Information Sciences Letters
*An International Journal*

http://dx.doi.org/10.18576/isl/130212

# Deepfake Image Forensics: Towards Efficient and Reliable Detection

*Manar M. Hafez*[*], *Mamdouh Mohamed, Youssef Hesham, Mahmoud M. Gomaa, Ghidan Sadek and Alyaa Amer*

Department of Software Engineering, College of Computing and Information Technology, Arab Academy for Science, Technology and Maritime Transport, Giza, Egypt

Received: 15 Dec. 2023, Revised: 24 Jan. 2024, Accepted: 5 Feb. 2024
Published online: 1 Mar. 2024

**Abstract:** In an era marked by the exponential growth and sophistication of deepfakes, heightened concerns among the public about the increasing role of these manipulative digital creations in disseminating disinformation have become evident. The surge in AI-driven research focusing on deepfakes, spanning creation methods, detection techniques, and datasets, underscores the urgency to address the challenges posed by these deceptive technologies. While deep learning methods have proven successful in identifying and preventing deepfakes, the rapid evolution of deep-fake technology continues to present critical challenges, especially in the realm of image forgery detection. To address these challenges head-on, our paper introduces an advanced deepfake image forensics approach leveraging Convolutional Neural Networks (CNN), incorporating ResNet50, DenseNet121, and InceptionV3. These three prominent deep learning models are strategically combined in an innovative ensemble stacking methodology, surpassing traditional approaches to significantly enhance detection performance. High-level features are meticulously extracted from two distinct base models, DenseNet121 and InceptionV3, which are then synergistically combined and input into a Logistic Regression (LR) meta-model, serving as a robust classifier for final predictions. To ensure the reliability of our proposed approach, we conducted extensive training and fine-tuning using a large dataset comprising 70,000 real human faces from the Flickr dataset. This dataset was enriched with an additional 70,000 manipulated faces generated through the style-GAN technique. Through comprehensive evaluation, our approach demonstrated an exceptional accuracy of 98.7% in detecting image forgeries. Additionally, the Error Level Analysis (ELA) technique was applied to the dataset and employed across the three deep learning models. This advanced deepfake image forensics system stands as a significant contribution to ongoing efforts in combating deepfake risks. By empowering users to authenticate digital images and detect potential manipulations, it plays a pivotal role in safeguarding against misinformation and fraudulent activities.

**Keywords:** Deep-fake, Image forensics, Ensemble stacking, Convolutional Neural Networks (CNN), ResNet50, DenseNet121, InceptionV3, Error Level Analysis (ELA).

## 1 Introduction

In our digitally driven society, the human face reigns supreme as our species' greatest distinguishing feature. This distinction, however, has become the focal point of a technological revolution, giving rise to the multidimensional phenomena of deepfakes [1]. Face synthesis technology has grown at an unpredictable rate, and the security vulnerabilities associated with face modification have gained attention. Individuals' faces can now be smoothly switched with those of others thanks to deep learning algorithms, generating a misleading authenticity that tests our discernment. Deepfake technology, which is part of the broader subject of artificial intelligence, propels the overlay of one person's face onto another's, which is choreographed by a symphony of algorithms, particularly generative adversarial networks (GANs) [2]. As a result, high-resolution deepfake images blur the border between reality and deception.

The widespread use of cell phones and the exponential proliferation of social networking platforms has accelerated the spread of deepfake content, transforming it into a global threat. Not long ago, the human eye was capable of detecting deepfake images due to indicative signs such as the" pixel collapse phenomenon," [3] which causes artificial visual inconsistencies in skin tones or facial contours. Deepfake technology, on the other hand, has evolved exponentially, with previously noticeable seems now expertly concealed by the technology that created them. This technology has seen extensive usage in the unauthorized placement of famous Hollywood celebrities' faces onto explicit imagery and videos. Furthermore, it has been employed to generate deceptive content and spread rumors concerning political figures [4–6]. For instance, in 2018, a fabricated video featuring Barack Obama was circulated, falsely attributing words to him

[*]Corresponding author e-mail: m.mohamed@aast.edu

that he never actually spoke [7]. Moreover, deepfakes have been employed in the context of the 2020 US election to manipulate videos of Joe Biden, including instances where his tongue appeared to be sticking out. Also, celebrities and public figures like Taylor Swift, Gal Gadot, Emma Watson, Meghan Markle, and many others have turned into the main targets of the potentially harmful effects of deepfake technology.

