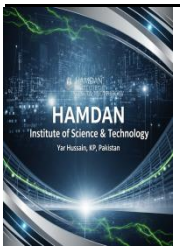

Journal of Fuzzy Intelligence

Journal homepage: www.mathfuzzyjournal.com

Vol: 02 Issue: 01, 2026, pp.84-90

Techniques and Methodologies for the Detection of DeepFake: A Review

Salman Ahmed[1], Wajid Ali[2, *], Amal Kumar Adak[3]

[1]Department of Computer Science, NUML, Islamabad, 4400, Pakistan
[2]Department of Mathematics, Air University Islamabad,4400, Pakistan
[3]Department of Mathematics, Ganesh Dutt College, Begusarai, India
[*]Corresponding author: wajidali00258@gmail.com

Available online: 19 April 2026

ABSTRACT

Deepfake technology has established itself as a powerful tool for generating highly realistic synthetic media, including manipulated images, videos, and speech. However, the misuse of this technology raises serious concerns related to misinformation, identity theft, and harm to society. As deep-fake content becomes more sophisticated, traditional visual inspection by humans is no longer sufficient for reliable detection. In this study, we perform a comparative evaluation of the main deepfake generation techniques and analyze their influence on the performance of existing detection methods. The research aims to identify which deepfake techniques are most effective at evading detection and assess their potential impact on industrial and real-world applications. Based on this analysis, we discuss key challenges and propose future research directions to improve deepfake detection frameworks.

Keywords

Deepfake Detection; Synthetic Media; Deep Learning; Video Forgery; Multimedia Forensics; Zero-Day Attacks

1. Introduction

The word deepfake is circulating around for a long time and have gain a lot of attention of the re- searchers, intelligence agencies, and many other organizations. What exactly is deepfake and why is it getting a lot of attention in the market now a day? These are the few questions that are surfacing in everyone's minds with a little knowledge of the technical industry. Deepfake emerged in late 1990s and has been evolving since then. As computational power and ease of access to such powerful technology is made public and cheap, deepfake progressed at a very fast pace and gained a lot of attention. The key point for this technology to emerge as an important research field is when the video of previous president of United States" Barak Obama" surfaced on the internet saying things that he actually didn't and created a lot of hype in the society as well as in the political environment [1]. This event was the turning point for deep-fake technology, and a lot of research has been done on it since then. The technology to create highly accurate and realistic videos emerged and it became essential to develop a working and accurate technique to detect such forged videos.

Deepfake technology can be used to make fake pornography, videos of someone doing an illegal act or passing unethical comments. It is very important that we as a community also use technology wisely and post everything on social media. This helps in such a way that the less data they have about you, less are the chances of their success of forging realistic videos or images of oneself doing or saying illegal things. It is especially important that we should not believe everything we see on social media and do some research of our own before forwarding and helping them spread fake content on the platforms. In this era of technology everything is becoming online. All of us have social media accounts where we share a lot of personal stuff and are easily accessible to everyone controlling those sites and other organizations that are responsible for their securities and many other such organizations. Even the general public can access the data you share about your personal life on social media platforms. It is very easy to make a forged image or video of someone with the data available on the internet and the advanced techniques that are now available and can be used very easily with very little knowledge. This is a very important problem as the people with ulterior motives can use or maybe are using such techniques to manipulate someone individually or an organization. This is very dangerous in the hands of people with dark motives as they can use this technology to spread fake news to manipulate the community against a single entity, a government or even some religion. It is important to put a stop to this as people tend to accept every and anything they see on the internet without any research. Which requires a method to detect such fake content on

the social media with better accuracy so that the spread of fake news through videos and images can be stopped prior to any damage to the society? A lot of researchers out there are working to develop a technique or methodology to detect forged videos and help society to use technology in a safe environment [2][3][4][5][6]. With the development of such detection techniques the technology of making fake videos is also progressing. This is now a race in which both are progressing forward. The problem of forged images is not a new one and has been there since the technology of images emerged and there have been many techniques to detect them. Since such kind of forgery was done using tools like Photoshop etc. so, there was no need to use the advanced and complicated techniques as it was easy to distinguish between the photo-shopped and real images and videos.

