Performance Evaluation of Enhanced DCGANs for Detecting Deepfake Audio Across Selected FoR Datasets

[1] Jovelin M. Lapates
Technological Institute of the Philippines
Aurora Blvd., Cubao, Quezon City, Philippines
qjmlapates@tip.edu.ph

[2] Bobby D. Gerardo, D.Eng
Northern Iloilo State University
Estancia, Iloilo, Philippines
bgerardo@nisu.edu.ph

[3] Ruji P. Medina, Ph.D
Technological Institute of the Philippines
Aurora Blvd., Cubao, Quezon City, Philippines
ruji.medina@tip.edu.ph

***Abstract*—As the generation of deepfake audio becomes more advanced, the need for effective detection techniques has grown significantly. This study explores the application of an enhanced DCGAN for detecting deepfake audio, with a focus on the Fake or Real (FoR) dataset. The model utilizes convolutional layers to process spectrograms derived from audio signals. By incorporating batch normalization in both the generator and discriminator, the model addresses training instability and improves convergence. The audio data, including real human speech and deepfake renditions generated through Retrieval-based Voice Conversion (RVC), were preprocessed using Audacity and Sonic Visualizer. The enhanced DCGAN was then trained on these spectrograms, leveraging adversarial training to improve detection capabilities. Performance evaluation of the model involved key metrics such as accuracy, precision, recall, and F1-score. The results revealed an accuracy rate of 98%, marking a 6.96% improvement over the standard DCGAN. Furthermore, the enhanced DCGAN exhibited superior precision and recall in distinguishing between real and fake audio samples.**

***Keywords*— DCGANs, deepfake audio, FoR dataset, spectrogram**

I. Introduction

As the technology behind deepfake audio continues to advance, the need for effective detection methods becomes increasingly critical. Using Deep Convolutional Generative Adversarial Networks (DCGANs) is one promising approach. Enhanced versions of these networks have shown potential in improving the accuracy and reliability of deepfake audio detection.

Technological advancements have enabled the use of Generative Adversarial Networks (GANs) [1] for creating deep fake audio, utilizing a generator and a discriminator that compete to improve the model's performance. GANs involve two neural networks, where the generator creates data samples and the discriminator distinguishes between real and generated samples. DCGANs, an extension of GANs, incorporate convolutional and deconvolutional layers, making them particularly effective for handling image data. DCGAN uses convolutional operations and spatial up-sampling to enhance image generation and addresses the mode collapse issue, where the generator biases toward a limited set of outputs [2]. This architecture, which allows DCGANs to generate highly realistic images by capturing intricate spatial hierarchies, as demonstrated by [3], is also stable in training and serves as a foundation for other GAN architectures [2]. Through adversarial training, the generator and discriminator continuously refine their abilities, leading to the production of images that closely resemble real data.

Although DCGAN is originally designed for image data, it can be adapted for audio by transforming the audio into a 2D representation, such as a spectrogram—a visual representation of sound frequencies over time. The model can then be trained on these spectrograms in much the same way it processes images, enabling the generator to create new spectrograms that can be converted back into audio waveforms [4]. However, applying DCGAN to audio may require additional preprocessing and adjustments to achieve effective results, but it has been successfully used in tasks like music generation and voice synthesis.

The increasing prevalence of deep fake audio has led to the use of Explainable AI (XAI) techniques to identify fake frequencies, which helps in reviewing and improving deep fake audio creation using GANs [5]. Various deep learning methods for detecting deep fakes have been reviewed, including CNNs with fully connected networks, hybrid models, and generative models like Autoencoders and GANs, as well as Recurrent Neural Networks such as LSTM, GRU, and Transformers [6]. DCGANs use convolutional layers for both downsampling and upsampling, employing convolutional and transposed convolutional layers for generating and discriminating functions [7].

Despite limited exploration of joint audiovisual representation learning for deepfake detection [8], the rapid advancement in generating realistic fake audio [9] with sophisticated neural networks has surpassed traditional detection methods, highlighting the need for more advanced tools. Among these tools, DCGANs have demonstrated potential in improving detection capabilities for deepfake audio. This study aims to assess the performance of enhanced DCGANs in detecting deepfake audio across selected FoR datasets, providing valuable insights into their effectiveness and potential applications in addressing the growing challenge of deepfake audio.

