JCTA Journal of Computing Theories and Applications ISSN:3024-9104

Research/Review Article

Multi-Descriptor Texture Fusion with Deep Residual Learning for Kinship Identification from Facial Images

Munzza Bibi, Wakeel Ahmad *, Syed M. Adnan, and Asifa Bibi

Department of Computer Science, Faculty of Information Engineering and Telecommunication, University of Engineering and Technology Taxila, Taxila 47050, Pakistan;
e-mail : munzza.bibi@students.uettaxila.edu.pk; wakeel.ahmad@uettaxila.edu.pk; syed.adnan@uettaxila.edu.pk; asifa.bibi2@students.uettaxila.edu.pk
* Corresponding Author : Wakeel Ahmad

Abstract: Kinship identification from facial images aims to determine biological relationships between individuals based on shared facial characteristics. However, subtle kinship-related facial cues are often obscured by variations in illumination, pose, age, and facial expressions, making reliable kinship classification challenging. To address this problem, this study proposes a hybrid framework that integrates handcrafted texture descriptors with deep features for multiclass kinship identification. A preprocessing pipeline consisting of image resizing to 224 × 224 pixels, noise reduction, intensity normalization, and facial region extraction is first applied to improve image consistency and feature quality. Three complementary local texture descriptors, namely Local Binary Pattern (LBP), Local Ternary Pattern (LTP), and Local Directional Pattern (LDP), are then employed to capture fine-grained facial texture and directional information. These handcrafted representations are fused with 2048-dimensional deep features extracted from a ResNet-50 model. The resulting 4562-dimensional pair representation is classified using a Support Vector Machine (SVM) under a four-class setting comprising father–son, father–daughter, mother–son, and mother–daughter relationships. Experiments on the KinFaceW-I dataset demonstrate that the proposed hybrid framework achieves a mean accuracy of 82.97%, with mean precision, recall, and F1-score of 82.99%, 83.00%, and 82.98%, respectively. The results further show consistent performance across all four relationship categories and competitive performance against existing methods, demonstrating the effectiveness of combining complementary local texture descriptors with deep semantic representations.

Keywords: Deep Learning; Facial Kinship Recognition; Feature Fusion; Hybrid Framework; KinFaceW-I; Machine Learning; ResNet-50; Support Vector Machine.

Received: April, 28th 2026
Revised: August, 4th 2026
Accepted: August, 5th 2026
Published: August, 18th 2026

1. Introduction

Kinship identification from facial images has emerged as an important research area in computer vision because of its wide range of practical applications. Beyond academic interest, kinship recognition plays a significant role in real-world scenarios such as missing person identification [1], forensic investigations [2], [3], child trafficking prevention [4], [5], genealogy analysis [6], immigration verification [7], family reunification [8], and family-based photo organization on social media platforms [9]. In forensic and law enforcement applications, automated kinship analysis can assist investigators in establishing potential biological relationships when direct identity evidence is unavailable. Similarly, genealogy and ancestry systems can benefit from facial analysis techniques to infer familial connections from historical or unstructured image collections.

Unlike conventional face recognition, which primarily aims to determine the identity of an individual, facial kinship identification seeks to infer biological relationships between different individuals based on shared facial characteristics [7]. This distinction makes kinship identification particularly challenging because inherited facial traits can be subtle and may be obscured by variations in age, facial expression, illumination, pose, and genetic diversity. Consequently, reliable kinship identification requires computational models capable of capturing

discriminative hereditary cues while remaining robust to variations in real-world facial appearance [10].

A major challenge in kinship identification is the substantial visual variability observed among related individuals. Variations in illumination, pose, facial expression, and age differences between related individuals can introduce considerable intra-family facial variability [11]. These factors may obscure shared facial characteristics and increase the difficulty of distinguishing genuine familial resemblance from incidental visual similarity. Although deep learning approaches have substantially advanced facial analysis, relying exclusively on deep representations may limit their ability to generalize effectively across the diverse conditions encountered in kinship recognition [12]. Therefore, developing representations that are sensitive to subtle genetic cues while remaining robust to visual variations remains an important research problem.

Traditional handcrafted texture descriptors, including Local Binary Pattern (LBP) [13], Local Ternary Pattern (LTP) [14], and Local Directional Pattern (LDP) [15], can capture fine-grained local facial texture and have demonstrated strong discriminative capability in various facial analysis tasks. However, handcrafted descriptors generally have limited capacity to encode high-level semantic information and complex facial structures. In contrast, deep convolutional neural networks such as ResNet-50 can learn hierarchical and highly informative representations of facial appearance. Nevertheless, deep networks may lose fine-grained texture information through progressive spatial downsampling and can be affected by variations in environmental conditions [16]. These complementary characteristics suggest that combining handcrafted and deep representations may provide a more comprehensive representation of facial characteristics [17]–[20].

Despite this potential, existing kinship recognition approaches still face several limitations. First, deep models primarily learn global and hierarchical facial representations, which may result in the loss of fine-grained texture patterns and local edge information during successive convolution and pooling operations. Such micro-level characteristics may contain useful hereditary cues for distinguishing related individuals. Second, existing hybrid approaches often rely on a limited number of handcrafted descriptors, such as standard LBP, which may not sufficiently capture directional structures or remain robust under noise, illumination changes, and facial expression variations. These limitations motivate the integration of multiple complementary local descriptors with deep facial representations within a unified framework.

To address these limitations, this study proposes a multi-scale feature fusion framework that integrates LBP, LTP, LDP, and ResNet-50 for multiclass kinship identification. LBP captures local surface texture patterns, while LTP introduces a threshold-based mechanism to improve robustness to small intensity variations and noise. LDP complements these descriptors by capturing directional edge responses across eight orientations using Kirsch masks. The resulting handcrafted representations are then fused with deep features extracted from ResNet-50 to jointly represent fine-grained local characteristics and high-level facial structure. Through this combination, the proposed framework aims to preserve complementary information that may be overlooked when individual descriptors are used independently.

The primary objective of this study is to investigate whether the combination of complementary handcrafted texture descriptors and deep facial representations can provide a more discriminative representation for multiclass kinship classification. In particular, the study evaluates the individual and combined contributions of LBP, LTP, LDP, and ResNet-50 and examines whether their fusion improves the classification of four kinship relationship categories: father–son (F-S), father–daughter (F-D), mother–son (M-S), and mother–daughter (M-D). The proposed framework is evaluated on the KinFaceW-I dataset using a multiclass Support Vector Machine (SVM) classifier. The main contributions of this study are summarized as follows:

- A hybrid kinship identification framework that integrates handcrafted texture descriptors (LBP, LTP, and LDP) with deep hierarchical features extracted from ResNet-50 to construct a richer facial representation.
- A standardized preprocessing pipeline incorporating noise reduction, normalization, facial cropping, and image resizing to 224 × 224 pixels to improve input consistency and facilitate feature extraction.

- A multiclass SVM-based classification strategy for distinguishing four kinship relationship categories: father–son, father–daughter, mother–son, and mother–daughter.
- A systematic ablation study evaluating individual descriptors, their progressive combinations, ResNet-50 alone, and the complete hybrid configuration to examine the contribution of different feature components to the classification performance.
- A comprehensive experimental evaluation on the KinFaceW-I dataset, including class-wise performance analysis and comparison with existing kinship recognition methods.

The remainder of this paper is organized as follows. Section 2 reviews related studies on facial kinship identification, handcrafted and deep facial features, and feature fusion approaches. Section 3 describes the proposed framework, including preprocessing, feature extraction, feature fusion, and multiclass classification. Section 4 presents the experimental setup and discusses the evaluation results, comparative analysis, and ablation study. Finally, Section 5 concludes the study, discusses its limitations, and outlines directions for future research.

2. Literature Review

This section reviews representative studies on facial kinship identification, with particular emphasis on deep learning, handcrafted feature extraction, feature fusion, and approaches designed to address variations in age, illumination, and facial appearance. Wu et al. [21] proposed a facial kinship verification approach based on remote photoplethysmography (rPPG) signals extracted from facial videos. Their method uses a one-dimensional convolutional neural network (1D-CNN) enhanced with a kinship contrastive loss and an attention mechanism to learn discriminative kinship-related information from multiple facial regions. Experiments on the UvA-NEMO Smile Database demonstrated the feasibility of using rPPG signals for kinship verification, providing an alternative to conventional image-based facial analysis.

