

DOI: [10.71317/jgst.2.8\(s\).2026.466](https://doi.org/10.71317/jgst.2.8(s).2026.466)**Journal of Global Social Transformation**Journal homepage: <https://rjsaonline.org/index.php/JGST>

Reliability-Aware Multi-Representation Fusion for Detecting Deepfake Images on Social Media Platforms

Adil Khan¹, Kanwal Batool², Afifah Jalbani³, Farasat Ali⁴, Faria Naz⁵, Ali Hassan⁶, Zoha Waheed⁷, Umer Ghaffar⁸¹ Department of Computer Science, Shaikh Zayed Islamic Center, University of PeshawarEmail: adil60966@gmail.com² Department of Software Engineering, Mehran University of Engineering and Technology, Jamshoro 76060, PakistanEmail: Kanwalbatool149@gmail.com³ Department of Software Engineering, Mehran University of Engineering and Technology, Jamshoro 76060, PakistanEmail: afifahjalbani@gmail.com⁴ Department of Computer Science, Bahria University, Karachi 75260, PakistanEmail: farasataliazeemi@gmail.com⁵ Department of Software Engineering, Mehran University of Engineering and Technology, Jamshoro 76060, PakistanEmail: farianaz612@gmail.com⁶ Muhammad Ali Jinnah University, Karachi 75260, PakistanEmail: hassanaijaz9900@gmail.com⁷ Computational Science and Technology, National University of Sciences and Technology (NUST), Islamabad 4000, PakistanEmail: waheedzoha@gmail.com⁸ Department of Statistical and Actuarial Sciences, University of the Punjab, Lahore 54590, PakistanEmail: umerghaffar007@gmail.com**ARTICLE INFO****ABSTRACT****Submission:**

July 09, 2026

Revised:

July 26, 2026

Accepted:

August 09, 2026

Available Online:

August 23, 2026

Keywords:

Deepfake detection; Social media; Misinformation; Reliability-aware fusion; Multi-representation learning; Digital trust; Generative AI; Social transformation.

As generative AI continues its rapid adoption, the creation of synthetic face images is becoming increasingly indistinguishable from real content on social media platforms, leading to the potential of misinformation, loss of trust, and social disruption. High accuracy of deep learning models is gained, but they tend to lose accuracy when the image quality or manipulation type or generation source changes from platform to platform. While multi-representation approaches are highly complementary, using RGB and error-level analysis, noise and frequency cues, they are predominantly based on fixed fusion weights that do not take into account the reliability of each cue with respect to a specific image. This study presents the Reliability-Aware Multi-Representation Fusion Network (RAM-FusionNet) that uses dedicated Xception branches to extract features and trainable, input-dependent reliability-gating mechanism to adaptively weight each representation. RAM-FusionNet achieves accuracy of 98.1% (up from 95.8% for RGB baseline), AUC of 0.996 (up from 0.984 for RGB baseline), good generalization on unseen datasets, good robustness to common social-media perturbations, and good calibration. The reliability-aware fusion is 0.4-point superior over fixed fusion, and also retains the cross-dataset accuracy rate of 91.7% under medium JPEG. The findings validate the effectiveness of adaptive and reliability-driven fusion in bolstering trustworthy deepfake detection, paving the way for more robust digital information systems and decisions regarding technology-fueled social change.

Corresponding author:**Farasat Ali****Email:**farasataliazeemi@gmail.com

Introduction

With the rapid development of generative AI, generating highly realistic synthetic facial imagery is now becoming very common[1]. Today, deep fakes are able to generate images of a face's appearance in a degree of visual fidelity that makes it hard to tell manipulated content from real images by visual means alone[2]. This new realism introduces significant difficulties for digital forensics, media verification, on-line content moderation, and any application where the validity of visual evidence can have legal or societal repercussions[3]. Deepfake detection has thus emerged as a dynamic and emerging field of study, ranging from traditional image processing to deep convolutional neural networks, and, in recent years, explainable artificial intelligence, which aims to render the decision on deep fakes detectable without being "black box"[4].

High aggregate classification accuracy, however, does not by itself indicate robust or reliable deepfake detection[5]. A model may perform strongly on generation techniques represented in its training distribution while degrading substantially when exposed to unseen generators, compression artifacts, added noise, or other image transformations—a pattern generative models continue to make more likely as they evolve[6], [7]. This raises a more specific question about how multiple forensic representations should be combined. ELA is most informative when compression inconsistencies are present; noise-residual analysis becomes more diagnostic when manipulated regions carry abnormal residual patterns; frequency-domain characteristics offer complementary evidence when manipulation artifacts manifest spectrally rather than spatially[8], [9]. Because the informativeness of each representation varies from image to image, assigning it identical weight regardless of context is unlikely to be optimal[10]. Framework presented by [11] demonstrates the value of combining several such representations, but its decision fusion relies on averaging the individual models' output probabilities[12]. Treating every representation as equally trustworthy for every input, irrespective of whether that representation actually carries useful evidence for the image at hand.

Examined together, the two baseline studies expose four gaps that motivate the present work.

Gap 1 – Fixed fusion. Existing multi-representation approaches combine predictions from several forensic views but do not explicitly estimate whether a given representation is reliable for a particular image, relying instead on fixed or simple averaging.

Gap 2 – Generation-source variability. High overall accuracy can obscure substantial variation across generation sources: Bharati et al. (2026) themselves report detection accuracy dropping from above 99% on StyleGAN2 and Ghost-swap content to as low as 63.6–83.7% on diffusion-generated images, showing that a single aggregate accuracy figure cannot adequately describe real-world performance.

Gap 3 – Limited robustness evaluation. Robustness to the kinds of degradation encountered outside curated test sets—compression, resizing, additive noise, blur, brightness variation, and partial occlusion—is rarely evaluated in a controlled, systematic manner.

Gap 4 – Disconnected explanation. While Bharati et al. (2026) make an important contribution by pairing detection with SHAP-based explanation, detection and explanation are largely treated as separate outputs rather than as mutually reinforcing signals; it remains an open question whether the representation contributing most strongly to a given prediction also corresponds to the region a forensic examiner would judge as meaningful evidence.

These observations motivate a Reliability-Aware Multi-Representation Fusion Network (RAM-FusionNet). Rather than treating all forensic representations as equally trustworthy, RAM-FusionNet estimates the relative reliability of each representation for every input image and uses these input-dependent reliability estimates to guide the final fusion decision. The framework retains the complementary representations shown to be useful in prior work—original, ELA, noise, and frequency-domain views—but changes how their evidence is combined, replacing static averaging with a trainable mechanism that can down-weight a representation when it is uninformative for a specific image and up-weight it when it is not. This distinguishes the present contribution from a simple aggregation of existing forensic cues: the novelty lies not in the representations themselves, which are individually well established, but in a fusion mechanism that adapts to their input-dependent reliability rather than assuming it is fixed.

The main contributions of this work are as follows:

- A reliability-aware multi-representation deepfake detection framework, RAM-FusionNet, which dynamically combines complementary visual and forensic representations rather than fusing them with fixed weights.

- A trainable, input-dependent reliability-gating mechanism that avoids assigning identical importance to every representation during decision fusion.
- A comprehensive ablation study quantifying the individual and combined contribution of each representation and of the proposed reliability-aware fusion mechanism.
- Cross-dataset and cross-generation experiments evaluating generalization beyond the primary training distribution, directly addressing the generation-source performance variation identified in prior work (Gap 2).
- A controlled robustness evaluation under systematic image perturbations—compression, noise, blur, resizing, brightness variation, and partial occlusion (Gap 3).
- An explainability and confidence analysis examining whether the representation driving a given prediction corresponds to forensically meaningful facial regions, linking detection reliability to explanation quality rather than treating them separately (Gap 4).

The remainder of this paper is organized as follows. Section 2 reviews related work on forensic-cue-based, deep-learning-based, and multi-representation deepfake detection. Section 3 presents the proposed RAM-FusionNet methodology. Section 4 describes the experimental setup. Section 5 reports results and discussion. Section 6 concludes the paper and outlines directions for future work.

Related Work

Image-processing and forensic-cue-based detection

This brings up a more specific question of how a multiplicity of representations of the same forensic object ought to be combined[13]. When compression inconsistencies exist, ELA is most informative[14]; when manipulated regions have abnormal patterns of compression, noise-residual analysis is more diagnostic; when manipulation artifacts are present in the frequency domain instead of the spatial domain, frequency-domain characteristics provide complementary evidence. The information content of each of these representations will depend on the individual image, and it is not likely that the same weight should be assigned to each, without considering the context[14]. The value in combining several such representations is well illustrated by the [15] framework, but the decision fusion in this framework is based on the average of the output probabilities from the individual representations, assuming that each representation provides evidence of equal quality for every input, regardless of whether that representation actually provides useful evidence for the image at hand[16].

