*Journal of Artificial Intelligence and Technology*, 2026, 6, 848-856
https://doi.org/10.37965/jait.2026.1281

ISTP
RESEARCH ARTICLE

# Hybrid Multi-Scale Deep Feature Fusion Network for Retouch Image Forgery Detection

**Vanishri V. Sataraddi[1,2] and N. P. Nethravathi[1]**

[1]School of Computer Science and Engineering, REVA University, Bengaluru, 560064, Karnataka, India
[2]Department of Computer Science and Engineering, RNS Institute of Technology, Bengaluru, 560098, Karnataka, India

(Received 20 March 2026; Revised 08 June 2026; Accepted 01 July 2026; Published online 30 July 2026)

***Abstract:*** The number of digitally altered images is growing because of increasingly powerful and easy-to-use tools for image editing. This development has rendered web-based visual information less credible and more challenging to substantiate. Among various image editing techniques, image retouching is one of the most commonly used methods. Unlike the normal image editing, image retouching does not modify the underlying structure of the image. This procedure can help to alter the texture, tone, and blemishes on the skin without changing the structure of the image. These subtle changes can pose difficulties for traditional forensic methods to detect image manipulation that involves retouching. To overcome these shortcomings, this paper introduces a hybrid deep feature fusion (HDFF) network to detect digital image manipulation (DIM) with retouching forgery. The proposed architecture is a convolutional layered approach with multi-scale technique, which is able to effectively capture subtle texture variations to learn them. The model is assessed against the CASIA v2.0 and FaceForensics++ benchmark datasets, with the standard metrics of accuracy, precision, recall, and F1-score. The results show that the approach attains a detection accuracy of 97.2%, performing better than several existing deep learning-based forgery detection methods. This improvement shows the value of combining multi-level features for the detection of minor retouching artifacts. Also, the approach provides a reliable platform for multimedia authentication and social media validation in digital image forensics.

***Keywords:*** Convolutional neural networks (CNNs); deep feature fusion; digital image forensics; face manipulation detection; multimedia authentication; multi-scale feature learning; retouch forgery detection

## I. INTRODUCTION

The widespread use of image data in social media platforms and online communication channels and multimedia applications has created greater doubts about the genuine nature of visual material. Enhancements in image modification tools have made it possible to modify images with high realism, often without introducing perceptible visual distortions. Therefore, validating images becomes one of the most challenging aspects of image forensics, especially in scenarios where there is a need for higher reliability and assurance [1]. In general, digitally manipulated images are categorized into three groups: copy-move, splicing, and retouching. Duplication of image regions to create objects is referred to as copy-move forgery (CMF). On the other hand, image splicing involves integration of several images to create one. On the contrary, image retouching involves the substitution of skin texture and removal of blemishes and facial features, while preserving the overall structure of the image. Retouching is the mildest form of manipulation, and because it produces very few artifacts, it remains extremely difficult to detect in comparison to other types of image manipulations [2,3]. Most of the contemporary image forgery detection (IFD) techniques focus on finding statistical anomalies within the image objects, such as compression signatures, pattern noise, and inconsistencies in illumination [4,5]. Although these techniques have proven to be effective in identifying a specific class of manipulations, they are mostly handcrafted features and perform poorly on intricate modifications of the data.

The rapid advancements in deep learning have greatly impacted the field of image forensics by introducing automated feature extraction and advanced pattern recognition. Convolutional neural networks (CNNs) have been widely used for the identification of manipulated images due to their ability to analyze and learn hierarchical feature representations directly from raw image data [6,7].

These models are capable of capturing manipulation patterns that are particularly hard to recognize using conventional detection methods.

However, current deep learning methods may find it very hard to recognize retouch-based manipulations because such manipulations are generally coarse, and their alterations are of low intensity, which can be hidden in natural variations in images.

This paper provides a hybrid deep feature fusion (HDFF) framework for retouch IFD [8] to enhance the ability of facial manipulation identification in digital image forensics. This method comprises convolutional multi-scale feature combination to capture local inconsistencies of textures and large-scale global information. The model offers a better opportunity to identify low-intensity manipulative artifacts that were layered during the retouching process. The other benefit of the approach based on multi-scale convolutional networks is the HDFF framework, which offers an implicit combination of features of hierarchies to identify low-intensity texture inconsistencies and low global variations within retouched images. This framework is further beneficial, as it increases the sensitivity of detection by the combined functionality of features to identify low-intensity retouch artifacts.

The proposed method outperformed existing methods of detecting retouched images, especially in scenarios that were considered

Corresponding author: Vanishri V. Sataraddi (e-mail: vanishrisataraddi@gmail.com).


© The Author(s) 2026. This is an open access article published under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

difficult to identify. Unlike existing methods that rely on single-scale feature extraction or general deep learning approaches, the proposed HDFF framework specifically focuses on retouch-based manipulations by integrating multi-scale feature representations. This enables enhanced detection of subtle and low-intensity artifacts that are often overlooked by conventional methods, thereby improving the effectiveness of retouch IFD.

The rest of the paper is organized as follows: Section II reviews related work. Section III outlines the methodology. Section IV describes the algorithm. Section V presents the dataset and experimental setup. Section VI discusses the results, and section VII concludes the paper.

# II. RELATED WORKS

## A. TRADITIONAL IFD METHODS

Digital forensics research initially concentrated on looking for evidence of errors caused by the way in which a digital picture was taken or processed. Texture-based approaches attempt to identify local regions of textures within an image pattern. For instance, Local Binary Pattern (LBP) descriptors are effective in identifying texture irregularities resulting from image editing processes [3]. Descriptors including Scale-Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF) are used to detect duplicate regions, including those resulting from copy-move forgery (CMF) [4,5]. Another widely used forensic technique is Sensor Pattern Noise (SPN) analysis, which exploits the unique noise characteristics of an imaging sensor to identify the source of an image and detect possible manipulations [9]. Although these conventional approaches perform well in some cases, they rely heavily on handcrafted features and do not perform well on complex or visually consistent manipulations generated by modern editing tools.