Moreover, Deepfake victimization [8] has tentacles that reach across the United States and Asian societies, victimizing women and contributing to a culture life with misinformation, particularly on social media. As a result, deepfakes have emerged as a formidable threat to today's generation, undermining the very foundations of trust and credibility. Therefore, these malicious applications of deepfakes pose a significant threat to our society and can contribute to the dissemination of misleading information. As society grapples with the consequences of defamation, deception, scams, and insecurity. Therefore, researchers and other several multinational corporations have risen to the occasion, relentlessly pursuing solutions to detect deepfakes. For instance, Facebook Inc. and Google launched a research initiative in attempting the detect and prevent deepfakes [9, 10].

Therefore, in the hope of contributing to the ongoing fight against image forgery and verifying the authenticity of digital images, our work primarily focuses on developing three CNN architectures—DenseNet121, Inception V3, and Resnet50—. Then developing a stacked ensemble model that combines features from two effective classification models; DenseNet-121 and InceptionV3 to robustly distinguish between fake and real images from a large dataset.

The rest of the paper is organized as follows: Section 2 provides a summary of the relevant deep-learning detection techniques in the domain. Section 3 describes the dataset used for evaluating and testing the model. Section 4, is a detailed description of the methodology, followed by Section 5 which describes the evaluation metrics and results. Finally, Section 6 demonstrates the discussion and some conclusive points.

## 2 Related Work

In the rapidly changing field of deep learning technologies, the rapid evolution has brought about significant advancements as well as significant challenges in various areas. One particularly pressing challenge is the widespread emergence of deepfakes, which are highly sophisticated AI-generated media that convincingly mimic genuine content. These deepfakes pose significant risks to the authenticity and integrity of digital information. [11, 12].

Deepfakes are driven by deep neural networks and generative adversarial networks (GANs) [13], they have evolved from basic manipulations to highly realistic fabrications. While they were initially limited to entertainment purposes, these manipulated media have expanded their reach to disinformation campaigns, cyber threats, and violations of privacy [14]. As a result, there is an urgent need for a comprehensive and coordinated approach to address this issue.

The increasing complexity of editing tools has led to a significant rise in the study of passive digital image forgery detection in recent research [15]. Researchers have explored both conventional digital forensic techniques and modern data-driven approaches, with a particular emphasis on deep neural networks (DNNs) [16, 17]. Video forgery detection has specifically addressed spatiotemporal inconsistency analysis indicative of manipulations [18, 19]. Despite progress, precise localization of manipulated regions within media remains a challenge. Ensuring model robustness to diverse real-world manipulation methods is a critical focus [20, 21].

Adversarial training strategies, augmenting datasets with challenging forged examples, are widely used to improve model robustness and generalization capability [22, 23]. However, challenges persist in handling previously unseen manipulation methods, leading to the exploration of few-shot meta-learning approaches for quick adaptation [24].

Many researches have been conducted to investigate the incorporation of biometric anti-spoofing systems and forensic applications, but there are persistent challenges in establishing confidence in passive forensic authentication systems [25, 26]. As digital media becomes increasingly prevalent in society, ongoing research focuses on various areas, such as hybrid detection approaches, enhanced localization, model resilience, reliable datasets, and real-time system implementation [27, 28]. The growing concern regarding security and privacy risks associated with deepfakes has led to the development of numerous methods for their detection and mitigation [23, 29].

The originality of this paper lies in its purposeful choice of three carefully designed CNN architectures. — DenseNet121, Inception V3, and Resnet50— reflecting a commitment to a comprehensive and scientifically rigorous investigation. A comparative analysis of CNN-based techniques, highlighted in the current work, provides empirical evidence and insights into various CNN architectures [29, 30]. Our literature review provides a comprehensive overview of the current state of passive digital image forgery detection, with a specific focus on the escalating threats posed by deepfakes. The highlighted studies contribute to understanding manipulation techniques, model robustness, and emerging challenges in real-world scenarios, setting the stage for our unique contributions.

In our paper, we tackle the pressing need for improved deepfake detection methods by employing an approach that

involves the competent selection of three different CNN architectures, namely DenseNet, Inception V3, and Resnet50. To enhance the resilience of our detection model, we incorporate the Error Level Analysis (ELA) technique prior to applying the CNN architectures. Going beyond individual models, our methodology involves the development of a stacked ensemble model using logistic regression, which combines the strengths of multiple algorithms to enhance overall detection performance. These contributions establish our work as a valuable contribution to the current discussion on deepfake detection.

