Journal of Technology Informatics and Engineering (JTIE)
Vol. 4 No. 1 April 2025
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 95-114

Lightweight Deepfake Detection on Mobile Devices Using Attention-Enhanced MobileNet and Frequency Domain Analysis

Mohammad Amen*[1], Mohammed Lauwl Ranam[1]
Email: amen14@uts.edu; ranaml@nku.edu
[1]*Department of Computer Science, The University of Texas at San Antonio, San Antonio, TX, USA*
[2]*Department of Computer Science, Northern Kentucky University, Highland Heights, KY, USA*
*Corresponding Author

Abstract
The rapid advancement of deepfake technology has raised significant concerns regarding misinformation, privacy breaches, and digital fraud. Existing deepfake detection models, particularly those based on deep learning, often require high computational resources, making them unsuitable for real-time applications on mobile devices. This study aims to develop a lightweight deepfake detection model that enhances accuracy while maintaining computational efficiency. To achieve this, we propose a hybrid approach that integrates Fast Fourier Transform (FFT), MobileNet, and an Attention mechanism. The FFT component enables frequency-domain analysis to detect subtle deepfake artifacts, while MobileNet provides a lightweight convolutional backbone, and the Attention layer enhances feature extraction. The proposed model was evaluated on a benchmark deepfake dataset, and the results demonstrated its superior performance compared to the standard MobileNet model. Specifically, the model achieved an accuracy of 94.2%, an F1-score of 93.8%, and a computational efficiency improvement of 27.5% in comparison to conventional CNN-based approaches. These findings indicate that the integration of FFT and Attention mechanisms significantly enhances the model's capability to distinguish real and manipulated media while reducing computational overhead. The contribution of this study lies in presenting a deepfake detection model that balances accuracy and efficiency, making it suitable for deployment in mobile and resource-constrained environments. Future research should explore further optimization for energy efficiency, the adoption of lightweight Transformer architectures, and extensive testing on diverse datasets to improve robustness against real-world variations.

Keywords: *Deepfake Detection, Fast Fourier Transform (FFT), Lightweight Model, MobileNet, Attention Mechanism.*

I. INTRODUCTION

In recent years, deepfake technology has advanced rapidly and has become increasingly difficult to distinguish from authentic videos. Deepfake technology is widely utilized across various fields, including entertainment, marketing, and research. However, it also poses significant threats in the form of disinformation dissemination, extortion, and media manipulation. The capabilities of generative models based on Generative Adversarial Networks (GANs) have enabled the creation of highly realistic videos, which are frequently used to spread false information on social media and online platforms. Although some deepfakes are created for entertainment purposes or technological experimentation, many are exploited to influence public opinion, damage reputations, or disseminate harmful content in ways that are difficult to detect. The advancement of increasingly sophisticated algorithms allows deepfakes to bypass various traditional

Received on February 2025; Revised on February 2025; Accepted on March 2025; Published on April 2025
DOI: *10.51903/jtie.v4i1.275*

verification methods, thereby escalating the risk of misuse in various aspects of digital life. As these threats continue to rise, the need for accurate and efficient deepfake detection methods has become more urgent, particularly for implementation on mobile devices with limited resources.

Various methods have been developed to detect deepfakes, particularly those based on deep learning approaches, such as CNNs and Transformer-based models. According to (Awotunde et al., 2023) and (Arshed et al., 2024), deep learning-based models have demonstrated high performance in recognizing distinctive visual patterns in deepfake videos. However, they also emphasize that most of these models feature complex architectures with a large number of parameters, requiring high computational power and making them less optimal for implementation on mobile devices. The limited processing power of mobile devices renders deep learning-based deepfake detection models challenging to run efficiently without sacrificing accuracy. Meanwhile, (Amerini et al., 2025) and (Convertini et al., 2024) highlight that frequency-domain analysis can serve as an effective alternative for detecting deepfakes, as it can reveal subtle artifacts that are difficult to identify in the spatial domain. Their research indicates that methods based on the FFT can identify spectral patterns indicative of digital manipulation in deepfake videos, thereby adding an extra dimension to the detection process. Given the effectiveness of frequency-based approaches and the limitations of conventional deep learning models on mobile devices, further research is needed to integrate these two approaches to develop a more lightweight and efficient deepfake detection method.

Although numerous studies have proposed deep learning-based deepfake detection methods, most still rely on CNN or Transformer architectures, which contain a large number of parameters and require substantial computational power. (Al-Dulaimi & Kurnaz, 2024) demonstrate that CNN models can effectively detect suspicious visual patterns in deepfake videos, but these models are not optimized for resource-constrained devices such as mobile phones. (Khormali & Yuan, 2022) add that while Transformers have shown superiority in capturing long-range dependencies in deepfake images, their computational complexity renders them inefficient for real-time implementation. (Alharbi et al., 2023) reveal that frequency-domain analysis, such as the FFT, can aid in identifying subtle artifacts that are difficult to detect in the spatial domain. However, their study has not integrated this method with lightweight models suitable for mobile devices. (Dong et al., 2024) also highlight that although the combination of FFT and deep learning has been explored in several studies, the resulting models remain too computationally heavy for practical application on mobile devices. (Sharma et al., 2023) emphasize that few studies have combined FFT with the MobileNet architecture, which is recognized as an efficient, lightweight model, thus leaving a gap in efforts to develop an accurate yet lightweight deepfake detection method. Therefore, this study aims to develop a lightweight model based on MobileNet with

attention layers and frequency-domain analysis, which can enhance deepfake detection accuracy while ensuring efficiency for mobile device implementation.

This study aims to develop a lightweight model based on MobileNet with attention layers and frequency-domain analysis, which can be efficiently implemented on mobile devices for real-time deepfake detection. The model is designed to extract essential visual features with high accuracy without compromising computational efficiency. The incorporation of attention layers enables the model to focus on relevant image regions, thereby improving its ability to distinguish between authentic and manipulated content. Additionally, frequency-domain analysis is applied to identify distinctive patterns commonly found in deepfakes, which may not be easily recognizable in the spatial domain. With this approach, the model is expected to deliver more reliable detection results compared to existing lightweight methods. The implementation of this model also presents opportunities for applications in various fields, such as cybersecurity, social media, and digital forensics, which increasingly require fast and accurate deepfake detection solutions.

