International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V71I5P243

Volume 71 Issue 5, 435-460, May 2023

Review Article

Age Invariant Face Recognition Techniques : A Survey on the Recent Developments, Challenges and Potential Future Directions

Sonia Mittal[1], Sanskruti Patel[2]

[1]*Department of CSE, Institute of Technology, Nirma University, Gujarat, India*
[1,2]*Smt.Chandaben Mohanbhai Patel Institute of Computer Applications, CHARUSAT, Changa, Gujarat, India,*

[1]*Corresponding Author : sonia.mittal@nirmauni.ac.in*

Received: 23 February 2023 Revised: 07 May 2023 Accepted: 13 May 2023 Published: 25 May 2023

Abstract *- Face recognition has been a central research topic in recent years. Because it has many different application areas in automatic surveillance systems, passport control, missing person detection and criminal detection, many researchers have studied the problems of face recognition and authentication for many years. However, still, many issues remain to be addressed. The limitation of this topic includes the variability in poses, facial expressions, illumination, and a large age gap between the face images. This paper presents the survey for some commonly used techniques for Age Invariant face recognition (AIFR). A comprehensive overview of the most used age-invariant methods in face recognition is presented, with a comparative overview of different approaches and a comparison in accuracy. Some facial ageing datasets, with their essential characteristics in images, no. of subjects, and age gap, are briefly presented in a tabular view. Commonly used techniques for feature extraction and classification in this field are presented. Major topics covered are experimental results obtained from these methods, issues/challenges, the scope of future work, and conclusions.*

Keywords *- Age invariant face recognition, Cross age face recognition, Age progression, Convolutional Neural Network, Deep learning discriminative method, Generative method.*

1. Introduction

Depending upon a person's surroundings, habits, etc., a person's face changes in many different ways as they age. Even with many irregularities and dissimilarities, humans are pretty good at recognising the face of the same person at different ages. Because regardless of the exact way a person ages, certain general facial features (or a pattern) would not change significantly despite changes in facial appearance due to the ageing process. The relative effect of these changes also depends on the Age being considered, e.g., face structure-like shape variations affecting facial bones and muscles are more in the younger generation means in children up to 18 years. In contrast, texture changes are more prominent in adults [1].

Figure 1 shows two examples of how faces can change in approximately 05 to 40 years. In this paper, we present a comprehensive literature survey in the area of Age Invariant Face Recognition (AIFR) methods. Each article is classified by feature extraction method, feature classification methods, and dataset used. In the field of face recognition, along with other variations like illuminations, different poses, and expressions aging is one of the criteria which have been studied and experimented with extensively in the last few years by many researchers. There are broadly two problems in the field of face recognition.

Fig. 1 Some images from the FG-NET dataset [2] clearly show intra-class variations (e.g. illumination, expression, pose and aging)

First is the recognition of images of the same person at different ages; that is, the correct classification of a test image by a trained classification model. The second is the estimation of a person's Age, which is the calculation of a person's Age in a test image by a face recognition/estimation system. These two challenges are typically addressed with three types of methods: Generative, Discriminative, and Deep Neural Networks (DNN). For all three methods, the models are trained to learn (or derive) key features that are invariant over time/aging. These are then used for the identification of test or query images. In the discriminative methods, the focus is on using the invariant features to classify test images. In the generative methods, the invariant features are used to generate images at various ages, and then these are used to match with the test images. Generally, discriminative methods have been used for classification and generative methods for age estimation. Deep neural network-based methods are used for both classification and age estimation purposes. Human face images consist of many pixels. If the size of the image is n*m, then the dimension of the face is n*m; it is approximately 100 or 1000 dimensions or many more, depending on the size of the face image. So face image data sets are considered a high dimensional dataset. The face images with variability like illumination, pose, expression, and age gap are not linearly separable. The principal component analysis (PCA) technique does not perform well on this type of variable face dataset for classification/recognition types of problems, as PCA linearly separates the data point and is a linear classifier. Therefore face images are non-linearly separable. Here in Figure 2, the visualisation of face images of the same and different persons are shown in high dimensions.

As described above, face images lie on high dimensions and are non-linearly separable. Therefore, it becomes a difficult task to classify the face images. The main focus of any classification problem is extracting the discriminative features from the training datasets, which a classifier uses for recognition and verification purposes.

This paper is divided into 8 sections. Section 2 covers the related work in this domain, the good review papers are studied, and their contribution to the AIFR domain is shared. Section 3 comprises two subsections. In subsection 3.1, the feature extraction and classification techniques in discriminative methods are described. Then in subsection 3.2, the detailed review of papers on discriminative methods is covered. Section 4 is divided into 2 subsections. In subsection 4.1, the modeling of generative methods is explained, and in subsection 4.2, the key research papers are reviewed.

Similarly, section 5 is also divided into two subsections. Subsection 5.1 describes the feature extraction and classification techniques of deep neural networks (DNN), and in subsection 5.2, the various research papers in the field of DNN are briefly covered. The rank-1 recognition rate of these methods is presented in the respective methods section. In section 6, we have covered some issues related to the data sets for this domain and some challenges in developing the model. Section 7 includes the discussion and future direction in the area of the Face recognition problem domain. In the last section, 8, the conclusion of the review work is presented. In Table 1, some good survey papers are summarised on Age Invariant Face recognition methods. Table 2 highlights the difference between the three models: discriminative, generative and DNN are given with respect to their functionalities for quick reference. In Table 3, the various loss functions used in DNN are shown. Ass this review has considered all three approaches of AIFR, a summary is presented in Table 4, including their performance parameters, feature extraction, classification details, the dataset used, and corresponding recognition accuracy. In Table 5, the usage of various feature descriptors and classification techniques found in the different papers is given. Table 6 gives some critical information about well-established face invariant datasets like the number of persons, the average number of images per person, the aging periods of the persons, etc.

So, as mentioned in previous paragraphs, due to aging, pose variation, illumination changes, and the nonlinear distribution of face images, classification or clustering of facial images becomes difficult. To address these problems, a comprehensive review of the age-invariant face domain is conducted, and various insightful findings are presented. In this review, the main goal is to analyze various techniques at each stage of the classification problem.

Contribution of this review paper:

- A systematic overview of the development of generative models (GM) and discriminant models (DM) is presented.
- Detailed study of the various feature extraction and classification methods used for DM and GM is covered.
- The existing good survey papers in this domain are covered with their highlights about core findings.

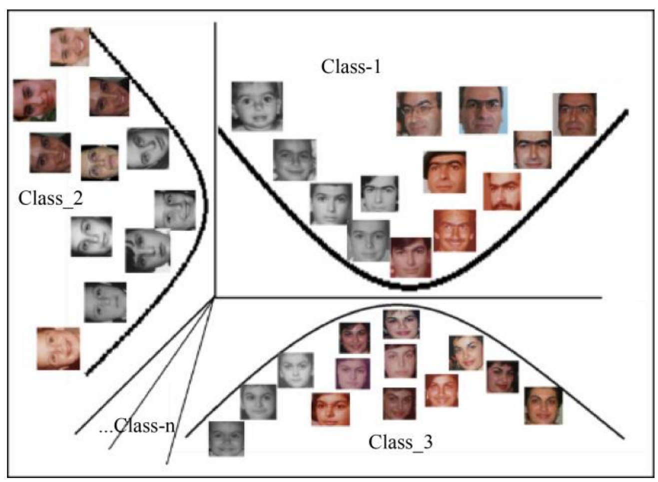


Fig. 2 Visualisation of face images in non-linear subspaces. Also, the figure shows face images in single-class and various classes in multi-dimension and non-linear subspaces

- The latest models based on Deep neural networks (DNN) are presented briefly with the network architecture functioning.
- The overview of various loss functions is presented used by DNN models.
- Key highlights of some good papers are presented in tabular form for comparative analysis.

- Various face aging data sets with the parameters like no. of images, no. of subjects, year, age gap, and number of images per subject.
- Future scope in this domain.

2. Related Work

Many researchers have presented review work in the Invariant face recognition domain during the last decade. Review criteria and parameters considered are different by each researcher. Here in this section, a description of some existing review work is presented. Table I summarises the highlights of some of the good review papers in this domain covering from 2009 to 2021. Ramanathan et al. [3] have presented Thompson's study [Shaw et al., [4]] of morphological changes associated with growth in biological forms in human faces by introducing a framework to embed biological forms such that it is used for studying transformations of different morphogenetic events. With inspired by Thompson's work, Shaw et al. [4] have explained facial growth with the help of mathematical conversions to characterize the facial growth event. For the outer contour of faces, two transformations, namely Cardioidal strain and Affine shear, is discovered and used.

- Cardioidal strain: Strain by which the face stretches downward and outward.
- Affine shear: When applied in the right proportion, it introduces a bulging and flattening in the jaw and a backward slant in the forehead.

The cardioidal strain transformation model was the central model in the initial days. Researchers have revisited this model and developed some variants of this transformation to describe the process of facial growth over the years. Further, the paper has included contributions from the Computer Vision domain like the feature selection process for shape/texture, 2D and 3D modeling for invariant aging. The Paper has categorised various approaches for this domain like Subspace-based modeling, 2D or 3D-based models, Machine learning-based modeling, and Image Feature-based modelling. The reviewers also presented a brief overview of facial anthropometric studies, which provide dense measurements of key facial landmarks in individuals of different ages and may play an important role in modeling the facial aging process. Various face-ageing datasets are also discussed in detail. Masi, Iacopo, et al. [5] have divided the paper into four parts, namely:

- The training process of the state-of-the-art (SOTA) systems and description of various public data sets used in the survey
- Face pre-processing part (detection, alignment, etc.)
- Architecture and loss functions used for transfer learning
- Face recognition for verification and identification and conclusion

In the paper, the authors have covered general face recognition problems. General face data sets include different types of variations like face rotation, different illumination effects, blurring effects due to motion in videos, face with different emotions, etc. The authors also discussed different types of images in the various face datasets like CASIAWebFace, VGGFace, UMDFaces, MS-Celeb-1M, VGGFace2, and IMDb-Face. The survey concludes with an overview of the SOTA results at a glance and some open issues currently overlooked by the community. According to Mortezaie et al. [6], research in the Age progression domain is classified into three groups: Age estimation, Age simulation, and Age-invariant face recognition approach.

In the first group of methods, Age is estimated when inputting the person's facial image. In the second group, a probabilistic or heuristic computational model is used to model the facial parameter, which captures the facial appearance across the various Age and hence learns the ageing pattern. In the third group, age-invariant features are extracted, with discriminative competency for robust face recognition. Further, the authors have divided age-invariant face recognition methods into three categories: generative, non-generative (discriminative), and the recently commonly used deep learning-based methods. In the following review paper from Sawant et al. [7], the authors have divided the paper into three parts: Facial Ageing datasets, Age Invariant Face Recognition methods, and the third part, the effect of ageing on the performance of face recognition systems. The authors have described the face aging datasets in detail, like FERET, FGNET, MORPH, Cross-age celebrity dataset (CACD), Pinellas County Sheriff's Office Longitudinal Study (PCSO_LS), and WhoIsIt (WIT). The authors have investigated the functioning of generative methods in detail and described the framework for 2D and 3D aging patterns. The active appearance model (AAM) (Cootes et al. 1995, 2001) has been found to be a widely used technique for learning both shape and appearance features, and principal component analysis (PCA) (Turk and Pentland 1991) enables dimensionality reduction.

So PCA coefficients of AAM are used for face analysis in low dimensions. AAM is also used by Lanitis et al. (2002) to describe craniofacial growth and gray level differences at different age periods. This review article also presents a discriminating approach to solving the AIFR problem. They described descriptions of age-invariant features such as ageing-induced face drift, gradient-oriented pyramid (GOP), scale shift (SIFT), local binary pattern (LBP), multiscale local binary pattern (MLBP) and presented a detailed overview of the various discriminative methods. Further, the Deep neural network is also covered with the description of the convolutional neural network (CNN) functionality with the discriminative approach for AIFR and the Deep learning-based generative approach using an auto-encoder for face reconstruction and face verification is explained. K. Baruni et al. [8] have also briefly described the facial aging datasets and three methods generative, discriminative and deep neural networks with their merits and demerits. It has given a survey of discriminative and deep learning-based methods and discussed the various feature fusion techniques used in various papers. Mei Wang and Weihong Deng [9] have focused mainly on deep neural network techniques for face recognition/verification and the age estimation domain. Extensive studies are done for the various loss functions like center loss, marginal loss, contrastive loss, softmax loss, and stochastic gradient descent(SGD) loss.

The various CNN architectures are discussed, and the evolution of CNN architectures from Alexnet, VGGNet, DeepFace, ResNet, GoogleNet, SENet, and the latest VGGFace2 have been discussed with their characteristics related to network architecture, discriminative and robustness for the face verification task. Abdulabas, Marem H, et al. [10] have divided the Face Aging Progression techniques broadly into two methods: Traditional and Deep learning based methods. Further, these two categories are divided into subcategories.

Traditional methods are divided into two models, the first model is based on physical models, and the second model covers prototype-based approaches. The physical model includes parameterization to reproduce muscle, skin, skull and other structural changes in the facial image. This model is computationally expensive because the model parameters lack generality; As a result, each face must be retrained. Another method is a prototype-based model, which creates a face with intermediate features (prototypes) of predetermined age ranges of facial features. Sometimes Personality specific characteristics are missing in this method. The Deep learning-based methods are divided into three subcategories: Translation, Sequence, and Condition-based. In Translation methods, Generative adversarial network (GAN) based models are covered. In Sequence, Recurrent Neural Networks (RNN) based models are used. And in condition-based approaches, add target age labels to the network as additional information to manage age formulation. And GAN and auto-encoders are used. An extensive review of deep learning-based approaches includes a short description of aging face datasets.