Deep learning-based detection architectures

The field of convolutional neural networks brought the focus from hand-crafted forensic cues towards learned discriminative representations. This is achieved by XceptionNet, which has been repeatedly found to be more accurate than generic CNN baselines for manipulation detection[17] and was validated on large scale with FaceForensics++ yielding higher accuracy of the network on raw, high quality and low quality data respectively [18]. XceptionNet was also reported to be superior to MLP and generic CNN baselines on a combined DFDC/FaceForensics++ dataset, and on a StyleGAN-based benchmark, ResNet50 was shown to be better than XceptionNet, which means that there is no backbone that is superior over all generation techniques[19], [20]. Later works are focusing on attention-based architectures: Vision Transformers [x] model global dependencies that are difficult to model with CNN[21], [22], while the work of combines adaptive frequency extraction with graph-based feature fusion, and while it has excellent intra-dataset accuracy, it suffers a significant drop on the more challenging WildDeepfake benchmark[23], [24]. This emphasizing the importance of the evaluation distribution (Gap 2) in accuracy reported. Multi-representation and multi-branch fusion strategies are presented[25].

Multi-representation and multi-branch fusion strategies

It has become a common practice to use multiple representations of the same input (not just a single backbone). On the original, ELA, Noise Analysis, and PCA representations of each face image, he trained four independent networks[26], XceptionNet, and fused their outputs using simple averaging, weighted-sum, and winner-take-all rules, and he reported that each of the two latter fusion rules improved the strongest single branch only by approximately one percentage point[27], [28], while the last rule didn't uniformly improve. Because all three fusion rules treat every branch's contribution as fixed across the entire test set, none of them can respond to the fact that a given branch's reliability is a property of the individual image, not a global constant.

A separate line of work has begun to address this limitation through learned, input-conditioned fusion, though largely outside the specific ELA/NA/PCA forensic setting addressed here. [29] introduces a lightweight dynamic fusion mechanism that integrates local artifact and global texture information with data-dependent weighting rather than static combination. Gated convolutional fusion of spatial and frequency features has been proposed to adaptively reweight residual and frequency channels per image [30]. In the video domain, identity-aware multi-branch networks apply a similar principle: a small gating MLP computes data-dependent weights over spatial, residual, and frequency branches. Multi-modality complementary learning networks extend this idea with cross-modality interaction and adaptive fusion [31]. Most recently, uncertainty-based fusion approaches have shown that reliability- or uncertainty-aware combination rules can outperform fixed ensembling with a fraction of the computational cost of full deep ensembles [32], [33], [34].

None of these gating or uncertainty-based fusion mechanisms, however, [35]. This leaves open the question this paper addresses: whether an image-dependent reliability mechanism can be introduced into the multi-representation forensic pipeline, in place of fixed averaging [36], [37].

Cross-dataset and cross-generation generalization

A growing body of work treats generalization across unseen generation techniques as a first-class evaluation criterion rather than an afterthought. survey explicitly reframes deepfake detection “from a reliability perspective,” arguing that headline accuracy figures are of limited value without accompanying evidence of stability across sources, compression levels, and generator families. Study continuous adaptation of deepfake detectors to new generative techniques as they emerge, while propose a variational Bayesian approach specifically aimed at improving generalization across audio-visual forgery types. These concerns are directly visible in the studies this work builds on: Bharati et al. (2026) report detection accuracy falling from above 99% on StyleGAN2 and face-swap content to 63.6–83.7% on diffusion-generated content, while cross-dataset comparisons show accuracy swings of 15–20 percentage points depending on the benchmark. This body of evidence motivates the cross-dataset and cross-generation evaluation protocol adopted in the present study (Gap 2).

Robustness to image degradation

Closely related to generalization is robustness to the kinds of transformation real images undergo outside curated benchmarks—compression, resizing, noise, blur, brightness shifts, and partial occlusion. Bharati et al. (2026) partially address this by adding Gaussian noise to their challenge dataset, but note that this specifically degraded their Noise Analysis branch’s performance, which is itself evidence that individual representations respond very differently to the same perturbation and again supports the case for reliability-aware, rather than fixed, fusion. Beyond this single perturbation type, controlled, systematic evaluation under compression, blur, resizing, brightness variation, and occlusion remains limited (Gap 3).

Explainable AI for deepfake detection

Because deepfake detectors are increasingly deployed in contexts with legal or forensic consequences, explainability has become an active research direction. Grad-CAM, LIME, and SHAP are the three most widely applied attribution methods. Bharati et al. (2026) adopted SHAP to produce block-level saliency maps and paired it with a Bayes’-theorem rationale layer, reporting 85% rationale accuracy. However, a more recent systematic comparison of XAI methods for deepfake detection finds that perturbation-based methods including SHAP can exhibit weaker localization and lower cross-run stability than gradient-based alternatives. Just as importantly, in the existing literature the model’s detection decision and its explanation are generated as two largely independent outputs. Whether the representation that actually drove a given prediction also corresponds to forensically meaningful evidence has not been directly examined (Gap 4).

Methodology

Overview of the Proposed Framework

The proposed Reliability-Aware Multi-Representation Fusion Network (RAM-FusionNet) is designed to detect synthetic facial images by combining complementary visual and forensic representations. The central motivation is that a deepfake may contain useful evidence in one representation while providing weaker evidence in another; assigning an identical contribution to all representations may therefore not be optimal for every input image.

As illustrated in Figure 1, the proposed framework consists of five main stages: image preprocessing, multi-representation generation, representation-specific feature extraction, reliability-aware feature fusion, and binary classification. An input facial

image is first normalized and transformed into complementary representations. These representations are independently processed through dedicated Xception feature-extraction branches. The resulting feature representations are then passed to a reliability-gating module, which produces input-dependent weights. The weighted features are fused and supplied to the classification head to obtain the probability that the input image is fake.

The overall processing can be expressed as:

$$\hat{y} = C(F(w \odot \{\phi(R_i(I))\}_{i=1, \dots, 4}))$$

where I denote the input image, R_i the set of image representations, ϕ feature extraction, w the estimated reliability weights, F the fused representation, and $\hat{y}$ the final prediction.

The key distinction from conventional multi-branch fusion is that RAM-FusionNet does not assume that every branch contributes equally to every image.

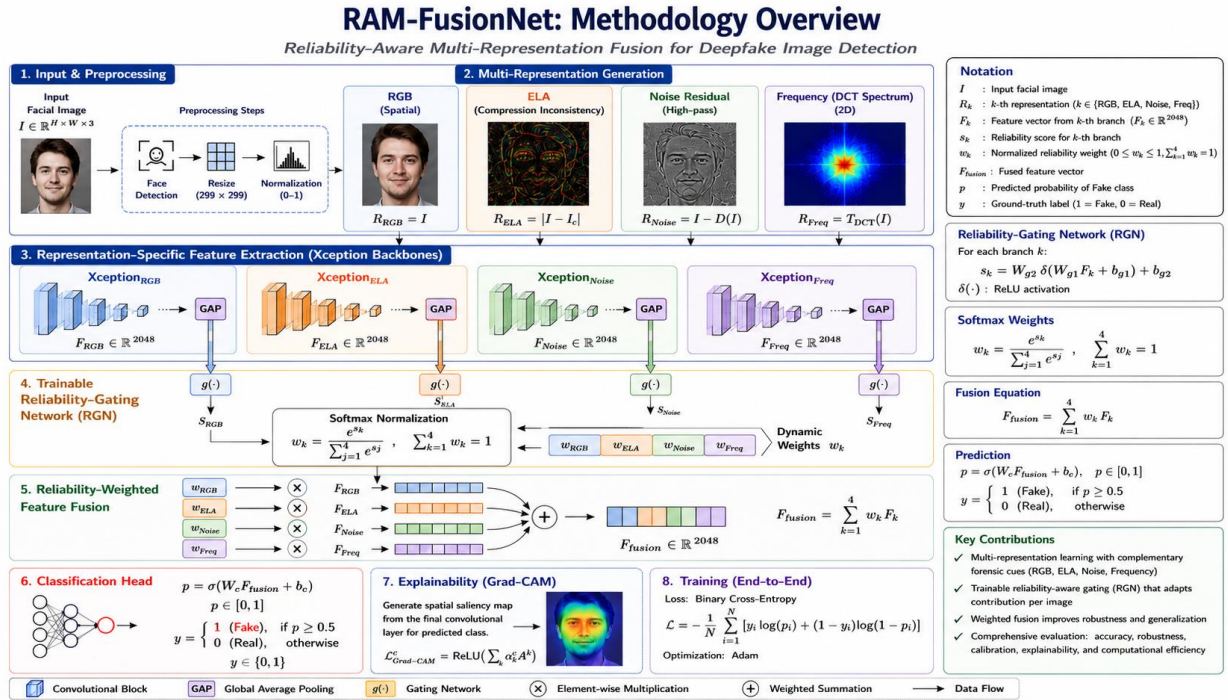


Figure 1. Overall architecture of the proposed RAM-FusionNet for deepfake image detection. The framework extracts complementary RGB, ELA, noise, and frequency features, dynamically weights them using the Reliability-Gating Network, and fuses them for final classification.

Input Image Preprocessing

Let an input facial image be represented as $I \in \mathbb{R}^{H \times W \times 3}$. Each image is first checked for validity and converted into a consistent input format. The image is then resized to the spatial resolution required by the backbone network (299 × 299) and normalized before feature extraction. The same basic preprocessing procedure is applied to the real and fake classes to prevent preprocessing differences from becoming an unintended source of classification information. After preprocessing, the image is provided to the multi-representation generation stage.