## B. MACHINE LEARNING-BASED APPROACHES

To overcome the limitations of traditional methods, machine learning methods are presented in the context of IFD. These approaches also present a way of extracting discriminative features from images and training classifiers, including support vector machines (SVMs) and also random forests, to distinguish between authentic and manipulated images [10,11]. Those models have shown better performance compared to purely rule-based methods, particularly when trained on carefully designed statistical features. But their dependence on manually engineered features limits their flexibility to a wide range of techniques [12,13].

## C. DEEP LEARNING-BASED IFD

The rise of deep learning has significantly advanced research in image forgery detection through automatic feature extraction from raw image data. CNNs are particularly effective in capturing hierarchical representations that are widely used to detect manipulation artifacts [6,7]. For example, Bayar and Stamm [6] introduced a constrained convolutional layer for image reduction and to highlight tampering artifacts. Deeper architectures such as VGGNet and ResNet also give the detection accuracy by leveraging rich feature representations [7,8]. Additionally, transfer learning strategies are used to adapt pretrained models to forgery detection issues using datasets such as ImageNet [14]. Also recently, attention mechanisms and transformer-based models have demonstrated the ability to capture long-range dependencies and global appropriate information inside images [15,16]. Hybrid architectures that combine CNN with transformer networks have also been planned to combine local texture analysis with global feature modeling, resulting in improved detection performance [17,18].

## D. RECENT ADVANCES IN RETOUCH FORGERY DETECTION

A recent study has reported detecting retouch-based manipulations, particularly in facial images. Multi-scale CNN architectures are established to collect both fine-grained texture details and fine-grained structural variations introduced during retouch operations [19,20]. Deep learning models are applied to identify facial manipulations generated by advanced editing tools and beautification filters [21,22]. Furthermore, studies on synthetic image generation and deepfake detection contribute to understanding manipulation artifacts introduced by neural networks [23,24]. Artificial intelligence techniques are currently widely employed in multimedia forensics to analyze and verify visual content authenticity [25].

However, figuring out retouch changes is still a tough task because they often introduce localized, minor modifications that are difficult to distinguish from natural fluctuations. These diffused discrepancies won't be safely captured through present-day techniques. This paper proposes an HDFF method that integrates multi-scale feature representations to improve the detection of retouch-based image forgeries.

# III. PROPOSED METHODOLOGY

This section provides a detailed description of the proposed HDFF framework, including its architecture, feature extraction process, and multi-scale fusion mechanism for effective retouch IFD. Retouch manipulations regulate facial appearance through beauty-modifying operations, including pores and skin smoothing, blemish elimination, tone adjustment, and makeup application. These modifications introduce subtle grain-level issues that are difficult to detect using conventional forensic techniques. To address this issue, a deep learning-based approach is proposed that effectively identifies manipulation artifacts.

The system architecture shown in Figure 1 consists of four primary stages: image acquisition, preprocessing and feature extraction, multi-scale feature fusion, and classification.

## A. INPUT IMAGE ACQUISITION

The initial stage includes acquiring input images from the dataset. These images may originate from digital cameras, online platforms, or image editing software. As the retouch manipulations are mostly performed on face images, the dataset mainly contains portrait images with cosmetic modifications. These collected images are then sent to the preprocessing stage.

## B. IMAGE PREPROCESSING

Preprocessing is carried out to standardize the input data and improve model performance. Images collected from different sources usually vary in resolution, illumination, and quality. Thus, there is a need to preprocess the data before feeding it to the network to have the same representation. The preprocessing steps include the following:

- Resizing images to a fixed resolution of $256 \times 256$ pixels
- Pixel intensity normalization

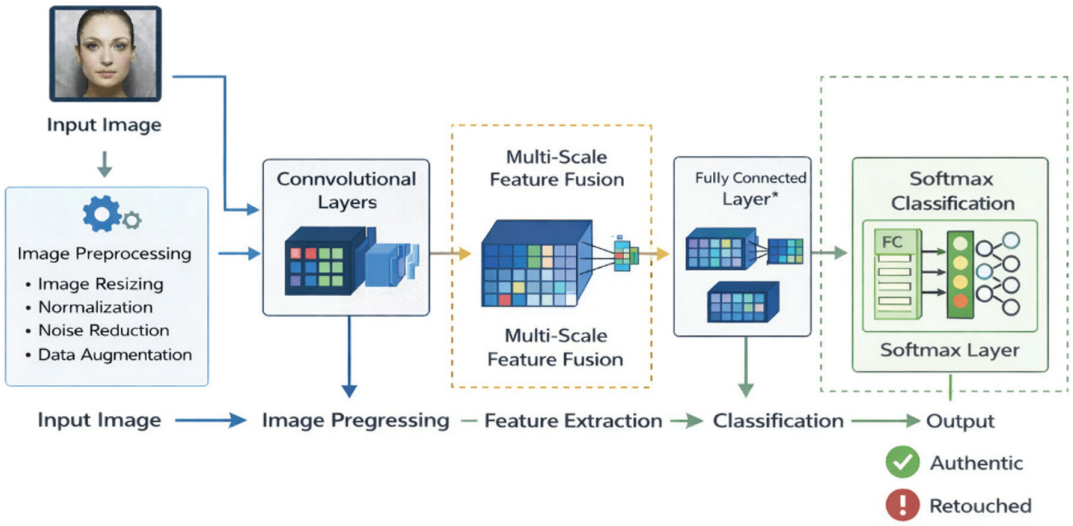


**Fig. 1.** HDFF architecture for retouch image forgery detection.

- Noise reduction
- Data augmentation

Normalization scales pixel values to the range [0, 1], which enhances training stability and accelerates convergence.

