EasyChair Preprint

№ 14205

Deep Fake Detection: Finding the Abnormal Patterns

Abdullah Farooq, Maymoona Naeem and Nouman Ahmed

July 28, 2024

Deep Fake Detection: Finding The Abnormal Patterns.

1st Abdullah Farooq
Department. of Cybersecurity
Air University
Islamabad, Pakistan
211133@students..au.edu.pk

2nd Maymoona Naeem
Department of Cybersecurity.
Air University
Islamabad, Pakistan
211045@students.au.edu.pk

3rd Nouman Ahmed
Department of Cybersecurity
Air University
Islamabad, Pakistan
211103@students.au.edu.pk

***Abstract*— Deepfakes are fake videos or images that look real. They are created using computer programs that can manipulate faces or voices to make people say or do things they never actually did.**

Advancement of deep learning techniques has led to the advancement of deep fakes as well, hyper-realistic digital forgeries that are now a threat to security, privacy, and the integrity of information. This paper explores the use of Support Vector Machine (SVM) classifiers with power spectrum and spatial frequency analysis for the detection of deep fakes.

By leveraging the frequency patterns that differentiate genuine images from manipulated ones, Approach aims to enhance the accuracy and reliability of deep fake detection in terms of image realism. Experimental results demonstrate that the proposed method effectively identifies deep fakes.

Keywords*— *Deep Fake, SVM, Logistic regression, FACTOR, datasets, GAN'S, Accuracy, Radial profile, Azimuthal Average, Fourier Analysis, Images, Videos, Audio-Visual, Spatial Frequency, Power Spectrum.

I. Introduction

Deep learning and generative adversarial networks (GANs) have revolutionized the development of synthetic media, which gave rise to convincing deep fakes. Manipulated media can mimic the appearance and voice of real individuals, making it difficult to differentiate between authentic and fabricated content. Deep fakes have repercussions and implications, from spreading misinformation and defaming individuals.

Over the last decades, the popularity of smartphones and growth of social networks have made digital images and videos very common digital objects. According to several reports, almost two billion pictures are uploaded every day on the internet. This tremendous use of digital images has been followed by a rise of techniques to alter image contents, using editing software like Photoshop for instance. The field of digital image forensics research is dedicated to the detection of image forgeries to regulate the circulation of such falsified contents. There have been several approaches to detect image forgeries. [1]

Detecting deep fakes is a critical challenge that needs sophisticated techniques capable of uncovering artifacts introduced during the manipulation process [2]. Traditional methods often fall short due to the ever-evolving nature of deep fake generation algorithms. In response, this paper proposes a novel deep fake detection framework that leverages Support Vector Machine (SVM) classifiers alongside power spectrum and spatial frequency analysis.

Power spectrum and spatial frequency analysis are powerful tools for examining the frequency domain characteristics of images. These techniques reveal inconsistencies, anomalies that are not apparent in the spatial domain. By extracting and analyzing frequency domain features, the proposed method seeks to capture the intrinsic properties that differentiate genuine images from manipulated ones.

Support Vector Machines (SVMs) are chosen for their effectiveness in classification tasks, particularly in scenarios involving high-dimensional feature spaces. The combination of SVM classifiers with frequency domain analysis aims to enhance the precision in deep fake detection. This paper details the methodology, experimental setup, and results of applying this approach to a dataset of real and fake images.

II. Motivation

Evaluation of several detection methods of Deepfakes, including lip-syncing approach and image quality metrics with SVM method are also introduced. Hence SVM would be used in our research. [3]

From a societal perspective, deep fakes disturb the trust in digital media, creating an environment where the authenticity of content is constantly questioned. This undermines public confidence in news sources, educational materials, and social media.

Economic implications include financial fraud, market manipulation, and damage to brand reputations.

In Political area**,** deep fakes represent a threat to national security. Manipulated videos and audio recordings can be used to discredit political figures, influence elections. These tactics can destabilize governments.