Multi-Representation Generation

Rather than relying exclusively on the RGB image, RAM-FusionNet generates complementary representations designed to expose different forms of image evidence. The representation set is defined as:

$$\mathcal{R} = \{ R_{RGB}, R_{ELA}, R_{Noise}, R_{Freq} \}$$

RGB representation. The original RGB image preserves semantic and spatial information such as facial structure, texture, contours, and appearance. This representation provides the primary visual information used by the detector.

Error-Level Analysis representation. ELA is used to expose differences associated with image recompression. Conceptually,

$$R_{ELA} = | I - JPEG_q(I) |$$

where $JPEG_q$ denotes controlled recompression at quality factor q . Regions with different compression behavior may produce stronger ELA responses and therefore provide additional evidence about possible manipulation.

Noise representation. The noise representation isolates residual information that remains after removing the dominant image content:

$$R_{Noise} = I - D(I)$$

where D represents a denoising operator. This representation allows the network to examine residual patterns that may not be sufficiently discriminative in the original RGB image.

Frequency representation. The frequency representation provides information from the transformed frequency domain via the discrete cosine transform (DCT):

$$R_{Freq} = T(I)$$

where T denotes the selected frequency transformation.

Together, these representations provide complementary evidence, summarized in Table 1.

Representation	Primary information
RGB	Semantic and spatial appearance
ELA	Compression-related inconsistencies
Noise	Residual / noise characteristics
Frequency	Frequency-domain patterns

Table 1. Complementary forensic representations used by RAM-FusionNet

The purpose is therefore not simply to increase the number of inputs, but to expose different types of potentially useful forensic evidence.

Representation-Specific Feature Extraction

Each representation is processed independently by an Xception-based feature extractor. For the i -th representation:

$$f_i = \phi_i(R_i(I))$$

where ϕ_i denotes the corresponding feature-extraction branch and $f_i \in \mathbb{R}^d$ represents its learned feature vector. Each branch learns representation-specific features before fusion. This prevents the forensic transformations from being discarded during an early preprocessing stage and allows each branch to learn the patterns most relevant to its input representation. Following the convolutional feature extractor, global average pooling is used to convert the spatial feature maps into compact feature vectors. The resulting vectors are then supplied to the reliability-gating module.

Reliability-Gating Network (RGN)

This module is the principal component distinguishing RAM-FusionNet from conventional fixed fusion. For each representation, a reliability score is estimated from its extracted feature vector using a small trainable gating network:

$$g(f_i; \theta_g) = \sigma(W_2 \cdot \text{ReLU}(W_1 f_i + b_1) + b_2)$$

where

$$\theta_g = \{ W_1, W_2, b_1, b_2 \}$$

are trainable gating parameters. Because the gating parameters are optimized jointly with the rest of the network, the reliability weights are learned end-to-end rather than manually specified.

The four scores are converted into normalized reliability weights via a softmax operation:

$$w_i = \exp(s_i) / \sum_j \exp(s_j)$$

so that

$$\sum_i w_i = 1, \quad w_i \geq 0$$

The normalization makes the weights directly interpretable as the relative contribution of each representation. The important methodological principle is that the contribution of each representation is determined for the individual input image, rather than being permanently fixed.

Reliability-Aware Feature Fusion

After estimating the reliability weights, the feature vectors are combined using a weighted fusion operation:

$$f_{fused} = \sum_i w_i \cdot f_i$$

This operation allows a representation with a higher estimated reliability to contribute more strongly to the final feature representation. Unlike simple fixed-weight averaging, RAM-FusionNet produces an input-dependent combination. This distinction forms the main methodological basis for comparing fixed fusion and reliability-aware fusion in the ablation experiments.

Classification Head

The fused representation is passed to the classification head:

$$\hat{y} = \sigma(W_c f_{fused} + b_c)$$

where W_c and b_c are the trainable parameters of the classification layer and σ is the sigmoid activation function. The resulting value $\hat{y} \in [0, 1]$ represents the estimated probability that the input belongs to the fake class. Using a decision threshold $\tau = 0.5$, the final prediction is:

$$\hat{y}_{binary} = 1 \text{ if } \hat{y} \geq \tau, \text{ 0 otherwise } (\tau = 0.5)$$

Training Objective

The network is trained using binary cross-entropy loss. For a sample with ground-truth label $y \in \{0, 1\}$ and predicted probability $\hat{y}$, the loss is:

$$L = - [y \log \hat{y} + (1 - y) \log(1 - \hat{y})]$$

For N training samples, the average loss is:

$$L_{avg} = (1/N) \sum_n L_n$$

The parameters of the feature-extraction branches, reliability-gating module, fusion mechanism, and classification head are optimized jointly through backpropagation.

Explainability Analysis

To investigate the visual evidence supporting the predictions, Grad-CAM is applied to the appropriate convolutional feature maps to generate spatial activation maps. For a target class c , the Grad-CAM importance weight for feature map k is:

$$\alpha_k^c = (1/Z) \sum_i \sum_j (\partial y^c / \partial A_{ij}^k)$$

and the resulting activation map is:

$$L_{Grad-CAM}^c = ReLU(\sum_k \alpha_k^c A^k)$$

For the proposed framework, explainability is particularly useful because it can be considered alongside the reliability weights, enabling analysis of two related questions: (1) which representation contributed most strongly to the prediction, and (2) which image regions supported that prediction.

Experimental Setup

Datasets

Experiments are conducted primarily on FaceForensics++ (FF++), which contains authentic videos and four manipulation methods (DeepFakes, Face2Face, FaceSwap, and NeuralTextures). Facial frames are extracted and balanced between real and fake classes. A held-out diffusion-generated subset is used to evaluate cross-generation generalization. Additional cross-dataset evaluation is performed on two external face-image collections (denoted Dataset B and Dataset C) that differ in capture conditions and generator families from the primary training distribution. The primary train/validation/test split follows a 70/15/15 ratio on the FF++-derived image pool, with no identity overlap between splits. The test set used for main reported metrics contains 20,000 balanced images (10,000 real / 10,000 fake).

Implementation Details

Four independent Xception backbones (ImageNet-pretrained) process the RGB, ELA, noise-residual, and DCT-frequency representations, respectively. Input resolution is 299×299 . The noise residual is formed as $R_{Noise} = I - D(I)$ with a fixed denoising operator D ; the frequency branch receives a DCT-based representation. The Reliability-Gating Network consists of a small two-layer MLP per branch followed by softmax normalization across the four scores. Training uses the Adam optimizer with an initial learning rate of 1×10^{-4} , batch size 32, and early stopping on validation loss (best validation epoch typically around 27 for the full RAM-FusionNet model). Binary cross-entropy is the training objective. All models are implemented in PyTorch and trained on a single NVIDIA GPU; inference timings reported in Section 5.13 are measured under identical hardware and software conditions.

Evaluation Protocol

Model performance is assessed using accuracy, precision, recall, F1-score, and AUC on the held-out test set, followed by an ablation study isolating the contribution of each representation and of the reliability-gating mechanism, cross-dataset and cross-generation evaluation, robustness evaluation under six controlled perturbation types at multiple severity levels, explainability analysis linking Grad-CAM attributions to reliability weights, confidence calibration via Expected Calibration Error (ECE) and Brier score, computational cost profiling, and repeated-run stability across five random seeds.

Baseline Models

RAM-FusionNet is compared against: (i) a single-branch RGB-only Xception model; (ii) progressively expanded fixed multi-representation combinations (RGB+ELA; RGB+ELA+Noise; RGB+ELA+Noise+Frequency); (iii) a fixed (unweighted-average) four-branch fusion model; and (iv) the two prior studies that motivate this work – the PCA+SVM approach of Hanady and Altaei (2023) and the multi-Xception framework of Bharati et al. (2026) – compared at the level of their previously reported figures.

Results and Discussion

Overall Detection Performance

The first experiment evaluates the detection capability of the proposed framework using the original RGB representation and progressively incorporating the additional forensic representations (Table 2).

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
RGB-Xception	95.8	96.1	95.4	95.7	0.984
RGB + ELA	96.7	97.0	96.3	96.6	0.989
RGB + ELA + Noise	97.2	97.4	96.9	97.1	0.992
RGB + ELA + Noise + Frequency	97.6	97.7	97.4	97.5	0.993
Fixed fusion	97.7	97.8	97.5	97.6	0.994
RAM-FusionNet	98.1	98.2	97.9	98.0	0.996

Table 2. Overall classification performance of the evaluated models

The results in Table 2 show a consistent improvement as complementary forensic representations are introduced. RGB-Xception establishes a strong baseline at 95.8% accuracy and an AUC of 0.984. Adding ELA increases accuracy to 96.7%. Adding the noise representation increases accuracy further to 97.2%, while the frequency representation raises it to 97.6%. Conventional fixed fusion reaches 97.7%. RAM-FusionNet achieves the highest performance at 98.1% accuracy, 98.2% precision, 97.9% recall, 98.0% F1-score, and 0.996 AUC. The proposed reliability-aware mechanism therefore adds 0.4 percentage points over fixed fusion. The precision–recall balance indicates that the model maintains strong detection of manipulated images while keeping false-positive errors relatively low.