Jiang et al. [22] introduced AutoSox, a framework designed to expand kinship verification datasets through synthetic facial image generation. The approach generates facial variations across different ages, genders, ethnic characteristics, and imaging conditions to improve dataset diversity. The authors augmented KinFaceW-I using ten styles and incorporated KinFace-CSFE, a Siamese-network-based feature extraction module combining channel-wise processing and hybrid-kernel spatial attention. YOCO-based augmentation was further employed to simulate challenging real-world imaging conditions. Experiments on KinFaceW-I, KinFaceW-II, and Syn-KinFaceW-I reported accuracies of 82.7%, 94.1%, and 83.2%, respectively.

Fibriani et al. [23] proposed SSCNetViT, which combines a Siamese Sequential Classification Network with a Vision Transformer for kinship verification based on facial microexpressions. Using the L32 backbone, the framework achieved an average accuracy of 90.07% on the LaIndo and FIW datasets, with reported accuracies of 99.5% for the BB relationship and 84.0% for the FD relationship. The B16 and B32 configurations achieved accuracies of 88.0% and 83.3%, respectively. These results indicate the potential of combining sequential classification and transformer-based representations for kinship analysis.

Oruganti et al. [24] developed a childhood image kinship recognition approach that combines the Curvelet Transform (CLT) for sub-band feature extraction with CNN-based texture descriptors. The method employs a threshold-optimized classifier and was evaluated on KinFaceW-I, KinFaceW-II, FIW, TSKinFace, UBKinFace, and childhood image datasets. The reported results demonstrate the applicability of multiscale frequency-domain features for kinship recognition under substantial age-related variations. Yan and Liu [25] proposed a lightweight CNN incorporating multi-scale feature fusion and channel-spatial attention. The method achieved accuracies of 83.85% on KinFaceW-I and 91.75% on KinFaceW-II, demonstrating the potential of combining multi-scale representations with attention mechanisms for facial kinship analysis.

Nader et al. [26] proposed a kinship verification framework that fuses color and texture features extracted from multiple color spaces. Their approach combines Local Binary Pattern (LBP), Scale-Invariant Feature Transform (SIFT), Heterogeneous Auto-Similarities of Characteristics (HASC), Dense Color Histogram (DCH), and Color Correlogram (CC). The framework consists of preprocessing, feature extraction, normalization, feature fusion, representa-

tion, and verification stages. Experiments on KinFaceW-I and KinFaceW-II achieved accuracies of 79.54% and 90.65%, respectively, demonstrating the potential of combining complementary handcrafted representations for kinship verification.

Othmani et al. [27] addressed class imbalance in kinship recognition using a deep learning framework based on ResNet-50, one-hot encoding, and data augmentation. Evaluated on the Families in the Wild (FIW) dataset, the method demonstrated improved recognition performance compared with the considered existing approaches, highlighting the importance of addressing class imbalance in kinship recognition.

Zhu et al. [28] introduced Distance and Direction Based Deep Discriminant Metric Learning (D-DDML), which incorporates both distance and directional information within a multitask learning framework. The method aims to improve the discrimination between kin and non-kin pairs while reducing computational requirements. Evaluations on KinFaceW-I, KinFaceW-II, Cornell, and UBKinFace demonstrated competitive performance against existing approaches. Kim et al. [29] proposed RA-GAN, a kinship verification framework based on face age transformation to address age-related variations. The method generates age-progressed facial images and aims to reduce age discrepancies between parent and child images. Experiments on KinFaceW-I, KinFaceW-II, and FIW showed improvements in kinship recognition performance across father–son, father–daughter, mother–son, and mother–daughter relationships.

Chen et al. [30] proposed Deep Discriminant Generation-Shared Feature Learning (D2GFL), which employs a two-stream architecture for extracting parent- and child-specific features together with a shared feature-learning module. By incorporating local and global family identity cues, the method reported accuracy improvements of 3.5%, 1.5%, and 0.8% on KinFaceW-I, Cornell, and TSKinFace, respectively. A deep learning approach based on MobileNetV2 was developed to segment key facial and human regions, including hair, head, clothing, torso, and skin, for kin and non-kin classification [31]. The method was evaluated on a self-collected dataset and reported improvements in accuracy and F-score, indicating its potential for family-photo identification and related forensic applications.

Mukherjee et al. [32] introduced FuseKin, which combines kinship information from multiple images of the same individual. The framework employs weighted multi-sample fusion and patch-based discriminant analysis to improve feature selection. Experiments on Family101 and FIW demonstrated competitive performance compared with existing approaches. Abbas et al. [33] proposed a facial kinship verification framework that combines a Siamese neural network with a Lifespan Age Transformation (LAT) algorithm to reduce domain-mismatch errors in transfer learning and align parent–child images across different ages. The framework achieved an overall accuracy of 76.38%, demonstrating the potential of age alignment for kinship verification.

Liu et al. [34] introduced an Age-Invariant Adversarial Feature (AIAF) learning framework that uses adversarial learning to disentangle identity- and age-related facial representations. An Identity Feature Weighted (IFW) module is subsequently used to emphasize purified kinship-related features. The approach was evaluated across multiple benchmark datasets and reported competitive verification performance. A baseline accuracy of approximately 71% on the FIW MM benchmark [35] further highlights the difficulty of kinship recognition under challenging conditions. Table 1 summarizes these representative approaches and their corresponding datasets and architectures.

Table 1. Summary of Representative State-of-the-Art Kinship Identification Methods.

Ref.	Year	Methodology	Dataset(s)	Classifier / Architecture
[21]	2025	rPPG + Attention + Contrastive Loss	UvA-NEMO Smile	1D-CNN
[22]	2025	AutoSox + KinFace-CSFE	KinFaceW-I, KinFaceW-II, Syn-KinFaceW-I	CNN with CSFE
[23]	2025	SSCNet-ViT	FIW, LaIndo	SSCNet + ViT
[24]	2024	Curvelet Transform + CNN	KinFaceW-I, KinFaceW-II, UB-KinFace, FIW	CNN + Curvelet
[25]	2024	Multi-scale CNN + Attention	KinFaceW-I, KinFaceW-II	Multi-scale CNN

Table 1 cont.

[26]	2023	HASC + CC + DCH	KinFaceW-I, KinFaceW-II	Gentle AdaBoost
[27]	2023	Deep CNN	FIW	ResNet-50
[28]	2023	D-DDML	KinFaceW-I, KinFaceW-II, Cornell, UBKinFace	D-DDML + Triplet Loss
[29]	2023	Face Age Transformation	KinFaceW-I, KinFaceW-II, FIW	ResNet-50, VGG
[30]	2022	D2GFL	KinFaceW-I, Cornell, TSKinFace	DNN
[31]	2022	MobileNetV2	Self-collected	MobileNetV2
[32]	2022	NRML + PDA + WMSF	Family101, FIW	NRML
[33]	2022	LAT + Siamese Network	RFIW	Siamese Network
[34]	2022	AIAF + IFW	KinFaceW-I, KinFaceW-II, FIW	Fully Connected Network

Based on Table 1, existing studies have explored deep representations, attention mechanisms, age transformation, synthetic augmentation, and handcrafted feature fusion for kinship recognition. However, the systematic combination of multiple complementary local texture descriptors (LBP, LTP, and LDP) with deep ResNet-50 representations remains insufficiently investigated, particularly for multiclass classification of the four kinship relationships in KinFaceW-I. This motivates the proposed framework, which evaluates both the individual and combined contributions of these feature components.

3. Proposed Method

3.1. Overview of the Proposed Framework

The overall workflow of the proposed hybrid framework is illustrated in Fig. 1. The framework is designed for multiclass facial kinship classification and consists of four main stages: preprocessing, parallel feature extraction, feature fusion, and classification. Given a pair of facial images, the preprocessing stage performs image resizing, noise removal, intensity normalization, and facial cropping to obtain standardized facial inputs.