The issue of individual privacy is also at the forefront of the deep fake dilemma. Personal images and videos can be maliciously altered to create embarrassing content, causing severe emotional distress and reputational harm.

Given these widespread and severe implications, the motivation behind this research is clear: to contribute or at least take part as a concept of future research work. By integrating power spectrum and spatial frequency analysis with SVM classifiers, we aim to provide a solution that not only enhances the accuracy of detection but also contributes to the broader effort of safeguarding society, the economy, political systems, and individual privacy from the effects of digital forgeries.

III. Background and Literature Review

The Paper [3] focuses on the need of automated detection of deep fake leading towards development of the first publicly available Deep fake video dataset, generated from "VidTIMIT" videos using GAN-based software. Two versions of Deep fakes have been produced. Low quality

(LQ) High quality (HQ) Vulnerability of state-of-the-art face recognition systems (VGG and Face net) to Deep fakes is assessed. Some detection methods are evaluated such as: lip-sync inconsistency detection. Image quality metrics with SVM classifiers. Utilizes 129 IQM features. Both VGG and Face net-based systems exhibit high vulnerability to Deep fake videos. Deep fakes, indicating reasonable accuracy in identifying tampered videos. Lip-sync-based approaches fail to detect Deep fakes.

The study in [4] proposes both spatial and temporal features for detection. Celeb-DF dataset is used Preprocessing crops faces isolating relevant facial features. DFT extracts discriminative features cropped at initial stage for classification. Deep fake videos often contain inconsistent temporal artifacts due to frame-by-frame manipulation. RCN addresses this challenge by combining CNN (Convolutional Neural Network) for feature extraction and LSTM (Long Short-Term Memory) for temporal sequence analysis. 3D CNNs capture spatial and temporal information effectively. This architecture introduces shortcut connections to facilitate information flow, enabling effective representation learning from video data. I3D offers a state-of-the-art approach to spatiotemporal learning by inflating 2D ConvNet architectures to 3D. It uses pre-trained models on successful 2D architectures. The proposed algorithms achieve remarkable ROC-AUC scores and accuracy rates, outperforming baseline methods on the Celeb-DF dataset.

Focus of this paper [1] is on Two network architectures: Meso-4 MesoInception-4. Training data is processed in batches, with slight random transformations applied to improve generalization. The networks are trained on datasets specifically curated for Deepfake and Face2Face techniques, containing forged and real face images extracted from videos. The evaluation considers each frame independently, and the results are analyzed for different compression levels. Meso-4 and MesoInception-4 networks demonstrate high detection rates, with scores of over 90% for both Deep fake and Face2Face techniques. Image aggregation further improves detection accuracy, with scores exceeding 98% for Deep fake detection.

This paper [5] also leverages deep learning techniques to combat manipulated facial images. The researchers constructed a dataset named FaceForensics++ by applying four state-of-the-art face manipulation methods to 1,000 pristine videos downloaded from the internet. As a result, the dataset contains over 1.8 million images from 4,000 manipulated videos. The study indicates a correlation between video quality and detection accuracy, with lower quality videos having decreased performance. The paper evaluates automated forgery detection methods on the test set. The algorithms outperform human observers by a margin, showcasing the effectiveness of deep learning techniques in detecting facial manipulations. Results indicate varying detection accuracy across different manipulation methods and video quality levels. For steganalysis features, handcrafted features from high-pass images are mentioned. These features are then fed into a Support Vector Machine (SVM) classifier. The XceptionNet architecture is highlighted as outperforming other variants in detecting fakes. The benchmark dataset is used to evaluate the performance of forgery detection models in a standardized setting.

