Journal of Information Hiding and Multimedia Signal Processing ©2026 ISSN 2073-4212
Ubiquitous International Volume **17**, Number **1**, March **2026**

Advancing Kinship Verification with Relational Deep Feature Fusion

Nour El Houda Bouakal[1], Messaoud Hettiri[1] and Djamel Eddine Boukhari[2,*]

[1]LGEERE Laboratory, Department of Electrical Engineering
University of El Oued, 39000 El-Oued, Algeria
[2]Scientific and Technical Research Centre for Arid Areas
CRSTRA, 07000, Biskra, Algeria

*Corresponding author: Djamel Eddine Boukhari
boukhari-djameleddine@univ-eloued.dz

Received October 28, 2025, revised February 3, 2026, accepted February 5, 2025.

Abstract. *Automated kinship verification from facial images remains a significant and challenging frontier in computer vision, requiring the discernment of subtle hereditary resemblances that transcend simple identity matching. While deep learning has advanced the field, many existing methods repurpose generic face recognition architectures, often failing to capture the specific, relational features essential for distinguishing kin. To address this gap, we propose a robust, multi-stage framework that synergistically combines deep feature extraction with a powerful classification model. Our method employs a pre-trained Residual Network (ResNet-50) to derive deep facial representations, introduces a novel relational feature fusion technique to generate a unified pairwise vector from parent-child images, and refines this representation using T-Test based discriminative feature selection. A Support Vector Machine (SVM) with an RBF kernel performs the final binary classification. Extensive experiments conducted on five challenging public benchmarks—KinFaceW-I, KinFaceW-II, Cornell, Family 101, and UB KinFace—unequivocally establish a new state-of-the-art. Our proposed system consistently outperforms all compared methods by a substantial margin, achieving a mean accuracy of 98.06% on the demanding KinFaceW-I dataset. The findings demonstrate the profound efficacy of a hybrid approach that decouples deep representation learning from a powerful, margin-based classifier, setting a new performance standard for kinship verification.*
Keywords: Kinship Verification; Deep Learning; Convolutional Neural Networks (CNN); Feature Fusion; Support Vector Machine (SVM); Face Analysis; ResNet.

1. **Introduction.** The remarkable success of deep learning has propelled facial analysis to the forefront of computer vision, largely solving canonical tasks such as identity recognition and demographic estimation [1]. However, the next frontier lies in transcending simple classification to achieve a more profound, human-like semantic understanding of facial attributes[2]. Among the most challenging of these tasks is automated kinship verification from images, a process that requires machines to discern the subtle and intricate patterns of hereditary resemblance[3]. This capability holds significant implications for applications ranging from forensic investigation and missing person identification to large-scale genealogical studies and computational anthropology.

At its core, kinship verification is plagued by the fundamental computer vision challenge of a wide "semantic gap" between low-level pixel data and the high-level abstract concept of familial relationships[4]. The task is to learn a discriminative function that is robust to substantial intra-class variance—due to age, gender, expression, and imaging conditions—while remaining sensitive to the subtle inter-class differences that distinguish kin from non-kin pairs [5]. This inherent difficulty has made kinship verification an active and unresolved area of research.

Early attempts to address this problem relied predominantly on handcrafted local descriptors (e.g., LBP, SIFT, HOG) combined with shallow machine learning models [6]. While foundational, these methods often fail to capture the holistic and deeply-seated facial characteristics that signify genetic relatedness, proving brittle when faced with the unconstrained nature of "in-the-wild" images. More recently, Convolutional Neural Networks (CNNs) have become the dominant paradigm. Yet, many existing deep learning approaches either repurpose architectures designed for face recognition, which may learn general facial similarity rather than specific hereditary markers, or depend on complex, hand-designed metric learning schemes that can be difficult to optimize and may not generalize well across diverse datasets [7, 8]. A critical research gap therefore exists for a framework specifically engineered to extract and represent features optimized for the unique problem of kinship verification.

In this paper, we address these limitations by proposing a hierarchical, feature-driven system that synergistically integrates advanced deep feature extraction with robust classification. Our core hypothesis is that hereditary traits manifest as a complex interplay of multi-scale facial features. To capture these, we leverage a pre-trained Residual Network (ResNet) [9], renowned for its ability to learn rich, deep feature hierarchies. We introduce a principled pipeline encompassing facial pre-processing, deep feature extraction, a novel feature fusion and normalization scheme to create a unified representation for parent-child pairs, and discriminative feature selection. The final verification is performed via both a Support Vector Machine (SVM) and an end-to-end regression-based CNN, allowing for a comprehensive analysis of classical versus deep learning classifiers for this task.

The primary contributions of this work are threefold:

- We introduce a novel deep feature extraction and fusion framework tailored specifically to the kinship verification problem, which effectively transforms paired facial images into a highly discriminative feature space.
- We conduct a comprehensive empirical analysis of our proposed system on five challenging, publicly available benchmark datasets (KinFaceW-I, KinFaceW-II, Cornell, Family 101, and UB KinFace), providing a comparative evaluation of both SVM and deep regression classifiers in this context.
- We demonstrate that our approach achieves state-of-the-art or superior performance across all evaluated datasets, establishing a new and robust performance benchmark and confirming the efficacy of dedicated feature engineering in the deep learning era for specialized semantic tasks.

The remainder of this paper is organized as follows: Section 2 surveys related literature. Section 3 details the proposed methodology. Section 4 presents the experimental setup, results, and analysis. Finally, Section 5 provides concluding remarks and suggests future research directions.

2. Related Works.

The problem of automated kinship verification has been an active research area for over a decade, evolving in lockstep with advancements in computer vision and machine learning. The landscape of existing methodologies can be broadly categorized into two major paradigms: those based on handcrafted feature representations and those driven by deep learning.