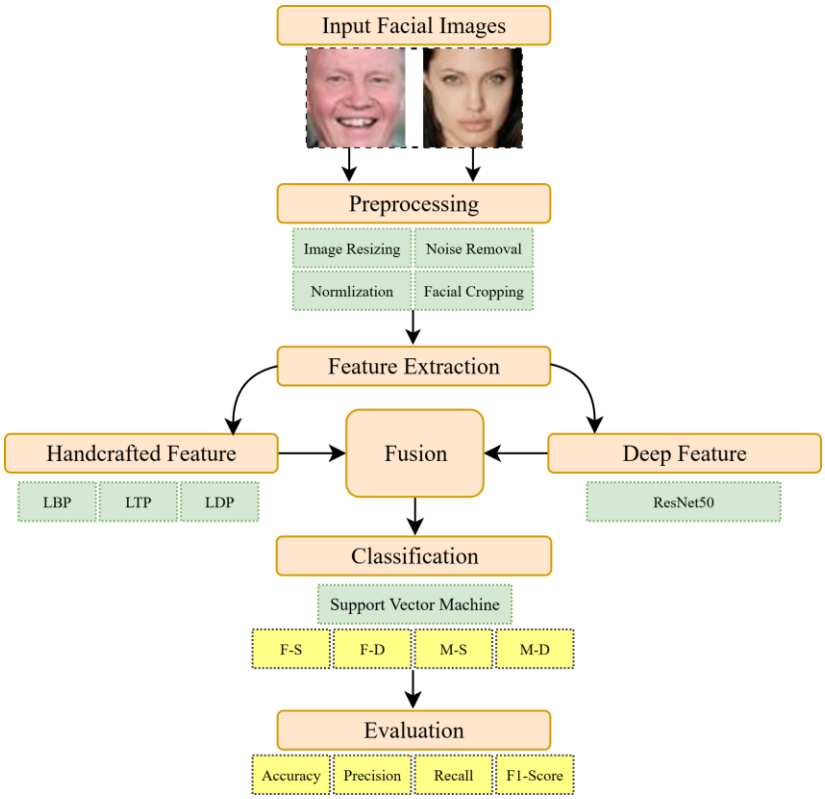

Figure 1. End-to-End Overview of the Proposed Framework Pipeline

Following preprocessing, two complementary types of facial representations are extracted in parallel. The first branch employs three handcrafted local texture descriptors, namely Local Binary Pattern (LBP), Local Ternary Pattern (LTP), and Local Directional Pattern (LDP), to capture fine-grained texture and directional information. The second branch uses ResNet-50 to extract high-level deep features representing broader facial structure and semantic information. The handcrafted and deep feature vectors are subsequently concatenated to form a unified representation for each facial image and then combined at the pair level. The resulting pair-level feature representation is classified using a Support Vector Machine (SVM) with an RBF kernel under a multiclass one-vs-one (OvO) strategy. The classifier predicts one of four kinship relationship categories: father–son (F-S), father–daughter (F-D), mother–son (M-S), and mother–daughter (M-D). The final classification performance is evaluated using accuracy, precision, recall, and F1-score.

3.2. Dataset

This study uses the KinFaceW-I dataset, introduced by Lu et al. [36], as the primary benchmark for facial kinship analysis. The dataset contains 1,066 facial images representing individuals and their relatives collected under uncontrolled imaging conditions. It includes variations in pose, illumination, background, facial expression, ethnicity, age, and partial occlusion, making it suitable for evaluating kinship classification under diverse facial conditions. Representative samples are shown in Fig. 2.

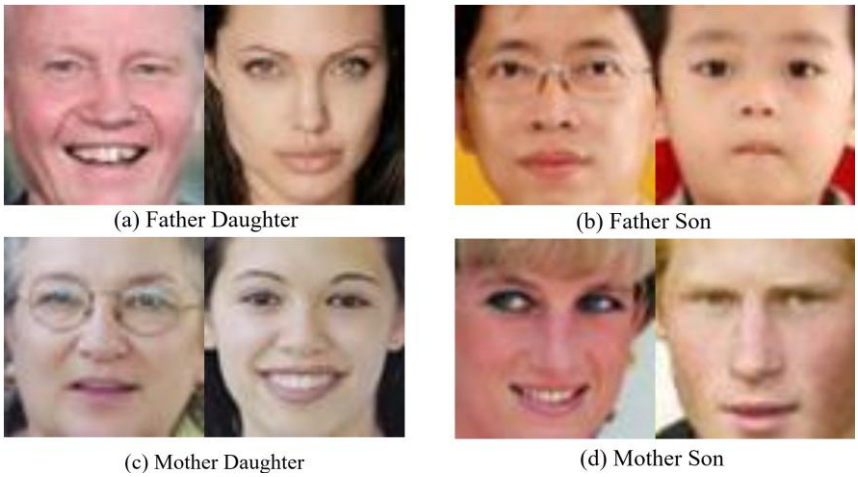

Figure 2. Sample Images from the KinFaceW-I Dataset

The dataset contains four parent–child relationship categories: father–son (F-S), father–daughter (F-D), mother–son (M-S), and mother–daughter (M-D), comprising 156, 134, 116, and 127 pairs, respectively. The detailed class-wise distribution and corresponding training and testing partitions are presented in Table 2.

Table 2. Class-wise Details of the KinFaceW-I Dataset

Kinship Relation	Original Pairs	Training Pairs (70%)	Testing Pairs (30%)	Augmented Training Pairs (5×)	Final Total Pairs
F-S	156	109	47	545	592
F-D	134	94	40	470	510
M-S	116	81	35	405	440
M-D	127	89	38	445	483
Total Pairs	533	373	160	1,865	2,025
Total Images	1,066	746	320	3,730	4,050

Note: The 70–30 split uses standard rounding to the nearest whole pair. The augmented training set represents five times the original training set size, while the testing set remains unchanged.

To address the limited sample size, data augmentation was applied exclusively to the training set. The original dataset was first divided into 70% training and 30% testing pairs before augmentation, resulting in 373 training pairs and 160 testing pairs, corresponding to 746 and 320 facial images, respectively. The training images were then augmented using geometric and photometric transformations, including horizontal flipping, random rotation within ±15°, brightness adjustment, scaling, and translation. This increased the training set from 746 to 3,730 images, while the original 320 testing images remained unchanged. This split-before-augmentation strategy prevents transformed versions of testing images from being introduced into the training set, thereby reducing the risk of data leakage and preserving the independence of the evaluation set. The final dataset consists of 4,050 images, including 3,730 augmented training images and 320 original testing images.

3.3. Preprocessing

The preprocessing stage is applied to standardize the facial images before feature extraction. Each image undergoes four sequential operations: resizing, Gaussian smoothing, intensity normalization, and facial region-of-interest (ROI) extraction. First, all input facial images are resized to 224 × 224 pixels to provide a consistent spatial resolution and match the input requirements of the ResNet-50 architecture. A Gaussian smoothing filter is then applied to reduce high-frequency noise and minor image distortions that may affect subsequent feature extraction. Next, pixel intensity values are normalized to the range [0, 1] to provide a consistent intensity scale across the dataset and reduce variations caused by different illumination conditions. Finally, a Haar Cascade classifier is used to detect and extract the facial region of interest. This step removes irrelevant background information and focuses the subsequent feature extraction on the facial region containing the most relevant visual information. The resulting images have a standardized resolution, intensity range, and facial region, and are subsequently used as inputs for both handcrafted texture and deep feature extraction.

3.4. Feature Extraction

Following preprocessing, two complementary types of features are extracted from each facial image in parallel: handcrafted texture features and deep hierarchical features. The handcrafted branch uses LBP, LTP, and LDP to capture fine-grained local texture and directional patterns, whereas the ResNet-50 branch extracts higher-level facial representations. This parallel strategy is designed to preserve complementary local and global facial information for subsequent feature fusion.

3.4.1. Handcrafted Feature Extraction

Three local texture descriptors, namely LBP, LTP, and LDP, are employed to characterize complementary facial texture and structural information. Their parameter settings and resulting feature dimensions are summarized in Table 3.

