IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 25, Issue 4, Ser. 2 (July. – August. 2023), PP 60-68
www.iosrjournals.org

The Emergence of Deepfake Media and All-Inclusive Study of technological advances and Media Integrity Threat Challenge

Nikhat Fatima[1], Dr. Sameena Banu[2]
[1,2]Assistant professor, CSE Dep of Faculty of Engineering and Technology, Khaja Bandanawaz University.

Abstract: *Deepfake technology, first developed in 2017, has evolved significantly since then, generating convincing false media using artificial intelligence (AI), particularly deep learning and machine learning approaches. Initially popular in entertainment and face-swapping apps like Facelab and FaceApp, deepfake media has evolved from low-quality still photos to high-definition videos, making it harder to distinguish real from altered information. This technology has helped spread false information, including fake news, particularly involving public figures targeted. The development of deepfake technology raises ethical and social questions regarding misinformation, privacy, and authenticity. As AI continues to develop, it has produced clearer, more convincing images and videos, making it even more difficult to distinguish fake from real media. The term "deepfake" is derived from the combination of "deep" (from deep neural networks) and "fake," meaning artificial intelligence-generated false media including audio and video.*

Keywords: *Deepfake, Machine Learning, Deep learning and AI.*

I. Introduction

The technology of deepfake media has advanced significantly since its debut in 2017. According to a major UK newspaper, the word was coined by a social media user who substituted famous people's faces in a number of pornographic movies [1]. The novel aspect of this technology, which enables users to create amusing material, was the original driving force for the birth of several faceswapping programs, such as Facelab and FaceApp . The whole community is becoming more aware of this technology's broad possibilities as well as any potential downsides. The rapid development of artificial intelligence (AI), especially in the fields of machine learning (ML) and deep learning (DL), has accelerated the technology's evolution and aided in the dissemination of false information across our society. Deepfake media's early versions were simple and frequently connected to static pictures of poor quality. Higher-quality photos and videos are now given priority due to recent developments in deep learning model training and the growth of open-source content creation techniques. As we get closer to a critical level, it gets harder to distinguish between fake and real media. The seriousness and dangers of prominent public personalities whose dishonesty has enabled the spread of false information through fake news have been highlighted in recent news headlines. A "deepfake" is a media synthesis method that makes use of artificial intelligence [2]. Artificial methods for creating fake information, including fake photos and movies, have become more prevalent in recent years. "Deepfake" is a phrase that combines the terms "Deep" and "fake" to refer to artificial material produced using Deep Neural Networks (DNNs). As seen in Figure 1, the deepfake technique produces realistic-sounding and accurate audio and video, which makes it difficult for people to recognize authenticity, when using deepfake content,

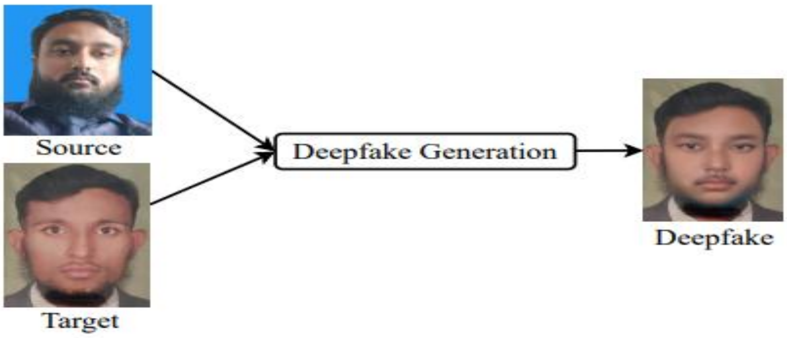

Figure 1 Deep fake technique

The search for truth has become even more important in the digital sphere. Given the prevalence of dangerous deepfakes and the ease with which they may be produced, controlling and reducing these technologies is extremely difficult. Numerous techniques have been developed to identify deepfakes. Deep learning is used in many systems, and it is necessary to distinguish between useful and potentially detrimental uses of deep learning methods. In order to mitigate the hazards connected with deepfakes, the US Defense Advanced Research Projects Agency (DARPA) launched the MediFor research project to improve the techniques for detecting counterfeit digital visual information [4]. Microsoft Corp., Facebook Inc., and the Partnership on AI alliance recently launched a deepfake detection project to encourage further research and development in the detection and mitigation of deepfakes used by dishonest people. Computer perception is now more focused on feature engineering because too deep learning technologies, which have lessened the need for human input techniques like hand and face instructions. For deep neural networks (DNNs) to produce better results, large amounts of high-quality data are required. For this, a data training system is required [5]. The large datasets acquired from Facebook and Google make it easier to create a model that is both well-trained and of high quality. When used on well-structured datasets, deep learning models perform better.