## 3 Dataset

The dataset used in this research was sourced from Kaggle, encompassing a comprehensive collection of 70,000 real human faces retrieved from the Flickr dataset, meticulously curated by Nvidia Corporation. This dataset was augmented with an additional 70,000 fake images, extracted from a pool of one million fabricated faces generated through the style-GAN technique. The main motive behind the integration of these two distinct datasets, is to increase the diversity of the training data, and this helps the model generalize better to unseen data, improves its robustness, and reduces overfitting. All images underwent a uniform resizing process, resulting in a standardized resolution of 256x256 pixels.

The collected dataset was divided into three distinct sets: training, validation, and test sets. This partitioning scheme involved allocating 70% of the dataset for training, 15% for validation, and 15% for testing. Specifically, the training set consisted of a total of 100,000 images, evenly split between 50% real images and 50% fake images. The validation set included 20,000 images, balanced between fake and real images. As for the test set, it comprised 20,000 images, equally divided into real and fake images.

Generally, the main rationale behind this dataset division strategy is to ensure that the model can generalize well to new data, prevent overfitting, allow for early stopping, and provide an unbiased estimate of the model's performance. Also, we employed data augmentation as the groundwork for the precise evaluation and analysis of the proposed deepfake detection models. A sample of the dataset is seen in Figures 1 and 2.

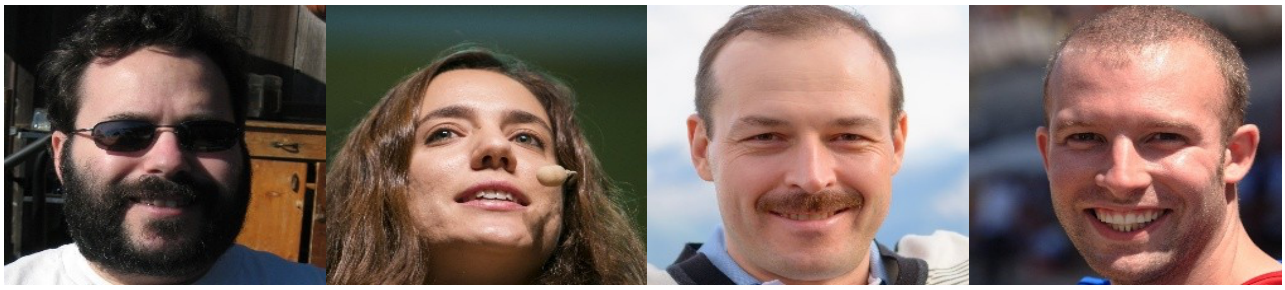

**Fig. 1:** The real images.

**Fig. 2:** The fake images.

## 4 Methodology

Our automated deepfake image forensics model depicted in Figure 3 is divided into the following stages; (1) A preprocessing stage which is split into two steps: data preparation and Error Level Analysis (ELA). This process helps to enhance data quality by identifying error levels in images. (2) A model learning stage, where we employ 3 CNN architectures, DenseNet121, Inception V3, and ResNet50. (3) A model ensemble stage, which is a judgmental analysis between the CNN approaches. This stage employs both stacking and average element-wise attributes. Through stacking, individual models are seamlessly integrated to form a meta-model using logistic regression, capturing diverse insights and augmenting overall predictive capability. (4) A model evaluation stage, which considers some performance metrics such as accuracy, recall, precision, and F1-score.

In the following sections, we will provide a comprehensive explanation of the integrated algorithms in our model. We

will also provide a detailed explanation of the model to deepen the understanding of the complexities involved and pave the way for future developments in this ever-evolving field.

### 4.1 Error Level Analysis (ELA)

Error Level Analysis (ELA) [31, 32] works by saving an image again with a set error rate, usually 95%, and then finding the differences between the original and the changed images. When minimal change is observed, it indicates that the pixels have reached their local minima for error at that quality level. Conversely, a substantial alteration signals that the pixels are not at their local minima and essentially remain in their original state. ELA reveals distinct error-level variations across an image, strongly suggesting some form of digital manipulation. Specific areas of interest include the lips, shirt, and eyes, all of which exhibit significantly different error levels compared to their surroundings. This discrepancy implies potential alterations in colors and brightness within these regions.

### 4.2 Convolutional Neural Network

CNN [33] is a specialized category within machine learning, belonging to the broader family of artificial neural networks employed across various applications and data formats. Within the realm of deep learning algorithms, CNN stands out as a distinctive network architecture designed specifically for tasks involving image recognition and the manipulation of pixel data.

While deep learning encompasses diverse neural network types, CNNs excel in the identification and classification of objects, making them the architecture of preference. This unique suitability positions CNNs as particularly effective in applications demanding computer vision, including pivotal scenarios like object recognition in self-driving cars and facial recognition technology.