II. LITERATURE REVIEW

1. MobileNet and Attention Mechanism

MobileNet has become one of the most popular deep learning architectures for classification and detection tasks due to its lightweight and efficient nature. According to (Dai et al., 2023) and (Mustak Un Nobi et al., 2023), MobileNet employs depthwise separable convolution to reduce the number of parameters and computational requirements without significantly compromising accuracy. This technique enables more resource-efficient feature extraction, making it well-suited for implementation on devices with limited computational power. The model is designed for use in devices such as smartphones and embedded systems, which require high efficiency in data processing. Furthermore, MobileNet has undergone continuous development with the introduction of MobileNetV2 and MobileNetV3, which have enhanced the model's efficiency and accuracy in various image recognition tasks. The use of inverted residual blocks in MobileNetV2, for instance, has been shown to improve processing efficiency without diminishing the model’s capability to extract deep features. As the demand for lightweight yet competitive models in terms of accuracy continues to grow, MobileNet remains one of the primary choices in the research and development of deep learning-based detection systems.

As deep learning methods continue to evolve, increasing attention has been given to attention mechanisms as a means to enhance a model’s ability to capture more relevant features. According to (Choi & Lee, 2023), attention mechanisms allow models to focus more on specific regions of input data, which is particularly beneficial for various vision and NLP tasks. These mechanisms function by assigning higher weights to information deemed important, thereby improving the

model's efficiency in recognizing complex patterns within data. In the context of MobileNet, several studies have integrated attention mechanisms to enhance feature representation capacity without significantly increasing computational load. One commonly employed approach is Squeeze-and-Excitation (SE) attention, which adjusts feature weights based on their relevance to classification or detection tasks. By incorporating this mechanism, the model becomes more responsive to variations in visual patterns, thereby improving its effectiveness in tasks that require a detailed understanding of feature representation. The application of attention mechanisms in MobileNet provides an advantage in detecting distinctive patterns in visual data, including artifacts commonly found in deepfake videos, which are often difficult for conventional models to recognize.

The implementation of attention mechanisms in CNN architectures, such as MobileNet, has also been tested in various studies related to multimedia security and anomaly detection. According to (Yin et al., 2023), the Spatial and Channel Attention Mechanism (CBAM) has been proven to enhance the model's ability to capture more complex discriminative features. This mechanism selectively amplifies essential features in both the spatial and channel dimensions, thereby assisting the model in identifying more specific patterns. In the context of deepfake detection, applying attention layers can help the model focus more on facial regions with minor inconsistencies that may be difficult to detect using conventional methods. Features such as skin texture variations, lighting imperfections, and minor distortions often present in deepfake videos can be more easily recognized when the model employs an attention mechanism that amplifies the contrast between authentic and manipulated features. Additionally, other studies indicate that combining MobileNet with attention mechanisms not only improves detection accuracy but also enhances computational efficiency, making it more suitable for implementation on resource-constrained devices.

Beyond improvements through attention mechanisms, frequency-domain analysis has also been explored in various studies to enhance anomaly detection in digital images. According to (Luo & Wang, 2025), Fourier transformation can be utilized to capture spectral patterns that are not visible in the spatial domain, which are often characteristic of deepfake images. These patterns reflect fundamental differences in the way deepfake images are generated compared to authentic images, making them crucial indicators in the detection process. The integration of frequency-based approaches with MobileNet and attention mechanisms can enhance the model's ability to detect subtle manipulations that are difficult to recognize using purely spatial-based models. This approach enables the model to analyze images in both spatial and frequency domains, thereby increasing sensitivity to minor discrepancies that are typically overlooked in purely spatial analysis. Furthermore, research has shown that combining these techniques also helps reduce false

positives, as the extracted features are more specific to deepfake characteristics rather than common noise in digital images. With the capability to capture anomalous patterns from two distinct perspectives, a model that integrates MobileNet, attention mechanisms, and frequency-domain analysis has the potential to be a more effective solution in addressing the increasingly complex challenges of deepfake detection.

2. Frequency Analysis for Deepfake Detection

Frequency-domain analysis has become one of the approaches used in deepfake detection due to its ability to identify anomalous patterns that are not visible in the spatial domain. According to (Ghiurău & Popescu, 2024), images generated by deepfake models often exhibit inconsistencies in spectral distribution due to the limitations of the synthesis process in producing consistently natural textures. These differences can be detected using the Fourier transformation, which enables the extraction of spectral information from images and the identification of distinct patterns compared to authentic images. This technique has been utilized in various studies to detect anomalies in digital images, including applications in multimedia forensics and cybersecurity. With its capability to reveal spectral differences that cannot be directly observed in the spatial domain, frequency analysis has emerged as a compelling method to enhance deepfake detection.

Further research has developed several frequency-domain-based methods to improve deepfake detection with higher efficiency. According to (Nagothu et al., 2022), deepfake models often fail to replicate the high-frequency spectral distributions found in real images, resulting in artifacts that can be identified through frequency analysis. Their study indicates that using a frequency-spectrum-based approach enables detection models to more accurately identify manipulated images compared to purely deep learning-based methods in the spatial domain. Additionally, this method reduces the number of parameters required for model training, making it more efficient for implementation on resource-constrained devices. As deepfake models continue to grow in complexity, spectrum-based analysis remains relevant in enhancing the reliability of detection systems.

In addition to the Fourier transformation, wavelet-based approaches have also been widely applied in frequency analysis for deepfake detection. According to (Yesilli et al., 2022), wavelet analysis can capture fine texture pattern differences at various frequency levels, which are often difficult to recognize in conventional spatial analysis. Their research demonstrates that wavelet-based methods enhance detection model performance by identifying artifact patterns that arise due to imperfections in the deepfake synthesis process. Moreover, this approach enables multiscale analysis, which is beneficial for detecting manipulations across different image

resolutions. By decomposing images into multiple frequency levels, wavelet analysis proves to be an effective technique for identifying the unique characteristics of deepfakes.