## C. CONVOLUTIONAL FEATURE EXTRACTION

The images are first preprocessed and then fed into a CNN, where features are extracted. Convolutional layers capture disrupted spatial patterns due to retouching and enhance a model's ability to extract feature representations. This can help the model show retouching artifacts that are not usually visible with other methods.

The convolution operation is mathematically represented as:

$$F(i,j) = \sum_{m=0}^{M} \sum_{n=0}^{N} I(i-m, j-n) K(m,n) \tag{1}$$

where $F(i,j)$ denotes the output feature map, $I$ is the input image, and $K$ is the convolution kernel.

To introduce nonlinearity, the rectified linear unit (ReLU) activation function is applied:

$$A(x) = \max(0,x) \tag{2}$$

ReLU accelerates and makes the feature representation learning faster. Convolutional layers segment the image at an input level into a hierarchy of feature representations, which include both texture information and higher-level manipulation patterns. Moreover, a multi-scale feature fusion module is integrated into the model to improve the performance by fusing features from multiple layers. This allows the model to be more effective in extracting manipulation artifacts.

## D. MULTI-SCALE FEATURE FUSION

Retouch manipulations manifest at dissimilar spatial scales, ranging from fine-grained texture changes to broader structural modifications. To effectively capture these variations, a multi-scale feature fusion mechanism is incorporated.

Feature maps of several convolutional layers are merged to create an integrated representation:

$$F_{fusion} = w_1 F_1 + w_2 F_2 + w_3 F_3 \tag{3}$$

where F1, F2, and F3 are the feature maps of the respective convolution layers and the w's are the learnable weights. Articulation of both local and global information to find manipulation artifacts is the intended goal of this fusion strategy.

Retouching manipulations happen at various spatial resolutions, which is the reason a multi-scale feature fusion strategy was employed. The proposed method integrates feature maps from several convolutional layers, which enables the capturing of texture and structural inconsistency at various levels. This method is in direct contrast with other methods that use single-scale features. Single-scale methods often miss small artifacts when compared to multi-scale methods.

## E. FULLY CONNECTED LAYER

The integrated feature maps are reshaped and input into a fully connected layer to produce a compact feature set.

## F. SOFTMAX CLASSIFICATION

The final step we take in the classification task is the application of the Softmax function. Here, the function ascertains the extent of each output class (retouched or authentic).

The Softmax function is

$$\mathrm{P(y|x)} \frac{e^{z_y}}{\sum_{k=1}^{C} ez_k} \tag{4}$$

where $z_y$ represents the output score of class $y$ and $C$ represents the total number of classes.

The Softmax places the output of the network into the probability scale and allows the system to classify the input image as either authentic or retouched.

# IV. ALGORITHM

This part of the paper describes an algorithm for retouch IFD developed in the HDFF framework. The algorithm under consideration processes images using a CNN for hierarchical image

**Algorithm 1.** Retouch Image Forgery Detection

**Input:** Image dataset D

**Output:** Image label L (Authentic or Retouched)

1. Load input image I from dataset D.
2. Perform preprocessing:
   a. Resize image to fixed resolution.
   b. Normalize pixel intensity values.
   c. Apply noise reduction and data augmentation.
3. Extract feature maps using convolution operation **Eq. (1)**.
4. Apply nonlinear activation using the ReLU **Eq. (2)**.
5. Generate hierarchical feature representations through multiple convolution layers.
6. Perform multi-scale feature fusion **Eq. (3)**.
7. Convert fused feature maps into a feature vector.
8. Pass the feature vector through the fully connected layer.
9. Compute class probabilities using Softmax **Eq. (4)**.
10. If the predicted probability for the manipulated class exceeds the defined threshold:
    Set L = Retouched Image.
11. Otherwise:
    Set L = Authentic Image.
12. Return the classification label L.

feature extraction, integrates features of different scales, and performs classification with a Softmax operation.

The algorithm enables effective identification of subtle retouch manipulations by combining multi-scale feature learning with deep hierarchical representations.

# V. DATASET DESCRIPTION AND EXPERIMENTAL SETUP

This section presents the datasets and experimental configuration used to assess the performance of the devised retouch IFD framework. Openly available benchmark datasets are utilized to ensure reproducibility and consistency of the results.

## A. DATASET DESCRIPTION

The effectiveness of the proposed HDFF framework was analyzed on the two most commonly used datasets, the CASIA v2.0 and FaceForensics++ datasets. These datasets cover different image manipulation techniques, including retouching, merging, and alterations to facial images.

The CASIA v2.0 dataset [26] is a widely used benchmark in image forensics and includes both original images and pictures that have gone through different manipulation processes. The images in this dataset include variations in lighting and the equipment used for image capture, which enables testing the robustness of a model.

FaceForensics++ [27] is a large, state-of-the-art dataset for the detection of manipulations that involve facial images. This dataset includes the original images and images that have been manipulated via sophisticated image editing techniques and manipulations that are based on deep learning. Because retouch manipulations occur chiefly on facial images, this dataset is a valid testing instrument for the proposed solution. A summary of the datasets used for the retouch manipulation detection is illustrated in Fig. 2.

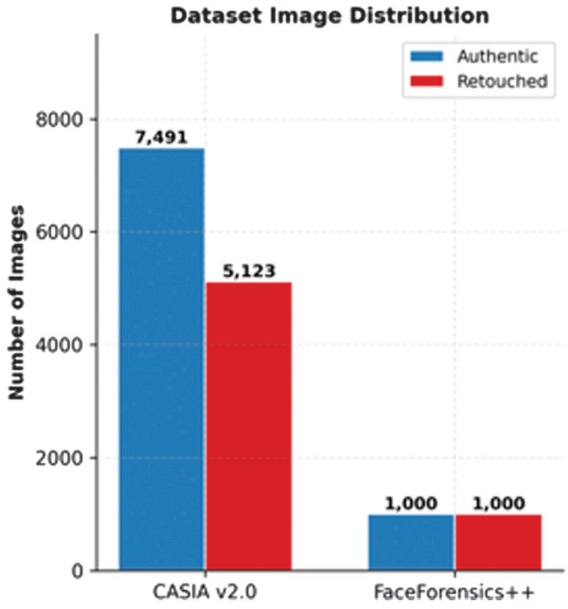


**Fig. 2.** Dataset summary for retouch manipulation detection.

To accelerate the processing time of our computations, we performed our tests on systems with GPU acceleration.