1. Local Binary Pattern (LBP)

LBP [13] is used to capture fine-grained local texture patterns by comparing each neighboring pixel γ_j with the corresponding reference pixel γ_c within a local neighborhood. Given M neighboring pixels and radius R, the LBP code at each pixel is computed as

$$\mathrm{LBP}_{M,R} = \sum_{j=1}^{M} 2^{j-1}\phi(\gamma_j - \gamma_c) \tag{1}$$

where the thresholding function $\phi(\cdot)$ is defined as:

$$\phi(z) = \begin{cases} 1 & \text{if } z \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

In this study, M=8 and R=1 are used. The resulting LBP codes are accumulated into a histogram to represent the local texture distribution of each facial image. To reduce histogram sparsity, a uniform LBP mapping is applied. With eight neighbors, 58 uniform patterns are retained, while all non-uniform patterns are grouped into a single additional bin, resulting in a 59-dimensional feature vector, denoted as X_{LBP}. The LBP encoding process is illustrated in Fig. 3.

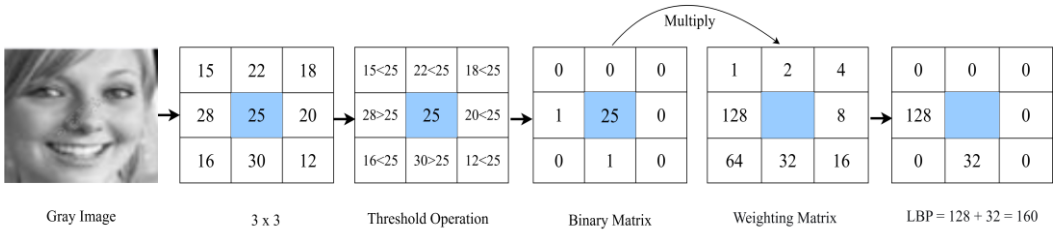

Figure 3. Illustration of LBP Encoding Using a 3 × 3 Neighborhood

LBP provides an efficient representation of local facial texture and is relatively insensitive to monotonic grayscale variations. However, its strict binary thresholding can make it more sensitive to local noise. This limitation motivates the use of LTP as a complementary texture descriptor.

2. Local Ternary Pattern (LTP):

LTP [14] extends the binary encoding of LBP by introducing a tolerance threshold τ. Rather than assigning a binary state to every neighboring pixel, LTP assigns three states according to the intensity difference between the neighboring pixel P_n and the center pixel C_p:

$$T(P_n; C_p; \tau) = \begin{cases} +1 & P_n \geq C_p + \tau \\ 0 & |P_n - C_p| < \tau \\ -1 & P_n \leq C_p - \tau \end{cases} \tag{3}$$

where τ controls the tolerance to small intensity variations. The resulting ternary pattern is decomposed into positive and negative binary patterns, which are computed using Eqs. (4) and (5).

$$\text{LTP}^+ = \sum_{n=0}^{N-1} s^+(p_n - p_c) \cdot 2^n, \quad s^+(z) = \begin{cases} 1 & \text{if } z \geq \tau \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

$$\text{LTP}^- = \sum_{n=0}^{N-1} s^-(p_n - p_c) \cdot 2^n, \quad s^-(z) = \begin{cases} 1 & \text{if } z \geq -\tau \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

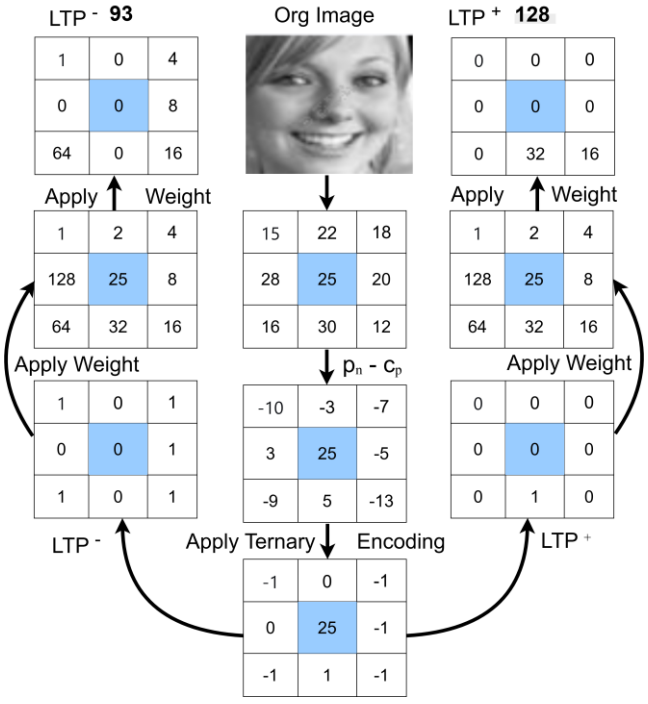

Figure 4. Illustration of LTP Encoding Using a 3 × 3 Neighborhood

In this study, τ=3 is used, while the neighborhood configuration follows R=1 and P_n =8. Each positive and negative binary pattern is represented using the same 59-bin uniform mapping as LBP. Consequently, the two components produce 118 dimensions (59 + 59), denoted as X_{LTP}. This dual representation provides greater tolerance to small local intensity variations while retaining complementary texture information.

3. Local Directional Pattern (LDP):

LDP [15] is incorporated to capture directional edge information that is not explicitly represented by LBP and LTP. For each pixel location (x, y), the image is convolved with eight 3×3 Kirsch compass masks $k_0, k_1, \ldots, k_7$, producing eight directional responses $r_0, r_1, \ldots, r_7$. The K strongest responses are then used to construct the LDP code according to:

$$\mathrm{LDP}^{(k)}(X,Y) = \sum_{i=0}^{7} I\,(r_i * r_{(k)}).\,2^i \tag{6}$$

where the indicator function $I(\cdot)$ is defined as $I(z) = \begin{cases} 1, & z \geq 0 \\ 0, & \text{otherwise} \end{cases}$.

The resulting LDP codes are accumulated into a histogram over the image to construct the final feature representation:

$$\mathrm{X}_{\mathrm{LDP}}(l) = \sum_{X=1}^{M}\sum_{Y=1}^{N} \delta\big(\mathrm{LDP}^{(K)}(X,Y), l\big) \tag{7}$$

where $\delta(a,b)$=1 if $a = b$ and 0 otherwise. In this study, K=3 is used to select the three most prominent directional responses, resulting in a 56-dimensional feature vector, denoted as X_{LDP}. By capturing directional edge structures, LDP provides complementary information to the local intensity patterns captured by LBP and LTP. The LDP encoding process is illustrated in Fig. 5.

15	22	18
28	25	20
16	30	12

Image

For k-th Selection (k = 3)

Direction	0	1	2	3	4	5	6	7
Response	-83	-3	-43	37	-11	-115	-147	109
Bit	0	1	0	1	0	0	0	1
Weight	1	2	4	8	16	32	64	128
Value	0	2	0	8	0	0	0	128

LDP = 2 + 8 + 128 = 138

Figure 5. Illustration of LDP Encoding Using a 3 × 3 Neighborhood

3.4.2. Deep Feature Extraction Using ResNet-50

ResNet-50 [16] is used to extract high-level facial representations that complement the local texture information obtained from the handcrafted descriptors. The model is pretrained on the ImageNet dataset and receives the preprocessed facial images as input. Features are extracted from the Global Average Pooling (GAP) layer, which provides a compact representation of high-level facial structure and appearance.

The resulting ResNet-50 representation contains 2048 dimensions and is normalized before feature fusion. Let $X_{ResNet50}$ denote the resulting deep feature vector. The ResNet-50

architecture used for feature extraction is illustrated in Fig. 6, and the configurations of all feature extraction components are summarized in Table 3.