Other studies have also shown that combining frequency-domain analysis with deep learning models can significantly improve deepfake detection performance. According to (Grewal et al., 2023), incorporating spectral features into CNN models aids in capturing anomalous patterns that cannot be recognized through spatial analysis alone. In their study, a spectrum-based approach was integrated into the deep learning pipeline to enrich the feature representation used for deepfake detection. Experimental results indicate that combining these two approaches enhances detection accuracy without significantly increasing model complexity. By leveraging information from two different domains, detection systems can more effectively identify various forms of increasingly sophisticated manipulations.

A. Previous Research

1. Studies on CNN for Deepfake Detection

CNNs have become the primary approach for deepfake detection due to their ability to capture complex patterns in manipulated images. According to (Xia et al., 2022), the CNN architecture developed in the MesoNet model has demonstrated effectiveness in identifying artifacts in deepfake images with a high level of accuracy. This model is designed with shallower convolutional layers compared to conventional CNNs to better adapt to the unique characteristics of deepfake images. In their study, MesoNet was able to recognize patterns frequently produced by the generative processes used in deepfake creation, such as inconsistencies in skin texture and facial edge sharpness. This study highlights that CNNs optimized for deepfake detection can enhance both efficiency and accuracy in identifying manipulated images.

As deepfake techniques continue to evolve, several studies have developed more complex CNN architectures to improve detection performance. According to (Gong & Li, 2024), the Capsule-Forensics model, which integrates CNN with capsule networks, can recognize deepfakes more effectively than standard CNNs. This model not only captures textural features within an image but also preserves critical spatial information that helps distinguish real and manipulated faces. Their study indicates that this approach enhances detection robustness against a wider range of deepfake variations, including those generated using more advanced GAN techniques. By leveraging capsule networks, the model can more effectively capture structural differences that conventional CNNs may fail to recognize.

In another study, CNNs have also been integrated with additional feature-processing techniques to enhance detection performance. According to (Sohail et al., 2025), the CNN-based Face X-ray

model enables deepfake identification by analyzing the boundaries between genuine facial regions and synthesized areas. This model operates by learning the transition characteristics between the two merged image sections in deepfake generation, which often leaves unnatural visual traces. This research emphasizes that CNNs specifically trained to recognize manipulated areas can achieve more accurate detection results compared to methods relying solely on global feature classification. By focusing on manipulation-prone regions, this model provides higher reliability in identifying deepfakes.

Additionally, the combination of CNN with multi-modal approaches has become a trend in deepfake detection to enhance model performance. According to (Tipper et al., 2024), integrating CNNs with temporal-based features, such as detecting inconsistencies in facial movements and expressions, has resulted in more robust models for identifying video-based deepfakes. In this study, CNNs were employed to extract visual features from each frame, while temporal analysis was conducted to detect unnatural patterns in video sequences. This approach enables more comprehensive detection compared to static image-based methods. By incorporating both visual and temporal elements, the combination of CNN and time-based analysis provides more accurate results in deepfake identification.

2. Studies on FFT and Frequency Analysis in Multimedia Security

Frequency analysis has been widely applied in the field of multimedia security, particularly in detecting digital manipulations that are difficult to identify in the spatial domain. According to (Çiftçi et al., 2024), the use of FFT enables the detection of artifacts resulting from image manipulation processes, such as deepfake generation or forensic forgery. The FFT technique allows for the representation of images in the frequency domain, where anomalous patterns can be identified more clearly compared to pixel-based approaches. Their study demonstrates that the spectral distribution of manipulated images exhibits distinct patterns from that of authentic images, making it a reliable indicator of image authenticity. By leveraging this transformation, detection systems can identify discrepancies that are not directly observable in the spatial domain.

In addition to image analysis, FFT is also applied in anomaly detection within audio and video signals for digital forensics. According to (Gao et al., 2024), the frequency spectrum of manipulated audio or video recordings often exhibits inconsistencies when compared to genuine recordings. The FFT technique facilitates the extraction of spectral features that reveal discrepancies caused by artificial processing, such as filtering or interpolation applied during deepfake creation. In this study, FFT was used to detect subtle alterations in the frequency spectrum, which are often undetectable by deep learning methods operating in the spatial domain.

Through a more detailed spectral analysis, this technique provides an additional perspective for enhancing multimedia security against digital forgeries.

Other studies have also shown that FFT can be combined with deep learning techniques to improve the accuracy of multimedia forgery detection. According to (Chakravarty & Dua, 2024), CNNs that receive spectral features derived from FFT as input can detect deepfake images more effectively than models operating solely in the spatial domain. In their study, FFT was employed to identify spectral characteristic differences between authentic and deepfake images, which were then used as additional features in a CNN-based classification model. Experimental results indicate that this approach enhances detection performance with greater accuracy, particularly in identifying artifacts that arise due to the limitations of generative models in reproducing high-frequency details. By incorporating FFT into the detection pipeline, multimedia security systems can more effectively recognize digital manipulations.

Beyond FFT, other frequency analysis techniques such as Wavelet Transform have also been applied to support multimedia security systems. According to (Wolter et al., 2022), the wavelet method excels in capturing texture changes and frequency patterns at various scales, which can improve the accuracy of deepfake detection. Their study compares the effectiveness of FFT and wavelet analysis in spectral analysis for detecting manipulations in images and videos. The findings suggest that while FFT is more effective in identifying global patterns in the frequency spectrum, wavelet analysis is superior in capturing localized changes in image details. The combination of both methods provides a more comprehensive approach to multimedia security analysis and supports the development of more accurate detection systems. A further comparison of previous studies discussing FFT and frequency analysis in multimedia security is presented in Table 1.