#### Residual neural network (ResNet50)

Residual neural network (ResNet-50) [34] is a convolutional neural network comprising of 50 layers, serving as a foundational architecture for numerous computer vision tasks. ResNet is a classic neural network architecture widely employed as a backbone in various computer vision applications. The key innovation introduced by ResNet is its capability to facilitate the training of exceptionally deep neural networks, surpassing 150 layers, which marked a significant breakthrough in the field. Also, the presence of skip connections in ResNet enables the training of extremely deep networks, eases the training process, and prevents overfitting, allowing the model to generalize well to unseen data.

#### Inception-v3

Inception-v3 [35] is a convolutional neural network architecture hailing from the Inception family, characterized by notable enhancements. These improvements encompass the implementation of label smoothing, the utilization of factorized 7 x 7 convolutions, and the incorporation of an auxiliary classifier. The latter serves the purpose of propagating label information to lower levels of the network. Additionally, the model incorporates batch normalization for layers within the seedhead. These refinements collectively contribute to the efficiency and effectiveness of Inception-v3 in handling complex visual recognition tasks.

#### DenseNet

Dense Convolutional Network (DenseNet) [36] is a deep learning network that enhances training efficiency by increasing the depth of deep learning layers. It also establishes compact connections between layers. Each layer is intricately connected to all subsequent layers in the network. The connectivity pattern ensures that the first layer links to the 2nd, 3rd, 4th, and so forth, fostering optimal information flow across the network.

Unlike ResNets, which aggregate features through summation, DenseNet employs concatenation to combine features. Each layer receives input from all preceding layers and transmits its own feature maps to all subsequent layers. Consequently, DenseNet requires fewer parameters than conventional convolutional neural networks, eliminating the need to learn extraneous feature maps.

### 4.3 Logistic Regression (LR)

Logistic Regression (LR) [37] is a model that predicts the probability of a discrete outcome based on input variables. Typically applied to binary outcomes (e.g., true/false, yes/no), this method is widespread. However, multinomial logistic regression extends its applicability to scenarios with more than two possible discrete outcomes. It serves as a valuable analytical tool, especially in classification problems where the goal is to ascertain the best-fit category for a new sample.