The term "deepfake" describes the alteration of face characteristics or emotional emotions, as [6]. The face picture of one person is replaced with another in deepfake movies, which may violate public domain rights and endanger the person who is impacted. Recent developments in this technology have led to imitations that closely mimic real articles, making it more difficult to distinguish between real and fake images and videos. By combining, swapping out, or imposing pictures or movies for misleading ends, artificial intelligence (AI) may produce deepfake images or videos [7].

Facial emotions, facial alignment, and face classification are the main obstacles in deepfake detection. The crucial first step in the deepfake detection procedure is facial feature analysis. Deepfake defections require a high degree of formal communication to be implemented [8]. In many applications, including automated immigration systems, intelligent inspection systems, and identity verification systems, face recognition plays a crucial self-regulating role. Within the topic of deepfake detection, face recognition and face verification are separate subfields. Face recognition technology finds the image that most closely matches the samples that are presented [9]. Facial feature recognition has become more popular in both practical and scholarly applications. A notable development in deep learning technology has sparked the modification of face characteristics. On the other hand, facial wrapping depends on the façade, the face's emotion, movement, and general look. It is really difficult to find a realistic face in these situations. Large-scale data collection and classification is a difficult procedure that takes a lot of time to complete successfully. Even though the publicly accessible datasets are expensive and have a high failure tolerance, they cannot accommodate any changes. The lack of sufficient facial training datasets is addressed by face data augmentation.

DeepFakes are incredibly lifelike simulated pictures and movies produced by combining computer vision algorithms, such as autoencoders and Generative Adversarial Networks (GANs), with deep learning techniques. Using deep learning methods with artificial media makes it easier to edit images or movies, enabling anybody to make changes without needing to know anything about machine learning. To create a new version of the data that retains comparable qualities for computer systems and human interpretation, the original data is altered. Public confidence in digital media has decreased as a result of the rise of DeepFakes, which have raised doubts about the veracity of visual material. In the absence of deep learning methods, research aimed at identifying or detecting unlawfully changed material is categorized as conventional research. Generative deep models may be used to create DeepFakes, which poses serious problems for conventional detection techniques. To close this gap and preserve public trust in digital multimedia, essential research in DeepFake detection is required. A method called FaceSwap1 is intended to produce DeepFake movies with actual people acting in modified situations. It's frequently difficult for viewers to tell the difference between real and fake content. In addition to negatively impacting the lives of those targeted, the use of these technologies may increase political instability, enable acts of terrorism, violence, or civil unrest, and aid in the spread of hate speech and false information [10]. The synthesis and improvement of human facial features is one of the uses of AI-driven DeepFakes in computer vision and graphics.

Motivation

The rapid advancement of deepfake technologies has presented a dual-edged sword in the realm of digital media. While these technologies offer groundbreaking possibilities in content creation and entertainment, their misuse poses serious threats to information integrity, privacy, and societal trust. The proliferation of convincing fake media has made it increasingly difficult for individuals and systems to differentiate between authentic and manipulated content, emphasizing the urgent need for robust detection mechanisms. Our work is driven by the mission to address this critical challenge. By leveraging innovative techniques and methodologies, we aim to bridge the gap between existing detection limitations and the escalating sophistication of deepfake generation. This research not only contributes to advancing the field of digital forensics but also reinforces the broader goal of safeguarding digital ecosystems and fostering trust in an era of rapid technological evolution.

Research contribution made

- **Development of a Discrepancy-Aware Forgery Detection Network (DAFDN)**

This research introduces a novel Discrepancy-Aware Forgery Detection Network (DAFDN), which integrates Feature Representation Extractor (FRE) and Feature Refinement Module (FRM) to generate unbiased feature representations. By incorporating Attention-Guided Feature Rectification (AGFR) and Discrepancy-Aware Interaction Module (DAIM), the model effectively captures both regional and channel-level inconsistencies to enhance deepfake detection accuracy.

- **Enhanced Deepfake Localization with Region-Aware Forgery Detection (RAFD)**

The proposed framework improves forgery localization by leveraging Region-Aware Forgery Detection (RAFD) and Channel Discrepancy Analysis (CDA). These components allow the model to detect subtle facial manipulations, warping artifacts, and inconsistencies in deepfake videos, significantly improving detection performance across multiple datasets.

- **Comprehensive Performance Evaluation on Benchmark Datasets**

The research evaluates DAFDN on challenging deepfake datasets, including Celeb-DF, WildDeepfake, and DFDC, demonstrating superior detection accuracy compared to state-of-the-art methods. The proposed model outperforms existing approaches in detecting highly realistic, compressed, and manipulated videos, showcasing its robustness and effectiveness in real-world deepfake detection scenarios.