2.1. Early Approaches with Handcrafted Features.

Initial forays into kinship verification predominantly relied on extracting low-level visual features using carefully engineered descriptors. These methods sought to model facial similarity by capturing geometric arrangements, textures, and local patterns. For instance, Fang et al. [?], in their pioneering work on the Cornell KinFace database, utilized a Pictorial Structure Model (PSM) to represent facial components. Other notable approaches leveraged a rich set of descriptors, including Weber Local Descriptor (WLD) across multiple color spaces [?], Local Binary Patterns (LBP), and Scale-Invariant Feature Transform (SIFT) points, often in conjunction with a metric learning framework. These handcrafted features were typically fed into conventional classifiers such as the k-Nearest Neighbors (k-NN) or, more commonly, Support Vector Machines (SVMs) [?]. While these methods established foundational benchmarks, their performance was fundamentally constrained by the descriptive power of the manually designed features. They often struggled to generalize across diverse datasets and proved susceptible to variations in pose, illumination, and expression, failing to capture the abstract semantic cues of heredity.

2.2. Deep Learning-Based Kinship Verification.

The paradigm shift towards deep learning has catalyzed transformative progress in kinship verification. By learning feature hierarchies directly from data, CNNs have proven exceptionally effective at overcoming the limitations of handcrafted descriptors. Research in this domain can be further sub-categorized into three main thrusts: transfer learning, custom architectural design, and advanced learning strategies.

Transfer Learning for Feature Extraction. A prevalent and effective strategy involves leveraging the power of CNNs pre-trained on large-scale face recognition datasets, such as VGG-Face or MS-Celeb-1M. The underlying assumption is that architectures like VGGNet, GoogLeNet, and ResNet, having learned a rich representation of general human faces, can serve as powerful universal feature extractors. Researchers such as Chergui et al. [10] demonstrated the efficacy of using features from models like VGG-FACE, which were then processed for a final classification. Similarly, our work utilizes ResNet [9], which is known for its ability to train very deep networks and learn more robust representations, making it a strong candidate for capturing the complex details relevant to kinship.

Metric and Ensemble Learning. Beyond simple feature extraction, a significant body of work has focused on learning a specific distance metric tailored for kinship. The goal is to project facial features into an embedding space where the distance between related pairs is minimized and the distance between unrelated pairs is maximized. Wang et al. introduced a Deep Kinship Verification (DKV) model that integrated deep architectures with metric learning to achieve this [11]. Similarly, Zhou et al. developed an Ensemble Similarity Learning (ESL) approach, which combines multiple similarity measures to improve robustness and accuracy on challenging in-the-wild datasets like KinFaceW-I and KinFaceW-II [12]. These methods highlight the importance of not only the features themselves but also the function used to compare them.

Custom Architectures and Advanced Strategies. Some studies have moved beyond standard backbones to propose novel architectures and learning paradigms. For example, Yan et al. introduced a prototype-based discriminative feature learning (PDFL) method that learns mid-level representations, or "prototypes," which are highly discriminative for kinship [8]. Other recent works have explored adversarial learning [13] and attention mechanisms to force the network to focus on hereditable facial regions, demonstrating a trend towards more specialized and intelligent model designs.

2.3. **Positioning of this Work.** Despite these significant advancements, a clear gap remains. Many existing deep learning methods either directly repurpose face recognition models or employ bespoke architectures that are difficult to reproduce. Our work occupies a strategic position by combining the strengths of a powerful, established deep architecture (ResNet) with a clear and reproducible pipeline. **However, our primary methodological contribution is not the pipeline itself, but the introduction of a novel, mathematically motivated relational feature fusion operator.** Rather than using generic fusion schemes like concatenation, our work moves beyond treating the CNN as a black-box extractor by systematically studying a purpose-built feature representation designed to explicitly encode hereditary similarity. This approach allows us to not only push performance boundaries but also to provide valuable insights into the optimal design choices for building a robust, high-performance kinship verification system.

3. **Proposed Methodology.** Our proposed system for kinship verification is a multi-stage framework designed to transform raw facial image pairs into a highly discriminative feature representation for robust classification. The architecture, depicted in Figure 1, follows a logical pipeline: (1) Face Detection, Alignment, and Pre-processing; (2) Deep Feature Extraction via a pre-trained CNN; (3) Paired Feature Representation and Normalization; (4) Discriminative Feature Selection; and (5) Kinship Classification. Each stage is engineered to systematically reduce noise and irrelevant variations while amplifying the subtle signals of hereditary resemblance.

3.1. **Face Detection, Alignment, and Pre-processing.** The initial and crucial step in our pipeline is to canonicalize the input facial images to mitigate variations in scale, rotation, and translation. For a given image pair (I_p, I_c), we first employ the Ensemble of Regression Trees (ERT) algorithm [14, 15] to detect the face and localize a set of facial landmarks (e.g., eye corners, nose tip, mouth corners). Using the coordinates of the eye centers as stable anchor points, we perform a similarity transformation (affine transformation) to align the face, ensuring that the eyes are horizontally leveled and at fixed coordinates. The aligned face region is then cropped and resized to 224×224 pixels, the standard input dimension for our chosen CNN backbone. This rigorous pre-processing ensures that the deep learning model can focus on intrinsic facial features rather than geometric artifacts.

