.4 Gabor Filters

Gabor filters have been commonly used in facial age estimation to extract wrinkles and edge features, because of their ability to determine both the orientation and magnitude of wrinkles and lines [51]. It is characterized by its robustness to noise caused by aspects such as glasses and beard [52]. Here, on each pixel of the face image, filter banks of 2D Gabor filters, including five frequencies and eight orientations, were used in our approach as illustrated in Fig. 13. The Gabor filter formula [53] is defined as follows:

$$G(x,y) = \frac{f^2}{\pi\gamma\eta} \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp(j2\pi f x' + \phi)$$

$$x' = x\cos\theta + y\sin\theta$$

$$y' = -x\sin\theta - y\cos\theta \tag{8}$$

where θ denotes the orientation, f represents the frequency, σ refers to standard deviation of Gaussian envelope, ϕ is the offset of phase, and γ is spatial aspect ratio that specifies the ellipticity to support Gabor function. The local skin regions shown in Fig. 12 were also used here to extract Gabor features.
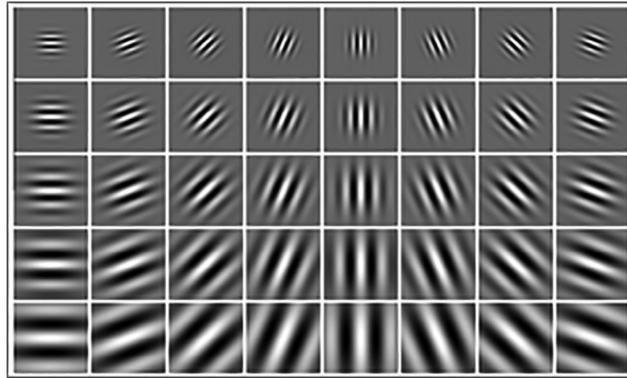
Figure 13: Used 2D Gabor filters in five scales and eight orientations

Finally, all computer-vision features were rescaled using Z-score standardization Eq. (1), then, due to the high dimensionality of shape, appearance, LBP and Gabor features, PCA was performed as a feature subset selection method to determine and select the most effective components. Tab. 5 summarizes the determined number of the most effective principal components for each feature type.

Table 5: Feature dimensions after applying PCA for features selection

Type of Features	Shape	Appearance	LBP	Gabor
Number of Features	20	200	20	50

2.4 Hierarchical Classification and Regression

Estimating human facial age may be a problem of multi-classification [54–56], regression [57,58], or a hybrid of both techniques [18,59]. Many results of previous research explorations have proved that a hybrid or hierarchical age estimation approach outperformed single-stage approaches [60].

In this work, we applied a multi-level hierarchical age estimation approach which classifies images into their appropriate age groups (as high-level age estimation), then predicts an exact age value using regression (as low-level age estimation). A single (one-vs-one) SVM multi-classifier was trained to classify images into one of four age groups: from newborn to toddler (0-3 years), from pre-adolescence to adolescence (4-16 years), from young adult to adult (17-39), and from middle-aged to senior (40-69 Year).

In the regression stage, four SVM regression models were trained. Each regressor was trained using a specific age group to be responsible for predicting ages within the range of this age group. To reduce the effect of misclassification, the regressors were trained with a range overlapping between each adjacent regressors. The range of overlapping among age groups was experimentally selected for better regression performance. Fig. 14 shows the structure of our multi-level age estimation.
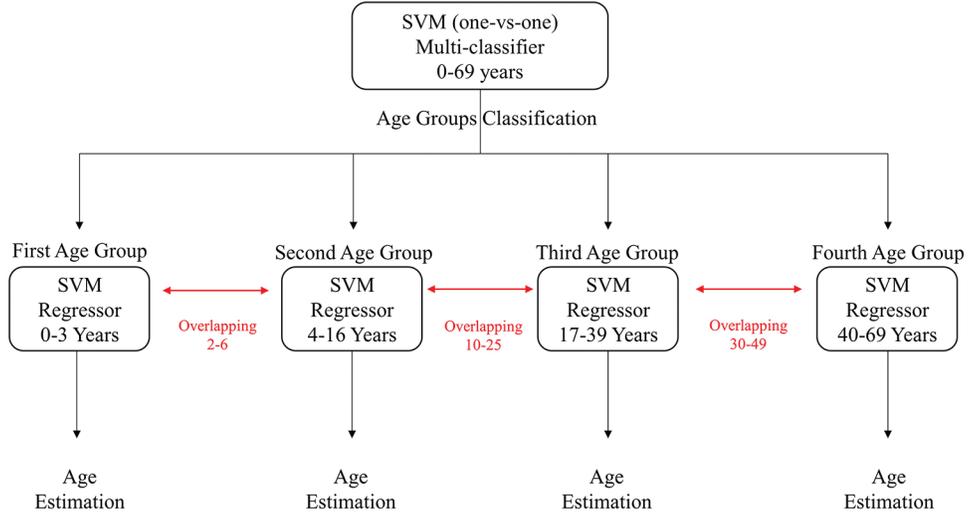
Figure 14: Multi-level hierarchical classification and regression with age groups overlapping to reduce the effect of misclassification on age prediction