II. Related work

Artifacts found in both the region based and frequency domains have revealed important information about the pixel formation in the spatial domain that constitutes the overall image over time, or the frequency representation including low- or high-frequency components in the frequency domain, which relates to the rate of change in pixel information. Significant statistical data that might reveal the location of the tampering could be produced by a break in the surrounding pixel formation between the old and new content. Furthermore, the synthesis method used to create a deepfake naturally produces face blending inconsistencies, which result in detectable artifacts remaining in the picture data [11]. Similar to how a fingerprint is extracted from a photograph, the camera model NoisePrint efficiently extracts and compares noise signatures from images using the Photo-Response Non-Uniformity approach [12]. Furthermore, the use of frequency and spatial domains as built-in machine learning characteristics has made it possible to create innovative detection methods that can extract information from complicated data with little assistance from humans. There are difficulties in choosing the right features for training, especially when the underlying pipeline used to create deepfakes is dynamic. This technique is usually used in conjunction with a fully connected layer and a binary classifier. Applying this to unknown data might lead to insufficient generalizations. An important advancement in deepfake detection has been made possible by the use of artificial neurons that mimic human brain activity to create a model that can learn intricate multi-dimensional patterns from complicated datasets [13]. This makes it possible to obtain a more thorough feature representation, which is not possible with conventional machine learning techniques. [14].

Every frame of a real video must have a consistent fingerprint or artifact, according to a method put out by [15]. Deepfakes always produce inconsistent artifacts because of the changed face areas. An Artifact Discrepant Data Generator (ADDG) and a Deepfake Artifact Disagreement Detector (DADD) are used in a self-supervised deepfake identification approach to find inconsistencies in the produced data. By using well-established processing techniques to modify the face region of real video frames, the ADDG creates synthetic examples. Using a multi-task learning approach, the DADD links each sub-task to a unique category of created data and combines the sub-tasks to produce the desired outcomes. These methods are advantageous since the visual artifacts are sufficiently specified.

[16] presented a strategy that combines deep learning and machine learning techniques to efficiently categorize deepfake pictures. Convolutional Neural Networks (CNNs) are used for feature extraction, whereas the ELA approach detects image alternations. K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) are used to classify the pictures. The accuracy of the model varies as noise is introduced into the data. The significant processing power needed to apply deep learning techniques might be problematic in real-time situations. [17] suggested a model that used Alex Net and Shuffle Net in combination with the ELA to distinguish between real and fake photos. Even if the dataset is 2041 in size, the small number of pictures may limit the model's ability to generalize to other datasets or real-world scenarios. The study focused on recognizing deepfakes in photos because the model showed difficulties in detecting deepfakes in videos. Using ELA methods might be difficult when working with different picture formats or compression. In these

circumstances, the strategy put forward by [17] is inappropriate. [18] applied the ELA to the pictures, this approach emphasized the differences in compression levels and pinpointed the regions that needed improvement. ELA was used as a forensic approach to identify the variations in the changed photographs. Dropout layers were used to reduce overfitting, however the model's performance on unknown data was still not ideal.

Due to constraints on processing power and production time, deepfake algorithms are limited in their capacity to produce face pictures up to a certain size. In order to match the face arrangement of the source, affida warping is necessary. Face warping produces a variety of aberrations due to the disparity in resolution between the warped facial area and the surrounding features. According to [15], realistic photos are trained using the variational auto-encoder, which classifies them as synthetic along with other pictures. The blending limitations for face swapping method detection were defined by [6]. Another approach uses neural networks, including regular neural networks, customized deep networks, and other variants, to detect the fake traits. The effectiveness of the neural approaches was impressive. Among the uses of deepfaked products are extortion and interest termination. The term "deepfake" describes a real-time digital impersonation of a UK CEO that is used to transmit sensitive data or carry out an urgent financial transaction. The integrity of national policies and procedures is seriously threatened by Deepfake technology, which must be acknowledged in order to solve the problem of nations with no public disagreement [19].

In their fashion presentations, corporations may use a variety of models with a range of body shapes, heights, and skin tones. Furthermore, they could work with attractive models who don't always meet the criteria for glamor models. Deepfakes also enable users to produce highly customized material that may be used as models. Customers may evaluate items before making selections by using the technology to provide virtual try-ons. Apart from that, it creates customized fashion ads that change according to the target demographic, weather, and time of day [20] presented the Deep Convolutional Generative Adversarial Network (DCGAN), a more stable computational architecture, to improve training stability. Instead of using pooling and batch normalizing approaches, the researchers used deep convolutional networks, showing enhanced picture synthesis performance by using an arithmetic vector. A year later, in an effort to improve the precision and dependability of learning results, researchers at National Vision Instrument and Advanced Graphics (NVIDIA) introduced a novel network design called the Progressively Growing Generative Adversarial Network (ProGAN). In the end, inferior quality data improves an algorithm's properties during training. StyleGAN is a network variation that is based on ProGAN. According to a study of the literature, the researchers modified the generator approach by using Adaptive Instance Normalization (AdaIN) to execute creator training at each CNN layer. Typically, the developer uses the given vector to create a consistent posture or style.

