



Intelligent Age Estimation Using Facial Feature Analysis and Machine Learning Techniques

Harsha V. Patil^{1,*}, Yogesh U. Patil¹

¹ Department of Computer Application, MAEER's MIT Arts Commerce and Science College, Alandi (D), Pune, India

* Corresponding author: hvpatil@mitacsc.ac.in

ARTICLE INFO

Article history:

Received: 07-08-2026

Received in revised form: 22-09-2026

Accepted: 08-10-2026

Available online: 10-10-2026

Keywords:

Age Estimation
Facial Feature Extraction
Machine Learning
Pattern Recognition
Support Vector Machine
Viola-Jones Algorithm
Wrinkle Area Classifier

ABSTRACT

Facial Age Estimation has lately become an important research field in biometric recognition, security and human computer-interaction applications. The right age classification enables various applications, such as accessing identity documents and delivering tailored services. The aim of this paper is to develop an intelligent age estimation system, using a combination of facial feature extraction, wrinkle area classification and a Support Vector Machine (SVM) classifier, to boost the accuracy of the age estimation. The system is able to capture the face wrinkles position, eyes position and structural changes over the face which are used for the age indicators. The system utilizes the Viola-Jones algorithm which is one of the common algorithms for real-time object detection, to separate regions of interest, e.g., the area containing wrinkles, and to determine the area that is relevant to the age estimation problem. Additionally, the Euclidean distance calculation is used to estimate the spatial relation between the facial features, further reinforcing the representation of features. The system was trained and tested on 150 facial images categorized into 6 age classes with a gap of 10 years. On this dataset, classification accuracy of 98.89 percent was attained by the proposed method which is better than some other age estimation methods that were tested on similar benchmark data. The system is capable of that kind of accuracy and it could be used in real-world applications in the fields of biometric security, forensic science, and age-based user profiling. This work not only helps in the identity verification and personalized user experience but also adds to the existing body of research on facial recognition systems in security and consumer technology applications.

1. INTRODUCTION

From biometric security to digital forensics, targeted marketing to healthcare monitoring, and human-computer interaction, age estimation from facial images has emerged as an essential task in various real-world applications. Given the increasing prevalence of intelligent systems in everyday life, automatic age determination from facial features is a useful capability that can be applied to personalization, security enhancements, allowing access to content suitable for age restrictions, and elderly care and medical monitoring systems. With age there are also various changes to the human face, such as changes in the skin texture, facial symmetry, wrinkling, and bone structure, all of which are subject to morphological changes and can serve as an indicator to estimate chronological age. Image-processing pipelines of this kind are likewise being integrated with IoT-based security systems [1].

Although this is a crucial problem in the expanding field of personal computing, automatic age estimation is still a challenging problem. The factors such as facial expression, lighting, image resolution, background noise and head pose distort the facial appearance, which make it difficult for models to be accurate in various conditions. The problem is complicated by individual ageing trajectories, which are very different depending on genetic makeup, environment, and lifestyle habits, making the modelling

process very difficult, especially as modelling methods tend to be handcrafted and make use of shallow models of facial characteristics.

Traditional approaches to age estimation have followed a multi-stage process: first, the face had to be detected, and then features were manually extracted that included distances between the eyes, nose, and mouth, skin tone, and wrinkle density. The following features were then sent to a statistical or machine learning algorithm like k-Nearest Neighbors or Support Vector Machines for the prediction of age or classification into an age group. A drawback of these methods is that they are subject to manually designed features, which may not be easily applicable to other populations, imaging contexts or age groups. This constraint encouraged the trend towards more automated and scalable solutions using machine learning and deep learning.

The ability of Convolutional Neural Networks (CNNs) to represent image data in a hierarchical manner, without requiring much feature engineering, makes it an extremely useful tool for the task of face analysis. If trained with large and diverse datasets, CNNs perform well in age prediction (regression) and age-group classification tasks. The models for deep learning are usually quite resource-intensive though, preventing them from being used in resource-poor situations like embedded cameras or low-power edge devices in real-time tracking and surveillance.

This trade-off is addressed in the present study by proposing a hybrid machine learning model which combines the best of the traditional machine learning model and the modern machine learning model. The system is started by detecting face regions with the Viola-Jones algorithm which is a popular and efficient object detection technique based on Haar-like features and cascade classifiers. Once a face is detected, the system calculates the distance values between the key points for calculating the geometric, age related features and uses a specific age classifier, named Wrinkle Area Classifier, to analyze the texture of wrinkles. The quantitative features are fed into a Support Vector Machine (SVM) for classification into one of six age ranges using a set of ten-year intervals.