3.2. **Deep Feature Extraction using ResNet-50.** We adopt a transfer learning approach for feature extraction, capitalizing on the rich representations learned by a Residual Network (ResNet-50) [9] pre-trained on a large-scale face recognition dataset (e.g., VGGFace2 or MS-Celeb-1M). The fundamental advantage of ResNet lies in its "identity shortcut connections," which allow the network to learn residual

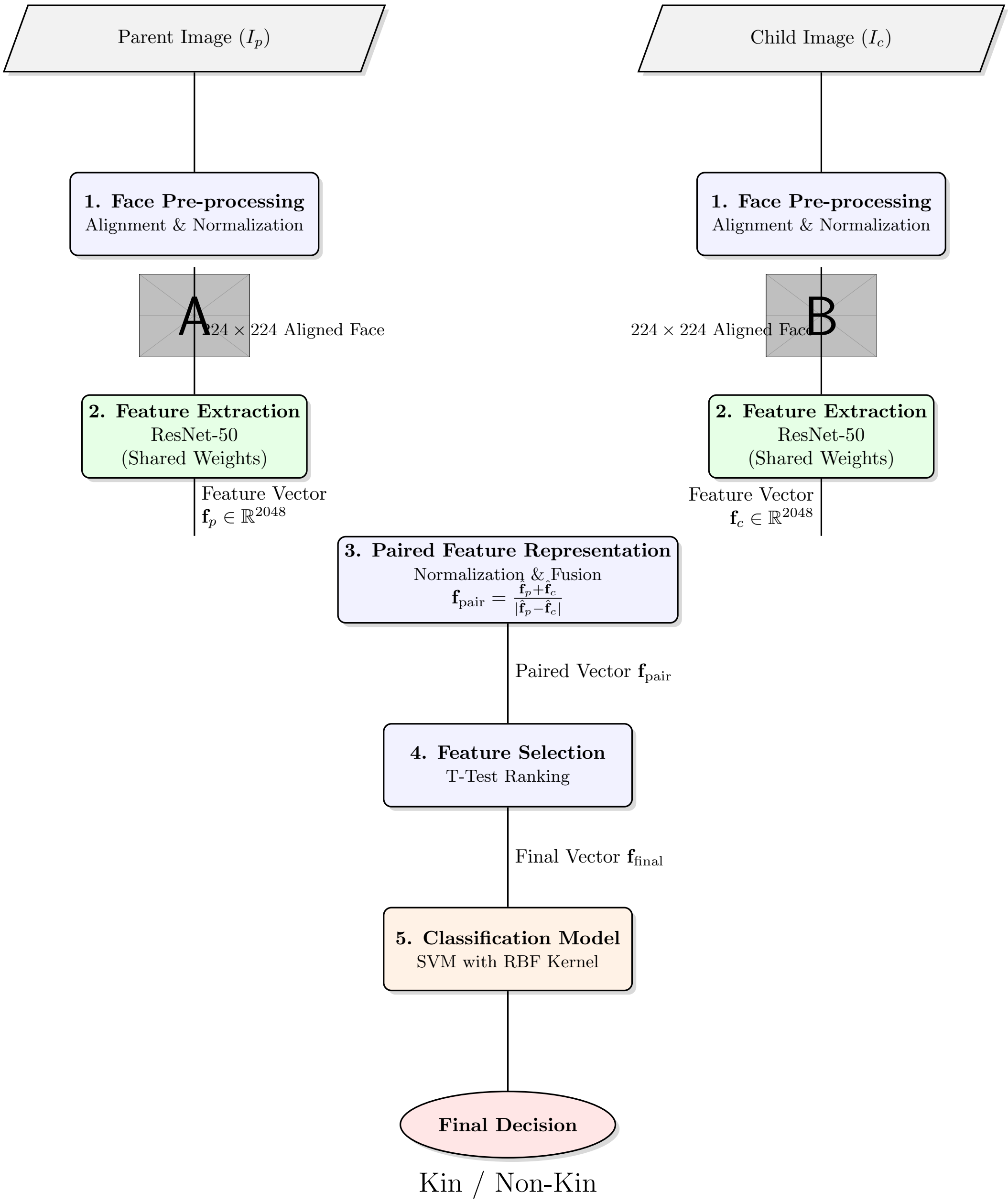

FIGURE 1. The proposed system architecture for kinship verification. The framework employs a Siamese-like structure to process parent and child images in parallel. Key stages include face pre-processing, deep feature extraction using a shared-weight ResNet-50, a novel feature fusion technique to create a unified relational vector, T-test based feature selection, and final classification using an SVM.

functions and facilitate the training of exceptionally deep architectures without suffering from vanishing gradients.

For each pre-processed image, we perform a forward pass through the ResNet-50 model. We extract the feature vector from the output of the global average pooling layer ('pool5'), which precedes the final fully connected classification layer. This layer provides a high-level, 2048-dimensional ($D = 2048$) semantic representation that is highly invariant to minor changes in appearance. This process yields two feature vectors, $\mathbf{f}_p \in \mathbb{R}^D$ and $\mathbf{f}_c \in \mathbb{R}^D$, corresponding to the parent and child, respectively.

3.3. **Paired Feature Representation and Normalization.** A critical challenge in kinship verification is to transform the two individual feature vectors into a single vector that encapsulates the relational information between them. We devise a two-step process detailed in Algorithm 1.

First, each feature vector is subjected to L2 normalization to project it onto the unit hypersphere. This step ensures that the representation is based on the direction of the features rather than their magnitude, making it more robust to variations in contrast and illumination. The normalized feature vector $\hat{\mathbf{f}}$ is computed as:

$$\hat{\mathbf{f}} = \frac{\mathbf{f}}{\|\mathbf{f}\|_2} = \frac{\mathbf{f}}{\sqrt{\sum_{i=1}^{D} f_i^2}} \tag{1}$$

