Journal of Informatics and Web Engineering

Vol. 3 No. 3 (October 2024) eISSN: 2821-370X

Sibling Discrimination Using Linear Fusion on Deep Learning Face Recognition Models

Rita Goel[1*], Maida Alamgir[2], Haroon Wahab[3], Maria Alamgir[4] , Hassan Ugail[5], Irfan Mehmood[6], Amit Sinha[7]

[1,2,3,5,6] University of Bradford, Richmond Rd, Bradford BD7 1DP, United Kingdom.

[4]Manchester Metropolitan University, Ormond, Lower Ormond St, Manchester M15 6BX, United Kingdom.

[7]ABES Engineering College, 19th KM Stone, NH-09, Ghaziabad, Uttar Pradesh 201009, India.

**corresponding author: (r.goel1@bradford.ac.uk; ORCiD: 0009-0002-5290-686X)*

Abstract - Facial recognition technology has revolutionised human identification, providing a non-invasive alternative to traditional biometric methods like signatures and voice recognition. The integration of deep learning has significantly enhanced the accuracy and adaptability of these systems, now widely used in criminal identification, access control, and security. Initial research focused on recognising full-frontal facial features, but recent advancements have tackled the challenge of identifying partially visible faces, a scenario that often reduces recognition accuracy. This study aims to identify siblings based on facial features, particularly in cases where only partial features like eyes, nose, or mouth are visible. Utilising advanced deep learning models such as VGG19, VGG16, VGGFace, and FaceNet, the research introduces a framework to differentiate between sibling images effectively. To boost discrimination accuracy, the framework employs a linear fusion technique that merges insights from all the models used. The methodology involves preprocessing image pairs, extracting embeddings with pre-trained models, and integrating information through linear fusion. Evaluation metrics, including confusion matrix analysis, assess the framework's robustness and precision. Custom datasets of cropped sibling facial areas form the experimental basis, testing the models under various conditions like different facial poses and cropped regions. Model selection emphasises accuracy and extensive training on large datasets to ensure reliable performance in distinguishing subtle facial differences. Experimental results show that combining multiple models' outputs using linear fusion improves the accuracy and realism of sibling discrimination based on facial features. Findings indicate a minimum accuracy of 96% across different facial regions. Although this is slightly lower than the accuracy achieved by a single model like VGG16 with full-frontal poses, the fusion approach provides a more realistic outcome by incorporating insights from all four models. This underscores the potential of advanced deep learning techniques in enhancing facial recognition systems for practical applications.

Keywords—Face Recognition, Sibling Identification, Linear Fusion, Deep Learning, Pre-trained Models, Confusion Matrix

Received: 27 June 2024; Accepted: 29 August 2024; Published: 16 October 2024

1. INTRODUCTION

Facial recognition has evolved into a highly effective and logical method for human identification. Unlike other techniques such as signatures, hand maps, voice identification, and speech recognition, facial recognition is favoured for its contactless nature. This technology's modern applications leverage Deep Learning, making facial recognition more sophisticated and accurate. It is extensively used in various real-world scenarios, including criminal identification, access control, and security systems, notably in biometric security [1].

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2024.3.3.14

Published by MMU Press. URL:
https://journals.mmupress.com/jiwe

The advancement of face recognition technology, particularly through deep learning, has significantly improved its capabilities. By utilising large datasets, deep learning models can learn intricate facial features, enhancing the accuracy and reliability of recognition systems. In recent decades, researchers have mainly concentrated on recognising full-frontal faces or significant portions of the face [2]. However, the more intriguing and nascent area of research involves occluded faces, where only parts of the face, such as the eyes, nose, or mouth, are visible. This partial visibility often leads to lower performance and accuracy in facial recognition systems.

Recognising faces that look very similar, such as those of siblings [3], doppelgangers, and stranger twins, poses an additional challenge, especially when only partial facial features are available. Addressing this challenge is crucial for advancing face recognition technologies. One effective approach to improve recognition accuracy and robustness is integrating multiple sources of information. Linear fusion methods have gained popularity for combining diverse feature representations [4]. In the context of deep learning-based face recognition models, linear fusion involves integrating feature representations extracted from different modalities or sources through linear operations, such as weighted averaging.

This study aims to develop a framework for distinguishing between images of siblings based on various facial features. The process begins with normalising the input image pairs. Then, the embeddings of these images are extracted using state-of-the-art face recognition models. These images are classified as the same or different based on the similarity measures used. Subsequently, a weight is applied to the results of all the face recognition models used, and linear fusion is applied to aggregate opinions from all the models. Finally, the model's accuracy is assessed through confusion matrix analysis, yielding more reliable and realistic outcomes.

In this research, four state-of-the-art models— VGG19, VGG16, VGGFace, and FaceNet—are employed to differentiate between image pairs of siblings across five variables: full-frontal face orientation, and cropped regions of the eyes, nose, mouth, and forehead. While many face recognition models exist, such as Deepface and the DeepID series, the selected models are considered the best candidates due to their proven accuracy and extensive training on large datasets. In this study, custom datasets were generated by cropping the facial regions of siblings, creating challenging cases to assess the models' accuracy and robustness.

The primary objective is to develop a reliable framework for accurately recognising or differentiating between siblings or look-alikes in scenarios such as surveillance or criminal investigations where other biometric methods might be unavailable. The experimental results show that the proposed approach is highly effective for sibling differentiation, indicating its potential for broad application in real-world situations, especially when only partial facial features are accessible. This paper is organised into five sections. Section 2 reviews related work on sibling face recognition and kinship verification. Section 3 details the framework developed for this study. Section 4 outlines the datasets, evaluation matrices, and experiments, describing how performance is analysed. Section 5 depicts the results achieved in this research. Finally, Section 6 presents the concluding remarks.

2. LITERATURE REVIEW

2.1 Related Work

Despite the scattered and inconsistent literature on siblings and kinship verification, numerous innovative ideas have emerged to address this issue. Researchers have explored various approaches, leveraging advancements in machine learning and deep learning technologies. Significant progress has been achieved in developing effective solutions for kinship verification by combining various data sources and utilising advanced algorithms. These innovations are crucial for applications in security, forensic investigations, and social networking, where accurately identifying familial relationships provides valuable insights. Continued exploration in this field promises further advancements, leading to more robust and efficient kinship verification systems [5],[6].

2.1.1 Kinship Verification Techniques

Addressing the challenge of kinship verification, Fang et al. [7] introduced an innovative lightweight facial feature extraction and selection algorithm designed to enhance kinship verification processes. This algorithm works by classifying image pairs as either "related" or "unrelated" based on frontal facial images of 150 well-known public figures and celebrities. Using the K-Nearest Neighbors (KNN) classification method, the algorithm analyses features such as eye colour and skin tone to determine familial relationships. Remarkably, this approach achieved results superior to human performance in accurately verifying kinship. This advancement highlights the potential of machine learning techniques to outperform traditional methods in recognising subtle familial traits, paving the way for more reliable and efficient kinship verification systems in various applications.

Lamba et al. [8] studied face recognition for look-alikes in three stages, beginning with an analysis of human recognition capabilities to understand how well people can identify look-alikes. This was followed by a comparison with ten existing face recognition algorithms to evaluate their effectiveness in distinguishing between similar-looking individuals. Finally, the researchers proposed an innovative approach to enhance face verification accuracy. Their method involved extracting features from overlapping facial regions and integrating this information at the match score level. This approach yielded slightly better results than existing algorithms, demonstrating the potential to improve the efficiency of both human and algorithmic recognition in accurately identifying look-alikes. This research underscores the effectiveness of combining human intuition with advanced computational techniques to refine face recognition technologies.

Similar to Lamba et al., Xia et al. [9] utilised transfer learning to enhance kinship verification accuracy by creating an extensive new database featuring facial images of children and parents of various ages. Their advanced transfer learning method effectively addressed distribution gaps between the facial features of children and older parents. This innovative approach enhanced kinship verification by accurately aligning and transferring knowledge across generational differences in facial characteristics.