Since the emergence of the generative adversarial networks (GANs) and the advanced powerful computers with high resolution graphic cards it is becoming easy to make forged images or videos that cannot be distinguished from the real ones using the previous techniques or through naked human eyes. It requires more advanced techniques to distinguish the fake images and videos from the real ones. So, the research started using different techniques and some of them gained excellent results which are discussed in Literature review section.

2. Literature Review

The Afchar et al [2] proposed technique using convolutional neural networks (CNNs) and GANs to detect the forged videos and are regarded as one of the earlier researchers back in 2018 to do so. Authors named those techniques Meso-4 and MesoInception-4. Both of these techniques were used to detect the forged videos generated using two well-known techniques DeepFake and Face2Face [7]. They collected data from various sources and generated the deepfakes using the two well-known techniques back then DeepFake and Face2Face. They achieved very good accuracy at that time of 0.969 and 0.984 for deepfake using Meso-4 and MesoInception-4 respectively and for the Face2Face they achieved the accuracy of 0.953 for both techniques.

After the emergence of the first successful technique a lot of other researchers proposed their techniques, one of them were Cozzolino et.al. [3]. They proposed another solution to this problem of forging videos known as 'Forensic Transfer.' As per the authors, the techniques based on the CNNs are only good for detecting the videos from the domain they are trained on and do not perform well on the videos generated through different approaches. They proposed an alternative that would work as a classifier and check for the anomalies in the images and set them aside from the real images chunk. The anomaly would be the altering of videos using any technique the model has not seen during training. Their Proposed technique achieved superior results and is considered as one of steppingstone for future development.

David Guera and Edward J. Delp proposed a technique to detect fake videos using recurrent neural networks (RNN) [4]. They created a pipeline in which first a CNN is used to extract the feature from the frames known as frame level features and then trained a RNN to classify the fake and real videos using those extracted features. They used the HOHA-dataset [8] to train and evaluate their model. The dataset contains 300 real videos and 300 forged videos collected from the movies. They achieved very promising results with the accuracy of 0.971 and stated that their model can detect the fake videos in 2 seconds.

A group of researchers proposed a technique of detecting the forged videos using the eye blinking [9]. This technique counts the number of times a person blinks in a minute. This is one of the fundamental techniques in the detection of the fake videos even before the evolution of deepfakes. Authors trained the Long-term Recurrent Convolutional Networks (LRCN) [10] using the dataset they collected on their own and achieved the accuracy of 0.99. Authors also suggested that this technique can be enhanced in future using more sophisticated techniques.

In 2019 another group of researchers proposed a technique to detect deepfakes based on the fact that deepfakes are generated through splicing synthetic regions of the face onto other images [11]. They proposed that the estimation of the 3D head poses can be used against the deepfake generators and detect them. They trained a classifier based on the feature of head poses and achieved an accuracy of 0.89. In this study [5], authors stressed upon the fact that the images generated through the GANs are different from the images captured through the camera. They used this cue and created a GAN that can capture the difference in image pixels and then passed those features through a support vector machine (SVM) which is used as a classifier to differentiate those images.

In this study [6], authors proposed a technique that can possibly challenge the forgery techniques and also provided a dataset containing around 1.8 million entries. According to the authors, the lack of good datasets and features are the main reason that most of the techniques do not perform well in real world scenarios. Authors claimed in this study that the features they crafted are better than the hand-crafted features being used in detectors and provide more accurate results. They also claimed that their dataset and the benchmarks features will be very helpful in the future of the deepfake detection.