2.5 Evaluation Metrics

The accuracy measure is used to evaluate the performance of SVM in age group classification inferred as:

$$Accuracy = \frac{C}{N} \times 100\% \tag{9}$$

where C is the number of correctly classified images and N refers to the number of all tested images.

The Mean Absolute Error (MAE) and Cumulative Score (CS) evaluation metrics have been commonly used to evaluate the performance of specific age estimation [61]. MAE is defined as the mean of absolute difference between ground-truth and predicted ages, which is expressed as:

$$MAE = \frac{\sum_{i=1}^{N} |a\hat{g}e_i - age_i|}{N} \tag{10}$$

where N refers to the number of tested images, $a\hat{g}e_i$ is the predicted age of image i and age_i is the ground-truth age of image i.

The CS is the ratio of the number of images whose errors are less than or equal to the threshold value, divided by the total number of test images. CS enables performance comparison at different absolute error levels, which can be defined as:

$$CS = \frac{N_{e \leq x}}{N} \times 100\% \tag{11}$$

where N_e is the number of images whose estimation error is less than the defined threshold x, and N is the total number of test images.

3 Experiments and Results

Each person in FG-NET dataset has several images at different age stages. Therefore, dividing the dataset into 80% for training and 20% for testing like [62,63], appears to be not an ideal way to test our models, due to the possibility of having very similar images belonging to one person in both training and testing stages. Hence, to prevent that, we have followed a Leave One Person Out (LOPO) validation technique as applied in [64–68] to unbiasedly and effectively evaluate our proposed approach. In LOPO, all images belonging to one person are included into testing set while the remaining images of other people in the dataset are used to train the model. As such, the LOPO strategy is repeated as per the number of individuals in the database with a different individual one-by-one used for testing each time, while the final results are calculated from all the estimates.

The accuracy metric is used here to evaluate the performance of SVM in classifying images into their respective age groups. As shown in Tab. 6, we carried out several classifications with various combination of features. The accuracy results were firstly presented using computer-vision-based features alone and secondly after being augmented with our novel semantic features. All proposed 32 semantic features were included in the classification stage based on deduced significance analysis by ANOVA. A significant improvement was observed in all classification experiments after supplanting the vision-based classifiers with the semantic features. The best obtained accuracy result was 75.15% when adding semantic features against only 63.87% when using vision features alone. We will rely on these experiments later to perform regression process.

Table 6: Accuracy of age group classification using different combinations of vision features, before and after adding semantic features

Vision-based Features	Before adding semantic features %	After adding semantic features %
Ratios + Coordinates + Appearance	68.36	**69.26**
Ratios + Coordinates + LBP	63.87	**75.15**
Ratios + Coordinates + Gabor	64.37	**71.76**
Ratios + Appearance + LBP	67.56	**71.86**
Ratios + Appearance + Gabor	65.17	**66.67**
Ratios + LBP + Gabor	62.28	**69.16**
Coordinates + Appearance + LBP	67.27	**71.36**
Coordinates + Appearance + Gabor	65.47	**67.66**
Coordinates + LBP + Gabor	64.37	**71.06**
Appearance + LBP + Gabor	64.17	**66.17**

Fig. 15a shows the confusion matrix illustrates how images are classified when computer-vision features are used in isolation. It is clear that age group classification is achieved with high misclassification rate, especially in the fourth age group for people of middle-aged to senior (40-69 Year) ages. On the other hand, Fig. 15b provides the confusion matrix for the same age group classification approach but after being augmented with the proposed semantic features. As can be observed, the accuracy of age classification was significantly increased in all age groups. Furthermore, it was reported that a majority of misclassified images have been assigned to their adjacent age groups, indicating a closer age group

