SEMARAK ILMU
PUBLISHING
202103268166(003316878-P)

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

SEMARAK ILMU PUBLISHING
Journal of
ADVANCED RESEARCH IN APPLIED SCIENCES AND ENGINEERING TECHNOLOGY
ISSN 2462 - 1943

Wavelet Transform-based Methods for Forensic Analysis of Digital Images

Muhammad S. Mandisha[1,*], Mohamed A. Hussien[1], Amr K. Shalaby[1], Omar M. Fahmy[2]

[1] Faculty of Computer Science, Arab Academy for Science, Technology, and Maritime Transport, Egypt
[2] Electrical Engineering Department, Badr University in Cairo, Egypt

ARTICLE INFO

Article history:
Received 22 June 2023
Received in revised form 10 February 2024
Accepted 6 March 2024
Available online 25 April 2024

Keywords:
GANs; Image and video manipulation; Gradient boosting-based classifiers; Image forensics; Forgery detection

ABSTRACT

In recent years, Generative Adversarial Networks (GANs) have been utilized in many applications of our daily lives to create digital media that was previously impossible. This paper utilizes GANs in multimedia forensics, where the precision and accuracy of image classification are crucial. The objective of the study is to detect sophisticated manipulated images generated by advanced deep learning methods using classical transformation methods. The method combines the classical wavelet transform for feature extraction with a classifier model to distinguish between real and fake images. The proposed method extracts unique features that differentiate between fake and real images, which are then fed to several gradient-boosting-based classifiers. The proposed method was tested using the FaceForensics++ dataset, which contains low-resolution video sequences of faces. The proposed model achieves an accuracy of 95% in detecting manipulated videos compared with 87% accuracy of other classical techniques.

1. Introduction

The widespread use of social media networks and developments in deep learning, notably GANs, have significantly improved the generation of media content. However, the creation of modified or fake images poses significant challenges to various fields, including security, forensics, and the media. Several literature sources, including [1,2], have extensively discussed the utilization of GAN techniques to generate synthetic images that closely resemble real ones, making it challenging to differentiate between the two. In addition, in [3,4], researchers have focused on advanced GAN methods for producing high-resolution photorealistic images and videos, even from low-resolution data. Moreover, GANs have gained significant traction in design contests, as highlighted by Shahriar [5]. The author conducted a comprehensive survey exploring the application of GANs across various artistic mediums, such as visual arts, music, and literary text generation. The survey showcases the potential of GANs in generating creative works that may not be easily distinguishable as machine-generated. Furthermore, Han *et al.* [6] introduced a specialized GAN technique called Depth Extraction Generative Adversarial Network (DE-GAN) for art generation. DE-GAN enables the creation of artwork that exhibits depth and complexity, challenging the perception of whether it was generated by a machine or an artist. In a different vein, Meng *et al.* [7] presented GAN techniques

[*] *Corresponding author.*
E-mail address: M.Ibrahim33815@student.aast.edu

https://doi.org/10.37934/araset.44.1.4654

specifically tailored for the intricate and time-consuming task of icon design. Their work demonstrates the versatility of GANs in tackling diverse design challenges. Overall, these studies and surveys exemplify the wide-ranging capabilities of GANs in generating realistic and aesthetically pleasing outputs across various domains, pushing the boundaries of artificial creativity. However, this mixing of artistic work generated by humans and machines can create an unfair ground for such contests. It may lead to forgery works in other fields, particularly in critical and safety-related situations. Many research studies have introduced deep learning techniques for image classification and detection, Deepika *et al.,* [8]. Additionally, approaches such as those described in [9,10] have been employed to differentiate between real and manipulated images, addressing this particular challenge. These techniques have demonstrated promising accuracy gains. However, according to [11,12], altering deep learning methods to solve deep fake classification problems has some challenges, including the need for large amounts of labeled data and the possibility of overfitting. Additionally, the requirement for specific hardware, such as Graphics Processing Units (GPUs), poses a significant challenge in the widespread adoption of deep learning. Training deep neural networks demands substantial computational power to enhance the learning process speed when compared to Central Processing Units (CPUs), as mentioned by Munanday *et al.*[13] However, the cost and availability of GPUs present barriers for many individuals and organizations interested in utilizing deep learning techniques.