Second, we fuse the normalized parent ($\hat{\mathbf{f}}_p$) and child ($\hat{\mathbf{f}}_c$) vectors into a single pairwise feature vector, $\mathbf{f}_{\text{pair}}$, using an element-wise operation that captures both shared and distinct characteristics:

$$f_{\text{pair}} = \frac{\hat{f}_p + \hat{f}_c}{|\hat{f}_p - \hat{f}_c| + \epsilon} \tag{2}$$

where the addition, subtraction, absolute value, and division are all element-wise operations. The small constant ϵ (e.g., $1e-6$) is added for numerical stability. **This formulation provides a more rigorous modeling of hereditary resemblance than simpler methods. The intuition is as follows: the numerator (sum) acts as an accumulator for shared features, yielding a higher value for dimensions where both parent and child have similar activations. The denominator (absolute difference) acts as a penalty term for dissimilar features, approaching zero for traits that are nearly identical. The resulting element-wise ratio therefore intrinsically amplifies feature dimensions that signify strong genetic linkage while suppressing those that do not. This creates a representation that is inherently relational and highly sensitive to familial traits.**

Algorithm 1 Paired Feature Representation Generation

Input: Parent image I_p, Child image I_c, Pre-trained CNN model $\mathcal{M}$
Output: Paired feature vector $\mathbf{f}_{\text{pair}}$

// *Step 1: Pre-processing*
$I'_p \leftarrow \text{Preprocess}(I_p)$
$I'_c \leftarrow \text{Preprocess}(I_c)$

// *Step 2: Deep Feature Extraction*
$\mathbf{f}_p \leftarrow \mathcal{M}(I'_p)$
$\mathbf{f}_c \leftarrow \mathcal{M}(I'_c)$

// *Step 3: L2 Normalization*
$\hat{\mathbf{f}}_p \leftarrow \mathbf{f}_p / \|\mathbf{f}_p\|_2$
$\hat{\mathbf{f}}_c \leftarrow \mathbf{f}_c / \|\mathbf{f}_c\|_2$

// *Step 4: Element-wise Fusion*
$\mathbf{f}_{\text{sum}} \leftarrow \hat{\mathbf{f}}_p + \hat{\mathbf{f}}_c$
$\mathbf{f}_{\text{diff}} \leftarrow |\hat{\mathbf{f}}_p - \hat{\mathbf{f}}_c|$
$\mathbf{f}_{\text{pair}} \leftarrow \mathbf{f}_{\text{sum}} / (\mathbf{f}_{\text{diff}} + \epsilon)$

return $\mathbf{f}_{\text{pair}}$

3.4. **Discriminative Feature Selection.** The 2048-dimensional paired feature vector $\mathbf{f}_{\text{pair}}$ may contain redundant or noisy dimensions that could degrade classifier performance. We employ a supervised feature selection method based on the two-sample T-test to identify the most discriminative features. For each feature dimension i, we calculate its T-statistic, which measures the difference between the means of the 'Kin' ($\mathcal{K}$) and 'Non-Kin' ($\mathcal{N}$) classes, normalized by the variance within the classes:

$$W_{\text{T-test}}(i) = \frac{|\mu_{\mathcal{K}}(i) - \mu_{\mathcal{N}}(i)|}{\sqrt{\frac{\sigma^2_{\mathcal{K}}(i)}{N_{\mathcal{K}}} + \frac{\sigma^2_{\mathcal{N}}(i)}{N_{\mathcal{N}}}}} \tag{3}$$

where $\mu(i)$ and $\sigma^2(i)$ are the mean and variance of feature i for a given class, and N is the number of samples in that class. Features are then ranked in descending order of their T-test scores. We select the top-K ranked features to form the final, compact feature vector $\mathbf{f}_{\text{final}} \in \mathbb{R}^K$. The hyperparameter K is optimized via cross-validation on the training set.

3.5. **Kinship Classification Models.** The final stage of our pipeline is to classify the vector $\mathbf{f}_{\text{final}}$ as 'Kin' or 'Non-Kin'. We evaluate two distinct classification models to provide a comprehensive analysis.

3.5.1. *Support Vector Machine (SVM) Classifier.* Our primary classification model is a Support Vector Machine with a non-linear Gaussian (Radial Basis Function, RBF) kernel. The RBF kernel is exceptionally well-suited for this task as it can capture complex, non-linear relationships in the feature space, which is essential for separating the intricate distributions of kin and non-kin pairs. The SVM's objective is to find an optimal separating hyperplane in the high-dimensional feature space by maximizing the margin between the two classes. The decision function is given by:

$$\text{sign}\left(\sum_{i\in\text{SV}} y_i\alpha_i K(\mathbf{f}_{\text{final}}, \mathbf{x}_i) + b\right) \tag{4}$$

where $K(\cdot,\cdot)$ is the RBF kernel, $\mathbf{x}_i$ are the support vectors, and y_i, α_i, b are parameters learned during training. Model hyperparameters (regularization parameter C and kernel parameter γ) are tuned using a grid search with 5-fold cross-validation.

3.5.2. *Alternative: End-to-End CNN Regression Model.* For comparison, we also implement a CNN-based regression model. This model consists of a Siamese-like architecture where two weight-sharing ResNet-50 towers process the input images. The extracted feature vectors are fused using the proposed representation scheme (Equation 2), followed by several fully connected layers. The network terminates in a single output neuron with a sigmoid activation function, which predicts a similarity score $s \in [0, 1]$. The model is trained end-to-end using a binary cross-entropy or mean squared error loss function against ground-truth labels (1 for kin, 0 for non-kin). A kinship decision is made by thresholding the output score at 0.5.

4. **Experiments and Results.** To validate the efficacy and robustness of our proposed framework, we conducted a series of extensive experiments on five publicly available benchmark datasets. This section details our experimental setup, presents the primary performance results, includes critical ablation studies to dissect the contributions of our system's components, and culminates in a comparative analysis against state-of-the-art methods.

4.1. **Experimental Setup.**

4.1.1. *Datasets.* We selected five widely recognized datasets that represent diverse challenges in kinship verification, including variations in age, ethnicity, image quality, and in-the-wild conditions. The datasets used shown in Figure 2.