Shadrikov [10] introduced a novel approach for automated kinship recognition by utilising RetinaFace for face detection and ArcFace for face verification. This method set a new standard in the field by achieving outstanding results in the Families in the Wild Data Challenge. By leveraging these advanced algorithms, their system demonstrated remarkable accuracy and efficiency in identifying familial relationships in uncontrolled environments, highlighting its potential for real-world applications in kinship recognition.

Vieira et al. [11] created a comprehensive dataset for their study, consisting of four sub-datasets containing high-quality images captured under controlled conditions. These images were specifically curated to identify siblings in paired images. The study focused on examining how various facial attributes contribute to sibling classification, comparing the effectiveness of automated classification systems with human abilities. By integrating features from diverse sources, the researchers achieved significantly higher classification accuracies than human performance, demonstrating the potential of advanced algorithms to outperform human recognition in sibling identification tasks.

Matthews and colleagues [12] utilised the database created by Vieira et al. to create a comprehensive framework for identifying siblings through image feature matching. Their approach involved calculating a similarity metric for input image pairs based on the quantity of matching features in the combined images. This innovative technique significantly reduced error rates and processing times in sibling prediction, highlighting the effectiveness of their method in accurately detecting sibling relationships from images.

Recent research on kinship verification by Yan and Song [13] introduced a deep relational network that utilises multi-scale facial image data to enhance verification accuracy. Their method incorporates two convolutional neural networks (CNNs) with shared parameters to extract features at various scales, significantly improving the system's ability to discern kinship from image pairs. The findings highlighted the network's effectiveness in accurately identifying familial relationships, underscoring its potential for advanced kinship recognition.

The Dynamic Feature Matching (DFM) method has been introduced as a solution for partial face recognition by He et al. [14]. This technique integrates Fully Convolutional Networks (FCN) with sparse representations to improve the extraction of features. The FCN plays a key role by producing a feature map that identifies more distinct features within images. A crucial aspect of their approach is the use of the VGG-Face model, which supplies features that the FCN then processes. Their method demonstrated impressive classification accuracies, achieving 97% on the CASIA-NIR-Distance dataset and 74% on the YouTube dataset.

Extensive research has been conducted on partial face recognition by Li et al. [15], focusing on challenges such as changes in lighting, disguises, and occlusions in frontal-view images. The researchers proposed a novel method that extracts a dynamic subspace from images to pinpoint each individual's unique facial features. These distinctive features were then utilised to create a recognition system using the k-nearest neighbour (K-NN) algorithm. Following the approach of earlier studies, their experiments employed databases like ORL and Extended Yale B. The findings were promising, showing improved recognition rates by effectively using partial facial information.

More recently, Lahasan et al. [16] proposed a framework named Optimised Symmetric Partial Facegraph (OSPF) for face recognition. This framework was tested under various challenging conditions, such as face occlusion, differing facial expressions, and variations in lighting. The experimental results demonstrated that using partial facial data can significantly improve recognition rates.

Elmahmudi et al. [17] have recently introduced a method for face recognition that focuses on using partial or incomplete facial data. In their approach, features are extracted from fragmented facial images and then compared to a database of full frontal images. Their results indicated that although recognition rates were lower when using isolated facial parts, combining these parts into a complete probe significantly improved recognition accuracy.

2.1.2 Fusion Approaches in Face Recognition

Many researchers have used fusion approaches for face recognition [18],[19] as they offer several advantages in face recognition. It allows leveraging complementary information from multiple modalities, improving accuracy, robustness, and performance in challenging conditions. It also provides flexibility by combining different modalities based on application requirements [20].

Basiri et al. in 2019 [21] presented an innovative approach that integrates four deep learning models with a traditional supervised machine learning model for sentiment analysis of COVID-19-related tweets from different countries. Additionally, they examined Google searches related to COVID-19 to gain deeper insights into shifting sentiment patterns. By applying both classical and deep learning models, they aimed to enhance confidence in sentiment classification. Their results demonstrated that the proposed fusion models achieved promising performance in classifying sentiments and could be effectively utilised to analyse the collected coronavirus-related data.

Iqbal et al. in 2021 [22] presented deep learning model fusion with novel classifiers to classify brain tumour segmentation. They have trained both networks in different ways to make it inquisitive about using both models together and find how the group of these models can impact the results compared to their applications. They used the benchmark dataset BRATS2015. Their result comparison shows the improved segmentation accuracy by combining the models over individual model performance.

Gao et al. [23] presented a survey on deep learning for multimodel data fusion to provide fundamentals of multimodal deep learning fusion method and to motivate new multimodal data fusion techniques of deep learning. They summarised the architecture of the widely used deep learning models to explain multimodel deep learning and then encapsulated the state-of-the-art multimodel data fusion and deep learning models. Their review suggests that the state-of-the-art models have made some progress, but they are still in the preliminary stage and ineffective in some complex scenarios.

In 2022, Zheng et al. [24] conducted a state-of-the-art survey on the fusion of deep learning models and fuzzy systems to show that fuzzy systems can help improve the accuracy of deep learning models. They first described the results of the relevant publications and conventional deep learning algorithms, then constructed the graphic form of the fusion of fuzzy systems into deep learning models. Their review indicates that this field is an emerging area of research that is gaining considerable attention. Fuzzy systems, in particular, have a notable impact on deep learning models, especially in the areas of classification, prediction, natural language processing, and automation. This fusion of technologies is being applied across various fields, including computer science, natural language processing, medical systems, intelligent energy management systems, and the manufacturing industry.

Sun and Lv [25] developed a hierarchical feature fusion method aimed at improving face recognition accuracy. This approach utilises supervisory information to learn both shallow and deep facial features, which are then integrated to boost recognition performance under varying lighting conditions and occlusions. The researchers conducted their experiments using the Labeled Faces in the Wild (LFW) and AR face databases, applying feature fusion at each layer of their end-to-end network. Their findings indicated that this method significantly enhanced face recognition performance, demonstrating the effectiveness of fusing multi-level features to create a more robust recognition system.

2.2 Summary

Based on the literature above, it is evident that integrating fusion techniques in face recognition can enhance the reliability and accuracy of results. Specifically, combining multiple sources of information can improve effectiveness, especially when facial images are incomplete, such as containing only partial features like the nose or eyes. In such cases, individual methods may perform poorly, resulting in lower recognition rates. This limitation is largely due to the structure of the methods and their training on specific types of data. Furthermore, existing face databases typically do not provide isolated facial parts, which hinders the ability of models to learn distinct features for different facial regions.

3. LINEAR FUSION FOR SIBLING DISCRIMINATION

Recognising siblings with high facial similarity can be challenging for face recognition models, which might incorrectly identify them as the same person. The difficulty increases when only partial facial features, such as the nose or eyes, are available. This work introduces a novel approach to distinguishing between siblings by employing a linear fusion of four cutting-edge face recognition models, each analysing different facial features of siblings. Comprehensive experiments are conducted to determine the model's accuracy, which is evaluated based on various similarity measures. Interestingly, in complex scenarios, the most accurate model might yield poor results, while a less accurate one might perform better. To address this, linear fusion integrates the results from all four models, incorporating their collective insights to deliver more robust and realistic outcomes. The proposed architecture for sibling discrimination using linear fusion is illustrated in Figure 1.