This research proposes an approach that leverages classical wavelet transform for feature extraction and a gradient boosting-based classifier model to address these challenges. The wavelet transform is a powerful signal-processing method that can efficiently extract features from images and videos. Such a method differentiates the unique patterns and artifacts present in manipulated data by utilizing wavelet transform-based feature extraction.

The proposed method has several advantages over deep learning-based approaches. First, it does not require vast amounts of labeled data, making it a cost-effective and practical solution for detecting manipulated data. Second, wavelet transform-based feature extraction is computationally efficient and can be applied to low-resolution data, making it suitable for detecting manipulated data in low-quality scenarios. Third, our approach can be easily implemented and does not require specialized hardware or deep learning expertise.

To assess the proposed method, The experiments are conducted on the FaceForensics++ dataset [14], which consists of low- resolution video sequences of faces. Our method achieved an accuracy of 95% in detecting manipulated face images, demonstrating the potential of our approach in detecting manipulated data in low-quality scenarios.

This paper is organized as follows: Part 1 of Section 2 describes the proposed method. The simulation experiments are mentioned in part 2 of section 2. The results of the proposed model are described in section 3 and finally, section 4 concludes the paper.

2. Methodology

2-1) Data Processing

In this section, the discussion will focus on the proposed method presented in Fig. 1, which is composed of two main parts: image preprocessing for feature extraction and classification for forgery detection.

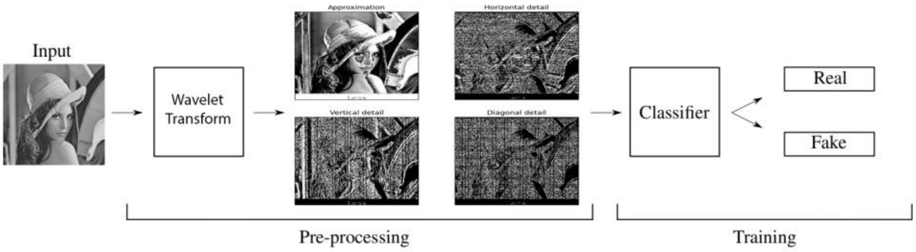

Fig. 1. The first part represents feature extraction for gray-scale face images and the second part represents the classifier which determines whether the face is real or fake

2-1-1) Image Preprocessing

This method aims to utilize the powerful use of Discrete Wavelet Transform DWT for image feature representation. This approach decomposes signal or data set into a set of basis functions that are localized in both time and frequency, providing a more detailed analysis of the image features as shown in Fig.2. Because the basis functions are localized, analyzing time and frequency in greater detail is possible, leading to a more thorough understanding of the image features as discussed in [15,16].

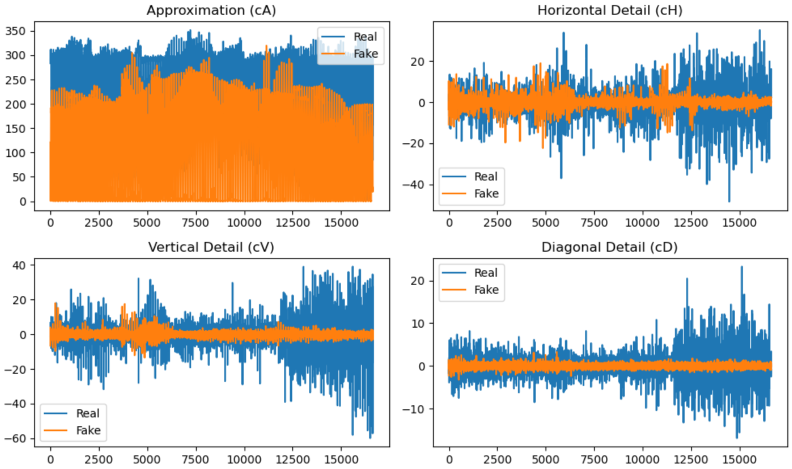

Fig. 2. The signal graph shows the dynamic range discrepancy between the features of a real face image sample compared to a fake one

