Celal Bayar University Journal of Science

DeepFake Detection Using Fine-Tuned CNN Architectures

Dilber ÇETİNTAŞ[1], Zehra YÜCEL[2*]

[1] University of Malatya Turgut Özal, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, Malatya, Türkiye
[2] Necmettin Erbakan University, Department of Computer Technologies, Konya, Türkiye
* zkarhan@erbakan.edu.tr
* Orcid No: 0000-0002-2863-9119

Received: August 9, 2024
Accepted: November 28, 2024
DOI: 10.18466/cbayarfbe.1530209

Abstract

Synthetic images have gained significant popularity, producing high-quality visuals that are challenging to distinguish from real images. Computer-generated images have become increasingly realistic and misleading as artificial intelligence models advance. The easy dissemination of synthetic images online has raised concerns about their potential misuse. An automated detection system has become essential to safeguard personal privacy. Such a system is also critical for preventing manipulation, maintaining social order, and preserving the authenticity of images. This study compares lightweight and dense models for real-fake classification tasks. In the first phase, the performance of lightweight models on the dataset is analyzed, followed by an assessment of dense models in the second phase. When the best-performing lightweight model, EfficientNetV2B0, is combined in a hybrid with the top dense model, DenseNet201, an 88% accuracy rate is observed. Moreover, a hybrid of the two most effective dense models, DenseNet121 and DenseNet201, achieved an accuracy of 89% on the test dataset. Experimental results indicate that DenseNet networks excelling in finer details achieve preferable outcomes on synthetic data.

Keywords: Artificial Intelligence, DeepFake, Fine-Tuned CNN, Real-Fake Distinction, Synthetic Images,

1. Introduction

People generally tend to trust their own eyes and ears when communicating. For this reason, audio and visual evidence have traditionally been regarded as reliable, despite a long history of forgeries such as photo tampering [1]. However, advancements in artificial intelligence (AI) have undermined this trust. The ability to generate synthetic data with great realism, often without detectable visual traces, poses a significant challenge to the authenticity of digital content. In recent years, synthetic image generation has seen rapid progress, making it increasingly difficult to distinguish between real and fake images [1, 2].

Research in this area suggests that detection methods must also advance as generative models evolve. Preventing the misuse of synthetic images depends primarily on their accurate detection. Misusing of such images can create false narratives and misinformation, leading to manipulation and deception. The manipulation of visual evidence can greatly impact individual and societal decision-making. This makes detecting synthetic images crucial in fields like cybersecurity, digital forensics, and media verification. Preparing for and defending against the threats of synthetic data requires a detector that can classify images well [3, 4].

Studies in this area initially focused on pixel-level analysis and then explored various analytical approaches. Later studies emphasized geometric anomalies, reflections, and the uniform distribution of light. The consistency of metadata has also been taken into account [5, 6, 7].

Several studies have focused on the analysis of intrinsic patterns and artifacts present in synthetic images. Zhang et al. [8] and Frank et al. [9] investigated the detection of fake images generated by GANs by analyzing their frequency spectrum in the Fourier domain. Many detection methods remain GAN-specific, with relatively fewer studies addressing more recent generative models, such as diffusion models. However, Corvi et al. [2] analyzed the fingerprints left by diffusion models, a newer and increasingly popular class of generative models. Their work emphasizes the difficulty in distinguishing synthetic images produced by diffusion models from authentic ones and highlights the fundamental differences.

Another area of interest is the generalization of detection models across multiple datasets. Wang et al. [10]

observed that detection models performed well on a dataset generated using 11 different GAN-based image generation models. They emphasized that data augmentation improves performance but noted that there remains a need for model generalization across different generation techniques. Guarnera et al. [11] proposed a self-supervised learning-based method for distinguishing images generated by various GAN models (GDWCT, StarGAN, AttGAN, StyleGAN, StyleGAN2). Their approach involves analyzing features at different scales and employing the Expectation-Maximization (EM) algorithm to facilitate the detection process. This method allows for the effective differentiation of synthetic images by leveraging the inherent characteristics across different GAN architectures. In the context of diffusion models, Somepalli et al. [12] explored the relationship between dataset size and the similarity of diffusion models' output to training data. They demonstrated that larger training datasets enable the generation of higher-quality images.