Reconstruction Challenge is discussed in this [6]. The challenge was divided into two main tasks: 1. Deepfake Detection Task. 2. Deepfake Reconstruction Task. CelebA and FFHQ for real face images and various GAN-generated deepfake images were used. No participant was able to propose a solution for the reconstruction task within the given deadline, highlighting the complexity of the challenge. Participants' solutions were evaluated based on their ability to accurately classify real and deep fake images. Solutions: Multiple teams utilized deep learning approaches, particularly convolutional neural networks (CNNs), to achieve the best classification results. Biometria Team, has demonstrated a high expressive power that allows for generalization across different application contexts and effective recognition of artifacts throughout the entire image. However, its performance was affected by further manipulations of test data, such as compression and resizing. For instance, resizing the OpenForensics test images to 55% resulted in a notable decrease in accuracy to 89.30%. During the challenge test, this approach misclassified live samples that contained manipulations outside the facial region. While the method examines the entire spectrum of the image and detects manipulations, even in the background or the subject's hair, it was primarily designed to detect images manipulated to harass or persecute a victim. Thus, although these samples were considered incorrectly classified for the competition, in a real-world application context, this functionality could be valuable for detecting manipulations in multimedia files representing individuals.

This review [7] aimed to summarize existing research, techniques, and datasets for Deepfake detection, categorize detection techniques, analyze experimental evidence, and provide guidelines for future research and practices in the field. After completing the review of all the studies, the outcomes were reported in a suitable form to the distribution channel and target audience. Machine Learning-based Methods Achieve up to 98% accuracy in detecting Deepfakes. Utilize CNNs, RNNs, and ensemble learning techniques, achieving over 99% accuracy. Use techniques like PRNU, Expectation-Maximization, etc. Employ blockchain technology for decentralized verification. FaceForensics, Celeb-DF, and DFDC are popular datasets. Special artifacts, face landmarks, spatio-temporal consistency widely used. CNNs (e.g., XceptionNet, ResNet), RNNs (e.g., LSTM), SVM, kMeans clustering prevalent. Accuracy, ROC curve, AUC, recall, precision are commonly used. Deep learning-based methods outperform others, achieving 89.73% accuracy and 0.917 AUC on average. Deep learning models show better efficiency than non-deep learning models in Deep fake detection. The review acknowledges several limitations and challenges, such as construct validity (potential missing studies), internal validity (data extraction and analysis errors), and external validity (inconsistencies in reported results). Addressing these challenges is crucial for advancing research in deepfake detection.

The paper [8] proposes using deep neural networks, specifically a modified network structure called the Common Fake Feature Network (CFFN), for effective fake image detection. A classifier network is trained to recognize fake face images based on learned CFFs. The proposed approach involves training CFFN using contrastive loss first, followed

by training the classifier using crossentropy loss. For fake face image detection, the researchers extracted images from the CelebA dataset, they also utilized images generated by five state-of-the-art GANs: DCGAN, WGAP, WGAN-GP, LSGAN, and PGGAN. Each GAN generated 40,000 fake images of size 64x64 pixels, resulting in a total of 200,000 fake images. Real images were randomly selected from CelebA, resulting in a balanced dataset of 200,000 real and 200,000 fake images. For fake general image detection, the researchers used three state-of-the-art GANs: BigGAN, SA-GAN, and SN-GAN. They generated 100,000 fake images of size 128x128 pixels with each GAN, resulting in a total of 300,000 fake images. Real images were randomly selected from the ILSVRC12 dataset, resulting in a dataset of 300,000 real and 300,000 fake images. For fake face image detection, the proposed method outperformed existing techniques across all tested GANs. The researchers trained a Convolutional Feature Fusion Network (CFFN) for fake image detection, leveraging a pairwise learning strategy. Visualization techniques were employed to interpret fake image features by mapping feature responses to the image domain. However, the researchers acknowledged limitations related to the need for retraining with new generators and the challenge of collecting training samples for GANs with undisclosed technical details.