While traditional Fourier transform methods have been proposed in some literature, as proposed by Durall *et al.,*[17], there are limitations and drawbacks to using them for feature extraction and classification. As mentioned in the previous reference [17], the model is highly sensitive to the number of features, leading to overlapping standard deviations of real and deep fake statistics at lower feature numbers, resulting in misclassified samples. Moreover, Fourier transform lacks information about the timing or location of signal components, is sensitive to noise, and has limited ability to capture complex relationships in non-linear classification problems, as mentioned in [18,19]. An alternative method is proposed here to address these limitations, which involves the use and comparison of two wavelets from the Wavelet Family, namely "Daubechies" and "Haar," for feature extraction and classification. The Decomposition wavelet filters (Daubechies) allow multi-resolution signal analysis that is sensitive to fine details in the image.

$$\psi_{a,b}(t) = |a|^{\frac{-1}{2}}\psi\left(\frac{t-b}{a}\right), \quad a, b \in \mathbb{R} \text{ dan } a \neq 0, \tag{1}$$

Where ψ represents a mother wavelet, a represents the scaling (dilation) parameter that determines the degree of compression or scale, and b represents the translation parameter that specifies the time location of the wavelet [20].

Similarly, the Haar wavelet also provides multi-resolution signal analysis, enabling the extraction of features that are sensitive to fine and coarse details in the signal. This capability has been extensively demonstrated in various critical scenarios. For instance, the Haar wavelet has proven its effectiveness as a non-separable watermarking algorithm for copyright protection of digital content, as proposed by A. Razak *et al.* [21]. This algorithm aims to embed or extract a watermark in a manner that preserves the integrity of the data while providing additional information and protection. The successful application of the Haar wavelet in this context further highlights its efficacy in feature extraction and its suitability for non-separable watermark algorithms. The Haar wavelet's mother wavelet function can be described as shown in Fig. 3.

$$\psi(x) = \begin{cases} 1, & 0 \leq x < 1/2, \\ -1, & \frac{1}{2} \leq x < 1, \\ 0, & \text{others}, \end{cases}$$
(2)

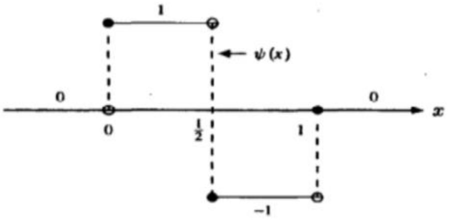

Fig. 3. Graph of Haar wavelet

2-1-2) Classification Algorithms:

In this study, two tree-based classifiers were proposed as the main classifiers for distinguishing between fake and real images. The effectiveness of tree-based classifiers has been demonstrated in various challenging classification tasks. For instance, a model for Distributed Denial-of-Service (DDoS) Attack Detection was proposed by Goparaju and Rao [22], utilizing a combination of convolutional neural network and decision tree classifier. An accuracy of over 96% was achieved by this model. This emphasizes the ability of tree-based classifiers to handle complex classification problems and achieve high accuracy rates.

XGBoost: (eXtreme Gradient Boosting) is used, which is a powerful algorithm that forms a strong learner by using decision trees as weak learners to classify input data as mentioned by Elsheakh *et al.,* [23]. L1 and L2 regularization are implemented in XGBoost's built-in mechanism to prevent overfitting and improve model generalization. Furthermore, the identification of the most important predictors in the input data is enabled by its feature importance measure. The effectiveness of XGBoost has been demonstrated in numerous studies, surpassing classical classifiers, as proposed by Maheshwari *et al.* [24]. They conducted a comparison of multiple classifier accuracies for credit card fraud detection and found that XGBoost outperformed other classical classifiers. This further highlights the superior performance and capabilities of XGBoost in various classification tasks.

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} \iota\left(\mathcal{Y}_{i,}\widehat{\mathcal{Y}_{\iota}}^{(t-1)} + \mathrm{f}_t(\mathbf{x}_i)\right) + \Omega(\mathrm{f}_t\ ) \tag{3}$$

CatBoost: is a gradient-boosting algorithm that uses a decision tree variant known as "gradient-boosted trees" to gradually boost the performance of the model. The algorithm updates the model with new trees after each iteration, fixing any mistakes the earlier trees had made. The algorithm specifically computes the gradient of the loss function with respect to the model's output and uses this data to train a new tree that is optimized to reduce residual error as concluded by Sadek *et al.,*[25].

$$h(\mathbf{x}) = \sum_{j=1}^{J} b_j\ \ \mathbb{1}_{\{\mathbf{x}\in R_j\ \}} \tag{4}$$

2.2) Experiments

In this section, the discussion will be divided into two parts: the first part will focus on the dataset used for experimentation and testing the hypothesis, while the second part will cover the model training process.