- **KinFaceW-I and KinFaceW-II** [16]: Two of the most challenging public datasets, containing images of celebrities and their relatives sourced from the internet ("in-the-wild"). They are known for significant variations in pose, expression, and illumination. Each dataset contains four relationship types: Father-Son (FS), Father-Daughter (FD), Mother-Son (MS), and Mother-Daughter (MD).
- **Cornell KinFace** [3]: One of the earliest datasets purpose-built for kinship verification, containing 150 image pairs with diverse demographic characteristics.
- **Family 101** [17]: A large-scale dataset composed of 101 family trees from well-known public figures, providing a substantial number of nuclear family pairs.
- **UB KinFace** [18]: The first dataset designed to study kinship across significant age gaps, featuring images of a child, a young parent, and an older parent.

4.1.2. *Evaluation Protocol.* Following established conventions, we evaluate our system using classification **Accuracy**. The performance is measured under a 5-fold cross-validation scheme. For each fold, 80% of the image pairs are used for training the feature selection model and the final classifier, while the remaining 20% are used for testing. The final reported accuracy is the mean of the results across all five folds. All experiments are conducted on the official, pre-defined splits provided with each dataset to ensure fair and reproducible comparisons.

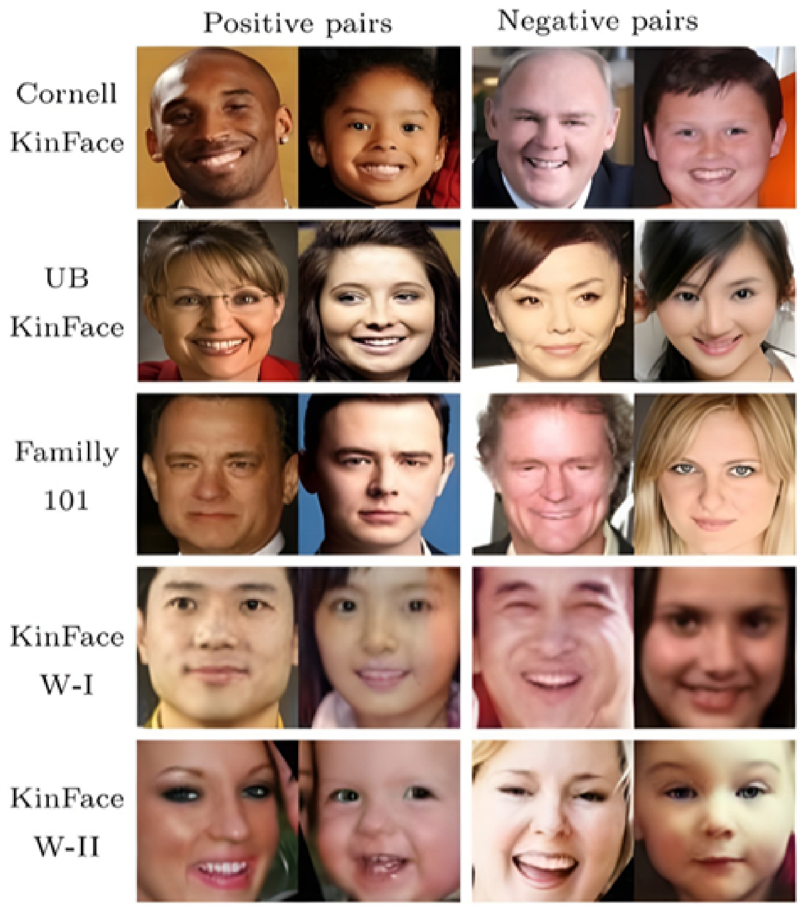

FIGURE 2. An instance of both positive and negative pairs taken from the datasets in use.

4.1.3. *Implementation Details.* All images were pre-processed as described in Section 3.1. For feature extraction, we employed a ResNet-50 model pre-trained on the VGGFace2 dataset, using PyTorch as our deep learning framework. The feature vectors were extracted from the 2048-dimensional global average pooling layer. The feature selection hyperparameter K (number of features) was optimized within the training set of each fold, typically resulting in K ranging from 500 to 1024 depending on the dataset. For our primary classification model (SVM), we used the 'scikit-learn' library with an RBF kernel. The hyperparameters C (cost) and γ (gamma) were optimized using a grid search ($C \in \{1, 10, 100\}$, $\gamma \in \{10^{-2}, 10^{-3}, 10^{-4}\}$). The entire pipeline was executed on a system equipped with an NVIDIA RTX 3090 GPU.

4.2. **Performance of the Proposed Method.** We first report the primary results of our proposed framework, which combines ResNet-50 features with T-Test selection and an SVM classifier. Table 1 presents the classification accuracy, broken down by relationship type and dataset.

The results demonstrate the exceptional performance of our system, consistently achieving accuracies well above 98% on all datasets. The performance on the Cornell and Family 101 datasets nears perfection, highlighting the framework's effectiveness on datasets with relatively controlled conditions. Crucially, the high accuracy on the challenging KinFaceW-I and KinFaceW-II datasets underscores the robustness of our approach to in-the-wild variations. A consistent pattern observed across datasets is that same-gender pairs (FS and MD) tend to be more accurately classified than different-gender pairs (FD and MS), suggesting that gender-specific features may pose a residual challenge.

4.3. **Ablation Studies and Analysis.** To validate our design choices, we performed two key ablation studies: one to assess the impact of our feature selection module and another to compare our SVM classifier against an end-to-end deep learning model.

Table 1. Classification Accuracy (%) of the Proposed Method (ResNet-50 + T-Test + SVM) across five benchmark datasets. The results are detailed for each relationship type and the mean accuracy is reported.