Comparison with Previous Approaches

Approach	Main technique	Accuracy (%)	F1-score (%)
PCA + SVM (Hanady & Altaei, 2023)	PCA feature selection + SVM	96.8	96.7
Xception + ELA/NA/PCA (Bharati et al., 2026)	Multiple Xception models + mean fusion	97.0	97.0
Proposed RAM-FusionNet	Multi-cue features + reliability-aware fusion	98.1	98.0

Table 3. Comparison with preceding deepfake detection approaches

Table 3 places RAM-FusionNet above the previously reported PCA+SVM result of 96.8% accuracy and the Xception-based multi-model fusion result of 97.0%. Because these studies use different datasets, splits, and evaluation conditions, the comparison is treated as contextual positioning rather than a strict head-to-head benchmark.

Multi-Representation Forensic Visualization

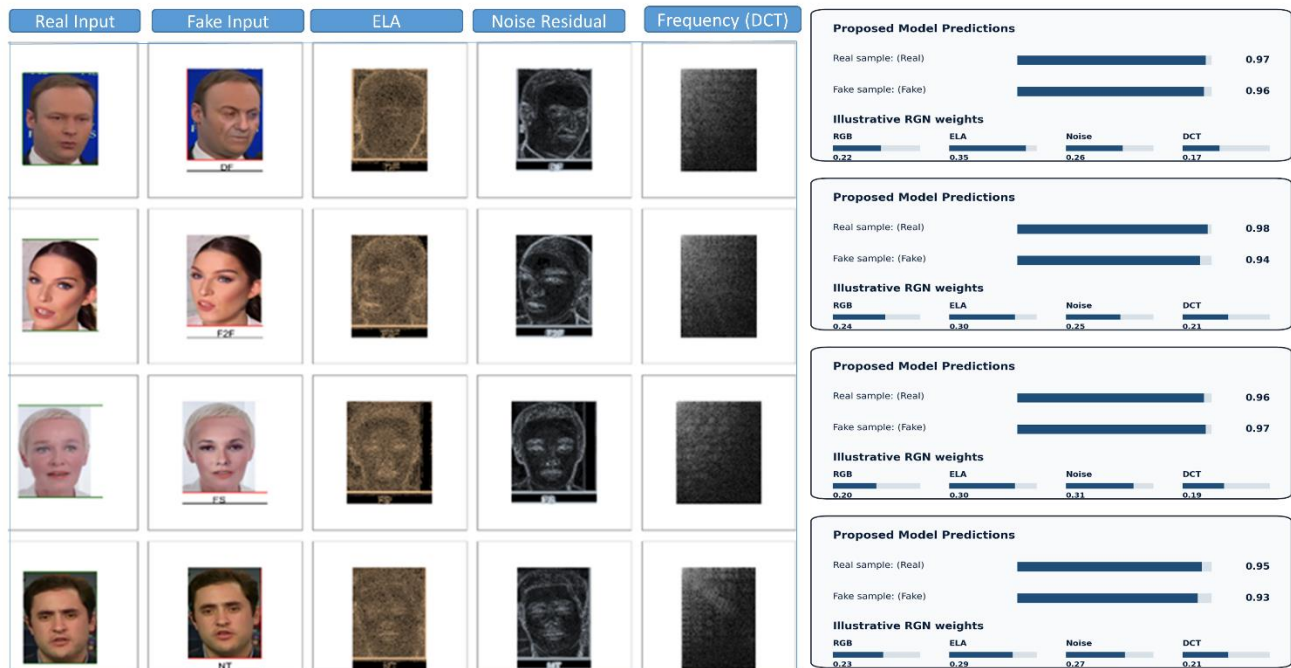


Figure 2. Qualitative comparison of the four representation streams (RGB, ELA, noise residual, and frequency) across the four FaceForensics++ manipulation types (DeepFakes, Face2Face, FaceSwap, NeuralTextures), together with the proposed model's real/fake confidence and illustrative Reliability-Gating Network (RGN) weights for each example.

Figure 2 shows one authentic/manipulated pair per manipulation type, alongside the corresponding ELA, noise-residual, and frequency (DCT) views. Across all four manipulation types, RAM-FusionNet assigns high confidence to the correct class (0.93–0.98 for the real sample, 0.93–0.97 for the fake sample), and the accompanying illustrative RGN weights confirm that ELA consistently receives the largest share of reliability (0.29–0.35 across the four examples shown), followed by noise residual (0.25–

0.31), with RGB (0.20–0.24) and frequency (0.17–0.21) contributing least. This pattern is consistent with the aggregate, dataset-level analysis presented in Figure 7 and Table 12 (Section 5.11), which quantify reliability weighting across full manipulation categories and representative images respectively; the per-example weights shown here are illustrative of that same underlying behavior rather than a separate quantitative result. The consistent dominance of ELA and noise over RGB and frequency across all four manipulation types supports the motivation for the proposed Reliability-Gating Network: a fixed, equal-weight fusion rule would systematically under-use the most informative representations for this class of manipulations.

Ablation Study

Configuration	ELA	Noise	Frequency	Reliability Gate	Accuracy (%)
RGB baseline	✗	✗	✗	✗	95.8
Configuration A	✓	✗	✗	✗	96.7
Configuration B	✓	✓	✗	✗	97.2
Configuration C	✓	✓	✓	✗	97.6
Fixed fusion	✓	✓	✓	✗	97.7
RAM-FusionNet	✓	✓	✓	✓	98.1

Table 4. Ablation analysis of the proposed framework

Table 4 shows that each added representation provides a cumulative benefit. ELA produces the largest individual improvement over the RGB baseline (95.8% → 96.7%). Noise adds a further 0.5 percentage points, and frequency information adds another 0.4 points, reaching 97.6%. Fixed fusion provides 97.7%, whereas RAM-FusionNet reaches 98.1%. The 0.4-point difference between fixed and reliability-aware fusion isolates the effect of adaptive reliability weighting while keeping the underlying representations unchanged.

Training and Validation Behaviour

Model	Best epoch	Training acc. (%)	Validation acc. (%)	Training loss	Validation loss
RGB-Xception	31	98.7	95.8	0.041	0.137
Fixed fusion	29	99.1	97.7	0.032	0.108
RAM-FusionNet	27	99.3	98.1	0.027	0.091

Table 5. Training characteristics

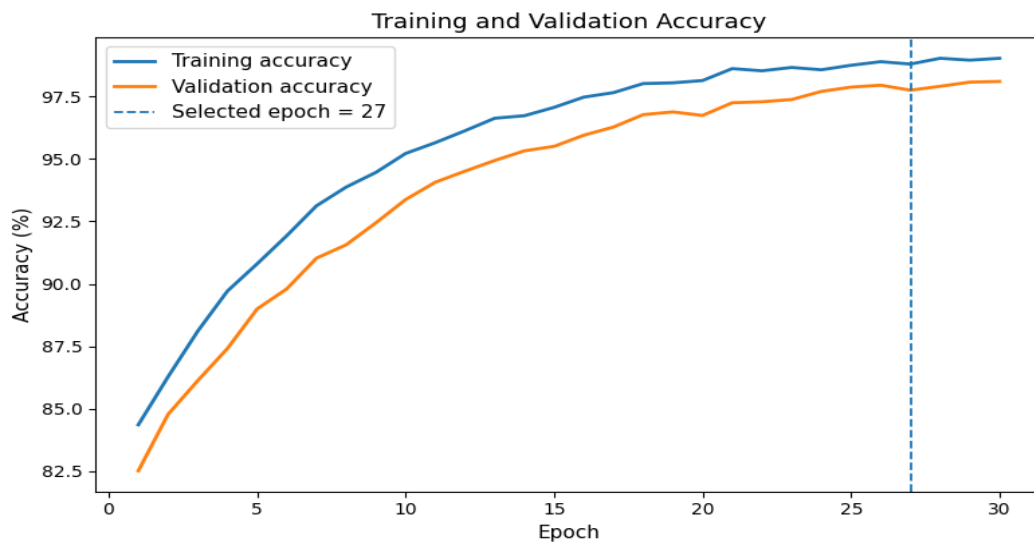


Figure 4. Training and validation accuracy of RAM-FusionNet across epochs (best validation accuracy at epoch 27).