Table 1 List of all approaches, their results, and advantages/disadvantages

Approaches	***Methodology***	***Dataset***	***Accuracy***	***Advantages /Disadvantages***
[2]	Meso-4 MesoInception-4	Face Forensics (FF) And privately collected data from different sources	0.98 for deepfake videos 0.95 for Face2Face videos (Detection rate)	Very novel technique but works only for deepfakes generated through these two techniques
[4]	CNN and LSTM	HOHA dataset	0.967 (20frames) 0.971 (40frames)	Does not give good results on the low-quality videos
[9]	Deep Neural Network LRCN (long term re-current convolutional network)	N/A	Eye-blinking "34.1/min(real) 3.4/min(fake)"	Very easy to overcome this. weakness in deepfakes
[11]	SVM classifier using head poses GNN	UADFV dataset	AUROC 0.890 (frames) 0.974 (videos)	N/A
[5]	SVM	N/A	0.92	N/A
[6]	Xception-net (CNN), LSTM	FF++	0.81	Novel Technique is used here
[12]	MLP (NN) Classifier, Logistic Regression	Celeba, ProGAN, Glow	0.84 (MLP) 0.83 (LogReg)	Does not account for the real world and noisy data
[16]	Features extractor algo. and SVM classifier	Private dataset	HOG 0.945, SURF 0.90, KAZE 0.765	A novel approach to that time and achieved acceptable results
[18]	ModalityDissonance Score (MDS)	DFDC and DeepFake-TIMIT	0.915 (DFDC) 0.979 (DF-TIMIT LQ) 0.968(DF-TIMIT HQ)	Results cannot be verified due to very less detail
[19]	CNN	Celeba, AFW, FDDB	0.95 discrete 0.74 continuous	Works best only for the method it was trained on
[20]	Voila jones face detection technique	Private dataset	0.88	N/A
[21]	CNN, LSTM	Face Forensics++(FF++), Deepfake, TIMIT, US-DFV, Celeb-D	0.7767	The accuracy of the pro-posed technique is not best
[22]	(ILSVRC) 2012-pre-trained ResNet-18	FF++, Dessa, Celeb-DF, Google DFD	0.9223	An outstanding technique but consumes many computational power
[23]	CNN, LSTM	FF++, Celeb-DF, Deepfake Detection Challenge (DFDC)	0.84 with & 0.75 without transfer learning	Need a huge amount of data to achieve good results
[24]	Xception-net L R P a n d LIME	FF++	0.9017	No cross-data evaluation shown. No way to check the biases in results
[25]	SVM, KNN	Celeba	0.9267 (ATTGAN), 0.8840 (GDWCT), 0.9317 (STARGAN), 0.9957(STYLEGAN,0.9981(STYLEGAN2)	Needs to be evaluated on real world data
[26]	Biological signal m o d e l i n g , Feature extraction Classifier	FF, FF++, Celeb-DF, Private ("in the wild")	0.9939	A quite simple technique yet gives very good results
[27]	Face X-ray	FF++	0.9773 (Face2Face)	Does not work well on low. quality data
[28]	Pairwise self-consistency learning (PCL)	FF++	0.9805 (in-dataset) 0.9218 (cross dataset)	Extremely easy to overcome and does not perform well on low quality data

Visual artifacts are the root of the standard computer vision problem solving. These simple artifacts can be used to detect the deepfakes generated through advanced GANs [12], even though they did not perform better than the existing techniques like Afchar et. al. [2]. The technique achieved good results even using such simple features of the images. Authors used three datasets for their model training and evaluation in this study which are ProGAN [13], CelebA [14], and Glow [15]. Another way of solving the deepfake problem is to use the SVM regression on different features extracted through different techniques [16]. This also shows a simple way of exploiting the weakness of the deepfakes, but the results were not up to par or better than the existing techniques.

After a long race with the deepfakes a promising technique surfaced [17] in which authors used the optical flow of the frames in a video instead of performing experiments on single frames and trained CNN on it to distinguish between real and fake content. They used the FaceForencis++ [6] dataset to train their model and achieved an accuracy of 0.8161.

In another study [18], authors proposed a completely fresh idea for the detection of deep-fake videos as they used an exception maximization algorithm to extract the features of the frame and pass them through the convolutional generative process. They created their own dataset using the available tools for deepfake generation such as GDWCT, STARGEN, STYLEGAN/2 etc. and achieved very good detection accuracy. Their methodology has many limitations. They used the idea of the kernel impression on the images to distinguish reality from fake images.