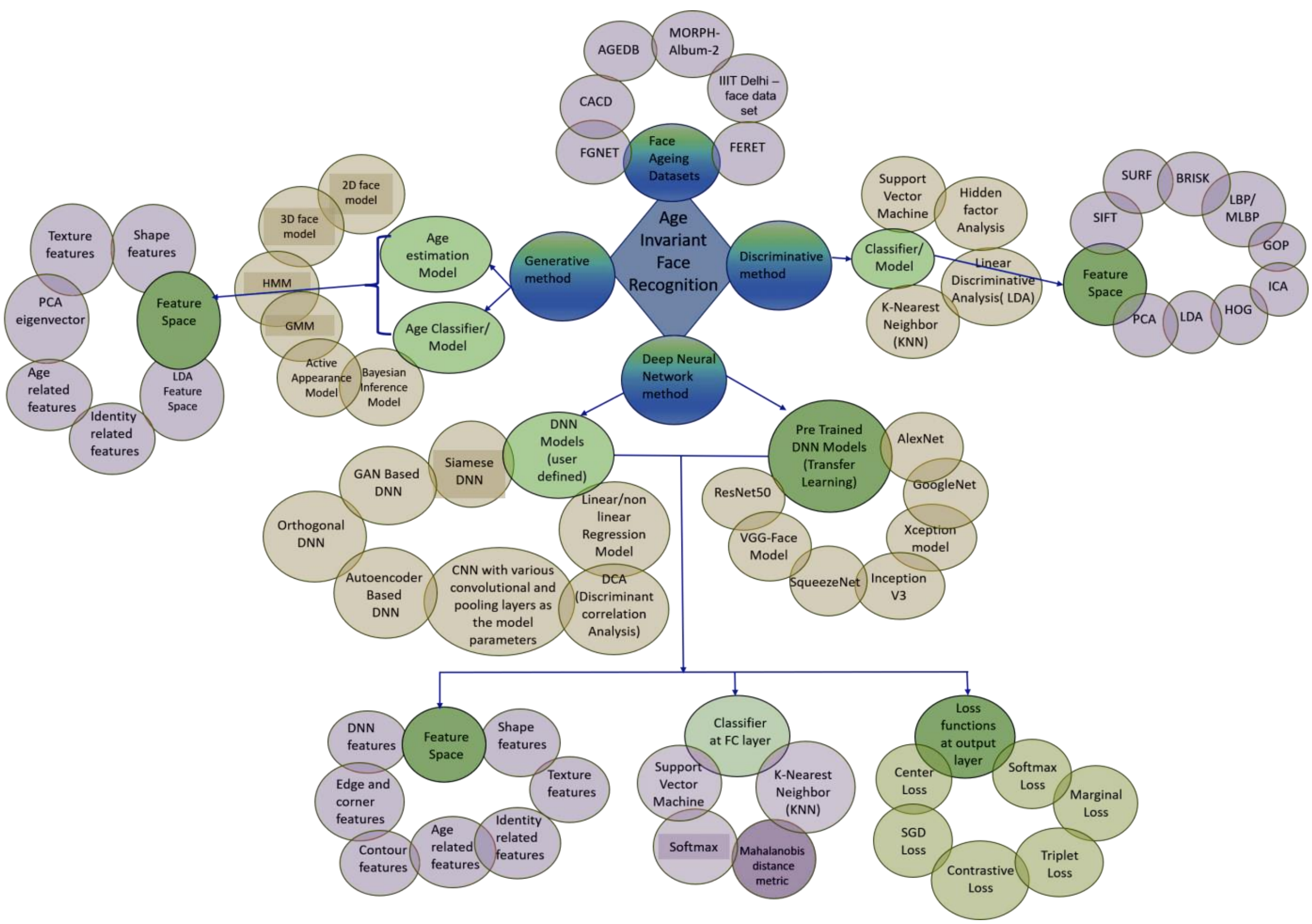


Fig. 3 Graphical View of Age Invariant Face Recognition Taxonomy.

Table 1. Survey papers on Age-Invariant face recognition and face recognition methods

Paper	Year	Core concept	Methods included
Ramanathan et.al. [3]	2009	Morphological changes associated with the growth of the human face are modeled, and also covered mathematical transformations like Cardioidal strain and Affine shear are for facial growth.	The survey includes papers based on computer vision, like subspace learning, machine learning, and feature-driven methodologies.
Masi, Iacopo, et al.[5]	2018	The various face datasets are discussed. Different face alignment and data augmentations methods are described.	The survey covers the papers based on Deep neural network architecture-based techniques. The authors have presented a comprehensive survey of the various loss functions.
Mortezaie et.al. [6]	2019	Performance comparisons on Various datasets related to age-invariant FR are presented.	Survey based on three discriminative, generative, and deep learning methods for age-invariant face recognition.
Sawant, et. Al.[7]	2019	It discusses about the biological changes due to age progression on the human face structure and skin.	Three categories of methods for Face recognition are covered, and the performance results are presented for various aging datasets.
K. Baruni et. al. [8]	2021	The advantages and shortfalls are covered with a discussion on pre-processing methods through which the performance has improved.	Survey based on three discriminative, generative, and deep learning methods for age-invariant face recognition.
Mei Wang, Weihong Deng [9]	2021	Various loss functions, feature extraction, and classification techniques are covered for DNN. And face invariance like illumination, pose, facial emotions, and aging effect is studied in the survey.	The focus is on Deep neural network-based techniques only. All types of datasets used in face recognition (FR) applications have been covered.
Abdulabas, Marem H, et. al. [10]	2022	Face Aging Progression techniques are broadly divided into two methods, which are based on Traditional and Deep learning. Further, these two are divided into various subcategories.	The brief functionalities of Prototype, Translation, Sequence and Condition-based methods are given. And the review of all types of categories is considered, including a short description of aging face datasets.

2.1. Conclusion of Related Work

After reviewing the various survey papers in the Age invariant face recognition (AIFR) domain, which presents the taxonomy of AIFR, figure 3 gives a comprehensive graphical view of the AIFR taxonomy for better understanding. The AIFR has four major frontiers, the face aging datasets, which have face images that consist of variations in many dimensions along with aging, and the three methods: generative, discriminative, and deep neural network. The figure is self-explanatory. The generative methods address two types of problems Age estimation and classification/recognition. In this category, various researchers described and used, e.g. hidden Markov model (HMM), gaussian mixture model (GMM), 2D face model, 3D face model, active appearance model (AAM), and bayesian inference models.

Similarly, in the Discriminative category, support vector machine (SVM), k-nearest neighbor (K-NN), hidden factor analysis(HFA), and linear discriminative analysis (LDA) were the models used extensively. The deep neural network(DNN) category is divided into two parts. Some users have defined and designed their own DNN with combinations of various layers like convolutional layer, pooling layers and activation functions etc., and pre-trained DNN is also used prominently by many researchers with some customization at the last layers, i.e. classification and loss function. All three models have their own feature space and are shown in the figure with different features, and some are common, like shape, texture, edge, and contour. The detailed study is covered in the respective method sections.

3. Discriminative Methods (DM)

3.1. Feature Extraction and Classification Techniques in Discriminative Methods

The key features for the Generative and Discriminative methods are extracted from the datasets, which can be categorised as either local or global. Local or regional features are related to a local patch of the image, e.g. in a facial image, it can be like nose, left eye, right eye, mouth, forehead, and so on. Therefore, these local face patches describe the person's unique facial characteristics. Whereas the global features are extracted from the entire image. E.g., the texture of the image or the shape of the image. It can be obtained by considering some visual properties of the whole image or concatenating the regional features. Global features are useful for object detection, while local features are useful for detecting and identifying objects.

The classification of local and global features is shown in Figure 4. Many methods are available to extract local and global features from the images; some of the prominent local key features are Scale-invariant feature transform (SIFT), Speeded robust features (SURF), Binary Robust invariant scalable key points (BRISK), Local Binary Pattern (LBP). SIFT [11] features extract salient points and describe the gradient properties around the salient point. SURF [12] is based on a hessian blob detector. BRISK [13] is known for its rotation invariant characteristics and faster computation than SIFT and SURF for hard real-time applications. It takes a concentric circular shape to sample pixels to determine the gradients, and these local gradients of pair of circles are used to calculate the angle of the key points. LBP [14] is a very efficient visual feature descriptor to describe the image's texture.

It compares the intensity value of all surrounding pixels to the middle pixel and thresholds it. LBP is used to describe the local spatial pattern. Figure 5 shows the LBP face generated from the input face image. This LBP face has texture information which will be useful for classification. LBP shows invariance property against illumination, which means if the image is undergoing monotonic grey scale change, LBP computed faces eliminate those variances caused by this. There are many variants used by the researchers. One of the variants of LBP is multi-scale local binary patterns (MLBP)[15]; extended LBP is used to compute the LBP descriptor at different radii, e.g. if we take four, then 1, 3, 5, 7 radii are used. Over-Complete Local Binary Patterns (OCBP) takes overlapping patches in horizontal and vertical directions and extracts uniform pattern from all the blocks; modified Local Binary Patterns (mLBP) uses a 3*3 window, i.e., 9-pixels average values are compared to the current pixel's intensity value.

Many more LBP variants are available in the literature, providing efficient spatial descriptor information. A complete survey of LBP variants is available in [16]. Just like, local features are used to describe the tiny details of the image, global features are also useful to get the overall characteristics of the image. It is useful to denote the class or category of the images. Some of the important global key feature descriptors are Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Histogram of gradients (HOG), Independent Component Analysis (ICA), and so on. In machine learning, PCA and LDA have also considered dimensionality reduction techniques. PCA [17], [18] preserves important features from the data. PCA is a technique to reduce features by discarding correlated features from the datasets and preserving only concise and important uncorrelated features. PCA is also known as an unsupervised technique for feature selection and extraction. LDA [19] is known to be a supervised feature extraction technique to classify the data.

LDA finds the subspace such that the distance will be a minimum within-class scatter, and the distance will be maximised between classes. In a nutshell, the aim of LDA is to find the lower dimensional subspace when projecting the high dimension data to the lower dimension, the class discrimination information is preserved, and the dimensions of the original data get reduced. Hence both PCA and LDA are used for the linear transformation of feature space for dimensionality reduction. PCA ignores the class labels and determines the data set's orthogonal directions of maximum variance. It is known as principal components. And LDA computes the directions, i.e. axes, whereby projecting the different class data provides the maximum separability. Independent Component Analysis (ICA) is also considered a linear transformation technique to extract statistically independent features, which are a mixture of various signals [20]. Many researchers have used ICA successfully in many problem domains like an electroencephalogram (EEG), Magnetoencephalography (MEG), and hidden pattern extraction from financial data and face recognition. PCA can separate linearly pair-wise pixels correlations or dependencies in face images, i.e., PCA finds maximum variance directions, whereas ICA can find the correlations in higher-order statistical subspaces.

This is why ICA is known as a generalisation of PCA [21]. For the face images, the facial characteristics are available at higher dimensions; in PCA, there are chances to lose some of the nonlinear information, which can be extracted and preserved through ICA. Histogram of gradient (HOG) [22] is a feature descriptor used to extract the features from the images, but into the corresponding bins, instead of the frequency of gradients. The various intensities across the block will be normalised. Now the concatenation of the histogram of local blocks is combined as a HOG descriptor for the regional level. SIFT feature descriptor, as discussed earlier, is a local feature used to describe the salient point in a local patch and compute the gradients. Still, SIFT is scale-invariant feature points obtained using the difference of gaussian (DoG) keypoint detector. After the feature extraction, the next step in the recognition task is to apply the feature selection techniques to select the correlated data and remove the noise from the features. Some researchers have directly used the extracted features described in the previous steps for the model classification. Some researchers used bag-of-words, random feature selection techniques, PCA, or fusion of the various features extracted from different methods for the final feature selection. The next step is to use these features to train the classifier model to predict the results. As face recognition is supervised machine learning, the predicted output will be compared with the actual label, and the model is evaluated by computing the prediction error. The parameter updation is done based on the parameter feedback. This training process is executed in terms of batches of inputs. The whole training data set is divided into batches. And the error is evaluated across the batches. When this process is completed for the whole cycle of the input dataset, it is termed one epoch. This process continues with the number of iterations till the error is nullified or is less than some threshold value. This is applicable to discriminative and deep learning-based methods. In generative methods, the flow is different.

The features are extracted to reconstruct the face image for a specific age. In any technique, the face identity and its variations are kept apart, like Age, illumination, rotation, or scale, by means of extracted features and model classifier. The very prominent and robust technique found in literature and used by many researchers is the Support Vector Machine (SVM). SVM [23], [24], [25] is a classification technique used extensively for face images. It attempts to map the data to the high dimensional system even when data is not linearly separable by dividing the data through hyperplane. It is more effective when no. of dimensions are more as compared to no. of instances. SVM has been used for classification in combinations with features extracted by PCA, LDA, Gabor features, or even Deep learning features. SVM can be successfully used for multi-class classification problems also.