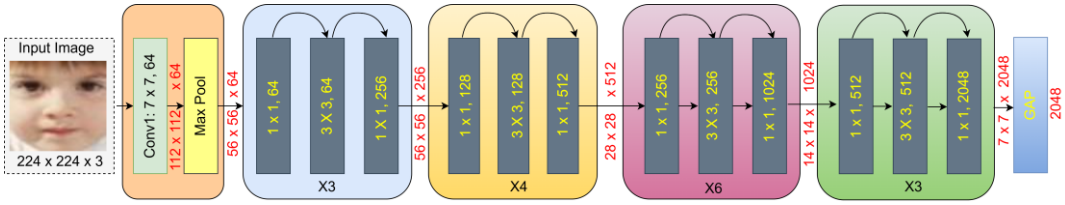

Figure 6. ResNet-50 Architecture for Deep Feature Extraction

Table 3. Configuration of the Feature Extraction Components

Component	Parameter	Value / Setting	Dimension
LBP	Radius (R), Neighbors (M), Type	R=1, M=8, Uniform	59
LTP	Radius (R), Neighbors (P_n), Threshold (τ)	R=1, P_n=8, τ=3	118
LDP	Mask size, Selected responses (K)	3×3 Kirsch, K=3	56
ResNet-50	Output layer	Global Average Pooling	2048

3.5. Feature Fusion

The extracted handcrafted and deep features are concatenated to preserve their complementary information in a unified representation. For each facial image, the individual feature vector is defined as

$$X_{Ind} = [X_{LBP} \| X_{LTP} \| X_{LDP} \| X_{ResNet50}] \tag{8}$$

and its dimensional composition is

$$X_{Ind} = [\ 59 \ \| \ 118 \ \| \ 56 \ \| \ 2048 \ ] \tag{9}$$

where ‖ denotes vector concatenation. With 59 LBP, 118 LTP, 56 LDP, and 2048 ResNet-50 features, each facial image is represented by 2,281 features.

Because the proposed framework performs multiclass kinship relationship classification, each sample is represented by a pair of parent and child facial images. The feature vectors of the two images are therefore concatenated side-by-side to preserve the complete representation of both individuals:

$$X_{Fusion} = X_{Parent} \| X_{Child} \tag{10}$$

where X_{Fusion} denotes the resulting pair-level representation. This produces a total of 4,562 features per kinship pair.

Feature concatenation is adopted because the four feature groups represent different types of information. LBP and LTP primarily capture local intensity and texture patterns, LDP provides directional edge information, and ResNet-50 provides higher-level facial representations. Concatenation retains these representations independently rather than forcing them into an element-wise mathematical interaction that could suppress information from individual descriptors. It therefore provides the classifier with access to the complete set of local and global characteristics.

No dimensionality reduction, such as Principal Component Analysis (PCA), is applied because reducing the fused representation could discard fine-grained information captured by the handcrafted descriptors. The resulting 4,562-dimensional pair representation is directly provided to the SVM classifier. To prevent differences in feature magnitude from disproportionately affecting the classifier, feature-wise normalization is applied to the fused representation. This scaling is particularly important because ResNet-50 activations and histogram-based handcrafted descriptors may have different numerical ranges. The normalization ensures that no feature group dominates the SVM optimization solely because of its scale.

3.6. Multiclass Classification

The fused pair-level representation integrates local texture information from LBP, LTP, and LDP with high-level features extracted by ResNet-50. This representation is provided to an SVM classifier for multiclass kinship relationship classification. SVM is selected because of its established use in kinship recognition [26], [37], [38] and its suitability for high-dimensional feature representations. Given the relatively small size of KinFaceW-I, an SVM provides a lightweight classification strategy while reducing the risk of overfitting associated with more parameter-intensive classifiers.

The classification task is formulated as a four-class problem, with the relationship labels defined as father–son (F-S), father–daughter (F-D), mother–son (M-S), and mother–daughter (M-D). This formulation differs from binary kinship verification, which predicts whether a pair is kin or non-kin. Accordingly, the proposed framework directly predicts the specific kinship relationship rather than using binary positive/negative pair terminology. For the multiclass classification task, SVM learns separating decision boundaries between the four relationship categories. The decision hyperplane is expressed as

$$W^T X + b = 0 \tag{11}$$

where w represents the normal vector to the hyperplane and b denotes the bias term. To accommodate nonlinear relationships in the high-dimensional fused feature space, a Gaussian Radial Basis Function (RBF) kernel is employed:

$$k(x_i, x_j) = \exp\left(-\gamma \left\| x_i , x_j \right\|^2\right) \tag{12}$$

The proposed framework uses a one-vs-one (OvO) strategy to distinguish the four kinship categories. Under this strategy, binary SVM classifiers are constructed for each pair of relationship classes, resulting in $K(K-1)/2 = 6$ binary classifiers for K=4 classes. For a new sample x, the final relationship label is determined through majority voting across the binary classifiers:

$$\mathrm{F}(x) = \arg\max_{c} \sum_{m=1}^{6} I(f_m(x) = c), \tag{13}$$

where $I(\cdot)$ denotes the indicator function and $h_m(x)$ represents the prediction of the m-th binary classifier.

The SVM hyperparameters, including C, which controls the trade-off between the classification margin and misclassification error, and γ, which determines the width of the RBF kernel, are optimized using grid search combined with cross-validation. The selected configuration is then used to classify unseen kinship pairs.

4. Results and Discussion

This section presents the experimental results of the proposed framework, including the evaluation metrics, individual and fused feature configurations, overall classification performance, comparison with existing methods, and ablation analysis.

4.1. Evaluation Metrics

The proposed framework is evaluated on the KinFaceW-I dataset using accuracy, precision, recall, and F1-score for the four kinship relationship categories: father–son (F-S), father–daughter (F-D), mother–son (M-S), and mother–daughter (M-D). These metrics provide complementary perspectives on multiclass classification performance. For each relationship category, precision measures the proportion of samples predicted as a particular class that are correctly assigned to that class, whereas recall measures the proportion of actual samples from that class that are correctly identified. The F1-score represents the harmonic mean of precision and recall and therefore provides a balanced measure when both types of errors are considered. Overall accuracy measures the proportion of all kinship pairs that are correctly classified across the four relationship categories. The metrics are calculated as follows:

$$\text{Overall Accuracy} = \frac{\text{Total Correct Predictions}}{\text{Total Samples}} \tag{14}$$

$$\text{Precision} = \frac{\text{Correct Predictions for Class}}{\text{All Predictions for Class}} \tag{15}$$

$$\text{Recall} = \frac{\text{Correct Prediction}}{\text{All Actual Samples}} \tag{16}$$

$$\text{F1 Measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{17}$$

Because the proposed task is formulated as four-class kinship relationship classification, the reported class-wise precision, recall, and F1-score provide additional information beyond overall accuracy and help characterize how consistently the framework distinguishes the individual relationship categories.

4.2. Evaluation Results Using Individual and Fused Feature Descriptors

The performance of the individual handcrafted descriptors, their combinations, and the ResNet-50 representation is evaluated to examine their respective contributions to kinship classification. The results are summarized in Table 4.

Table 4. Classification results for individual and fused feature models

Model	Kinship Category	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
LBP	F-D	64.33	32.98	41.33	36.69
	F-S	67.67	34.29	32.00	33.10
	M-D	66.00	30.99	29.33	30.14
	M-S	68.00	33.85	29.33	31.43
LTP	F-D	72.33	45.00	48.00	46.45
	F-S	63.67	27.03	26.67	26.85
	M-D	65.00	29.73	29.33	29.53
	M-S	67.00	33.33	32.00	32.65
LDP	F-D	68.33	38.37	44.00	40.99
	F-S	69.33	38.96	40.00	39.47
	M-D	73.67	46.97	41.33	43.97
	M-S	68.67	36.62	34.67	35.62
LBP-LTP	F-D	70.33	41.86	48.00	44.72
	F-S	68.00	41.46	68.00	51.52
	M-D	74.67	49.18	40.00	44.12
	M-S	75.67	53.33	21.33	30.48
LBP-LTP-LDP	F-D	76.67	54.55	40.00	46.15
	F-S	73.67	47.96	62.67	54.34
	M-D	78.00	57.14	48.00	52.17
	M-S	69.00	39.29	44.00	41.51
ResNet-50	F-D	78.00	56.52	52.00	54.17
	F-S	78.67	56.32	65.33	60.49
	M-D	78.33	57.14	53.33	55.17
	M-S	73.67	47.30	46.67	46.98

Among the handcrafted descriptors, LBP and LTP provide relatively limited performance when used individually. LBP achieves accuracies ranging from 64.33% to 68.00%

across the four relationship categories, while LTP ranges from 63.67% to 72.33%. LDP performs better overall among the individual handcrafted descriptors, achieving its highest accuracy of 73.67% for the M-D category. This result suggests that directional information provides useful complementary information for distinguishing facial kinship relationships. The ResNet-50 representation provides a stronger individual baseline than the handcrafted descriptors, with accuracies of 78.67%, 78.00%, 73.67%, and 78.33% for F-S, F-D, M-S, and M-D, respectively. This indicates that the deep representation captures discriminative facial characteristics that are not sufficiently represented by the individual local texture descriptors.