estimation to the actual age group, meaning a smaller range of error unlike the traditional vision-based methods, which were often found randomly assigning a far falsely age group in such misclassification cases. This demonstrates the potency of these semantic features and how they can be capable to distinguish several age groups with less possible error.
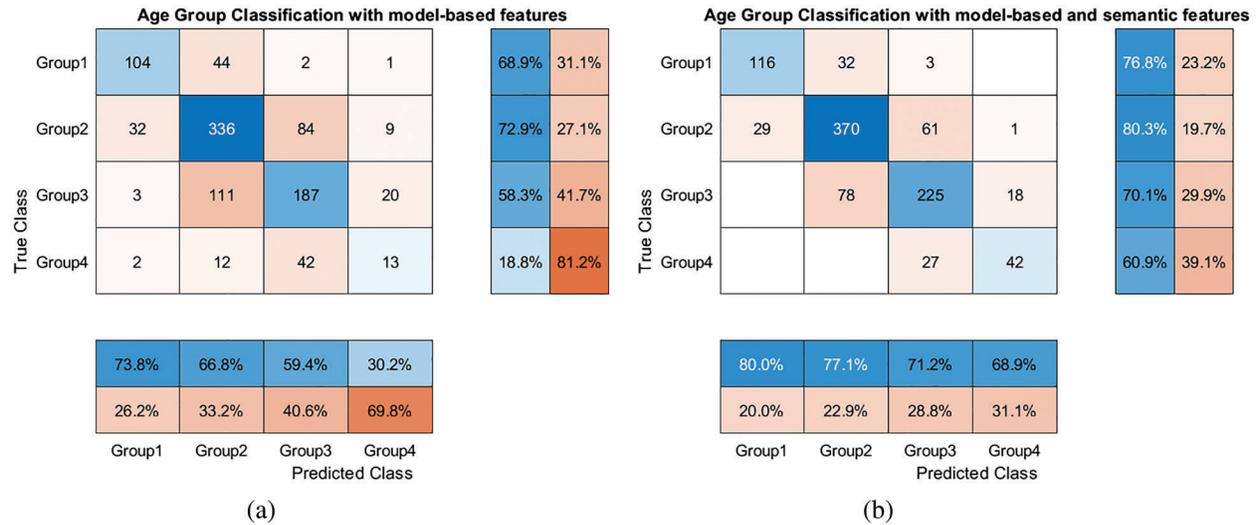
Figure 15: (a) Confusion matrix of age groups classification using vision-based features alone, (b) Confusion matrix of age groups classification using vision-based and semantic features

The MAE measurement evaluates the performance of the used regressors in estimating exact ages. In the regression stage, we have eliminated gender, mustache, and beard semantic features from our experiments due to their weak effect and relationship to the exact age value as proved by Pearson's correlation coefficient analysis.

Tab. 7 displays the results of MAE using various combinations of features. The experiments show that all results improved after combining semantic features with traditional computer-based features. The best MAE we have achieved is 4.41 years old, compared to 5.71 years when semantic features are excluded.

Table 7: Mean absolute error (MAE) for exact age prediction before and after adding semantic features

Vision-based Features	Before adding semantic features	After adding semantic features
Ratios + Coordinates	6.54	**4.66**
Ratios + Appearance	5.71	**4.41**
Ratios + LBP	6.41	**4.58**
Ratios + Gabor	6.65	**5.02**
Coordinates + Appearance	5.64	**4.53**
Coordinates + LBP	6.45	**4.62**
Coordinates + Gabor	6.53	**5.19**
Appearance + LBP	5.84	**4.71**
Appearance + Gabor	6.44	**5.52**
LBP + Gabor	7.12	**5.61**

Tab. 8 summarizes the MAE produced by each age group separately to clarify the positive impact of semantic features on each regressor. The results show how semantic features decrease MAE at all regressors, while a clear improvement was remarkably noticed with older ages within the conducted experiments. This reflects the viability and great sensitivity of our proposed semantic features in capturing and analyzing details of soft facial tissues, which are essential signs of aging in adults and seniors.

Table 8: (MAE) for each regressor separately and their overall MAE

Age estimation performance	MAE 1	MAE 2	MAE 3	MAE 4	Overall MAE
Without Semantic Features	3.18	4.59	7.38	14.93	5.71
With Semantic Features	**2.03**	**3.55**	**5.91**	**8.95**	**4.41**

The distribution of images in FG-NET dataset is not balanced for all ages, where the number of images decreases as the age increases, since regression models usually need sufficient samples for training to be able to predict ages more efficiently. As a result, the MAE gets worse as the average of ages within the age group increases.

The cumulative score was computed at different error levels ranging from 0 to 10 years. Consequently, 83.53% of images have a level error less than or equal to 10 years when using computer-vision features, while the cumulative score percentage improved to reach 90.12% after augmenting them with semantic features.