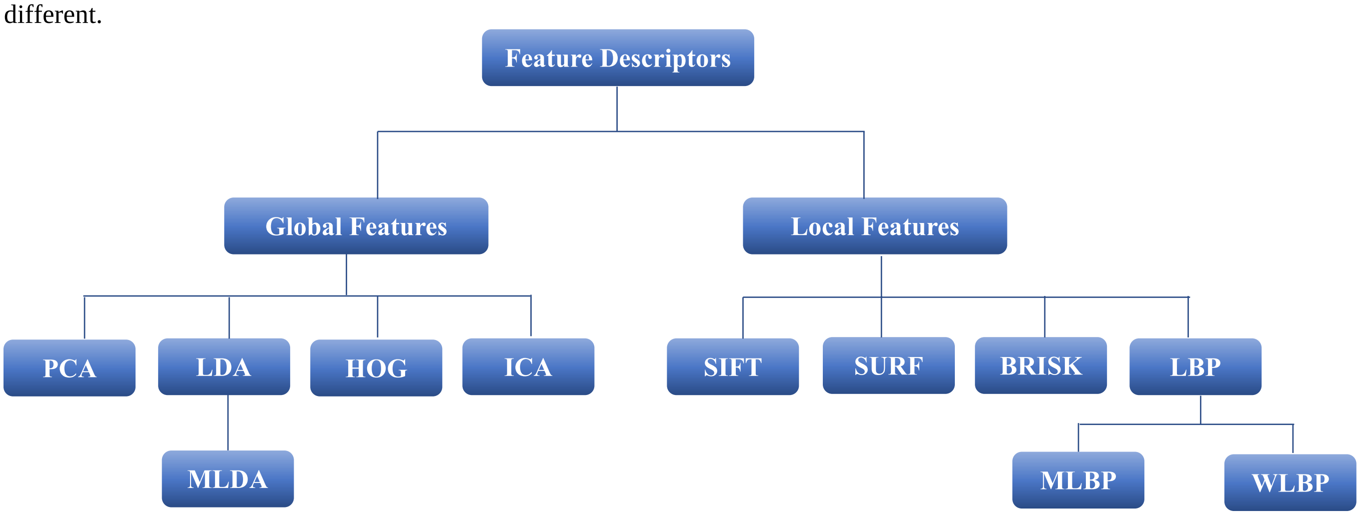


Fig. 4 Classification of var ious global and local features

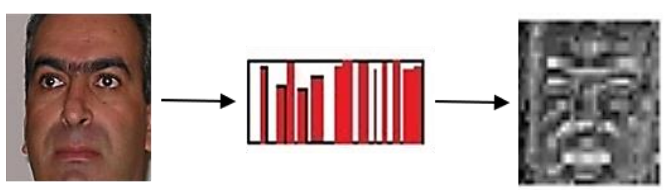


Fig. 5 LBP face generation from the input face image

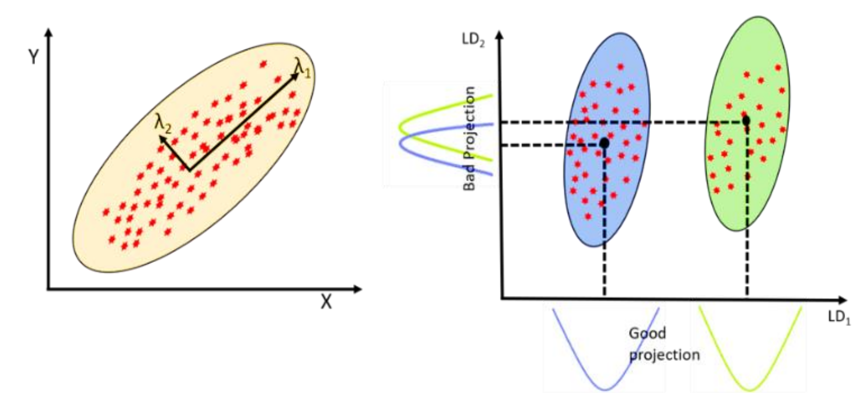


Fig. 6 Principal Component Analysis (PCA) (left), Linear Discriminant Analysis (LDA) (right).

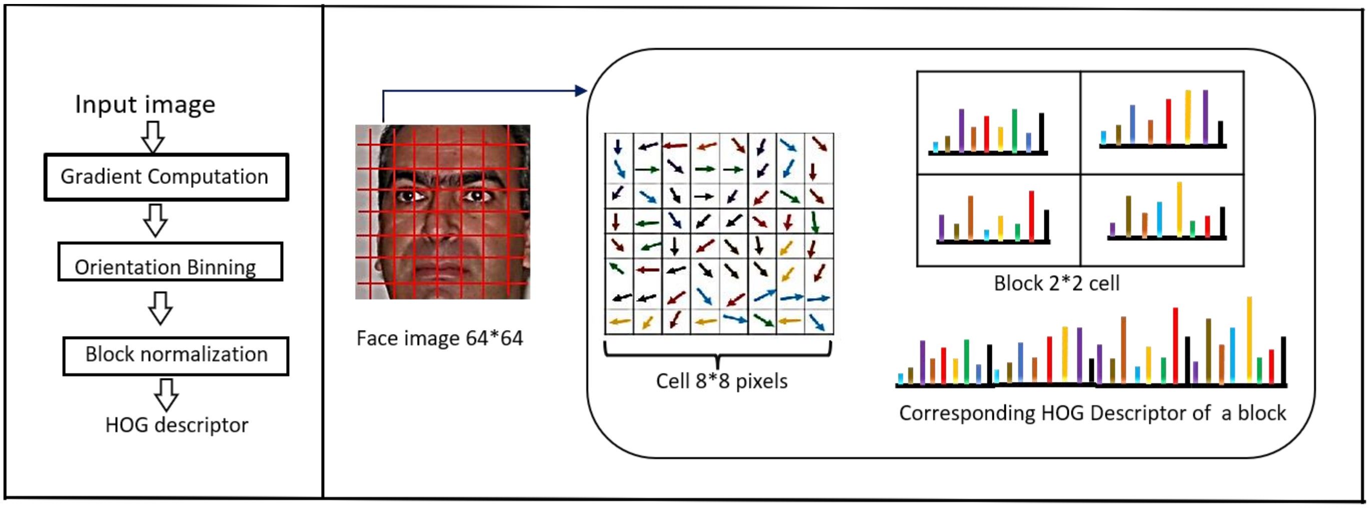


Fig. 7 Flow diagram of the various stages of HOG feature construction

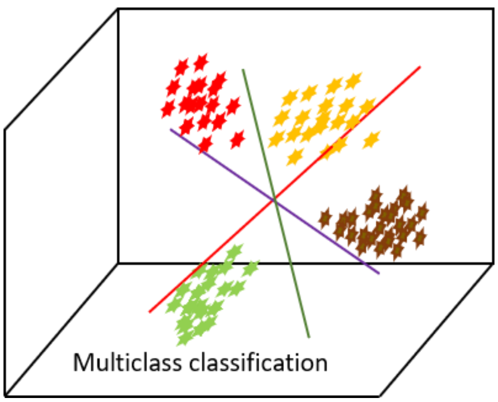


Fig. 8 Multiclass classification

The virtualization of Multi-class classification can be seen in Figure 8. From the diagram, it is clear that in multi-dimension, the different data classes can be separated with hyper-planes. The other classifier is a k-nearest-neighbour (KNN), the simplest machine learning classifier for supervised learning. KNN uses similarity metrics to classify the new data point. The new data point will be allocated to the class to which it is near compared to other classes. Many researchers use KNN [26], [27], [28] to classify face images with different types of features like PCA, LDA, Gussian-based, or self-organising maps (SOM). KNN is a non-parametric technique because it makes no assumptions for underlying datasets like the Gaussian mixture model. Some of the disadvantages of KNN is it will not work with large datasets and high dimension datasets. It is sensitive to noisy and missing data.

The next multi-class classifier is the softmax function. It takes n real or integer value as input and converts it into probabilities value between 0 to 1. It converts a small value into a small probability and a large value into a large probability. Hence it can be used for multiclass classification. It is a generalisation of the sigmoid function, which is used for logistic regression. Deep learning networks use the softamx function as the last layer for classification purposes. And sigmoid function is used for binary classification problems. These two functions are used primarily as activation functions in deep neural networks. A detailed comparison of activation functions of Deep neural networks is found in [29].

3.2. Review work on Discriminative Methods

Discriminative methods do not rely on face modeling as compared with generative methods. In discriminative approaches, the first hand, local and global features are extracted from the faces, such as Gradient Orientation Pyramid (GOP), Scale-Invariant Feature Transform (SIFT) [11], and Multi-Scale Local Binary Pattern (MLBP) [14], [15]. And various classification methods can be used to match the appropriate class further. Figure 9 shows the steps of a standard discriminative model for face recognition. Face features are extracted, the classifier uses these features to classify the training data, and error is estimated by comparing it with the actual output label. The error propagated back to the classifier model.

In discriminative models usually, researchers have used Linear discriminant Analysis (LDA), Discriminative subspace learning, Hidden factor analysis, Bacterial foraging method, Support vector Machine (SVM), and K nearest neighbors (KNN) for the classification of test data. Z. Li et al. [30] used the multi-feature discriminant analysis method to refine the feature space and enhance recognition performance. They used scale-invariant feature transform (SIFT) and Multidimensional Local Binary Patterns (MLBP) to describe multi-feature discriminate. These two local feature spaces are processed in a unified framework using the MFDA method (multi-feature discriminant analysis). MFDA is an improvement of LDA. The performance report of this method with recognition results on the FG-NET dataset is 47.0%, whereas good results for the MORPH dataset with 83.9%. D. Sungatullina et al. [31] have used multiple local feature descriptors like Scale Invariant Feature Transform (SIFT), Local Binary Pattern (LBP), and Gradient Orientation Pyramid (GOP) for each face image. PCA is employed for dimensional reduction. Then, by projecting three features, a hidden discriminant subspace is learned. The method is repeated until the controller estimates the optimised parameters of the hidden subspace such that when multiple features are projected into the hidden discriminant subspace, where the intra-class variability of each feature is minimized, and the interclass variability of each feature and the different features of the same individual is maximized correlation.

This approach uses more than one local descriptor; thus, more discriminative information is extracted. After projecting those features, an optimised detection model is obtained by finding a common hidden subspace. The performance of this method is good, with 91.8% recognition results on the FG-NET dataset, whereas the average results for the MORPH dataset with 65.2%. However, this approach requires more computational resources to determine the between-class dispersion and use the within-class dispersion matrices in the hidden subspace. BC Chen et al. [32] used a knowledge-based approach and introduced a new coding method called Cross-Aged Reference Coding. Here, it uses images of different people as a reference set and encodes the local features of the test image into an n-dimensional feature. This method converts local features into age-invariant representations. As a result, two images of individuals of the same and different ages have similar features in the new comparison condition. Therefore, this method achieves high face recognition and retrieval accuracy among different age groups. On the MORPH dataset, it gives 92.8% recognition results. D. Yadav et al. [33] present a fusion algorithm for bacterial foraging.

In this paper, the local binary pattern (LBP) feature is calculated for different regions of the face, the whole face, the right and left periocular region, the binocular region and the mouth region. The LBP plots of these five regions are now compared using the chi-square distance between the two images to calculate a goodness-of-fit score. A further bacterial-based aging method is used to calculate the weight of each face region. This weighted sum is used for classification. On the FG-NET dataset, the recognition results achieved is 64.5%. S. Biswas et al. [34] studied the effect of aging on human faces and observed that the appearance of the human face changes coherently. SIFT measures the coherency in feature drifts and presents a new feature descriptor. The pattern of coherency in the feature drifts is used to classify the two images. The same subject's coherent pattern exhibits similarity, whereas a different subject's coherency shows a different pattern. If the two images belong to the same subject, drifts in features follow a coherent pattern which may not be the case if the images belong to different subjects. H. Ling et al. [35] analysed the ageing effects on human faces. In this paper, gradient orientation is used as it exhibits illumination invariant properties. The Gradient oriented pyramid (GOP) is used for face descriptors. And SVM is used for the classification of test faces. A robust and discriminative face descriptor is designed by utilizing only gradient and discarding the magnitude of the image. The combination of GOP and SVM gives the best results for FG-NET datasets. Dihong Gong et al. [36] presented the hidden factor analysis (HFA) model. The discriminant information in this model is learned through the probabilistic model with two latent factors: an age-invariant identity factor and an age factor affected by the aging process.

The Expectation Maximization (EM) approach will determine the model parameters. In general, the convergence of the EM model is both very slow and prone to convergence to local optima. In this study, the initialization is done to converge in fewer iterations. Also, Age and identity are modeled separately, and Histograms of Oriented Gradients (HOG) are used as local feature descriptors. On FG-NET, it gives 49% recognition results. It can be tested using another local feature descriptor like LBP or combined to improve performance. Si, Junyong et al. [37] have extended the work of [36] further to improve the performance of the Hidden factor analysis model by including the progressive feature parameter and presenting the feature progressing model.

In [36], age-related and identity-related subspaces are learned. But this EM model is used to optimize for age-invariant identity components, meaning identity preserving. Whereas not much emphasis is given to age-component modeling. Hence considering this [37] divided the human Age into some m groups. For each age group, a separate subspace is learned as compared to [36], in which a single subspace across all age groups is considered. [37] has given the concept of age progression features of face images through m age group. In their model, subspace decomposition represents progressive changes in facial features rather than a single age-invariant identity subspace across all age groups.