The results also show an improvement when handcrafted descriptors are combined. The LBP-LTP configuration achieves its highest accuracy of 75.67% for M-S, while adding LDP produces the LBP-LTP-LDP configuration, which reaches 78.00% for M-D and 76.67% for F-D. These results provide empirical evidence that combining different local descriptors can preserve complementary texture and directional information that may not be captured by a single descriptor. Overall, the results indicate that the handcrafted descriptors and ResNet-50 provide different levels of facial information. The handcrafted descriptors focus on local texture and directional patterns, whereas ResNet-50 provides a higher-level representation. The subsequent fusion of these feature types is therefore investigated as the proposed hybrid configuration.

4.3. Evaluation Results Using Hybrid Multi-Descriptor Fusion

The performance of the complete hybrid framework is evaluated using the fused LBP, LTP, LDP, and ResNet-50 representation. The class-wise results are summarized in Table 5. The framework achieves its highest performance for the mother–daughter (M-D) category, with precision, recall, F1-score, and accuracy of 87.91%, 87.29%, 87.60%, and 87.29%, respectively. The father–son (F-S) category follows with an accuracy of 83.26% and an F1-score of 83.25%. The mother–son (M-S) and father–daughter (F-D) categories show relatively lower performance, with accuracies of 81.06% and 80.28%, respectively. The corresponding F1-scores are 80.51% and 80.55%, indicating that the differences are also reflected when precision and recall are considered rather than accuracy alone.

Across the four categories, the framework obtains mean precision, recall, F1-score, and accuracy of 82.99%, 82.97%, 82.98%, and 82.97%, respectively. The relatively close values of these metrics indicate that the model does not rely solely on correct predictions for a particular class, although some differences remain among the individual relationship categories. The class-wise variation is further examined through the confusion matrix in the subsequent analysis. The proposed framework is formulated as a four-class kinship relationship classification problem, in which each facial pair is assigned to one of the F-S, F-D, M-S, or M-D categories. Therefore, the reported results should be interpreted as relationship classification performance rather than conventional binary kinship verification.

Table 5. Performance Metrics of the Proposed Hybrid Multi-Descriptor Fusion Framework

Metric	M-S (%)	F-S (%)	M-D (%)	F-D (%)	Average (%)
Precision	79.96	83.24	87.91	80.83	82.99
Recall	81.06	83.26	87.29	80.28	82.97
F1-score	80.51	83.25	87.60	80.55	82.98
Accuracy	81.06	83.26	87.29	80.28	82.97

4.4. Comparison with Existing Methods

Table 6 compares the proposed framework with representative methods that have reported results on the KinFaceW-I dataset. The reported values for the compared methods are taken from their respective publications, while the result for the proposed framework is obtained from the experiments conducted in this study. Therefore, the comparison is intended to provide a benchmark-level positioning of the proposed method rather than a reimplementation-based comparison.

The proposed framework achieves an overall accuracy of 82.97%, which is higher than most of the methods listed in Table 6. Its performance is comparable to Jiang et al. [22] (82.70%), Kim et al. [29] (82.40%), and Zhu et al. [28] (82.10%). The highest reported accuracy in the table is achieved by Yan and Liu [25] at 83.85%, which remains 0.88 percentage

points above the proposed framework. Thus, the proposed method demonstrates competitive performance on KinFaceW-I but does not achieve the highest reported accuracy among the compared methods.

The comparison also shows that the proposed framework performs better than several earlier approaches, including NRCML-LE [39] (66.30%), EPSM [40] (72.60%), NESN-KVN [41] (78.63%), and AdvKin [42] (78.68%). However, differences in model architecture, training procedures, preprocessing, augmentation, and experimental implementation may affect the reported values. Consequently, the results should be interpreted as a comparison with previously reported benchmark results, rather than as evidence of strict superiority under identical reimplementation conditions. More importantly, the comparative results indicate that the proposed relatively simple feature-fusion framework can achieve performance close to several more complex deep learning and attention-based approaches. This provides motivation for further investigating the contribution of the individual feature components through the ablation analysis presented in the next section.

Table 6. Performance comparison with existing methods on KinFaceW-I

Reference	Year	Methodology	Accuracy (%)
Jiang et al. [22]	2025	AutoSox + KinFace-CSFE	82.70
Oruganti et al. [24]	2024	CLT + CNN	81.01
Yan and Liu [25]	2024	Multi-scale CNN + Attention	83.85
Nader et al. [26]	2023	HASC + CC + DCH	79.54
Zhu et al. [28]	2023	D-DDML	82.10
Kim et al. [29]	2023	Face Age Transformation	82.40
Yan et al. [39]	2017	NRCML-LE	66.30
Meenpal et al. [40]	2023	EPSM	72.60
Wang et al. [41]	2020	NESN-KVN	78.63
Duan et al. [42]	2020	AdvKin	78.68
Sun et al. [43]	2025	FDFN	81.30
Nazari et al. [44]	2025	FNN	82.09
Su et al. [45]	2023	FaCoRNet	81.55
Proposed	2026	Multi-Descriptor Texture + Deep Fusion	82.97

4.5. Ablation Analysis

To examine the contribution of the individual feature components, an ablation analysis was conducted using the same experimental setting as the proposed framework. The results are presented in Table 7, covering standalone handcrafted descriptors, their progressive combinations, the ResNet-50 baseline, and the complete hybrid representation.

Among the standalone handcrafted descriptors, LBP, LTP, and LDP obtain mean accuracies of 66.50%, 67.00%, and 70.00%, respectively. LDP provides the strongest performance among the three handcrafted descriptors, particularly for the M-D category, where it achieves 73.67%. This result suggests that directional information provides useful structural cues that are not fully captured by intensity-based local patterns alone. ResNet-50 achieves a higher mean accuracy of 77.17% when used independently, indicating the contribution of high-level facial representations to the classification task.

Progressive fusion of the handcrafted descriptors further improves the results. Combining LBP and LTP increases the mean accuracy to 72.16%, while the addition of LDP raises it to 74.33%. The complete fusion of LBP, LTP, LDP, and ResNet-50 achieves the highest mean accuracy of 82.97%, representing an improvement of 5.80 percentage points over the ResNet-50-only configuration.

These results provide empirical support for retaining both handcrafted and deep feature representations. In particular, the progressive improvement from the handcrafted combinations, followed by the additional gain obtained after incorporating ResNet-50, indicates that the feature groups provide complementary information. The results also support the use of direct feature concatenation without applying dimensionality reduction before classification.

Table 7. Ablation analysis of different feature extraction configurations on the KinFaceW-I dataset

Configuration	F-S (%)	F-D (%)	M-S (%)	M-D (%)	Mean Accuracy (%)
LBP Only	67.67	64.33	68.00	66.00	66.50
LTP Only	63.67	72.33	67.00	65.00	67.00
LDP Only	69.33	68.33	68.67	73.67	70.00
LBP + LTP	68.00	70.33	75.67	74.67	72.16
LBP + LTP + LDP	73.67	76.67	69.00	78.00	74.33
ResNet-50 Only	78.67	78.00	73.67	78.33	77.17
Proposed	81.06	83.26	87.29	80.28	82.97

4.6. Discussion

The results indicate that the effectiveness of the proposed framework is primarily associated with the complementary characteristics of the feature representations. The handcrafted descriptors and ResNet-50 operate at different levels of facial representation: LBP and LTP describe local intensity patterns, LDP emphasizes directional structures, while ResNet-50 provides higher-level facial representations. Their combination therefore provides a broader representation than any individual component.