2-2-1) FaceForensics++ Dataset

The FaceForensics++ dataset [14] is a large-scale benchmark dataset designed for evaluating the performance of face manipulation detection methods. The dataset contains a diverse set of low-resolution quality video sequences featuring real human faces, which have been manipulated using various automated face manipulation techniques as shown in Fig.4. These techniques include DeepFakes, Face2Face, FaceSwap, and NeuralTextures, among others. It comprises more than 300 original sequences from many paid actors in various scenes, all of which have a visible, mostly frontal face without obstructions, making it possible for automated tampering methods to create realistic forgeries.

This dataset was inspired by the dataset used in [17]. The aim was to compare and validate our method against the Fourier transform method using the same dataset.

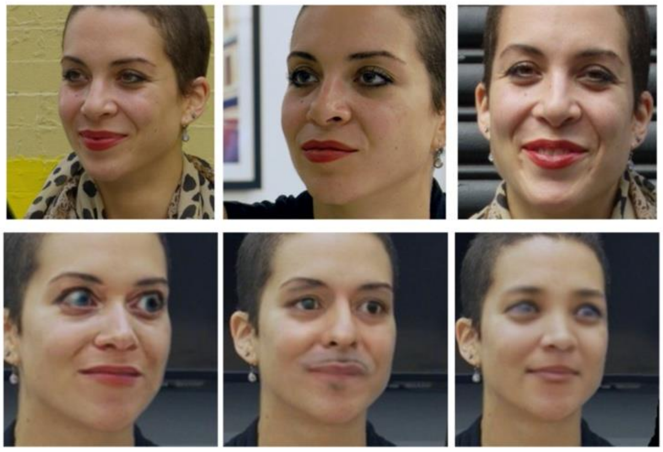

Fig. 4. shows examples from the dataset, with the first row displaying real images and the second row showing generated fake images

2.2-2) Model training process:

The dataset used in the study consisted of frames extracted from videos, representing images of faces. The first step in the model pipeline involved cropping the interior faces from the images, as illustrated in Fig. 5. This resulted in cropped faces of varying sizes, reflecting the diverse nature of the scenes in the videos. Following the cropping of interior faces from the images, each image in the dataset was converted to grayscale and normalized. This involved scaling the image's pixel values to a range of 0 to 1 to ensure consistency across the dataset. Next, two wavelet filters, the Haar filter, and the Daubechies filter, were applied to each normalized grayscale image to extract features.

The wavelet transformation decomposes the image into a set of coefficients that represent different frequency components of the image. In this study, a wavelet decomposition level of 4 was chosen to balance between capturing sufficient detail in the image and avoiding noise as shown in Fig.6. The resulting coefficients from the wavelet transformation were used as features for subsequent analysis.

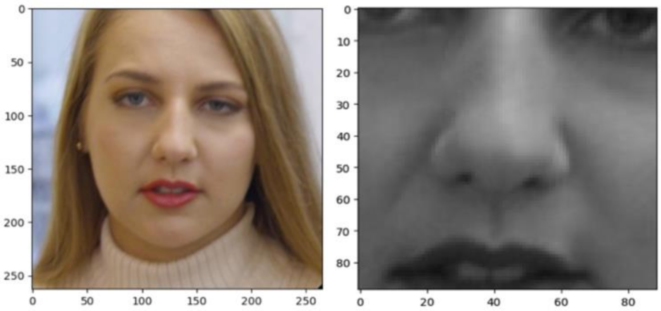

Fig. 5. shows a sample of the raw dataset (left) and the same image after being cropped and converted to a grayscale version (right)

As the faces in the dataset had varying sizes, the number of features extracted from each image differed. To tackle this problem, the number of extracted features for each image was recorded. This maximum number of features was then used to standardize the feature set for the entire dataset. Specifically, for images that had fewer features than the maximum, padding with zeros was applied to ensure that all images had the same number of features. By applying padding to the coefficient matrices, the dataset was standardized, and each image had the same number of features. This step was essential to ensure that the input to the classification model was consistent across all images, regardless of their size or the number of features extracted.