The CNN-based transfer learning approach is widely used for detecting fake images. Malolan et al. [13] successfully distinguished between real and fake images with an accuracy of 94.33% using a CNN-based method on the FaceForensics dataset [14]. They also incorporated various explainable AI techniques, such as layer-wise relevance propagation (LRP) and local interpretable model-agnostic explanations (LIME), to enhance the interpretability of their model. Ranjan et al. [15] proposed a transfer learning-based CNN framework tested on three different datasets—DeepFakeDetection (DFD) [16], Celeb-DF [17], and the DeepFake Detection Challenge (DFDC) [18]. Additionally, they compiled a custom dataset for model evaluation, achieving 95.86% accuracy in distinguishing real and fake images.
Nida et al. [19] achieved 92.09% accuracy in detecting real and fake images using CNN-based models on the Real and Fake Face Detection dataset [20], after performing image normalization and Error Level Analysis (ELA) to enhance feature extraction prior to training.

Most studies on fake-vs.-real image detection has focused on facial images, while deepfake research involving other types of images (e.g., nature, vehicles, etc.) remains limited. Although considerable attention has been given to detecting GAN-generated images, particularly in ensuring robust detection across different types of synthetic data and applying existing deep-learning models, there is comparatively less research on images produced by diffusion models. In our study, we focus on detecting images generated by generative models, such as diffusion models, and extend the scope beyond facial images to include various other types of imagery. we aim for a generalizable detector.

Contributions of this Study:

- Comprehensive Evaluation: We provide a performance comparison of various fine-tuned CNN architectures for synthetic image detection.
- Superior Model Performance: Our experiments demonstrate that DenseNet models achieve the highest accuracy, outperforming other models.
- Synthetic Image Detection: Unlike many existing studies that focus solely on GAN-generated images, we provide a more generalized framework capable of detecting synthetic images from a variety of sources, including those generated by diffusion models.
- Analysis: We offer an analysis of each model's strengths and weaknesses in comparison to existing methods.

The rest of this paper; Section 2 presents the dataset, techniques, and models used. Section 3 presents the analyses and results of the experiments. In the last section, general conclusions and implications are presented.

2. Materials and Methods

This section contains information about the techniques, methods, and data sets used in the classification of synthetic data and real data.

2.1. Data Set

The CIFAKE [21] dataset consists of two groups: real images and fake images. The real image set consists of the CIFAR-10[22] dataset. The CIFAR-10 dataset contains 60,000 RGB images of 10 classes (deer, ship, horse, frog, aircraft, automobile, bird, cat, dog and truck) in 32x32 size. It contains 6,000 data from each of its classes.
The set of fake images of the CIFAKE dataset is obtained by applying LDM to the images in the CIFAR-10 dataset. A set of 60,000 synthetic images equivalent to 60,000 real datasets is obtained. The resolution of all images is 32x32 px. In order to achieve differentiation, the researchers used different prompt variations for each class.
In this study, evaluations were conducted using randomly generated subsets from the CIFAKE dataset. Among the created subsets, the best results were achieved with a training set containing 7,000 fake and 7,000 real images, and a test set with 1,058 fake and 1,058 real images. To ensure generalizability, datasets with varying distributions were utilized.

Celal Bayar University Journal of Science
Volume 21, Issue 1, 2025, p 121-128
Doi: 10.18466/cbayarfbe.1530209

2.2. Pre-Trained Models

In this section, MobileNet [23], DenseNet121[24], EfficientNetV2B0[25] are discussed and briefly introduced.

MobileNet is widely used to create lightweight deep convolutional neural networks, providing an efficient architecture that does not require extensive computational resources. It is particularly prevalent in computer vision applications like object detection and classification. The architecture includes down-sampling with features derived from the previous layer and concludes with an average pooling layer. A SoftMax function is used in the final layer for classification purposes. MobileNet's architecture relies on depthwise separable convolutions, incorporating batch normalization and ReLU activation functions. The entire MobileNet model consists of 28 layers [23,26].