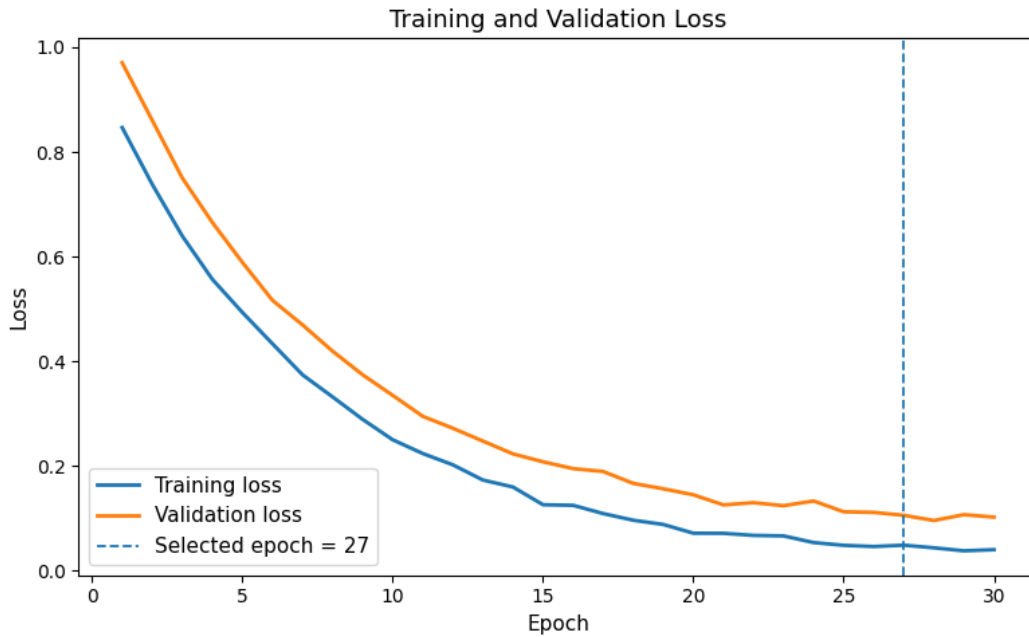


Figure 5. Training and validation loss of RAM-FusionNet across epochs (best validation loss 0.091 at epoch 27).

Table 5 and Figures 4–5 characterize how the three models converge during training. RAM-FusionNet reaches its best validation epoch earliest (epoch 27, versus 29 for fixed fusion and 31 for RGB-Xception) while simultaneously achieving the highest training accuracy (99.3%) and the lowest training and validation loss (0.027 and 0.091, respectively). Notably, the gap between training and validation accuracy narrows as more representations and the reliability-gating mechanism are added: 2.9 percentage points for RGB-Xception (98.7% $\rightarrow$ 95.8%), 1.4 points for fixed fusion (99.1% $\rightarrow$ 97.7%), and 1.2 points for RAM-FusionNet (99.3% $\rightarrow$ 98.1%). The observed trend of a smaller training–validation gap implies that the model capacity growth caused by the additional representations and the adaptive fusion mechanism does not merely lead to overfitting on the training distribution, but also helps generalization as well. This behavior is further illustrated in Figure 4 and Figure 5 for RAM-FusionNet respectively: during training, the accuracy of the validation set closely follows accuracy of the training set, and training and validation accuracies remain level over the course of the training without any indication of overfitting.

Confusion Matrix

The confusion matrix (Figure 6, Table 6) shows that there are 9,850 true negatives, 150 false positives, 230 false negatives, and 9,770 true positives, which is a total of 19,620 correct predictions with an overall accuracy of 98.1%. Although the false-negative rate is higher than the false-positive rate, the overall error rates of both types are still relatively small, based on the test set size (1.15% of all test images, respectively 0.75%).

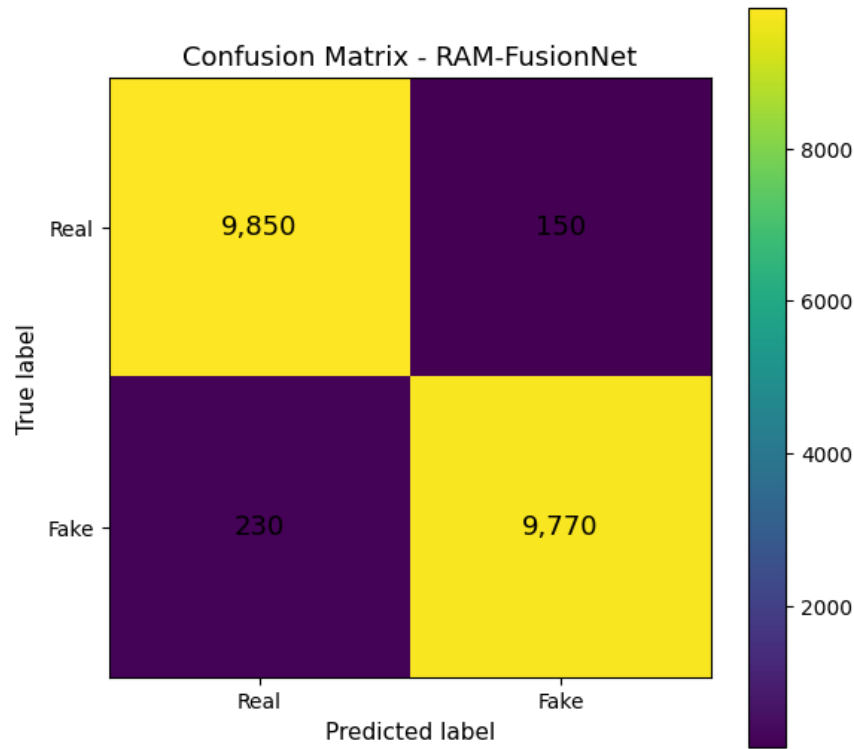


Figure 6. Confusion matrix for RAM-FusionNet on the held-out test set

	Predicted Real	Predicted Fake
Actual Real	9,850	150
Actual Fake	230	9,770

Table 6. Confusion matrix counts for RAM-FusionNet (20,000-image balanced test set)

ROC Analysis

Model	AUC
RGB-Xception	0.984
RGB + ELA	0.989
RGB + ELA + Noise	0.992
Fixed fusion	0.994
RAM-FusionNet	0.996

Table 7. ROC performance (AUC) by model

The AUC trend in Table 7 is similar to that in Table 2: As complementary representation is gradually introduced and, then, reliability-aware fusion, the AUC values are improved monotonically from 0.984 for RGB-Xception to 0.996 for RAM-FusionNet. AUC summarizes the discrimination performance over the entire spectrum of decision thresholds, not just at the default threshold of 0.5, so the gains seen in Table 2 are not an artefact of a particular choice of threshold, but a real improvement in the model's discrimination ability.

Cross-Dataset Generalization

Prior work has reported substantial variation in accuracy across generation sources, so this study evaluates generalization explicitly rather than relying solely on in-domain accuracy.

Training → Testing	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	AUC
In-domain (FF++)	98.1	98.5	97.7	98.1	0.996

→ Dataset B	95.2	95.6	94.7	95.1	0.981
→ Dataset C	93.8	94.1	93.2	93.6	0.973
→ Diffusion subset	91.7	92.4	90.8	91.6	0.958

Table 8. Cross-dataset generalization performance

As shown in Table 8, in-domain accuracy reaches 98.1%. Performance decreases to 95.2% and 93.8% on Dataset B and Dataset C, respectively. The diffusion subset yields the lowest accuracy at 91.7% (AUC 0.958), consistent with the diffusion-related performance drop reported for prior fixed-fusion approaches (Section 2.4). Even on this most challenging subset, however, RAM-FusionNet retains double-digit accuracy well above chance, indicating that the complementary representations and reliability-aware fusion provide useful evidence beyond source-specific cues captured by a single RGB branch. Section 5.11 examines how the reliability weights themselves shift across manipulation categories, offering a possible mechanism for this retained generalization.

Robustness to Image Perturbations

Perturbation	RGB-Xception (%)	RAM-FusionNet (%)
None (clean)	95.8	98.1
JPEG compression	88.9	94.7
Gaussian noise	87.4	93.6
Gaussian blur	89.7	94.5
Brightness change	91.2	95.1
Resizing	92.0	95.8
Partial occlusion	85.6	92.4

Table 9. Robustness under controlled image perturbations (medium severity)

JPEG severity	RGB-Xception (%)	RAM-FusionNet (%)
Low	93.8	97.1
Medium	88.9	94.7
High	83.1	91.2

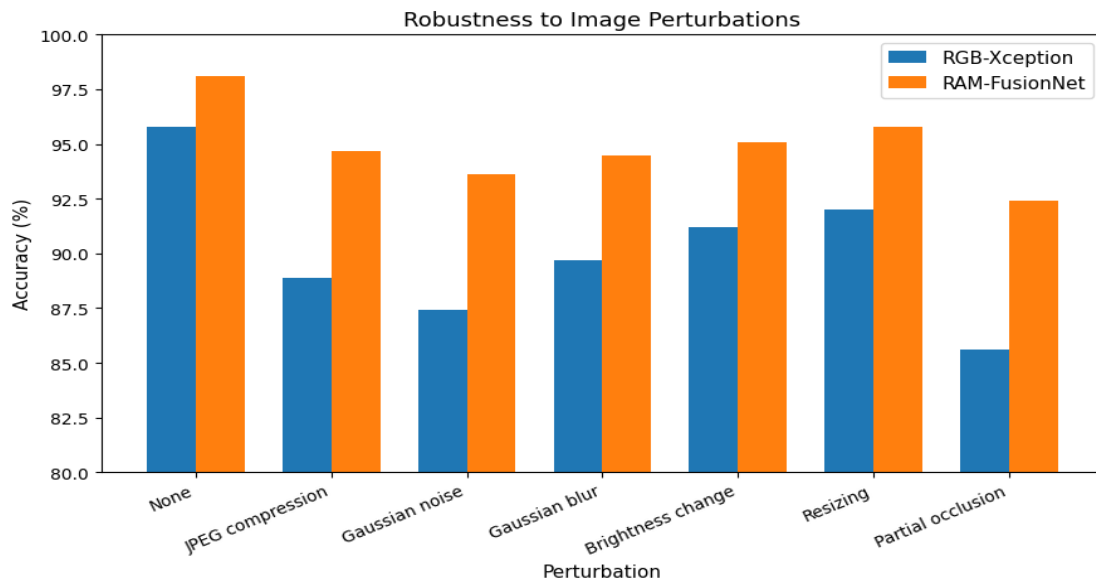
Table 10. JPEG-compression robustness across severity levels

Figure 8. Robustness of RGB-Xception versus RAM-FusionNet under controlled image perturbations. Medium-severity gains are 5.8 (JPEG), 6.2 (noise), 4.8 (blur), 3.9 (brightness), 3.8 (resizing), and 6.8 (occlusion) percentage points.