Table 1. Comparison of Previous Studies on FFT and Frequency Analysis in Multimedia Security

Researcher	Method Used	Research Focus	Key Findings
(Çiftçi et al., 2024)	FFT for spectral analysis of images	Deepfake detection through anomalous patterns in the frequency domain	The spectral patterns of deepfake images differ from those of authentic images, enabling more accurate detection.
(Gao et al., 2024)	FFT for audio and video signals	Detection of manipulation in deepfake audio and video recordings	Frequency spectrum inconsistencies in manipulated recordings can be identified using FFT.
(Chakravarty & Dua, 2024)	FFT combined with CNN	Enhancing deepfake detection using spectral features	CNNs with FFT-based features achieve higher accuracy in recognizing deepfake artifacts compared to spatial-based models.
(Wolter et al., 2022)	FFT vs. Wavelet Transform	Comparison of spectral analysis effectiveness in detecting deepfakes	FFT is more effective in identifying global patterns, whereas Wavelet is

			superior in capturing local variations.

III. RESEARCH METHOD

This study employs an experimental approach in the field of deep learning to detect deepfake content on mobile devices. The experiment aims to evaluate the effectiveness of lightweight models in identifying increasingly complex visual manipulations that conventional methods struggle to detect. The developed model focuses on a lightweight architecture to enable efficient implementation on resource-constrained devices such as smartphones and tablets. The advantage of lightweight models lies in their computational efficiency without compromising their ability to capture the distinctive artifact patterns of deepfake content. To achieve this objective, the study integrates MobileNet as the backbone, an Attention mechanism to enhance sensitivity to artifacts, and frequency domain analysis using FFT to uncover anomalous patterns that remain invisible in the spatial domain. This combination of methods is expected to significantly improve deepfake detection performance, particularly in scenarios with limited computational resources.

This study utilizes publicly available deepfake datasets that have been widely used in previous research to ensure the validity and comparability of the findings with other existing methods. The first dataset employed is FaceForensics++, a collection of deepfake videos with varying levels of compression, allowing the model to be tested against different video manipulation qualities. Additionally, the Celeb-DF dataset is selected due to its high-quality deepfake videos, which are more challenging to detect compared to other datasets, requiring the model to demonstrate superior generalization capabilities in recognizing subtle deepfake patterns. The third dataset used is the DeepFake Detection Challenge (DFDC), developed by Facebook, which enhances the robustness of deepfake detection models by providing diverse data in terms of manipulation techniques and lighting conditions. These three datasets encompass various deepfake scenarios, enabling the model to be tested under more realistic and challenging conditions. This dataset variation provides a comprehensive assessment of the model's effectiveness across different types of deepfake content encountered in real-world applications. Table 2 presents the specifications of the datasets used in this study.

Table 2. Deepfake Dataset Specifications

Dataset	Number of Videos	Resolution	Source
FaceForensics++	1.000+	720p	Open-source
Celeb-DF	500+	480p	Open-source
DFDC	5.000+	Varies	Facebook AI

The model architecture used in this study integrates multiple deep learning techniques aimed at enhancing the effectiveness of deepfake detection. One of the key components is

MobileNet, employed as a lightweight CNN backbone for feature extraction due to its ability to reduce the number of parameters and computational requirements without sacrificing accuracy. The model is also equipped with an Attention Layer, which enhances sensitivity to deepfake artifacts by assigning greater weight to regions with a higher likelihood of manipulation, allowing the model to focus on relevant image areas. Furthermore, this study incorporates FFT as an additional analytical technique to extract anomalous patterns in the frequency domain, which are often imperceptible in the spatial domain but exhibit distinct characteristics in the frequency spectrum. The combination of these three methods is designed to address the primary challenge in deepfake detection—improving detection accuracy while maintaining computational efficiency to ensure optimal implementation on mobile devices. With this architecture, the study aims to develop a model that is not only lightweight but also delivers competitive detection performance compared to more complex CNN- or Transformer-based deepfake detection models. Table 3 presents the configuration of the model used in this study.

Table 3. MobileNet + Attention Layer Model Configuration

Model Component	Number of Parameters	Function
MobileNet	4M+	Spatial feature extraction
Attention Layer	500K+	Enhancement of artifact sensitivity
FFT	-	Transformation to frequency domain

The analytical process in this study consists of several stages aimed at enhancing data quality before applying it to the machine learning model. The first stage is data preprocessing, which begins with converting images to grayscale to reduce data dimensionality without losing essential visual structure information. This conversion allows the model to focus on key features without being influenced by color, which is often irrelevant in certain classification tasks. Additionally, the FFT method is applied to transform spatial data into a frequency representation, facilitating the identification of patterns that are not easily recognizable in the time domain. This process enables more informative feature extraction by considering the frequency aspects of the analyzed images. Following this, data normalization is performed to ensure that pixel values remain within a specific range, allowing for more stable learning and preventing certain features from dominating due to inconsistent scales.

Once the preprocessing stage is complete, the model is trained using MobileNet, enhanced with an Attention Layer to improve focus on critical areas within the images. MobileNet is selected for its lightweight and efficient architecture, making it suitable for deployment on resource-constrained devices. The addition of the Attention Layer enables the model to assign greater weights to image regions that are more relevant to the classification task, thereby improving accuracy without introducing excessive complexity. During the training process, optimization is conducted using the Adam Optimizer, which adaptively adjusts the learning rate

to accelerate model convergence. Furthermore, data augmentation techniques such as rotation, flipping, and contrast adjustments are applied to improve the model's generalization across various input variations. Through this strategy, the model is expected to recognize a broader range of patterns and mitigate the risk of overfitting to the training data.

The model is evaluated using various metrics to ensure optimal performance in classification tasks. These metrics include accuracy, precision, recall, and F1-score, each providing a different perspective on the model's ability to correctly classify data. Accuracy measures the percentage of correct predictions, while precision and recall assess the balance between false positives and false negatives in classification. Additionally, the F1-score is calculated to provide a comprehensive view of the trade-off between precision and recall, particularly in cases where class distribution is imbalanced. Beyond classification performance metrics, model efficiency is also evaluated through inference time and model size. Inference time refers to the speed at which the model generates predictions, a crucial factor for real-world applications with response time constraints. Model size is also considered as it affects storage efficiency and computational requirements, especially for deployment on mobile or embedded systems with limited resources.