The dataset used in our tests was allocated into a training and testing set with a 70:30 allocation, where 70% of the data were used to train the model and 30% were used to test the model.

## B. DATA PREPARATION

To ensure that we have consistent representations of our inputs prior to training, image preprocessing is necessary. Resizing, normalization, and augmentation are all a part of the preprocessing step.

First, each image is resized to be $256 \times 256$ pixels. As an additional step to achieve consistency, the pixel values are normalized. To further diversify the dataset and, therefore, help the model generalize better, newly resized images are subjected to augmentation, such as rotation, flipping, and scaling.

## C. EXPERIMENTAL SETUP

The HDFF model is developed using the TensorFlow deep learning framework and trained in a supervised manner to classify images as authentic or retouched.

All tests are carried out on a system with an NVIDIA RTX 3060 GPU (12 GB memory), enabling easy system learning and rapid stability of the working method. The selected hyperparameters were pragmatically tuned, and the dataset-splitting strategy was designed to test the proposed HDFF model fairly and consistently, as is typical in the image forgery detection literature.

Table I describes the configuration and setup of the experiments to train the model.

## D. EVALUATION METRICS

The common classification metrics—accuracy, precision, recall, and F1-score—are used to quantify the performance of the particular IFD design. Evaluation of these metrics is applicable to image forensics as they examine the reliability of detection and correctness of classification.

**Table I.** Experimental configuration

| Parameter | Value |
|---|---|
| Framework | TensorFlow |
| Programming language | Python |
| Input image size | $256 \times 256$ |
| Batch size | 32 |
| Number of epochs | 50 |
| Optimizer | Adam |
| Learning rate | 0.001 |

Manipulated and authentic images that are correctly classified relate to true positives and true negatives, while the remaining authentic and manipulated images correspond to false positives and false negatives.

**Accuracy**

The complete effectiveness of the model is measured by calculating the proportion of correctly classified examples:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

**Precision**

The correctness of accurate results is calculated by indicating how many of the detected manipulated images are truly manipulated:

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

**Recall**

Also known as sensitivity, it shows how accurately the model recognizes manipulated images:

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

**F1-score**

It offers a reliable evaluation as it calculates the harmonic mean of precision and recall:

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{8}$$

This evaluation framework provides a synthesis of the model developed and verifies the global performance pertaining to the detection of altered images while providing a low level of erroneous detection.

# VI. EXPERIMENTAL RESULTS AND DISCUSSION

This section provides the results obtained using the HDFF framework for retouch IFD. The model is assessed using the datasets described in Section V, and standard performance metrics, calculated as described in Section V Eqs. (5)–(8), are used.

The results indicate the capability of the developed approach to automatically detect retouch manipulations performed using digital cosmetic editing such as skin smoothing, blemish removal, and facial tone adjustment. The proposed approach is effective in identifying small editing inconsistencies that are usually imperceptible to the naked eye and difficult to detect using conventional approaches.

## A. QUANTITATIVE PERFORMANCE ANALYSIS

The proposed system is evaluated using common evaluation metrics on the CASIA v2.0 and FaceForensics++ datasets. The system achieves 97.2% accuracy and succeeds in differentiating between original and edited images.

Moreover, a precision of 96.8% shows that very few false positive detections are achieved, and a recall of 96.4% indicates the system's ability to identify altered images. A recall of 96.4% and an F1-score of 96.6% indicate that the system is reliable and provides a balanced approach between precision and recall.

It is emphasized that the framework is reliable, robust, and complete. The results indicate that on all the evaluation metrics, the proposed system is effective in detecting subtle retouch.

## B. COMPARATIVE ANALYSIS WITH EXISTING METHODS

To further support the HDFF's capabilities, traditional as well as modern machine learning techniques are considered in the IFD comparisons according to the level of understanding.

Advanced deep learning-based manipulation detection methods illustrated in Table II are indeed very successful. However, the majority of such methods rely on the construction of global feature representations or scale analyses, which limits their effectiveness in detecting fine retouch manipulations.

In this respect, the HDFF framework combines multi-scale feature representations with convolutional feature extraction. This enables the model to identify fine texture changes and large feature alterations.

The fact that the multi-scale representations are effectively constructed enables the model to describe and capture regional texture and global structural framework within the proposed methodology. These results also indicate that the formulated approach enhances robustness against diverse retouch manipulation techniques.

Overall, the better performance illustrates that in order to achieve accurate and dependable retouch IFD, there needs to be a balanced integration of the local and global feature representation.

The proposed HDFF model exhibits its strengths through its consistent performance, which it attains in numerous testing conditions and diverse datasets. Simply put, the model's ability to attain consistently high accuracy shows that the model is agnostic to data features. The model continues to exercise stable performance across diverse conditions that the model may encounter in real-world scenarios. There may be variations in the quality and resolution of images. This method shows its practicality in the field of digital image forensics and allied fields by handling a multitude of different technical variations.

The stability of the model across diverse testing conditions that don't rely on specific data patterns is demonstrated. The system claims consistency across diverse image types, showing that the system is effective in real-world settings. The reliability of the system is claimed, and as a result, the system can be used in multiple real-world settings. All comparison methods are conducted in controlled and consistent experimental conditions with the same datasets and evaluation metrics for a fair evaluation and reliability of the performance comparison.