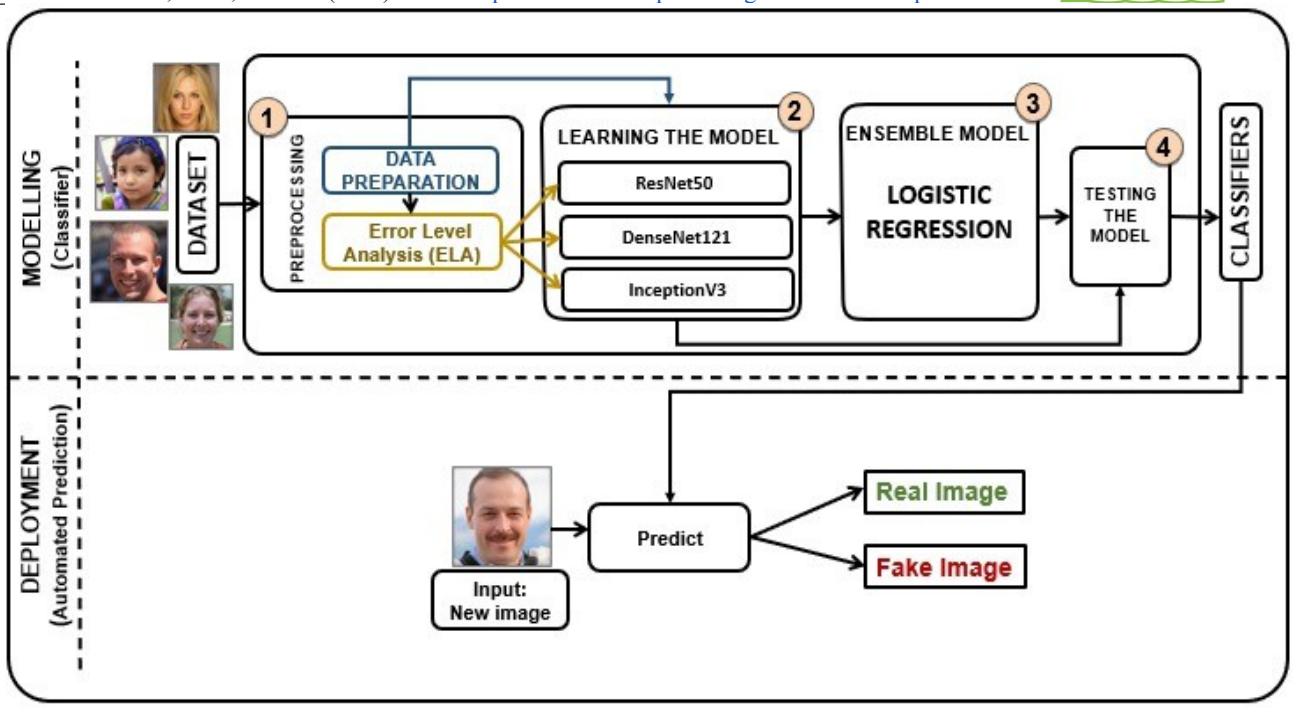


**Fig. 3:** Our proposed Model

## 5 Evaluation and Results

For implementing the model, we utilized the Python programming language, along with popular libraries and frameworks such as TensorFlow, Keras, and scikit-learn. The chosen implementation environment was Spyder, providing a user-friendly IDE for efficient coding and experimentation. The evaluation and results were calculated using performance parameters such as accuracy, precision, recall, and F1-Score [38].

Accuracy is a key metric for evaluating classification models. Informally, it represents the fraction of correct predictions made by our model. Formally, accuracy is defined by the following equation:

$$Accuracy = \frac{\#\ of\ correct\ predictions}{Total\ \#\ of\ predictions} \tag{1}$$

Accuracy can also be calculated in terms of positives and negatives, as illustrated by Equation 2.

$$Accuracy = \frac{\mathrm{TP+TN}}{TP+TN+FP+FN} \tag{2}$$

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

Precision, also known as positive predictive value, measures how well the model predicts positive values out of all the positive values predicted by the model, as expressed in Equation 3.

$$\mathrm{Precision} = \frac{\mathrm{TP}}{TP+FP} \tag{3}$$

Recall is employed to assess how effectively the model identifies true positives. A high recall indicates that the model excels at recognizing true positives. Conversely, a low recall value suggests that the model experiences a considerable number of false negatives, as illustrated in Equation 4.

$$\mathrm{Recall} = \frac{\mathrm{TP}}{TP+FN} \tag{4}$$

F1-Score represents the harmonic mean of precision and recall, offering a more robust estimate than the accuracy metric for wrongly categorized cases. This is expressed in Equation 5.

$$\mathrm{F1-Score} = 2 * \frac{\mathrm{Precision*Recall}}{\mathrm{Precision+Recall}} \tag{5}$$

To assess the performance of our model, we will demonstrate the outcome of the following two processes: (i) Directly employing three Convolutional Neural Network (CNN) models—ResNet50, DenseNet121, and Inception-V3 on a dataset comprising 140,000 genuine and manipulated facial images, detailed in Table 1. (ii) Utilizing ELA as a

preprocessing step prior to inputting the data into the model, as depicted in Table 2.

The results outlined in Table 1 are obtained by training each model for 10 epochs. From the table, we can observe the effectiveness of different architectures in distinguishing real from fake faces. Upon careful examination, DenseNet121 and InceptionV3 emerge as top performers, achieving impressive accuracy scores of 98.405% and 98.105%, respectively. Notably, DenseNet121 stands out with the highest precision, recall, and F1-score, solidifying its role as the primary model for this specific task. These findings emphasize the pivotal role of CNN architecture selection, even with extensive datasets, in achieving superior image classification results.

The table also shows the classification accuracy of the ensemble average element-Wise model, exhibiting a notable increase by 0.3%, reaching 98.7%. The model combines the strengths of both DenseNet121 and InceptionV3 using a weighted averaging method, and is trained only for 10 epochs, with data augmentation. This outperformance shows the capability of the ensemble model for providing further enhancement.

The stacked ensemble model, another aspect of our evaluation, utilized the capabilities of DenseNet121 and InceptionV3 by forwarding the output to the logistic regression (LR). To elaborate further, the base models (InceptionV3 and DenseNet121) are first trained on the original 140k dataset, and their predictions are utilized as input features. These predictions from the base models on the original dataset then function as input features for our stacking dataset. Each base model's predictions create distinct feature columns in the stacking dataset, and we incorporate the true target label associated with the predictions from the original dataset. This model attained remarkable results, scoring the highest in recall and F1-score at 98.9% and 98.7%, respectively. In comparison with DenseNet121 alone, the stacked ensemble model achieved notable increases in accuracy, precision, recall, and F1-score by 0.25%, 0.1%, 0.5%, and 0.3%, respectively. Similarly, when compared with InceptionV3, the stacked ensemble model demonstrated improvements in accuracy, precision, recall, and F1-score by 0.55%, 0.4%, 0.8%, and 0.6%, respectively. These results highlight the efficacy of ensemble model in consistently delivering better performance metrics, showing its potential for enhancing classification capabilities.