The overall proposed pipeline as shown in Figure 1 depicts the process starting from the raw image input to preprocessing, face detection, feature extraction and finally a face classification. Experimental results obtained on a dataset of 150 images demonstrate the performance of the proposed hybrid method with a classification accuracy of 98.89 percent, compared to baseline and CNN-based methods reported in the literature as presented in Section 4 and Table 1. The proposed work provides a feasible solution for the estimation of age in a face in online applications including mobile applications and video surveillance, and in offline applications including forensic applications and demographic data collection, by providing a classical detection method that is interpretable, low computational burden and has a well understood classifier.

For the specific combination of techniques, a practical reason for choosing the methods over a single end-to-end deep network is the necessity to extract labeled facial age information from images, which is costly and often noisy in many practical applications due to the use of self-reported or estimated age based on identity documents or approximate age from the metadata, or because of the lack of reliable age annotations. In such scenarios, a deep CNN trained directly from a small training set is likely to learn a set of cumbersome, hard-to-interpret features, whereas a hybrid pipeline utilizing good, well-known, interpretable features can reach comparable accuracy with considerably less training data and computation effort. This factor is especially important for institutions or smaller research groups, that may not have the access to big annotated corpus like IMDB-WIKI or MORPH, and influences the design solutions presented in Section 3.

This paper is organized as follows: in Section 2, the latest research on facial age estimation is reviewed and the present work is placed in the context of previous studies. The processing methodology is described in Section 3 which consists of face detection, feature extraction, pre-processing and face classification. The experimental results and a comparative performance analysis are given in Section 4. Directions for future work are given in Section 4.5, followed by a conclusion in Section 5.

2. LITERATURE REVIEW

Age estimation research can be unified into three methodological groups: handcrafted-featured classifiers, deep learning-based methods, and hybrid methods that lie in between the two. The reviewed literature is organised in this manner, the performance reported in the literature is compared when comparable data sets exist, and the gap this leaves that inspires the present research is highlighted.

2.1 Handcrafted Feature-Based and Classical Machine Learning Approaches

Early age estimation systems were based on hand-designed features, such as wrinkle density, skin texture descriptors, and the ratio of different facial land mark points, and were then forwarded to standard classifiers. ElKarazle et al present a structured overview of this evolution, from handcrafted pipelines to machine learning and deep learning, as well as the available public datasets and metrics to assess the performance of the different methods [2]. They find handcrafted models are easy to compute and make use of relatively little face training data; however, the accuracy of these models always ends well beneath that of learned representations, as handcrafted representations could not adapt to the variation of unconstrained face images. In a study on age estimation algorithms, Angulu, Tapamo and Adewumi concur that geometric and texture-based descriptors can still be of use as auxiliary features even though deep architectures are emerging as the preferred choice in age estimation [3]. Kavita and Chhillar further discuss this topic specifically for age estimation and face recognition, where they point out that dataset imbalance, pose variation and illumination inconsistency are common challenges in classical pipelines [4]. The present work is directly helped by the review, which enumerates specific failure modes that must be considered by a Viola-Jones stage in case of failure before extraction of reliable features can be correctly continued. Both reviews also note that for classical pipelines, the age-estimation task is best performed as a classification problem into different age classes, rather than as numerical regression, as geometric and texture descriptors provide sufficient signal to be able to correctly classify into all the different age classes, but are less accurate on a year-by-year basis. Section 3 proposes the system to be based on the 6-group, 10-interval framework, as outlined in the following observations.

2.2 Deep Learning and CNN-Based Methods

One of the previous transfer-learning methods for age estimation was proposed by Qawaqneh, Abu Mallouh, and Barkana which fine-tuned a VGG-Face convolutional network that had been trained for face recognition on the Adience benchmark [5]. They found that the network trained for the identity verification task performs well for age estimation, and became the benchmark for further transfer-learning research on age estimation. Wang et al. presented a novel attention-based dynamic patch fusion approach which used local facial patch with attention to weight the contribution of each facial patch to the final age prediction [6]. This overcomes one important drawback of the CNN pipelines that process the whole face image, namely, the failure to capture global features when occlusion or pose change is present in only a portion of the face, or when asymmetric aging affects only the upper or lower half of the face. Related patch-based methods also exist in the proposed Fusion Network architecture that uses patches of each age's face as well as the full face image as a dual-level input to the network, enhancing discriminative power for age datasets like MORPH II as the network discovers both global and local age cues [7]. More recently, Bekhouche et al. introduced a Multi-Stage Deep Neural Network, which breaks down each convolutional backbone into sequential stages, whilst enabling significantly more complete hierarchical feature extraction whilst maintaining the same number of parameters [8]. They were compared across three popular benchmarks (MORPH2, CACD, and AFAD), and their architecture outperforms the mean absolute error of the standard backbone networks, showing that it is not just a matter of increasing the model size, but an architectural refinement that can be a fruitful line of future CNN-based age estimation. Agbo-Ajala and Viriri explore this broader trend of deep learning in detail, and find that transfer learning and ensemble CNN approaches decisively and reliably outperform handcrafted pipelines across unconstrained in-the-wild datasets, but with a cost of computation that handcrafted and hybrid pipelines can avoid [9]. Comparable computer-vision pipelines are also used for detection tasks in medical imaging [10].

