Applied AI Letters

WILEY

LETTER OPEN ACCESS

Vision Transformer-Enhanced Multi-Descriptor Approach for Robust Age-Invariant Face Recognition

Justice Kwame Appati[1] | Emmanuel Tsifokor[1] | Daniel Kwame Amissah[1] | David Ebo Adjepon-Yamoah[2]

[1]Department of Computer Science, University of Ghana, Legon, Ghana | [2]Department of Computer Science and Information Systems, Ashesi University, Berekuso, Ghana

Correspondence: Daniel Kwame Amissah (damissah002@st.ug.edu.gh)

Received: 27 February 2025 | **Revised:** 11 May 2025 | **Accepted:** 13 June 2025

Funding: The authors received no specific funding for this work.

Keywords: age-invariant face recognition | dimensionality reduction | feature extraction | local binary patterns | machine learning | scale-invariant feature transform | vision transformers

ABSTRACT

This study presents a robust age-invariant face recognition framework, addressing challenges posed by age-related facial variations. Evaluated on the FGNet and Morph II datasets, the system integrates Viola-Jones for face detection, SIFT and LBP for feature extraction, and Vision Transformers (ViTs) for global feature representation. Feature fusion and dimensionality reduction (KPCA, IPCA, UMAP) enhance efficiency while retaining key discriminative information. Using Random Forest, KNN, and XGBoost classifiers, the model achieves 96% accuracy, demonstrating the effectiveness of combining traditional and deep learning techniques in advancing age-invariant face recognition.

1 | Introduction

Biometrics involves analyzing and measuring distinct physical and behavioral traits to verify or identify individuals [1]. The field is classified into two primary divisions: physical and behavioral biometrics [2]. Physical biometrics quantify distinctive physical characteristics specific to individuals, including fingerprint, iris, and face recognition [3]. Face recognition technology analyzes facial features and their arrangement relative to each other [4]. This technology has broad applications, including smartphone unlocking and surveillance [5], making it a contactless and unique biometric identification method [6]. Face recognition involves several stages. It begins with data acquisition, where facial images are captured using cameras or other imaging devices, categorized as either 2D or 3D. During feature extraction, algorithms identify key facial landmarks, such as the eyes, nose, mouth, and jawline. These features are converted into a numerical face template, encoding distinctive patterns [7]. Matching follows, where verification (one-to-one comparison) or identification (one-to-many matching) occurs. Verification is often used in device authentication, while identification supports security and surveillance systems by matching an individual's face against a database.

Facial recognition is used in diverse applications. In security, it helps monitor and identify people in settings like airports and stadiums. For access control, it secures access to physical spaces and devices. In law enforcement, it aids in locating suspects and missing persons using facial databases [6]. This technology offers non-invasiveness, ease of implementation, and scalability but faces limitations such as privacy concerns, susceptibility to spoofing, and potential biases based on ethnicity, age, or gender, which affect system reliability and accuracy. Environmental factors, facial expressions, and the natural aging process can also lead to misidentification [8].

Age-related changes challenge the accuracy of face recognition systems. Natural aging processes cause a decline in skin elasticity, resulting in wrinkles and other changes that impact recognition based on texture and shape [9]. Alterations

in fat distribution and bone structure can affect the contours of the cheeks, jawline, and eyes, making it challenging for algorithms reliant on geometry to recognize older faces accurately. Makeup and facial hair variations across age groups may also impact accuracy if systems are not designed to account for these changes [10]. Aging also creates inconsistencies within databases, as older photos may not reflect current appearances, leading to increased false rejection rates (FRR) [11]. This issue is particularly evident in operational systems such as passport control and surveillance databases where templates are static and rarely updated. Without updated templates, recognition performance degrades over time. When models are trained primarily on younger faces, their generalization capability decreases for older faces, affecting performance [12]. Solutions to aging-related accuracy issues include longitudinal data gathering, which involves tracking individuals over time to create dynamic templates that adjust for age-related changes. Updated templates can improve accuracy by accounting for progressive facial changes. Advanced algorithms, particularly deep learning models trained on diverse datasets, enhance performance by adapting to age-related variations. Multimodal approaches—such as integrating face recognition with other biometric modalities like voice or fingerprint recognition—introduce additional verification levels to offset aging effects [13]. Age-invariance in face recognition is critical for maintaining accurate identification over time, as facial features naturally evolve [14]. This concept requires algorithms to adapt to ensure consistent performance across ages. The aging process impacts skin elasticity, resulting in wrinkles, and alters bone density and facial fat distribution, making recognition challenging due to modified facial contours and proportions [15]. Age-invariance requires systems to recognize individuals accurately despite age-related changes in skin texture and bone structure.

Age-related challenges affect feature extraction and matching, leading to higher false rejection and acceptance rates, especially if templates are not updated to reflect current appearances [16]. Systems trained on datasets lacking age diversity also risk biased accuracy, particularly underperforming for older age groups. Age-invariant solutions involve longitudinal data collection for updated references and emphasize extracting stable features like bone structure. High-quality images improve age-invariant feature extraction, and regular system calibration helps maintain consistency [17]. To address these issues, enhanced algorithms, such as deep learning models trained on diverse age datasets, improve generalization across age groups. Multimodal biometrics, which combine face recognition with other modalities, offer additional verification and reduce exclusive reliance on facial features [18]. High-resolution images support accurate extraction of stable features despite aging effects. Continuous system calibration and maintenance ensure stable performance over time, especially as users' appearances evolve.

Despite various solutions, achieving accurate and robust face recognition across different age groups remains a significant and ongoing challenge. Prior studies have introduced algorithms that struggle under real-world conditions, where variations in pose, illumination, blur, and distance frequently distort facial features and compromise recognition accuracy [19]. To address these limitations, this paper proposes a multi-feature classification framework that integrates Local Binary Patterns (LBP), Scale-Invariant Feature Transform (SIFT), and a modified Vision Transformer (ViT) architecture to improve age-invariant face recognition performance. The rest of the paper is structured as follows: Section 2 presents related work. Section 3 describes the proposed methodology. Section 4 presents and discusses the experimental results. Section 5 concludes the paper and outlines directions for future work. The main contributions include:

i. Hybrid Feature Extraction Pipeline: The paper introduces a novel approach by combining three distinct feature extraction methods—Local Binary Patterns (LBP), Scale-Invariant Feature Transform (SIFT), and Vision Transformer (ViT)—for age-invariant face recognition. This combination leverages the complementary strengths of texture-based (LBP), keypoint-based (SIFT), and deep learning-based (ViT) features, providing a rich, multi-dimensional representation of facial characteristics across different age groups.

ii. Improved Age-Invariant Face Recognition: The proposed framework significantly enhances face recognition accuracy across age groups, particularly for older individuals. By integrating multiple feature extraction techniques, the model is more robust to age-related changes in facial structure, such as wrinkles, skin texture, and bone structure alterations, which are typically challenging for traditional recognition systems.

2 | Related Studies

Huang, Zhang, & Shan, [20] introduced the MTL-Face framework, a multi-task system for age-invariant face recognition (AIFR) and face synthesis to improve model interpretability. It used attention-based feature decomposition to separate age and identity features, and an identity-conditional module with weight-sharing for smooth age transitions in synthesized faces. Selective fine-tuning enhanced AIFR by using high-quality synthesized faces. The model achieved accuracy rates of 96.45% on AgeDB-30, 95.98% on CALFW, 99.58% on CACD-VS, and 95% on FGNet, with both tasks benefiting each other. However, its reliance on synthesized faces might introduce artifacts, potentially affecting real-world performance. Hou, Qiu, Wan, & Tang, [18], developed a method to encode features into an age-invariant representation by learning a space map, with the aim of removing aging effects while preserving personalized facial traits. The method involved two stages: first, mapping features to reduce aging effects, and then applying temporal and boundary constraints during encoding to improve recognition accuracy. Tested on the Cross-Age Celebrity Dataset (CACD) using LBP and deep features, it reportedly outperformed human accuracy in face verification on the CACD-VS dataset. However, the reliance on high-dimensional features may lead to increased computational complexity. Facial images frequently exhibit variations in appearance, including scale, translation, rotation, lighting, and color cast. Sajid et al. [9] used a preprocessing technique combining a luminance model with Gaussian difference filtering to segment facial images into 128 × 64-pixel sections.

These sections were mirrored to create half-face images that highlight asymmetric facial variations for demographic analysis. Three CNNs were trained to classify age, gender, and race: Network A for age, and Networks G and R for gender and race, respectively. The approach was evaluated on the MORPH II and FERET datasets. A potential limitation is that the method's reliance on asymmetric facial features may overlook symmetric variations, possibly reducing classification accuracy. Riaz et al. [21] proposed a 3D gender-specific aging model designed to be robust to variations in aging and facial position. This model used 2D facial photos from datasets including PCSO, BROWNS, Celebrities, Private, and FG-NET. The method was tested on FG-NET and MORPH-Album2, employing the VGG face CNN descriptor for matching. The system achieved 83.89% rank-1 accuracy and 43.24% TAR on FG-NET, and PCA and LDA were used to improve face verification performance. LDA particularly enhanced accuracy. Tripathi & Jalal, [13], introduced a robust local descriptor for age-invariant face recognition (AIFR) to address the limitations of fixed encoding in descriptors. The method computes the successive kth neighboring difference of pixel intensities across localized face regions, focusing on face detection, feature extraction, and classification using a Support Vector Machine. The Rank-1 recognition performance was evaluated via a 5-fold cross-validation. A 4×4 local region was used to calculate differences, with pixel intensities ordered in a matrix to form histograms as discriminative feature vectors. This approach effectively handles variations in pose, illumination, and expression (PIE) alongside age. The proposed approach was evaluated on FGNET and achieved 80.25% accuracy. However, the increased complexity of texture and shape may introduce computational inefficiencies.