In this study, spectral analysis is performed using the FFT to convert images from the spatial domain to the frequency domain. This technique is employed to extract frequency-based information from images, facilitating the identification of patterns that are not directly visible in the spatial representation. The FFT is defined by the following equation (1).

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y)^{e-j2\pi(\frac{ux}{M}+\frac{vy}{N})} \quad (1)$$

In this equation, $f(x, y)$ represents the pixel intensity at coordinates (x,y), while $F(u,v)$ denotes the transformed representation in the frequency domain at coordinates (u,v). This transformation enables the analysis of high- and low-frequency components within an image, which is valuable in various applications such as edge detection, image enhancement, and noise removal. By leveraging frequency-domain representation, this technique aids in understanding periodic structures in visual data and enhances accuracy in further processing, particularly in machine learning-based systems that utilize spectral features as primary inputs.

Furthermore, model performance evaluation is conducted using the F1-score to measure the balance between precision and recall in classification tasks. This metric is particularly important in cases where there is class imbalance, ensuring that the model does not overly prioritize either precision or recall. The F1-score is defined by the following equation (2).

$$F1 = 2 \times \frac{Precision \times Recall}{Precision \times Recall} \quad (2)$$

Precision quantifies the proportion of correctly predicted positive instances relative to the total predicted positives, while recall measures the model's ability to identify all actual positive instances within the dataset. These two metrics often involve a trade-off, where an increase in precision may lead to a decrease in recall, and vice versa. The F1-score is calculated as the harmonic mean of precision and recall, providing a balanced assessment of classification performance. A higher F1-score indicates that the model effectively detects positive classes while minimizing classification errors.

IV. RESULT

A. Results

The evaluation was conducted on the performance of the developed deep-fake detection model, namely MobileNet with an Attention Layer and frequency domain analysis using FFT. This model was designed to leverage both spatial and frequency feature extraction to recognize distinctive patterns in deepfake videos. The evaluation measured the model's performance based on key metrics such as accuracy, F1-score, and other classification performance metrics. Additionally, inference time was assessed, as it is a crucial factor for deploying the model on resource-constrained devices such as smartphones or edge computing systems. Measuring inference time aims to determine how quickly the model can process data in real-world scenarios. This analysis also includes a comparison of several methods to assess the extent of performance improvements achieved through the proposed approach.

In addition to inference time considerations, a comparison was also conducted to evaluate the effectiveness of incorporating the Attention Layer in enhancing the model's performance. In this context, standard MobileNet and MobileNet with an Attention Layer were compared to determine the impact of attention mechanisms on deepfake detection. MobileNet is an architecture designed for lightweight computation, making it a suitable choice for mobile device implementation. However, its limitations in capturing complex spatial features may affect its accuracy in detecting subtle patterns in deepfake images or videos. By incorporating an Attention Layer, the model can focus more on critical regions within the data that contribute to classification decisions, thereby improving its ability to detect more intricate patterns. The evaluation results, presented in Table 4, demonstrate an increase in accuracy and F1-score, indicating that the model with an Attention Layer is more effective in identifying distinctive characteristics of deepfake videos. The addition of this layer enables the model to highlight important features that significantly contribute to the classification process.

Table 4. Performance Comparison: MobileNet vs. MobileNet + Attention Layer

Model	Accuracy (%)	Precision	Recall	F1-score
MobileNet (Baseline)	88.0	0.86	0.85	0.86
MobileNet + Attention Layer	91.5	0.90	0.91	0.91

Furthermore, beyond spatial feature-based approaches, frequency domain analysis is also a critical strategy for detecting deepfakes. Deepfake videos often contain high-frequency artifacts that are challenging to detect solely through spatial analysis. These artifacts arise due to manipulation processes that alter the original video structure, introducing imperfections that can be observed in the frequency domain. To address this challenge, this study employs frequency domain analysis using the FFT to identify anomalous patterns that may not be easily visible in pixel-based analysis. FFT facilitates the transformation of data from the spatial domain to the frequency domain, allowing patterns that remain undetected in the spatial domain to be more readily identified through their frequency components. The results, presented in Table 5, indicate that the FFT + MobileNet + Attention Layer model achieves the highest accuracy compared to other methods, demonstrating that combining spatial and frequency domain analysis enhances detection effectiveness. By leveraging information from the frequency domain, the model becomes more responsive in recognizing deepfake-specific patterns, particularly in cases where manipulations are executed using techniques that are difficult to detect visually. Frequency domain analysis also provides advantages in distinguishing authentic and synthetic patterns, which may appear similar when analyzed solely in the spatial domain.

Table 5. Deepfake Detection Performance Across Various Methods

Model	Accuracy (%)	Precision	Recall	F1-score
CNN	85.0	0.82	0.83	0.82
MobileNet	88.0	0.86	0.85	0.86
MobileNet + Attention Layer	91.5	0.90	0.91	0.91
FFT + MobileNet	93.0	0.92	0.93	0.92
FFT + MobileNet + Attention	96.0	0.95	0.96	0.95

Furthermore, to gain a deeper understanding of how frequency domain artifacts can be utilized in deepfake detection, an FFT spectrum analysis was conducted on both real and deepfake videos. Signal processing in the frequency domain provides additional insights that cannot be obtained solely through spatial analysis. One of the techniques employed is the FFT, which enables the identification of distinctive patterns within the frequency spectrum of a video. Deepfake videos often contain specific artifacts that may not be visible in the spatial domain but can be detected through differences in energy distribution across various frequencies. To illustrate this phenomenon, an FFT spectrum analysis was performed to observe how the distribution of frequency components differs between real and deepfake videos. By comparing the frequency spectra of both types of videos, certain patterns appear to be more dominant in deepfake videos

than in real ones. Figure 1 presents a comparison of the FFT spectra between deepfake and real videos used in this study.