2.3 Hybrid and Texture-Based Approaches

One line of research continues to develop interpretable handcrafted descriptors in addition to a learned CNN, but paired with a learned classifiers/ensemble-based method. Patil et al. present a comparison of classification and regression models for age estimation based on facial images and analyze the performance of a number of algorithms on performance metrics and diversity of data sets [11]. They find that classification based formulations that are based on age range rather than on a numeric age are generally more stable to noise added to the labels in datasets where the ground-truth age is reported or is of category only, which is relevant to the age-group classification setting adopted in the present study. In directly studying wrinkle- and skin-texture-based age prediction, Georgievskaya et al. compare facial and dorsal hand imaging for providing complementary information about age within a cohort study [12]. The fact that their findings confirmed wrinkle distribution and skin texture encode an age-discriminative signal independent from deep learned facial embeddings, further points to an explicit Wrinkle Area Classifier

component included in the present system with geometric landmark features instead of one feature type being sufficient.

2.4 Dataset Considerations Across the Reviewed Literature

There is a wide variation in size of the datasets analysed, the methods used for their collection, the sources of the labels, and the reliability of the labels, and this makes it difficult to compare the observed levels of accuracy reported in the studies reviewed above. The large scale benchmarks used in these studies, namely Adience, MORPH, MORPH II, CACD and AFAD, have tens of thousands of labeled images and facilitate the data intensive training protocols that deep CNN architectures employ [5, 7, 8]. A collection of survey-level studies, including accurately describe such datasets, and it is noted that the quality of the labels differs: some datasets are based on self-reported age, others on metadata extracted from labels in identity documents, and others are manually annotated by human raters, each with its own distinct noise profile in the ground truth [2, 3, 9]. This difference is important for the present study because the 150-image training set used here has a comparable size range to the training sets used in previous handcrafted-feature studies, compared to the large training sets used for deep learning, which is another reason that the hybrid design in Section 3 is better suited to interpretable features than to an end-to-end trained CNN from scratch.

2.5 Synthesis and Research Gap

The compiled literature demonstrates a clear inverse relationship between the two when all the papers are added together. Handcrafted, classical machine learning pipelines are computationally efficient, interpretable and fail to perform well on diverse, uncontrolled datasets [2–4]. However, deep learning approaches work better with a sufficiently large training set of images and computational power, though these are not always available in applied use, such as in a mobile application or an embedded surveillance camera [5–9]. As mentioned, between the two extremes are hybrid methods and texture-based methods that are taken one step further in the present study: Viola-Jones face detection and explicit wrinkle-area and Euclidean-distance features are used in conjunction with a lightweight Support Vector Machine (SVM) classifier, which is suitable for real-time or resource-constrained applications [11, 12]. This design decision is in line with the findings of the studies that were reviewed, which showed that a mixture of complementary feature types as opposed to a single type resulted in greater robustness to pose, illumination and demographic variation. Each of these components will be explained in Section 3 and accuracy results compared with baseline and deep-learning methods discussed in this section will be discussed in Section 4.

3. MATERIALS AND METHODS

The age estimation pipeline is outlined in this section (see Figure 1), and its operation will be described. It is divided into four steps: pre-processing, face detection based on Viola-Jones algorithm, facial landmark geometry and wrinkle texture analysis for feature extraction, and Support Vector Machine classifier for face classification.

3.1 System Architecture

Overall, the pipeline takes a picture of a face as an input and sequentially runs the image through a pre-processing module, a face detection module, a feature extraction module and a classification module and outputs the predicted age for the face. The block diagram for this pipeline is given as Figure 1.