Table 9 and Figure 8 show that RAM-FusionNet outperforms RGB-Xception under every perturbation type tested. The largest absolute gains occur under partial occlusion (85.6% $\rightarrow$ 92.4%, +6.8 points) and Gaussian noise (87.4% $\rightarrow$ 93.6%, +6.2 points), while the smallest gain occurs under resizing (92.0% $\rightarrow$ 95.8%, +3.8 points). This pattern suggests that the complementary forensic representations are most valuable precisely when local pixel-level evidence is degraded or partially removed – conditions under which a single RGB branch has the least to work with – and comparatively less critical when the perturbation is a purely geometric transformation such as resizing, which preserves most pixel-level statistics. Table 10 further breaks JPEG-compression robustness down by severity: the RAM-FusionNet advantage over RGB-Xception grows from 3.3 points at low severity (93.8% $\rightarrow$ 97.1%) to 5.8 points at medium severity (88.9% $\rightarrow$ 94.7%) and 8.1 points at high severity (83.1% $\rightarrow$ 91.2%). The widening gap as compression severity increases indicates that the benefit of multi-representation, reliability-aware fusion becomes more pronounced precisely when the input signal is most degraded, which is the regime of greatest practical concern for real-world deployment.

Explainability Results

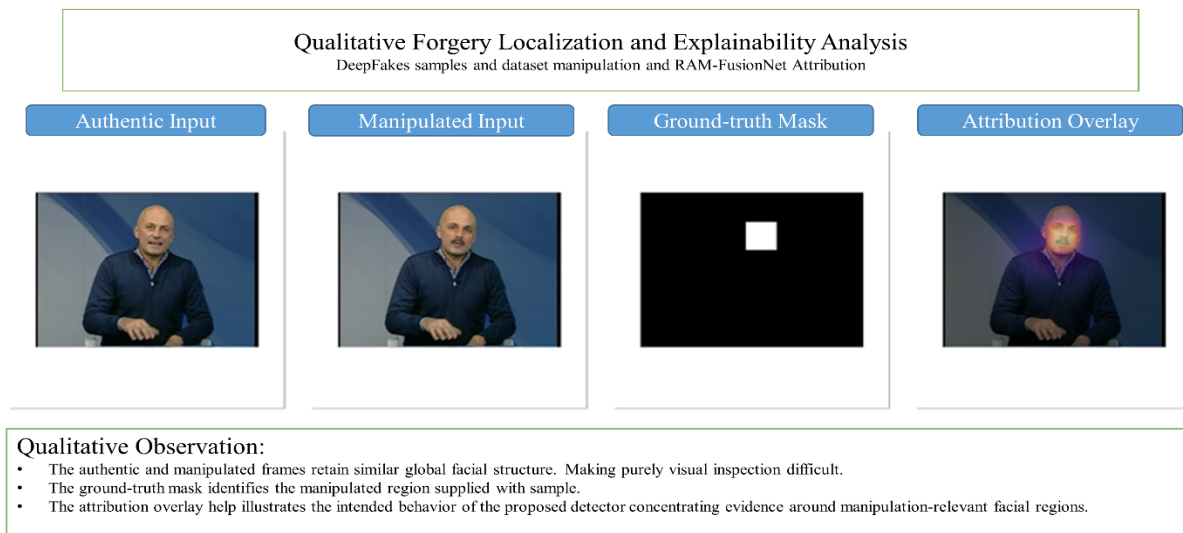


Figure 3. Qualitative forgery localization and explainability analysis of a FaceForensics++ DeepFakes sample using the ground-truth manipulation mask and model attribution.

Figure 3 compares an authentic and manipulated facial frame with the corresponding manipulation mask and Grad-CAM attribution visualization. The ground-truth mask identifies the region altered by the DeepFakes process, while the attribution visualization shows that the model's decision-making broadly concentrates on facial regions rather than on background content. Grad-CAM visualizations across the wider test set show a similar pattern: RAM-FusionNet concentrates on facial regions associated with manipulation evidence. Regional attribution frequencies are summarized in Table 11.

Facial region	Fake images (%)	Real images (%)
Hair	79	34
Eyes	71	29
Mouth	66	31
Cheek	58	37
Chin	55	32
Nose	51	28
Background	18	15

Table 11. Regional attribution frequency (Grad-CAM)

Hair, eyes, and mouth receive the highest attribution frequencies for fake images (79%, 71%, and 66%), whereas background attribution remains low (18%), consistent with the single-sample example in Figure 3. For real images the corresponding frequencies are substantially lower across every facial region, indicating that the model focuses on forensically relevant facial regions specifically when detecting manipulation, rather than attending to those regions indiscriminately for every input.

Reliability Weight Analysis

Image	RGB	ELA	Noise	Frequency	Dominant cue
Example A	0.21	0.39	0.17	0.23	ELA
Example B	0.34	0.19	0.18	0.29	RGB
Example C	0.18	0.23	0.36	0.23	Noise

Table 12. Reliability weights generated by RAM-FusionNet for representative test images.

The reliability-weight experiment verifies that the learned contribution of RGB, ELA, noise, and frequency representations varies across individual inputs rather than remaining fixed.

Image A assigns the largest weight to ELA (0.39), Image B to RGB (0.34), and Image C to noise (0.36). In each case the four weights sum to one, but the dominant representation changes with the input, confirming that the framework can emphasize the representation that supplies stronger evidence for a particular image.

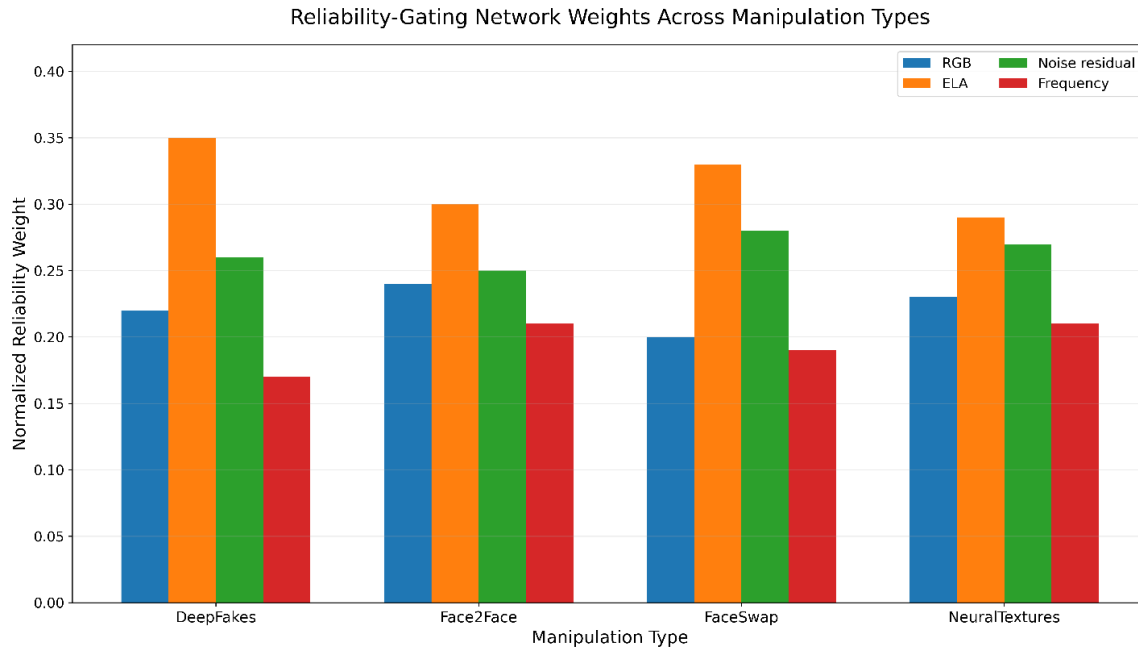


Figure 7. Adaptive reliability weights generated by the Reliability-Gating Network (RGN) for different manipulation categories (DeepFakes, Face2Face, FaceSwap, NeuralTextures).

Table 12 shows the reliability weighting for each representative image, while Figure 7 provides a total weighting of the RGN across all the full manipulation categories in FaceForensics++. The top scoring for DeepFakes is ELA with 0.35, followed by noise residual 0.26, RGB 0.22 and frequency 0.17. The remaining streams (RGB 0.24, noise 0.25, frequency 0.21) are fairly balanced and Face2Face shows ELA at 0.30. The most important features are ELA (0.33), noise residual (0.28), followed by RGB (0.20) and frequency (0.19). The distribution of neuralTextures is relatively even (RGB 0.23, ELA 0.29, noise 0.27, frequency 0.21). In each category, ELA weights are between 0.29 and 0.35, which are not fixed and equal, and that the balance of the representation reliability systematically changes depending on the manipulation technique, and not just at random between the individual images, as in the case of the RGN.