Table 1

Reference	Method	Advantages	Disadvantages	Research Gap
Nguyen et al. [2019]	CNN-based deepfake detection	Good generalization on standard datasets	Fails against adversarially modified deepfakes	Vulnerability to adversarial attacks
Afchar et al. [2018]	MesoNet for face forgery detection	Lightweight and computationally efficient	Struggles with highly compressed videos	Poor performance on low-quality videos
Tolosana et al. [2020]	Visual artifacts and physiological analysis	Detects subtle physiological inconsistencies	Requires high-resolution input for effective detection	Limited to high-quality datasets
Li et al. [2020]	Face warping artifact detection	Effective against low-quality deepfakes	Cannot detect unseen deepfake methods	Ineffective against hybrid deepfakes
Agarwal et al. [2021]	Biometric-based forensic detection	Leverages facial biometrics for detection	Sensitive to pose and lighting variations	Pose and illumination challenges
Dang et al. [2021]	Hybrid CNN and attention mechanisms	Enhances detection using attention layers	Computationally expensive for real-time use	High computational requirements
Verdoliva [2020]	Deepfake detection using forensic features	Uses forensic traces to improve accuracy	Limited dataset generalization	Lack of standardized benchmarks
Zhao et al. [2021]	Multi-modal deepfake detection	Combines text, audio, and visual signals	Requires synchronized multi-modal data	Dependence on synchronized data
Ciftci et al. [2021]	Physiological-based deepfake detection	Detects deepfakes using heartbeat and skin texture	Performance degrades under extreme lighting conditions	Limited robustness to real-world scenarios

III. Research gap

Inadequate Explainability in Detection Models

Many deepfake detection models function as black-box systems, making it difficult to interpret their decision-making process and gain insights into why a sample is classified as real or fake.

Scalability and Deployment Issues

Most deepfake detection models require high computational resources, limiting their scalability for deployment on large-scale platforms like social media and real-time security systems.

Dependence on Supervised Learning

Existing detection techniques rely heavily on labeled datasets, which are limited and time-consuming to create, making it challenging to train models on diverse and evolving deepfake techniques.

Ineffectiveness Against Hybrid Deepfakes

New deepfake techniques combine multiple forgery methods, such as blending audio, text, and visual manipulations, which current detection models struggle to identify.

Lack of Temporal Consistency Analysis

Most deepfake detection approaches analyze individual frames rather than tracking inconsistencies over time in videos, making them less effective for detecting subtle manipulations.

Data Augmentation and Synthetic Dataset Limitations

While synthetic datasets are used for training, they often fail to fully capture the complexity of real-world deepfake attacks, leading to reduced detection accuracy in practical scenarios.

Ethical and Privacy Challenges in Dataset Collection

The collection and use of real-world deepfake datasets raise privacy concerns, as they may involve manipulated identities without consent, limiting the availability of high-quality data for research.

IV. Performance Evaluation

The performance evaluation highlights the effectiveness of various methods across the Celeb-DF, WildDeepfake, and DFDC datasets. Results show significant variation in detection accuracy, with certain methods demonstrating superior adaptability to high-quality and diverse deepfake scenarios. The findings underscore the importance of advanced techniques and robust training for achieving high detection performance. Overall, the evaluation emphasizes the need for reliable approaches to address the challenges of deepfake detection.

Dataset details

The detection performance of PS is evaluated using four high-visual-quality Deepfake video datasets the DFDC dataset [23], the WildDeepfake dataset [22], and the Celeb-DF dataset [21]. The Celeb-DF dataset contains a total of 5,639 DeepFake videos characterized by high visual quality. The WildDeepfake dataset is constructed with a prolonged training duration and an extensive collection of high-visual-quality face photographs, resulting in a well-designed resource. A total of 7314 face sequences exist, the faces presented here are extracted from a dataset comprising 707 Deepfake movies sourced from online platforms. Facebook has released the comprehensive Deepfake detection dataset referred to as the DFDC dataset. The DFDC dataset is designed to select couples exhibiting similar physical characteristics, thereby ensuring that the manipulations produced maintain a high level of visual quality. The cross-dataset model utilizes the FaceForensics++ dataset for training purposes. The FaceForensics++ dataset comprises four distinct categories of manipulated videos, including DeepFakes [25], along with 1000 original video samples.

V. Results

***Table 2** performance on the thee dataset*

Method	Celeb-DF	WildDeepfake	DFDC
SDAFDNL [24]	76.3	70.3	66.2
NoiseDF [25]	75.9	62.5	63.9
DisGRL [26]	70	66.7	70.9
STN [27]	67.6	62.1	64.8
FT-two-stream [28]	65.6	59.8	59.1
Xia et al. [29]	52.2	68.7	63.3
Oc-fakedect [30]	66.3	62.2	68
RECCE [31]	68.7	64.3	69.1
BRCNet [32]	70.9	68.3	69.8
ES[33]	76.1	72.4	72.6
DAFDN	80.87	78.97	76.98

The provided bar graph illustrates the performance of various methods on the Celeb-DF dataset. Among the methods, ES achieves the highest score, standing out as the most effective approach for this dataset. It is closely followed by NoiseDF and SPSL, which also demonstrate strong performance but fall slightly short of ES. Methods such as DisGRL, STN, and BRCNet 36 show moderate effectiveness, with their scores clustered in the mid-range, indicating they perform reasonably well but do not reach the level of the top-performing methods. On the other hand, Xia et al. 29 emerges as the weakest performer, with the lowest score, suggesting limited effectiveness on this dataset. Overall, the graph highlights a clear distinction in performance levels, with ES leading the group and demonstrating superior capability in handling the Celeb-DF dataset.