The method uses stable characteristics during the aging period and gradual changes in function at each age. Therefore, this method overcomes the performance problem of previous HFA and other discrimination methods with small age differences and noise. A partial Least Square is used to extract the latent factor of input facial features corresponding to each age group (m). On Morph - II, it gives 89.74%, and on FG-NET, it provides 74.7% recognition accuracy, respectively. Xiaonan Hou et al. [38] proposed a robust feature coding method to map the original feature to an age-invariant state. On a spatial map, they constrain the mean characteristics of different age distributions as closely as possible in age-invariant space. Large-scale LBP is used as a feature descriptor, and PCA is for dimensionality reduction.

3.3. Conclusion of the Review Work on Discriminative Methods

Here it can conclude that discriminative approaches are divided into two phases first, local and global features extraction by means of various techniques. And secondly, it uses a classification technique to classify the test sample. In AIFR first low-level feature space is to be learned. The number of features used by researchers is generally more than one to build a strong discriminant feature space.

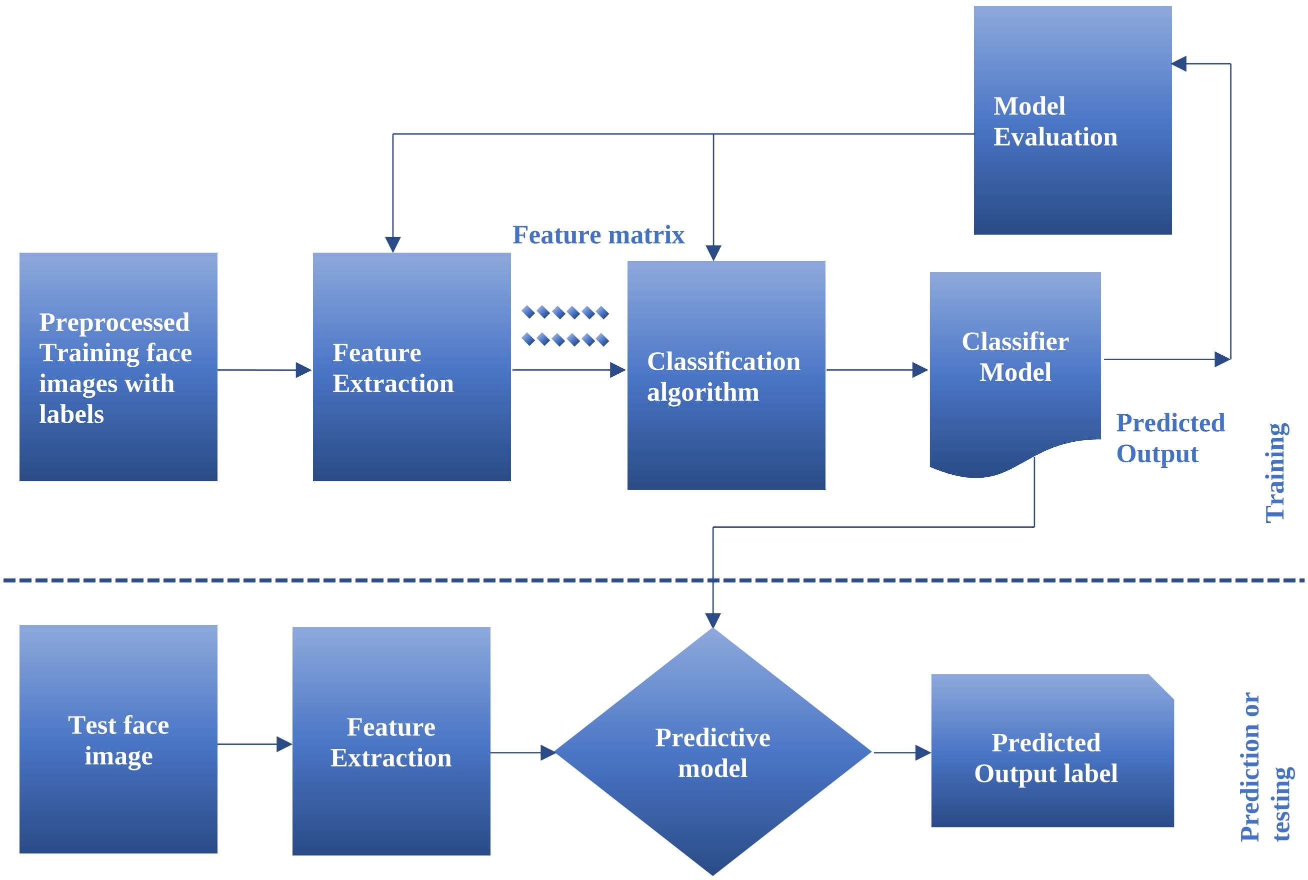


Fig. 9 Various steps in the discriminative model for face recognition

Hence, features are combined with some fusion techniques to learn joint discriminant feature space. And the training images will be projected to these feature spaces to learn corresponding low-level feature spaces. This redundancy would also be removed, e.g. with some random sampling method etc. So that computation time and memory storage can be used optimally. Finally, a classification technique is used to classify the test sample based on unique features per class.

Discrimination methods have advantages and disadvantages. These methods rely on robust feature descriptors and discriminatory learning/classification methods. Discrimination methods are generally less computationally intensive and provide optimal results quickly. The model's accuracy is better than the generative method whenever the training data is large. This is because discriminant models converge to smaller asymptotic errors. So, as the number of training examples increases, the error rate decreases for the discriminative models.

4. Generative Methods (GM)

4.1. Feature Space and Modeling of Generative Methods

In generative methods, 2D or 3D face models are synthesized by incorporating personal identity and aging property. Aging models can compensate for the aging process in face matching or age estimation. This method is widely used in age estimation problems. But researchers have used it for facial aging classification problems as well.

Some of the steps are given in Figure 10. Generally, this model primarily uses two underlying models, shape and texture aging, at given Age. From the training datasets, the features for texture and shapes are extracted for various age groups. And then used for fitting shape and texture model parameters by some algorithms; in the literature survey, some methods like the Gaussian mixture model, Hidden Markov, PCA eigenvectors, Bayesian inference or LDA feature subspace were mentioned. The matching process results will determine the error in the model parameters.

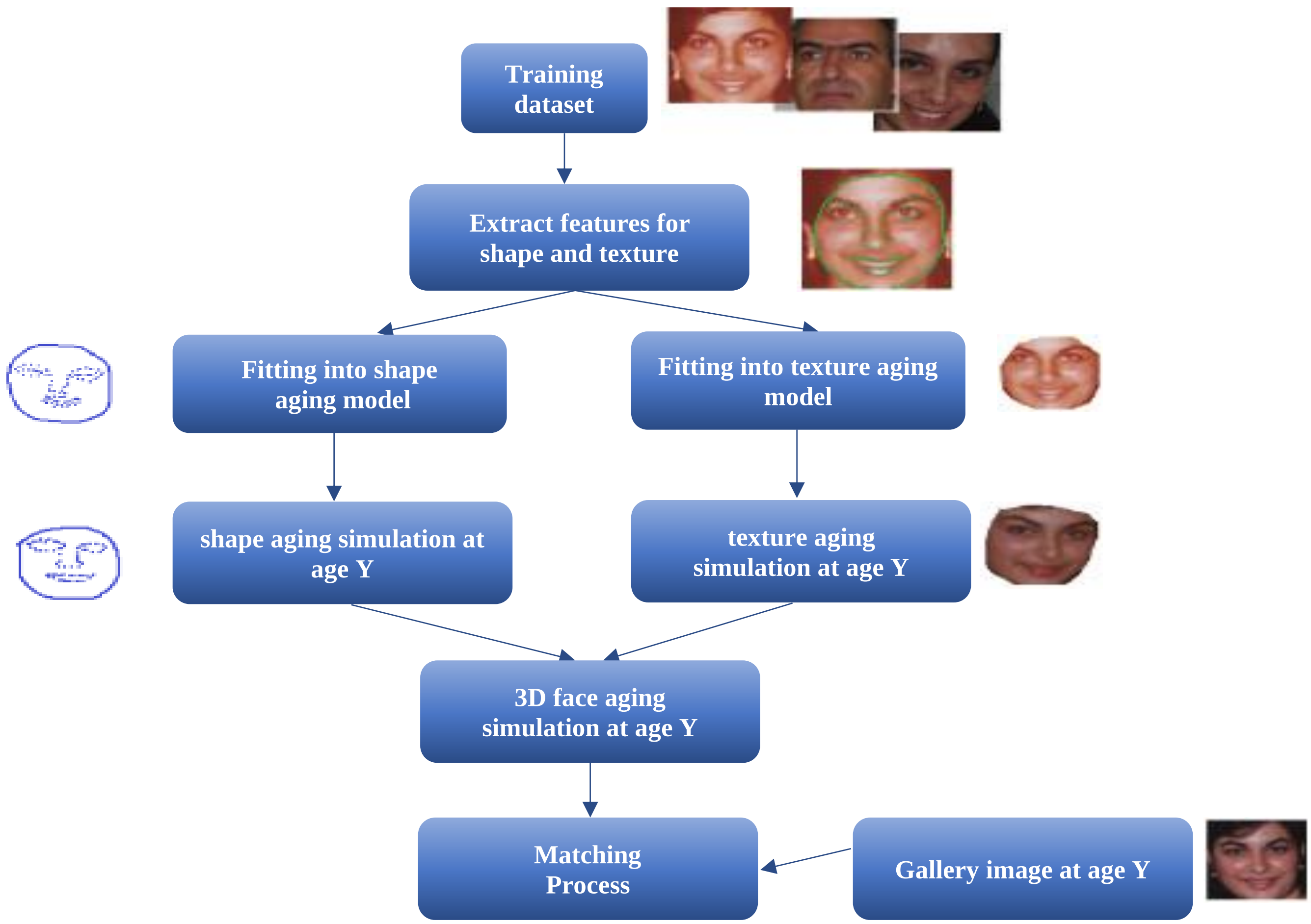


Fig. 10 Process of generative model for face recognition at various stages

The process will be iterated to some threshold error value such that the model will converge as per the penalty function. Here we are giving the approach covered by Lanitis et al.[1] in their paper to model face aging. In Lanitis et al.[1] statistical model parameters and mean example (X_m) are learned from training, and they have used the following equations to present their model[1]:

$$X = X_m + P\,b \tag{1}$$

Where X is the training example, X_m is the mean example, P is the matrix of eigenvectors, and b is a vector of weights. The weights are computed from Equation 1 as follows:

$$b = P^{-1}(X - X_m) \tag{2}$$

Parameter b and quadratic function are used for age estimation. The aging function is defined as,

$$agef = f(b) \tag{3}$$

Here f is the aging function, and agef is the estimated age calculated by f. The aging function represents the relationship between the face image and the actual age. The appearance of a person of a certain age is reconstructed using

$$b = f^{-1}(age) \tag{4}$$

The model parameters b are calculated and classified using the Mahalanobis distance between the probe and gallery images for a given test image. Some of the research in this area is explained in the following subsection.

4.2. Review work on Generative Methods

A. Lanitis et al. [1] studied the effects of aging on facial appearance and built an age transformation model for unseen images that can accurately estimate Age. U. Park et al. [39] present a 3D face aging model and simulation method for aging face detection. They studied 2D facial images and converted them into 3D models, and modeling the shape and texture change during aging is presented as follows: They modeled the effects of aging in two different age groups, i.e. under 18 and over 18. For the MORPH and FG-NET datasets, it gives recognition of 66.4% and 37.4%, respectively. H. Zhou et al. [40] presented a variational model of human identity and aging using probabilistic LDA. Different approaches are used to obtain human identity features combined with maximised correlations for face recognition.

Impressive results are obtained from MORPH and FG-NET datasets, 95.62% and 88.23%. J. Suo et al. [41] presented a component and dynamic method for modeling human facial aging. The And-OR diagram is constructed to represent facial and identity characteristics and simulates the aging pattern of each age group. In this way, a hierarchical AND-OR diagram is modeled, where the "And" nodes are used to represent composite details of facial features, and the "OR" branch simulates large gaps in facial features. The parse graph is computed from an observed image by Bayesian inference from coarse to fine. Further, face aging is modeled as a Markov process on the parse graph. Finally, they have proposed a face aging simulation and prediction algorithm. J. Wang et al. [42] presented an age simulation in which the shapes and textures of the human face were changed from the initial Age to the target age using shape and texture vectors.

They derived a polynomial relationship between the feature vector of shape and texture and the corresponding Age. By adapting a 3D morphable model to a 2D aging face image, U. Park et al. [43] presented a 3D model to learn the aging patterns of shape and the corresponding texture in the 3D domain. In Geng et al. [44] paper, the key idea is to model global aging patterns and personality-based facial aging patterns. Both feature vectors are combined for face reconstruction and age estimation. For the FG-NET dataset, the mean absolute error reported for the age estimation task is 7.5%.

4.3. Conclusion of the Review Work on Generative Methods

The above section discusses approaches based on generative methods. Fewer approaches to reproductive methods have been described in the literature. Ideally, generative methods model the distribution of individual classes (i.e. images belonging to a single individual). Mittal et al.,[86] In a generative approach, 2D or 3D face modeling is performed based on training datasets. Face modeling requires strong parametric assumptions, which may not meet the requirements when the number of classes and images in the class increases. As datasets grow with variations in facial images in more than one dimension, such as aging, lighting effects, expressions and poses, it is difficult for generative modeling to handle these variations together. In a generative model, the hypothesis made for selecting parameters will often not successfully model the whole range of variations.