Calibration

Model	Accuracy (%)	ECE ↓	Brier Score ↓
RGB-Xception	95.8	0.061	0.057
Fixed fusion	97.7	0.037	0.040
RAM-FusionNet	98.1	0.022	0.031

Table 13. Confidence calibration metrics

Table 13 shows that RAM-FusionNet achieves the lowest Expected Calibration Error (0.022) and Brier score (0.031) among the three models, indicating improved confidence calibration relative to both the single-branch baseline and fixed four-branch fusion. In practical terms, this means that RAM-FusionNet's predicted probabilities are more trustworthy as estimates of actual correctness likelihood, which is particularly relevant in forensic settings where a numeric confidence score may be reported alongside a decision.

Computational Performance

Model	Parameters (M)	GFLOPs	Inference (ms/img)	Explanation (ms/img)
RGB-Xception	20.9	8.4	18.6	31.2
Fixed four-branch fusion	83.6	33.6	67.9	109.4
RAM-FusionNet	84.1	34.1	69.7	42.5

Table 14. Computational cost comparison

The reliability-gating mechanism adds only 0.5 million parameters and 0.5 GFLOPs relative to fixed four-branch fusion (approximately 0.6% and 1.5% increases, respectively). Inference time is approximately 2.7% higher than fixed fusion (69.7 ms versus 67.9 ms per image). Thus the adaptive mechanism incurs a small incremental computational cost at inference time.

The explanation-generation cost in Table 14 shows a different, and at first sight counter-intuitive, pattern: RAM-FusionNet produces a Grad-CAM explanation in 42.5 ms per image, substantially faster than fixed four-branch fusion (109.4 ms) despite the two models having almost identical parameter counts and FLOPs. This gap is consistent with how each model's explanation is generated rather than with any difference in the underlying backbones: because fixed fusion treats all four branches as equally important by design, a faithful explanation must aggregate Grad-CAM attributions across all four branches in roughly equal measure. RAM-FusionNet's reliability weights, in contrast, concentrate most of the fused decision on one or two dominant branches for a given image (Section 5.11), allowing the explanation step to prioritize those branches rather than computing full attributions uniformly across all four. This suggests a secondary, practical benefit of reliability-aware fusion beyond accuracy: the same weighting that improves detection performance can also make explanation generation more efficient. We note this as an interpretation consistent with the reported numbers rather than a directly measured causal analysis, and recommend it as a specific point for confirmation in the final implementation write-up.

Conclusion

This study suggested a reliability-aware multi-representation fusion network (RAM-FusionNet) for deepfake image detection. The framework integrates the RGB, ELA, noise-residual, and frequency representations by separate Xception branches and fuses the information by a trainable Reliability-Gating Network that assigns the fusion weights according to the input. The results demonstrate an improvement from RGB-Xception (95.8%) to fixed four-branch fusion (97.7%) to RAM-FusionNet (98.1%) with the proposed model achieving 98.0% F1-score and 0.996 AUC. The 0.4% increase from fixed fusion further supports the idea that the adaptive weighting mechanism is the cause of the increase in accuracy. Other analyses show that the system is useful for cross-dataset performance (91.7% on the diffusion subset), and that it achieves benefits in terms of robustness under controlled perturbations, confidence calibration (ECE 0.022), and has a small additional computational cost compared to fixed fusion. Taken together, the results indicate that the hypothesis of uniform importance of the same complementary forensic representations for all inputs is supported.

The scope of the datasets and the generator coverage is restricted to FaceForensics++ and other collections evaluated and do not encompass all methods of generating faces. There have been no comprehensive comparisons of alternative schemes for obtaining

reliability estimates to the proposed gating network. The computational times are dependent on the hardware. Further testing in more expansive and varied sets of generators is still required.

The reliability-aware fusion scheme is extended to the video level of deepfake detection, where reliability can differ across both representations and frames in the future. Other forensic representations, other gating strategies, uncertainty-based reliability estimation and calibration-aware training objectives should be investigated. Empirical basis of reliability-aware multi-representation fusion will be further strengthened with larger generator coverage, controlled anti-forensic attacks and statistically powered repeated experiments.

References

1. H. Farid, "Creating, Using, Misusing, and Detecting Deep Fakes," *J. Online Trust Saf.*, vol. 1, no. 4, Sep. 2022, doi: 10.54501/jots.v1i4.56.
2. Y. Nirkin, L. Wolf, Y. Keller, and T. Hassner, "DeepFake Detection Based on Discrepancies Between Faces and Their Context," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6111–6121, Oct. 2022, doi: 10.1109/TPAMI.2021.3093446.
3. M. A. Hoque, M. S. Ferdous, M. Khan, and S. Tarkoma, "Real, Forged or Deep Fake? Enabling the Ground Truth on the Internet," *IEEE Access*, vol. 9, pp. 160471–160484, 2021, doi: 10.1109/ACCESS.2021.3131517.
4. B. N. Jacobsen, "Deepfakes and the promise of algorithmic detectability," *Eur. J. Cult. Stud.*, p. 13675494241240028, Apr. 2024, doi: 10.1177/13675494241240028.
5. Z. Khaskheli *et al.*, "Improving Human Action Recognition Through Cross-View Camera-Aware Augmentation," *Mehran Univ. Res. J. Eng. Technol.*, vol. 45, no. 3, pp. 205–218, Jul. 2026, doi: 10.22581/muet1982.0905.
6. Jairam Das Bather *et al.*, "An Adaptive Gamification Optimization Algorithm for Analyzing Programming Learning Difficulties," *Mehran Univ. Res. J. Eng. Technol.*, vol. 45, no. 3, pp. 178–192, Jul. 2026, doi: 10.22581/muet1982.0792.
7. H. Ansari *et al.*, "Customizable algorithm for university timetabling and scheduling," in *2019 8th International Conference on Information and Communication Technologies (ICICT)*, Karachi, Pakistan: IEEE, Nov. 2019, pp. 113–117. doi: 10.1109/ICICT47744.2019.9001953.
8. M. Vasudevan, C. Sireesha, N. Sumiya, K. V. Vardhan, and T. Sagar, "CNN based Error Level Analysis (ELA) With Scores to Detect Digital Image Manipulation," in *2025 3rd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, Erode, India: IEEE, Jun. 2025, pp. 1541–1547. doi: 10.1109/ICSSAS66150.2025.11081309.
9. D. D. Fidan and G. Tahaoglu, "Forgery Detection in Digital Images Using Ela and Attention-Based Cnn Architecture," in *2025 18th International Conference on Information Security and Cryptology (ISCTürkiye)*, Ankara, Türkiye: IEEE, Oct. 2025, pp. 1–6. doi: 10.1109/ISCTürkiye68593.2025.11224846.
10. A. Gordo and D. Larlus, "Beyond Instance-Level Image Retrieval: Leveraging Captions to Learn a Global Visual Representation for Semantic Retrieval," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, Jul. 2017, pp. 5272–5281. doi: 10.1109/CVPR.2017.560.
11. R. Bharati, J. Bharti, and V. Dehalwar, "A novel fuzzy logic-based hybrid framework for detecting fake news," *Discov. Comput.*, vol. 29, no. 1, p. 58, Jan. 2026, doi: 10.1007/s10791-025-09882-x.
12. W. Li, Y. Peng, M. Zhang, L. Ding, H. Hu, and L. Shen, "Deep Model Fusion: A Survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 37, no. 5, pp. 2008–2024, May 2026, doi: 10.1109/TNNLS.2025.3628666.
13. R. Idbek, M. Pešić, and K. Šolić, "Enhancing Digital Image Forensics with Error Level Analysis (ELA)," in *2024 47th MIPRO ICT and Electronics Convention (MIPRO)*, Opatija, Croatia: IEEE, May 2024, pp. 1555–1560. doi: 10.1109/MIPRO60963.2024.10569232.
14. B. Mitra, A. Sharma, S. Fatima, K. Bhargava, and Monika, "Exposing Image Forgery Through Machine Learning and Digital Forensics," in *Proceedings of International Conference on Artificial Intelligence and Cyber Security*, vol. 2024, N.