After a considerable improvement was proved for the new proposed semantic features by all means in the carried out experiments, we have used them alone to train a classifier and regressors on age estimation task for validating their pure viability and performance in isolation. The received results were 61.88% classification accuracy, 6.15 MAE, and 83.23% CS for age prediction. Fig. 16 shows and compares the CS performance when using computer-vision features, semantic features, and the fusion of both semantic and vision features.
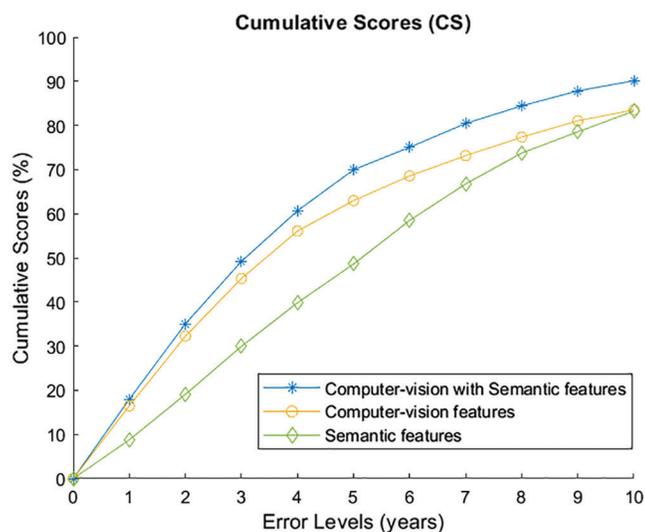
Figure 16: Cumulative Score of computer-vision features, semantic features, and augmenting computer-vision with semantic features
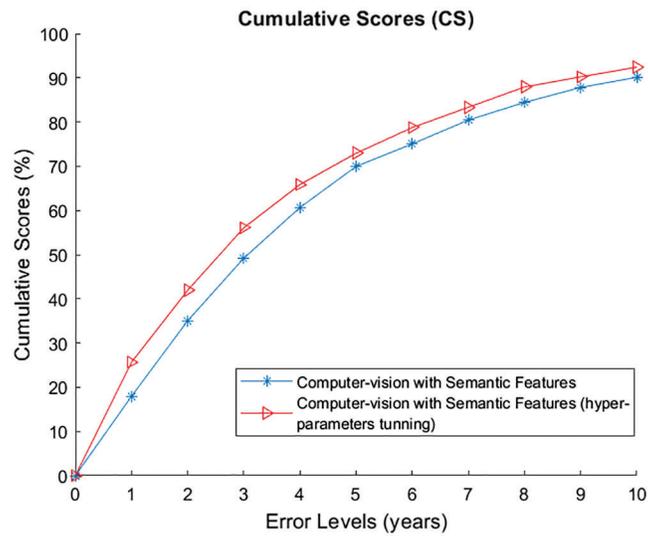
Figure 17: Cumulative Score of our proposed augmented approach before and after optimization process

Table 9: The best MAE obtained before and after tuning the hyperparameters of each regressors

Proposed Approach Performance	MAE 1	MAE 2	MAE 3	MAE 4	Overall MAE
Before Optimization	2.03	3.55	5.91	8.95	4.41
After Optimization	**1.31**	**3.06**	**5.62**	**7.75**	**3.90**