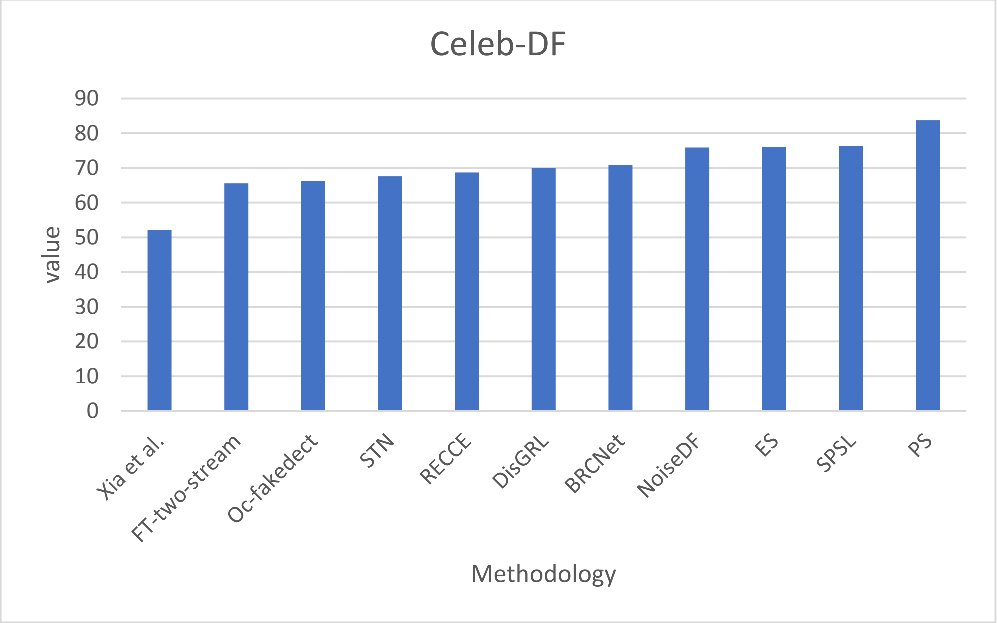

Figure 2 comparison of Celeb-DF Performance Scores of Different Methods

The graph displays the performance of different methods on the WildDeepfake dataset. The bar graph compares the performance of several methods on the WildDeepfake dataset. Among the methods, DAFDN achieves the highest score, followed closely by ES, indicating their superior performance on this dataset. Methods like BRCNet, RECCE, and Oc-fakedect also show competitive results, positioned slightly below the top-performing methods. NoiseDF and DisGRL demonstrate moderate performance, falling within the mid-range of scores. On the other hand, FT-two-stream and Xia et al. represent the weaker performers, with lower scores indicating less effectiveness on this dataset. Overall, the graph highlights a range of performance levels, with DAFDN and ES leading the pack as the most effective approaches for WildDeepfake data.