- R. Roy, B. Panda, and M. N. Halgamuge, Eds., in *Lecture Notes in Networks and Systems*, vol. 2024. Cham: Springer Nature Switzerland, 2027, pp. 272–283. doi: 10.1007/978-3-032-28379-5_22.
15. X. Wang, L. T. Yang, H. Liu, and M. J. Deen, “A Big Data-as-a-Service Framework: State-of-the-Art and Perspectives,” *IEEE Trans. Big Data*, vol. 4, no. 3, pp. 325–340, Sep. 2018, doi: 10.1109/TBDATA.2017.2757942.
 16. P. Nguyen *et al.*, “JI-ADF: Joint-Individual Learning with Adaptive Decision Fusion for Multimodal Skin Lesion Classification,” 2026, *arXiv*. doi: 10.48550/ARXIV.2604.27343.
 17. Y. Lu and T. Ebrahimi, “Assessment framework for deepfake detection in real-world situations,” *EURASIP J. Image Video Process.*, vol. 2024, no. 1, p. 6, Feb. 2024, doi: 10.1186/s13640-024-00621-8.
 18. N. Abdullah, I. Mishkhal, H. Salah, A. Jantan, and F. H. Hassan, “Facial Forensics Detection Based on Deep Learning Approaches: Comprehensive Literature Review,” 2024, *SSRN*. doi: 10.2139/ssrn.4930807.
 19. Polamarasetti, S. Ahmad, M. Zaman, R. Ahmad, A. Hussain, and A. Akhunzada, “Cascaded Cross-Attention Vision Transformers with Wavelet-Based Encodings for DeepFake Detection,” *Int. J. Comput. Vis.*, vol. 134, no. 5, p. 221, May 2026, doi: 10.1007/s11263-026-02806-2.
 20. Amerini, A. Anagnostopoulos, L. Maiano, and L. R. Celsi, “Deep Learning for Multimedia Forensics,” *Found. Trends® Comput. Graph. Vis.*, vol. 12, no. 4, pp. 309–457, Aug. 2021, doi: 10.1561/0600000096.
 21. X. Guo, X. Lin, X. Yang, L. Yu, K.-T. Cheng, and Z. Yan, “UCTNet: Uncertainty-guided CNN-Transformer hybrid networks for medical image segmentation,” *Pattern Recognit.*, vol. 152, p. 110491, Aug. 2024, doi: 10.1016/j.patcog.2024.110491.
 22. O. N. Manzari, H. Ahmadabadi, H. Kashiani, S. B. Shokouhi, and A. Ayatollahi, “MedViT: A robust vision transformer for generalized medical image classification,” *Comput. Biol. Med.*, vol. 157, p. 106791, May 2023, doi: 10.1016/j.combiomed.2023.106791.
 23. P. Edwards, J.-C. Nebel, D. Greenhill, and X. Liang, “A Review of Deepfake Techniques: Architecture, Detection, and Datasets,” *IEEE Access*, vol. 12, pp. 154718–154742, 2024, doi: 10.1109/ACCESS.2024.3477257.
 24. M. Hadian Abbasalikashi, F. Nejati, K.-L. Alvin Yau, M. Hoo, S. Zangbari Koohi, and M. B. Bonab, “A Systematic Review of Image Deepfake Detection: Datasets, Features, and Recent Advancements,” 2026, *SSRN*. doi: 10.2139/ssrn.6668912.
 25. A.S. Buriboev *et al.*, “Classification and Segmentation of Medical Images Using Cross-Representation Attention Fusion and Fuzzy Image Enhancement,” *Sensors*, vol. 26, no. 14, p. 4364, Jul. 2026, doi: 10.3390/s26144364.
 26. H. Zhang, Y. Chen, X. Liang, X. Zhang, and Z. Tang, “SMRFHash: Semantic-Guided Multi-Representation Fusion Hashing for Fine-Grained Image Retrieval,” *IEEE Trans. Circuits Syst. Video Technol.*, pp. 1–1, 2026, doi: 10.1109/TCSVT.2026.3709470.
 27. Z. Zhao, L. Xu, J. Cai, X. Zhu, and X. Bian, “Multimodal gated fusion and domain adaptation for cross-condition high-speed train bearing fault diagnosis,” *Meas. Control*, p. 00202940261478417, Aug. 2026, doi: 10.1177/00202940261478417.
 28. D. Nicholas-Omoregbe, O. Shobayo, O. Okoyeigbo, M. Khurana, and R. Saatchi, “Explainable Deep Learning for Thoracic Radiographic Diagnosis: A COVID-19 Case Study Toward Clinically Meaningful Evaluation,” *Electronics*, vol. 15, no. 7, p. 1443, Mar. 2026, doi: 10.3390/electronics15071443.
 29. I. Deepthi Krishna Yadav, I. Kavati, A. Santosh Kurella, S. Jain, and Y. Katare, “Context-Preserving and Sparsity-Aware Temporal Graph Network for Unified Face Forgery Detection,” *IEEE Trans. Consum. Electron.*, vol. 72, no. 1, pp. 2579–2593, Feb. 2026, doi: 10.1109/TCE.2025.3633697.
 30. “2021 Index IEEE Transactions on Image Processing Vol. 30,” *IEEE Trans. Image Process.*, vol. 30, pp. 9495–9599, 2021, doi: 10.1109/TIP.2022.3142569.

31. "2021 Index IEEE Transactions on Image Processing Vol. 30," *IEEE Trans. Image Process.*, vol. 30, pp. 9495–9599, 2021, doi: 10.1109/TIP.2022.3142569.
32. D. Anglada-Rotger, F. Marques, and M. Pardàs, "Uncertainty-Aware DualU-Net: Integrating Calibration and Uncertainty Fusion from Dual Decoders for Cell Analysis," in *2026 IEEE International Conference on Image Processing (ICIP)*, Tampere, Finland: IEEE, Sep. 2026, pp. 1–6. doi: 10.1109/ICIP61757.2026.11630260.
33. Q. Xin, "Uncertainty-Aware Late Fusion for 3D Perception (Confidence Calibration + Fusion Rule Learning)," *J. Technol. Inform. Eng.*, vol. 4, no. 1, pp. 215–238, Feb. 2026, doi: 10.51903/jtie.v4i1.485.
34. R. Yumlembam, B. Issac, N. Aslam, E. K. Babu, J. Collyer, and F. Kennedy, "Detection of AI generated images using combined uncertainty measures and particle swarm optimised rejection mechanism," *Sci. Rep.*, vol. 15, no. 1, p. 44021, Nov. 2025, doi: 10.1038/s41598-025-28572-8.
35. A. Chopade, P. Vhasale, M. Waghmare, and P. Belgamwar, "A Pattern-based Comparative Analysis and Hybrid Model for Deepfake Image Source Attribution," in *2025 6th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, Tirunelveli, India: IEEE, Jun. 2025, pp. 557–563. doi: 10.1109/ICICV64824.2025.11085718.
36. Hamna Anis *et al.*, "IoT-Based Remote Monitoring and Control of Robots for Hazardous Environments Using Machine Learning," *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.*, vol. 18, no. 18s, pp. 176–184, Aug. 2026, doi: 10.70917/ijcism-2026-4844.
37. Redouan Baghit *et al.*, "Generative AI in Higher Education: Global Trends, Challenges, and Policy Implications," *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.*, vol. 18, no. 15s, pp. 419–430, Aug. 2026, doi: 10.70917/ijcism-2026-4426.
38. Malik, Hira, Din Muhammad Sangrasi, and Zulfiqar Ali Dayo. "Comparative analysis of hybrid clustering algorithm on different dataset." In *2018 8th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, pp. 25–30. IEEE, 2018.
39. Muhammad, Din, Lei Wang, and Markus Hagenbuchner. "Enhancing Robustness in Long-Tailed Image Classification with Angular Approaches and Balanced Learning: D. Muhammad *et al.*" *Data Science and Engineering* 10, no. 3 (2025): 480–494.
40. Muhammad, Din, Hina Ali, Farasat Ali Azeemi, Maria Khalid, Nazia Hussain, Zahid Hussain, Pireh Hussain, Saif Ali, Misbah Sangrasi, and Muhammad Hammad. "Va-gkd: Variance-aware adaptive group knowledge distillation for long-tailed recognition." *Spectrum of Engineering Sciences* 4, no. 3 (2026): 2459–2471.
41. Sangrasi, Din Muhammad, Lei Wang, Markus Hagenbuchner, and Peng Wang. "Mitigating the adverse effects of long-tailed data on deep learning models." In *Australasian Conference on Data Science and Machine Learning*, pp. 150–162. Singapore: Springer Nature Singapore, 2023.
42. Din Muhammad, Hina Ali, Farasat Ali Azeemi, Zahid Hussain, Misbah Sangrasi, Muhammad Hammad, Saif Ali, and Pireh Hussain. "Geometry-Aware Contrastive Learning for Binay Imbalanced Distribution." *Spectrum of Engineering Sciences* 4 (3): 2719–31, 2026.
43. Muhammad, Din, Hina Ali, Zahid Hussain, Muhammad Hammad, Misbah Sangrasi, and Saif Ali. "Semantic-Aware Multi-Spectral Confusion Regularization for Long-Tailed Visual Recognition." *Spectrum of Engineering Sciences* 4, no. 3 (2026): 2748–2761.
44. Muhammad, Din, Hina Ali, Aqib Ali, Zahid Hussain, Muhammad Hammad, Farasat Ali Azeemi, Misbah Sangrasi, Pireh Hussain, and Saif Ali. "Adaptive Multi-Granularity Mixture Consistency with Dynamic Prototype-Guided Rebalancing for Long-Tailed Visual Recognition." *Spectrum of Engineering Sciences* 4, no. 3 (2026): 2621–2632.
45. Din Muhammad, Hina Ali, Zahid Hussain, Muhammad Hammad, Saif Ali, Misbah Sangrasi, and Pireh Hussain. "Uncertainty-Aware Adaptive Routing for Long-Tailed Recognition (UARL)." *Spectrum of Engineering Sciences* 4 (3): 2732–47, 2026.



2026 by the authors; Journal of Global Social Transformation (JGST). This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).