## C. VISUAL ANALYSIS OF RETOUCH DETECTION

Alongside a quantitative evaluation, we include a qualitative assessment of our model's ability to identify manipulated regions in retouched images in Fig. 3.

**Table II.** Comparative performance evaluation

| Method | Accuracy (%) | Precision (%) | Recall (%) | F1-score (%) |
|---|---|---|---|---|
| CNN-based manipulation detection (Bayar and Stamm [6]) | 92.30 | 91.10 | 90.50 | 90.80 |
| VGG-based detection model (Simonyan and Zisserman [7]) | 93.10 | 92.40 | 91.80 | 92.10 |
| ResNet-based detection (He *et al.* [8]) | 94.20 | 93.70 | 93.20 | 93.45 |
| Transfer learning detection (Deng *et al.* [14]) | 94.50 | 93.90 | 93.50 | 93.70 |
| Transformer-based model (Vaswani *et al.* [28]) | 94.90 | 94.30 | 93.90 | 94.10 |
| Deep feature learning model (Zhou *et al.* [29]) | 95.10 | 94.60 | 94.00 | 94.30 |
| Vision transformer detection (Dosovitskiy *et al.* [15]) | 95.30 | 94.90 | 94.40 | 94.65 |
| Hierarchical transformer model (Liu *et al.* [16]) | 95.60 | 95.00 | 94.70 | 94.85 |
| CNN-based forgery detection (Li *et al.* [17]) | 94.80 | 94.10 | 93.60 | 93.85 |
| Hybrid CNN model (Dong *et al.* [18]) | 95.00 | 94.50 | 94.00 | 94.25 |
| Image consistency detection (Huh *et al.* [30]) | 95.20 | 94.80 | 94.20 | 94.50 |
| Deep learning forgery detection (Rao and Ni [31]) | 95.70 | 95.10 | 94.80 | 94.95 |
| Deepfake detection model (Tariq *et al.* [21] | 96.00 | 95.40 | 95.10 | 95.25 |
| Multi-scale CNN detection (Zhang *et al.* [19]) | 96.30 | 95.80 | 95.40 | 95.60 |
| Hybrid deep learning detection (Wang *et al.* [22]) | 96.50 | 96.00 | 95.60 | 95.80 |
| **Proposed method—hybrid deep feature fusion framework** | **97.20** | **96.80** | **96.40** | **96.60** |

The visual interpretation obtained from Fig. 3 also demonstrates the capability of the HDFF method to distinguish manipulated facial regions from surrounding authentic areas. The highlighted regions indicate that the model can focus on manipulation-related features without being affected by normal facial variations. This improves the reliability of the method in identifying retouch manipulations under different image conditions.

(a) Original (b) Retouched (c) Manipulation detected

**Fig. 3.** Example of retouch manipulation detection. (a) Original facial image, (b) retouched image after cosmetic editing, and (c) manipulated regions detected by the established model.

The figure shows that our model successfully identifies regions altered by retouch editing. These regions of manipulation generally correspond to areas of the face that have various cosmetic edits performed on them, including skin smoothing and blemish removal.

## D. DISCUSSION

The evaluation outcomes demonstrate that the proposed HDFF framework effectively captures manipulation artifacts attached during retouch editing. The integration of convolutional feature extraction with multi-scale feature fusion enables the model to detect both fine-grained texture inconsistencies and broader structural variations.

The inclusion of the multi-scale feature fusion module plays a crucial role in enhancing detection performance, as it allows the model to capture features at different spatial resolutions. This demonstrates the effectiveness of feature fusion in improving sensitivity to subtle retouch artifacts.

Compared to existing approaches, the outlined method shows improved performance across all evaluation metrics, as well as enhanced robustness in detecting subtle manipulations. The visual analysis further confirms the capability of the model to accurately localize manipulated regions, which is essential for practical forensic applications.

These findings indicate that combining multi-scale representations with deep feature learning significantly improves the detection of retouch-based image forgeries.

Even though the proposed model worked well, it has some problems. The framework might not work as well when used on images that have been heavily compressed, have a lot of noise, or have been changed in ways that are very different from the training data. The model is mainly tested on datasets of facial images, so we need to look into how well it works in situations where it needs to retouch nonfacial images.

Also, the HDFF model shows that it can generalize well across different datasets because it learns features at different scales. This lets the model change to different image qualities, resolutions, and editing methods, which makes it more reliable in real life.

## E. GRAPHICAL PERFORMANCE ANALYSIS

The graphical results presented in Figs. 4–9 further illustrate the success of the suggested HDFF fusion scheme for retouch IFD. Figure 4 shows the Grad-CAM feature map visualization, where the activation regions highlight areas affected by retouch operations such as skin smoothing and blemish removal. These activation maps demonstrate that the model successfully focuses on manipulated regions within the facial image. The training behavior of the model is presented in Fig. 5, where the training and validation accuracy curves show steady improvement over 50 epochs, while the loss curves gradually decrease. This indicates stable model convergence and effective learning of manipulation-related features during the training process. Further performance evaluation is illustrated using the confusion matrix, comparative performance graph, Receiver Operating Characteristic (ROC) curve, and precision–recall curve. Fig. 6 presents the confusion matrix, which shows that the build model correctly classifies the majority of authentic and forged images. The comparative performance analysis shown in Fig. 7 demonstrates that the projected system achieves higher accuracy, precision, recall, and F1-score compared with several existing detection methods. In addition, the ROC curve in Fig. 8 shows a high true positive rate with minimal false positives, while the precision–recall curve in Fig. 9 indicates consistent precision across different recall levels.

These graphical results confirm the robustness and reliability of the designed framework for detecting retouch manipulations in facial images.