After extracting features from the images, several classifiers were applied to the features, and the accuracy scores for each classifier were recorded. The classifier with the best performance for the classification task was identified as the final step. In this study, XGBoost was used as the primary classifier, and a grid search was conducted to identify the optimal hyperparameters. After iterating through several design options, the following hyperparameters resulted in the best performance: 'learning_rate': 0.1, 'max_depth': 7, 'n_estimators': 1000, 'reg_alpha': 0.1.The hyperparameters were also tuned for CatBoost, as 'max_depth': 11 and 'n_estimators': 2000.

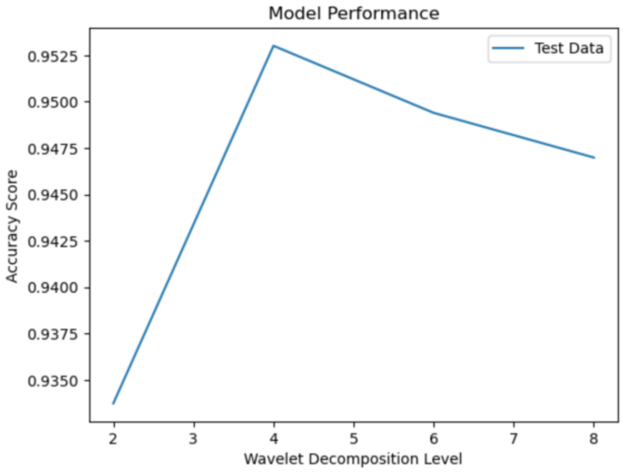

Fig. 6. shows the relation between wavelet decomposition level and accuracy score on test data

The experiments were carried out using a GPU P100 for the model training process. The training was completed in 148 seconds. Moreover, the model exhibited exceptional space efficiency, requiring only 48 bytes of memory, with a memory usage of only 12.66 megabytes during training. This indicates that the model is highly efficient in terms of memory consumption, making it suitable for applications with limited resources or large datasets.

3. Results

Our results showed that using the db2 wavelet transformation with XGBoost classifier achieved the highest classification accuracy of 95%. This finding supports our hypothesis that wavelet transforms are more effective for feature extraction and classification in detecting deepfake images. On the other hand, the highest accuracy achieved by using the Fourier transform was 87% with an SVM classifier as shown in Table 1. Although this accuracy is lower than that achieved with the wavelet transform, it still suggests that Fourier transform can also be effective in detecting manipulated facial images.

Our study demonstrates the importance of selecting appropriate feature extraction techniques and machine learning algorithms for detecting deepfake images. The results of our study can be useful for developing more accurate and effective techniques for detecting manipulated images in real-world scenarios.

Table 1
Accuracy of various classifiers compared with Fourier transform and wavelet transform methods

Classifiers	Accuracy		
	Wavelet Family		Fourier
	db2	Haar	Fourier transform
XGBClassifier	0.953	0.87	0.83
SVM	0.85	0.7	0.87
CatBoostClassifier	0.93	0.89	0.83
Random Forest	0.91	0.88	0.84

The effectiveness of the proposed methodology was demonstrated by utilizing a low-resolution dataset. Despite the lower quality of sophisticated manipulated dataset images as shown in Fig. 7, the proposed model can detect it with high accuracy.

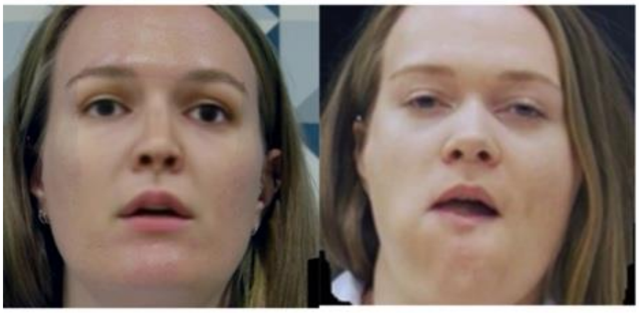

Fig. 7. the left image represents a fake image, while the right image represents a real image

4. Conclusions

The potential advantages of utilizing wavelet transform over Fourier transform for feature extraction are highlighted in our study, with a classification accuracy of 95% achieved using XGBoost classifier. This surpasses the performance of the Fourier transform, which yielded only 87% accuracy. Furthermore, high effectiveness in distinguishing between fake and real images is demonstrated by our proposed method, as evidenced by the high precision and recall achieved in the classification of manipulated images in the FaceForensics++ dataset.