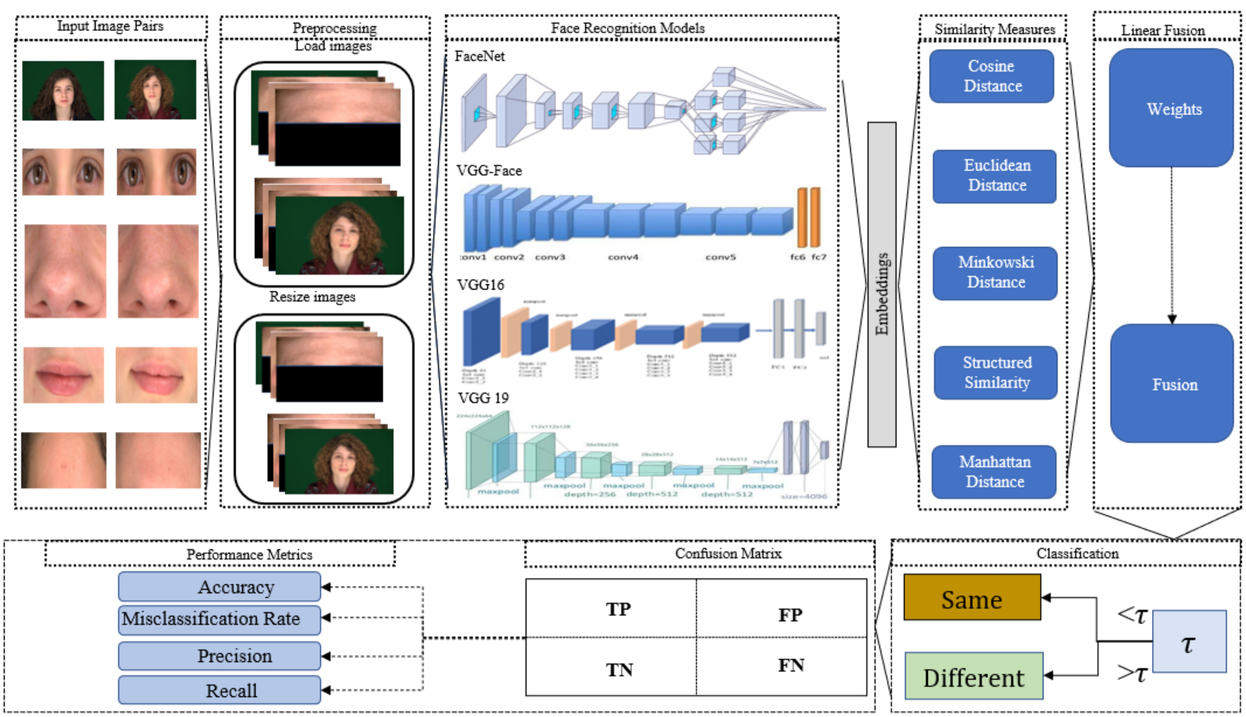

Figure 1. Architectural Overview Of The Framework Used To Differentiate Between Siblings Employing State-of-the-art Face Recognition Models and Linear Fusion Among Them

3.1 Face Recognition Models

In this work, four state-of-the-art face recognition models— VGG19, VGG16, VGGFace, and FaceNet—are utilised to discriminate between siblings based on various facial features. FaceNet, developed by Google researchers in 2015, achieved top results on multiple face recognition benchmark datasets. FaceNet is available as a pre-trained, third-party open-source implementation [26],[27]. It extracts high-quality facial features, referred to as face embeddings, which are utilised to train a face recognition system. The model learns to map face images directly into a condensed Euclidean space, where distances indicate similarity between faces — small distances indicate the same person, while large distances indicate different people.

FaceNet employs a deep convolutional network that is optimised directly for generating embeddings, avoiding the intermediate bottleneck layers used in earlier deep learning methods. This approach results in a highly robust and effective face recognition system, with its general face embeddings applicable to a variety of tasks. As explained by Florian Schroff and his team at Google [26], FaceNet uses a triplet loss function to ensure that vectors representing the same identity are closely aligned while vectors for different identities remain distinctly separated. The network predicts a 128-dimensional vector, or embedding, for each image. By focusing on directly creating image embeddings rather than extracting them from an intermediate layer, FaceNet has achieved significant effectiveness. Trained on a large dataset, FaceNet is widely recognised for its accuracy, which has been evaluated in this study.

Another widely used model in face recognition is VGGFace, developed by the Visual Geometry Group (VGG) at the University of Oxford [28]. VGGFace models are designed for face recognition and have been demonstrated on benchmark computer vision datasets. For face verification, VGGFace calculates embeddings of a face image and compares them with embeddings of other faces using similarity measures such as Cosine distance or Euclidean distance. It determines a match by comparing these distances to a predefined threshold. The model comprises 11 layers, including eight convolutional layers and then three fully connected layers, and was trained on a large dataset containing 2.6 million face images. With 38 layers and an input image size of 224×224, VGGFace is known for its ease of training and popularity in visual computing. This work aims to evaluate VGGFace's accuracy in challenging scenarios, such as distinguishing between siblings or similar-looking faces.

Another prominent CNN model is VGG16, which achieves around 92% accuracy on the ImageNet dataset, which includes over 14 million images across 1000 categories. Created by Simonyan and Zisserman at the University of Oxford [29], VGG16 was a leading model in the ILSVRC-2014 competition [30]. This model enhances accuracy by using multiple smaller 3×3 filters instead of the larger kernels employed in AlexNet [31]. Like VGGFace, VGG16 takes a 224×224 RGB image as input for its first convolutional layer. Featuring 16 convolutional layers, VGG16 is known for its consistent architecture. Although it was trained on an Nvidia Titan GPU over 2-3 weeks, which makes the training process relatively slow, this approach results in exceptionally high accuracy. However, it also necessitates a high-end system configuration and leads to a larger model size.

Another model used in this work is VGG19, an extension of the VGG model [32], which comprises 19 layers: 16 convolutional layers, three fully connected layers, five max-pooling layers, and one softmax layer. This pre-trained network excels in classifying images into 1000 object categories, spanning items like keyboards, mice, pencils, and various animals. With an input image size of 224-by-224, VGG19 uses small (3x3) convolution filters, which significantly enhance feature representation compared to earlier configurations. Similar to VGG16, VGG19 is known for its slow training process but delivers exceptionally high accuracy. However, this entails a trade-off of increased model size and demanding computational requirements. VGG19 is selected for these experiments due to its reputation for high accuracy, particularly in tasks involving the classification of partial faces.

These four models are considered to be the most efficient in the field of face recognition. Using pre-trained models for face recognition is advantageous for various reasons. Pre-trained models require less training and effort to build the frameworks. Another advantage is that pre-trained models are significantly more accurate as they are trained on extensive data and can classify images in various objects.

In this work, these models compare images of siblings with a high degree of similarity in their facial appearances across siblings. The models have been used in this framework to conduct various experiments. To begin with, various similarity measures, such as Cosine similarity and Euclidean distance, are employed to assess the accuracy of the pre-trained models for distinguishing between siblings. The accuracy is evaluated by calculating the embeddings of the images of different face parts of the siblings.

3.2. Dataset

The experiments in this study utilise SiblingsDB, curated by Vieira et al. [11], which includes diverse datasets depicting individuals in sibling relationships. SiblingsDB consists of two main databases, with HQfaces being the primary focus. HQfaces comprises 92 pairs of high-quality images featuring 184 individuals. Among these, 79 pairs feature profile images, and 56 pairs include both smiling frontal and profile pictures. The images were professionally taken under controlled conditions with uniform backgrounds and lighting, boasting a resolution of 4256×2832 pixels.

The subjects in the images are voluntary students and employees of Politecnico di Torino, along with their siblings, spanning an age range from 13 to 50 years, with an average age of 23.1 years. The age differences between siblings vary, with a maximum of 30 years and an average of 4.6 years. The dataset comprises entirely Caucasians, with 57% being male, and subjects were instructed not to wear makeup. The dataset is organised into three subsets: HQf (expressionless frontal images), HQfp (expressionless frontal and profile images), and HQfps (expressionless and smiling frontal and profile images).

This work involved a comprehensive dataset of 204 images, comprising 102 pairs of sibling images with full-frontal poses. To enhance the diversity and robustness of our dataset, we included sibling pairs with varied ages, genders, and ethnic backgrounds. Additionally, we expanded the dataset by creating subsets of images cropped to focus on specific facial regions such as noses, eyes, mouths, and foreheads. The criteria for cropping involved precise measurements to ensure consistency:

- **Eyes**: Cropped to include the eye region from brow to mid-cheek.
- **Noses**: Focused on the area from the bridge to the tip.
- **Mouths**: Included the region from the upper lip to the chin.
- **Foreheads**: Cropped from the hairline to just above the eyebrows.