This paper proposes [9] leveraging fact checking, adapted from fake news detection, to detect zero-day deep fake attacks. It introduces FACTOR. [10] Assumption of paper is that current generative models cannot accurately encode false facts into fake media. FACTOR computes the truth score between media using off-the-shelf features, effectively distinguishing between real and fake media. FACTOR formulates facts as statements comparing the content of two media. It relies on off-the-shelf encoders to measure similarity between media, with low truth scores indicating false facts. The approach involves comparing a test image to a reference set of authentic images of the claimed identity. The similarity between the test image and the images in the reference set is measured using cosine similarity over facial features. Low similarity scores indicate a falsehood, signaling a fake image. The performance of the proposed method is evaluated using the truth score, which measures the similarity between media. FACTOR utilizes a reference set of real images from claimed identities and evaluates them against fake images generated by deepfake methods. The method calculates truth scores based on image similarities. FACTOR consistently demonstrates superior performance to supervised methods across all categories of fake videos. It outperforms all supervised baselines in terms of average AP and ROC-AUC. Limitations as given by the paper: i)- Image realism is not catered. ii)- Does not deal with cases where the original and claimed identities are identical, but other attributes are manipulated, such as changes in facial expressions, age or other non-identity-related features. iii)- Facts must be falsifiable iv)- Unconditional deep fakes do not include facts v)- Supervised approaches work well on previously seen attacks vi)- No pretrained encoders for non-standard facts.

The paper [2] outlines the process of training a Convolutional Neural Network (CNN) model using both positive (real) and negative (synthetic) examples. The CNN models used include VGG16, ResNet50, ResNet101, and ResNet152. UADFV, DeepfakeTIMIT. The performance of each CNN model (VGG16, ResNet50, ResNet101, ResNet152) is evaluated on both UADFV and DeepfakeTIMIT datasets. AUC scores are reported for image-based and video-based evaluations. Results demonstrate the effectiveness of the proposed method, with ResNet50 achieving the highest performance across datasets and evaluation methods. The process begins with collecting positive (real) face images from the Internet. RoIs encompasses both the face area and its surrounding regions, as the aim is to expose the artifacts between the fake face area and its surroundings. RoIs are selected based on facial landmarks and resized to a standard size (224 × 224) before feeding them into the CNN models for training. The CNN models are trained using a dynamic approach, where negative examples are generated dynamically during the training process to enhance diversity. Training hyperparameters such as batch size, learning rate, and optimization method are specified. Models are fine-tuned using hard mining strategy to focus on challenging examples. Performance evaluation is conducted using Area Under Curve (AUC) metric for both image-based and video-based evaluations. AUC scores are computed for each CNN model on different datasets, including UADFV and DeepfakeTIMIT, to assess the effectiveness of the proposed method.

The DeepFake Detection Challenge (DFDC) Dataset is introduced as a response to the need for dataset for training and evaluating Deepfake detection models. [11] Previous datasets lacked the scale, diversity, and ethical considerations necessary for model training and evaluation. The DFDC Dataset addresses these limitations by: Over 48,000 videos featuring 3,426 paid actors were recorded specifically for the dataset, ensuring consent and ethical considerations. Existing Deepfake datasets are categorized into three generations. The DFDC Dataset contains videos recorded in natural settings, with participants consenting to their inclusion in a machine learning dataset. Training data augmentation techniques were applied to enhance model generalization. The top-performing solutions in the competition use a combination of face detection algorithms (like MTCNN), feature extraction architectures (like EfficientNet and Xception), and ensemble methods to achieve high detection accuracy.

IV. Methodology

In this Paper Support Vector Machines SVMs [3] A type of ensembling technique [8] integrated with basic image realism approach to detect deepfake through temporal coherence and geometrical proportions by using three datasets is used. Celeb-df, face forensics ++, and ffhq. Radial profile, azimuthal average, Fourier transform analysis are mathematical concepts used to train SVM and LR algorithms on training, testing, real, and fake data. Model learns the difference of dimensions, frequencies, and most importantly spatial frequencies with respect to power spectrum between real and fake images after it is trained on the above written three datasets. Then model is given unseen images to detect deep fake based upon spatial frequency and power spectrum.