II. RELATED WORKS

The literature on DCGANs is extensive and influential, focusing on their performance and characteristics.

The emergence of deepfake audio, which involves generating highly realistic synthetic audio using advanced neural networks, has significantly outpaced traditional detection methods, highlighting the urgent need for more sophisticated detection tools.

TABLE 1. COMPARISON OF DEEPFAKE AUDIO DETECTION APPROACHES

Deepfake Detection Tool and Author	Datasets Used	Model Used	Accuracy
DeepSonar [10]	ASVspoof 2019, TIMIT	Custom ultrasonic waveform model	~92%
Resemblyzer[11]	LibriSpeech, VoxCeleb	Resemblyzer model (embedding-based neural network)	~85%
DeFake[8]	VoxCeleb2, FakeAVCeleb	Multimodal deep learning model (audio-visual)	~88%
AutoVC-based Detector [12]	VCTK Corpus, LibriTTS	AutoVC model (voice conversion-based)	~80%
WaveFake [13]	WaveFake (a combination of multiple datasets)	Waveform analysis model with deep learning	~83%
SE-ResNet50 [14]	ASVspoof 2019, LA & PA datasets	SE-ResNet50 (deep learning architecture)	~87%

Deepfake detection applications employ various techniques with distinct strengths and weaknesses. DeepSonar performs well in controlled conditions but struggles in noisy environments, Resemblyzer is effective for specific languages but can miss high-quality deepfakes, and DeFake integrates visual and audio data, though it is computationally intensive; AutoVC-based Detector, WaveFake, and SE-ResNet50 each have limitations in generalization and validation across diverse deepfake types and real-world scenarios.

DCGANs have shown promise in improving the detection of such forgeries due to their ability to learn complex features and patterns from data. An improved DCGAN, by [15], features a double branch structure aimed at extending SWIR-HSI spectra. This enhancement improved classifier accuracy, as evaluated using a 1-NN classifier. The effectiveness of this improved DCGAN was further verified by three commonly used classifiers: decision tree (DT), random forest (RF), and support vector machine (SVM), as demonstrated by [16].

SA-SN- CGAN method [17], generates synthetic images to meet training requirements by integrating a DCGAN [18], [19] with self-attention (SA) and spectral normalization (SN). This method combines these techniques for synthetic image generation, uses PSNR-CWT preprocessing for signal-to-image conversion, and enhances fault diagnosis performance, especially in small-sample cases, through fine-tuning transfer learning.

Data augmentation methods are crucial for generating new datasets from existing ones [20]. However, to address unbalanced data, data augmentation is necessary, but surrogate data can often be physically invalid, leading to inaccurate predictions. To mitigate this, [21] developed a deep learning model with physical constraints to predict porosity in laser metal deposition (LMD), utilizing a DCGAN for data generation.

To create large datasets affordably, [22] utilizes a DCGAN. This approach generates synthetic images that deceive the discriminator into believing they are real, followed by additional pre-processing techniques. The study explores a synthetic speech dataset called Fake or Real (FoR) by [23], which enhances the diversity of deepfake speech data for this purpose.

III. METHODS

To effectively assess the performance of the enhanced DCGAN, particular samples from the Fake or Real (FoR) dataset [24] were carefully selected. This targeted selection ensures that the chosen data aligns with the objectives of the evaluation, allowing for a focused assessment of the model's ability to generate realistic outputs and improve upon the existing architecture.

TABLE 2. DATASETS SUBJECTED FOR PRE-PROCESSING, TRAINING, TESTING AND VALIDATION

FoR	Speech	Length (MM:SS)
Real	Barack Obama	10:00
Fake	Biden-to-Obama	10:00
	Linus-to-Obama	9:30
	Musk-to-Obama	10:00
	Ryan-to--Oobama	1:33
	Taylor-to-Obama	10:00
	Trump-to-Obama	10:00

All audio data were preprocessed using Audacity, where each audio file was trimmed to a length of 3 seconds. The preprocessed audio was then converted into spectrograms using Sonic Visualizer, and data augmentation techniques [25] such as time stretching and pitch shifting were applied before inputting the spectrograms into the DCGAN.