In comparison with discriminative models, generative models converge faster. Thus, generative models are preferable when training datasets are small. Even though the generative models converge faster, they converge to a higher asymptotic error, which means the error rate may increase with the number of datasets. Discriminative methods and generative methods are different in how they utilise the input data points X and label Y to predict the new data point label. A discriminant model tries to optimise the prediction of Y (the class label) from X (the input test data), while a generative model tries to optimise the joint prediction of X and Y . This is why discriminative models work better—generative models in conditional prediction tasks. But at the same time, it requires several optimisation techniques and is sensitive to the initialisation of input parameters. Examples of generative models are Naïve Bayes classifier, generative adversarial network (GAN), and Discriminative models are linear regression, support vector machine (SVM), etc. A comparison of these two models is summarized in Table 2, along with the Deep neural network model (DNN) comparison. However, the generative method's important and strong application area is the face reconstruction for the Age for which the face image is unavailable. This is the limitation of the discriminative methods; the GM can be used for age estimation and age modeling, whereas the DM methods are useful for recognition tasks only.

5. Deep Neural Network (DNN) Based Methods

5.1. DNN Feature Extraction and Classification Techniques

In traditional approaches to extract the low-level and high-level features from the images generally, handcrafted features like canny edge detectors and Harris corner detectors are used for edges and corner detection, respectively. Histogram of oriented gradients (HOG) and statistical methods are used for blob detection, and finally, objects have been detected by mixing all these together in one or some other ways. The problem with these features is that nothing is guaranteed about how many corners and edges are good for classifying faces or objects in general. This way of extracting the feature is called manual feature engineering. With the advent of fast Graphical Processing Units (GPU), the face recognition field has also explored and experimented with Deep neural networks (DNN) and Convolutional neural networks (CNN). In this section, some of the CNN and DNN architecture which have been used for Age invariant face recognition, are explored. CNN architectures are broadly divided into feature extraction and classification tasks.

In this, the feature extraction task is performed by stacks of convolutional layers and pooling layers. The lower levels of convolutional layers extract some meaningful low-level features like horizontal, vertical, and slanted edges. Subsequently, higher layers of the convolutional network will connect low-level features to construct some blob or close curve, e.g. small geometric object-like features that can be used as discriminative features for classification tasks. In standard CNN, a flattened layer is used to convert the output of the convolutional layers into a single vector. And one or more fully connected layers (FC) perform the classification task. The softmax/sigmoid function calculates the class probability in the last output layer. The class with the highest probability is considered the predicted class of the input. CNN uses different loss functions to calculate the loss or error between the predicted and actual outputs. A backpropagation algorithm is used to update weights and biases.

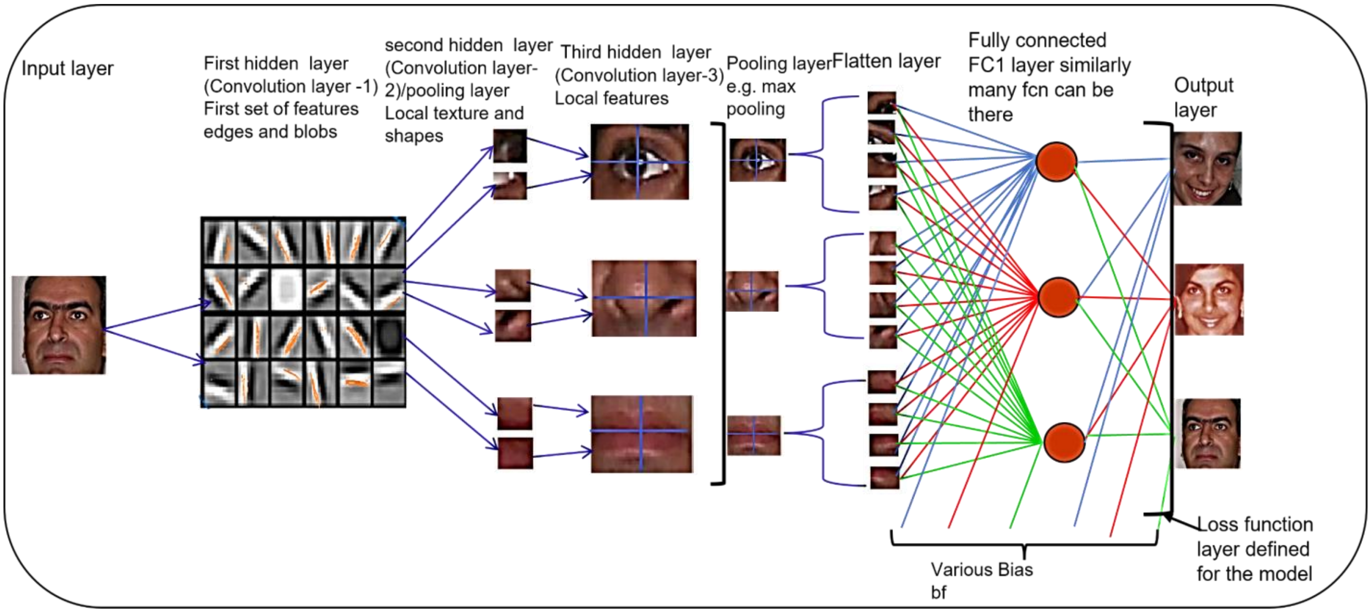


Fig. 11 Feature extraction through various hidden layers and classification by fully connected layers in deep neural network

Figure 11 shows all the steps used in a typical CNN network. Here in the architecture of CNN, the vital element is the loss function because the loss function measures the algorithm's performance. How well the algorithm fits data/ classify or recognises the test sample. Through the loss function, the error is determined, and the parameter/weights of the CNN get adjusted in the direction of the optimal feature subspace.

Various loss functions are evolved with the progress of machine/ deep learning. Out of many, the important ones are sigmoid, softmax, cross-entropy, contrastive loss, triplet loss, center loss, N-pair loss, and so on. The objective of each loss function is different. Each loss function provides measures depending on its functionality. e.g. The Cross-entropy is the loss function based on the probability calculation of the class outcomes. Cross-entropy loss is a classification loss that operates on class probabilities produced by the network independently for each sample, whereas contrastive loss is a metric learning loss, which operates on the data point produced by the network and their positions relative to each other. Cross entropy and sigmoid loss are not having the much discriminative characteristic, and they provide only the class information to which the sample may belong. It will not provide the subspace embedding and is not sensitive about intra-class and inter-class boundaries. To address the intra-class problem more effectively, the contrastive loss Hadsell, Raia et al. [45] and triplet loss Schroff et al. [46] proposed to that ensured the extra compactness in intraclass samples and a marginal separability between inter-class samples.

The resultant problem, however, is that the number of training pairs and triplets can theoretically go up to $O(N_2)$, where N is the total number of training samples. To get the best results, careful design of the subset of the training set is the extra burden in this type of loss function. Whereas with the softmax function, many researchers (Krizhevsky et al., [47]; Kaiming et al., [48]) have used it in CNN due to its simplicity and probabilistic interpretation. To empower CNN/DNN, the softmax function is equipped with more discriminative capability by introducing large marginal softmax (L-Softmax) (Liu, Weiyang et al., [49]). The L-softmax loss is based on angular separability; the generated features are angularly separated with some margin, leading to good classification. A short comparison of various loss functions with their definition, equation, inputs, and characteristics is given in Table 3 for ready reference. In the table, the softmax loss and large-marginal softmax loss are pictorially depicting the compactness of the L-softmax, providing more discriminative boundaries.

5.2. Review work on Deep Neural Network

Various studies show [50], [51], [52], [53], [54] the deep neural network uses discriminative methods in the area of the face aging process also. A. A. Moustafa et al. [50] have given features for face images. They used Discriminant correlation analysis (DCA) for a real-time fusion of face-feature in parallel for multi-discriminant correlation analysis. The k-nearest neighbor and support vector machine (SVM) is used for classification at the last layer. H. Huang et al. [51] designed a deep learning approach similar to the Siamese neural network, which attempts to learn a label distribution loss function that should have class conditional probability distributions similar when labels are of the same class. Y. Wen et al. [52] designed center loss, a new supervision signal for face recognition tasks. The centre loss maintains the center for deep features per class and simultaneously incurs a penalty score when the centre and the distance are away from some threshold. Y. Wang et al. [53] used a deep neural network to decompose face features into two components ages related and identity-related. The model has used linear regression techniques to learn the age-related component, and a variant of the A-softmax function is used for identity loss determination. It gives 98.75%, 58.21%, 99.5%, and 99.47% recognition accuracy on MORPH-II, FG-NET, CACD-VS, and LFW datasets. The model is not performing well on the FG-NET dataset, as FG-NET consists of large age variations, whereas it gives good results on other datasets. M. Nimbarte and K. Bhoyar [54] used a 7-layer CNN model.

A multidimensional support vector machine (MSVM) is used at the fully connected layer for classification. The recognition accuracy of their method on the FG-NET and the MORPH dataset is 76.6% and 92.5%, respectively. Similarly, J. Deng et al. [55] try to capture deep features to minimise marginal loss and enhance the discriminative power of the network. For example, an intra-class variance must be minimum and inter-class variance maximises with a joint combination of softmax loss. Y. Wen et al. [56] introduced a robust deep face recognition framework for cross-age by utilising the concept of Latent sub-space with CNN (LF-CNN). Their CNN model trained to learn Latent Identity called Latent Identity Analysis (LIA) instead of stochastic gradient descent (SGD) for parameter updation. LIA is used for the fully connected (FC) layer for parameter updation. The key to this model is to divide any face into two main factors: identity-related and age-related components (representing the age variations). The model parameters are learned by the maximum likelihood function. Finally, the joint objective function is used for softmax and contrastive loss functions. They registered a good accuracy of 97.51%, 88.1%, and 99.5% on MORPH Album-II, FG-NET, and CACD-VS datasets. S. Bianco [57] presented a new method for face recognition by incorporating large age gaps in Siamese deep coupled networks, and to calculate the losses, It uses a contrastive function. They have enhanced the method by including a feature injection layer at the fully connected layer. L. Boussaad and A. Boucetta [58] have tried various deep-learning models for their experiments and compared five popular deep-learning-based face characterisation architectures: AlexNet, GoogleNet, Inception V3, SqueezeNet, and ResNet50. Pose correction is performed to extract the feature vector. Then finally, three classifiers are used for classification: discriminant analysis, K-nearest neighbour, and support vector machines. Subsequently, it is concluded that AlexNet is best at handling variations in ageing effects. A. A. Moustafa et al. [59] use a pretrained standard VGG-Face model for Transfer Learning in order to optimise deep features and avoid overfitting. Genetic algorithm is used to extract the most appropriate and applicable features for classification. The method is evaluated on different distance metrics (Euclidian, Manhattan, Correlation, and Cosine) for the K-Nearest Neighbor (KNN) classifier. The recognition accuracy on FG-NET and MORPH-II datasets are 86.2% and 96%, respectively. Y. Li [60] used a deep convolutional network to learn features, including the Mahalanobis metric as the distance metric and threshold function. And a fully connected layer on the top of the deep architecture. They have reported 93.6% accuracy on the MORPH-II dataset.

Khawar Islam et al. [61] have coined the term shallow Age-Invariant Face Learning, which means pair of face images are grouped such that the gap of Age is kept large and tested the state-of-the-art methods (SOTA) for face verification in place of the complete data-sets. They investigated the performance of six of the most influential and cited deep face models, state-of-the-art methods (SOTA), such as Open-Face, VGG-Face, Face-Net, Deep-ID, and Deep-Face. , and ArcFace under different metric learning approaches. They reorganize the three cross-age data into negative and positive pairs with different age groups. The Facial Aging Dataset (FGNET) covers all age groups with aging and occlusion, the Longitudinal Dataset (MORPH-Album 2) for illumination and low resolution, and the Age Database (Age-DB) for pose and noise. In the AIFR field, there is very little research on low AIFR topics. Their studies and experiments concluded that Arc-Face and Face-Net outperform other SOTA for shallow AIFR problems. Most face recognition algorithms result in over-fitting with respect to feature dimensions and degenerate for shallow AIFR. So matching image pairs for a large age gap is still a challenging problem in the computer vision domain. L., Boussaad, et al. [62] have analysed the aging process on the performance of two algorithms.

The first algorithm is the study and implementation of the feature transformation via a two-dimensional discrete cosine (2D-DCT) and Kernel Fisher Analysis (KFA); k-NN is used for classification, which measures Euclidean distance between two features. And the other approach is tested in two different ways: first, they considered the face as a single entity, then viewed the face as an independent component set. They have experimented with the output of Alex-Net pre-trained CNN by fusing it with discriminant Correlation analysis (DCA) and without fusion. And in both cases, SVM is used for classification purposes. They performed experiments by dividing the FG-NET datasets into three age groups, 0 to 18, 19 to 40, and more than 40 years. Among the three approaches as given above, the last one, DCA plus SVM, gives a good performance in the last age groups; for the initial periods of age groups, the reason for obtaining the low recognition rate is to difference in facial features in faces because of the major growth of the facial muscles during the childhood and adulthood and the relatively small number of images for training the CNN. Du Lingshuang et al. [63] presented an Identity Difference Feature Extractor (IDFE). The identity Consistency Loss (ICL) regularisation function is derived and used at the output layer for each image.

Inspired by contrastive loss, their loss function penalises the features and assigns large weights to the images with large age gaps. Hence the classifier learns the distribution of the identity-related features and inter-identity features along with variations like illumination, poses, and noises. The model gives a good recognition accuracy of 97.5% on the CACD-VS dataset. Wang et al. [64] factorise the facial features into two uncorrelated elements or parts, age-related and identity-related components, through a deep feature factorisation framework. They have proposed a decorrelated learning method to minimise the correlation between decoupled features of age and the identity of a similar person.