A. Image Realism

Deepfake detection relies on understanding and identifying features that distinguish real images and videos from artificially generated ones. Real images exhibit natural textures and fine details that can be challenging for generative

models to replicate perfectly. Skin texture, hair strands, and other fine details often reveal deepfakes. Realistic images have consistent lighting and shadows that align with the scene's light sources. Deepfakes may struggle to accurately mimic these aspects, leading to unnatural lighting effects or shadow inconsistencies. Human facial expressions and movements are complex and nuanced. Deepfakes sometimes produce unnatural or exaggerated expressions and jerky or unnatural movements. Detection systems can analyze these elements to identify irregularities indicative of deepfakes. In many deepfakes, eye reflections might not match the lighting environment, and blinking patterns can appear unnatural. Eyes might also exhibit a dead look, lacking the small involuntary movements present in real videos. Human faces and bodies follow certain geometric proportions. Deepfakes can sometimes produce slight distortions in these proportions. For instance, the alignment of facial features might be slightly off, or the proportions of the body might look unnatural. Deepfakes might exhibit slight mismatches between lip movements and spoken words, leading to audio-visual synchronization issues. This discrepancy can be detected through analysis. [9] The context in which the person appears can also provide clues.

B. FACTOR

In Paper [8] FACTOR is employed that caters Contextual Information of pre encoded parameters such as true information or misinformation in alignment with audio-visual tasks in which claimed identities are true or not is detected. Out of limitations of paper: If an attacker simply copies the claimed face onto the observed image, it will correspond to the claimed face identity, although this would result in an unrealistic appearance. To mitigate this, they recommend ensembling their method with a simple image realism-based approach which will easily catch such crude attacks. ii) Their method does not deal with cases where the original and claimed identities are identical, but other attributes are manipulated, such as changes in facial expressions, age, or other non-identity-related features. These tasks are left for future work.

C. Spatial Frequency and Power Spectrum

Deepfake detection using spatial frequency and power spectrum analysis is an absolute on point way to find the patterns and inconsistencies in the frequency domain that often emerge due to the nature of deepfake generation processes.

Spatial frequency refers to the level of detail present in an image. High spatial frequencies mean fine details and edges, whereas low spatial frequencies relate to smooth and broad regions. Natural images have a specific distribution of spatial frequencies, with a balance between low, medium, and high frequencies that correspond to real-world textures and details. The power spectrum of an image is obtained by performing a Fourier Transform, which distributes the image into its frequency components. The power spectrum displays how the power (variance) of an image is distributed across different spatial frequencies. In a typical power spectrum of a natural image, the power decreases as the spatial frequency increases. Deepfakes often have unnatural patterns within the frequency domain because the generative models (such as GANs) used to create them may not perfectly capture the natural distribution of spatial frequencies. GAN-generated images might have too much or too little power in certain frequency bands or exhibit artificial regularities or noise that are not present in real images. Deep fake algorithms sometimes leave inconsistencies that are not easily noticeable in the spatial domain (the image itself) but become apparent in the frequency domain. Grid-like patterns, periodic noise, or unnatural smoothness in high-frequency regions. Comparative Analysis can be done by comparing the power spectrum of a suspected deepfake to that of a real image to detect anomalies. Examining the slope can reveal deep fakes. Log power can also be used, but in our research, we aim to examine the slope only**.** Real images typically have a certain statistical regularity in their power spectrum, while deepfakes deviate from this. One way to spot deepfakes is by looking at the image in a different way — not just as a picture, but by examining the underlying patterns in the image's details. Imagine a picture made up of lots of tiny details and smooth areas. High spatial frequency means lots of tiny details (like edges of objects or textures). Low spatial frequency means smooth areas (like the sky or a wall). To see these details and smooth areas differently, we use power spectrum. The power spectrum is like a special graph that shows how much detail is in the picture at different levels. Think of it as separating a song into its different notes to see which ones are loudest.