Finally, we used the features of the second experiment as shown in Tab. 7 that achieves the best MAE (4.41 years); a Bayesian optimization was performed for each regressor independently, because each of them deals with a different age range. The Bayesian optimization intends to search for optimum hyperparameters that give a minimum error. We obtained 3.90 MAE and 92.42% CS at 10 years' error level. Fig. 17 and Tab. 9 compare the CS and MAE before and after optimization process, respectively. Fig. 18 provides samples of good and poor age estimation of some images along with their respective actual and estimated age value.
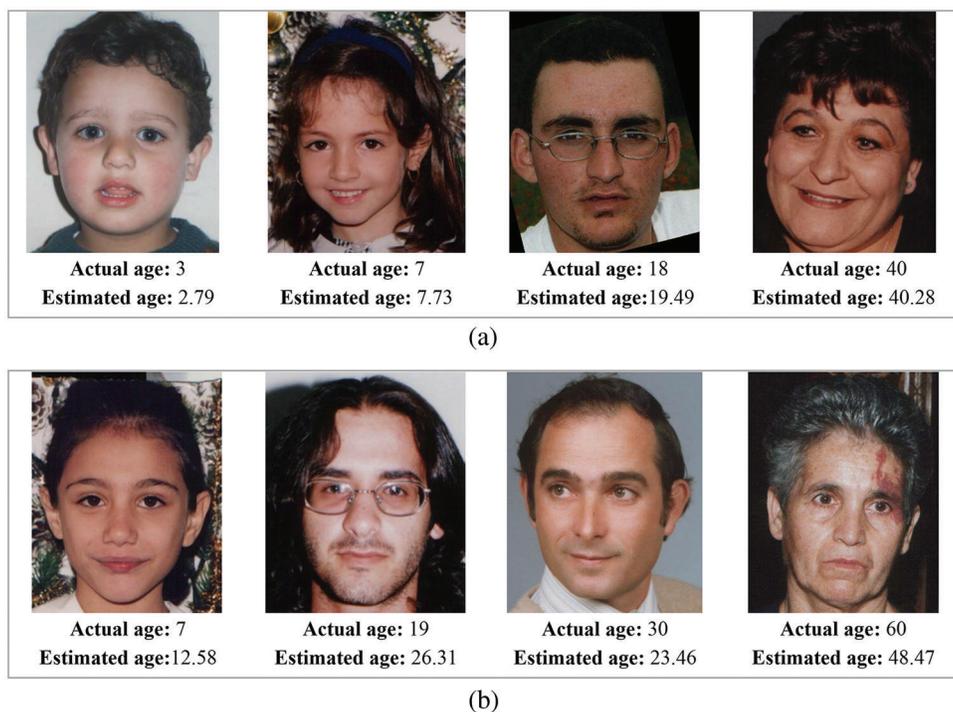
Figure 18: Sample results of different age estimations; (a) provides good estimations, (b) provides poor estimations

4 Conclusion and Future Work

In this paper, we investigated the problem of human age estimation from face images. We proposed a set of novel semantic features representing various facial aging characteristics or signs. We examine significance and effectiveness of each semantic feature using diverse statistical analysis tests, then we exploit powerful features to augment traditional computer-vision-based features. A hierarchical estimation approach was applied by performing classification to predict age group followed by regression to estimate the exact age. FG-NET aging dataset with variety in expressions, races, resolutions, poses and illumination was used to evaluate the performance of our proposed approach.

Experimental results proved a significant and remarkable improvement in age estimation performance, by all means, occurred at the results of both stages: age group (high-level) classification; and exact age (low-level) prediction, when fusing and supplementing traditional computer vision-based features with our proposed human-based semantic features, especially with ages in the fourth age group ranging from middle-aged to senior (*i.e.*, from 40 to 69 years old).

This initial research study may establish further research tracks to be undertaken by researchers in the future utilizing our obtained results and findings in how to supplement computer vision-based features with human-based features for augmented facial-based human age estimation.

As future work, increased discrimination for more robust age estimation capabilities can be investigated by fusing our proposed semantic features with deep features, where it is highly expected, based on this research's results and conclusions, that our semantic features would be most likely capable to augment/ enhance the performance when fused with the most recent models based on deep learning and CNN, and to outperform their performance when used in isolation. Learning for automatic semantic face feature

extraction and annotation is a potential desired future work that helps researchers to improve computer-vision techniques and bridge the semantic gap between machines and humans automatically.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1]   M. C. Voelkle, N. C. Ebner, U. Lindenberger and M. Riediger, "Let me guess how old you are: Effects of age, gender, and facial expression on perceptions of age," *Psychology and Aging*, vol. 27, no. 2, pp. 265–277, 2012.

[2]   A. A. Shejul, K. S. Kinage and B. E. Reddy, "Comprehensive review on facial based human age estimation," in *2017 Int. Conf. on Energy, Communication, Data Analytics and Soft Computing*, Chennai, India, pp. 3211–3216, 2017.

[3]   M. Georgopoulos, Y. Panagakis and M. Pantic, "Modeling of facial aging and kinship: A survey," *Image and Vision Computing*, vol. 80, no. 14, pp. 58–79, 2018.

[4]   R. Angulu, J. R. Tapamo and A. O. Adewumi, "Age estimation via face images: A survey," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, pp. 179, 2018.

[5]   M. Kaur, R. K. Garg and S. Singla, "Analysis of facial soft tissue changes with aging and their effects on facial morphology: A forensic perspective," *Egyptian Journal of Forensic Sciences*, vol. 5, no. 2, pp. 46–56, 2015.

[6]   A. Othmani, A. R. Taleb, H. Abdelkawy and A. Hadid, "Age estimation from faces using deep learning: A comparative analysis," *Computer Vision and Image Understanding*, vol. 196, no. 8, pp. 102961, 2020.