The recognition rate achieved on FG-NET was 94.15%, and MORPH-II was 98.93%. Zhizhong Huang et al. [65] have improved the linear factorisation module of [64]; in the later model one-dimensional feature vector, the identity-related component is generated that lacks spatial information of face. Therefore, the former author has proposed an attendance-based features decomposition which is more semantically correlated in high-dimensional feature space. GAN is used to train the model. The 94.78% recognition accuracy is obtained on FG-NET.

In El Khiyari. et al. [66] used pertained VGG-Face network to extract the facial features, and three classifiers were used to evaluate performance: Principal Component Analysis (PCA), Linear discriminant analysis (LDA), Nearest neighbour classifier, and subspace discriminant method, which is an ensemble of 200 weak learners and decision tree. The performance of these classifiers is evaluated, and out of which, the Subspace discriminant method outperforms others. On FGNET and MORPH datasets, 80.6% and 92.2% results were obtained, respectively. Jian Zhao et al. [67] proposed an autoencoder-based generative adversarial network (GAN) that learns discriminative and robust person-specific facial representations free from age variance. Cross entropy loss is used to calculate the losses. The recognition accuracy tested on FG-NET, CACD, and MORPH-II benchmark data sets, and rates are 93.20%, 99.76%, and 99.65%, respectively.

Furthermore, they have prepared and collected a new large-scale cross-age Face Recognition (CAFR) benchmark dataset that can be used for age-invariant face recognition research. Here we have presented some of the prominent deep neural network-based methods, which are used in Age invariant face Recognition tasks; it is observed that training a deep neural network from scratch is a time-consuming task, and to achieve good accuracy of the deep neural network large number of datasets are essential. Hence extensive computational resources and labeled datasets for training the supervised models are required. Most of the time, it is not a practical solution in real-world applications. So, one can use a pre-trained network's activation layers as a feature extractor, and as per the requirements, different classifiers can be used.

Some well-known models in this category are Visual Geometry Group (VGG) models from the University of Oxford, VGG16, VGG19, ResNet50 from Facebook, InceptionV3 from Google, and Xception. The Xception model is an extension of the inception architecture. Many researchers are using these pre-trained models for various facial modeling tasks like facial expression, facial pose, gender, aging, and so on. The concept of using a pre-trained network layer is known as transfer learning. As given in the previous paragraph, various known architectures are available. They have different structures in terms of the number of convolution layers, number of max pooling layers, number of hyperparameters, and fully connected layers. They are available in a number of different libraries like Keras, Pytorch and Tensorflow, etc. Each model is trained on the ImageNet dataset containing about 1.2 million images. Research has recently accelerated in computer vision using Deep neural networks.

5.3. Conclusion of the Review Work on Deep Neural Network Methods

It is evident that for any face-related application or in any computer vision problems, the user can apply either pre-trained or custom-designed DNN. The conclusions drawn after reviewing the DNN-based papers are the selection of layer parameters viz a number of filters, the filter size, and the no. of neurons at the hidden layer, the no. of the hidden layer makes the important design criteria. But more emphasis should be given to the selection and design of the loss function at the fully connected layer. The features are good if they provide good inter-class separability and intra-class compactness. As in the case of many computer vision tasks like face recognition, there are significant intra-class variations, making the classification task more challenging. Various loss functions are discussed in the feature and classification subsection. Among them, The center loss and L-softmax loss are promising loss functions with compactness among the identical class distributions, whereas there is a marginal gap between different class distributions. Out of many applications of GAN, one of the applications is data augmentation.

Nowadays, generative adversarial network (GAN) [68] is used to generate age synthetic face images for different age groups, to solve the problem of shallow data. In many age-invariant datasets, a sufficient number of face images for different age groups are not available. The ideal dataset should contain an adequate number of face images to absorb the age progression phenomenon. Here GAN is used in this kind of data augmentation application. In other forms, GAN [65][69], [68], [70] is used to encode the identity-related features, and the decoder decodes the face images into different age group face images of the same identity. The aging parameter of each individual will be learned and encoded by the encoder, and the decoder generates the face images in the required age group by decoding the parameter.

Table 2. Comparison between three models based on general Characteristics

Characteristics of Model	Discriminative Model	Generative Model	DNN Model
Learning	Decision Boundaries	Probabilistic model	Gradient-based back-propagation
Asymptotic Error	Small Error	Converges Faster parameters	Depends on the learning rate and other hyper-parameters
Feature estimation/feature construction	P (y | x)	P (x, y)	Automatic Feature Engineering
Explain-ability	Generally low to medium	Express complex relationship	complex to explain, tools are available to explain
Label data set requirements	Yes	Not necessary	Yes, for supervised, and not necessary for unsupervised
Supervised/Unsupervised	Suitable for supervised problems (classification)	Suitable for unsupervised problems (clustering)	Suitable for both problems
Classification function	Support Vector Machine, Linear Regression	Naïve Bayesclassifier, Generative Adversarial Network (GAN),	Softmax for multi-class, sigmoid for binary classification, Contrastive loss, Triplet loss, Center loss

Table 3. Comparison between various Loss Functions of Deep Neural Network

Loss Function	Cross entropy Loss	Softmax Loss /L-Softmax Loss	Contrastive Loss	Triplet Loss	Center Loss
Definition	A defect is calculated, which penalizes the probability based on how far it is from the actual expected value. The penalty is logarithmic in nature, giving a high score for large differences close to one and a low score for small differences approaching zero.	Softmax is an activation function that returns probabilities for each class, and these probabilities sum to 1.	Works with similarity labels to learn distance mapping. Maximizes the distance between the negative pair of images and minimizes the distance between the positive pair of images. i.e. the goal of contrastive loss is an embedding space in which similar sample pairs stay close to each other while dissimilar ones are far apart.	Works with three data items, the Anchor, the Positive, and the Negative data. Maximizes the distance between the anchor and the negative and minimizes the distance between the anchor and the positive embedding.	Center Loss adds a new regularization term to Softmax Loss to pull features into proper class centers. Therefore, the center loss clusters the samples on the appropriate class centers. It also solves the sampling problem because you don't have to do difficult sample mining. (In case of triplet loss)
input /pair ofinput	Softmax probability of each class as input to this function.	Softmax Activation plus a Cross-Entropy Loss.	Pair of positive or negative data/images (features).	The Anchor (ground truth), the Positive, and the Negative data/image(features).	The images/data(features)
Remarks	Cross-entropy loss compares the outcome's probability with the desired output. It can be used when the model output can be measured in terms of the probability distribution. The	Softmax loss only learns separable features that are not discriminative.	Contrastive loss is advantageous when similarity relationships are known. For supervised and unsupervised learning, Contrastive learning can be applied.	A Triplet loss can be used if the data has negative /positive correlations. Triplet loss performance is based on the training data point (triplet) selection process. Triplet Loss does not have the side effect of forcing anchor and positive	Compared to triplet loss, the correct choice of training data points is the key to success for triplet loss, whereas the choice of training data points is not critical for center loss, so it is not

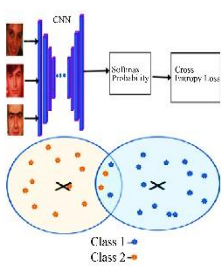
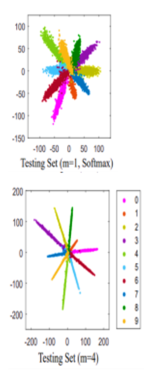
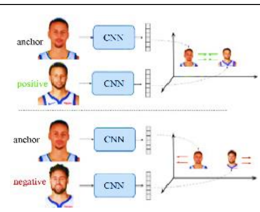
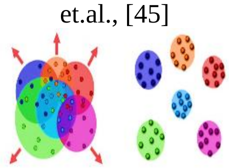
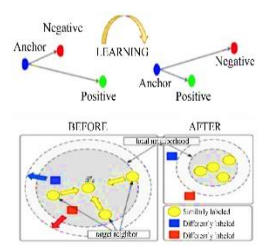
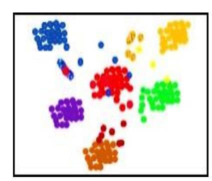
	purpose of this loss is to minimize the difference between the actual output and the desired output loss, i.e. the smaller the loss, the better the model. A perfect model has almost zero cross-entropy loss.			samples to be encoded at the same point in vector space as Contrast Loss. Unlike contrast loss, it allows Triplet Loss to tolerate intra-class variance because the triplet essentially forces the distance between the anchor and any positive to a certain limit.	difficult to train. The joint function of center loss and softmax loss will prevent embedding from collapsing.
Loss Function equation	$L_{CE} = -\sum_{i=1}^{n} t_n \log(p_i), f$	$s(y)_i = \frac{\exp(y_i)}{\sum_{j=1}^{n}\exp(y_i)}$	$yd^2 + (1-y)\max(margin - d, 0)^2$	$\max(\lVert f_a - f_p\rVert^2 - \lVert f_a - f_n\rVert^2 + m, 0)$	$\tau_{center} = \tau_{softmax} + \frac{\lambda}{2}\sum_{i=1}^{y} \lVert z_i - c_{y_i}\rVert_2^2$
Description	where t_i is the truth label, and p_i is the softmax probability of the i_{th} class	Here y_i is the i_{th} input vector. And n is the no. of classes.	where d is the Euclidean distance between the two image features (suppose their features are $f_1 and f_2$): $d = \lVert f_1 - f_2 \rVert_2$. The margin term is used to "tighten" the constraint: The two dissimilar image pairs should remain at least a "margin" distance away, or a loss will be incurred.	In the above formula, m is a boundary term used to "stretch" the difference in distance between similar and dissimilar pairs within a triplet. f_a, f_p, f_n are the feature embeddings for the anchor, positive and negative images.	The $c_{yi} \epsilon R^d$ denotes the $y_i th$ class center of deep features. Ideally, the c_{yi} keeps on updating as the deep features change means when even new face images are introduced. In each iteration averaging the features of every class in each iteration is considered.
Related Diagrams	 Cross Entropy Loss	 Softmax loss Liu, Weiyang et. al. [49]	 Contrastive Loss Hadsell et.al., [45]  Raia et.al.[45], Rippel, Oren, et.al., [72]	 Triplet Loss F. Schroff et.al. [46]	 CenterLossWen, Yandong [73]

Table 4. Some existing techniques with comparison regarding different feature descriptors and classifiers. It also gives Rank-1 recognition rate on various datasets

Paper	Feature descriptor	Classifier	Data sets used	Efficiency
Z. Li, U. Park and A. K. Jain (2011) [30]	**Discriminative:** Patch-based local feature SIFT and multi-scale local binary patterns (MLBP)	LDA-based fusion rule.	FG-NET MORPH – II	47.0% 83.9%
D.Sungatullina,	**Discriminative:** Patch-based three types	Multi-view-based discriminative	MORPH	65.2%

J. Lu, G.WangandP. Moulin (2013) [31]	of local features such as SIFT, LBP, and GOP, used	latent subspace learning (MDL)	FG-NET	91.8%
DihongGong et al. (2013) [36]	**Discriminative:** HOG features descriptor. PCA and LDA (for dimension reduction)	HFA (hidden factor analysis) is based on a Probabilistic model with latent factors.	FG-NET MORPH – II	69% 91.14%
Si, Junyong et.al., (2015) [37]	**Discriminative:** HOG features descriptor. PCA and LDA (for dimension reduction)	Feature progressing model designed based on regression method. And Cosine similarity is used to check similarity.	FG-NET MORPH-II	74.7% 89.74%
B. Chen, C. Chen and W. H. Hsu (2014) [71]	**Discriminative:** Data-driven method is used. LBP (for sixteen prominent landmarks from faces)	Encoded local features used into reference space using Cross Age Reference Coding.	CACD MORPH – II	87.6% 92.8%
Yadav D, et al. (2013) [33]	**Discirminative:** LBP descriptors are computed for facial regions like right and left periocular, binocular and mouth regions.	Chi-square distance metric was used to calculate the match score. A bacterial foraging-based classification method was used.	FG-NET IIIT Delhi	64.5% 54.3%
Xiaonan Hou et.al., (2016) [38]	**Discriminative:** High-dimension LBP descriptors	Reference set encoding is used corresponding to each age range, and cosine similarity is used to compute the matching scores between images.	MORPH - II CACD	94.5% 64.0%
U. Park, Y. Tong and A. K. Jain (2010) [39]	**Generative:** Learn aging patterns based on PCA coefficients in separated 3D shapes.	By 3D shape modelling closets, the possible weighted sum of the shapes at any age is computed. And the state-of-the-art commercial face matcher FaceVACS is used.	FG-NET MORPH – II	37.4% 66.4%
HuilingZhou,Kin-Man Lam (2018) [40]	**Generative:** Model human identity and aging variables using probabilistic LDA.	Expectation-Maximization is used iteratively for identity subspace.	FG-NET MORPH – II	88.23% 95.62%
Li, Ya, et al. (2015) [60]	The deep convolutional network learns features, distance metrics and threshold functions simultaneously.	It uses the Mahalanobis metric and distance thresholds jointly for classification.	MORPH-II	93.6%
YandongWen et al. (2016) [52]	CNN is used for new supervision signals.	Centre loss for the face recognition task.	AGEDB	95.9%
Wen et al. (2016) [56]	CNN: Use latent Identity. They design a Latent Identity Analysis (LIA) method to learn weight and bias parameters for the LF-CNN.	Used joint objective function for softmax loss and contrastive loss.	FG-NET MORPH – II	88.1% 97.51%
El Khiyari. et al. (2016) [66]	Pretrained VGG-Face model used to extract face features.	Subspace discriminant method used for classification, Which is an ensemble of 200 weak learners.	FG-NET MORPH-II	80.6% 92.2%
Deng, J. Y. et al. (2017) [55]	It uses CNN for marginal loss to enhance the discriminative power of the deeply learned features.	Joint softmax and marginal loss are used for trainable classification functions and optimised by standard SGD.	AGEDB	98.1%
Bianco (2017) [57]	Siamese network (Deep CNN) is used for feature extraction.	The feature injection layer is inserted at the fully connected (fc) layer. A contrastive loss function is used to optimize the loss and classification.	LAG	84.95%