Zhang, Wana, Gao, & Tao, [22] developed a backbone network to extract facial features, followed by the extraction of age-dependent and age-independent features through fully connected layers. These features were classified using an Age Discriminator and ID Discriminator. Python libraries like dlib, OpenCV, and shutil were used for face detection and alignment, with images resized to 96×112 pixels. The model achieved 0.984 accuracy on the CACD-VS dataset and 98.45% on the Morph dataset. Moustafa, Elnakib, & Areed, [19] developed an age-invariant facial recognition system comprising four stages: preprocessing, feature extraction, feature fusion, and classification. It utilized the Viola-Jones method and frontal face alignment for detection, extracted features with the VGG-Face model, and integrated them through multi-discriminant correlation analysis. The system assessed K-nearest neighbors and support vector machines for classification. The system employed feature-level fusion to get superior recognition outcomes compared to decision-level fusion. Multimodal discriminant correlation analysis (MDCA) was employed for fusion, capable of integrating any quantity of input pairs. Experiments performed on two established face-aging datasets, FGNET and MORPH, reveal a Rank-1 recognition accuracy of 81.5% for FGNET and 96.5% for MORPH. Ren, Wang, & Li, [23] proposed a deep learning system including several attention processes to extract age-invariant identification features obscured by age-variant ageing characteristics. The methodology focused on enhancing cross-age facial recognition by employing a hierarchical local pooling mechanism and the Multiple Attention Mechanism Network (MAM-CNN). It introduced a Step Local Pooling (SLP) block for refined feature extraction. The preprocessing involved sorting facial images into six age categories and training on the CASIA-WebFace dataset using ResNet-50 as the backbone network. This approach utilized residual learning to address gradient vanishing and integrated various attention methods to improve identification features. The model was further refined with cross-age datasets, emphasizing critical facial regions to reduce age-related distractions. Experiments on the MORPH and CACD-VS datasets showed performance improvements of 96.0% and 52.0% over state-of-the-art algorithms. In the work of Okokpujie, John, Ndujiuba, Badejo, & Osaghae, [24], during preprocessing, four types of noise were incorporated into the dataset to enhance the extraction of trait ageing facial characteristics and the classification training model. The model represented a modification of the pre-trained convolutional neural network architecture Inception-ResNet-v2. It attained a recognition accuracy of 99.94%, with a mean square error of 0.0158 and a mean absolute error of 0.0637. The network aimed to enhance intra-class subject recognition among various age groups through data augmentation techniques. The model utilized a pre-trained CNN architecture (Inception-ResNetv-2) in conjunction with the Softmax classifier for multiclass classification.

The work of Wu, Du, & Hu, [25] proposed a Parallel Multi-path Age Distinguish Network (PMADN) model to overcome the shortcomings of conventional age-based facial recognition techniques. The model comprised two sequential networks: an Age Distinguish Mapping Network (ADMN) and a Cross-Age Feature Recombination Network (CFRN). ADMN categorized facial traits into distinct age groups via parallel multi-path fully linked layers, which they contended are superior at extracting identity features within a narrow age range. CFRN nonlinearly recombined mapped data to obtain age-invariant characteristics advantageous for identity classification, circumventing the straightforward linear amalgamation of identity and age factors employed in current methodologies. The algorithm integrated transfer learning, using solely the age label and a pre-trained conventional face recognition network for training, thus rendering it appropriate for extensive ageing face datasets. The PMADN model integrates the decomposition method with the age synthesis algorithm, presenting a novel non-linear recombination model to depict the cross-age characteristics of the human face. It developed an Age Distinguish Mapping Network (ADMN) that maps facial features into distinct age subspaces and utilizes a Cross-Age Feature Recombination Network (CFRN) to extract age-invariant features. Experiments conducted on three prominent face ageing datasets, MORPH Album 2, CACD-VS, and Cross-Age LFW, illustrate the efficacy and benefits of the proposed model. Agrawal, Kumar, Kumar, & Thomas [26] introduced a new age function modeling approach that combined local features derived from texture descriptors. The method involved image normalization, feature extraction, and dimensionality reduction using PCA. They utilized the Extreme Learning Machine (ELM) classifier to assess the processed images against the originals, resulting in enhanced generalization performance. The initial process involved image normalization, followed by feature extraction using the LPQ texture descriptor, WLD, and LBP. The

Galactic Swarm Optimisation (GSO) algorithm was employed to implement the optimal feature selection technique. The ELM classifier was then employed for age function modeling. Their research concentrated on local characteristics, including methods such as Local Binary Pattern, Weber Local Descriptor, and Local Phase Quantisation for feature extraction. The feature fusion level was employed for concatenating features, integrating multi-scale LPQ, WLD, and LBP texture descriptors with the retrieved texture features. The concatenation method augmented the feature vector's dimensions, whereas PCA was employed to diminish dimensionality, hence enhancing data variances. The global representation of features was obtained to diminish dimensionality, and the classification method is conducted on the retrieved features. The proposed project consists of two phases: verification and updating. During the verification phase, testing is performed, whereas in the updating phase, the dataset is refined utilizing the GSO method. Their approach achieved an accuracy of 95.0%. Li & Lee [27] proposed an innovative attention-based feature decomposition model, capable of acquiring more discriminative feature representations while minimizing the interference attributed to ageing. The study implemented a channel attention block-based feature decomposition module to extract age-invariant identity features from facial representations. This end-to-end approach directly learned these characteristics, reducing the training complexity of multi-stage methods. A direct sum loss function was introduced to diminish the influence of age-related features. The proposed AEFN system consisted of five components: a convolutional neural network (CNN) backbone, the EFDM, an age classifier, an identity classifier, and a direct sum module, which collaboratively decomposed age-invariant facial features. The EFDM generated the identity-specific feature map, while the classifiers and direct sum module produced the age-invariant features. To augment discriminative identification power, Curricular Face loss was employed to oversee the acquisition of the identity-preserving component and guarantee the retention of identity-preserving information. An auxiliary age discriminator guided the decomposition process to reveal intrinsic age indicators. The SoftMax and cross-entropy losses were used in the age classification task to align predicted and actual ages. Various cross-age datasets were employed for training and evaluation, achieving average accuracies of 96.0%, 99.0%, and 98.0% on the CALFW, CACD-VS, and AgeDB-30 datasets, respectively. A study by Wang H., Gong, Li, & Liu [28] proposed a method to exclude age-related components from attributes that are conflated with both identification and age information. The methodology involved decomposing a composite facial feature into two distinct components: one dependent on identity and the other on age. The researchers established the Decorrelated Adversarial Learning (DAL) method, which utilized a Canonical Mapping Module (CMM) to determine the maximum correlation between paired data generated by a backbone network. Both the backbone network and the factorization module were trained to produce features that reduced correlation, resulting in a significant decrease in the characteristics related to age and identity. The identity-dependent and age-dependent characteristics were regulated by signals retaining identification and age to guarantee the accuracy of the information contained. Comprehensive experiments were performed on widely used public-domain facial ageing datasets to validate the efficacy of the proposed method. The Decorrelated Adversarial Learning (DAL) algorithm calculates the canonical correlation between paired features of decomposed components, employing stochastic gradient descent (SGD) optimization. The multi-task training method facilitated the acquisition of deconstructed features via three primary supervision modules: age discriminator, identity discriminator, and DAL regularizer. The model achieved accuracies of 94.50%, 98.30%, and 99.40% on the FGNET, MORPH Album 2, and CACD-VS datasets, respectively.

The body of research on age-invariant face recognition (AIFR) includes subspace learning (Mittal & Patel, 2023), deep learning techniques [15], and disentangled representation [18] approaches. Although each category has made significant progress, they present notable shortcomings. Subspace-based methods like PCA and LDA rely on handcrafted features and are vulnerable to intra-class variations, reducing their robustness under aging. Deep learning-based models, despite their high performance on controlled datasets, often struggle under large age gaps due to their inability to fully disentangle age-related and identity-preserving features. Methods based on representation disentanglement aim to separate identity and age factors but tend to overfit on constrained datasets, limiting their applicability in real-world scenarios. Many still depend on shallow or handcrafted features, which lack the expressive power to model long-term facial changes. Deep models trained with limited aging supervision frequently fail to perform reliably under extreme age differences, as well as variations in pose, lighting, or occlusions. Moreover, CNN-based models, though effective at capturing fine-grained details, are insufficient in modeling global dependencies necessary for consistent identity representation across different age stages. Additionally, transformer-based networks remain underexplored in the AIFR domain, especially in their capacity to combine localized and global facial representations.

Furthermore, some areas remain unaddressed in the existing literature. Few studies have developed unified frameworks that integrate both local and global features. The potential of transformer architectures, particularly for achieving age-invariant recognition, is yet to be fully exploited. Moreover, there is limited emphasis on evaluating model performance in unconstrained or real-world settings. Specifically, there is a lack of approaches that leverage the strengths of both handcrafted descriptors, such as LBP and SIFT, and transformer models to capture the complementary aspects of facial representation. To address these challenges, this study proposes a hybrid framework that integrates LBP and SIFT for extracting robust local texture features with Vision Transformers (ViT) for modeling long-range dependencies. This combined approach targets the key limitations of existing methods by enhancing both local robustness and global consistency. The local descriptors improve the model's resilience to aging variations, while the ViT ensures comprehensive structural understanding, thereby promoting better generalization across extreme age gaps and diverse environmental conditions.

3 | Materials and Method

Figure 1 illustrates the methodological approach that will be employed in this investigation. The methodology comprises

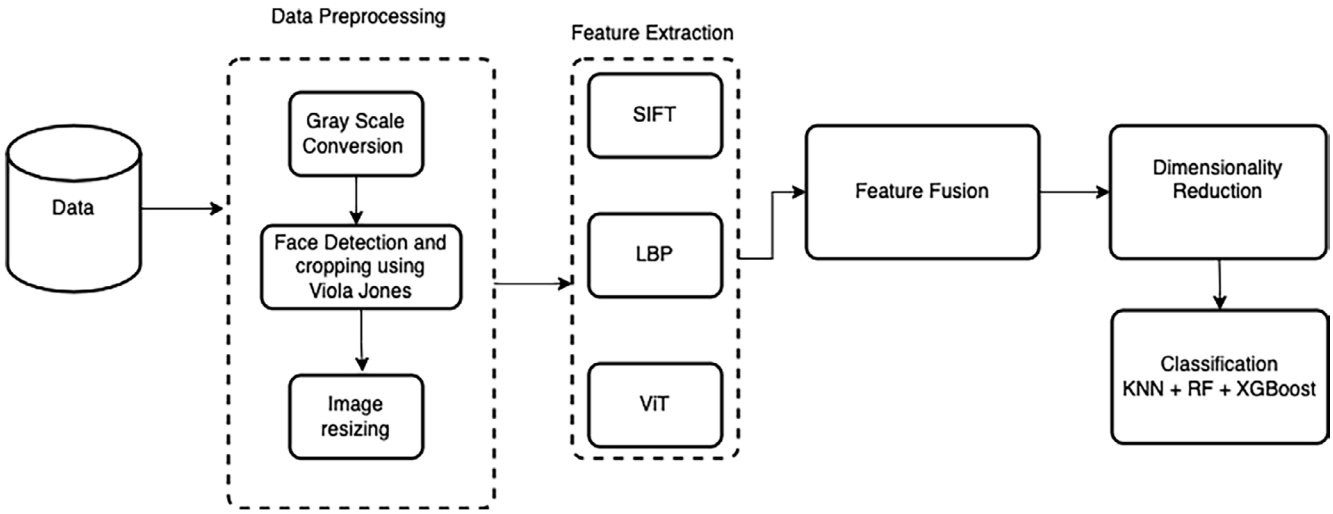

FIGURE 1 | Methodological approach adopted in this study.

pre-processing, feature extraction using ViTs, and classification phases, which will be addressed in depth in the following sections. By combining SIFT with Local Binary Patterns (LBP) and Vision Transformers (ViTs), a complementary feature extraction approach was established. SIFT captured local stable features, LBP provided local texture information, and ViTs captured global contextual representations, enhancing model accuracy and robustness. SIFT's integration thus ensures a comprehensive, distortion-resilient approach to feature extraction [29].