[7]   N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," in *2006 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'06)*, New York, NY, USA, pp. 387–394, 2006.

[8]   T. Dhimar and K. Mistree, "Feature extraction for facial age estimation: A survey," in *2016 Int. Conf. on Wireless Communications, Signal Processing and Networking (WiSPNET)*, Chennai, India, pp. 2243–2248, 2016.

[9]   S. Kumar, S. Ranjitha and H. Suresh, "An active age estimation of facial image using anthropometric model and fast ica," *Journal of Engineering Science & Technology Review*, vol. 10, no. 1, 2017.

[10] F. B. Alvi, R. Pears and N. Kasabov, "An evolving spatio-temporal approach for gender and age group classification with spiking neural networks," *Evolving Systems*, vol. 9, no. 2, pp. 145–156, 2018.

[11] L. F. Porto, L. N. C. Lima, A. Franco, D. M. Pianto, C. E. M. Palhares *et al.,* "Estimating sex and age for forensic applications using machine learning based on facial measurements from frontal cephalometric landmarks," *arXiv preprint arXiv:1908.02353,* 2019.

[12] P. Thukral, K. Mitra and R. Chellappa, "A hierarchical approach for human age estimation," in *2012 IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Kyoto, Japan, pp. 1529–1532, 2012.

[13] M. Iqtait, F. Mohamad and M. Mamat, "Feature extraction for face recognition via active shape model (asm) and active appearance model (aam)," in *IOP Conf. Series: Materials Science and Engineering*, Tangerang Selatan, Indonesia, pp. 12032, 2018.

[14] S. A. Rizwan, A. Jalal, M. Gochoo and K. Kim, "Robust active shape model via hierarchical feature extraction with sfs-optimized convolution neural network for invariant human age classification," *Electronics*, vol. 10, no. 4, pp. 465, 2021.

[15] P. C. S. Reddy, K. Sarma, A. Sharma, P. V. Rao, S. G. Rao *et al.,* "Enhanced age prediction and gender classification (eap-gc) framework using regression and svm techniques," Materials Today: Proc., 2020.

[16] A. A. Shejul, K. S. Kinage and B. E. Reddy, "Cdbn: Crow deep belief network based on scattering and aam features for age estimation," *Journal of Signal Processing Systems*, vol. 92, no. 10, pp. 1–19, 2020.

[17] V. Martin, R. Seguier, A. Porcheron and F. Morizot, "Face aging simulation with a new wrinkle oriented active appearance model," *Multimedia Tools and Applications*, vol. 78, no. 5, pp. 6309–6327, 2019.

[18] T. Wu, Y. Zhao, L. Liu, H. Li, W. Xu *et al.,* "A novel hierarchical regression approach for human facial age estimation based on deep forest," in *2018 IEEE 15th Int. Conf. on Networking, Sensing and Control*, Zhuhai, China, pp. 1–6, 2018.

[19] J. K. Pontes, A. S. Britto Jr, C. Fookes and A. L. Koerich, "A flexible hierarchical approach for facial age estimation based on multiple features," *Pattern Recognition*, vol. 54, no. 10, pp. 34–51, 2016.

[20] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li and H. Dai, "Learning from facial aging patterns for automatic age estimation," in *Proc. of the 14th ACM Int. Conf. on Multimedia*, Santa Barbara, CA, USA, pp. 307–316, 2006.

[21] T. K. Sahoo and H. Banka, "Multi-feature-based facial age estimation using an incomplete facial aging database," *Arabian Journal for Science and Engineering*, vol. 43, no. 12, pp. 8057–8078, 2018.

[22] P. Punyani, R. Gupta and A. Kumar, "Neural networks for facial age estimation: A survey on recent advances," *Artificial Intelligence Review*, vol. 53, no. 5, pp. 3299–3347, 2020.

[23] S. Taheri and Ö. Toygar, "On the use of dag-cnn architecture for age estimation with multi-stage features fusion," *Neurocomputing*, vol. 329, no. 1, pp. 300–310, 2019.

[24] S. Taheri and Ö. Toygar, "Multi-stage age estimation using two level fusions of handcrafted and learned features on facial images," *IET Biometrics*, vol. 8, no. 2, pp. 124–133, 2019.

[25] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed *et al.,* "Going deeper with convolutions," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, Massachusetts, pp. 1–9, 2015.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv: 1409. 1556, 2014.

[27] P. Smith and C. Chen, "Transfer learning with deep cnns for gender recognition and age estimation," in *2018 IEEE Int. Conf. on Big Data (Big Data)*, Seattle, WA, USA, pp. 2564–2571, 2018.