To further reduce computational costs, MobileNet employs width and resolution multipliers. The width multiplier thins the model by a specified factor; selecting a smaller multiplier yields a faster, smaller model but may lead to some information loss. The resolution multiplier, on the other hand, adjusts the input image resolution. Choosing a smaller resolution provides a more compact and efficient model, though potentially at the cost of detail. It is therefore critical to select these multipliers carefully to balance computational efficiency with minimal information loss [23,26]. MobileNet has been improved and extended with newer versions, such as MobileNetV2[27] and MobileNetV3[28], aiming to enhance accuracy and speed. MobileNetV2 introduced inverted residuals to achieve an expansion-filtering-squeezing mechanism [27]. In MobileNetV3, squeeze-and-excitation layers were added to the initial block structure further enhance performance [28].

Densely Connected Convolutional Networks (DenseNet), unlike other architectures, contain dense blocks of layers. Each layer is directly connected to each subsequent layer in a feed-forward manner [24]. This architecture encourages the reuse of features when addressing the gradient fading problem and therefore reduces the number of parameters. As a result, DenseNet offers a powerful approach in scenarios where small differences are important [24, 29]. There are several versions of DenseNet used in object recognition applications, for example, DenseNet121, DenseNet160, and DenseNet201. Numbers in these versions represent the number of layers in the model. The general representation of the DenseNet121 architecture is shown in Figure 1. In DenseNet121, except for the first layer that receives the input image, each subsequent convolutional layer creates an output feature map by taking the output of the previous layer [24].

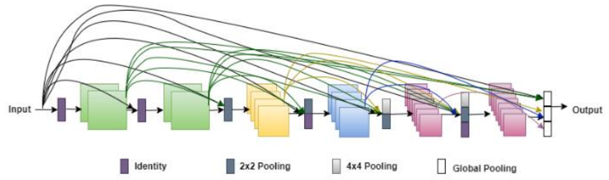


Figure 1. General Structure of DenseNet121 Architecture

EfficientNetV2B0 is an upgraded model of EfficientNet, proposed by Google Brain. It employs a mobile-dimensional convolutional network based on training-driven neural architecture search and scaling. This model outperforms previous models in terms of training speed, accuracy, and parameter efficiency. EfficientNetV2B0 demonstrates that deep convolution is slow in the early layers but becomes more efficient in later stages [25, 30]. [25, 30]. The general structure of the EfficientNetV2B0 architecture is shown in Figure 2. Architecture consists of a combination of MBConv block [31] to balance the expressive power and computational cost, and Fused-MBConv [32] to further speed up the process.

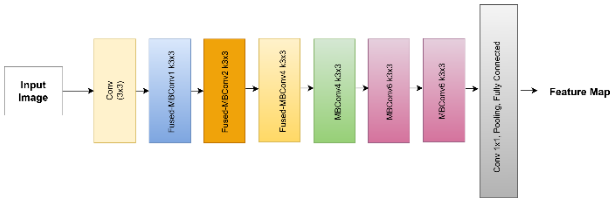


Figure 2. General Structure of EfficientNetV2B0 Network [25]

2.3. Evaluation Metrics

The confusion matrix was used to evaluate performance. In this matrix, rows indicate real classes and columns indicate predicted classifications. To measure evaluation metrics, tp (true positive), tn (true negative), fp (false positive), fn (false negative) values are used.

Precision: It is defined as the ratio of the data determined as positive among the predicted ones to the total number of positives. Calculation of the precision value is given in equation 1.

$$Precision=tp/(tp+fn) \quad \textbf{(2.3.1)}$$

Recall: It provides information about the number of data that are actually predicted as positive among the data that should be predicted as positive. Calculation of the sensitivity value is specified in equation 2.

$$Recall=tp/(tp+fp) \quad \textbf{(2.3.2)}$$

F1-Score: It is calculated by the harmonic mean of precision and sensitivity calculations. It takes values between 0 and 1, with 1 indicating that the best result has been achieved. Calculation of the F1 score value is given in equation 3.

F1-Score=2*tp/(2tp+fp+fn) **(2.3.3)**

Accuracy: It is the ratio of correctly predicted data to all data. Calculation of Accuracy value is given in equation 4.

Accuracy=(tp+tn)/(tp+tn+fp+fn) **(2.3.4)**

3. Results and Discussion

In the initial approach, the input layers of various lightweight models, including MobileNetV2, MobileNetV3Small, MobileNetV3Large, and EfficientNetV2B0, were adjusted to the 32x32 image size. In the second stage, this adjustment was applied to dense models, namely DenseNet121, DenseNet169, and DenseNet201. These models were then trained with frozen parameters. Subsequently, additional layers were appended to the architecture, including a GlobalAveragePooling2D layer, a Dense layer with 1024 units and ReLU activation, a Dropout layer with a rate of 0.5, and a final Sigmoid layer. The Adam optimization algorithm was used to train the models with a learning rate of 1e-5 and an epoch count of 50. Experimental evaluations were then conducted using this configuration.