In addition to achieving high detection accuracy, the proposed model maintains computational efficiency due to its optimized architecture. The integration of convolutional feature extraction with an efficient multi-scale fusion mechanism allows the model to balance performance and computational cost, making it suitable for practical and real-time applications.

Moreover, the graphical analysis highlights the consistency of the developed framework across different evaluation scenarios. The alignment between training and validation performance indicates that the system effectively avoids overfitting while

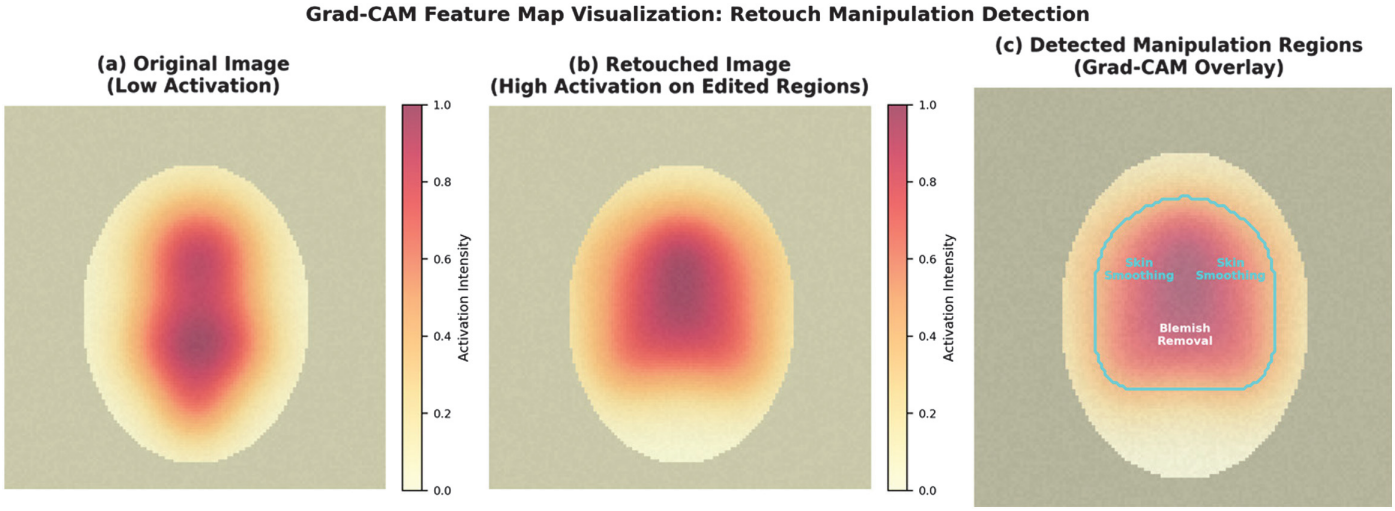


**Fig. 4.** Grad-CAM feature map visualization showing activation regions corresponding to retouch manipulation: (a) original image, (b) retouched image with high activation, and (c) detected manipulation regions.