[28] H. Han, C. Otto and A. K. Jain, "Age estimation from face images: Human vs. Machine performance," in *2013 Int. Conf. on Biometrics*, Madrid, Spain, pp. 1–8, 2013.

[29] H. Han, C. Otto, X. Liu and A. K. Jain, "Demographic estimation from face images: Human vs. Machine performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1148–1161, 2015.

[30] J. Alarifi, J. Fry, D. Dancey and M. H. Yap, "Understanding face age estimation: Humans and machine," in *2019 Int. Conf. on Computer, Information and Telecommunication Systems*, Beijing, China, pp. 1–5, 2019.

[31] E. S. Jaha, "Augmenting gabor-based face recognition with global soft biometrics," in *2019 7th Int. Sym. on Digital Forensics and Security*, Barcelos, Portugal, pp. 1–5, 2019.

[32] E. S. Jaha and M. S. Nixon, "Soft biometrics for subject identification using clothing attributes," in *IEEE Int. Joint Conf. on Biometrics*, Clearwater, FL, USA, pp. 1–6, 2014.

[33] S. Samangooei and M. S. Nixon, "Performing content-based retrieval of humans using gait biometrics," *Multimedia Tools and Applications*, vol. 49, no. 1, pp. 195–212, 2010.

[34] G. Panis, A. Lanitis, N. Tsapatsoulis and T. F. Cootes, "Overview of research on facial ageing using the FG-NET ageing database," *IET Biometrics*, vol. 5, no. 2, pp. 37–46, 2016.

[35] V. Carletti, A. Greco, G. Percannella and M. Vento, "Age from faces in the deep learning revolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 9, pp. 2113–2132, 2020.

[36] O. F. Osman and M. H. Yap, "Computational intelligence in automatic face age estimation: A survey," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 3, no. 3, pp. 271–285, 2019.

[37] Figure eight, 2020. Available: https://appen.com/figure-eight-is-now-appen/.

[38] A. M. Albert, K. Ricanek Jr and E. Patterson, "A review of the literature on the aging adult skull and face: Implications for forensic science research and applications," *Forensic Science International*, vol. 172, no. 1, pp. 1–9, 2007.

[39] S. R. Coleman and R. Grover, "The anatomy of the aging face: Volume loss and changes in 3-dimensional topography," *Aesthetic Surgery Journal*, vol. 26, no. 1_Supplement, pp. S4–S9, 2006.

[40] A. Porcheron, E. Mauger and R. Russell, "Aspects of facial contrast decrease with age and are cues for age perception," *PLoS One*, vol. 8, no. 3, pp. e57985, 2013.

[41] K. H. Liu, S. Yan and C. C. J. Kuo, "Age estimation via grouping and decision fusion," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 11, pp. 2408–2423, 2015.

[42] C. E. P. Machado, M. R. P. Flores, L. N. C. Lima, R. L. R. Tinoco, A. Franco *et al.,* "A new approach for the analysis of facial growth and age estimation: Iris ratio," *PLOS One*, vol. 12, no. 7, pp. e0180330, 2017.

[43] M. Alamri and S. Mahmoodi, "Facial profiles recognition using comparative facial soft biometrics," in *2020 Int. Conf. of the Biometrics Special Interest Group*, Darmstadt, Germany, pp. 1–4, 2020.

[44] T. F. Cootes, G. J. Edwards and C. J. Taylor, "Active appearance models," in *European Conf. on Computer Vision*, Berlin, Heidelberg, pp. 484–498, 1998.

[45] X. Gao, Y. Su, X. Li and D. Tao, "A review of active appearance models," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 2, pp. 145–158, 2010.

[46] G. Tzimiropoulos and M. Pantic, "Optimization problems for fast aam fitting in-the-wild," in *Proc. of the IEEE Int. Conf. on Computer Vision*, Sydney, NSW, Australia, pp. 593–600, 2013.

[47] Y. Liang, X. Wang, L. Zhang and Z. Wang, "A hierarchical framework for facial age estimation," *Mathematical Problems in Engineering*, vol. 2014, no. 6, pp. 1–8, 2014.

[48] P. Koruga, M. Bača and J. Ševa, "Application of modified anthropometric model in facial age estimation," in *Proc. ELMAR-2011*, Zadar, Croatia, pp. 17–20, 2011.

[49] J. Jagtap and M. Kokare, "Human age classification using facial skin aging features and artificial neural network," *Cognitive Systems Research*, vol. 40, pp. 116–128, 2016.