3.1 | Dataset Selection & Description

The FG-NET Aging Database is a dataset that consists of 1002 photos of 82 individuals, with each individual having between 6 and 18 images taken at different ages. These images capture the aging process from birth to 69 years of age, with a higher concentration of images in the younger age brackets (below 30 years). It is utilized for the purpose of estimating age, recognizing faces, and doing research on age progression and regression. The photographs exhibit variations in lighting, facial expression, and stance, and each image is accompanied by annotations indicating age and facial landmarks. The dataset can be used for non-commercial research purposes and is valuable; however, it has limited diversity and size. Figure 2 represents sample images from the FGNet database. Table 1 shows the partition scheme for the FGNet database (Figure 3).

The MORPH II Dataset contains approximately 55,000 facial images from around 13,000 unique individuals. Most subjects have 2 to 5 images captured at different times, making it suitable for longitudinal studies such as age progression and estimation. The age range spans from 16 to 77 years, with the majority of samples falling between 20 and 50 years. The dataset is relatively balanced compared to FG-NET but still exhibits a bias toward middle-aged males, particularly those of African-American descent. According to the dataset creator, the dataset starts at age 16 and so the age labels are scaled such that age 16 maps to label 0, age 17 to label 1, and so on, up to age 77 mapping to label 61. This results in 62 distinct age labels.

3.2 | Image Processing

We used OpenCV to convert images to grayscale, simplifying data processing for face detection and age estimation by focusing on intensity rather than color. The Viola-Jones [30] algorithm was applied for efficient, real-time face detection using Haar-like features, integral images, AdaBoost, and a cascade classifier. Haar-like features capture image edge and texture variations critical for facial detection. Integral images accelerate feature computation, while AdaBoost selects optimal features by training weak classifiers to enhance accuracy. The cascade classifier progressively discards non-facial regions for speed and efficiency. We trained the model on annotated face and non-face images, applying a 1.1 scaling factor for multi-scale detection, with a minimum detection size of 30×30 pixels to optimize accuracy and reduce false positives. Figure 4 shows an image before and after facial detection.

3.3 | Multi-Feature Extraction

3.3.1 | The Scale-Invariant Feature Transform (SIFT)

SIFT, developed by David Lowe, is a robust method for extracting distinctive local image features that are invariant to scale, rotation, and partial occlusion [31]. In this study, SIFT is employed to extract stable keypoint-based features that aid in consistent facial representation under varying imaging conditions.

3.3.2 | Local Binary Patterns

LBP is a simple and efficient texture descriptor widely used in image analysis for its robustness to illumination changes and ability to capture local texture variations [32]. In this study, LBP is used alongside SIFT and Vision Transformers

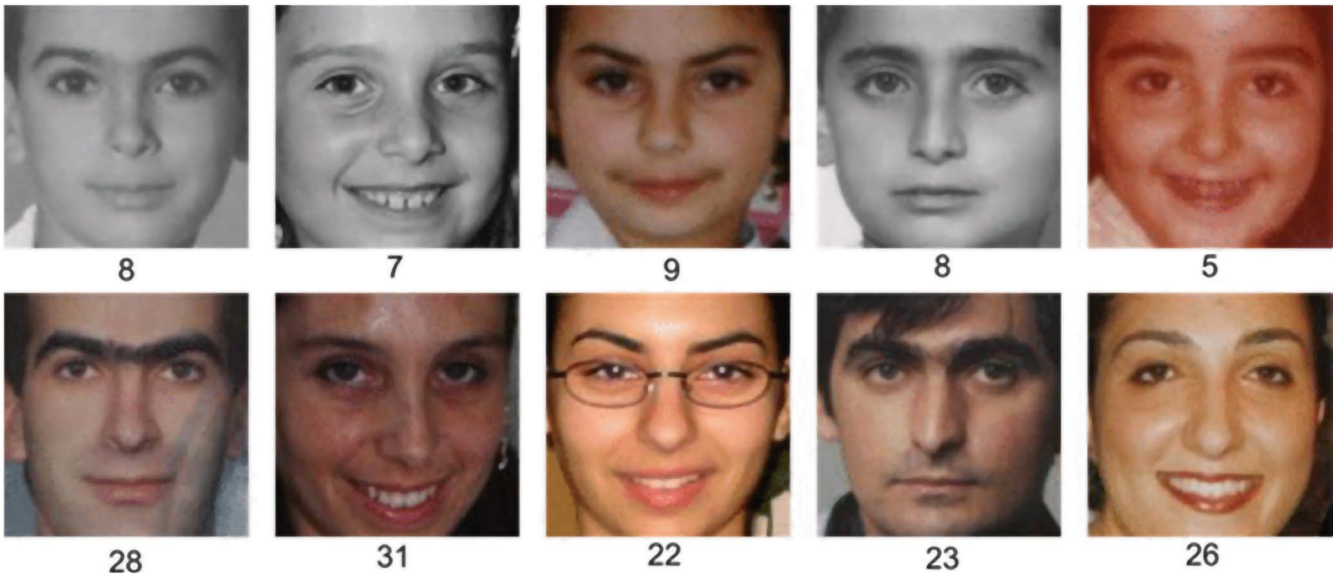

FIGURE 2 | Examples of face images from five different persons across different ages in the FG-NET database with different age values, where each column is face images of the same person captured at different age values and the number below each image is the age value of the person [14].

TABLE 1 | Partition scheme for FGNET database.

Age group	Age range
1	0–2
2	3–5
3	6–8
4	9–12
5	13–16
6	17–21
7	22–30
8	31–69

(ViT) in a multi-feature fusion framework to enhance texture representation.

3.3.3 | Vision Transformer

The Vision Transformer (ViT) architecture by Dosovitskiy et al. [33] is a transformer-based model originally designed for natural language processing and adapted to image recognition. Unlike convolutional neural networks (CNNs), which rely on convolutional layers, ViTs employ self-attention mechanisms to process images divided into non-overlapping patches. Each patch is embedded in a high-dimensional space and enriched with positional information, allowing the model to learn spatial relationships. These embedded patches are processed by a Transformer encoder comprising multi-head self-attention layers and feed-forward networks, stabilized by layer normalization and residual connections. A classification token is appended to the sequence, with the encoder output used for final classification. Trained on large datasets, ViTs can achieve or even exceed CNN performance on multiple benchmarks, particularly when pre-trained on massive datasets and fine-tuned for specific tasks. The use of Vision Transformers (ViTs) in this study for feature extraction alongside Scale-Invariant Feature Transform (SIFT) and Local Binary Patterns (LBP) provides a robust multifeature fusion framework. ViTs offer a unique capability to capture long-range dependencies and global contextual relationships within images, making them ideal for high-level feature representation. Unlike SIFT and LBP, which focus on local textures and patterns, ViTs can capture broader, complex patterns across entire images. This complementary blend of features—local descriptors from SIFT and LBP, and global features from ViTs—enhances the model's ability to discern complex visual patterns, supporting a more comprehensive analysis in visual recognition tasks.

3.4 | Feature Fusion

Feature fusion is a crucial process in machine learning and computer vision, aiming to integrate information from multiple sources to create a unified, enriched representation that enhances model performance. It is particularly useful when different data sources provide complementary insights. Three primary fusion techniques exist: early fusion, late fusion, and hybrid fusion [34]. Early fusion, also known as feature-level fusion, combines raw features from different sources into a single feature vector before they enter the model. While this approach can lead to high dimensionality and redundancy, it simplifies the model architecture by enabling it to work with a single, cohesive feature set [35]. Late fusion, in contrast, processes features separately through distinct models, combining the outputs through averaging, voting, or concatenation, which allows specialization but adds complexity to training. Hybrid fusion merges early and late fusion advantages by integrating features at intermediate stages [36]. In this study, early fusion was chosen to combine features extracted from SIFT, LBP, and Vision Transformers (ViTs). This approach was selected due to its ability to provide a comprehensive, integrated feature representation that captures diverse aspects

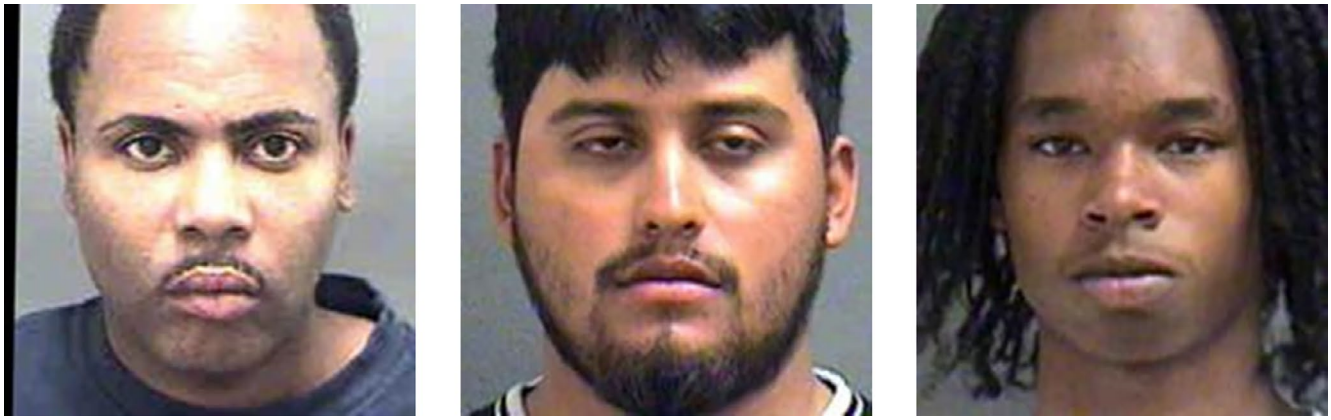

FIGURE 3 | Sample images from the MORPH II dataset.