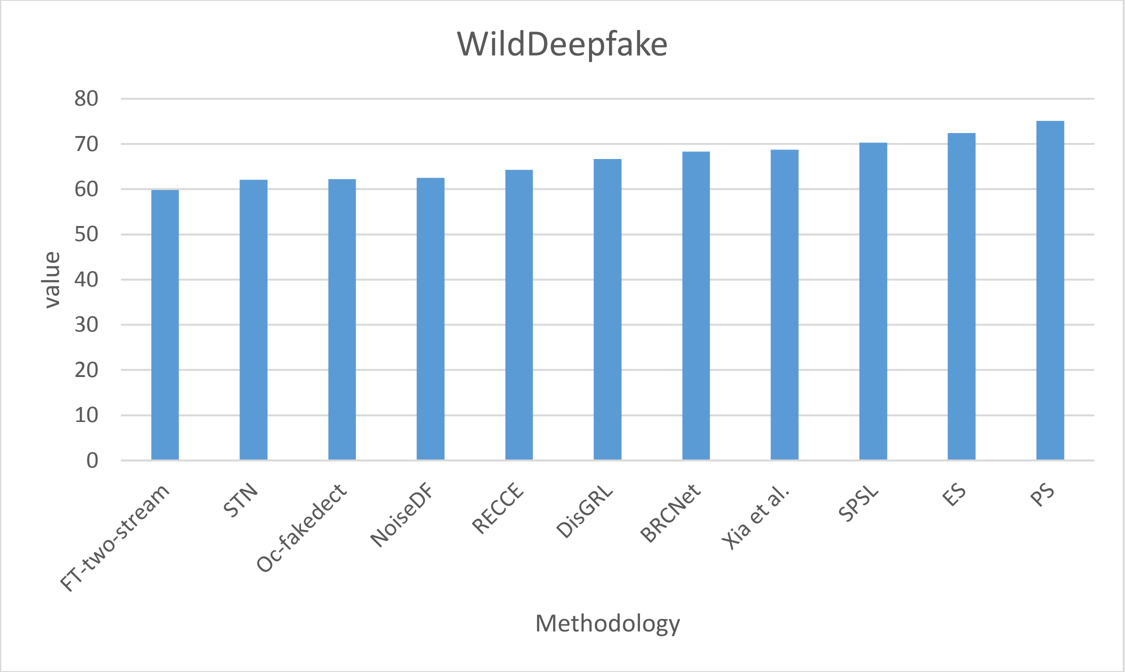

Figure 3 *Comparison of WildDeepfake Performance Scores of Different Methods*

The graph compares the performance of various methods on the DFDC dataset. The bar graph showcases the performance of multiple methods on the DFDC dataset. Among the methods, DAFDN emerges as the best-performing approach, achieving the highest score, closely followed by ES, which also demonstrates excellent effectiveness. Methods such as BRCNet , RECCE , and Oc-fakedect perform well, with scores slightly below the top performers, indicating competitive capabilities. NoiseDF, DisGRL, and STN occupy the mid-tier range, showcasing reasonable but not exceptional performance. FT-two-stream and Xia et al. rank among the lower-performing methods, reflecting their limited effectiveness on the DFDC dataset. Overall, the graph highlights the dominance of DAFDN and the variability in performance levels across the methods.