Wang et al. (2018) [53]	Orthogonal Deep Features Decomposition.	Stochastic gradient descent (SGD) algorithm was used to optimize the two types of losses.	FG-NET MORPH – II CACD-VS	58.21% 98.75% 99.5%
Nimbarteand Bhoyar (2018) [54]	Deep learning with CNN.	Support vector machine (SVM) classifier.	FG-NET MORPH	76.6% 92.5%
Wang et.al., (2019) [64]	Deep feature factorization learning framework and Decorrelated Adversarial Learning (DAL).	The DAL calculates the canonical correlation between the paired features of the decomposed components. And the stochastic gradient descent method is used to optimize the loss.	FG-NET MORPH-II	94.5% 98.97%
Hai Huang et al. (2019) [51]	Discriminative features based on class conditional probability distributions for similar labels or different labels.	The softmax loss is followed by distribution loss.	MORPH – II LFW CACD-VS	97.1% 99.46% 99.10%
Jian Zhao et al. (2020) [67]	A network similar to the generative adversarial network (GAN) that provides a Representation of Learning sub-Net	Cross entropy loss is used to calculate the losses.	FG-NET CACD MORPH-II	93.20%, 99.76% 99.65%
Boussaad et al. (2020) [58]	Various pretrained CNN networks are used and compared with each other. Alexnet gives the best results.	Three classifiers, K-nearest neighbor, discriminant analysis and support vector machines, are used.	FG-NET	98.21%
Moustafa, A. Elnakib, and N. F. F. Areed (2020) [59]	Feature extracted through a pretrained VGG face model is optimized using a Genetic algorithm.	K-nearest neighbor (KNN) classifier.	FG-NET MORPH – II	86.2% 96%
Moustafa,A.A., Elnakib, A.& Areed, N.F.F (2020) [50]	VGG-Face model used to extract compact face features.	Features are fused using the real-time feature-level multi-discriminant correlation analysis.	FG-NET MORPH – II	81.5% 96.5%
Islam et.al., (2021) [61]	State-of-the-art (SOTA) models Face-Net, VGGFace, Open-Face, Deep-Face, Deep-ID, and ArcFace are used to extract features.	SOTA model's classifier is used to verify the shallow age-invariant face images. Best models: ArcFace - FG-NET VGG-Face - AGEDB, and FaceNet - MORPH are giving good results.	FG-NET AGEDB MORPH	67.82% 60.02% 97.06%
L., Boussaad, et.al., (2021) [62]	Two methods are compared. First method: two-dimensional discrete cosine transform (2D-DCT) fused with a Kernel Fisher Analysis (KFA). Second method: Alex-Net's fc layer output as descriptors are fused by a discriminant Correlation analysis (DCA) and SVM	K-nearest neighbor (KNN) and SVM classifier (for age group greater than 40)	FG-NET	98.25%
Du Lingshuang et.al., (2021) [63]	An Identity Difference Feature Extractor (IDFE) is used to obtain an identity difference feature vector.	Direct Cross-age Verification Network (DCVN) is introduced, and a new loss function is derived. The loss function learns the identity difference distributions of positive and negative image pairs.	CASD-VS	97.5%

Table 5. Usage of various feature descriptors and classification techniques found in the paper

Papers	Types of Feature Descriptor used							Type of Classification Method					Face Aging Dataset used					Type of Machine Learning Method		
	LBP	SIFT	GOP	HOG	PCA	LDA	Deep Learning Features	K-NN	SVM	Sigmoid	Softmax	Other Encoded	FG-NET	AGEDB	MORPH-Album-2	CACD	IIIT Delhi	Discriminative	Generative	Deep Learning
Park et al., (2011)[30]	✓	✓				✓						✓	✓		✓			✓		
Sungatullina et al., (2013)[31]	✓	✓	✓			✓						✓	✓		✓			✓		
Gong et al., (2013) [36]				✓	✓	✓						✓	✓		✓			✓		
Chen, et al., (2015) [71]	✓											✓			✓	✓		✓		
Yadav et al. (2013)[33]	✓											✓	✓				✓	✓		
Wu, Yongbo, et al.,(2018)[3]	✓				✓		✓					✓	✓		✓	✓		✓		
Park et al., (2010)[39]					✓							✓	✓		✓				✓	
Zhou and Lam (2018) [40]						✓							✓		✓				✓	
Moustafa et al. (2020) [50]							✓						✓		✓					✓
Huang et al. (2019)[51]							✓				✓				✓	✓				✓
Wen et al. (2016) [52]							✓				✓			✓						✓
Deng, et al. (2017) [55]							✓			✓	✓			✓						✓
Wen et al. (2016) [56]							✓				✓		✓		✓					✓
Bianco (2017) [57]							✓				✓									✓
Wang et al. (2018) [53]							✓			✓			✓		✓	✓				✓
Nimbarte and Bhoyar (2018) [54]							✓		✓				✓		✓					✓
Boussaad and Boucetta (2020)[58]							✓		✓						✓					
Moustafa et al. (2020) [59]							✓						✓		✓					✓
Li et al. (2015) [60]							✓	✓				✓			✓					✓

Table 6. Summary related to the age-varying face datasets

Age varying datasets	Year	Age range	No. of Subjects	Total No. of Images	No. of images per subject (on average)
FG-NET [2]	2004	1 to 69 years	82 persons	1002	12 images per person.
CACD (Cross Age Celebrity Dataset) [71]	2015	16 to 62 years	2,000persons (of celebrities)	160,000	80 images per person
AGEDB [75]	2017	7 to 85 years	568 persons (of celebrities)	16,488 images	29 images per person.
MORPH-Album–2 (standard) [76]	2006	18 to 76 years	20,000 persons	78,000	4 images per person
MORPH-Album-2 (publicly available) [76]	2006	16 to 77 years	13,618 persons	55,134	4 images per person
IIIT Delhi – face data set [33]	2013	4 to 88 years	102 persons	2,600	4 images per person
PCSO_LS [77]	2017	18 to 83 years	18,007	147,784	8 images per person
WhoIsIt (WIT) [78]	2014	1 to 80 years	110	1109	10 to 12 images per person
LARGE AGE-GAP DATASET (LAG) S. Bianco, [57]	2016	Large range (10 to 60 years)	1,010	3,828	3.8 images per person

6. Issues and Challenges

In this survey, we have addressed the issues mainly related to the effects of age variation on the performance of the algorithms. And for that, the age-related dataset is required to train and test the models. The present dataset like FGNET, AGEDB, MORPH-II, and CACD-VS, which are available, have majorly following issues:

- The training dataset for the AIFR problem requires face images in progressive time domain mean age progressive images are needed. But existing datasets have imbalanced images per person. Then it becomes difficult to train the model and to get the optimal results. And another factor that makes modelling more difficult is that each person has his/her age pattern based on genes, health, environment, ethnicity, upbringing, culture etc.
- The face images have variations like Pose, illumination, blurring, and expressions, making the recognition task more difficult. ·
- The current data set, like in FG-NET images with large age gaps, is available for some subjects. So, for each age group, progressively facial images are a must to conduct extensive research in this domain.

The other challenges are related to the problem domain:

- During the childhood phase, the part of the data, which is between ages 0 to around 7,8 years, the facial features are yet to evolve and are still stable. Especially in this age group, the person's identity is not mature. At this age, it is tough to associate the face of this age (0 to 7) with the correct same person's elder face, even with human eyes.
- The effects of human facial aging are not constant over time. During childhood, the shape of the face changes more than the texture of the face. During the teenage years, facial muscle changes slowly and stabilize. The facial structure in adulthood changes slowly until age 50, after which variations such as wrinkles appear.
- Wrinkles and lines at various places of the face and changes in pigmentation are more visible. So, modelling a system that correctly recognizes human faces at all age levels with variations is difficult.

So, there is a need to study the problem and methods related to this domain. And we have presented a comprehensive review of the research work done by various researchers that will be useful for the upcoming researchers targeting this domain.

7. Discussion about Future Directions

The methods covered in this review are mainly based on supervised based learning. Labelled datasets are required to train the supervised learning models to classify or predict the respective class. In contrast, unsupervised learning is applied where a labelled dataset is missing. It provides a technique that learns similar features in the data or images and divides them into clusters based on similarities. These algorithms discover hidden patterns or data groupings without the need for human intervention. In the domain of face recognition, the new challenges are unsupervised manner learning and clustering of face images across various classes. Many researchers [79], [80], [81], [82], [83], [84] have used threshold clustering, k-means, mean shift, DBSCAN (Density-Based Spatial

Clustering of Applications with Noise), approximate rank-order, spectral clustering, and graph-based clustering approaches but still have not achieved impressive results. There are still challenging problems like gender identification, age determination, and clustering of face images when the image consists of a variety of variations like aging, illumination, expression, etc. are the active area of researchers. As discussed in Section 5, the shallow AIFR, the face verification for a large age gap, is also a challenging problem. Therefore, more robust, reliable, and fast face recognition methods across aging and other appearance variations with a limited number of datasets are required to accommodate the issues and challenges in the area of supervised and unsupervised learning techniques.

8. Conclusion

The previous sections present a detailed review of the experimental results from some existing techniques with an analysis of the model characteristics and performance benchmarking on the most important facial aging datasets (FGNET, AGEDB, MORPH-II, CACD). The rank-1 recognition accuracy is considered to measure the performance. And many researchers have achieved comparatively good results. The AIFR techniques discussed in sections 3, 4, and 5 addresses face recognition across the variation in the time domain frame of reference. As described already in section 3, Discriminative approaches depend on appropriate distinct local and global features. These approaches address the major AIFR, but performance would be degraded due to some issues like shallow aging data sets and blurring of the images.

Moreover, another important phase that can be emphasised is pre-processing, which improves accuracy and dimensionality reduction. PCA is widely used for this. The Generative techniques are broadly divided into two steps. The first is synthesizing the face image by modeling an individual's identity and age parameters at a certain age and then optimizing the aging model for different age groups. The optimizing process is a difficult task. Secondly, when synthesizing the face images, the inherent noise of the model parameters may have a negative impact on recognition accuracy.

Moreover, constructing a generative model with good accuracy involves complex tasks of age progression. Hence, these methods are not suitable for real-time face recognition. In the last few years, we have witnessed impressive research in the area of Deep Learning for computer vision problems. The methods proposed in [50] - [60], [85] are the recently developed deep learning approaches for AIFR. They offered a good discriminative feature extraction model with various loss functions used to update parameter updation and optimization errors. To get high performance on DNN enormous amount of training on huge datasets is needed, resulting in a lot of time for training a DNN. Otherwise, DNN will not perform well.

Hence complete training in deep learning networks using massive datasets remains a challenge. Transfer learning and domain adaptation could solve this problem. And also, the combination of the approaches mentioned above (Discriminative and Generative) is being developed and used for the appropriate performance of AIFR. In many studies of deep learning-based methods, pre-trained deep networks are fused with discriminative feature-based classification layers and used successfully in various applications like security systems, bio-metric recognition, attendance systems, etc. The AIFR domain, generative adversarial network (GAN), is also widely used. In a GAN, the generative and discriminant network is the two integral parts of the network. Because of the success of GAN techniques, the generative techniques now again become prominent while addressing real-world problems, as mentioned above.