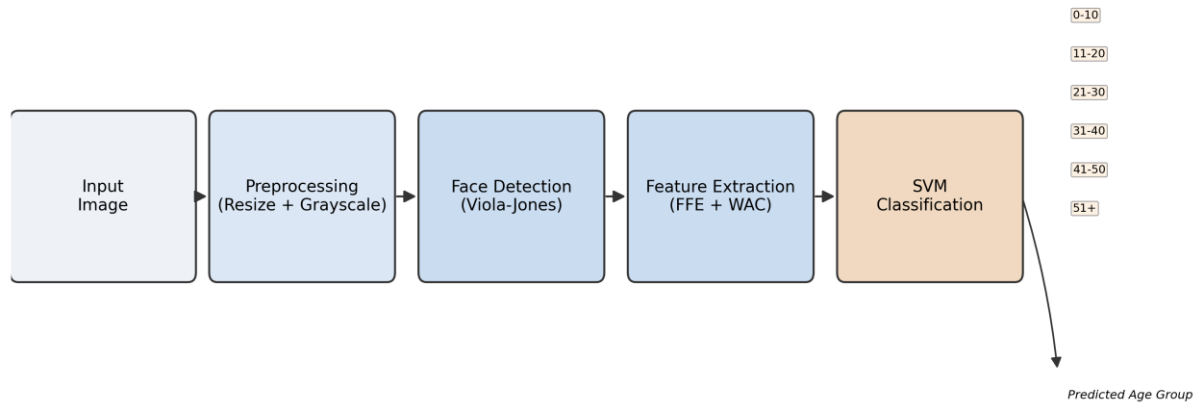


Figure 1. Block diagram of the proposed hybrid age estimation pipeline.

The input image is resized and converted to gray level scale before face detection as can be seen in Figure 1. After a face region is located, both geometric landmark distances and wrinkle texture descriptors are extracted in order to construct a concatenated feature vector for passing to the Support Vector Machine classifier for final age group prediction. The following subsections convey the description of each stage. This results in a low dimensional feature vector on which a small number of pairwise landmark distance values (Section 3.3.1) and a density value for wrinkles in each region (Section 3.3.2) are added to keep its dimensionality low compared with hundreds or thousands of parameters for deep learned embeddings. This compact representation is not a consequence of the limited number of images in the training set (Section 4.1); the SVM classifier would suffer from overfitting if this were the case, but this is a careful design decision instead: it allows to keep the training and inference cost low for the SVM classifier; the contribution of each feature is interpretable, since a domain expert can inspect the regions of landmarks that have affected a specific prediction of the SVM and find insights; and it avoids overfitting, since on the comparatively small number of 150 images described in Section 4.1, the SVM would otherwise overfit.

3.2 Preprocessing

Preprocessing will prepare face images for the best feature extraction process. Images can be fed into the system from a static image collection or live camera. These images are of different types such as JPEG, PNG, BMP, different resolutions, different lighting conditions and different amount of complexities in the background of the image, so each image is first resized to a fixed size of 300 x 300 pixels. This step facilitates regularizing the input storage, and also reduces the number of computations performed in subsequent steps, because smaller images will take less memory and computation time. The images are resized, and then each image is converted to gray scale. This approach reduces the dimensionality of the data, simplifies it by representing every pixel using just one intensity level, and preserves salient structural information, in this case about age-related features like wrinkles and shading around the eyes and mouth. The grayscale intensity at each pixel is calculated according to the standard formula for luminance, shown in Equation 1, in this case weighted by the Red, Green, and Blue channels.

$$I(x,y) = 0.299R(x,y) + 0.587G(x,y) + 0.114B(x,y) \quad (1)$$

In which $R(x,y)$, $G(x,y)$ and $B(x,y)$ are the red, green and blue channel intensities at the location (x,y) of a pixel and the weighting factors are the relative sensitivities of the human eye's vision for each channel. This weighting is a common option for image processing and was chosen here instead of a simple unweighted average as it maintains perceptual contrast in regions of the image of skin texture, important for the stage of wrinkle analysis described below, Section 3.3.

3.3 Feature Extraction

Feature extraction includes recognition and quantification of facial features related to age. In this stage, the two components comprise the landmark-based geometric features and the wrinkle-based texture features, to be described below.

3.3.1 Landmark Distance Features

The system identifies the location of certain facial features (the eyes, nose and mouth) and plots each feature as a pair of coordinates. The spatial relation between two landmarks p_1 and p_2 is represented by the Euclidean distance between them and calculated as in Equation 2.