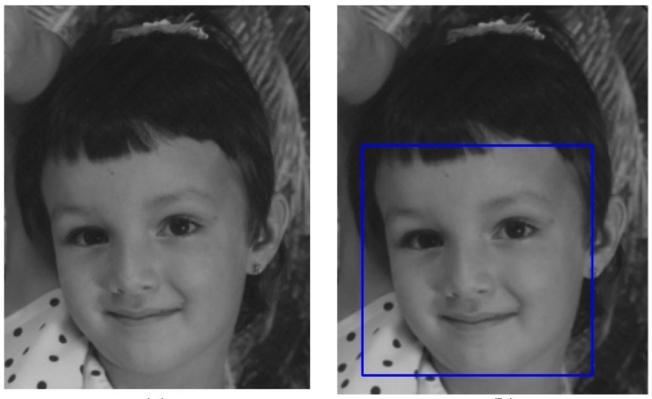

FIGURE 4 | Image with bounding box after Viola-Jones Detection. (a) Original image. (b) Face detected with Viola-Jones algorithm.

of the data at the initial stage. By concatenating these features into a unified vector, early fusion maximizes complementary insights from each extraction method—SIFT's local texture, LBP's spatial patterns, and ViTs' high-level visual features—thereby improving the model's ability to discern complex patterns with greater accuracy. Additionally, early fusion reduces model complexity by streamlining the input to a single feature set, thus optimizing computational efficiency in training and inference.

3.5 | Dimensionality Reduction

3.5.1 | Incremental Principal Components Analysis

In this study, Kernel Principal Component Analysis (KPCA) was first applied to the dataset following feature extraction using SIFT, LBP, and Vision Transformers (ViTs). KPCA extends traditional PCA by utilizing the kernel trick to manage nonlinear relationships within the data, identifying principal components that capture complex patterns beyond the reach of standard PCA. This approach improves the dataset's representation of complex structures, making it suitable for applications requiring detailed feature analysis. By capturing non-linear relationships in facial features, particularly those representing age-related changes, KPCA transforms the data into a feature space where these patterns are more linearly separable, simplifying subsequent processing by IPCA. Incremental Principal Component Analysis (IPCA) was then used to further reduce dimensionality. IPCA is especially advantageous for large datasets that exceed memory capacity or are processed in smaller batches, as in streaming contexts. This technique enables efficient iterative updates of principal components, allowing for dimensionality reduction while preserving significant variance from the combined feature set. The combined use of KPCA followed by IPCA provides robust feature extraction through fusion while effectively managing the dimensionality of complex datasets. This two-step process ensures that both linear and nonlinear structures are adequately represented, thus enhancing the overall performance of subsequent analysis and modeling efforts.

3.5.2 | Uniform Manifold Approximation and Projection (UMAP)

The Uniform Manifold Approximation and Projection (UMAP) algorithm is a powerful non-linear dimensionality reduction technique grounded in manifold learning and topological data analysis [37]. By constructing a weighted graph based on local neighborhood relationships and a chosen distance metric, UMAP generates a fuzzy topological structure that preserves both local and global features of high-dimensional data. This balance is achieved by minimizing a cross-entropy loss function, enabling accurate representation of the data's complex topology. Compared to methods such as t-SNE, UMAP exhibits superior scalability and computational efficiency, effectively capturing global patterns while being applicable to various data types. These characteristics make UMAP particularly suitable for tasks such as high-dimensional data visualization, feature reduction, and clustering in fields like bioinformatics and natural language processing [38]. The underlying mathematical framework of UMAP combines elements from graph theory, differential geometry, and stochastic optimization to faithfully preserve data structure during the dimensionality reduction process. Then to minimize the discrepancy between the representations in the high-dimensional space and their low-dimensional counterparts, an objective function is formulated based on a

cross-entropy loss. UMAP was used in this study because of its ability to effectively capture and preserve both local and global structures in high-dimensional data. Unlike linear dimensionality reduction techniques, UMAP is well-suited for datasets with complex, non-linear relationships, such as those encountered in age-invariant face recognition. After feature extraction using SIFT, LBP, and ViTs, UMAP was applied to reduce the dimensionality while retaining the critical variations in facial features across different age groups. This is particularly important for age-invariant recognition, as subtle changes in facial appearance must be accurately represented. UMAP's capacity to maintain discriminative features ensures that the key characteristics of the face are preserved, improving the performance and robustness of the recognition model.

3.6 | Classification Algorithms and Model Training

K-Nearest Neighbors (KNN), Random Forest (RF), and XGBoost are used to evaluate age-invariant face recognition. KNN is selected for its simplicity and effectiveness in classification tasks based on proximity in feature space. RF, as an ensemble method, offers improved robustness and generalization. XGBoost is used for its ability to model complex non-linear relationships efficiently. We apply 5-fold cross-validation to ensure reliable performance evaluation across all classifiers. The task is formulated as a binary verification problem, where the model learns to distinguish between same-identity and different-identity instances. True Positives represent correctly identified same-identity instances, while False Positives indicate incorrect matches. This approach ensures that each classifier learns identity-discriminative features rather than memorizing specific instances.

3.7 | Model Evaluation

For this study, our model was evaluated using accuracy, precision, recall, f1 Score, and Area Under the Curve (AUC) [39].

For a binary classification problem, accuracy can be expressed mathematically as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision in classification and retrieval systems is the ratio of accurately identified relevant instances (true positives) to the total relevant occurrences. Its mathematical form is:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

Recall is defined as the ratio of true positives (TP) to the sum of true positives and false negatives (FN). Mathematically, it is represented as:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{3}$$

The F1 score combines precision and recall into a single metric. The formula for the F1 score is:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

The Area Under the Curve (AUC) measures binary classification model performance. It measures the area under the Receiver Operating Characteristic (ROC) curve, which compares TPR and FPR at various thresholds. The AUC is the integral of the ROC curve from 0 to 1. Mathematically, it can be expressed as:

$$AUC = \int_0^1 \text{TPR}(\text{FPR})\text{d}(\text{FPR}) \tag{5}$$

4 | Results and Analysis

4.1 | Overview

This section presents and discusses the findings of this study. We also present an analysis of our findings. We also compare our findings to existing studies. For this study, all experiments were conducted on a personal computer with 32GB of RAM, core i7 CPU with an Nvidia Geforce RTX 1650 with 4GB of video memory running Ubuntu operating system.

4.2 | Empirical Evidence on FG-Net Data

4.2.1 | Using IPCA for Dimensionality Reduction on the FGNet Data

Figure 5 illustrates how IPCA affects K-Nearest Neighbors (KNN) classification accuracy. The testing accuracy (95.0%) slightly surpasses the training accuracy (93.0%), implying good generalization with limited overfitting. Training precision is lower at 91.5% compared to testing precision of 94.5%, suggesting potential class imbalance or noise in the data. High recall values of approximately 95.0% across both training and testing data indicate the model's sensitivity in identifying a wide range of faces across age groups. However, a reduced F1 score in the training data (92.0%) compared to 94.5% in testing indicates a trade-off between precision and recall due to the complexities in the training dataset. The ROC curve for IPCA-KNN (as shown in Figure 5) further demonstrates the model's proficiency, achieving an Area Under the Curve (AUC) of 96.0%, reflecting a high true positive rate with minimal false positives, emphasizing the model's robust discriminatory ability for the age-invariant face recognition task.

Figure 6 demonstrates Random Forest (RF) classifier performance post-IPCA. The training metrics achieve perfect scores (accuracy, precision, recall, and F1 score of 100%), suggesting potential overfitting due to flawless pattern learning on the training set. Testing metrics are slightly reduced, with an accuracy and F1 score of 96.0%, while test precision remains high at 99.0%, indicating an effective reduction of false positives. However, the recall decreases to 93.0% on the test set, pointing to some missed true positives, which results in a trade-off where

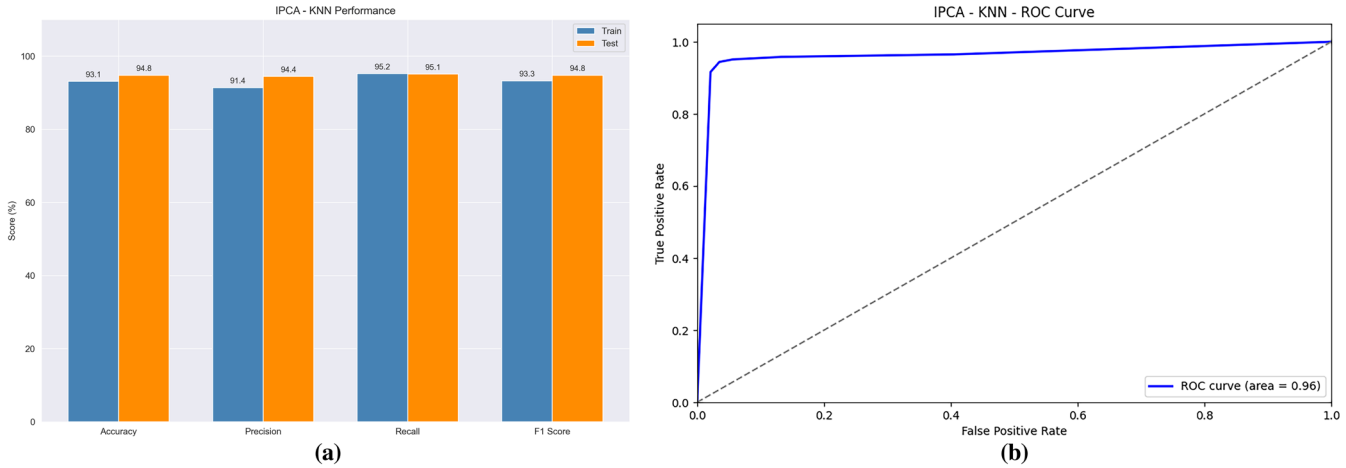

FIGURE 5 | Performance of KNN on FGNET data using IPCA. (a) Performance values. (b) ROC curve.