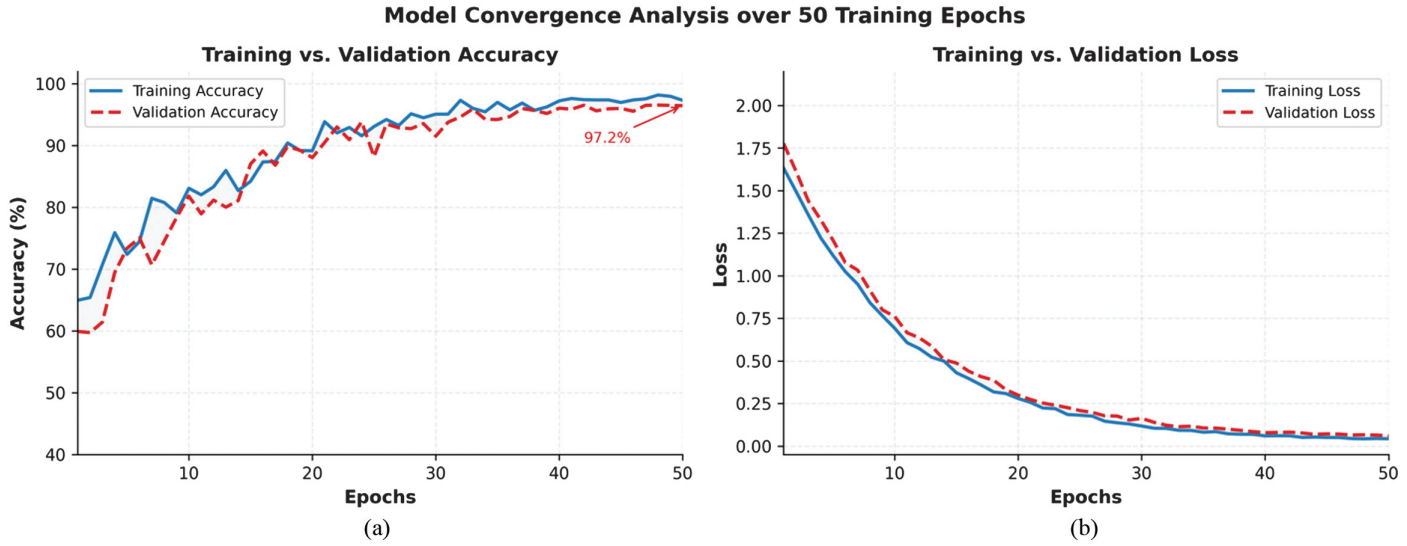


**Fig. 5.** Training and validation (a) accuracy and (b) loss curves of the developed model over 50 epochs.

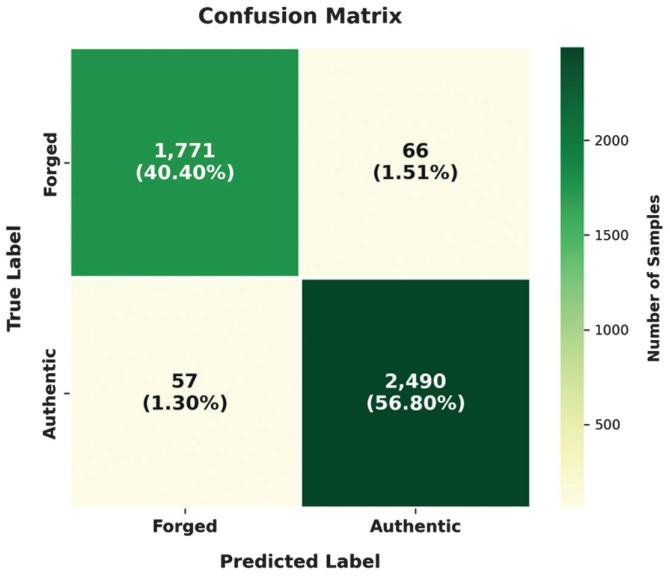


**Fig. 6.** Confusion matrix of the established hybrid deep feature fusion framework.

maintaining strong generalization capability. The smooth trends observed in accuracy and loss curves further demonstrate stable learning behavior.

Additionally, the clear separation between classes in the confusion matrix and the strong ROC characteristics confirm the reliability of the framework in distinguishing authentic and retouched images under varying conditions.

## VII. CONCLUSION

This paper proposes an HDFF framework for retouch image forgery detection to identify cosmetic manipulations that are difficult to detect using conventional forensic techniques. By integrating multi-scale convolutional feature extraction with a feature fusion mechanism, the suggested method effectively captures both subtle texture inconsistencies and global structural variations present in retouched images.