**Table 1:** Results of the implemented models

| Models | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| **ResNet50** | 94.835% | 92.6% | 97.4% | 94.9% |
| **DenseNet121** | 98.405% | 98.40% | 98.405% | 98.404% |
| **InceptionV3** | 98.105% | 98.13% | 98.10% | 98.10% |
| **Ensemble Average Element-Wise** | **98.7%** | **98.6%** | 98.6% | 98.6% |
| **Stacked Ensemble Model** | 98.65% | 98.5% | **98.9%** | **98.7%** |

In our pursuit of refining the classification process, we use ELA as a preprocessing stage before inputting data into our models. ELA, commonly employed in image forensics, aims to identify regions of interest in digital images. How- ever, our findings, presented in Table 2, reveal an unexpected result. While ResNet50, DenseNet121, and InceptionV3 demonstrated significant performance when trained directly on the dataset for 10 epochs (as detailed in Table 1), the integration of the ELA resulted in a reduction in performance with all models. Notably, in ResNet50, there was an increase in precision of 1.33%, and the ensemble average element-wise model provided a slight improvement in precision by 0.03%. These findings suggest that ELA, which was adopted with the intention of improving the models' performance, actually had a negative impact on their classification abilities. Although ELA may be useful for certain image analysis tasks, its effectiveness in this image classification scenario appears to be limited. Therefore, it may be necessary to consider alternative approaches for enhancing our models' performance.

**Table 2:** Results of the models with ELA

| Models | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| **ResNet50** | 94.54% | 93.93% | 93.10% | 93.51% |
| **DenseNet121** | 95.88% | 95.53% | 96.26% | 95.90% |
| **InceptionV3** | 96.26% | 96.52% | 95.98% | 96.25% |
| **Ensemble Average Element-Wise** | 97.36% | **98.63%** | **98.51%** | **98.43%** |
| **Stacked Ensemble Model** | **97.89%** | 98.14% | 97.74% | 97.96% |

Finally, our proposed model demonstrates exceptional advantages in deepfake image forensics. With a commendable accuracy of 98.7%, the approach excels in accurately discerning manipulated images. The strategic use of ensemble models, particularly the stacked ensemble model, yields outstanding results in recall (98.9%) and F1-score (98.7%). A deliberate selection of ResNet50, DenseNet121, and InceptionV3 showcases a comprehensive approach to deep learning model choice. Despite challenges, the introduction of ELA reflects a commitment to exploring diverse image forensics techniques. Reliability is ensured through extensive training with a dataset of 70,000 real and manipulated faces. The incorporation of various evaluation approaches allows for a thorough understanding of model performance.

Despite mixed results, the inclusion of ELA signifies a willingness to experiment with different image analysis techniques. Overall, these advantages position the proposed methodology as a significant contribution to the evolving field of deepfake detection and image forensics, paving the way for further research and development.

## 6 Conclusions

This research tackles the crucial issues presented by the growing prevalence and escalating sophistication of deepfakes, which is causing concern among various stakeholders. Through an analysis of deepfake creation techniques, detection methods, and datasets, we were driven to offer a practical solution and establish a strong basis for further research endeavors within the field. Generally, our research builds on the success of deep learning methods to explore ways to address the challenges associated with deepfake images.

The proposed model combines three essential models: ResNet50, DenseNet121, and InceptionV3. Beyond individual models, we use ensemble techniques, including ensemble average element-wise and stacked ensemble models. Extracting features from DenseNet121 and InceptionV3, these features strengthen the classification process in a logistic regression meta-model within stacked ensemble models.

After a thorough evaluation, incorporating CNN-based models in an ensemble average-element-wise approach improved results by 0.3%. The stacked ensemble model further increased classification accuracy. This research has the potential to significantly contribute to combatting image fraud and manipulation, ensuring the integrity of digital visual content across diverse domains.

Finally, by sharing our findings and approach, we aim to establish a solid foundation for other researchers who are keen to explore the field of image forensics detection and contribute to its advancement.