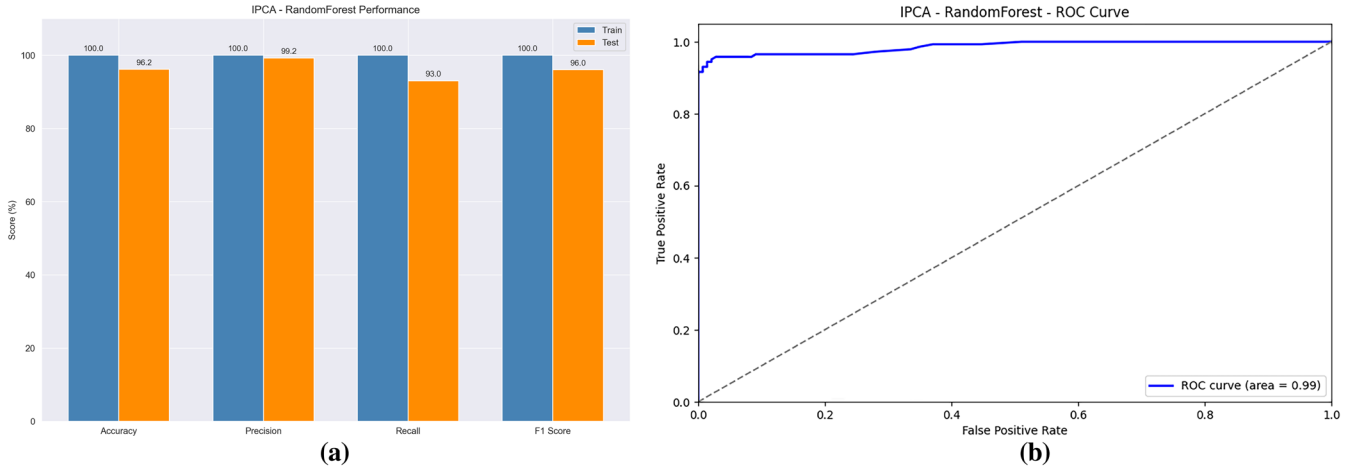

FIGURE 6 | Performance of Random Forest on FGNET with IPCA. (a) Performance values. (b) ROC curve.

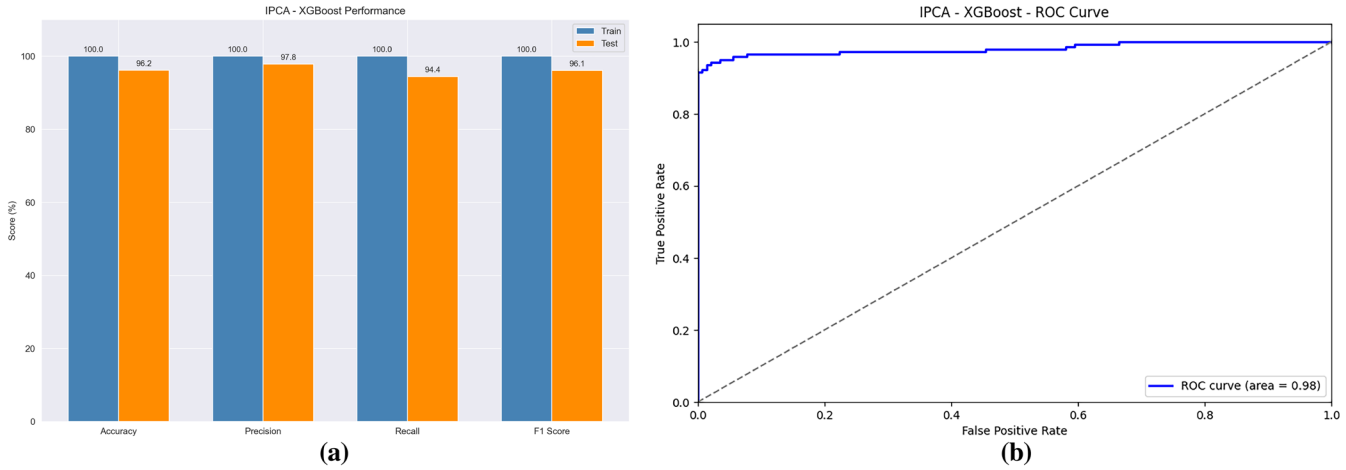

FIGURE 7 | Performance of XGBoost on FGNET with IPCA. (a) Performance values. (b) ROC curve.

precision is prioritized over recall. The ROC curve for IPCA-RF (as shown in Figure 6) reveals a high AUC of 99.0%, highlighting the model's effectiveness in differentiating faces across age categories. The curve's proximity to the top-left corner further emphasizes the model's capability to maintain a high true positive rate while keeping false positives low, making the approach reliable for applications where precise classification with minimal misclassification is crucial.

Figure 7 shows the IPCA-XGBoost classifier performance. Training metrics are perfect at 100%, which, while effective in categorizing all cases, indicate possible overfitting. In testing,

there is a slight dip in accuracy to 96.0% and precision to 98.0%, accompanied by a more notable decline in recall to 96.0%, reflecting a modest sensitivity drop for unseen data. The F1 score aligns closely with recall at around 96.0%, capturing the precision-recall trade-off observed with slightly diminished generalization. The ROC curve for IPCA-XGBoost (as shown in Figure 7) illustrates a high AUC of 98.0%, nearly perfect in distinguishing between positive and negative cases, and the curve's positioning close to the top-left indicates sustained high sensitivity at low false-positive rates. This high AUC validates IPCA-XGBoost's capability for accurate classification, with an effective balance of true positives and minimal false positives. Consequently, IPCA-XGBoost emerges as a dependable model for high-dimensional data, suitable for real-world settings where classification precision is paramount. The combined use of IPCA with XGBoost illustrates the advantages of this approach for predictive accuracy and efficient handling of high-dimensional data in age-invariant face recognition.

Table 2 presents a summary of performance values observed on FGNET data using IPCA as a dimensionality reduction technique.

4.2.2 | Using UMAP for Dimensionality Reduction on the FG-Net Data

The results in Figure 8 show that the KNN model maintains strong generalization capabilities with only a slight drop in accuracy from 91.7% on the training data to 91.0% on the test data. Precision on the training set reached 92.6%, though it declined to 90.0% on the test set, indicating a tendency for higher false positives on new data, suggesting potential overfitting. Conversely, recall improved on the test set, indicating a better detection rate of positive instances but at the cost of increased false positives. The F1 Score remained consistent across training and test sets, highlighting the model's balanced handling of new data. The ROC curve in Figure 8 with an AUC of 96.0% reflects strong model performance, as the KNN classifier reliably distinguishes between classes with minimal error. However, training data results are near perfect, raising concerns about overfitting, suggesting that while UMAP contributed to reducing dimensionality effectively, further optimization might improve generalization on new data. This could involve tuning hyperparameters or experimenting with cross-validation techniques to mitigate overfitting.

The Random Forest model's performance metrics, shown in Figure 9, demonstrate high accuracy (92.0%) on test data, though a discrepancy arises with a perfect score on the training data, pointing to overfitting. The ROC curve in Figure 9 exhibits an AUC of 98.0%, highlighting the model's strong predictive power, though the flawless training performance suggests an over-reliance on training patterns. UMAP's role in the high-dimensional reduction and clustering tasks appears to have improved computational efficiency and supported generalization on test data; however, further tuning or exploring alternative dimensionality reduction approaches could address the overfitting indicated by the discrepancy between training and test performances.

XGBoost also benefited from UMAP's dimensionality reduction, achieving balanced metrics on the test set with a high recall of 97.0% and an AUC of 97.0% as shown in Figure 10. The minor drop in precision to around 90.0% compared to a near-perfect

TABLE 2 | Performance values from our experiment on FGNet Dataset with IPCA.

Model	Train ACC	Train PRE	Train REC	Train F1	Test ACC	Test PRE	Test REC	Test	Test AUC
Random Forest	100%	100%	100%	100%	96.2%	99.3%	93.0%	96.0%	98.7%
KNN	93.1%	91.4%	95.2%	93.3%	94.8%	94.4%	95.1%	94.8%	96.2%
XGBoost	100%	100%	100%	100%	96.2%	97.8%	94.4%	96.1%	98.9%

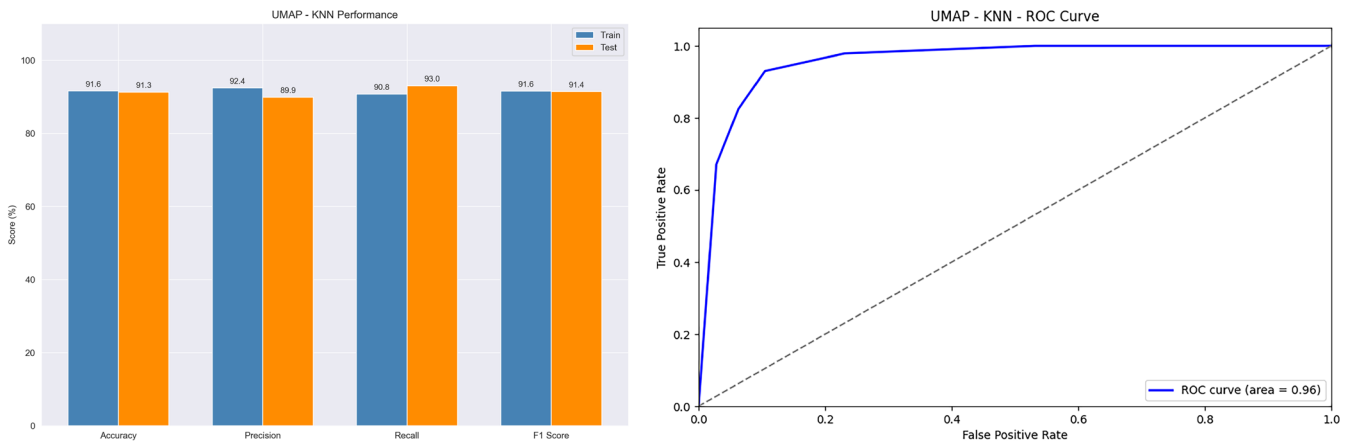

FIGURE 8 | Performance of KNN on FGNET with UMAP. (a) Performance values. (b) ROC curve.

training performance emphasizes the trade-off between true positive rate and false positives, beneficial in contexts where high sensitivity to positive detection is prioritized over minimizing false positives. While XGBoost demonstrated an excellent AUC, indicating substantial discriminative power, the overfitting on training data again points to potential benefits from fine-tuning or cross-validation strategies to reduce variability across datasets.

Table 3 presents a summary of performance values observed on FGNET data using IPCA as a dimensionality reduction technique.

4.3 | Empirical Evidence on MORPH II Dataset

4.3.1 | Using IPCA for Dimensionality Reduction on MORPH II

Figure 11 shows the performance values observed from our experimentation on the Morph II data.