In the literature, numerous studies on the classification of synthetic data have made significant contributions by evaluating different deep-learning models in this area. However, determining which models are more effective at distinguishing between real and fake data remains a challenging problem. This study compares the performance of lightweight and heavy models to analyze their impact on synthetic data classification. Lightweight models include EfficientNetV2B0, MobileNetV2, MobileNetV3Small, and MobileNetV3Large, while heavy models consist of DenseNet121, DenseNet169, and DenseNet201. The performance of these models was evaluated comparatively, with the results presented in Tables 1 and 2, model accuracy detailed through confusion matrices in Figures 3 and 4. In the matrices, the value "0" represents the "Fake" label, while "1" denotes the "Real" label.

Table 1. Performance Results Evaluation of Lightweight Models in Deepfake Detection

Models	Precision	Recall	F1-Score	Accuracy
MobileNetV2	0.30	0.73	0.42	0.59
MobileNetV3Small	0.71	0.48	0.57	0.47
MobileNetV3Large	0.72	0.82	0.77	0.78
EfficientNetV2B0	0.77	0.90	0.83	0.84

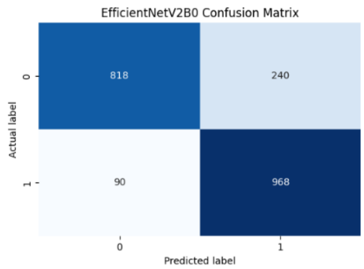
The experimental results reveal distinct performance patterns across the models tested. EfficientNetV2B0 demonstrated the highest performance among the models, achieving a precision of 0.77, recall of 0.90, F1-score of 0.83, and accuracy of 0.84. The model's high recall, coupled with robust precision and accuracy, underscores its ability to accurately identify relevant instances while maintaining a low false-positive rate. This balanced performance makes EfficientNetV2B0 the most reliable lightweight model in this study for achieving both precise and comprehensive classification.

Table 2. Performance Results Evaluation of Compact Models for Test Data

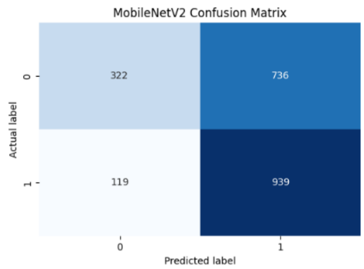
Models	Precision	Recall	F1-Score	Accuracy
DenseNet121	0.84	0.91	0.87	0.88
DenseNet169	0.83	0.89	0.86	0.86
DenseNet201	0.85	0.90	0.88	0.88

The performance results for the DenseNet models indicate strong classification abilities across all three variations. Overall, all three DenseNet models demonstrate excellent classification capabilities, with DenseNet201 showing a slight edge, particularly for tasks prioritizing both precision and recall.

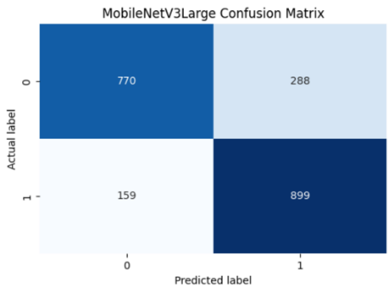
Celal Bayar University Journal of Science
Volume 21, Issue 1, 2025, p 121-128
Doi: 10.18466/cbayarfbe.1530209