These cropped subsets contained images of both siblings and non-siblings, further enriching the dataset and providing a more challenging and comprehensive basis for sibling discrimination analysis. To ensure a comprehensive analysis, we randomly selected non-sibling image pairs from the original dataset, adding variety to our sample. Figures 2 and 3 showcase sample images from the prepared datasets, illustrating the types of images used in our research. This approach allowed us to visually represent the data and highlight the distinctions between siblings and non-siblings based on specific facial features.

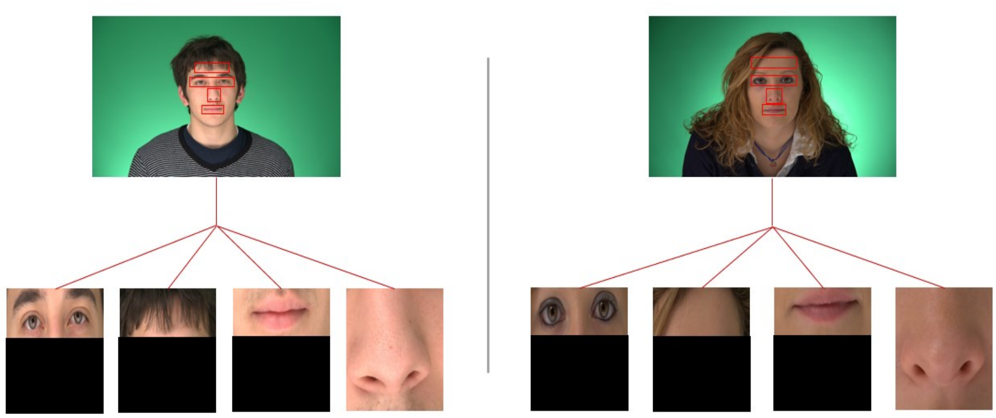

Figure 2. Sample Image Pair For Siblings' Full-Frontal Face, Foreheads, Eyes, Nose, and Mouth. The Images Are Taken From The HQf Dataset Of The Sibling Database, And Face Parts Are Cropped From The Full-frontal Image

3.3 Sibling Discrimination Using Linear Fusion

Linear fusion is helpful in some complex circumstances where the best-performing model has classified the siblings as the same person, and the non-best face recognition model has classified them differently. Linear fusion is a technique used to combine outputs from multiple models or feature extractors into a single, unified representation. This method involves merging these outputs by applying weighted contributions from each source to create an integrated vector. The process begins by obtaining the results from different models or components, which are then scaled according to their assigned weights. These weighted outputs are summed to form the final fused representation. By balancing the contributions from each source, linear fusion leverages the strengths and diverse perspectives of the individual models, thereby enhancing the overall performance and robustness of the system. This approach is particularly valuable for integrating complementary information to improve accuracy and effectiveness.

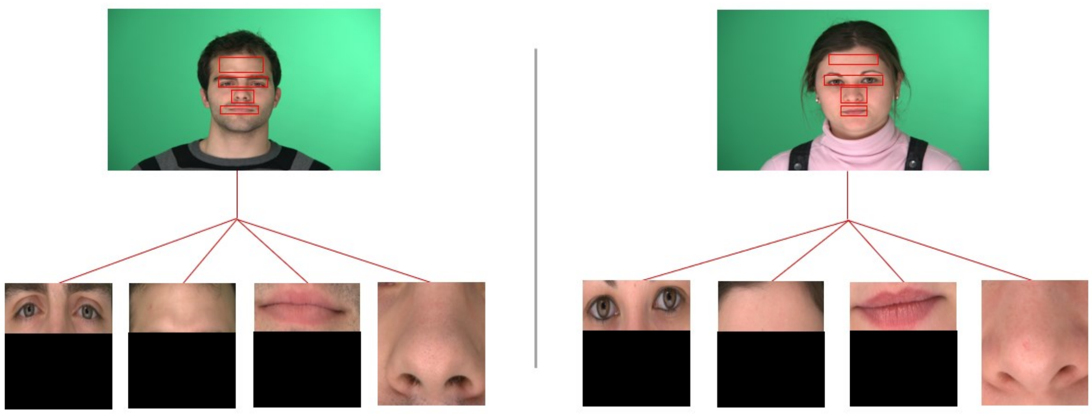

Figure 3. Sample Image Pair For Non-Siblings' Full-Frontal Pose, Eyes, Foreheads, Mouth, And Nose. The Images Are Randomly Taken From The HQf Dataset From The Sibling Database, And Face Parts Are Cropped From The Full-frontal Image

The siblings' image pairs of full-frontal face pose, cropped eyes, cropped nose, cropped mouth, and cropped forehead are taken as input image dataset P, containing p image pairs of different face variates. A threshold τ is defined for each face variate for each state-of-the-art face recognition model. Each image in pair i is transformed into an array using the TensorFlow preprocessing library according to the input requirements of the specific model used. The preprocessing steps include:

- **Resizing:** The image is resized to match the input shape required by the model.
- **RGB Conversion:** The resized images are converted to RGB format, as the model expects a three-channel colour image.
- **NumPy Conversion:** After converting to RGB, the images are converted to NumPy arrays to represent them as vectors.
- **Dimension Expansion:** The image dimensions are expanded to match the input shape expected by the model.

Once the images are preprocessed, their embeddings are obtained using the respective model for comparison. The distance between these embeddings is then calculated using the different similarity measures used in the framework. For each face variate used in the framework, a mean distance is calculated by taking an average of the results of all similarity measures for each state-of-the-art face recognition model.

Once a mean distance is calculated, a weight is applied to each model for each face variate using Equation (1).

$$w^{frm} = \frac{Pm^{frm}}{\sum Pm^{frm}} \tag{1}$$

Where, w^{frm} is the weight for the respective model,

frm is the respective Face Recognition Model, and,

Pm is the average distance calculated for each face recognition model on the respective variate.

After applying weights to the models, a linear fusion is applied to these weights to obtain a robust and realistic distance for each face variate, which has opinions from all state-of-the-art face recognition models used in this work. The linear fusion is applied as per Equation (2).

$$FPm^{fv} = \sum_{i=1}^{i=4} w^i * Pm^i \tag{2}$$

Where, FPm^{fv} Is the fused distance for the corresponding face variate,

w^i is the weight assigned to each model,

Pm^i is the average distance for each model on respective face variate, and

i is the face recognition model used.

The distances calculated are then classified as the same or different by comparing the distance with the threshold defined for the respective face variate. The same experiments are conducted on the face variates of the non-siblings to evaluate the various performance metrics for the framework. Algorithm 1 shows the process for differentiating between siblings using Linear Fusion.

Algorithm 1: Differentiate between siblings using Linear Fusion

```
Input: Image Pairs of Siblings P with p pairs
Ꞇ = threshold array defined for each similarity measure
mod = respective model
for i=1 to p, do
        for each image in pair i, do
                Img1 → read first image;
                Img1 → resize first image;
                Img1 →preprocess image;
                E1 = extractEmbeddings(mod(Img1));
                Img2 → read second image;
                Img2 → resize second image;
                Img2 →preprocess image;
                E2 = extractEmbeddings(mod(Img2))
        end
end
cosine → cosine_distance(E1, E2)
euc →  euclidean_distance(E1, E2)
ssim → structured_similarity(E1, E2)
manh →  manhattan_distance(E1, E2)
mink →  minkowski_Distance(E1, E2)
mean_dist = Average(distance)
robust_distance → linear_fusion(mean_dist)
for j= 1 to Ꞇ.length, do
        if  robust_distance < τ[j]
                distance_result →"Same"
        else
                distance_result →"Different"
```

end

3.4. Similarity Measures

In this study, images are categorised as identical or different using various similarity metrics: Cosine similarity, Euclidean distance, Structured Similarity, Manhattan distance, and Minkowski distance [33]. These metrics are chosen for their widespread use in machine learning and face recognition applications. Each similarity measure is compared against a predefined threshold to determine whether the image pair belongs to the same class or different classes. This approach leverages established methods to effectively compare image vectors, ensuring robust classification based on similarity criteria.