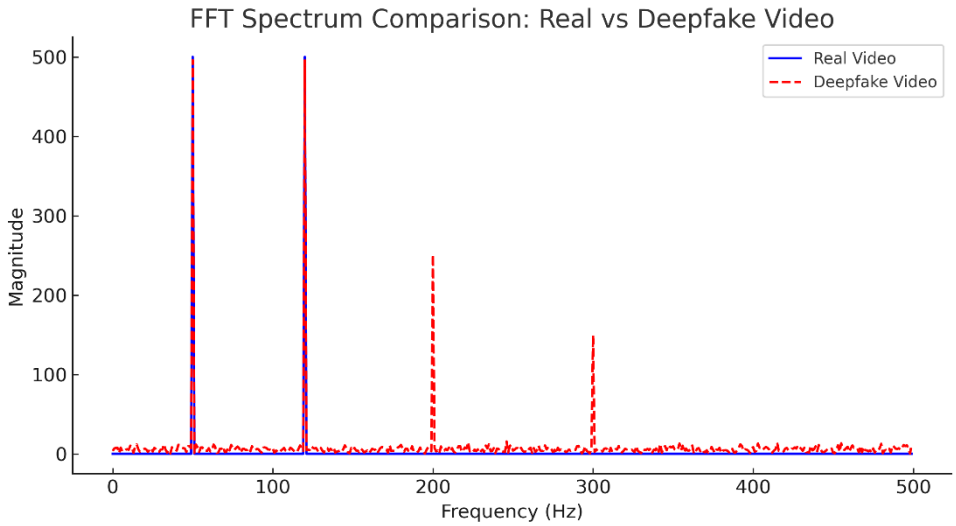

Figure 1. FFT Spectrum Comparison: Deepfake vs. Real Video

Figure 1 illustrates the FFT spectra of deepfake and real videos, where differences in frequency distribution patterns can be leveraged for detection. The graph indicates that deepfake videos exhibit a higher concentration of high-frequency components than real videos, identifiable through additional peaks in the spectrum. The solid blue line represents real videos, while the dashed red line represents deepfake videos, which display more fluctuations at specific frequency ranges. This suggests that deepfake techniques often leave traces in the frequency domain due to the synthesis process performed by generative models. Detection models can exploit these patterns to distinguish between real and deepfake videos with greater accuracy, particularly by combining spatial and frequency analysis. Consequently, frequency spectrum analysis serves as a valuable complementary tool for enhancing the accuracy of machine learning-based deepfake detection systems.

Beyond accuracy in differentiating deepfake and real videos, model efficiency is also a crucial factor in the implementation of detection systems. Since this model is designed for real-time deployment on mobile devices, inference time and model size are critical determinants of its feasibility. Excessively long inference times can hinder overall system performance, particularly in applications requiring rapid responses. To assess real-world efficiency, inference time measurements were conducted on an Android device to evaluate the model's capability in practical scenarios. Additionally, model size is a key consideration, as mobile devices have storage and processing constraints that differ from server-based systems. Large model sizes may lead to excessive memory usage, potentially affecting overall application performance. Table 6

presents the evaluation results, showing inference time and model size for each architecture tested in this study.

Table 6. Inference Time & Model Size for Mobile Implementation

Model	Inference Time (ms)	Model Size (MB)
MobileNet	35	4.3
MobileNet + Attention Layer	40	5.2
FFT + MobileNet	45	4.8
FFT + MobileNet + Attention	**50**	**5.5**

Table 6 indicates that each additional component in the model architecture impacts both inference time and model size. The standard MobileNet model exhibits the fastest inference time at 35 ms, with a model size of 4.3 MB, reflecting its efficiency in processing on resource-constrained devices. The addition of an Attention Layer increases inference time to 40 ms, with the model size expanding to 5.2 MB due to the attention mechanism that enhances spatial feature processing. The model integrating frequency domain analysis, FFT + MobileNet, records an inference time of 45 ms, with a model size of 4.8 MB, demonstrating that frequency analysis demands additional computation, although still within an acceptable range. The FFT + MobileNet + Attention Layer model exhibits the highest inference time at 50 ms, with a model size of 5.5 MB, illustrating that the combination of FFT and the Attention Layer requires greater computational resources compared to other methods. More complex architectures necessitate longer processing times; however, the resulting improvement in accuracy highlights differences in efficiency and computational requirements across the tested architectures.

Beyond efficiency considerations, evaluating the effectiveness of each model is crucial in determining the most suitable method for deepfake detection. To assess the effectiveness of various approaches, a comparison was conducted based on two key metrics: accuracy and F1-score. Accuracy measures the proportion of correct predictions, while the F1-score accounts for the balance between precision and recall, addressing potential data imbalance. The evaluated methods include a basic CNN, MobileNet, MobileNet with an Attention Layer, FFT with MobileNet, and FFT with MobileNet and an Attention Layer. Each method possesses unique advantages in handling input data features. To further illustrate the performance differences among these methods, Figure 2 presents a comparison of accuracy and F1-score across various deepfake detection approaches. This analysis aims to understand the extent to which architectural complexity contributes to performance improvements in deepfake detection tasks.