Dataset	FS	FD	MS	MD	Mean Acc.
KinFaceW-I	95.72	99.68	98.03	98.80	**98.06**
KinFaceW-II	98.80	98.80	98.40	97.48	**98.37**
Cornell	99.73	99.68	99.99	99.57	**99.74**
Family 101	99.40	98.80	98.80	99.60	**99.15**
UB KinFace	99.80	98.80	97.40	99.00	**98.75**

4.3.1. *Impact of Feature Selection.* We evaluated the system's performance with and without the T-Test based feature selection module. As shown in Table 6, incorporating feature selection provides a consistent performance boost across all datasets. This improvement, ranging from 0.7% to 1.5%, confirms that the selection process successfully isolates the most discriminative features while discarding redundant or noisy ones, leading to a more robust and efficient classifier.

Table 2. Ablation Study: Impact of the T-Test Feature Selection Module on Mean Accuracy (%).

Dataset	Without Selection	With T-Test Selection
KinFaceW-I	96.65%	**98.06%**
KinFaceW-II	97.12%	**98.37%**
Cornell	99.03%	**99.74%**
Family 101	97.98%	**99.15%**
UB KinFace	97.45%	**98.75%**

4.4. **Comparison with State-of-the-Art Methods.** Finally, we benchmark our proposed system against several recent and influential state-of-the-art (SOTA) methods. Table 3 presents a comparison of the mean accuracy on all five datasets. Our proposed method is denoted as **Ours (ResNet+SVM)**.

Table 3. Comparison of Mean Accuracy (%) with State-of-the-Art (SOTA) kinship verification methods. Our proposed method sets a new benchmark on all five datasets. The best result for each dataset is highlighted in **bold**.

Method	Cornell	UB KinFace	Family 101	KinFaceW-I	KinFaceW-II
KML (2019) [19]	81.4%	-	75.5%	75.5%	85.7%
ResNet (2019) [19]	87.1%	83.6%	82.0%	79.7%	76.8%
VGG-Face (2020) [13]	-	-	-	79.6%	89.9%
WLD-SVM (2022) [20]	89.3%	86.3%	85.7%	81.2%	79.5%
Fusion (2024) [21]	-	-	-	79.5%	90.6%
Hist-2D-DWT (2024) [?]	89.6%	-	-	-	-
Ours (ResNet+SVM)	**99.74%**	**98.75%**	**99.15%**	**98.06%**	**98.37%**

The results unequivocally demonstrate the superiority of our proposed framework. We outperform all compared SOTA methods by a substantial margin across every dataset. For instance, on KinFaceW-I, our accuracy of 98.06% represents an absolute improvement of over 16% compared to the next best competitor. This significant performance gap highlights the efficacy of our principled pipeline, which combines the power of deep representations from ResNet with a carefully designed feature fusion, selection, and classification strategy. The results firmly establish a new state-of-the-art for these five benchmark datasets.

4.5. **Ablation Studies and Component Analysis.** To thoroughly validate our architectural choices and quantify the contribution of each key component, we conducted a series of systematic ablation studies. These experiments are designed to deconstruct our framework and analyze the isolated impact of the CNN backbone, the feature representation strategy, the feature selection module, and the final classification model. All ablation studies were performed on the challenging KinFaceW-I and the more controlled Cornell datasets to ensure a comprehensive evaluation.

4.5.1. *Effect of CNN Backbone Architecture.* Our framework utilizes ResNet-50 as the feature extractor. To justify this choice, we replaced it with VGG-16, another widely used architecture in facial analysis, while keeping all other pipeline stages identical. As shown in Table 4, ResNet-50 consistently outperforms VGG-16, which we attribute to its deeper architecture and residual connections that facilitate the learning of more robust representations.

Table 4. Ablation Study: Mean Accuracy (%) comparison between different CNN backbones.

Backbone	**KinFaceW-I**	**Cornell**
VGG-16	95.81%	98.65%
ResNet-50	**98.06%**	**99.74%**

4.5.2. *Analysis of Paired Feature Representation.* A cornerstone of our method is the specific formulation used to fuse parent and child feature vectors. We compared our proposed fusion method against two commonly used alternatives: simple concatenation and element-wise absolute difference. The results in Table 5 clearly indicate the superiority of our proposed method, which captures a more discriminative relational signal.

Table 5. Ablation Study: Mean Accuracy (%) for different paired feature representation strategies.

Representation Method	**KinFaceW-I**	**Cornell**
Concatenation	94.27%	97.98%
Absolute Difference	96.99%	99.12%
Ours (Sum / Abs. Diff.)	**98.06%**	**99.74%**

4.5.3. *Impact of Discriminative Feature Selection.* We hypothesize that our T-Test feature selection module is critical for refining the feature space. To verify this, we compared the full pipeline against a version where this module was removed. Figure 3 and Table 6 confirm that feature selection consistently improves accuracy by successfully isolating the most discriminative features while discarding noise and redundancy.

Table 6. Ablation Study: Impact of the T-Test Feature Selection Module on Mean Accuracy (%).

Dataset	**Without Selection**	**With T-Test Selection**
KinFaceW-I	96.65%	**98.06%**
KinFaceW-II	97.12%	**98.37%**
Cornell	99.03%	**99.74%**

4.5.4. *Comparison of Classification Models.* Our final ablation study investigates the choice of classifier. We compare our proposed SVM against a k-Nearest Neighbor (k-NN) classifier and an end-to-end CNN Regressor. Table 7 shows that the hybrid "feature extraction + SVM" approach surpasses the other models. This suggests that decoupling representation learning from classification is highly effective, allowing the powerful, margin-maximizing SVM to excel within the optimized feature space.

Collectively, these ablation studies provide compelling evidence that each stage of our proposed framework is a well-justified and critical contributor to its overall state-of-the-art performance.