These results suggest that our method has the potential to be applied in a wide range of image classification tasks, particularly in situations where the ability to accurately identify manipulated images is critical, such as in forensics, security, and authentication.

In summary, our proposed method is a promising approach to image processing and feature extraction that has the potential to enhance the accuracy and dependability of image classification in different applications which can be useful in preventing the spread of fake images in fields like journalism, social media, and advertising where trustworthy visuals are essential.

Acknowledgment

This research was not funded by any grant

References

[1] Saxena, Divya, and Cao. “Generative Adversarial Networks (GANs).” *ACM Computing Surveys* 54, no. 3 (May 8, 2021): 1–42. https://doi.org/10.1145/3446374

[2] Gragnaniello, Diego, Davide Cozzolino, Francesco Marra, Gianfranco Poggi, and Luisa Verdoliva. “Are GAN Generated Images Easy to Detect? A Critical Analysis of the State-of-the-Art.” *ArXiv (Cornell University)*, April 6, 2021. https://doi.org/10.48550/arxiv.2104.02617

[3] Karras, Tero, Timo Aila, Samuli Laine, and Jaakko Lehtinen. “Progressive Growing of GANs for Improved Quality, Stability, and Variation.” *ArXiv (Cornell University)*, October 27, 2017. https://doi.org/10.48550/arxiv.1710.10196

[4] Liu, Ming-Yu, Xun Huang, Jiahui Yu, and Arun Mallya. “Generative Adversarial Networks for Image and Video Synthesis: Algorithms and Applications.” *Proceedings of the IEEE* 109, no. 5 (May 1, 2021): 839–62. https://doi.org/10.1109/jproc.2021.3049196

[5] Shahriar, Sakib. “GAN Computers Generate Arts? A Survey on Visual Arts, Music, and Literary Text Generation Using Generative Adversarial Network.” *Displays* 73 (July 1, 2022): 102237. https://doi.org/10.1016/j.displa.2022.102237

[6] Han, Xinying, Yang Wu, and Rui Wan. “A Method for Style Transfer from Artistic Images Based on Depth Extraction Generative Adversarial Network.” *Applied Sciences* 13, no. 2 (January 8, 2023): 867. https://doi.org/10.3390/app13020867

[7] Meng, Nan, Jia Yang, and Haibo Wang. "Icon Art Design with Generative Adversarial Network under Deep Learning." *Wireless Communications and Mobile Computing* 2022 (September 6, 2022): 1–9. https://doi.org/10.1155/2022/3499570

[8] J. Deepika, P. Selvaraju, Mahesh Kumar Thota, Mohit Tiwari, Dunde Venu, K. Manjulaadevi and N. Geetha Lakshmi 2023. Efficient classification of kidney disease detection using Heterogeneous Modified Artificial Neural Network and Fruit Fly Optimization Algorithm. Journal of Advanced Research in Applied Sciences and Engineering Technology. 31, no. 3 (Aug. 2023), 1–12. DOI: https://doi.org/10.37934/araset.31.3.112

[9] Marra, Francesco, Diego Gragnaniello, Davide Cozzolino, and Luisa Verdoliva. *Detection of GAN-Generated Fake Images over Social Networks*, 2018. https://doi.org/10.1109/mipr.2018.00084

[10] Yu, Ning, Larry S. Davis, and Mario Fritz. *Attributing Fake Images to GANs: Learning and Analyzing GAN Fingerprints*, 2019. https://doi.org/10.1109/iccv.2019.00765

[11] Chen, Chen, Xinwei Zhao, and Matthew C. Stamm. "Generative Adversarial Attacks against Deep-Learning-Based Camera Model Identification." *IEEE Transactions on Information Forensics and Security*, January 1, 2022, 1. https://doi.org/10.1109/tifs.2019.2945198

[12] Eykholt, Kevin, Ivan Evtimov, Earlence Fernandes, Bo Li, Amir Rahmati, Chaowei Xiao, Atul Prakash, Tadayoshi Kohno, and Dawn Song. *Robust Physical-World Attacks on Deep Learning Visual Classification*, 2018. https://doi.org/10.1109/cvpr.2018.00175