The analysis reveals that Random Forest, despite achieving a high training accuracy (99.0%), suffered from overfitting, with test accuracy dropping significantly to around 79.9%. This discrepancy suggests Random Forest's inability to generalize

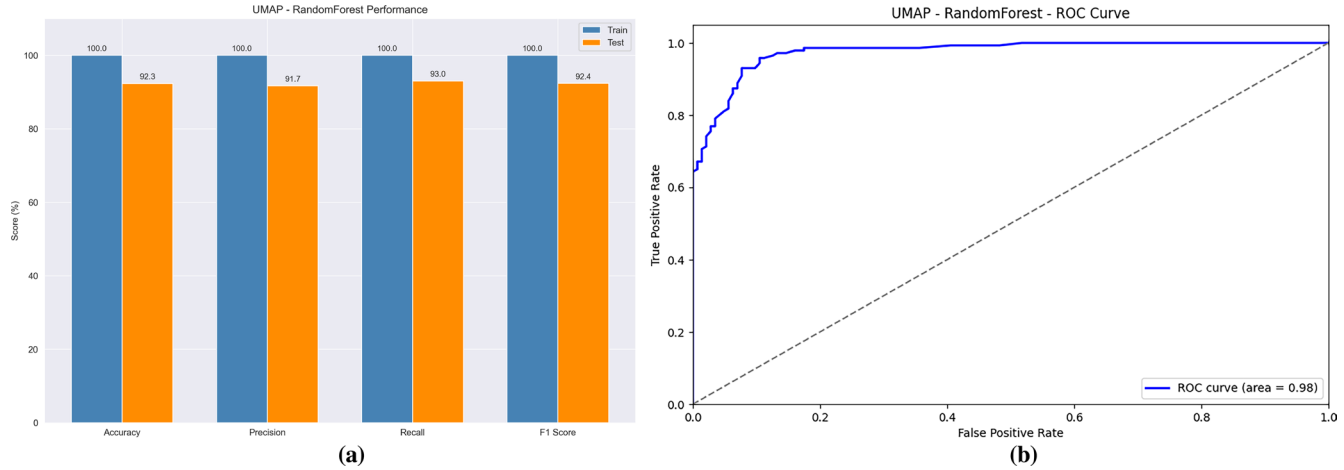

FIGURE 9 | Performance of Random Forest on FGNET with UMAP. (a) Performance values. (b) ROC curve.

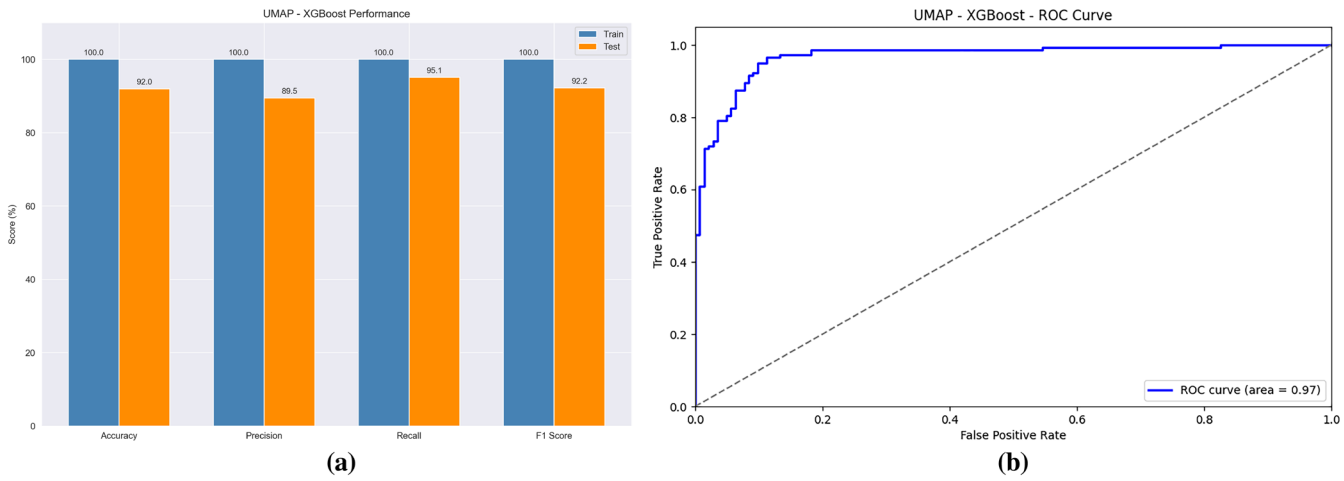

FIGURE 10 | Performance of XGBoost on FGNET with UMAP. (a) Performance values. (b) ROC curve.

TABLE 3 | Performance values from our experiment on FGNet Dataset with UMAP.

Model	Train ACC	Train PRE	Train REC	Train F1	Test ACC	Test PRE	Test REC	Test F1	Test AUC
Random forest	100%	100%	100%	100%	92.3%	91.7%	93.0%	97.5%	97.5%
KNN	91.6%	92.4%	90.7%	91.6%	91.3%	89.9%	93.0%	91.4%	95.9%
XGBoost	100%	100%	100%	100%	91.9%	89.5%	95.1%	92.2%	96.7%

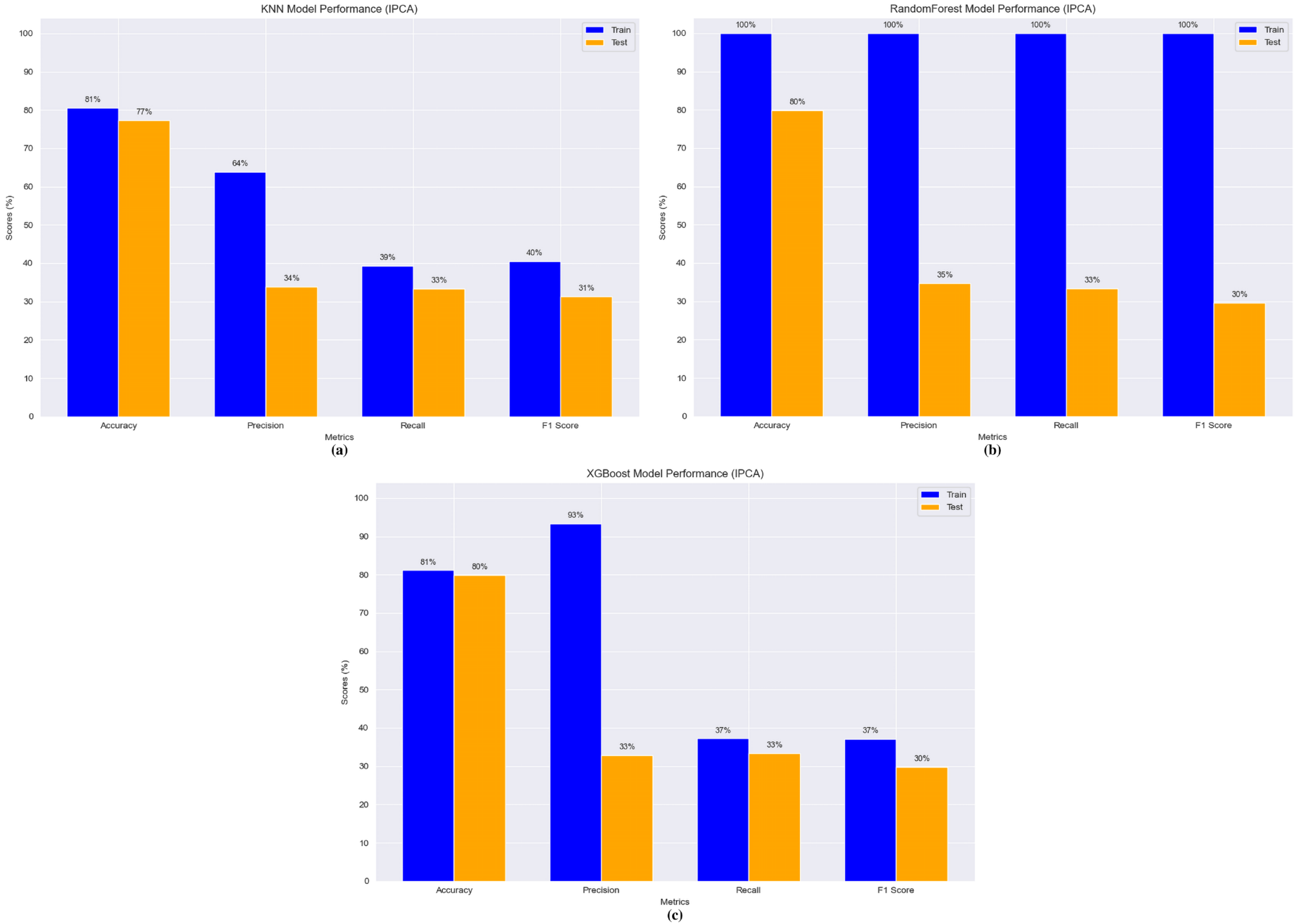

FIGURE 11 | Performance of IPCA on MORPH II data. (a) Performance with KNN classifier. (b) Performance with Random Forest classifier. (c) Performance with XGBoost classifier.

well, as it learns specific patterns from the training data without effectively adapting to new, unseen data. In contrast, K-Nearest Neighbors (KNN) and XGBoost demonstrated better generalization, each achieving more balanced results: KNN with training accuracy at approximately 80.6% and test accuracy around 77.3%, while XGBoost recorded 81.2% for training and 79.9% for testing. These results indicate that KNN and XGBoost, though lower in training accuracy compared to Random Forest, were more reliable in maintaining test performance, showcasing their capability to generalize more effectively in an age-invariant face recognition setting. Precision metrics further emphasized this trend, where Random Forest displayed a steep drop in test precision to below 40.0%, suggesting a high false positive rate in the test data and reinforcing its tendency toward overfitting. Conversely, KNN and XGBoost maintained more stable precision, indicating better handling of the trade-offs between precision and recall across the data sets. This stability in precision highlights these models' suitability for tasks requiring balanced performance across varied conditions. Random Forest's recall scores supported this conclusion; while nearly perfect in training, its recall dropped significantly on the test set, highlighting its lack of robustness for generalization. In contrast, KNN and XGBoost preserved their recall scores across both data sets, proving more reliable for real-world applications where unseen data could lead to unexpected variations. The F1 scores, which harmonize precision and recall, further illustrated the performance trends. Random Forest's high training F1 score was met with a sharp decline in the test F1, reinforcing the model's overfitting tendency. KNN and XGBoost, however, maintained consistent F1 scores, highlighting their balance between precision and recall even in unseen contexts. These findings support the suitability of KNN and XGBoost over Random Forest for age-invariant face recognition tasks, as their stable F1 scores reveal a more robust performance under varied conditions. Applying Incremental Principal Component Analysis (IPCA) for dimensionality reduction produced mixed outcomes. While IPCA accelerated computation, some crucial age-invariant features might have been lost in the dimensionality reduction process, potentially impacting model performance. Notably, the limited generalization ability of Random Forest, despite its high training metrics, suggests that it may not be optimal for age-invariant face recognition applications. Instead, KNN and XGBoost, with their lower but consistent performance, seem more suitable for practical scenarios where reliability across test data is crucial. The comparable test performance of KNN and XGBoost implies that the choice between them could depend on factors beyond predictive performance, such as inference time or interpretability. Table 4 shows a comparison of the various classifiers after applying IPCA on MORPH II data.