Real pictures have a certain balance of details and smooth areas. This balance creates a natural pattern in the power spectrum. Deepfakes often mess up this balance. They might have too many details in some places or not enough in others. We show the computer lots of real and fake pictures, so it learns what to look for. Once it's trained, the computer can check new pictures and decide if they're real or fake based on their power spectrum. [12]

D. Feature Extraction and Classification

Trained models on a dataset of celeb-a, face forensics++, and ffhq of real and fake images, using features derived from their power spectra are used.

Validated the model's performance. By looking at the details and smooth areas in a picture through the lens of spatial frequency and power spectrum, it can tell if a picture is a deepfake. In this Model. SVM classifiers and LR to detect deepfakes on datasets are trained. Common Points to look at: Check if both lines on x,y axis generally decrease from left to right. A natural decrease (real images) usually indicates more power at low frequencies and less at high frequencies. Identify specific regions where the blue and orange lines diverge significantly. Look at the overall shape and smoothness of the lines. Real images usually have a smoother transition, while fake images might show abrupt changes or irregular patterns. Check for peaks (high points) and valleys (low points) in both lines. Compare the positions and magnitudes of these features. Real images should have a predictable pattern, while fake images might show unexpected peaks. Compare the slopes of the lines. The slope of the power spectrum in real images typically follows a specific decay rate.

E. Using Mathematical Concepts

This paper used the concepts of radial profile, and azimuthal where radial profile calculates the average intensity of pixel

values at each radial distance from the center of the image, which provides a measure of image's frequency content. Azimuthal average means averaging the pixel values at each radial distance from the center of the image along concentric circles, essentially capturing the average intensity in each radial direction. Fourier analysis is a mathematical technique used to decompose complex signals or functions into simpler components, typically sine and cosine waves, through the Fourier transform. The azimuthal average function calculates the azimuthally averaged radial profile of a 2D image. Fake image data from the dataset_celebA directory is processed. The Fourier transform and magnitude spectrum are computed for each image. Radial profiles are calculated using the azimuthal Average function. The processed data is saved in a pickle file. Real image data is then processed. Like fake image data, Fourier transforms, magnitude spectra, and radial profiles are computed. The processed data is saved in a separate pickle file. .pkl file contains balanced real and fake images. Initialization of arrays is done to store the mean and standard deviation of the PSD for both real and fake data. Each sample in the dataset is iterated and separates them based on their labels. Samples with a label of 0 are considered fake, while samples with a label of 1 are considered real. The PSD values of each sample are stored in separate arrays psd1 and psd2. For each feature, it computes the mean and standard deviation of the PSD values for both fake and real data. It creates a plot showing the meaning of the PSD for both real and fake data. The x-axis represents spatial frequency, and the y-axis represents the power spectrum. This methodology was applied to all three datasets.