[50] S. Bekhouche, F. Dornaika, A. Benlamoudi, A. Ouafi and A. Taleb-Ahmed, "A comparative study of human facial age estimation: Handcrafted features vs. Deep features," *Multimedia Tools and Applications*, vol. 79, no. 35, pp. 26605–26622, 2020.

[51] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park and J. Kim, "Age estimation using a hierarchical classifier based on global and local facial features," *Pattern Recognition*, vol. 44, no. 6, pp. 1262–1281, 2011.

[52] Z. Chai, Z. Sun, H. Mendez-Vazquez, R. He and T. Tan, "Gabor ordinal measures for face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 1, pp. 14–26, 2014.

[53] M. Haghighat, S. Zonouz and M. Abdel-Mottaleb, "CloudID: Trustworthy cloud-based and cross-enterprise biometric identification," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7905–7916, 2015.

[54] A. Singh, Sonal and C. Kant, "Face and age recognition using three-dimensional discrete wavelet transform and rotational local binary pattern with radial basis function support vector machine method," *International Journal of Electrical Engineering & Education*, vol. 58, no. 2, pp. 0020720920988489, 2021.

[55] E. Torres, S. L. Granizo and M. Hernandez-Alvarez, "Gender and age classification based on human features to detect illicit activity in suspicious sites," in *2019 Int. Conf. on Computational Science and Computational Intelligence*, Las Vegas, NV, USA, pp. 416–419, 2019.

[56] P. C. S. Reddy, S. G. Rao, G. Sakthidharan and P. V. Rao, "Age grouping with central local binary pattern based structure co-occurrence features," in *2018 Int. Conf. on Smart Systems and Inventive Technology*, Tirunelveli, India, pp. 104–109, 2018.

[57] M. E. Malek, Z. Azimifar and R. Boostani, "Facial age estimation using zernike moments and multi-layer perceptron," in *2017 22nd Int. Conf. on Digital Signal Processing*, London, UK, pp. 1–5, 2017.

[58] X.-Q. Zeng, R. Xiang and H. X. Zou, "Partial least squares regression based facial age estimation," in *2017 IEEE Int. Conf. on Computational Science and Engineering and IEEE Int. Conf. on Embedded and Ubiquitous Computing*, Guangzhou, China, pp. 416–421, 2017.

[59] R. Angulu, J. R. Tapamo and A. O. Adewumi, "Age estimation with local statistical biologically inspired features," in *2017 Int. Conf. on Computational Science and Computational Intelligence*, Las Vegas, NV, USA, pp. 414–419, 2017.

[60] S. Silvia and S. H. Supangkat, "A literature review on facial age estimation researches," in *2020 Int. Conf. on ICT for Smart Society*, Bandung, Indonesia, pp. 1–6, 2020.

[61] W. Shen, Y. Guo, Y. Wang, K. Zhao, B. Wang *et al.,* "Deep differentiable random forests for age estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 2, pp. 404–419, 2021.

[62] H. Zhang, Y. Zhang and X. Geng, "Practical age estimation using deep label distribution learning," *Frontiers of Computer Science*, vol. 15, no. 3, pp. 1–6, 2021.

[63] Z. Shah and M. Roy, "Dcnn based dual assurance model for age estimation by analyzing facial images," in *2020 IEEE 17th India Council Int. Conf. (INDICON)*, New Delhi, India, pp. 1–6, 2020.

[64] J. C. Xie and C. M. Pun, "Deep and ordinal ensemble learning for human age estimation from facial images," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2361–2374, 2020.

[65] H.-T. Q. Bao and S.-T. Chung, "A light-weight gender/age estimation model based on multi-taking deep learning for an embedded system," in *Proc. of the Korea Information Processing Society Conf.*," in *Korea Information Processing Society*, pp. 483–486, 2020.

[66] H. Pan, H. Han, S. Shan and X. Chen, "Mean-variance loss for deep age estimation from a face," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, Utah, pp. 5285–5294, 2018.

[67] S. Feng, C. Lang, J. Feng, T. Wang and J. Luo, "Human facial age estimation by cost-sensitive label ranking and trace norm regularization," *IEEE Transactions on Multimedia*, vol. 19, no. 1, pp. 136–148, 2017.

[68] O. Guehairia, A. Ouamane, F. Dornaika and A. Taleb-Ahmed, "Deep random forest for facial age estimation based on face images," in *020 1st Int. Conf. on Communications, Control Systems and Signal Processing*, El Oued, Algeria, pp. 305–309, 2020.