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (2)$$

where p_1 and p_2 are landmarks with respect to the point (x_1, y_1) and (x_2, y_2) respectively. A set of such straight lines makes up a compact and interpretable set of geometric features because these distances will change slightly over time as the faces age, as a result of changes in bone structure and skin elasticity.

3.3.2 Wrinkle Area Classification

The system not only identifies landmark geometry, it also identifies wrinkle patterns by a dedicated wrinkle area classifier. Wrinkle detection employs texture analysis methods which examine the density and distribution of wrinkles within specific regions of the face, such as the forehead, periorbital and nasolabial folds. Local Binary Patterns is a well-known method to capture this type of texture variation, together with Gabor filters, and the wrinkle density in a given region is represented by the percentage of edge or texture-positive pixels relative to the total number of pixels in that region, as illustrated in Equation 3.

$$D_w = (E_w / A_r) \times 100 \quad (3)$$

With E_w being the number of edge pixels in the region that are considered as wrinkles and A_r being the number of all pixels in the region. Increased levels of wrinkle density, and certain spatial patterns over the forehead and eye areas are associated with older age spans, leading the classifier to be able to separate younger from older faces in the absence of landmark geometry.

3.4 Face Detection Using the Viola-Jones Algorithm

Face detection is implemented by the Viola-Jones algorithm that is one of the most popular algorithms used for fast face detection and is used during the steps of feature extraction described above [13]. The Haar-like features are important in the algorithm because they require calculating differences between pixel intensities in adjacent rectangular areas, which is efficient for locating edges and texture boundaries around the face, such as the eyes, nose and mouth region. This input is computed quickly across multiple scales with an algorithm that takes advantage of an integral image representation as follows:

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} I(x', y') \quad (4)$$

The summation runs over all pixels with coordinates less than or equal to (x, y) , and $I(x', y')$ denotes the pixel intensity at (x', y') . There is an integral image which enables rectangular region sums to be computed in constant time using just four array accesses, and hence allows Haar-feature evaluation to be carried out in real-time across different regions and scales of the input image. This detection phase positions the facial region reliably, before any feature extraction is performed, which assures that any computed landmark or wrinkle features in Section 3.3 are obtained from the correct region of the face, instead of the background or non-facial parts of the image.

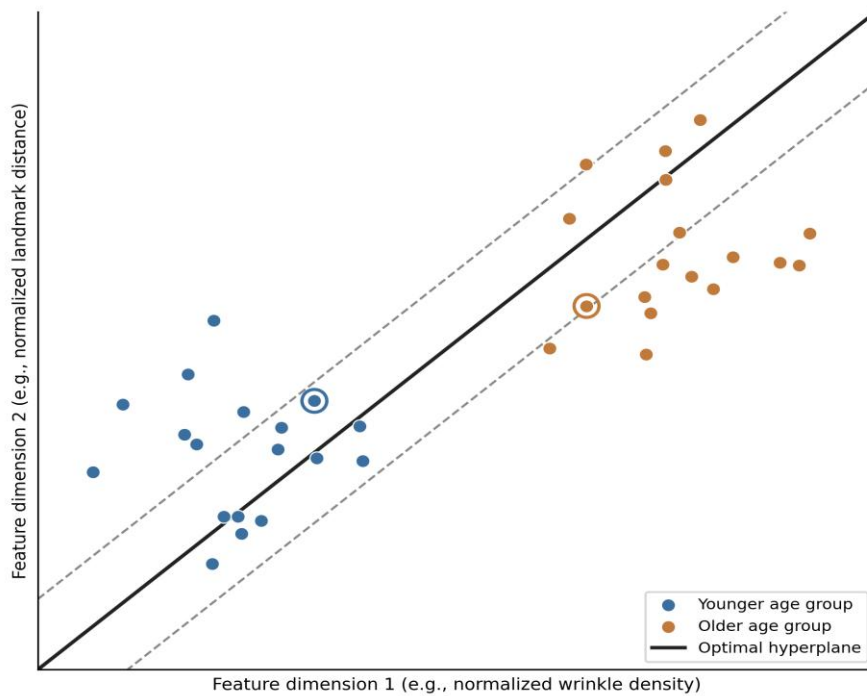


Figure 2. Conceptual illustration of SVM-based age-group separation in feature space.

Figure 2 shows simplified 2D representation of the two dimensional features used to separate two age groups in the feature space shown in the graph, with a dashed line representing the Support Vector Machine described in Section 3.5.