Cosine similarity [34] quantifies the similarity between two non-zero vectors in an inner product space. It is computed as in Equation (3).

$$\cos(\theta) = \frac{A \cdot B}{||A|| \, ||B||} = \frac{\sum_{i=1}^{n} A_i \, B_i}{\sqrt{\sum_{i=1}^{n} Ai^2} \, \sqrt{\sum_{i=1}^{n} Bi^2}}, \tag{3}$$

where A_i and B_i denote the components of vectors A and B, respectively.

The Euclidean distance [35] between two vectors a and b in the Euclidean space represents the straight line between points a and b and can be computed as in Equation (4).

$$\mathrm{Euc}(a, b) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2} \quad . \tag{4}$$

In contrast, Structured Similarity [36] measures the similarity between two images by evaluating the relationship between two images, a and b, of a common size N×N \times N×N. It is computed as in Equation (5).

$$\mathrm{SSIM}(a, b) = \frac{(2\mu_a\mu_b + s_1)\,(2\sigma_{ab} + s_2)}{({\mu_a}^2 + {\mu_b}^2 + s_1)\,({\sigma_a}^2 + {\sigma_b}^2 + s_2)}, \tag{5}$$

with:
μ_a the average of a,
μ_b the average of b,
${\sigma_a}^2$ the variance of a,
${\sigma_b}^2$ the variance of b,
σ_{ab} he covariance of a and b,
$s_1 = (j_1K)^2$, $s_2 = (j_2K)^2$ two variables to stabilise the division when the denominator is small,
K is the pixel value's dynamic range,
$j_1 = 0.01$ and $j_2 = 0.03$ by default.

Manhattan distance [33] is the distance between two vectors measured along axes at right angles. In a plane with p_1 at (a_1, b_1) and p_2 at (a_2, b_2), it can be calculated as,

$$\text{Manhattan distance(D)} = (|a_1 - a_2| + |b_1 - b_2|)\,. \tag{6}$$

The Minkowski distance [33] is a metric in a vector space that serves as an extension that encompasses both Euclidean distance and Manhattan distance. The distance of order p where p is an integer between points, P and Q is measured and can be computed as follows:

$$\text{Minkowski Distance (D)} = \left(\sum_{i=1}^{n} |P_i - Q_i|^p\right)^{\frac{1}{p}} \quad . \tag{7}$$

The similarity between pairs of images, whether of siblings or non-siblings, is determined using the methods described earlier. Based on predefined thresholds for each similarity measure and each face recognition model used in the study, these images are classified as depicting the same person or different individuals. After classification, they are input into a confusion matrix to assess performance metrics such as accuracy, precision, and misclassification rate. This evaluation helps gauge the effectiveness of each face recognition model in correctly identifying familial relationships and distinguishing between similar faces.

3.5. Evaluation Metrics

When evaluating the efficiency of the models mentioned above on sibling datasets, accuracy, precision, and misclassification rates [37] are computed for each model across different datasets: full-frontal face, cropped eyes, nose, and forehead. This assessment converts the problem into a two-class classification task using non-sibling datasets from the experiments, generating confusion matrices for each model with each dataset used. These advanced evaluation metrics are employed in this study to assess the effectiveness of face recognition models thoroughly. Unlike simple accuracy calculations (correct predictions/total predictions), these metrics provide detailed insights necessary for a deeper understanding of the model's performance in recognising familial relationships accurately.

A confusion matrix [38] provides a summary of prediction results from a classification algorithm or face recognition model. It not only reveals the errors made by the classifier but also specifies the types of errors occurring. The rows of the matrix are labelled as "Positive" for event cases and "Negative" for non-event cases, with predictions categorised as "true" or "false".

In this study, experiments were conducted using image pairs of siblings and non-siblings, which are inherently distinct. Therefore, "True" cases are assessed using the sibling dataset, while "False" cases are evaluated using the non-sibling dataset:

- **True Positive (TP):** These occur when image pairs of siblings, correctly identified as different in reality, are also predicted as different by the models.
- **True Negative (TN):** These occur when models correctly identify image pairs of siblings as the same person, even though they are different.
- **False Positive (FP):** These cases arise when image pairs of non-siblings, correctly different in reality, are incorrectly classified as different persons by the models.
- **False Negative (FN):** These occur when models incorrectly classify image pairs of non-siblings as the same person.

This classification scheme provides a detailed assessment of the models' performance in distinguishing between siblings and non-siblings based on their facial features.

3.6. Novelty Claims

The key novelties of this study include:

- **Novel Framework for Partially Visible Faces:** This study presents a unique framework for sibling identification, specifically designed to accurately recognise siblings even when only partial facial features are visible.
- **Integration of Multiple Deep Learning Models:** By utilising advanced models such as VGG19, VGG16, VGGFace, and FaceNet and employing a linear fusion technique, the framework effectively combines insights from multiple models, enhancing overall discrimination accuracy.
- **High Accuracy Across Various Conditions:** The research demonstrates a minimum accuracy of 96% in sibling identification across different facial regions and conditions, outperforming traditional single-model approaches in handling real-world scenarios with incomplete facial data.

4. RESULTS AND DISCUSSIONS

In this work, a thorough series of experiments has been undertaken, focusing on sibling datasets and employing four state-of-the-art face recognition models: VGG19, VGG16, VGGFace, and FaceNet. The investigations encompass various facial features of siblings and employ frontal face images from the well-known SiblingsDB [11] dataset. Accessible to the research community upon request, SiblingsDB provides a diverse collection of high-quality sibling images. To facilitate the study, a subset of sibling image pairs has been extracted, including full frontal faces and cropped images for eyes, nose, and foreheads.

Simultaneously, a dataset for non-siblings has been curated using random images from SiblingsDB [11]. Similar to the sibling dataset, this non-sibling dataset includes cropped images of noses, eyes, and foreheads, forming distinct datasets for non-sibling image pairs. Experiments have been systematically conducted on these datasets, encompassing full-frontal faces, cropped noses, eyes, and foreheads for both siblings and non-siblings. Across each case, a variety of similarity measures have been applied to each state-of-the-art face model to assess its accuracy. The subsequent subsection will delve into the specifics of the datasets utilised, the conducted experiments on various datasets, and the detailed results obtained from these comprehensive evaluations.

4.1 Experiments on Different Face Parts

State-of-the-art face recognition models are effective while distinguishing two faces, but where only the part of the face is available to distinguish, the accuracy of the state-of-the-art face recognition models degrades. Combining a state-of-the-art face recognition model with a similarity measure yields optimal results for facial part analysis. However, there are some complex scenarios where the most accurate combination does not yield the best result. However, any other state-of-the-art face recognition model combination accurately discriminates the siblings using the same or another similarity measure. Hence, in this work, another set of experiments is conducted where the evaluation metrics for all four state-of-the-art face recognition models used are calculated, and a linear fusion is applied to those evaluation metrics, which combines the results from all the models and provides more robust results which have an opinion from all the models. These experiments have provided more robust results, which can be helpful in complex scenarios where the combination of the most accurate model and similarity measure has not accurately discriminated between the images. Figure 4 shows such complex scenarios for siblings and non-siblings where the most accurate combination has recognised the input image pairs as the same. Figure 4 (a) shows images from the datasets consumed: Face part of Sibling input image pair where the most accurate combination of state-of-the-art face recognition model and similarity measure has classified them as the same with the similarity metric score calculated by the respective face recognition model, while Figure 4 (b) shows images of the datasets used: Face part of non-sibling input image pair where the most accurate combination of state-of-the-art face recognition model and similarity measure has classified them as the same with the similarity metric score calculated by the respective face recognition model.