Future work can explore the integration of advanced deep learning models, investigate ensemble techniques, and utilize transfer learning for further performance improvement. Optimizing parameters for enhanced efficiency and reduced processing time is another avenue for future exploration. Additionally, expanding the system's capabilities to include additional image forgery detection techniques, such as copy-move and splicing forgery detection, can broaden its applicability and enhance its effectiveness in detecting a wider range of image manipulations.

## Conflicts of Interest Statement

*The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.*

## References

[1] Westerlund, M. (11/2019 2019) 'The Emergence of Deepfake Technology: A Review', *Technology Innovation Management Review*. Ottawa: Talent First Network, 9, pp. 40–53. doi: 10.22215/timreview/1282.

[2] Zhu, J.-Y. *et al.* (2017) 'Unpaired image-to-image translation using cycle-consistent adversarial networks', in *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232.

[3] Shiba, S., Aoki, Y. and Gallego, G. (2022) 'Event Collapse in Contrast Maximization Frameworks', *Sensors*, 22(14). doi: 10.3390/s22145190.

[4] Nataraj, L., Mohammed, T.M., Manjunath, B.S., Chandrasekaran, S., Flenner, A., Bappy, J.H., Roy-Chowdhury, A.K. (2019). Detecting gan-generated fake images using co-occurrence matrices. Electronic Imaging 31(5), 532–15321 (2019). https://doi.org/10.2352/ISSN.2470-1173.2019.5.MWSF-532

[5] Wang, S.-Y., Wang, O., Zhang, R., Owens, A., Efros, A.A. (2020). Cnn-generated images are surprisingly easy to spot. . . for now. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8692–8701 (2020). https://doi.org/10.1109/CVPR42600.2020.00872

[6] Hsu, C.-C., Lee, C.-Y., Zhuang, Y.-X. (2018). Learning to detect fake face images in the wild. In: 2018 International Symposium on Computer, Consumer, and Control (IS3C), pp. 388–391 (2018). https://doi.org/10.1109/IS3C. 2018.00104

[7] Nataraj, L. *et al.* (2019) 'Detecting GAN generated fake images using co-occurrence matrices', *arXiv preprint arXiv:1903. 06836*. https://doi.org/10.1177/2056305120903408

[8] Stavola, J. and Choi, K.-S. (2023). Victimization by Deepfake in the Metaverse: Building a Practical Management Framework, *International Journal of Cybersecurity Intelligence & Cybercrime*, 6(2), p. 2.

[9] Kwok, A.O.J., Koh, S.G.M. (2021). Deepfake: a social construction of technology perspective. Current Issues in Tourism 24(13), 1798–1802 (2021). https://doi.org/10.1080/13683500.2020.1738357

[10] Westerlund, M. (2019). The emergence of deepfake technology: A review. Technology Innovation Management Review 9, 40–53 (2019). Chap. 40. https://doi.org/10.22215/timreview/1282

[11] Rana, M. S. *et al.* (2022) 'Deepfake Detection: A Systematic Literature Review', *IEEE Access*, 10, pp. 25494–25513. doi: 10.1109/ACCESS.2022.3154404.

[12] Guarnera, L., Giudice, O., Battiato, S. (2020). Deepfake detection by analyzing convolutional traces. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 666–667 (2020)

[13] Gong, D., Goh, O.S., Kumar, Y.J., Ye, Z., Chi, W. (2020). Deepfake forensics, an ai-synthesized detection with deep convolutional generative adversarial networks. Int J 9(3), 2861–2870

[14] Lyu, S. (2020). Deepfake detection: Current challenges and next steps. In: 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 1–6 (2020). IEEE

[15] Anwar, M.A., Tahir, S.F., Fahad, L.G., Kifayat, K. (2023). Image forgery detection by transforming local descriptors into deep-derived features. Applied Soft Computing 147, 110730

[16] Elaskily, M.A., Dessouky, M.M., Faragallah, O.S., Sedik, A. (2023). A survey on traditional and deep learning copy move forgery detection (cmfd) techniques. Multimedia Tools and Applications, 1–27

[17] Hu, Z. *et al.* (2023) 'An Attention-Erasing Stripe Pyramid Network for Face Forgery Detection'. Research Square. doi: 10.21203/rs.3.rs-2717775/v1.

[18] Abd El-Latif, E. I. and Khalifa, N. E. (2023) 'COVID-19 digital x-rays forgery classification model using deep learning', *Int J Artif Intell*, 12(4), pp. 1821–1827.