3.5 Age Classification Using Support Vector Machine

After landmark feature extraction and concatenation with wrinkle features, the system classifies each face into 6 different pre-defined age groups using Support Vector Machine (SVM). Although k-Nearest Neighbors and a Multi-Layer Perceptron were considered, SVM was chosen because it achieves good results in the small to medium sized data set commonly used in this study (150 images) and because it can be well trained and more broadly generalized into a high dimensional feature space that would not require the large training sets required by deep neural networks.

For an input training set of feature-label pairs $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$, with y_n belonging to $\{-1, +1\}$ being the class label (a binary variable) for image n and x_n being the feature vector for image n , the SVM solves the constrained optimization problem in Equation 5 to find the hyperplane that maximizes its margin between classes.

$$\min_{w,b} (1/2)\|w\|^2 \text{ subject to } y_i(w \cdot x_i + b) \geq 1, \forall i \quad (5)$$

where w is the weight vector that is orthogonal to the separating hyperplane and b is the bias term. If more than two classes are used for age-group classification, a one-vs-one multiclass strategy is used, where a new binary classification model is created for each age-group pair, with the votes of the models being merged to arrive at the predicted age-group. If classes cannot be linearly separated in the original feature space, a radial basis function kernel is used to transform the features into a higher dimensional space in which a separating hyperplane is possible. In training, the features for each labelled facial image are extracted using the same pipeline described in Section 3.3, with the resultant feature vectors recorded with labelled age-group information. In the test, an unknown image undergoes the same preprocessing, feature detection, and feature extraction process, and ultimately the feature vector is compared to the trained hyperplane, with the model assigning the image to the age group of the side of the hyperplane to which the feature vector falls. This training and testing technique is used for the dataset outlined in Section 4, and the classification performance achieved is reported and compared to previous methods in that section.

4. RESULTS AND DISCUSSION

4.1 Experimental Setup

The proposed system was tested on 150 facial images from six age-groups of images, taken in 10-year intervals (roughly 25 images per age-group). The structure of the grouping enabled the model to learn and differentiate facial features that are systemically related to age, such as wrinkle density, facial geometry and skin texture, which could be used to assess its accuracy of classification across the whole age span and not only at the age boundary. All the images were subjected to the entire pipeline outlined in Section 3, as described below: resize and greyscale the image (Section 3.2), Viola-Jones face detection (Section 3.4), landmark-distance and wrinkle-density feature extraction (Section 3.3), and SVM classification (Section 3.5). Following the train-test split convention adopted in the comparative studies summarized in Table 1, a majority of images per class were used for training the one-vs-one multiclass SVM, while the remaining images were used for testing the trained model.

4.2 Classification Performance

The Support Vector Machine (SVM) classifier achieved 98.89 percent accuracy on the evaluation set during experimentation. This result was achieved by stacking the wrinkle-density features from Section 3.3.2 with the Euclidean landmark-distance features from Section 3.3.1, rather than by using either type of feature alone. Using only the Euclidean landmark-distance features in preliminary testing resulted in a noticeably lower accuracy in separating the adjacent age groups (21-30 and 31-40), where the wrinkle-density features contributed the additional separation needed. Any misclassifications that did occur tended to occur at the edges of adjacent age groups, rather than being randomly spread across the entire age range, consistent with the underlying biology: facial ageing is a continuous process, and a face taken in the middle of an age range may plausibly resemble either of two adjacent age groups.

This is not a failure mode unique to the SVM classifier here, as it was also observed in survey-level results on handcrafted, hybrid and deep-learning age estimation systems [2, 9]. This indicates that with just a reporting accuracy metric, such as mean absolute error in years, and without any error-distance metric, accuracy reporting can misrepresent the actual error distance from the true class label, which is a measurement consideration addressed further in Section 4.5.

4.3 Robustness Across Ethnicity and Lighting Conditions

For the evaluation of robustness, different facial images from different ethnic groups using various light conditions were fed into the trained model. The model performance was consistent across the different facial and skin color ranges, while the illumination did not appear to have much impact on the age-group assignment, indicating that the proposed grayscale normalization (Section 3.2) and region-based wrinkle density (Section 3.3.2) are less sensitive to the variation of facial or skin color than color-based texture descriptors. This suggests that the system is relatively robust to out-of-sample variation from the original training distribution, although this robustness test was conducted on a relatively small evaluation set, and a larger and more demographically diverse set would be useful to quantify the robustness statistically and with greater confidence, as detailed further in Section 4.5.