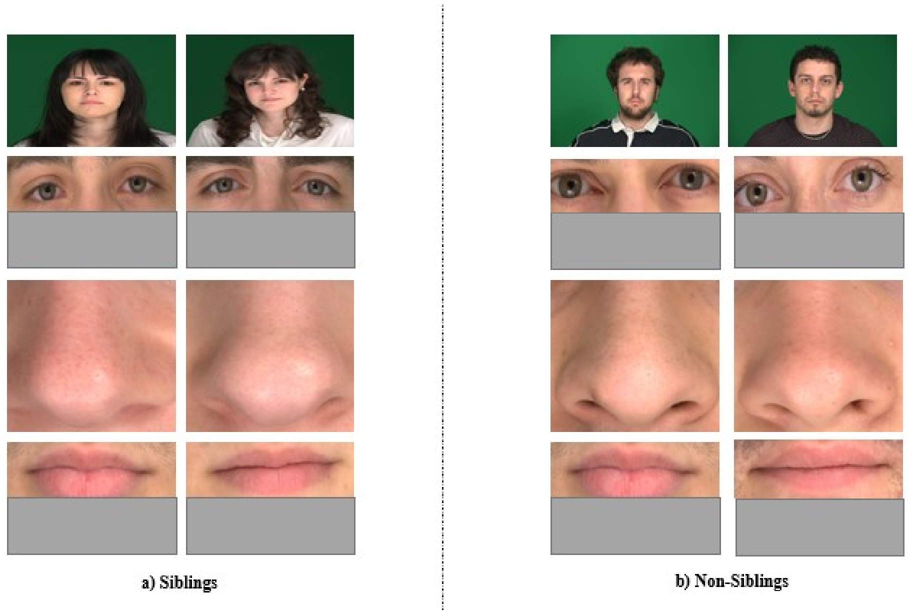

Figure 4. Complex Scenarios For (a) Siblings And (b) Non-siblings

Figure 4 shows images of some complex scenarios where the most accurate combination of model and similarity measure has not been effective. Hence, a linear fusion to take the opinion from all the models is helpful in such cases. Figure 5 shows the sibling discrimination performance using various evaluation metrics resulting from applying the linear fusion on the performance of all four state-of-the-art face recognition models: VGG19, VGG16, VGGFace, and FaceNet.

The analysis of the results demonstrates that employing linear fusion enhances the overall accuracy to a minimum of 88% by leveraging the combined performance of four advanced face recognition models. Specifically, when comparing full-frontal-face poses, the integration achieves a high accuracy of 96%. This outcome is attributed to prioritising VGG16 with cosine similarity, which had shown superior performance in previous experiments. For forehead comparisons, where the emphasis is on SSIM with VGG19, the accuracy reaches 93%.

In discriminating based on eyes, the fusion of results from all models achieves an accuracy of 92%. Here, the combination of VGGFace and Euclidean distance receives significant weighting. Similarly, for nose comparisons utilising FaceNet with Minkowski distance, the accuracy remains robust at 88%. Finally, in mouth comparisons, the fusion across all models achieves an accuracy of 91%, with VGG19 and SSIM being the most influential due to their previous excellent performance.

These findings underscore the effectiveness of combining multiple face recognition models through linear fusion to enhance accuracy across various facial features and poses, thereby demonstrating robust performance in discriminating between siblings and non-siblings in complex visual recognition tasks.

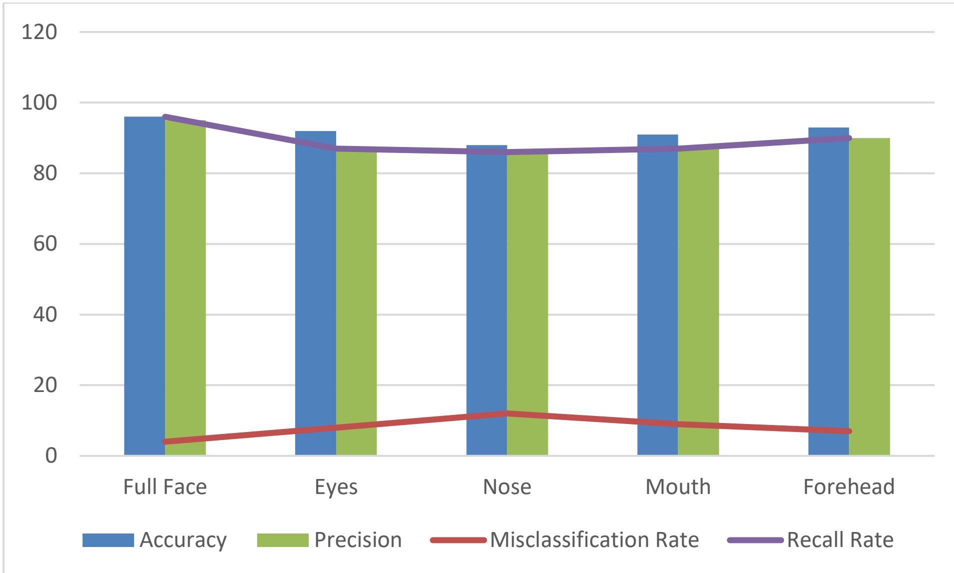

Figure 5. Performance Metrics Include Accuracy, Precision, Misclassification Rate, And Recall Rate For Discrimination Between Siblings Using The Fusion Of A State-of-the-art Face Recognition Model

4.2 Discussions

In this work, the focus is on addressing the intricate task of differentiating between individuals based on facial features, especially when dealing with similarities such as those between siblings, using state-of-the-art face recognition models. Building upon prior research, this study aims to enhance robustness in identification accuracy.

The experiments involved 204 pairs of siblings and non-siblings, evaluating the performance of four prominent face recognition models. A key methodology employed is linear fusion, which aggregates insights from all models rather than relying on individual model-similarity measure combinations. This approach enhances reliability by weighting each model's contribution based on its performance for each facial feature segment.

Specifically, weights are assigned to models according to their efficacy in differentiating between siblings across various facial parts. Models demonstrating superior performance receive higher weights, while those performing less effectively are given lower weights. This ensures a comprehensive assessment that considers the strengths of each model in different contexts.

Table 1 illustrates the specific models utilised, the corresponding weights applied to enhance accuracy in full-face comparisons of siblings, and the resultant accuracy achieved post-fusion. This comprehensive approach aims to achieve robust and reliable results in facial recognition, particularly in scenarios involving subtle differences between individuals with familial relationships.

Table 1. The Result Of The Fusion Applied To The Accuracy Of All Four Pre-trained State-of-the-art Face Recognition Models. The Fusion Accuracy Is Obtained By Applying The Weight To The Accuracy Of Each Model Based On Their Performance

Model	Accuracy	Weights	Fusion
FaceNet	0.91	0.236363636	0.215090909
VGGFace	0.96	0.249350649	0.239376623

VGG16	0.99	0.257142857	0.254571429
VGG19	0.99	0.257142857	0.254571429
Fused Accuracy			**0.96361039**

The application of fusion across all state-of-the-art face recognition models yields an F-measure of at least 0.86, underscoring the framework's robust precision and recall rates (see Figure 6). By aggregating insights from all four models, the framework achieves enhanced reliability even in challenging scenarios where individual models may falter.

Specifically, the framework achieves a high F-measure of 0.96 for discriminating full-frontal face poses, demonstrating near-perfect precision and recall. This implies that the model not only correctly identifies most true positives but also maintains low rates of false negatives. Similarly, in forehead discrimination, the F-measure is 0.90, with a false discovery rate (FDR) of 0.10. This indicates strong performance, as the model effectively identifies true positives with a manageable rate of false positives, while recall analysis shows that it captures a high proportion of actual forehead poses.

For eye detection, the framework achieves an F-measure of 0.87 with an FDR of 0.13. This suggests that the model is robust in distinguishing eyes, effectively balancing precision and recall to handle false discoveries appropriately.

In nose and mouth comparisons, the F-measures are 0.87 and 0.86, respectively. These scores indicate robust performance with high precision and recall, ensuring the model accurately identifies these features and minimises both false positives and false negatives.