A comparison Table 1 is given below to compare the most relevant studies in this domain with their advantages and disadvantages. In this table, methodologies that have been used in this study as well as datasets are also given for better overview. Moreover, results of the studies are also given in the table from which one can get idea of the effectiveness of the studies against the dataset and techniques that have been used.

3. Discussion

From the above given table and literature review discussion, it is evident that with the emergence of high computational power machines, social media, and deep learning, which is the sub-field of artificial intelligence, deep-fake technology has also emerged. It is not a bad field, but it is usage for ill purposes can be dangerous. That is one of the main reasons that researchers are trying to figure out a universal technique that can be used to detect the forged images and videos so that the misuse of this technology can be discouraged. In the last 6 to 7 years, many researchers proposed many techniques to detect deep-fake images and videos so that this problem can be addressed properly.

In this review study, I have reviewed all such studies in which such techniques have been proposed, and I have discussed them and provided their overview. To search the most relevant and authentic studies, I used some keywords to search the journal and conference papers such as," deepfake detection"," detection of deepfake video manipulation"," detection of deepfake forged images" and so on. Moreover, I also set the criteria to include only those studies which are presented at well-known conferences or published in well-known journals. After sorting all the studies, I selected 28 conference and journal papers to include in this review study, and all these papers are also visually illustrated in terms of number of papers selected from different below in Figure 1.

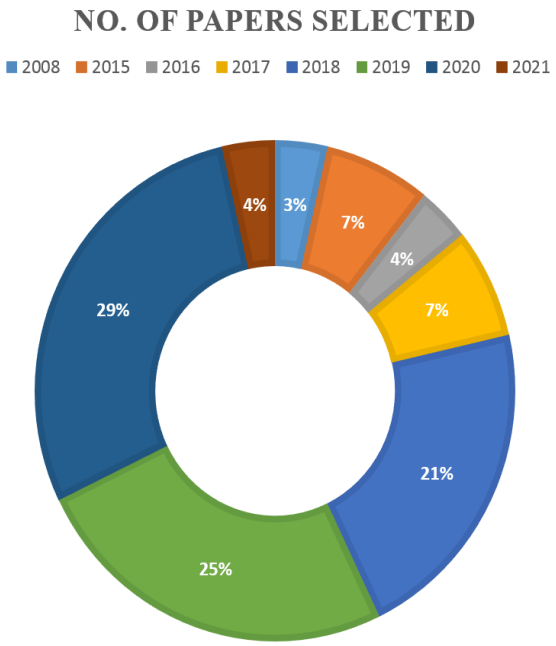

Figure 1: Selected Papers from different years for review.

4. Conclusions

The rapid advancement of deep-fake technology has introduced a new class of digital threats that significantly undermine trust in visual, audio, and multimedia content. As deep-fake generation methods continue to improve in realism and accessibility, their misuse poses serious risks to individuals, organizations, governments, and society at large. This review has comprehensively examined the evolution of deep-fake technologies and critically analyzed the state-of-the-art detection techniques developed to counter these threats. Through an extensive comparison of existing methodologies, it is evident that both traditional machine learning and modern deep learning approaches play a vital role in deepfake detection. Early detection systems relied heavily on handcrafted features, such as facial landmarks, eye-blinking patterns, and texture inconsistencies. However, these approaches often fail when confronted with high-quality synthetic media generated by advanced models. In contrast, deep learning-based techniques, particularly those utilizing Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks, have demonstrated superior performance by automatically learning discriminative spatial and temporal features from large-scale datasets.

Among the reviewed approaches, hybrid architectures that combine CNNs for spatial feature extraction with LSTMs for temporal sequence modeling have shown remarkable effectiveness in video-based deepfake detection. These models are particularly successful in capturing subtle temporal inconsistencies across frames, which are often imperceptible to the human eye. Despite their promising performance, such methods are still constrained by high computational costs, dependence on large, annotated datasets, and limited generalization to unseen deepfake generation techniques.