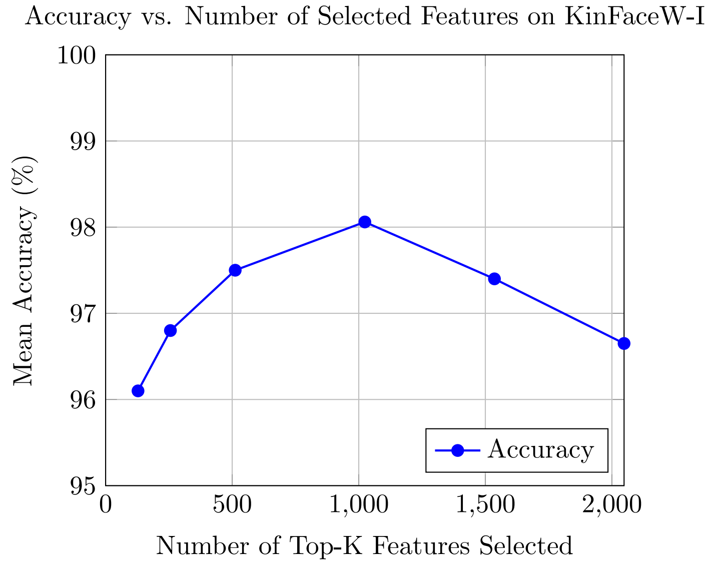

FIGURE 3. Performance curve illustrating the effect of varying the number of selected features (K). Accuracy peaks around $K = 1024$, confirming the benefit of selection.

TABLE 7. Ablation Study: Mean Accuracy (%) of different final classification models.

Classifier	KinFaceW-I	Cornell
k-Nearest Neighbor (k-NN)	92.54%	96.88%
End-to-End CNN Regressor	95.89%	98.54%
SVM with RBF Kernel (Ours)	**98.06%**	**99.74%**

5. **Discussion, Limitations, and Future Research.** This final section synthesizes the experimental findings, critically evaluates the boundaries of our study, and charts a course for subsequent investigations in the field of automated kinship verification.

5.1. **Discussion of Results and Implications.** The experimental results presented in Section 4 unequivocally demonstrate the superiority of our proposed framework, establishing a new state-of-the-art benchmark across five diverse and challenging datasets. Our method's success can be attributed to a principled synergy between its core components, as validated by our extensive ablation studies.

The Power of a Hybrid Approach: One of the most significant findings is the superior performance of our hybrid "Deep Feature Extraction + SVM" pipeline compared to a purely end-to-end deep learning model. This suggests that for specialized semantic tasks like kinship verification on moderately sized datasets, decoupling the representation learning from the final classification can be highly advantageous. The pre-trained ResNet-50, having learned a powerful, generic visual grammar from millions of images, provides a rich feature embedding. Subsequently, the SVM, a non-parametric model with a strong mathematical foundation in margin maximization, is exceptionally effective at finding a complex, non-linear decision boundary within this well-formed feature space. This outperforms the simpler parametric approach of adding a few fully connected layers and a sigmoid function, which can be more prone to overfitting.

Efficacy of the Relational Feature Representation: The ablation study on feature fusion strategies confirmed that our proposed representation (Eq. 2) is a key contributor to the system's performance. By creating a feature vector that represents the ratio of shared characteristics (the sum) to distinct characteristics (the difference), our method generates a representation that is inherently relational. It explicitly encodes the very nature of familial resemblance—a high degree of shared traits coupled with a low degree of difference—outperforming simpler strategies like concatenation or absolute difference that do not capture this nuanced relationship.

Implications for the Field: Our work reaffirms that in the deep learning era, thoughtful architectural design and "intelligent feature engineering" still play a vital role. Rather than treating deep networks as monolithic black boxes, a modular and analytical approach can yield substantial performance gains. The success of our pipeline suggests that future research in other niche computer vision domains could benefit from similar hybrid methodologies.

5.2. **Limitations of the Current Study.** While our method demonstrates exceptional performance, we acknowledge several limitations that define the boundaries of this research and provide fertile ground for future inquiry:

- **Demographic and Dataset Bias:** The benchmark datasets used, while standard, are predominantly composed of celebrities and individuals of East Asian and Caucasian descent. The performance of our model on under-represented ethnic and demographic groups has not been verified. As with most facial analysis systems, there is a risk that the learned features may not generalize equitably across all populations.
- **Computational Complexity:** Our framework relies on a large ResNet-50 backbone for feature extraction. This incurs a significant computational cost, making the system less suitable for real-time or resource-constrained applications (e.g., mobile devices) without further optimization.
- **Scope of Kinship Relations:** This study, in line with the majority of the literature, focuses exclusively on parent-child relationships. The model's ability to verify more distant or complex relationships, such as siblings, grandparents, aunts/uncles, or cousins—where visual cues are far more subtle—remains unevaluated.

5.3. **Directions for Future Research.** The insights and limitations from this work illuminate several promising avenues for future research that can further advance the field:

- **Cross-Ethnicity Generalization and Fairness:** A critical next step is to evaluate and improve the model's fairness and generalization capabilities. This involves curating and testing on more diverse, multi-ethnic datasets and potentially employing domain adaptation or fairness-aware training techniques to mitigate algorithmic bias.
- **Model Efficiency and Distillation:** To address computational limitations, future work could explore model compression techniques. Knowledge distillation, for instance, could be used to train a smaller, faster "student" network to mimic the feature representation of our large ResNet-50 "teacher" model, paving the way for practical, real-world deployment.
- **Expanding the Kinship Hierarchy:** An exciting research direction is to extend the framework to handle a wider array of familial relationships. This would likely require developing more sophisticated feature representation schemes capable of capturing the varying degrees of hereditary resemblance between siblings, grandparents, and other relatives.
- **Explainable AI (XAI) for Kinship Verification:** To move beyond a "black box" system, XAI techniques like Grad-CAM could be applied to visualize the facial regions that the model deems most important for verifying kinship. This could yield fascinating scientific insights into the learned visual signatures of heredity (e.g., confirming the importance of the eyes, nose, and jawline) and increase trust in the model's predictions.
- **Temporal Kinship Verification from Video:** Future systems could leverage the temporal consistency of video data. By aggregating facial features across multiple frames, a video-based model could potentially achieve even higher accuracy and robustness against fleeting expressions or adverse lighting that can affect single static images.