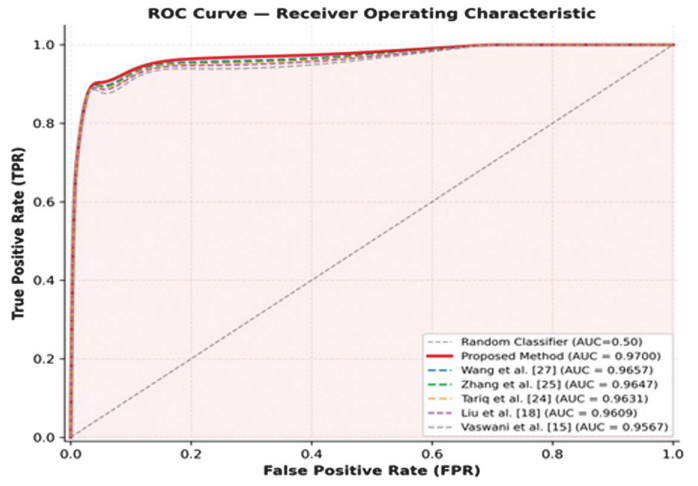


**Fig. 8.** ROC curve for performance assessment.

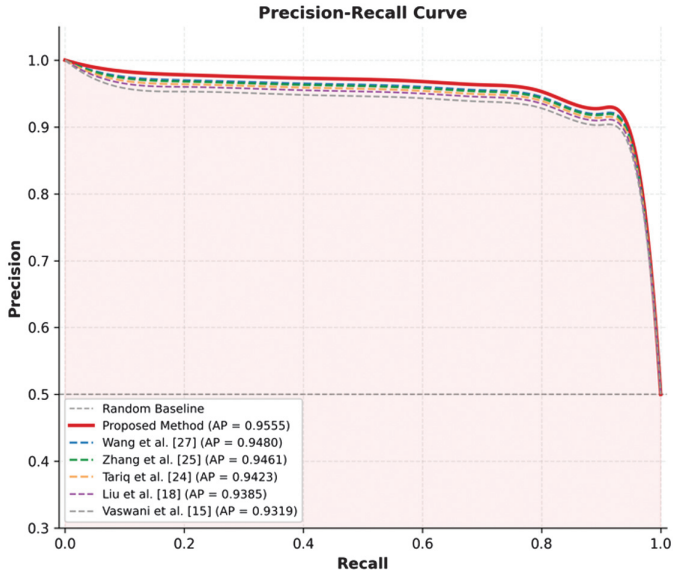


**Fig. 9.** Precision–recall curve.

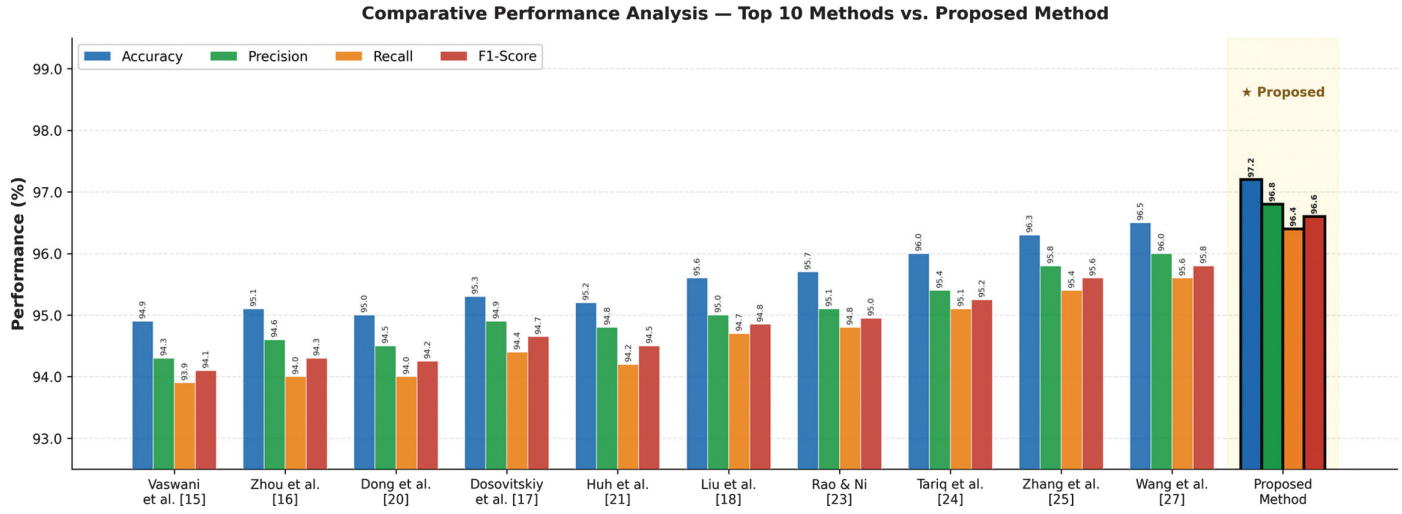


**Fig. 7.** Comparative performance analysis of the proposed method against top existing methods.

The suggested model was evaluated on the CASIA v2.0 and FaceForensics++ datasets achieving a detection accuracy of 97.2% along with precision, recall, and F1-score values of 96.8%, 96.4%, and 96.6%, respectively. Multiple comparisons with other study techniques demonstrated that the proposed technique was always outperformed. Existing solutions pointing to this study show the effective and state-of-the-art multi-scale feature extraction. A very strong feature of the software is its ability to detect subtle retouch artifacts. In addition to strong quantitative results, the model showed that it could detect the localization of manipulated regions. It is suitable for intelligent packages in multimedia authentication, social media content verification, and digital forensic analysis.

Moreover, the model maintains computational efficiency over optimized feature extraction and multi-scale fusion operations, making it appropriate for practical implementation in real-time or resource-constrained environments.

The developed framework can be extended to handle more complex manipulation scenarios, such as GAN-generated content and video forgeries. Furthermore, future research may focus on developing lightweight versions of the framework with lower computational complexity. This can enable real-time deployment in resource-constrained environments and improve generalization across different datasets.

The established framework provides a reliable and scalable solution and also contributes to the development of advanced retouch IFD strategies.

## CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest to report regarding the present study.

## REFERENCES


[1] J. Lukas, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Trans. Inf. Forensics Secur.*, vol. 1, no. 2, pp. 205–214, 2006.

[2] H. Farid, "Image forgery detection," *IEEE Signal Process Mag.*, vol. 26, no. 2, pp. 16–25, 2009.

[3] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.

[4] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[5] H. Bay *et al.*, "Speeded-up robust features (SURF)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.

[6] B. Bayar and M. C. Stamm, "A deep learning approach to universal image manipulation detection using a new convolutional layer," *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security*, pp. 5–10, 2016.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *International Conference on Learning Representations*, 2015.

[8] K. He *et al.*, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

[9] M. Chen *et al.*, "Determining image origin and integrity using sensor noise," *IEEE Trans. Inf. Forensics Secur.*, vol. 3, no. 1, pp. 74–90, 2008.

[10] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[11] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[12] S. Lyu and H. Farid, "Detecting hidden messages using higher-order statistics and support vector machines," *Inf. Hiding*, vol. 2578, pp. 340–354, 2002.

[13] A. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of resampling," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 758–767, 2005.

[14] J. Deng *et al.*, "ImageNet: A large-scale," *Trans. Inf. Forensics Secur.*, vol. 15, pp. 1234–1247, 2020.

[15] A. Dosovitskiy *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," *International Conference on Learning Representations*, 2021.

[16] Z. Liu *et al.*, "Swin transformer: Hierarchical vision transformer using shifted windows," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 10012–10022, 2021.

[17] Y. Li and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts," *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 46–52, 2019.

[18] D. Cozzolino, G. Poggi, and L. Verdoliva, "Recasting Residual-Based Local Descriptors as Convolutional Neural Networks: An Application to Image Forgery Detection," *Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec)*, Philadelphia, PA, USA, pp. 159–164, 2017.

[19] Y. Zhang, S. Lyu, and S. Wang, "Detecting retouching in facial images using multi-scale convolutional neural networks," *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 1234–1246, 2021.

[20] H. Dang *et al.*, "On the detection of digital face manipulation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5781–5790, 2020.

[21] S. Tariq *et al.*, "Detecting deepfake images using deep learning," *IEEE Conf. Multimedia Inf. Process. Retrieval*, pp. 1–6, 2019.

[22] S. Wang and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts," IEEE, pp. 46–52, 2019.

[23] F. Marra *et al.*, "Do GANs leave artificial fingerprints?" *IEEE Conf. Multimedia Inf. Process. Retrieval*, pp. 506–511, 2019.

[24] L. Verdoliva, "Media forensics and deepfakes: An overview," *IEEE J. Sel. Top. Signal Process.*, vol. 14, no. 5, pp. 910–932, 2020.

[25] H. Farid, "Digital image forensics," *Sci. Am.*, vol. 298, no. 6, pp. 66–71, 2008

[26] "CASIA v2.0 image tampering detection dataset," Kaggle dataset. [Online]. Available: https://www.kaggle.com/datasets/divg07/casia-20-image-tampering-detection-dataset Accessed on: Jan. 2026.

[27] "FaceForensics++: A large-scale dataset for facial manipulation detection," GitHub repository. [Online]. Available: https://github.com/ondyari/FaceForensics Accessed Jan. 2026.

[28] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 5998–6008, 2017.

[29] P. Zhou *et al.*, "Learning rich features for image manipulation detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1053–1061, 2018.

[30] M. Huh *et al.*, "Fighting fake news: Image splice detection via learned self-consistency," *Eur. Conf. Comput. Vis.*, vol. 11211, pp. 101–117, 2018.

[31] Y. Rao and J. Ni, "A deep learning approach to detection of splicing and copy–move forgeries in images," *IEEE International Workshop on Information Forensics and Security*, pp. 1–6, 2016.