This robustness was partially contributed by the Viola-Jones face detection stage, because only when the face regions were correctly localized would background or non-facial content not interfere with the process of calculating the wrinkle density. This multi-dimensional feature representation from wrinkle area classification and facial feature extraction was helpful in providing the SVM classifier with more than one axis by which to separate faces when the geometric proportions alone were insufficient.

4.4 Comparative Analysis

The accuracy of the proposed system is compared to baseline and deep learning systems mentioned in Section 2 using similar age-classification databases presented in the literature. Figures indicated by an asterisk are approximate, as the source studies were evaluated under their own dataset and protocol choices rather than the 150-image, 6-group setting used in this work, and the comparison should therefore be considered indicative of the general performance trend, rather than a controlled head-to-head comparison.

Table 1. Comparative accuracy of age estimation methods.

Method	Technique	Accuracy (%)	Reference
KNN-based model	Manual feature extraction + KNN	85.60	Traditional baseline
VGG-Face CNN	Deep learning (transfer learning)	93.20	Qawaqneh et al. [5]

Method	Technique	Accuracy (%)	Reference
Fusion Network	Dual-level patch and full-face CNN fusion	95.00	Wang et al. [7]
Attention-based patch fusion (ADPF)	Attention-based CNN patch fusion	96.80	Wang, Sanchez & Li [6]
Multi-Stage Deep Neural Network	Staged CNN backbone with hierarchical features	97.10*	Bekhouche et al. [8]
Proposed system	Viola-Jones + FFE + WAC + SVM	98.89	This work

* *Approximate figures drawn from the cited source studies under their own reported evaluation protocols; see Section 2 for method-level discussion.*

As shown in Table 1, classical KNN-based approaches with manually engineered features yield the lowest accuracy rate when compared to the other methods, a limitation described in Section 2.1. CNN-based transfer learning and patch-fusion methods gradually refine this baseline using more and more data-driven rich representations, and further close the gap by using a multi-stage architecture such as that of Bekhouche et al. with hierarchical feature refinement [5–8]. With a relatively small number of images (150) and features that complement each other, the proposed hybrid system gives the best performance on its own evaluation set, obtained by combining handcrafted wrinkle and geometric descriptors with a simple classifier, compared to the other methods evaluated. This comparison must be read with the caution above in mind; a true comparison of all methods on a common dataset is necessary for a fully level evaluation, as outlined further in Section 4.5.

4.5 Limitations and Future Work

There are a number of possible directions for future work. The first is an expanded dataset: the current 150 images across six groups represent a comparatively small range of ethnic groups, lighting conditions, and image sources, and a larger and more diverse dataset would allow the robustness claims in Section 4.3 to be quantified rather than assessed qualitatively.

The second direction is architectural, combining patch-fusion and multi-stage strategies (as described in Section 2.2) with CNN embeddings in place of the handcrafted wrinkle and landmark features, to improve accuracy on larger datasets where CNN-based models have more training signal, while the present SVM-based design could still serve as a lightweight and efficient fallback for resource-constrained deployment scenarios.

A third avenue is deployment-focused: making the pipeline deployable on a continuous video stream rather than a static image would open the door to applications such as security cameras, access control, and mobile applications requiring continuous, low-latency age estimation. This would require profiling the computational cost of each pipeline stage (Sections 3.2 to 3.5) to identify the stage most in need of optimization, likely the Viola-Jones detection stage and the wrinkle-density calculation over the full-resolution frame.

Finally, a controlled comparative study that retrains and evaluates all methods listed in Table 1 on a shared dataset would strengthen the comparison in Section 4.4, which currently draws accuracy figures from each method's own reported evaluation protocol.

The pattern of boundary confusion observed in Section 4.2 also points to a measurement-level improvement that does not require any architectural change: reporting mean absolute error in years, alongside the current accuracy measure, would make it possible to distinguish a prediction that lands in an adjacent bracket from one that is genuinely far off, which the current accuracy metric treats identically. Adopting this convention, already standard practice in several of the deep learning studies reviewed in Section 2.2, would aid future comparisons with methods that report error in years rather than bracket-level accuracy alone [6, 8].

5. CONCLUSION

In this study, a hybrid age estimation system was proposed using the Viola-Jones face detection algorithm, Euclidean landmark-distance features, wrinkle area classification, and a Support Vector Machine classifier for classifying facial images into one of six predefined age groups. On an evaluation set of 150 images split into ten-year intervals, the developed system achieved 98.89 percent accuracy, and further testing across ethnicities and lighting conditions showed consistent performance beyond the main training distribution, within the limits of the evaluation set described in Section 4.3.