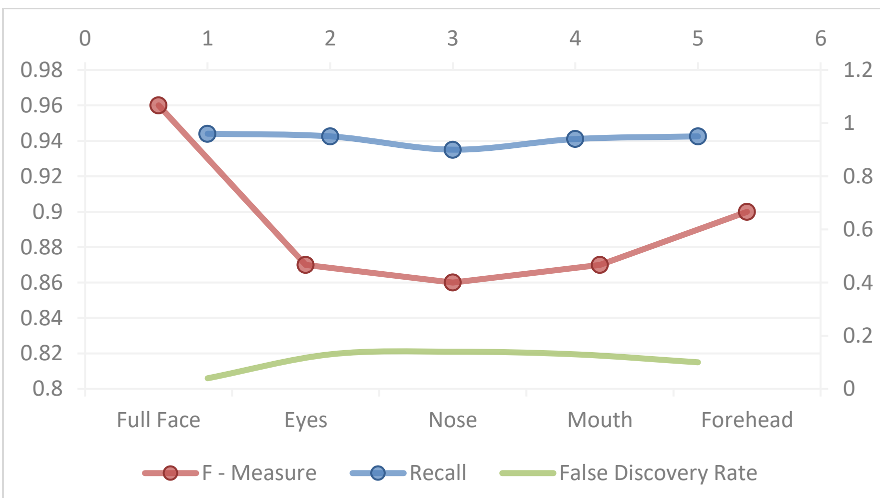

Figure 6. Analysis Of The Performance Of The Linear Fusion Framework Using F-Measure, Recall and FDR To Show The Robustness Of The Framework For Sibling Discrimination By Taking Opinions From All The Models Used In The Experiments

In summary, the fusion of multiple face recognition models enhances the framework's overall performance metrics. The high F-measure scores across various facial features and poses reflect the framework's effectiveness in achieving both high precision and recall, providing reliable and robust results for practical applications that require precise identification in complex visual recognition tasks.

5. CONCLUSION

This research presents a new framework developed to differentiate between siblings with high facial similarity. The framework tackles the challenge of inter-class similarity using advanced face recognition models and aggregating insights from four pre-trained models along with different similarity measures. Each model's performance is individually evaluated, and a linear fusion technique is employed to integrate various performance metrics, ensuring a robust outcome.

Experiments conducted in this study utilised custom-created datasets focusing on different facial features derived from the publicly available SiblingsDB database. The findings indicate that the suggested approach is highly effective in discriminating between siblings, showcasing its applicability in real-world scenarios where only partial facial information may be available.

Notably, while each state-of-the-art model demonstrates varying degrees of accuracy across different facial parts, certain combinations of models and similarity measures excel in specific discrimination tasks. However, complex cases arise where the optimal model-similarity measure combination fails to differentiate between siblings, as illustrated in Figure 4. In contrast, alternative model-similarity measure combinations successfully achieve accurate discrimination. Thus, leveraging insights from all models proves beneficial in overcoming such challenges.

Furthermore, applying linear fusion to aggregate the performance of all models enhances the robustness of results in these complex scenarios. This approach ensures that the framework maintains high discrimination accuracy across diverse facial features, underscoring its potential for practical applications in sibling identification and discrimination tasks.

This study primarily aimed to evaluate the practical performance of pre-trained face recognition models using linear fusion techniques. While cross-validation was not employed, this approach provided valuable insights into how these models perform in real-world sibling discrimination scenarios. However, to further enhance the robustness and generalizability of the findings, future research could benefit from incorporating cross-validation. Additionally, using a diverse and representative dataset could help ensure that the results are applicable across a broader range of sibling pairs.

ACKNOWLEDGEMENT

A version of this paper was presented at the fourth International Interdisciplinary Conference on Electrics and Computers, INTCEC 2024, held in the United States of America on 11th-13th June 2024.

The authors acknowledge *Dr. Pin Shen Teh* for providing feedback on this work.

FUNDING STATEMENT

This research received no specific grant from any funding agency

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Rita Goel: Conceptualization, Data Curation, Literature Review, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation.
Maida Alamgir: Formal Analysis, Literature Review, Validation, Visualisation, Writing –Review.
Haroon Wahab: Visualisation, Writing – Review & Editing.
Maria Alamgir: Formal Analysis, Review & Editing.
Hassan Ugail: Project Administration, Resources, Supervision.
Irfan Mehmood: Project Administration, Resources, Supervision.
Amit Sinha: Critical Analysis, Review.

CONFLICT OF INTERESTS

No conflicts of interest were disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/

REFERENCES

[1] V. Talreja, M. C. Valenti, and N. M. Nasrabadi, "Multibiometric secure system based on deep learning," *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2017, doi: 10.1109/globalsip.2017.8308652.

[2] P. Majumdar, A. Agarwal, R. Singh, and M. Vatsa, "Evading Face Recognition via Partial Tampering of Faces," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 11-20, doi: 10.1109/CVPRW.2019.00008.

[3] R. Goel, I. Mehmood, and H. Ugail, "A Study of Deep Learning-Based Face Recognition Models for Sibling Identification," *Sensors*, vol. 21, no. 15, p. 5068, 2021, doi: 10.3390/s21155068.

[4] J. W. Tanaka and D. Simonyi, "The 'Parts and Wholes' of Face Recognition: A Review of the Literature," *Quarterly Journal of Experimental Psychology*, vol. 69, no. 10, pp. 1876–1889, 2016, doi: 10.1080/17470218.2016.1146780.

[5] M. Almuashi, S. Z. Mohd Hashim, D. Mohamad, M. H. Alkawaz, and A. Ali, "Automated kinship verification and identification through human facial images: a survey," *Multimedia Tools and Applications,* vol. 76, no. 1, pp. 265-307, 2017, doi: 10.1007/s11042-015-3007-5.

[6] S. Zafeiriou, C. Zhang, and Z. Zhang, "A survey on face detection in the wild: Past, present and future," *Computer Vision and Image Understanding*, vol. 138, pp. 1–24, 2015, doi: 10.1016/j.cviu.2015.03.015.

[7] R. Fang, K. D. Tang, N. Snavely, and T. Chen, "Towards computational models of kinship verification," *IEEE International Conference on Image Processing*, 2010, doi: 10.1109/icip.2010.5652590.

[8] H. Lamba, A. Sarkar, M. Vatsa, R. Singh, and A. Noore, "Face recognition for look-alikes: A preliminary study," *International Joint Conference on Biometrics (IJCB)*, 2011, doi: 10.1109/ijcb.2011.6117520.

[9] S. Xia, M. Shao, and Y. Fu, *Kinship verification through transfer learning*. 2011. doi: 10.5591/978-1-57735-516-8/ijcai11-422.

[10] A. Shadrikov, "Achieving Better Kinship Recognition Through Better Baseline," *IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, 2020, doi: 10.1109/fg47880.2020.00137.

[11] T. F. Vieira, A. Bottino, A. Laurentini, and M. De Simone, "Detecting siblings in image pairs," *The Visual Computer*, vol. 30, no. 12, pp. 1333–1345, 2013, doi: 10.1007/s00371-013-0884-3.

[12] S. M. Mathews, C. Kambhamettu, and K. E. Barner, "Am I your sibling?"; Inferring kinship cues from facial image pairs, *2015 49th Annual Conference on Information Sciences and Systems (CISS)*, 2015, doi: 10.1109/ciss.2015.7086888.

[13] H. Yan and C. Song, "Multi-scale deep relational reasoning for facial kinship verification," *Pattern Recognition*, vol. 110, p. 107541, 2020, doi: 10.1016/j.patcog.2020.107541.

[14] L. He, H. Li, Q. Zhang, and Z. Sun, "Dynamic Feature Learning for Partial Face Recognition," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, doi: 10.1109/cvpr.2018.00737.

[15] S. Z. Li, D. Yi, Z. Lei, and S. Liao, "The CASIA NIR-VIS 2.0 Face Database," *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, doi: 10.1109/cvprw.2013.59.

[16] B. Lahasan, S. L. Lutfi, I. Venkat, M. A. Al-Betar, and R. San-Segundo, "Optimized symmetric partial facegraphs for face recognition in adverse conditions," *Information Sciences*, vol. 429, pp. 194–214, 2018, doi: 10.1016/j.ins.2017.11.013.