[19] Zhu, Y., Zhang, C., Gao, J., Sun, X., Rui, Z., Zhou, X. (2023). High-compressed deepfake video detection with contrastive spatiotemporal distillation. Neurocomputing 565, 126872

[20] Mehta, R., Kumar, K., Alhudhaif, A., Alenezi, F., Polat, K. (2023). An ensemble learning approach for resampling forgery detection using markov process. Applied Soft Computing 147, 110734

[21] Uma, S., Sathya, P. (2023). Forgery detection of digital images using teaching–learning–based optimization and principal component analysis. Sensing and Imaging 24(1), 1–16

[22] Zeng, P. *et al.* (2022) 'Multitask Image Splicing Tampering Detection Based on Attention Mechanism', *Mathematics*, 10(20). doi: 10.3390/math10203852.

[23] Agarwal, S., Cho, D.-J. and Jung, K.-H. (2023) 'Detecting Images in Two-Operator Series Manipulation: A Novel Approach Using Transposed Convolution and Information Fusion', *Symmetry*, 15(10). doi: 10.3390/sym15101898.

[24] Lin, Y.-K. and Yen, T.-Y. (2023) 'A Meta-Learning Approach for Few-Shot Face Forgery Segmentation and Classification', *Sensors*, 23(7). doi: 10.3390/s23073647.

[25] An, B., Lim, H. and Lee, E. C. (2023) 'Fake Biometric Detection Based on Photoplethysmography Extracted from Short Hand Videos', *Electronics*. MDPI, 12(17), p. 3605.

[26] Khudhair, Z.N., Mohamed, F., Rehman, A., Saba, T., et al. (2023). Detection of copy-move forgery in digital images using singular value decomposition. Computers, Materials & Continua 74(2)

[27] Patil, G., Shivakumara, P., Gornale, S.S., Pal, U., Blumenstein, M. (2023). A new robust approach for altered handwritten text detection. Multimedia Tools and Applications 82(14), 20925–20949

[28] Tan, X., Zhang, H., Wang, Z., Tang, J. (2023). Dual encoder network with efficient channel attention refinement module for image splicing forgery detection. Journal of Electronic Imaging 32(5), 053012–053012

[29] Rajkumar, R. (2022). Deep learning feature extraction using attention-based densenet 121 for copy move forgery detection. International Journal of Image and Graphics, 2350042

[30] Dong, F. *et al.* (2023) 'Contrastive learning-based general Deepfake detection with multi-scale RGB frequency

clues', *J. King Saud Univ. Comput. Inf. Sci.* USA: Elsevier Science Inc., 35(4), pp. 90–99. doi: 10.1016/j.jksuci.2023.03.005.

[31] Abd Warif, N.B., Idris, M.Y.I., Wahab, A.W.A., Salleh, R. (2015). An evaluation of error level analysis in image forensics. In: 2015 5th IEEE International Conference on System Engineering and Technology (ICSET), pp. 23–28 (2015). IEEE

[32] Gunawan, T. S. *et al.* (2017) 'Development of photo forensics algorithm by detecting photoshop manipulation using error level analysis', *Indonesian Journal of Electrical Engineering and Computer Science*, 7, pp. 131–137. Available at: https://api.semanticscholar.org/CorpusID:115236152.

[33] Rawat, W. and Wang, Z. (2017) 'Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review', *Neural Computation*, 29, pp. 2352–2449. Available at: https://api.semanticscholar.org/CorpusID:3290651.

[34] Wang, B. *et al.* (2023) 'Frequency Domain Filtered Residual Network for Deepfake Detection', *Mathematics*, 11(4), pp. 1–13. Available at: https://ideas.repec.org/a/gam/jmathe/v11y2023i4p816-d1059086.html.

[35] Dhar, A., Acharjee, P., Biswas, L., Ahmed, S., Sultana, A. (2021). Detecting deepfake images using deep convolutional neural network. PhD thesis, Brac University

[36] Sahib, I., AlAsady, T.A.A. (2022). Deep fake image detection based on modified minimized xception net and densenet. In: 2022 5th International Conference on Engineering Technology and Its Applications (IICETA), pp. 355–360 . IEEE

[37] Kumar, N., Pranav, P., Nirney, V. and Geetha, V., 2021, September. Deepfake Image Detection using CNNs and Transfer Learning. In 2021 International Conference on Computing, Communication and Green Engineering (CCGE) (pp. 1-6). IEEE.

[38] Dashti, M., Londono, J., Ghasemi, S., Tabatabaei, S., Hashemi, S., Baghaei, K., Palma, P.J., Khurshid, Z. (2024). Evaluation of the accuracy of deep learning and conventional neural network algorithms in the detection of dental implant type using intraoral radiographic images. A systematic review and meta-analysis. The Journal of Prosthetic Dentistry (2024).