4.3.2 | Using UMAP for Dimensionality Reduction for MORPH II Dataset

Figure 12 shows the performance values observed on the Morph II data after applying UMAP.

Random Forest achieved near-perfect training accuracy but exhibited a marked decrease in test performance, indicating overfitting, particularly in the high-dimensional feature space. Both KNN and XGBoost, however, displayed more stable performance between training and testing, with XGBoost slightly surpassing KNN on test data, suggesting its stronger generalization ability. This trend was consistent across multiple evaluation metrics, where Random Forest maintained high training precision but experienced a sharp decline in test precision, likely resulting in higher false positives in the test set. In contrast, XGBoost and KNN demonstrated more balanced precision, with XGBoost slightly outperforming KNN, making it potentially more reliable in minimizing false positives, which is essential in age-invariant face recognition. The recall analysis highlighted the overfitting in Random Forest, which showed significant disparity between training and test recall, with KNN and XGBoost achieving more consistent, albeit lower, recall values in the test phase. F1 scores echoed these findings, reinforcing the overfitting tendency in Random Forest, while KNN and XGBoost maintained steadier performance, though with somewhat reduced F1 values. These results suggest that Random Forest's high-dimensional feature space may have exacerbated its overfitting issue, indicating the need for additional regularization or refined feature selection strategies when using Random Forest for age-invariant recognition tasks. XGBoost's slight but consistent edge over KNN across these metrics suggests it may offer better generalization for MORPH II data after UMAP reduction, though all models showed a performance drop from training to testing, hinting that UMAP might not entirely capture age-invariant characteristics. This highlights the need for further optimization in initial feature extraction methods or refinement in UMAP to preserve age-invariant information. Average training and test performance metrics further reveal the strengths and limitations of UMAP for age-invariant recognition. Training accuracy and precision were high (around 87.35% and 86.24%, respectively), but recall and F1 scores averaged significantly lower (59.37% and 59.87%), suggesting that while models were precise, they lacked in consistently identifying true positives across age variations. In the test phase, the decline in performance was stark, with test accuracy averaging 78.91%, precision at 37.16%, and recall and F1 scores at 33.58% and 30.64%, indicating a substantial gap between training and testing, reinforcing the overfitting hypothesis and pointing to limitations in the UMAP-reduced space's ability to capture robust age-invariant features. When

TABLE 4 | Performance values from our experiment on MORPH II Dataset with IPCA.

Model	Train ACC	Train PRE	Train REC	Train F1	Test ACC	Test PRE	Test REC	Test F1
Random Forest	99.9%	99.9%	99.9%	99.9%	79.9%	35.2%	33.3%	29.7%
KNN	80.6%	63.9%	39.4%	40.5%	77.3%	33.9%	33.4%	31.3%
XGBoost	81.2%	93.3%	37.2%	37.1%	79.9%	32.9%	33.3%	29.7%

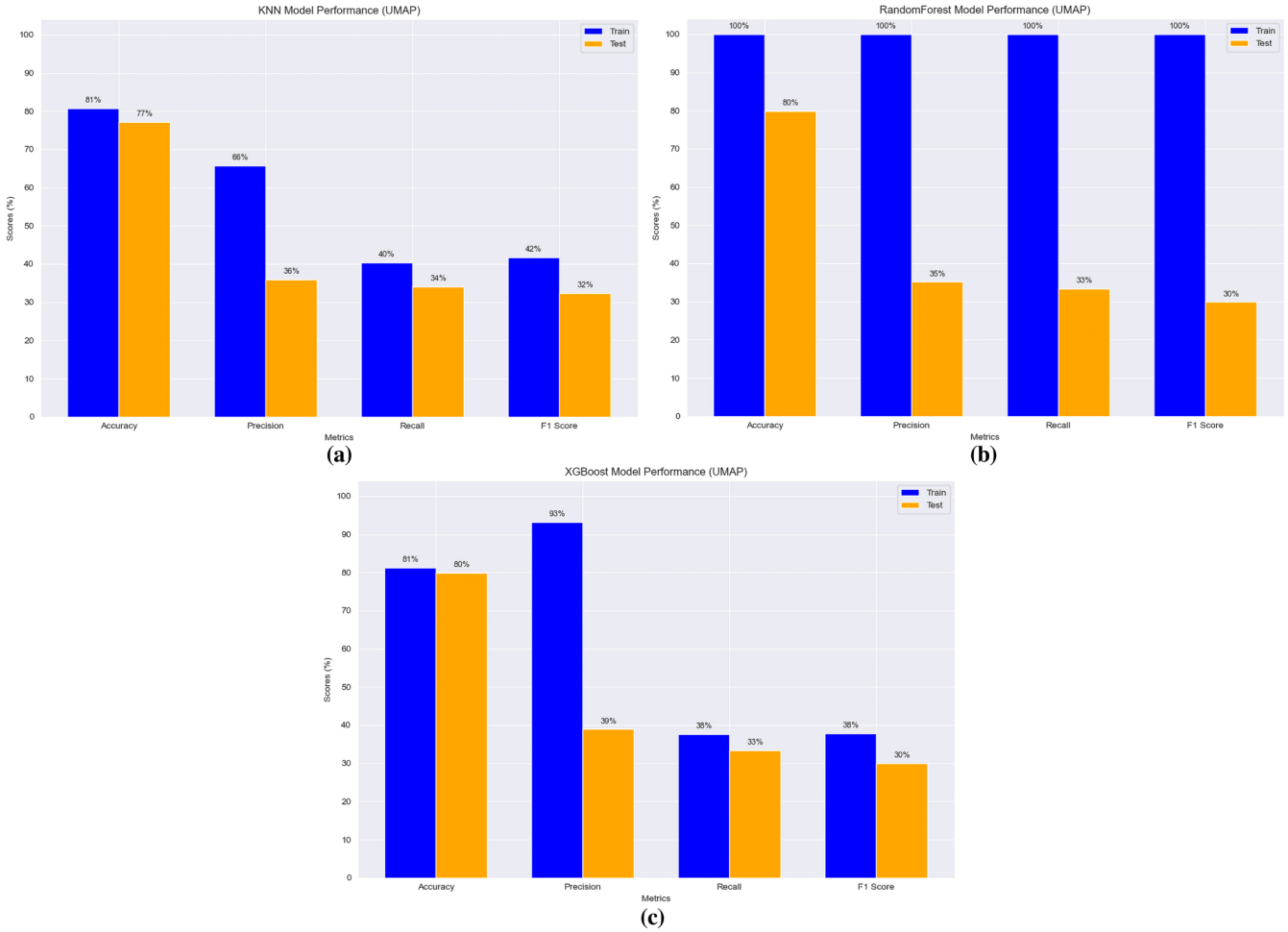

FIGURE 12 | Performance of UMAP on MORPH II data. (a) Performance with KNN classifier. (b) Performance with Random Forest classifier. (c) Performance with XGBoost classifier.

TABLE 5 | Performance values from our experiment on MORPH II Dataset with UMAP.

Model	Train ACC	Train PRE	Train REC	Train F1	Test ACC	Test PRE	Test REC	Test F1
Random Forest	99.9%	99.9%	99.9%	99.9%	79.8%	38.0%	33.4%	29.9%
KNN	80.7%	65.4%	40.3%	41.6%	77.1%	35.1%	33.9%	32.2%
XGBoost	81.3%	93.3%	37.8%	38.0%	79.8%	38.3%	33.4%	29.8%

comparing UMAP with IPCA, IPCA exhibited even more pronounced overfitting, especially with Random Forest, which performed well in training but failed to generalize. UMAP provided more stable test accuracy for KNN and XGBoost, with XGBoost again emerging as the most effective model. In terms of precision, Random Forest showed high training values with IPCA but experienced significant drops in test precision, suggesting a high false-positive rate. With UMAP, KNN and XGBoost maintained consistency across training and test precision, favoring XGBoost's reduced false positives, an advantage for real-world applications. Recall results were similar, with Random Forest suffering notable drops in IPCA but performing slightly better with UMAP, although both KNN and XGBoost struggled with maintaining high true positive rates on test data. F1 scores followed these trends, showing Random Forest's pronounced drop from training to testing, while KNN and XGBoost provided more consistent, albeit lower, F1 values with UMAP. The results suggest that UMAP, in contrast to IPCA, facilitates a better balance between precision and recall, enhancing model stability in practical applications. The distinction between UMAP and IPCA lies in their impact on model generalization. IPCA's overfitting issues, particularly with Random Forest, highlighted its limitations in capturing the underlying age-invariant structure, while UMAP's performance profile with XGBoost suggested that UMAP may preserve the structure more effectively for age-invariant features. From a computational perspective, IPCA offers efficiency and suitability for larger datasets due to its incremental updates, while UMAP excels at visualizing and preserving global structure, which aids in understanding complex distributions essential for face recognition tasks. Table 5 shows the comparison of the various classifiers on the MORPH II dataset after applying UMAP.

4.4 | Comparison With Existing Studies

Table 6 provides a comparative critique of the performance of several methods using the FGNet dataset, which is well acknowledged in the field of facial age estimation research. The analysis includes five studies, one of which is our own.

The accuracy attainment of 95.02% by Dhamija & Dubey [16] is indicative of a high level of precision. Likewise, Huang, Zhang, & Shan [20] documented a nearly identical outcome of 95.00%, with a little discrepancy probably attributable to minor errors in the methodology. Amal et al. (2023) attained a score of 93.00%, which, although lower, yet indicates a commendable level of quality. In the study, An & Wu, [12] found a notably reduced accuracy of 69.59%, indicating that their approach was less robust. With an accuracy of 96.00%, our strategy surpasses all others, demonstrating a significant enhancement in precision and solidifying its position as the top methodology for estimating facial age on the FGNet dataset. Next, we compare the performance of our model with existing studies that used the MORHP II dataset as shown in Table 7. Although there were a number of papers that used the MORPH II data for studies such as ours, we compared it to only one paper since that was the only full-text article in the domain that utilized accuracy as a metric.