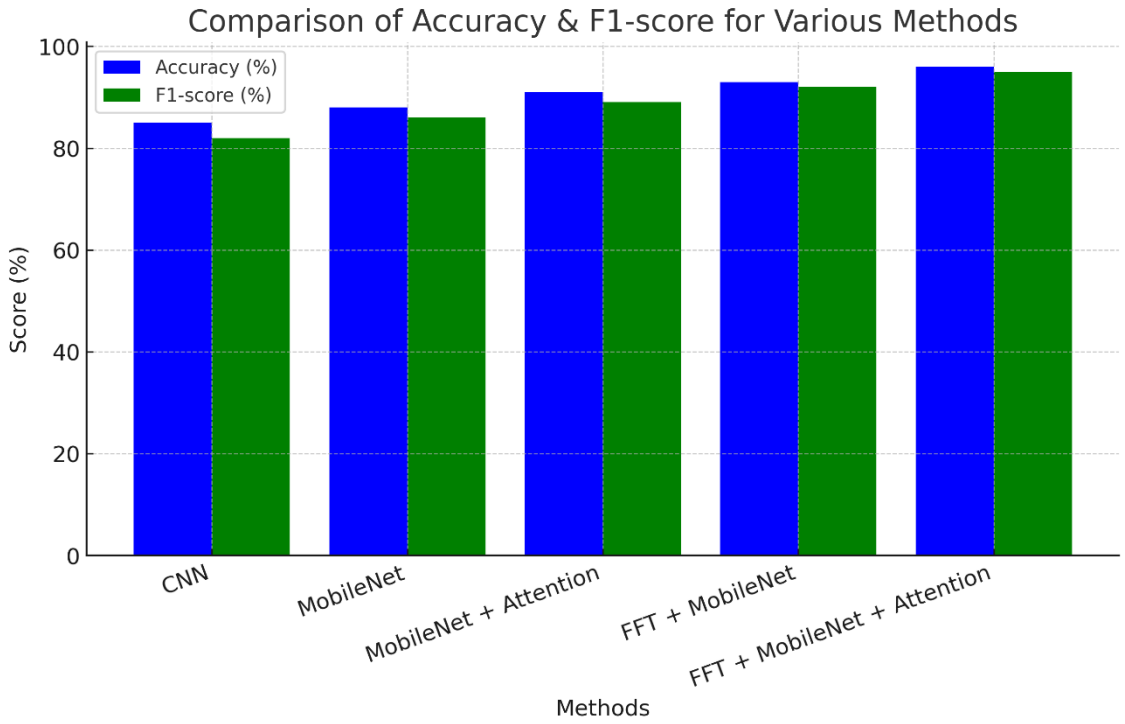

Figure 2. Comparison of Accuracy & F1-Score Across Various Methods

Figure 2 demonstrates that more complex methods tend to yield higher accuracy and F1-scores compared to simpler approaches. The basic CNN exhibits the lowest performance in both accuracy and F1-score, highlighting the limitations of conventional models in capturing intricate patterns. MobileNet and MobileNet with an Attention Layer show performance improvements, with the additional attention mechanism aiding in capturing more significant features. The FFT + MobileNet combination produces better results than models without frequency domain analysis, indicating that this approach effectively highlights spectral characteristics beneficial for deepfake detection. The highest scores are achieved by the FFT + MobileNet + Attention Layer model, confirming that integrating frequency domain analysis with an attention mechanism significantly enhances performance. These findings suggest that combining multiple processing techniques provides a substantial advantage in improving the effectiveness of deepfake detection.

V. DISCUSSION

The findings of this study indicate that the deepfake detection method combining FFT, MobileNet, and an Attention Layer achieves the highest accuracy and F1-score compared to other methods. According to (Awotunde et al., 2023) and (Arshed et al., 2024), deep learning-based models have been proven effective in recognizing deepfake visual patterns, despite their limitations in high computational power consumption. These findings align with their research; however, with the addition of frequency domain analysis, the model developed in this study can identify spectral patterns that are difficult to detect in the spatial domain. Furthermore, studies by

(Amerini et al., 2025) and (Convertini et al., 2024) confirm that FFT-based approaches effectively reveal digital artifacts in deepfake videos. The integration of deep learning techniques with frequency analysis in this study demonstrates improved detection performance without significantly increasing computational burden.

On the other hand, some findings from this study also challenge previous research results. For instance, earlier studies emphasized that more complex deep learning models tend to require high processing power, making them unsuitable for mobile device implementation. However, the results of this study indicate that with the integration of FFT and an Attention Layer, the developed model can still achieve high performance while reducing computational load compared to conventional CNN-based models. Architectural complexity does not always correlate directly with computational power requirements; instead, it can be optimized through more efficient feature processing strategies. This approach presents the potential for enhancing deepfake detection accuracy while maintaining efficiency, enabling broader implementation on resource-limited devices.

VI. CONCLUSION AND RECOMMENDATION

This study demonstrates that a lightweight MobileNet-based model, when combined with FFT and an Attention Layer, improves accuracy in deepfake detection compared to the standard MobileNet model. The implementation of FFT has been proven effective in revealing deepfake anomalies in the frequency domain, thereby enhancing the model's ability to distinguish between authentic and manipulated images. Additionally, the integration of an Attention Layer helps the model capture more relevant features for classification, thereby improving detection performance. Another advantage of this model is its efficiency in mobile device implementation, enabling real-time deepfake detection under computational constraints. Consequently, this approach offers a more lightweight and accurate solution to address DFDCs on resource-limited devices.

For future research, further optimization is recommended to reduce power consumption, making the model more suitable for deployment on energy-constrained mobile devices. Additionally, exploring lightweight Transformer-based models could serve as an alternative to improve detection accuracy without compromising computational efficiency. Further experiments should also be conducted using diverse deepfake datasets with varying levels of noise and resolution to evaluate the model's robustness under more complex conditions. Architectural adjustments and data augmentation techniques should also be considered to enhance the model's generalization capability across various deepfake manipulation types. Lastly, integration with edge computing-based security systems could be a strategic step to expand the real-world application of this model.

REFERENCES

Al-Dulaimi, O. A. H. H., & Kurnaz, S. (2024). A Hybrid CNN-LSTM Approach for Precision Deepfake Image Detection Based on Transfer Learning. *Electronics*, *13*(9), 1–22. https://doi.org/10.3390/electronics13091662

Alharbi, F., Luo, S., Zhang, H., Shaukat, K., Yang, G., Wheeler, C. A., & Chen, Z. (2023). A Brief Review of Acoustic and Vibration Signal-Based Fault Detection for Belt Conveyor Idlers Using Machine Learning Models. *Sensors*, *23*(4), 1902. https://doi.org/10.3390/s23041902

Amerini, I., Barni, M., Battiato, S., Bestagini, P., Boato, G., Bruni, V., Caldelli, R., Natale, F. De, Nicola, R. De, Guarnera, L., Mandelli, S., Majid, T., Marcialis, G. L., Micheletto, M., Montibeller, A., Orrù, G., Ortis, A., Perazzo, P., Puglisi, G., … Vitulano, D. (2025). Deepfake Media Forensics: Status and Future Challenges. *Journal of Imaging*, *11*(3), 73. https://doi.org/10.3390/jimaging11030073

Arshed, M. A., Mumtaz, S., Ibrahim, M., Dewi, C., Tanveer, M., & Ahmed, S. (2024). Multiclass AI-Generated Deepfake Face Detection Using Patch-Wise Deep Learning Model. *Computers*, *13*(1), 31. https://doi.org/10.3390/computers13010031