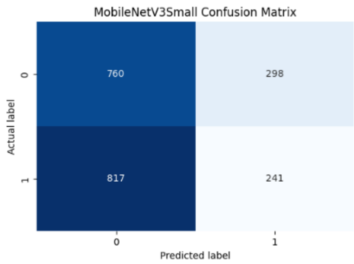
a) EfficientNetV2B0



b) MobileNetV2

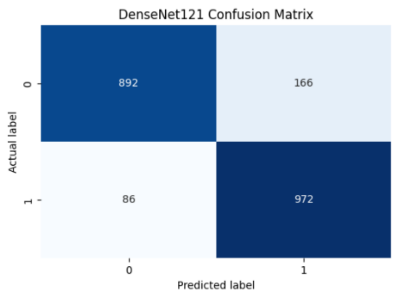


c) MobileNetV3Large

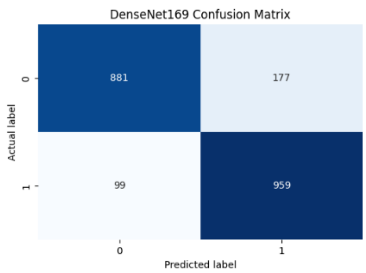


d) MobileNetV3Small

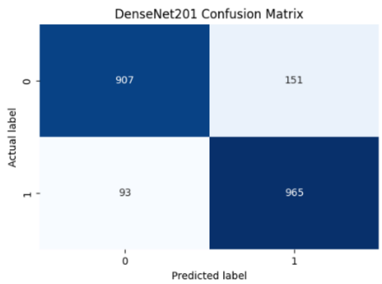
Figure 3. Confusion matrix obtained from lightweight fine-tuned CNN Architecture a)EfficientNetV2B0 b) MobileNetV2 c) MobileNetV3Large d) MobileNetV3Small



a) DenseNet121



b) DenseNet169



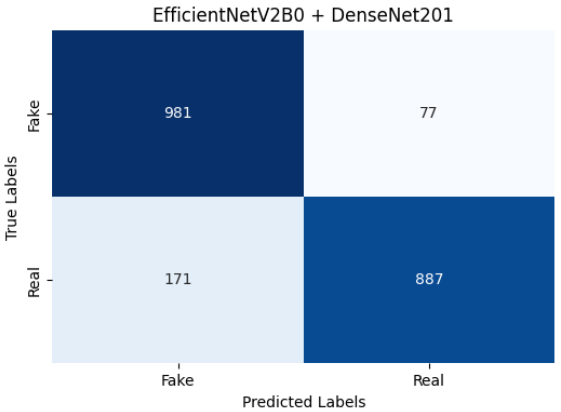
c) DenseNet201

Figure 4. Confusion matrix obtained from compact fine-tuned CNN Architecture a) DenseNet121 b) DenseNet169 c) DenseNet201

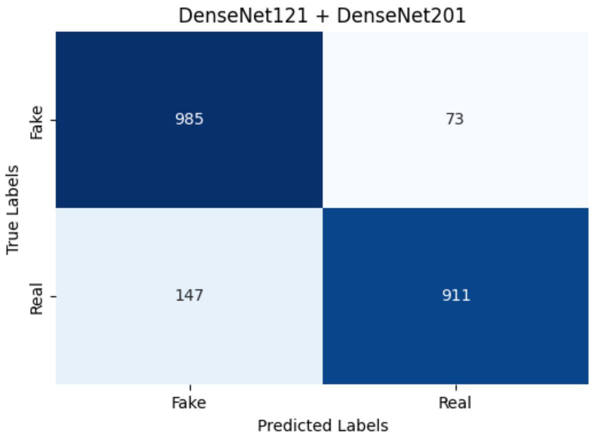
The analysis of the confusion matrices for various models, including EfficientNetV2B0, MobileNetV2, MobileNetV3 (Small and Large), DenseNet121, DenseNet169, and DenseNet201, highlights differences in classification performance. DenseNet models, particularly DenseNet201 and DenseNet121, demonstrate a balanced performance with lower false negative and false positive rates, indicating strong accuracy and reliability. EfficientNetV2B0 shows high accuracy but a relatively high false positive rate, which may impact its sensitivity. MobileNet models, especially MobileNetV2 and MobileNetV3Small, display higher error rates and appear less suitable for this dataset due to lower classification accuracy. Overall, DenseNet201 emerges as the optimal model for this task, achieving a robust balance between sensitivity and precision, making it a dependable choice for accurate classification. This study underlines the efficacy of DenseNet models, especially DenseNet201, as superior classifiers in terms of accuracy and balanced error rates.