4.1. Future Directions

Although existing deepfake detection methods have achieved encouraging results, several critical challenges remain and open promising avenues for future research. One of the most pressing issues is the heavy reliance of current models on large-scale labeled datasets. Collecting and annotating such datasets is not only time-consuming and expensive but also impractical in rapidly evolving real-world scenarios. Therefore, future research should focus on data-efficient learning strategies that can achieve high detection accuracy with limited training data. Techniques such as few-shot learning, semi-supervised learning, self-supervised learning, and transfer learning are particularly promising in this regard. Another major challenge is the detection of zero-day deepfakes synthetic media generated using previously unseen manipulation techniques. These deepfakes are especially dangerous, as they can bypass existing detection systems and spread widely before being identified. To address this issue, future detection frameworks should prioritize generalization and robustness rather than performance on specific datasets. Anomaly detection, domain adaptation, and continual learning approaches may help build systems capable of identifying novel and evolving deepfake attacks.

In addition, most current deepfake detection models are computationally intensive, making them unsuitable for real-time deployment on resource-constrained devices or large-scale online platforms. As a result, there is a growing need for lightweight and efficient detection models that balance accuracy with speed and computational cost. Model compression, knowledge distillation, and edge-based detection techniques represent important research directions for enabling real-time and scalable deepfake detection.

From an industry and cloud-computing perspective, integrating deep-fake detection systems into content moderation pipelines, social media platforms, and digital forensics tools remains an open challenge. Future work should explore end-to-end detection frameworks that seamlessly operate in cloud environments, ensuring scalability, privacy preservation, and robustness against adversarial attacks. Furthermore, multimodal detection approaches that jointly analyze visual, audio, and textual cues are expected to significantly enhance detection performance, particularly for complex and highly realistic deepfakes. Future research may explore the integration of deep reinforcement learning and cooperative multi-agent frameworks to develop adaptive deep-fake detection systems capable of responding dynamically to evolving and zero-day attacks [29,31]. Additionally, incorporating decision-making and uncertainty modeling techniques, such as fuzzy and intuitionistic aggregation methods, can enhance the robustness, interpretability, and reliability of deepfake detection models in real-world and cloud-based environments [30,32].

***Author contributions**:* All authors equally contributed to this article.
Funding Information: Funding information is not available.
Data Availability: All data generated or analyzed during this study are included in this published article.

References

[1] S. Suwajanakorn, S. M. Seitz, and I. Kemelmacher-Shlizerman, "Synthesizing obama," *ACM Transac- tions on Graphics (TOG)*, vol. 36, pp. 1 – 13, 2017.

[2] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "Mesonet: a compact facial video forgery detection network," *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1–7, 2018.

[3] D. Cozzolino, J. Thies, A. Ro¨ssler, C. Riess, M. Nießner, and L. Verdoliva, "Forensictransfer:

Weakly- supervised domain adaptation for forgery detection," *ArXiv*, vol. abs/1812.02510, 2018.

[4] D. Guera and E. J. Delp, "Deepfake video detection using recurrent neural networks," *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6, 2018.

[5] S. McCloskey and M. Albright, "Detecting gan-generated imagery using saturation cues," *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 4584–4588, 2019.

[6] A. Ro¨ssler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "Faceforensics++: Learn- ing to detect manipulated facial images," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1–11, 2019.

[7] J. Thies, M. Zollho¨fer, M. Stamminger, C. Theobalt, and M. Nießner, "Face2face: Real-time face capture and reenactment of rgb videos," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2387–2395, 2016.

[8] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, "Learning realistic human actions from movies," *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[9] Y. Li, M.-C. Chang, and S. Lyu, "In ictu oculi: Exposing ai created fake videos by detecting eye blinking," *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1–7, 2018.

[10] J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, T. Darrell, and K. Saenko, "Long-term recurrent convolutional networks for visual recognition and description," in *CVPR*, 2015.

[11] X. Yang, Y. Li, and S. Lyu, "Exposing deep fakes using inconsistent head poses," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8261–8265, 2019.

[12] F. Matern, C. Riess, and M. Stamminger, "Exploiting visual artifacts to expose deepfakes and face ma- nipulations," *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, pp. 83–92, 2019.