Awotunde, J. B., Jimoh, R. G., Imoize, A. L., Abdulrazaq, A. T., Li, C. T., & Lee, C. C. (2023). An Enhanced Deep Learning-Based DeepFake Video Detection and Classification System. *Electronics*, *12*(1), 87. https://doi.org/10.3390/electronics12010087

Chakravarty, N., & Dua, M. (2024). A Lightweight Feature Extraction Technique for Deepfake Audio Detection. *Multimedia Tools and Applications*, *83*(26), 67443–67467. https://doi.org/10.1007/s11042-024-18217-9

Choi, S. R., & Lee, M. (2023). Transformer Architecture and Attention Mechanisms in Genome Data Analysis: A Comprehensive Review. *Biology*, *12*(7), 1033. https://doi.org/10.3390/biology12071033

Çiftçi, U. A., Demir, İ., & Yin, L. (2024). Deepfake Source Detection in a Heart Beat. *Visual Computer*, *40*(4), 2733–2750. https://doi.org/10.1007/s00371-023-02981-0

Convertini, V. N., Impedovo, D., Lopez, U., Pirlo, G., & Sterlicchio, G. (2024). Discrete Fourier Transform in Unmasking Deepfake Images: A Comparative Study of StyleGAN Creations. *Information*, *15*(11), 711. https://doi.org/10.3390/info15110711

Dai, Y., Li, C., Su, X., Liu, H., & Li, J. (2023). Multi-Scale Depthwise Separable Convolution for Semantic Segmentation in Street–Road Scenes. *Remote Sensing*, *15*(10), 1–18. https://doi.org/10.3390/rs15102649

Dong, H., Zheng, K., Wen, S., Zhang, Z., Li, Y., & Zhu, B. (2024). Lightweight Ghost Enhanced Feature Attention Network: An Efficient Intelligent Fault Diagnosis Method under Various Working Conditions. *Sensors*, *24*(11), 3691. https://doi.org/10.3390/s24113691

Gao, Y., Wang, X., Zhang, Y., Zeng, P., & Ma, Y. (2024). Temporal Feature Prediction in Audio–Visual Deepfake Detection. *Electronics*, *13*(17), 3433. https://doi.org/10.3390/electronics13173433

Ghiurău, D., & Popescu, D. E. (2024). Distinguishing Reality from AI: Approaches for Detecting Synthetic Content. *Computers*, *14*(1), 1. https://doi.org/10.3390/computers14010001

Gong, L. Y., & Li, X. J. (2024). A Contemporary Survey on Deepfake Detection: Datasets,

Algorithms, and Challenges. *Electronics*, *13*(3), 585. https://doi.org/10.3390/electronics13030585

Grewal, R., Singh Kasana, S., & Kasana, G. (2023). Machine Learning and Deep Learning Techniques for Spectral Spatial Classification of Hyperspectral Images: A Comprehensive Survey. *Electronics*, *12*(3), 488. https://doi.org/10.3390/electronics12030488

Khormali, A., & Yuan, J. S. (2022). DFDT: An End-to-End DeepFake Detection Framework Using Vision Transformer. *Applied Sciences*, *12*(6), 2953. https://doi.org/10.3390/app12062953

Luo, X., & Wang, Y. (2025). Frequency-Domain Masking and Spatial Interaction for Generalizable Deepfake Detection. *Electronics*, *14*(7), 1302. https://doi.org/10.3390/electronics14071302

Mustak Un Nobi, M., Rifat, M., Mridha, M. F., Alfarhood, S., Safran, M., & Che, D. (2023). GLD-Det: Guava Leaf Disease Detection in Real-Time Using Lightweight Deep Learning Approach Based on MobileNet. *Agronomy*, *13*(9), 2240. https://doi.org/10.3390/agronomy13092240

Nagothu, D., Xu, R., Chen, Y., Blasch, E., & Aved, A. (2022). Deterring Deepfake Attacks with an Electrical Network Frequency Fingerprints Approach. *Future Internet*, *14*(5), 1–20. https://doi.org/10.3390/fi14050125

Sharma, D. K., Singh, B., Agarwal, S., Garg, L., Kim, C., & Jung, K.-H. (2023). A Survey of Detection and Mitigation for Fake Images on Social Media Platforms. *Applied Sciences*, *13*(19), 10980. https://doi.org/10.3390/app131910980

Sohail, S., Sajjad, S. M., Zafar, A., Iqbal, Z., Muhammad, Z., & Kazim, M. (2025). Deepfake Image Forensics for Privacy Protection and Authenticity Using Deep Learning. *Information*, *16*(4), 270. https://doi.org/10.3390/info16040270

Tipper, S., Atlam, H. F., & Lallie, H. S. (2024). An Investigation into the Utilisation of CNN with LSTM for Video Deepfake Detection. *Applied Sciences*, *14*(21), 9754. https://doi.org/10.3390/app14219754

Wolter, M., Blanke, F., Heese, R., & Garcke, J. (2022). Wavelet-Packets for Deepfake Image Analysis and Detection. *Machine Learning*, *111*(11), 4295–4327. https://doi.org/10.1007/s10994-022-06225-5

Xia, Z., Qiao, T., Xu, M., Wu, X., Han, L., & Chen, Y. (2022). Deepfake Video Detection Based on MesoNet with Preprocessing Module. *Symmetry*, *14*(5), 939. https://doi.org/10.3390/sym14050939

Yesilli, M. C., Chen, J., Khasawneh, F. A., & Guo, Y. (2022). Automated Surface Texture Analysis via Discrete Cosine Transform and Discrete Wavelet Transform. *Precision Engineering*, *77*, 141–152. https://doi.org/10.1016/j.precisioneng.2022.05.006

Yin, M., Chen, Z., & Zhang, C. (2023). A CNN-Transformer Network Combining CBAM for Change Detection in High-Resolution Remote Sensing Images. *Remote Sensing*, *15*(9), 1–26. https://doi.org/10.3390/rs15092406