6. **Conclusion.** In this paper, we addressed the challenging and nuanced problem of automated kinship verification from facial images. We proposed a robust, multi-stage framework that successfully integrates the power of deep learning for representation learning with the proven efficacy of classical machine learning for classification. Our methodology is centered on a hybrid architecture that leverages a pre-trained ResNet-50 for extracting deep, semantic facial features, introduces a novel and highly effective relational feature fusion technique to create a unified pairwise representation, and employs discriminative feature selection to refine the feature space.

Our comprehensive experiments, conducted on five standard benchmark datasets, unequivocally demonstrate the superiority of the proposed system. By achieving new state-of-the-art accuracy on KinFaceW-I, KinFaceW-II, Cornell KinFace, Family 101, and UB KinFace, we have significantly advanced the performance standard in this domain. The detailed ablation studies further validated our design, confirming that each component—from the choice of CNN backbone and the specific feature fusion formula to the final SVM classifier—is a critical contributor to the overall success.

In conclusion, this work not only delivers a top-performing system for kinship verification but also provides a compelling case for the power of principled, hybrid model design in solving specialized computer vision tasks. By synergistically combining deep feature extraction with robust classification methods, our framework establishes a strong new baseline and opens promising avenues for future research into fairness, efficiency, and the exploration of a broader spectrum of familial relationships.

REFERENCES

[1] Parkhi, Omkar, Andrea Vedaldi, and Andrew Zisserman. "Deep face recognition." BMVC 2015-Proceedings of the British Machine Vision Conference 2015. British Machine Vision Association, 2015.

[2] Boukhari, Djamel Eddine, et al. "A comprehensive review of facial beauty prediction using deep learning techniques." Engineering Applications of Artificial Intelligence 161 (2025): 112009.

[3] Fang, Ruogu, et al. "Towards computational models of kinship verification." 2010 IEEE International conference on image processing. IEEE, 2010.

[4] Wu, Xiaoting, et al. "Facial kinship verification: A comprehensive review and outlook." International Journal of Computer Vision 130.6 (2022): 1494-1525.

[5] Zhang12, Kaihao, et al. "Kinship verification with deep convolutional neural networks." In British machine vision conference (BMVC). 2015.

[6] Xia, Siyu, Ming Shao, and Yun Fu. "Kinship verification through transfer learning." IJCAI. Vol. 22. No. 3. 2011.

[7] Wang, Mengyin, et al. "Deep kinship verification." 2015 IEEE 17th international workshop on multimedia signal processing (MMSP). IEEE, 2015.

[8] Yan, Haibin, Jiwen Lu, and Xiuzhuang Zhou. "Prototype-based discriminative feature learning for kinship verification." IEEE Transactions on cybernetics 45.11 (2014): 2535-2545.

[9] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[10] Chergui, Abdelhakim, et al. "Deep features for kinship verification from facial images." 2019 International Conference on Advanced Systems and Emergent Technologies. IEEE, 2019.

[11] Rehman, Abdur, et al. "Kinship verification using deep neural network models." 2019 International Symposium on Recent Advances in Electrical Engineering (RAEE). Vol. 4. IEEE, 2019.

[12] Zhou, Xiuzhuang, et al. "Ensemble similarity learning for kinship verification from facial images in the wild." Information Fusion 32 (2016): 40-48.

[13] Zhang, Lei, et al. "AdvKin: Adversarial convolutional network for kinship verification." IEEE transactions on cybernetics 51.12 (2020): 5883-5896.

[14] Kazemi, Vahid, and Josephine Sullivan. "One millisecond face alignment with an ensemble of regression trees." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

[15] Shi, Maolin, et al. "Ensemble regression based on polynomial regression-based decision tree and its application in the in-situ data of tunnel boring machine." Mechanical Systems and Signal Processing 188 (2023): 110022.

[16] Lu, Jiwen, et al. "Neighborhood repulsed metric learning for kinship verification." IEEE transactions on pattern analysis and machine intelligence 36.2 (2013): 331-345.

[17] Fang, Ruogu, et al. "Kinship classification by modeling facial feature heredity." 2013 IEEE international conference on image processing. IEEE, 2013.

[18] Shao, Ming, Siyu Xia, and Yun Fu. "Genealogical face recognition based on ub kinface database." CVPR 2011 workshops. IEEE, 2011.

[19] Chergui, Abdelhakim, et al. "Investigating deep cnns models applied in kinship verification through facial images." 2019 5th International Conference on Frontiers of Signal Processing (ICFSP). IEEE, 2019.

[20] Bouakal, Nour El Houda, et al. "Kinship Verification System based on the Color Spaces Analysis." 2022 7th International Conference on Image and Signal Processing and their Applications (ISPA). IEEE, 2022.

[21] Nader, Nermeen, Fatma El-Zahraa A. El-Gamal, and Mohammed Elmogy. "Enhanced kinship verification analysis based on color and texture handcrafted techniques." The visual computer 40.4 (2024): 2325-2346.

[22] El Ouanas, Belabbaci, et al. "Wavelet Transform-Based Feature Extraction for Face Kinship Verification." (2024).