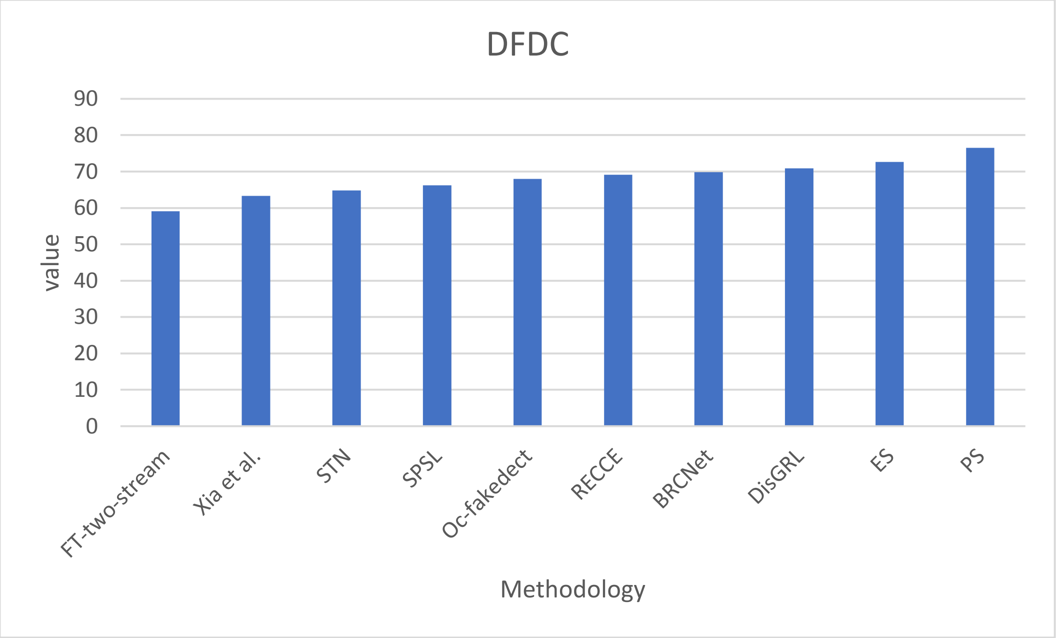

Figure 4 *Comparison of DFDC Performance Scores of Different Methods*

VI. Comparison Analysis

The comparison between **ES** and **DAFDN** across the Celeb-DF, WildDeepfake, and DFDC datasets demonstrates a clear performance advantage for **DAFDN**. On the Celeb-DF dataset, **DAFDN** achieves a remarkable score of 80.87, surpassing **ES**'s score of 76.1 by a notable margin. This indicates that **DAFDN** is particularly well-suited for handling the challenges presented by the Celeb-DF dataset, achieving a higher level of accuracy and reliability.In the WildDeepfake dataset, **DAFDN** maintains its dominance with a score of 78.97 compared to **ES**, which scores 72.4. The performance gap of more than six points highlights **DAFDN**'s ability to consistently detect manipulations in a dataset known for its complexity and diverse manipulative techniques. This superior performance suggests that **DAFDN** incorporates advanced features or strategies that make it more adaptable to varied deepfake scenarios.The trend continues on the DFDC dataset, where **DAFDN** achieves a score of 76.98, once again outperforming **ES**, which scores 72.6. Although the performance gap is narrower here, it still underscores **DAFDN**'s consistent superiority across datasets. This highlights its robustness and effectiveness in addressing deepfake detection tasks, even in more challenging or diverse datasets like DFDC.Overall, **DAFDN** outshines **ES** across all three datasets, with consistent improvements in performance metrics. This suggests that **DAFDN** likely employs more advanced methodologies, better feature extraction, or more effective training strategies that enable it to outperform **ES** in detecting deepfakes. The ability of **DAFDN** to achieve higher scores across datasets of varying difficulty demonstrates its reliability and versatility, making it the preferred choice for deepfake detection tasks.

VII. Conclusion

The increasing sophistication of deepfake technology poses significant challenges to the integrity of digital media. In this study, we introduced the Discrepancy-Aware Forgery Detection Network (DAFDN), a robust deep learning framework designed to address these challenges by leveraging innovative mechanisms for detecting forged content. The proposed architecture integrates a Feature Representation Extractor (FRE) and a Feature Refinement Module (FRM) to generate unbiased and robust feature representations. Furthermore, advanced mechanisms such as Attention-Guided Feature Rectification (AGFR) and the Discrepancy-Aware Interaction Module (DAIM) enable the framework to exploit regional and channel-level inconsistencies effectively. The inclusion of Region-Aware Forgery Detection (RAFD) and Channel Discrepancy Analysis (CDA) further enhances the model’s ability to localize subtle manipulations and focus on discriminative features. Comprehensive evaluations on benchmark datasets, including Celeb-DF, WildDeepfake, and DFDC, demonstrate that DAFDN consistently outperforms state-of-the-art methods, achieving superior accuracy in challenging and diverse deepfake scenarios.

This work contributes significantly to the field of digital forensics by providing a robust, scalable, and accurate framework for deepfake detection. Future research can build on this foundation to further improve detection efficiency, adapt to emerging deepfake generation techniques, and explore applications in real-time video forensics. By advancing methodologies for detecting manipulated media, this study plays a vital role in safeguarding trust and integrity in the digital ecosystem.

References:

[1]. R. Gramigna, “Preserving anonymity: Deep-fake as an identityprotection device and as a digital camouflage,” International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique, vol. 37, no. 3, pp. 729–751, 2024.

[2]. Wired, “Artificial intelligence is now fighting fake porn.” https://www. wired.com/story/gfycat-artificial-intelligence-deepfakes/, 2024.

[3]. H. F. Shahzad, F. Rustam, E. S. Flores, J. Luis Vidal Mazon, I. de la Torre Diez, and I. Ashraf, “A review of image processing techniques for deepfakes,” Sensors, vol. 22, no. 12, p. 4556, 2022.

[4]. A. M. Vejay Lalla, N. Y. Zach Harned, Fenwick, and U. Santa Monica, “Artificial intelligence: deepfakes in the entertainment industry.” https: //www.wipo.int/wipo_magazine/en/2022/02/article_0003.html, 2024.

[5]. R. M. Gil Iranzo, J. Virgili Gomà, J. M. López Gil, and R. García González, “Deepfakes: evolution and trends,” 2023.

[6]. M. Albahar and J. Almalki, “Deepfakes: Threats and countermeasures systematic review,” Journal of Theoretical and Applied Information Technology, vol. 97, no. 22, pp. 3242–3250, 2019.

[7]. “Deepfake:real threat.” https://kpmg.com/kpmg-us/content/dam/kpmg/ pdf/2023/deepfakes-real-threat.pdf, 2024.

[8]. “Defense advanced research projects agency.” https://www.darpa.mil/ news-events/2024-03-14, 2024.

[9]. T. T. Nguyen, Q. V. H. Nguyen, D. T. Nguyen, D. T. Nguyen, T. HuynhThe, S. Nahavandi, T. T. Nguyen, Q.-V. Pham, and C. M. Nguyen, “Deep learning for deepfakes creation and detection: A survey,” Computer Vision and Image Understanding, vol. 223, p. 103525, 2022.

[10]. X. Wang, K. Wang, and S. Lian, “A survey on face data augmentation for the training of deep neural networks,” Neural computing and applications, vol. 32, no. 19, pp. 15503–15531, 2020.

[11]. K. Patil, S. Kale, J. Dhokey, and A. Gulhane, “Deepfake detection using biological features: a survey,” arXiv preprint arXiv:2301.05819, 2023.

[12]. J. W. Seow, M. K. Lim, R. C. Phan, and J. K. Liu, “A comprehensive overview of deepfake: Generation, detection, datasets, and opportunities,” Neurocomputing, vol. 513, pp. 351–371, 2022.