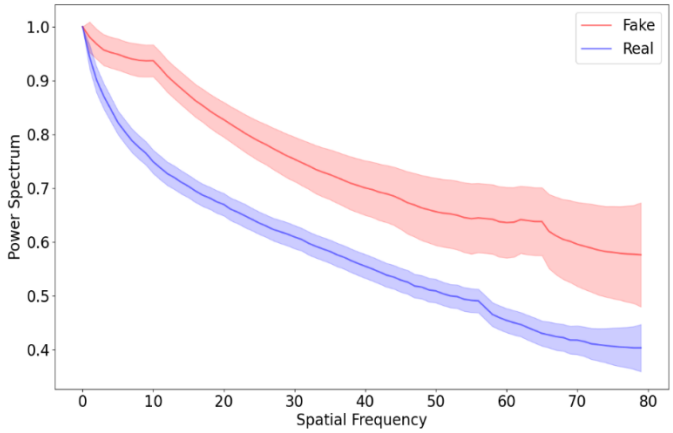

Fig 1. Celeb Dataset

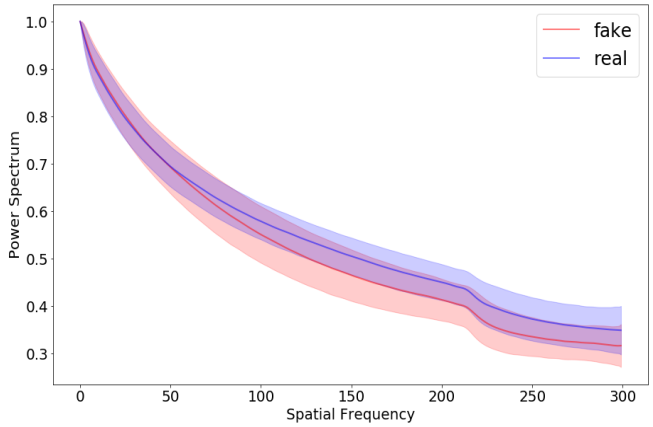

Fig 2. FaceForensics

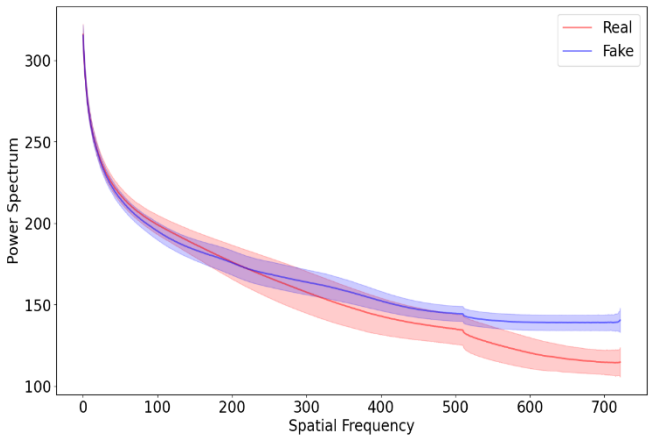

Fig 3. FFHQ

The above figures show the spatial frequency with respect to power spectrum. More the spatial frequency lesser the power spectrum. In real images the line is normal if there is not too much abnormality, and sudden surge or spike at any frequency level. To detect deepfakes. Generally, abnormality in this line is noticed, and as the above figures show the data is trained on different datasets of both the types of images. The fake images line spectrum is clearly different from the real ones, and this baseline foundation would be used to detect unseen data.

F. Unseen Data

An image is given and converted to grayscale for further processing. It computes the Fast Fourier Transform (FFT) of the input image. The magnitude spectrum is calculated by taking the absolute value of the FFT result. The code computes the azimuthally averaged 1D power spectrum of the magnitude spectrum using the azimuthal average function from the radial Profile module. This function calculates the average radial profile of the image's power spectrum, capturing information about the distribution of spatial frequencies. It visualizes the input image and its azimuthally averaged 1D power spectrum. The azimuthally averaged power spectrum gives insight into the distribution of spatial frequencies in the image, which is useful to detect structural characteristics of an image. Hence Indicating Possible deep fake and real images

V. DATASETS

A. Celeb-df

202,599 number of face images of various celebrities

10,177 unique identities, but names of identities are not given. 40 binary attribute annotations per image. 5 landmark locations

B. Face Forensics

Face Forensics++ is a forensics dataset consisting of 1000 original video sequences that have been manipulated with four automated face manipulation methods: Deepfakes, Face2Face, Face Swap and Neural Textures.

C. *Flickr-Faces-HQ (ffhq)*

Flickr-Faces-HQ (FFHQ) is a high-quality image dataset of human faces, originally created as a benchmark for generative adversarial networks (GAN.