[13] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *ArXiv*, vol. abs/1710.10196, 2018.

[14] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 3730–3738, 2015.

[15] D. P. Kingma and P. Dhariwal, "Glow: Generative flow with invertible 1x1 convolutions," in *NeurIPS*, 2018.

[16] F. F. Kharbat, T. Elamsy, A. Mahmoud, and R. Abdullah, "Image feature detectors for deepfake video detection," *2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA)*, pp. 1–4, 2019.

[17] I. Amerini, L. Galteri, R. Caldelli, and A. Bimbo, "Deepfake video detection through optical flow based cnn," *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 1205–1207,2019.

[18] K. Chugh, P. Gupta, A. Dhall, and R. Subramanian, "Not made for each other- audio-visual dissonance- based deepfake detection and localization," *Proceedings of the 28th ACM International Conference on Multimedia*, 2020.

[19] M. Luo, Y. Xiao, and Y. Zhou, "Multi-scale face detection based on convolutional neural network," *2018 Chinese Automation Congress (CAC)*, pp. 1752–1757, 2018.

[20] C.-C. Low, L.-Y. Ong, and V. C. Koo, "Experimental study on multiple face detection with depth and skin color," *2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, pp. 114–119, 2019.

[21] P. Chen, J. Liu, T. Liang, G. Zhou, H. Gao, J. Dai, and J. Han, "Fsspotter: Spotting face-swapped video by spatial and temporal clues," *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, 2020.

[22] S. Aneja and M. Nießner, "Generalized zero and few-shot transfer for facial forgery detection," *ArXiv*, vol. abs/2006.11863, 2020.

[23] P. Ranjan, S. Patil, and F. Kazi, "Improved generalizability of deep-fakes detection using transfer learning based cnn framework," *2020 3rd International Conference on Information and Computer Technologies (ICICT)*, pp. 86–90, 2020.

[24] B. Malolan, A. Parekh, and F. Kazi, "Explainable deep-fake detection using visual interpretability meth- ods," *2020 3rd International Conference on Information and Computer Technologies*

(ICICT), pp. 289– 293, 2020.

[25] L. Guarnera, O. Giudice, and S. Battiato, “Deepfake detection by analyzing convolutional traces,” *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2841– 2850, 2020.

[26] U. A. Ciftci and I. Demir, “Fakecatcher: Detection of synthetic portrait videos using biological signals,” *IEEE transactions on pattern analysis and machine intelligence*, vol. PP, 2020.

[27] L. Li, J. Bao, T. Zhang, H. Yang, D. Chen, F. Wen, and B. Guo, “Face x-ray for more general face forgery detection,” *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5000–5009, 2020.

[28] T. Zhao, X. Xu, M. Xu, H. Ding, Y. Xiong, and W. Xia, “Learning self-consistency for deepfake detec- tion,” *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 15003–15013, 2021.

[29] U. Rahim, W. Ali, and U. Ghanni. "Deep Reinforcement Learning for Robust USVs Navigation in Diverse Environmental Scenarios." *International Journal of Data Informatics and Intelligent Computing* 4.4 (2025): 1-10.

[30] M. Yaseen, W. Ali, U. Ghani, R. U. Khan, and A. K. Adak. "Advancing Aviation Safety and Sustainable Infrastructure: High-Accuracy Detection and Classification of Foreign Object Debris Using Deep Learning Models." *International Journal of Sustainable Development Goals* 1 (2025): 82-98.

[31] U. Rahim, M. Saeed, W. Ali, J. Nazar, and F. Nazar. "A Cooperative Heterogeneous Multi-Agent System Leveraging Deep Reinforcement Learning." *Knowledge and Decision Systems with Applications* 1 (2025): 112-124.

[32] A. Wajid, T. Shaheen, H. G. Toor, F. Akram, M. Z. Uddin, and M. M. Hassan. "An innovative decision model utilizing intuitionistic hesitant fuzzy aczel-alsina aggregation operators and its application." *Mathematics* 11, no. 12 (2023): 2768.