[17] A. Elmahmudi and H. Ugail, "Deep face recognition using imperfect facial data," *Future Generation Computer Systems*, vol. 99, pp. 213–225, 2019, doi: 10.1016/j.future.2019.04.025.

[18] Y. Zhu and Y. Jiang, "Optimization of face recognition algorithm based on deep learning multi feature fusion driven by big data," *Image and Vision Computing*, vol. 104, p. 104023, 2020, doi: 10.1016/j.imavis.2020.104023.

[19] S. Umer, B. C. Dhara, and B. Chanda, "Face recognition using fusion of feature learning techniques," *Measurement*, vol. 146, pp. 43–54, 2019, doi: 10.1016/j.measurement.2019.06.008.

[20] Z. Yujiao, L. W. Ang, S. Shaomin, and S. Palaniappan, "Dropout Prediction Model for College Students in MOOCs Based on Weighted Multi-feature and SVM," *Journal of Informatics and Web Engineering*, vol. 2, no. 2, pp. 29–42, 2023, doi: 10.33093/jiwe.2023.2.2.3.

[21] S. Nemati, R. Rohani, M. E. Basiri, M. Abdar, N. Y. Yen, and V. Makarenkov, "A Hybrid Latent Space Data Fusion Method for Multimodal Emotion Recognition," *IEEE Access*, vol. 7, pp. 172948–172964, 2019, doi: 10.1109/access.2019.2955637.

[22] S. Saleem, J. Amin, M. Sharif, M. A. Anjum, M. Iqbal, and S.-H. Wang, "A deep network designed for segmentation and classification of leukemia using fusion of the transfer learning models," *Complex & Intelligent Systems*, vol. 8, no. 4, pp. 3105–3120, 2021, doi: 10.1007/s40747-021-00473-z.

[23] J. Gao, P. Li, Z. Chen, and J. Zhang, "A Survey on Deep Learning for Multimodal Data Fusion," *Neural Computation,* vol. 32, no. 5, pp. 829-864, 2020, doi: 10.1162/neco_a_01273.

[24] Y. Zheng, Z. Xu, and X. Wang, "The Fusion of Deep Learning and Fuzzy Systems: A State-of-the-Art Survey," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 8, pp. 2783–2799, 2021, doi: 10.1109/tfuzz.2021.3062899.

[25] X. Sun and M. Lv, "Facial Expression Recognition Based on a Hybrid Model Combining Deep and Shallow Features," *Cognitive Computation*, vol. 11, no. 4, pp. 587–597, 2019, doi: 10.1007/s12559-019-09654-y.

[26] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, doi: 10.1109/cvpr.2015.7298682.

[27] M. S. Z. Ahmad, N. A. Ab. Aziz, and A. K. Ghazali, "Development of Automated Attendance System Using Pretrained Deep Learning Models," *International Journal on Robotics Automation and Sciences*, vol. 6, no. 1, pp. 6–12, 2024, doi: 10.33093/ijoras.2024.6.1.2.

[28] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," *British Machine Vision Conference,* 2015, doi: 10.5244/c.29.41.

[29] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv (Cornell University)*, 2014, doi: 10.48550/arxiv.1409.1556.

[30] C. Szegedy, S. E. Reed, D. Erhan, and D. Anguelov, "Scalable, High-Quality Object Detection," 2014. https://www.semanticscholar.org/paper/Scalable%2C-High-Quality-Object-Detection-Szegedy-Reed/4328ec9d98eff5d7eb70997f76d81b27849f3220

[31] A. Krizhevsky, I. Sutskever, G. E. Hinton, and University of Toronto, "ImageNet Classification with Deep Convolutional Neural Networks," *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, [Online]. Available:

https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[32] V. Sudha and T. R. Ganeshbabu, "A Convolutional Neural Network Classifier VGG-19 Architecture for Lesion Detection and Grading in Diabetic Retinopathy Based on Deep Learning," *Computers, Materials & Continua/Computers, Materials & Continua (Print)*, vol. 66, no. 1, pp. 827–842, 2020, doi: 10.32604/cmc.2020.012008.

[33] N. Shnain, Z. Hussain, and S. Lu, "A Feature-Based Structural Measure: An Image Similarity Measure for Face Recognition," *Applied Sciences*, vol. 7, no. 8, p. 786, 2017, doi: 10.3390/app7080786.

[34] A. R. Lahitani, A. E. Permanasari, and N. A. Setiawan, "Cosine similarity to determine similarity measure: Study case in online essay assessment," *International Conference on Cyber and IT Service Management*, 2016, doi: 10.1109/citsm.2016.7577578.

[35] A. Ultsch and J. Lötsch, "Euclidean distance-optimized data transformation for cluster analysis in biomedical data (EDOtrans)," *BMC Bioinformatics*, vol. 23, no. 1, 2022, doi: 10.1186/s12859-022-04769-w.

[36] I. Bakurov, M. Buzzelli, R. Schettini, M. Castelli, and L. Vanneschi, "Structural similarity index (SSIM) revisited: A data-driven approach," *Expert Systems With Applications*, vol. 189, p. 116087, 2021, doi: 10.1016/j.eswa.2021.116087.

[37] K. M. A. Parks, L. A. Griffith, N. B. Armstrong, and R. A. Stevenson, "Statistical Learning and Social Competency: The Mediating Role of Language," *Scientific Reports*, vol. 10, no. 1, 2020, doi: 10.1038/s41598-020-61047-6.

[38] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, 2020, doi: 10.1186/s12864-019-6413-7.

BIOGRAPHIES OF AUTHORS

	**Rita Goel** is a PhD research scholar at the Centre of Visual Computing. Her research encompasses machine learning, computer vision, and pattern recognition, with a particular emphasis on facial analysis. Rita's work involves developing and applying advanced computational techniques to improve the accuracy and efficiency of facial recognition systems. She aims to contribute to advancements in visual computing through her specialised research, addressing complex challenges in the field. She can be contacted at r.goel1@bradford.ac.uk.
	**Maida Alamgir** is a PhD student at the Centre for Visual Computing (CVC) at the University of Bradford, UK. Her research focuses on visual computing, machine learning, and facial analysis, particularly in their applications within the automobile industry. She can be contacted at m.alamgir@bradford.ac.uk.

	**Haroon Wahab** is a PhD student at the Centre for Visual Computing (CVC), University of Bradford, UK. His areas of research interests are representation learning, federated learning with applications in healthcare diagnostics, medical image analysis and decision support systems. He can be contacted at m.h.wahab@bradford.ac.uk.
	**Maria Alamgir** is a PhD student at the Manchester Metropolitan University, UK. Her research focuses on XR metaverse artificial intelligence, particularly in their applications within the healthcare industry. She can be contacted at maria.alamgir@stu.mmu.ac.uk
	**Dr. Irfan Mehmood** is currently serving as an Associate Professor at the University of Bradford, U.K. His sustained contribution to various research and industry collaborative projects gives him an extra edge to meet the current challenges faced in the field of multimedia analytics. Specifically, he has made a significant contribution in the areas of video summarisation, medical image analysis, visual surveillance, information mining, deep learning in industrial applications, and data encryption. He can be contacted at i.mehmood4@bradford.ac.uk.
	**Prof. Hassan Ugail** is the director of the Centre for Visual Computing in the Faculty of Engineering and Informatics at the University of Bradford, UK. He has a first-class BSc Honours degree in Mathematics from King's College London and a PhD in the field of geometric design from the School of Mathematics at the University of Leeds. Professor Ugail's research interests include computer-based geometric and functional design, imaging and machine learning. He can be contacted at h.ugail@bradford.ac.uk.
	**Dr. Amit Sinha**, Head and Professor of IT at ABES Engineering College, India, holds a PhD in Natural Language Processing. Specialising in NLP, Machine Learning, and Software Engineering, he excels in AI through extensive research and industry collaboration. His academic and practical expertise drives significant advancements in AI. Contact: amit.sinha@abes.ac.in.