VI. Results

A. *Celeb-A*

Model	Accuracy
SVM (Linear)	0.9995
SVM (RBF)	1.0000
SVM (Poly)	1.0000
Logistic Regression	0.9968

Figure 4. Accuracy

B. *Face Forensics*

Model	Accuracy
SVM	0.857
Logistic Regression	0.769

Figure 5. Accuracy 2.0

For FFHQ the accuracy is 1.0 for both SVM and LR, because training samples were just 1000, 500 fake and 500 real. For celeb A there were 2000, 1000 real and 1000 fake. For Face Forensics there were 3200 samples, 1600 fake and 1600 real.

VII. Evaluation

Evaluation is done based on MSE, and R squared error. The average is being shown in the figure below:

Iteration	Model	Mean Squared Error	R-squared
Iteration 1	SVM	0.0	
	Logistic Regression	0.0025	0.98999974999
Iteration 2	SVM	0.0	
	Logistic Regression	0.0	
Iteration 3	SVM	0.0	
	Logistic Regression	0.0	
Iteration 4	SVM	0.0	
	Logistic Regression	0.0025	0.98998773497
Iteration 5	SVM	0.0025	0.98999774949
	Logistic Regression	0.0025	0.98999774949
Iteration 6	SVM	0.0	
	Logistic Regression	0.0	
Iteration 7	SVM	0.0	
	Logistic Regression	0.0025	0.98999774949
Iteration 8	SVM	0.0	
	Logistic Regression	0.005	0.97979797979
Iteration 9	SVM	0.0025	0.98997493734
	Logistic Regression	0.0025	0.98997493734
Iteration 10	SVM	0.005	0.97988686360
	Logistic Regression	0.005	0.97988686360

Figure 6. Evaluation

MSE describes the variance in data and R-squared if it is high. It means model is almost best fit to the data.

VIII. Conclusion

The papers discussed present an overview of methodologies and approaches for detecting deepfake videos, addressing various aspects of image realism and leveraging advanced techniques such as spatial frequency analysis and feature extraction. Each paper contributes to the evolving landscape of deepfake detection. FACTOR, a method based on fact-checking principles, effectively distinguishes between real and fake media by computing truth scores. It demonstrates superior performance, particularly in zero-day attack scenarios, across different types of deepfake media, including face swapping, audio-visual deepfakes, and those generated from text prompts.

In addition to these papers, our proposed methodology in FACTOR combines SVM classifiers and LR algorithms with spatial frequency and power spectrum analysis for deepfake detection. By leveraging concepts from Fourier analysis and spatial frequency analysis, providing an alternative method especially the media that is flagged green by FACTOR could be checked by the technique proposed with more improvements.

IX. References

[1] Afchar, "MesoNet: a Compact Facial Video Forgery Detection Network," in *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, IEEE, 2018.

[2] S. Marcel, DeepFakes: a New Threat to Face Recognition? Assessment and Detection, 2018.

[3] George, "Deepfake Detection using Spatiotemporal Convolutional Networks," 2020.

[4] Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," 2019.

[5] L. Guarnera, O. Giudice, F. Guarnera, A. Ortis, G. Puglisi, A. Paratore, L. Bui, M. Fontani, D. Coccomini, R. Caldelli and e. al, "The Face Deepfake Detection Challenge.," *Journal of Imaging*, 2022.

[6] M. a. N. Rana, "Deepfake Detection: A Systematic Literature review," *IEEE Access*, vol. 10, 2022.

[7] Zhuang and C.-Y. Lee, "Deep Fake Image Detection Based on Pairwise Learning," *Applied Sciences.*, p. 10, 2020.

[8] Hoshen, "Detecting Deepfakes Without Seeing Any," 2023.

[9] Hoshen, "Github," Novembor 2023. [Online]. Available: https://github.com/talreiss/FACTOR/tree/master/audio-visual.

[10] Exposing Deepfake videos by Detecting Face warping artifacts, 2018.

[11] Ferrer, The DeepFake Detection Challenge (DFDC) Dataset, 2020.

[12] A. van der Schaaf, "Modelling the Power Spectra of Natural Images: Statistics and Information," *Vision Research,*, vol. 36, no. 17, pp. 2759-2770, 1996.