Overall, the DenseNet series models, particularly DenseNet201 and DenseNet121, stand out as models with the highest accuracy and balanced classification performance. While EfficientNetV2B0 also provides high accuracy, its relatively high false positive rate slightly limits its sensitivity. MobileNetV2 and

MobileNetV3Small demonstrate lower accuracy in comparison to other models, rendering them insufficient for the dataset. This analysis suggests that DenseNet201 could be preferred as an accurate and reliable classification model due to its strong balance between precision and sensitivity.



a) EfficientNetV2B0 + DenseNet201



b) DenseNet121+ DenseNet201

Figure 5. Deepfake detection results with DenseNet201

The hybrid model combining EfficientNetV2B0 and DenseNet201 achieved notable results in distinguishing between synthetic and real images confusion matrix, as shown in Figure 5. The model correctly identified 981 fake images (true positives) and 887 real images (true negatives), while misclassifying 77 fake images as real (false positives) and 171 real images as fake (false negatives). This performance yielded an overall accuracy of approximately 88%, indicating the model's strong capability in detecting both fake and real images.

When the two most successful dense models, DenseNet121 and DenseNet201, were combined in a hybrid approach, the classification performance demonstrated further improvements, particularly in accurately identifying real images. The confusion matrix reveals that the hybrid model correctly identified 985 synthetic (fake) images and 911 real images, resulting in 73 false positives (fake images misclassified as real) and 147 false negatives (real images misclassified as fake). This performance led to an overall accuracy of approximately 89%, highlighting the effectiveness of combining these dense networks. The results indicate that while both DenseNet121 and DenseNet201 individually excel at detecting fake images, their combined use enhances the model's ability to differentiate real images more accurately. This hybrid approach leverages the strengths of both models, providing a more robust and precise classification compared to their individual performances. The study demonstrates that integrating multiple dense architectures can effectively improve detection accuracy, particularly in challenging tasks that require high sensitivity to both synthetic and authentic visual patterns.
The method of synthetic data generation is an important limiting factor. Studies in the literature have predominantly utilized synthetic data generated with GANs. Our study addresses this gap by employing synthetic data generated through diffusion models. Another constraint is that successful results can be achieved even with input images at a resolution as low as 32x32. This new study, contributing to the limited body of research on diffusion models, compares lightweight and dense models to evaluate which is more effective in terms of overall performance. The findings provide valuable insights into the most suitable model types for different applications. The use of hybrid models on the newly developed CIFAKE dataset has proven to yield more effective results.

4. Conclusion

The hybrid use of dense networks contributed to achieving a more balanced and successful performance across both fake and real classes, significantly enhancing classification accuracy. The ensemble method, combining these models in a hybrid fashion, produced the most successful results on the given dataset, further demonstrating the effectiveness of integrating various model features in deep learning applications. This finding suggests that optimizing dense networks for specific classes can lead to higher performance in classification tasks, highlighting the potential of hybrid approaches for improved accuracy in such applications. It gives information about the performances of our study and lightweight-compact models. The results are promising in the field of classification. This may give an idea to scientists who will work in this field.

As a future study, we aim to improve the performance of these models by utilising other distinguishing features of synthetic data. At the same time, we aim to focus on a detector that can be generalised to the dataset produced by different diffusion models. Additionally, in image detection, learning long-range dependencies through Transformer models can significantly enhance the accuracy of synthetic data detection. As a potential direction for future research, a hybrid approach that combines CNN and Transformer models could be explored to leverage the strengths of both architectures.

Author's Contributions

All authors have contributed equally to this work. All authors have read, provided feedback on, and approved the final version of the manuscript.

Ethics

There are no ethical issues after the publication of this manuscript.

References

[1]. Dang, L.M., Min, K., Wang, H., Piran, M.J., Lee, C.H. and Moon, H., 2020. Sensor-based and vision-based human activity recognition: A comprehensive survey. *Pattern Recognition*, *108*, p.107561.

[2]. Corvi, R., Cozzolino, D., Zingarini, G., Poggi, G., Nagano, K. and Verdoliva, L., 2023, June. On the detection of synthetic images generated by diffusion models. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE.

[3]. Bonettini, N., Bestagini, P., Milani, S. and Tubaro, S., 2021, January. On the use of Benford's law to detect GAN-generated images. In *2020 25th international conference on pattern recognition (ICPR)* (pp. 5495-5502). IEEE.

[4]. Bird, J.J., Naser, A. and Lotfi, A., 2023. Writer-independent signature verification; Evaluation of robotic and generative adversarial attacks. *Information Sciences*, *633*, pp.170-181.