References

[1] A. Lanitis, C.J. Taylor, and T.F. Cootes, "Toward Automatic Simulation of Aging Effects on Face Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442–455, 2002. [CrossRef] [Google Scholar] [Publisher Link]

[2] FG-Net. [Online]. Available: http://www.fgnet.rsunit.com/

[3] Narayanan Ramanathan, Rama Chellappa, and Soma Biswas, "Computational Methods for Modeling Facial Aging: A Survey," *Journal of Visual Languages Computing*, vol. 20, no. 3, pp. 131–144, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[4] Robert Shaw, Michael McIntyre, and William Mace, "The Role of Symmetry in Event Perception," Essays in Honor of James J.Gibson, 1974. [Publisher Link]

[5] Iacopo Masi et al., "Deep Face Recognition: A Survey," *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[6] Zahra Mortezaie, and Hamid Hassanpour, "A Survey on Age Invariant Face Recognition Methods," *Jordanian Journal of Computers and Information Technology*, vol. 5, no. 2, pp. 87-96, 2019. [Google Scholar] [Publisher Link]

[7] Manisha M. Sawant, and Kishor M. Bhurchandi, "Age Invariant Face Recognition: A Survey on Facial Aging Databases, Techniques and Effect of Aging," *Artificial Intelligence Review*, vol. 52, pp. 981–1008, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[8] Kedimotse Baruni et al., "Age Invariant Face Recognition Methods: A Review," *2021 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] Mei Wang, and Weihong Deng, "Deep Face Recognition: A Survey," *Neurocomputing*, vol. 429, pp. 215–244, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Marem H. Abdulabas, and Noor D. Al-Shakarchy, "Face Information Forensics Analysis Based on Facial Aging: A Survey," *2022 International Conference on Intelligent Systems and Computer Vision (ISCV)*, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] David G. Lowe, "Distinctive Image Features from Scale Invariant Key Points," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004. [CrossRef] [Google Scholar] [Publisher Link]

[12] H. Bay et al., "Surf: Speeded up Robust Features," *European Conference on Computer Vision*, pp. 404–417, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[13] Stefan Leutenegger, Margarita Chli, and Roland Y. Siegwart, "Brisk: Binary Robust Invariant Scalable Key Points," *2011 International Conference on Computer Vision*, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[14] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002. [CrossRef] [Google Scholar] [Publisher Link]

[15] Srinivasa Reddy Konda, Vijaya Kumar, and Venkata Krishna, "Face Recognition using Multi Region Prominent LBP Representation," *International Journal of Electrical and Computer Engineering*, vol. 6, no. 6, p. 2781, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[16] Thierry Bouwmans et al., "On the Role and The Importance of Features for Background Modeling and Foreground Detection," *Computer Science Review*, vol. 28, pp. 26–91, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[17] Y. Danisman et al., "A Comparison of Eigenvalue Methods for Principal Component Analysis," *Applied Mathematics and Computation*, vol. 3, no. 13, pp. 316–331, 2014. [Google Scholar] [Publisher Link]

[18] Liton Chandra Paul, and Abdulla Al Sumam, "Face Recognition using Principal Component Analysis Method," *International Journal of Advanced Research in Computer Engineering Technology (IJARCET)*, vol. 1, no. 9, pp. 135–139, 2012. [Google Scholar] [Publisher Link]

[19] Fatma Zohra Chelali, A. Djeradi, and R. Djeradi, "Linear Discriminant Analysis for Face Recognition," *2009 International Conference on Multimedia Computing and Systems*, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[20] O. Déniz, M. Castrillon, and M. Hernández, "Face Recognition using Independent Component Analysis and Support Vector Machines," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2153–2157, 2003. [CrossRef] [Google Scholar] [Publisher Link]

[21] M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski, "Face Recognition by Independent Component Analysis," *IEEE Transactions on Neural Networks*, vol. 13, no. 6, pp. 1450–1464, 2002. [CrossRef] [Google Scholar] [Publisher Link]

[22] Chang Shu, Xiaoqing Ding, and Chi Fang, "Histogram of the Oriented Gradient for Face Recognition," *Tsinghua Science and Technology*, vol. 16, no. 2, pp. 216–224, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[23] Jiachen Chen, and W. Kenneth Jenkins, "Facial Recognition with PCA and Machine Learning Methods," *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[24] Thai Hoang Le, and Len Bui, "Face Recognition Based on SVM and 2DPCA," *arXiv preprint arXiv:1110.5404*, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[25] R. Indumathi, and N. Palanivel, "Pose Invariant Face Recognition using HMM and SVM using PCA for Dimensionality Reduction," *International Conference on Information Communication and Embedded Systems (ICICES2014)*, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[26] Pallabi Parveen, and Bhavani Thuraisingham, "Face Recognition using Multiple Classifiers," *2006 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06)*, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[27] Sasan Karamizadeh, and Shahidan M. Abdullah, "Race Classification using Gaussian-based Weight K-nn ALGORITHM for Face Recognition," *Journal of Engineering Research*, vol. 6, no. 2, pp. 103–121, 2018. [Google Scholar] [Publisher Link]

[28] Xiaoyang Tan et al., "Recognizing Partially Occluded, Expression Variant Faces from Single Training Image Per Person with SOM and Soft k-NN Ensemble," *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pp. 875–886, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[29] Chigozie Nwankpa et al., "Activation Functions: Comparison of Trends in Practice and Research for Deep Learning," *arXiv preprint arXiv:1811.03378*, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[30] Zhifeng Li, Unsang Park, and Anil K. Jain, "A Discriminative Model for Age Invariant Face Recognition," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1028–1037, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[31] Diana Sungatullina et al., "Multiview Discriminative Learning for Age-invariant Face Recognition," *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[32] Bor-Chun Chen, Chu-Song Chen, and Winston H. Hsu, "Face Recognition and Retrieval using Cross-age Reference Coding with Cross-age Celebrity Dataset," *IEEE Transactions on Multimedia*, vol. 17, no. 6, pp. 804–815, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[33] Daksha Yadav et al., "Bacteria Foraging Fusion for Face Recognition Across Age Progression," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 173–179, 2013. [Google Scholar] [Publisher Link]

[34] Soma Biswas et al., "A Non-generative Approach for Face Recognition Across Aging," *2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems*, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[35] Haibin Ling et al., "A Study of Face Recognition as People Age," *2007 IEEE 11th International Conference on Computer Vision*, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[36] Dihong Gong et al., "Hidden Factor Analysis for Age Invariant Face Recognition," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2872–2879, 2013. [Google Scholar] [Publisher Link]

[37] Junyong Si, and Weiping Li, "Age-invariant Face Recognition Using a Feature Progressing Model," *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[38] Xiaonan Hou, Shouhong Ding, and Lizhuang Ma, "Robust Feature Encoding for Age-invariant Face Recognition," *2016 IEEE International Conference on Multimedia and Expo (ICME)*, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[39] Unsang Park, Yiying Tong, and Anil K. Jain, "Age-invariant Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 947–954, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[40] Huiling Zhou, and Kin-Man Lam, "Age-invariant Face Recognition Based on Identity Inference from Appearance Age," *Pattern Recognition*, vol. 76, pp. 191–202, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[41] Jinli Suo et al., "A Compositional and Dynamic Model for Face Aging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 385–401, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[42] Junyan Wang et al., "Age Simulation for Face Recognition," *18th International Conference on Pattern Recognition (ICPR'06)*, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[43] Unsang Park, Yiying Tong, and Anil K. Jain, "Face Recognition with Temporal Invariance: A 3D Aging Model," *2008 8th IEEE International Conference on Automatic Face Gesture Recognition*, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[44] Xin Geng, Zhi-Hua Zhou, and Kate Smith-Miles, "Automatic Age Estimation based on Facial Aging Patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2234–2240, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[45] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality Reduction by Learning an Invariant Mapping," *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[46] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A Unified Embedding for Face Recognition and Clustering," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815–823, 2015. [Google Scholar] [Publisher Link]

[47] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[48] Kaiming He et al., "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016. [Google Scholar] [Publisher Link]

[49] Weiyang Liu et al., "Large-margin Softmax Loss for Convolutional Neural Networks," *arXiv:1612.02295*, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[50] Amal A. Moustafa, Ahmed Elnakib, and Nihal F.F. Areed, "Age Invariant Face Recognition Based on Deep Features Analysis," *Signal, Image and Video Processing*, vol. 14, pp. 1027-1034, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[51] Hai Huang et al., "Label Distribution Learning-Based Age-Invariant face Recognition," *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 214–222, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[52] Yandong Wen et al., "A Discriminative Feature Learning Approach for Deep Face Recognition," *European Conference on Computer Vision*, pp. 499–515, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[53] Yitong Wang et al., "Orthogonal Deep Features Decomposition for Age-invariant Face Recognition," *Proceedings of the European Conference on Computer Vision*, pp. 738-753, 2018. [Google Scholar] [Publisher Link]

[54] Mrudula Nimbarte, and Kishor Bhoyar, "Age Invariant Face Recognition using Convolutional Neural Network," *International Journal of Electrical & Computer Engineering*, vol. 8, no. 4, pp. 2126-2138, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[55] Jiankang Deng, Yuxiang Zhou, and Stefanos Zafeiriou, "Marginal Loss for Deep Face Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 60–68, 2017. [Google Scholar] [Publisher Link]

[56] Yndong Wen, Zhifeng Li, and Yu Qiao, "Latent Factor Guided Convolutional Neural Networks for Age-invariant Face Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4893–4901, 2016. [Google Scholar] [Publisher Link]

[57] Simone Bianco, "Large Age-gap Face Verification by Feature Injection in Deep Networks," *Pattern Recognition Letters*, vol. 90, pp. 36–42, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[58] Leila Boussaad, and Aldjia Boucetta, "Deep-learning Based Descriptors in Application to Aging Problem in Face Recognition," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 6, pp. 2975-2981, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[59] Amal A. Moustafa, Ahmed Elnakib, and Nihal F.F. Areed, "Optimization of Deep Learning Features for Age-invariant Face Recognition," *International Journal of Electrical Computer Engineering*, vol. 10, no. 2, pp. 1833-1841, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[60] Ya Li et al., "A Deep Joint Learning Approach for Age Invariant Face Verification," *CCF Chinese Conference on Computer Vision*, pp. 296–305, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[61] Khawar Islam et al., "Face Recognition using Shallow Age-invariant Data," *2021 36th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[62] Leila Boussaad, and Aldjia Boucetta, "The Aging Effects on Face Recognition Algorithms: The Accuracy According to Age Groups and Age Gaps," *2021 International Conference on Artificial Intelligence for Cyber Security Systems and Privacy (AI-CSP)*, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[63] Lingshuang Du, and Haifeng Hu, "Cross-age Identity Difference Analysis Model Based on Image Pairs for Age Invariant Face Verification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 7, pp. 2675–2685, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[64] Hao Wang et al., "Decorrelated Adversarial Learning for Age-invariant Face Recognition," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3527–3536, 2019. [Google Scholar] [Publisher Link]

[65] Zhizhong Huang, Junping Zhang, and Hongming Shan, "When Age-invariant Face Recognition Meets Face Age Synthesis: A Multi-task Learning Framework," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7282–7291, 2021. [Google Scholar] [Publisher Link]

[66] Hachim El Khiyari, and Harry Wechsler, "Face Recognition Across Time Lapse using Convolutional Neural Networks," *Journal of Information Security*, vol. 7, no. 3, pp. 141–151, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[67] Jian Zhao, Shuicheng Yan, and Jiashi Feng, "Towards Age-invariant Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 474-487, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[68] Siyao Chen et al., "Age-invariant Face Recognition Based on Sample Enhancement of Generative Adversarial Networks," *2019 6th International Conference on Systems and Informatics (ICSAI)*, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[69] Daksha Yadav et al., "Age Gap Reducer-gan for Recognizing Age-separated Faces," *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[70] Grigory Antipov, Moez Baccouche, and Jean-Luc Dugelay, "Face Aging with Conditional Generative Adversarial Networks," *2017 IEEE International Conference on Image Processing (ICIP)*, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[71] Bor-Chun Chen, Chu-Song Chen, and Winston H. Hsu, "Cross-age Reference Coding for Age-invariant Face Recognition and Retrieval," *European Conference on Computer Vision*, pp. 768–783, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[72] Oren Rippel et al., "Metric Learning with Adaptive Density Discrimination," *arXiv preprint arXiv:1511.05939*, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[73] Yandong Wen et al., "A Discriminative Feature Learning Approach for Deep Face Recognition," *European Conference on Computer Vision*, pp. 499–515, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[74] V. Muthuvel Vijai, and P.A. Mathina, "An Effective Ring Partition And Half toning Combined Face Morphing Detection," *International Journal of Computer and Organization Trends*, vol. 11, no. 4, pp. 10-14, 2021. [CrossRef] [Publisher Link]

[75] Stylianos Moschoglou et al., "AgeDB: The First Manually Collected, In-The-Wild Age Database," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 51–59, 2017. [Google Scholar] [Publisher Link]

[76] K. Ricanek, and T. Tesafaye, "Morph: A Longitudinal Image Database of Normal Adult Age-progression," *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[77] Lacey Best-Rowden, and Anil K. Jain, "Longitudinal Study of Automatic Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 1, pp. 148–162, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[78] Maneet Singh et al., "On Recognizing Face Images with Weight and Age Variaztions," *IEEE Access*, vol. 2, pp. 822–830, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[79] Yongbo Wu, Haifeng Hu, and Haoxi Li, "Age-invariant Face Recognition using Coupled Similarity Reference Coding," *Neural Processing Letters*, vol. 50, pp. 397–411, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[80] Shuai Shen et al., "Structure-aware Face Clustering on a Large-scale Graph with 107 Nodes," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9085–9094, 2021. [Google Scholar] [Publisher Link]

[81] Zhongdao Wang et al., "Linkage based Face Clustering via Graph Convolution Network," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1117-1125, 2019. [Google Scholar] [Publisher Link]

[82] Chunhui Zhu, Fang Wen, and Jian Sun, "A Rank-order Distance-based Clustering Algorithm for Face Tagging," *CVPR 2011*, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[83] Charles Otto, Dayong Wang, and Anil K. Jain, "Clustering Millions of Faces by Identity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 2, pp. 289–303, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[84] Leonardo Chang, Airel Pérez-Suárez, and Miguel González-Mendoza, "Effective and Generalizable Graph-based Clustering for Faces in the Wild," *Computational Intelligence and Neuroscience*, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[85] Hachim El Khiyari, and Harry Wechsler, "Age Invariant Face Recognition using Convolutional Neural Networks and Set Distances," *Journal of Information Security*, vol. 8, no. 3, 2017. [Google Scholar] [Publisher Link]

[86] Sonia Mittal, and Mukesh Patel, "Age Invariant Face Recognition using Discriminative Tensor Subspace Learning with Fuzzy based Classification," *Expert Systems*, 2023. [CrossRef] [Google Scholar] [Publisher Link]