[13] Munanday, A.P., Sazali, N., Wan Harun, W.S., Kumaran Kadirgama and Ahmad Shahir Jamaludin 2023. Analysis of Convolutional Neural Networks for Facial Expression Recognition on GPU, TPU and CPU. Journal of Advanced Research in Applied Sciences and Engineering Technology. 31, no. 3 (Aug. 2023), 50–67. DOI: https://doi.org/10.37934/araset.31.3.5067

[14] Rössler, Andreas, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. *FaceForensics++: Learning to Detect Manipulated Facial Images*, 2019. https://doi.org/10.1109/iccv.2019.00009

[15] Elmasry, Ramez M., Mohammed A.-m. Salem, Omar M. Fahmy, and Mohamed Abd El Ghany. *Image Enhancement Using Recursive Anisotropic and Stationary Wavelet Transform*, 2023. https://doi.org/10.1109/iwssip58668.2023.10180278

[16] Fahmy, Gamal, Mahmoud Fahmy, and O. M. Fahmy. *Bivariate Double Density Discrete Wavelet for Enhanced Image Denoising*, 2021. https://doi.org/10.1109/niles53778.2021.9600554

[17] Durall, Ricard, Margret Keuper, Franz-Josef Pfreundt, and Janis Keuper. "Unmasking DeepFakes with Simple Features." arXiv, 2019. https://doi.org/10.48550/ARXIV.1911.00686

[18] Islam, Md. Robiul. "SUPERIORITY OF WAVELET THEORY COMPARED TO FOURIER TRANSFORM." *Khulna University Studies*, October 16, 2022, 119–22. https://doi.org/10.53808/kus.2006.7.1.0512-ps

[19] Wirsing, Karlton. "Time Frequency Analysis of Wavelet and Fourier Transform." In *IntechOpen EBooks*, 2021. https://doi.org/10.5772/intechopen.94521

[20] Bahri, Syamsul, Lailia Awalushaumi, and Marliadi Susanto. *The Approximation of Nonlinear Function Using Daubechies and Symlets Wavelets*, 2018. https://doi.org/10.5220/0008521103000306

[21] Muhammad Khairi A Razak, Kamilah Abdullah and Suhaila Abd Halim 2023. Non-blind Image Watermarking Algorithm based on Non-Separable Haar Wavelet Transform against Image Processing and Geometric Attacks. *Journal of Advanced Research in Applied Sciences and Engineering Technology*. 29, no. 2 (Jan. 2023), 251–267. DOI:https://doi.org/10.37934/araset.29.2.251267

[22] Bhargavi Goparaju and Bandla Sreenivasa Rao 2023. Distributed Denial-of-Service (DDoS) Attack Detection using 1D Convolution Neural Network (CNN) and Decision Tree Model. *Journal of Advanced Research in Applied Sciences and Engineering Technology*. 32, 2 (Sep. 2023), 30–41. DOI:https://doi.org/10.37934/araset.32.2.3041..

[23] Elsheakh, Dalia M., Rawda A. Mohamed, O. M. Fahmy, Khaled Ezzat, and Angie R. Eldamak. "Complete Breast Cancer Detection and Monitoring System by Using Microwave Textile Based Antenna Sensors." *Biosensors* 13, no. 1 (January 4, 2023): 87. https://doi.org/10.3390/bios13010087

[24] Vikash Chander Maheshwari, Nurul Aida Osman and Norshakirah Aziz 2023. A Hybrid Approach Adopted for Credit Card Fraud Detection Based on Deep Neural Networks and Attention Mechanism. *Journal of Advanced Research in Applied Sciences and Engineering Technology*. 32, no. 1 (Sep. 2023), 315–331. DOI:https://doi.org/10.37934/araset.32.1.315331

[25] Sadek, Ahmed H., O. M. Fahmy, Mahmoud Nasr, and Mohamed K. Mostafa. "Predicting Cu(II) Adsorption from Aqueous Solutions onto Nano Zero-Valent Aluminum (NZVAl) by Machine Learning and Artificial Intelligence Techniques." *Sustainability* 15, no. 3 (January 21, 2023): 2081. https://doi.org/10.3390/su15032081