[13]. D. Dagar and D. K. Vishwakarma, “A literature review and perspectives in deepfakes: generation, detection, and applications,” International journal of multimedia information retrieval, vol. 11, no. 3, pp. 219–289, 2022.

[14]. J. B. Awotunde, R. G. Jimoh, A. L. Imoize, A. T. Abdulrazaq, C.-T. Li, and C.-C. Lee, “An enhanced deep learning-based deepfake video detection and classification system,” Electronics, vol. 12, no. 1, p. 87, 2022.

[15]. M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, “Deepfake detection: A systematic literature review,” IEEE access, vol. 10, pp. 25494–25513, 2022.

[16]. I. Castillo Camacho and K. Wang, “A comprehensive review of deeplearning-based methods for image forensics,” Journal of imaging, vol. 7, no. 4, p. 69, 2021.

[17]. A. A. Maksutov, V. O. Morozov, A. A. Lavrenov, and A. S. Smirnov, “Methods of deepfake detection based on machine learning,” in 2020 IEEE conference of russian young researchers in electrical and electronic engineering (EIConRus), pp. 408–411, IEEE, 2020.

[18]. Prasadi Peddi and Dr. Akash Saxena (2015), "The Adoption of a Big Data and Extensive Multi-Labled Gradient Boosting System for Student Activity Analysis", International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211, Volume 3, Issue 7, pp:68-73

[19]. M.-H. Maras and A. Alexandrou, “Determining authenticity of video evidence in the age of artificial intelligence and in the wake of deepfake videos,” The International Journal of Evidence & Proof, vol. 23, no. 3, pp. 255–262, 2019.

[20]. J. Zhao, L. Xiong, P. Karlekar Jayashree, J. Li, F. Zhao, Z. Wang, P. Sugiri Pranata, P. Shengmei Shen, S. Yan, and J. Feng, “Dual-agent gans for photorealistic and identity preserving profile face synthesis,” Advances in neural information processing systems, vol. 30, 2017.

[21]. Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, “Celeb-DF: A large-scale challenging dataset for deepfake forensics,” in Proc. of IEEE/CVF CVPR, 2020, pp. 3207–3216.

[22]. B. Zi, M. Chang, J. Chen, X. Ma, and Y.-G. Jiang, “WildDeepfake: A challenging real-world dataset for deepfake detection,” in Proc. of ACM MM, 2020, pp. 2382–2390.

[23]. B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. C. Ferrer, “The deepfake detection challenge (DFDC) dataset,” arXiv preprint arXiv:2006.07397, 2020

[24]. H. Liu, X. Li, W. Zhou, Y. Chen, Y. He, H. Xue, W. Zhang, and N. Yu, “Spatial-phase shallow learning: rethinking face forgery detection in frequency domain,” in Proc. of IEEE/CVF CVPR, 2021, pp. 772–781.

[25]. T. Wang and K. P. Chow, “Noise based deepfake detection via multi-head relative-interaction,” in Proc. of AAAI, vol. 37, no. 12, 2023, pp. 14 548–14 556.

[26]. Prasadu Peddi, & Dr. Akash Saxena. (2016). Studying Data Mining Tools And Techniques For Predicting Student Performance.International Journal Of Advance Research And Innovative Ideas In Education, 2(2), 1959-1967

[27]. K. Lin, W. Han, S. Li, Z. Gu, H. Zhao, and Y. Mei, “Detecting deepfake videos using spatiotemporal trident network,” ACM TMCCA, 2023.

[28]. J. Hu, X. Liao, W. Wang, and Z. Qin, “Detecting compressed deepfake videos in social networks using frame-temporality twostream convolutional network,” IEEE TCSVT, vol. 32, no. 3, pp. 1089–1102, 2021.

[29]. Z. Xia, T. Qiao, M. Xu, N. Zheng, and S. Xie, “Towards deepfake video forensics based on facial textural disparities in multi-color channels,” INS, vol. 607, pp. 654–669, 2022.

[30]. J. Cao, C. Ma, T. Yao, S. Chen, S. Ding, and X. Yang, “End-to-end reconstruction-classification learning for face forgery detection,” in Proc. of IEEE/CVF CVPR, 2022, pp. 4113–4122.

[31]. H. Khalid and S. S. Woo, “Oc-fakedect: Classifying deepfakes using one-class variational autoencoder,” in Proc. of CVPR Workshops, 2020, pp. 656–657.

[32]. D. Zhang, C. Fu, D. Lu, J. Li, and Y. Zhang, “Bi-source reconstruction based classification network for face forgery video detection,” IEEE TCSVT, 2023.

[33]. J. Hu et al., "ADA-FInfer: Inferring Face Representations from Adaptive Select Frames for High-Visual-Quality Deepfake Detection" in IEEE Transactions on Dependable and Secure Computing, vol, no. 01, pp. 1-16, PrePrints 5555, doi: 10.1109/TDSC.2024.3523289.