The ablation results support this interpretation. The handcrafted descriptors alone achieve lower mean accuracies than ResNet-50, but their progressive combination consistently improves performance from 66.50–70.00% for individual descriptors to 72.16% and 74.33% for the two- and three-descriptor configurations, respectively. Adding these local descriptors to ResNet-50 further increases the mean accuracy to 82.97%. This pattern suggests that the handcrafted descriptors retain local information that is complementary to the higher-level representation provided by the deep network.

The class-wise results also provide an indication of where the fused representation is more effective. Same-gender relationships, particularly M-D and F-S, obtain the highest accuracies in the final model. In contrast, F-D and M-S remain relatively more difficult to distinguish. The confusion matrix analysis in Fig. 7 similarly shows that several misclassifications occur between cross-gender relationship categories. These differences suggest that facial similarity across relationship categories cannot be captured equally well by a single representation and that local and global cues may contribute differently depending on the relationship type.

From a feature-representation perspective, LBP provides fine local texture patterns, LTP introduces tolerance to small intensity variations, and LDP emphasizes directional edge structures. ResNet-50 complements these descriptors by representing broader facial structure and appearance. Rather than assuming that any individual descriptor is sufficient, the proposed framework preserves these representations and allows the SVM classifier to use them jointly. This interpretation is consistent with the progressive gains observed in the ablation study. Overall, the findings suggest that the proposed fusion strategy is useful for representing kinship relationships from both local and higher-level facial information. However, the remaining differences between relationship categories also indicate that the framework does not completely resolve the difficulty of cross-gender classification. This limitation is important for future work, particularly for models that aim to learn more discriminative relationship-specific representations.

5. Conclusions

This study presented a hybrid framework for multiclass kinship identification by combining handcrafted local texture descriptors (LBP, LTP, and LDP) with deep features extracted using ResNet-50. The feature fusion strategy provides complementary local and high-level facial representations, enabling the model to distinguish four kinship categories: Father-Son (F-S), Father-Daughter (F-D), Mother-Son (M-S), and Mother-Daughter (M-D). Experiments on the KinFaceW-I dataset achieved an overall accuracy of 82.97%, with the ablation results supporting the contribution of combining handcrafted and deep features.

Despite these results, the study has several limitations. The evaluation is based on the KinFaceW-I benchmark, which contains relatively limited and semi-controlled facial data and therefore may not represent the variability encountered in unconstrained environments. In particular, the robustness of the framework under substantial illumination changes, occlusion,

pose variations, and large age differences remains to be further investigated. Future work will therefore consider more flexible feature-fusion strategies, including attention-based and vision-transformer architectures, to improve the representation of complementary facial cues. Evaluation on larger, unconstrained, and cross-age kinship datasets will also be important for assessing the generalizability of the framework beyond the current benchmark.

Author Contributions: Conceptualization: M.B. and W.A.; Methodology: S.M.A. and W.A.; Software: M.B. and A.B.; Validation: A.B. and M.B.; Formal Analysis: S.M.A., W.A, and M.B.; Investigation: A.B. and M.B.; Resources: W.A. and S.M.A.; Data Curation: A.B. and M.B.; Writing—Original Draft Preparation: A.B. and M.B.; Writing—Review & Editing: W.A. and S.M.A; Supervision: W.A. and S.M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The experimental data used in this study were obtained from the publicly available Kinship Face in the Wild (KinFaceW) dataset, which contains facial images collected for research on kinship verification from unconstrained facial images. The dataset is publicly hosted at https://www.kinfacew.com/download.html

Acknowledgments: The authors would like to thank the contributors of the KinFaceW dataset for making the source images publicly available. Google Gemini and ChatGPT were used solely for language polishing and grammatical improvement after completion of the research and original manuscript preparation. No generative AI tools were used for substantive content development, analysis, or scientific decision-making. The authors take full responsibility for the content, accuracy, and scientific integrity of the work.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] R. Li *et al.*, "Easykin: a flexible and user-friendly online tool for forensic kinship testing and missing person identification," *Int. J. Legal Med.*, vol. 137, no. 6, pp. 1671–1681, Nov. 2023, doi: 10.1007/s00414-023-03083-1.

[2] G. Cosenza, L. Castellino, S. Morelli, E. Alladio, and E. Pilli, "Kinship analysis and machine learning algorithms in forensic contexts: A new NGS panel," *Expert Syst. Appl.*, vol. 266, p. 126161, Mar. 2025, doi: 10.1016/j.eswa.2024.126161.

[3] E. Haimes, "Social and Ethical Issues in the Use of Familial Searching in Forensic Investigations: Insights from Family and Kinship Studies," *J. Law, Med. Ethics*, vol. 34, no. 2, pp. 263–276, Jan. 2006, doi: 10.1111/j.1748-720X.2006.00032.x.

[4] L. Han, "New Technologies in Combating Child Trafficking in China: Opportunities and Challenges for Children's Rights," *Peace Hum. Rights Gov.*, vol. 3, no. 3, pp. 389–414, 2019, doi: 10.14658/PUPJ-PHRG-2019-3-5.

[5] M. Saiz, M. J. Alvarez-Cubero, J. C. Alvarez, and J. A. Lorente, "Forensic Genetics Against Children Trafficking: Missing Children Genetic Identification," in *Handbook of Missing Persons*, Cham: Springer International Publishing, 2016, pp. 177–189. doi: 10.1007/978-3-319-40199-7_13.

[6] R. J. Quinlan and E. H. Hagen, "New Genealogy: It's Not Just for Kinship Anymore," *Field methods*, vol. 20, no. 2, pp. 129–154, May 2008, doi: 10.1177/1525822X07314034.

[7] 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," *Multimed. Tools Appl.*, vol. 76, no. 1, pp. 265–307, 2017, doi: 10.1007/s11042-015-3007-5.

[8] A. Durfey, "Leveraging AI and Genealogy for Family Reunification During Humanitarian Crises," in *Advanced Sciences and Technologies for Security Applications*, 2025, pp. 357–368. doi: 10.1007/978-3-031-86997-6_11.

[9] Q. Yuan and S. Xia, "Kinship Understanding for a Family Photo," in *2019 Chinese Control and Decision Conference (CCDC)*, Jun. 2019, pp. 5319–5323. doi: 10.1109/CCDC.2019.8832876.

[10] M. C. Mzoughi, N. Ben Aoun, and S. Naouali, "A review on kinship verification from facial information," *Vis. Comput.*, vol. 41, no. 3, pp. 1789–1809, Feb. 2025, doi: 10.1007/s00371-024-03493-1.

[11] S. A. Khan, M. Ishtiaq, M. Nazir, and M. Shaheen, "Face recognition under varying expressions and illumination using particle swarm optimization," *J. Comput. Sci.*, vol. 28, pp. 94–100, Sep. 2018, doi: 10.1016/j.jocs.2018.08.005.

[12] M. Chung, M. Wang, Z. Huang, and T. Okuyama, "Diverse sensory cues for individual recognition," *Dev. Growth Differ.*, vol. 62, no. 9, pp. 507–515, Dec. 2020, doi: 10.1111/dgd.12697.

[13] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002, doi: 10.1109/TPAMI.2002.1017623.

[14] Xiaoyang Tan and B. Triggs, "Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010, doi: 10.1109/TIP.2010.2042645.