[5]. Bayar, B. and Stamm, M.C., 2016, June. A deep learning approach to universal image manipulation detection using a new convolutional layer. In Proceedings of the 4th ACM workshop on information hiding and multimedia security (pp. 5-10).

[6]. Verdoliva, L., 2020. Media forensics and deepfakes: an overview. *IEEE journal of selected topics in signal processing*, *14*(5), pp.910-932.

[7]. Fridrich, J. and Kodovsky, J., 2012. Rich models for steganalysis of digital images. *IEEE Transactions on information Forensics and Security*, *7*(3), pp.868-882.

[8]. Zhang, X., Karaman, S. and Chang, S.F., 2019, December. Detecting and simulating artifacts in gan fake images. In *2019 IEEE international workshop on information forensics and security (WIFS)* (pp. 1-6). IEEE.

[9]. Frank, J., Eisenhofer, T., Schönherr, L., Fischer, A., Kolossa, D. and Holz, T., 2020, November. Leveraging frequency analysis for deep fake image recognition. In *International conference on machine learning* (pp. 3247-3258). PMLR.

[10]. Wang, S.Y., Wang, O., Zhang, R., Owens, A. and Efros, A.A., 2020. CNN-generated images are surprisingly easy to spot... for now. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8695-8704).

[11]. 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).

[12]. Somepalli, G., Singla, V., Goldblum, M., Geiping, J. and Goldstein, T., 2023. Diffusion art or digital forgery? investigating data replication in diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6048-6058).

[13]. Malolan, B., Parekh, A., & Kazi, F., 2020. Explainable deep-fake detection using visual interpretability methods. In *2020 3rd International conference on Information and Computer Technologies (ICICT)* (pp. 289-293). IEEE.

[14]. Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M., 2018. Faceforensics: A large-scale video dataset for forgery detection in human faces. *arXiv preprint arXiv:1803.09179*.

[15]. Ranjan, P., Patil, S., & Kazi, F., 2020. Improved generalizability of deep-fakes detection using transfer learning based CNN framework. In *2020 3rd international conference on information and computer technologies (ICICT)* (pp. 86-90). IEEE.

[16]. Dufour, N., & Gully, A. 2024. Contributing data to deepfake detection research. https://research.google/blog/contributing-data-to-deepfake-detection-research/

[17]. Li, Y., Yang, X., Sun, P., Qi, H. and Lyu, S., 2020. Celeb-df: A large-scale challenging dataset for deepfake forensics. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 3207-3216).

[18]. Dolhansky, B., 2019. The deepfake detection challenge (DFDC) pre view dataset. arXiv preprint arXiv:1910.08854.

[19]. Nida, N., Irtaza, A., & Ilyas, N. 2021. Forged face detection using ELA and deep learning techniques. In *2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST)* (pp. 271-275). IEEE.

[20]. Real and fake face detection, https://www.kaggle.com/ciplab/real-and-fake-face-detection (2024).

[21]. Bird, J.J. and Lotfi, A., 2024. Cifake: Image classification and explainable identification of ai-generated synthetic images. *IEEE Access*.

[22]. The CIFAR-10 dataset, https://www.cs.toronto.edu/~kriz/cifar.html (accessed at 13.03.2024).

[23]. Howard, A. G. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

[24]. Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q., 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

[25]. Tan, M. and Le, Q., 2021, July. Efficientnetv2: Smaller models and faster training. In *International conference on machine learning* (pp. 10096-10106). PMLR.

[26]. Wang, W., Li, Y., Zou, T., Wang, X., You, J., & Luo, Y. (2020). A novel image classification approach via dense-MobileNet models. Mobile Information Systems, 2020(1), 7602384.

[27]. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[28]. Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., & Adam, H. (2019). Searching for mobilenetv3. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1314-1324).

[29]. Bozkurt, F. (2021). Classification of blood cells from blood cell images using dense convolutional network. *Journal of Science, Technology and Engineering Research*, *2*(2), 81-88.

[30]. Zheng, T., Yang, X., Lv, J., Li, M., Wang, S. and Li, W., 2023. An efficient mobile model for insect image classification in the field pest management. *Engineering Science and Technology, an International Journal*, *39*, p.101335.

[31]. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520).

[32]. Gupta, S. and Akin, B., 2020. Accelerator-aware neural network design using automl. *arXiv preprint arXiv:2003.02838*.