From Table 7, our model underperformed when compared to the study of Abdulwahid [4], which indicated an accuracy of 86.0%. This suggests that the model is not generalized on the MORPH II data, suggesting that there are opportunities for further enhancement.

5 | Conclusion and Future Works

This study proposed a framework for age-invariant facial recognition to address challenges across age groups. The FGNet and Morph II datasets were used, and images were standardized through greyscale conversion, face detection using the Viola-Jones algorithm, and resizing. Feature extraction combined SIFT for invariant features, LBP for texture patterns, and ViT for global representations. Feature fusion followed by dimensionality reduction using Kernel PCA, Incremental PCA, and UMAP optimized the feature space. Random Forest, KNN, and XGBoost classifiers achieved high accuracy, demonstrating the framework's practical scalability.

Future work will focus on expanding the dataset with synthetic samples generated using GANs, improving preprocessing through posture correction and illumination normalization, and integrating CNNs or transformers with handcrafted methods to enhance feature representation. Further optimization of classifiers, exploration of advanced dimensionality reduction techniques, and GPU acceleration will also be considered to improve real-time performance.

TABLE 6 | Comparison with existing studies that used FGNET data.

Study	Accuracy
Dhamija & Dubey, [16]	95.0%
Huang, Zhang, & Shan [20]	95.0%
Aliα, Dass, Malik, Asirvadam, & Aziz [40]	93.0%
An & Wu [12]	69.5%
Ours	**96.0%**

TABLE 7 | Comparison with existing studies that used MORPH II data.

Study	Accuracy
Abdulwahid [4]	86.0%
Ours	**79.0%**

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

1. V. Vivek, K. R. Kumar, P. L. Kumari, S. Ahamad, R. Bansal, and A. Gupta, "Application of Biometric System to Enhance the Security in Virtual World," in *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (IEEE, 2022), 719–723.

2. M. HernandezdeMenendez, R. MoralesMenendez, C. A. Escobar, and J. Arinez, "Biometric Applications in Education," *International Journal on Interactive Design and Manufacturing (IJIDeM)* 15 (2021): 1–16.

3. A. K. Jain, A. Ross, and a. S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE Transactions on Circuits and Systems for Video Technology* 14, no. 1 (2004): 1–7.

4. A. A. Abdulwahid, "Classification of Ethnicity Using Efficient CNN Models on MORPH and FERET Datasets Based on Face Biometrics," *Applied Sciences* 13, no. 7288 (2023): 1–29.

5. L. Li, X. Mu, and S. L. Peng, "A Review of Face Recognition Technology," *IEEE Access* 8 (2020): 1–11.

6. S. Zhao, J. Li, and J. Wang, "Disentangled Representation Learning and Residual GAN for Age-Invariant Face Verification," *Pattern Recognition* 100 (2020): 1–10.

7. S. Sharma, M. Bhatt, and P. Sharma, "Face Recognition System Using Machine Learning Algorithm," in *2020 5th International Conference on Communication and Electronics Systems (ICCES)* (IEEE Xplore, 2020), 1162–1168.

8. J. Zhao, S. Yan, and J. Feng, "Towards Age-Invariant Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, no. 1 (2022): 474–487.

9. M. Sajid, T. Shafique, S. Manzoor, et al., "Demographic-Assisted Age-Invariant Face Recognition and Retrieval," *Symmetry* 10, no. 5 (2018): 148.

10. P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face Recognition by Humans: Nineteen Results all Computer Vision Researchers Should Know About," *Proceedings of the IEEE* 94, no. 11 (2006): 1948–1962, https://doi.org/10.1109/JPROC.2006.884093.

11. M. Sheng, Z. Ma, H. Jia, Q. Mao, and M. Dong, "Face Aging With Conditional Generative Adversarial Network Guided by Ranking-CNN," in

2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (IEEE Xplore, 2020), 314–319.

12. W. An and G. Wu, "Hybrid Spatial-Channel Attention Mechanism for Cross-Age Face Recognition," *Electronics* 60 (2024): 1–16.

13. R. K. Tripathi and A. S. Jalal, "A Local Descriptor for Age Invariant Face Recognition Under Uncontrolled Environment," in *2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC)* (IEEE Xplore, 2021), 513–517.

14. D. Sungatullina, J. Lu, G. Wang, and P. Moulin, "Multiview Discriminative Learning for Age-Invariant Face Recognition," in *International Conference on Automatic Face and Gesture Recognition* (IEEE, 2013), 1–6.

15. M. S. Shakeel and K.-M. Lam, "Deep-Feature Encoding-Based Discriminative Model for Age-Invariant Face Recognition," *Pattern Recognition* 93 (2019): 442–457.

16. A. Dhamija and R. Dubey, "A Novel Active Shape Model-Based DeepNeural Network for Age Invariance Face Recognition," *Journal of Visual Communication and Image Representation* 82 (2022): 103393.

17. M. F. Hirzi, S. Efendi, and R. W. Sembiring, "Literature Study of Face Recognition Using the Viola-Jones Algorithm," in *2021 International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)* (IEEE, 2021), 1–6.

18. X. Hou, X. Qiu, S. Wan, and T. Tang, "A Robust Feature-Processing Method for Age-Invariant Face Recognition," in *2019 3rd International Conference on Machine Vision and Information Technology (CMVIT 2019)* (Guangzhou: IOP Science, 2019), 1–8.

19. A. A. Moustafa, A. Elnakib, and N. F. Areed, "Age-Invariant Face Recognition Based on Deep Features Analysis," *Signal, Image and Video Processing* 14 (2020): 1027–1034.

20. Z. Huang, J. Zhang, and H. Shan, "When Age-Invariant Face Recognition Meets Face Age Synthesis: A Multi-Task Learning Framework and a New Benchmark," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, no. 6 (2023): 7917–7932.

21. S. Riaz, Z. Ali, U. Park, J. Choi, I. Masi, and P. Natarajan, "Age-Invariant Face Recognition Using Gender Specific 3D Aging Modeling," *Multimedia Tools and Applications* 78 (2018): 25163–25183.

22. J. Zhang, H. Wana, Y. Gao, and L. Tao, "An Age Invariant Face Recognition Algorithm Based on Decorrelated Feature Decomposition," in *2021 4th International Conference on Information Communication and Signal Processing, ICICSP 2021* (IEEE Xplore, 2021), 409–414.

23. X. Ren, J. Wang, and S. Li, "MAM: Multiple Attention Mechanism Neural Networks for Cross-Age Face Recognition," *Wireless Communications and Mobile Computing* 22 (2022): 11.

24. K. Okokpujie, S. John, C. Ndujiuba, J. A. Badejo, and E. N. Osaghae, "An Improved Age Invariant Face Recognition Using Data Augmentation," *Bulletin of Electrical Engineering and Informatics* 10, no. 1 (2021): 179–191.

25. Y. Wu, L. Du, and H. Hu, "Parallel Multi-Path Age Distinguish Network for Cross-Age Face Recognition," *IEEE Transactions on Circuits and Systems for Video Technology* 31, no. 9 (2021): 3482–3492.

26. S. Agrawal, S. Kumar, S. Kumar, and A. Thomas, "A Novel Robust Feature Extraction With GSO-Optimized Extreme Learning for Age-Invariant Face Recognition," *Imaging Science Journal* 67, no. 6 (2019): 319–329.

27. S. Li and H. J. Lee, "Effective Attention-Based Feature Decomposition for Cross-Age Face Recognition," *Applied Sciences* 12, no. 10 (2022): 4816.

28. H. Wang, D. Gong, Z. Li, and W. Liu, "Decorrelated Adversarial Learning for Age-Invariant Face Recognition," *Computer Vision and Pattern Recognition* (2019): 1–10.

29. Y. Li, W. Liu, X. Li, Q. Huang, and X. Li, "GA-SIFT: A New Scale Invariant Feature Transform for Multispectral Image Using Geometric Algebra," *Information Sciences* 281 (2014): 559–572.

30. P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision* 57 (2004): 137–154.

31. W. Cheung and G. Hamarneh, "N-SIFT: N-Dimensional Scale Invariant Feature Transform," *IEEE Transactions on Image Processing* 18, no. 9 (2009): 2012–2021.

32. M. Pietikäinen, "Local Binary Patterns," *Scholarpedia* 5, no. 3 (2010): 9775.

33. A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., *An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale* (ICLR, 2021), 1–22.

34. S.-C. Huang, A. Pareek, S. Seyyedi, I. Banerjee, and M. P. Lungren, "Fusion of Medical Imaging and Electronic Health Records Using Deep Learning: A Systematic Review and Implementation Guidelines," *npj Digital Medicine* 3, no. 136 (2020): 1–9.

35. A. Ayantayo, A. Kaur, A. Kour, et al., "Network Intrusion Detection Using Feature Fusion With Deep Learning," *Journal of Big Data* 10, no. 167 (2023): 1–24.

36. S. R. Stahlschmidt, B. Ulfenborg, and J. Synnergren, "Multimodal Deep Learning for Biomedical Data Fusion: A Review," *Briefings in Bioinformatics* 23, no. 2 (2022): 1–15.

37. M. W. Dorrity, L. M. Saunders, C. Queitsch, S. Fields, and C. Trapnell, "Dimensionality Reduction by UMAP to Visualize Physical and Genetic Interactions," *Nature Communications* 11 (2020): 1537.

38. Y. Yang, H. Sun, Y. Zhang, et al., "Dimensionality Reduction by UMAP Reinforces Sample Heterogeneity Analysis in Bulk Transcriptomic Data," *Cell Reports* 36, no. 4 (2021): 109442.

39. S. A. Atimbire, J. K. Appati, and E. Owusu, "Empirical Exploration of Whale Optimisation Algorithm for Heart Disease Prediction," *Scientific Reports* 4530 (2024): 1–22.

40. A. Aliα, S. C. Dass, A. Malik, V. Asirvadam, and A. Aziz, "Age Invariant Face Recognition Using Feature Level Fused Morphometry of Lip-Nose and Periocular Region Features," *London Journal of Research in Computer Science & Technology* 23, no. 1 (2023): 1–28.