[15] T. Jabid, M. H. Kabir, and Oksam Chae, "Local Directional Pattern (LDP) for face recognition," in *2010 Digest of Technical Papers International Conference on Consumer Electronics (ICCE)*, Jan. 2010, pp. 329–330. doi: 10.1109/ICCE.2010.5418801.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, vol. 2016-Decem, pp. 770–778. doi: 10.1109/CVPR.2016.90.
[17] F. Dornaika, I. Arganda-Carreras, and O. Serradilla, "Transfer learning and feature fusion for kinship verification," *Neural Comput. Appl.*, vol. 32, no. 11, pp. 7139–7151, Jun. 2020, doi: 10.1007/s00521-019-04201-0.
[18] M. A. Almuashi, S. Z. M. Hashim, N. Yusoff, and K. N. Syazwan, "A Fusion Schema of Hand-Crafted Feature and Feature Learning for Kinship Verification," in *Lecture Notes on Data Engineering and Communications Technologies*, 2021, pp. 1050–1063. doi: 10.1007/978-3-030-70713-2_94.
[19] L. R. Zuama, D. R. I. M. Setiadi, A. Susanto, S. Santosa, H.-S. Gan, and A. A. Ojugo, "High-Performance Face Spoofing Detection using Feature Fusion of FaceNet and Tuned DenseNet201," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 4, pp. 385–400, Feb. 2025, doi: 10.62411/faith.3048-3719-62.
[20] V. Kumar and A. Purohit, "Tuned Attention-Guided Fusion of Deep and Handcrafted Features for Distinguishing AI-Generated and Human-Created Artworks," *J. Futur. Artif. Intell. Technol.*, vol. 2, no. 4, pp. 629–647, 2026, doi: 10.62411/faith.3048-3719-291.
[21] X. Wu, X. Feng, C. Á. Casado, L. Liu, and M. B. López, "Exploring Facial Kinship Verification through Contactless Heart Activity Analysis," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2025, pp. 1–5. doi: 10.1109/ICASSP49660.2025.10888280.
[22] J.-X. Jiang, H. Jing, L. Zhou, Y. Li, and Z. Wang, "Multi-attribute balanced dataset generation framework AutoSyn and KinFace Channel-Spatial Feature Extractor for kinship recognition," *Neurocomputing*, vol. 613, p. 128750, Jan. 2025, doi: 10.1016/j.neucom.2024.128750.
[23] I. Fibriani, E. M. Yuniarno, R. Mardiyanto, and M. H. Purnomo, "SSCNetViT: A Hybrid Siamese Sequential Classification-ViT Model for Kinship Recognition in Indonesia Facial Microexpression," *Int. J. Intell. Eng. Syst.*, vol. 18, no. 1, pp. 832–846, Feb. 2025, doi: 10.22266/ijies2025.0229.59.
[24] M. Oruganti, T. Meenpal, and S. Majumder, "Kinship verification in childhood images using curvelet transformed features," *Comput. Electr. Eng.*, vol. 118, p. 109375, Aug. 2024, doi: 10.1016/j.compeleceng.2024.109375.
[25] C. Yan and Y. Liu, "Kinship verification based on multi-scale feature fusion," *Multimed. Tools Appl.*, vol. 83, no. 40, pp. 88069–88090, Mar. 2024, doi: 10.1007/s11042-024-18879-5.
[26] N. Nader, F. E.-Z. A. El-Gamal, and M. Elmogy, "Enhanced kinship verification analysis based on color and texture handcrafted techniques," *Vis. Comput.*, vol. 40, no. 4, pp. 2325–2346, Apr. 2024, doi: 10.1007/s00371-023-02919-6.
[27] A. Othmani, D. Han, X. Gao, R. Ye, and A. Hadid, "Kinship recognition from faces using deep learning with imbalanced data," *Multimed. Tools Appl.*, vol. 82, no. 10, pp. 15859–15874, Apr. 2023, doi: 10.1007/s11042-022-14058-6.
[28] X. Zhu, C. Li, X. Chen, X. Zhang, and X.-Y. Jing, "Distance and Direction Based Deep Discriminant Metric Learning for Kinship Verification," *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 19, no. 1s, pp. 1–19, Feb. 2023, doi: 10.1145/3531014.
[29] H. Kim, H. Kim, J. Shim, and E. Hwang, "A robust kinship verification scheme using face age transformation," *Comput. Vis. Image Underst.*, vol. 231, p. 103662, Jun. 2023, doi: 10.1016/j.cviu.2023.103662.
[30] X. Chen *et al.*, "Deep discriminant generation-shared feature learning for image-based kinship verification," *Signal Process. Image Commun.*, vol. 101, p. 116543, Feb. 2022, doi: 10.1016/j.image.2021.116543.
[31] T. Karnik, P. Shivakumara, P. N. Chowdhury, U. Pal, T. Lu, and N. B. Anuar, "A new deep model for family and non-family photo identification," *Multimed. Tools Appl.*, vol. 81, no. 2, pp. 1765–1785, Jan. 2022, doi: 10.1007/s11042-021-11631-3.
[32] M. Mukherjee, T. Meenpal, and A. Goyal, "FuseKin: Weighted image fusion based kinship verification under unconstrained age group," *J. Vis. Commun. Image Represent.*, vol. 84, p. 103470, Apr. 2022, doi: 10.1016/j.jvcir.2022.103470.
[33] A. Abbas and M. Shoaib, "Kinship identification using age transformation and Siamese network," *PeerJ Comput. Sci.*, vol. 8, p. e987, Jun. 2022, doi: 10.7717/peerj-cs.987.
[34] F. Liu, Z. Li, W. Yang, and F. Xu, "Age-Invariant Adversarial Feature Learning for Kinship Verification," *Mathematics*, vol. 10, no. 3, p. 480, Feb. 2022, doi: 10.3390/math10030480.
[35] J. P. Robinson, Z. Khan, Y. Yin, M. Shao, and Y. Fu, "Families in Wild Multimedia: A Multimodal Database for Recognizing Kinship," *IEEE Trans. Multimed.*, vol. 24, pp. 3582–3594, 2022, doi: 10.1109/TMM.2021.3103074.
[36] Jiwen Lu, Xiuzhuang Zhou, Yap-Pen Tan, Yuanyuan Shang, and Jie Zhou, "Neighborhood Repulsed Metric Learning for Kinship Verification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 331–345, Feb. 2014, doi: 10.1109/TPAMI.2013.134.
[37] S. Mahpod and Y. Keller, "Kinship verification using multiview hybrid distance learning," *Comput. Vis. Image Underst.*, vol. 167, pp. 28–36, Feb. 2018, doi: 10.1016/j.cviu.2017.12.003.
[38] H. Wu, J. Chen, X. Liu, and J. Hu, "Component-based metric learning for fully automatic kinship verification," *J. Vis. Commun. Image Represent.*, vol. 79, p. 103265, Aug. 2021, doi: 10.1016/j.jvcir.2021.103265.
[39] H. Yan, "Kinship verification using neighborhood repulsed correlation metric learning," *Image Vis. Comput.*, vol. 60, pp. 91–97, Apr. 2017, doi: 10.1016/j.imavis.2016.08.009.
[40] M. Oruganti, T. Meenpal, and S. Majumder, "Easy pair selection method for Kinship Verification using fixed age group images," *Image Vis. Comput.*, vol. 136, p. 104727, Aug. 2023, doi: 10.1016/j.imavis.2023.104727.
[41] S. Wang and H. Yan, "Discriminative sampling via deep reinforcement learning for kinship verification," *Pattern Recognit. Lett.*, vol. 138, pp. 38–43, Oct. 2020, doi: 10.1016/j.patrec.2020.06.019.
[42] L. Zhang, Q. Duan, D. Zhang, W. Jia, and X. Wang, "AdvKin: Adversarial Convolutional Network for Kinship Verification," *IEEE Trans. Cybern.*, vol. 51, no. 12, pp. 5883–5896, Dec. 2021, doi: 10.1109/TCYB.2019.2959403.
[43] S. Sun, Y. Yang, and H. Yan, "Kinship verification via Frequency Feature Decoupling and Fusion," *Pattern Recognit. Lett.*, vol. 193, pp. 1–7, Jul. 2025, doi: 10.1016/j.patrec.2025.04.002.
[44] A. Nazari, O. Borzoei, and M. E. Moghaddam, "Kinship Verification through a Forest Neural Network," *arXiv*. [Online]. Available: https://arxiv.org/abs/2504.18910

[45] W.-T. Su, M.-H. Chen, C.-Y. Wang, S.-H. Lai, and T. Chen, "Kinship Representation Learning with Face Componential Relation," in *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Oct. 2023, pp. 3097–3106. doi: 10.1109/ICCVW60793.2023.00335.