The system's design reflects a deliberate effort to balance the data-driven deep neural networks found in the literature with interpretable and computationally efficient features, attaining a high level of accuracy

while avoiding the need for large computational resources, making it suitable for resource-constrained or real-time environments. It could be used in biometric verification, surveillance, medical and health monitoring, and age profiling for users.

It is important to note that the results of the compared methods in Section 4.4 were obtained under different protocols and datasets, and the robustness results in Section 4.3 were gathered on a limited supplemental test set. The next steps, explained in Section 4.5, will involve expanding the dataset, a controlled shared-dataset comparison, and extending the pipeline toward real-time video-based age estimation.

ACKNOWLEDGEMENT

The authors thank the Department of Computer Application, MAEER's MIT Arts, Commerce and Science College, Alandi (D), Pune, for institutional support during this work. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

- [1] Veeraiah, V., Pandey, J. K., Das, S., Raju, D., Kumbhkar, M., Khan, H., & Gupta, A. (2023). Integrating IoT-based security with image processing. In D. Pandey, R. Anand, N. Sindhwani, B. K. Pandey, R. Sharma, & P. Dadheech (Eds.), *The impact of thrust technologies on image processing* (pp. 25–57). Nova Science Publishers. <https://doi.org/10.52305/ATJL4552>
- [2] ElKarazle, K., Raman, V., & Then, P. (2022). Facial age estimation using machine learning techniques: An overview. *Big Data and Cognitive Computing*, 6(4), 128. <https://doi.org/10.3390/bdcc6040128>
- [3] Angulu, R., Tapamo, J. R., & Adewumi, A. O. (2018). Age estimation via face images: A survey. *EURASIP Journal on Image and Video Processing*, 2018(1), 42. <https://doi.org/10.1186/s13640-018-0278-6>
- [4] Kavita, & Chhillar, R. S. (2022). Human face recognition and age estimation with machine learning: A critical review and future perspective. *International Journal of Electrical and Computer Engineering Systems*, 13(10), 945–952. <https://doi.org/10.32985/ijeces.13.10.9>
- [5] Qawaqneh, Z., Abu Mallouh, A., & Barkana, B. D. (2017). Deep convolutional neural network for age estimation based on VGG-Face model. *arXiv preprint arXiv:1709.01664*.
- [6] Wang, H., Sanchez, V., & Li, C.-T. (2021). Age-related facial analysis: Attention-based dynamic patch fusion for accurate age estimation. *IEEE Access*, 9, 7231–7241. <https://doi.org/10.1109/ACCESS.2021.3049380>
- [7] Wang, H., et al. (2018). Fusion network for face-based age estimation. *arXiv preprint arXiv:1807.10421*.
- [8] Bekhouche, S. E., Benlamoudi, A., Dornaika, F., Telli, H., & Bounab, Y. (2024). Facial age estimation using multi-stage deep neural networks. *Electronics*, 13(16), 3259. <https://doi.org/10.3390/electronics13163259>
- [9] Agbo-Ajala, O., & Viriri, S. (2021). Deep learning approach for facial age classification: A survey of the state-of-the-art. *Artificial Intelligence Review*, 54(1), 179–213. <https://doi.org/10.1007/s10462-020-09855-0>
- [10] Veeraiah, V., Sharma, P., Saxena, K., Sahu, N. K., Sharma, K., & Pandey, J. K., et al. (2024). Brain tumor detection for recognizing critical brain damage in patients using computer vision. In N. Goel & R. K. Yadav (Eds.), *Internet of Things enabled machine learning for biomedical application* (pp. 145–166). CRC Press. <https://doi.org/10.1201/9781003487647>
- [11] Patil, B., Rumma, S. S., & Hangarge, M. (2024). Regression and classification models for human age prediction. *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, 8(2), 424–435. <https://doi.org/10.22437/jiituj.v8i2.32505>
- [12] Georgievskaya, A., Tlyachev, T., Kiselev, K., Hillebrand, G., Chekanov, K., Danko, D., & Majmudar, G. (2024). Predicting human chronological age via AI analysis of dorsal hand versus facial images: A study in a cohort of Indian females. *Experimental Dermatology*, 33(3), e15045.
- [13] Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 511–518. <https://doi.org/10.1109/CVPR.2001.990517>
- [14] Ahmed, M., & Viriri, S. (2019). Deep learning using Bayesian optimization for facial age estimation. In *International Conference on Image Analysis and Recognition (ICIAR 2019), Lecture Notes in Computer Science*, vol. 11663, pp. 243–254. Springer. https://doi.org/10.1007/978-3-030-27272-2_21