3.1 Model Enhancement

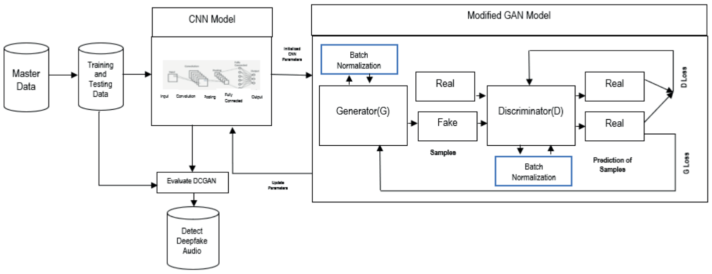

Fig 1. Enhanced DCGAN Architecture

Following the successful applications of convolutional neural networks (CNNs) [26], [27], [28] researchers recognized their potential and combined CNNs with GANs [3] to introduce DCGAN.

To design an enhanced DCGAN architecture tailored for audio data, it is essential to adapt the conventional DCGAN [29] to accommodate the unique characteristics of audio signals, such as their temporal structure and frequency content. The generator upsample a latent vector, typically sampled from a normal distribution, into audio-like structures such as waveforms or spectrograms, with extensive use of batch normalization to stabilize training and improve convergence. The discriminator, on the other hand, designed to handle the temporal relationships in audio data while distinguishing between real and fake samples [30], [31]. Standard GAN loss functions, like binary cross-entropy, are employed, with possible modifications [32] such as the Wasserstein loss to enhance training stability.

3.2 Training Procedure

Several hyperparameter were adjusted to enhance DCGAN for audio deepfake detection such as **batch size, number of layers and neurons,** and experiment with different **activation functions** and **loss functions** to optimize the model's ability [33] to generate and distinguish audio features.

3.3 Performance Evaluation

The performance of the enhanced DCGANs were assessed using metrics such as accuracy, precision, recall, and F1-score. Accuracy measures the proportion of correctly predicted instances (both true positives and true negatives) out of the total instances. Precision indicates how many of the predicted positives were actually correct. Recall measures how well the model identifies true positives. F1-score is the harmonic mean of precision and recall, providing a single measure that balances both concerns. These metrics are calculated for each FoR dataset to provide a comprehensive evaluation as follows.

$$accuracy = \frac{Number\ of\ correctly\ classified\ samples}{Total\ number\ of\ samples} x\ 100\% \qquad (1)$$

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \qquad (2)$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \qquad (3)$$

$$F1\ Score = 2x \frac{Precision * Recall}{Precision + Recall} \qquad (4)$$

Cross-dataset validation was conducted to test the model trained on one FoR dataset on another to assess generalization capabilities.

IV. EXPIRIMENTAL RESULS

For the experiments conducted in this study, the authors utilized Google Colab with a T4 GPU for computational resources, along with Python 3.6.6 and TensorFlow 1.10.0 for model development and training. MATLAB and Librosa were employed for additional data processing and analysis.

4.1 Fake vs Real Spectrogram Samples

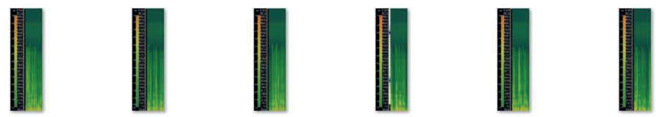

Fig. 2 Fake Spectrogram Samples

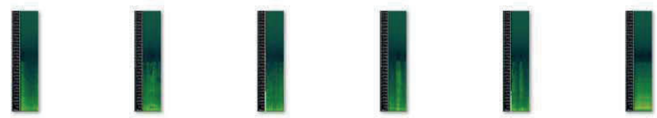

Fig. 3 Real Spectrogram Samples

The figures represent preprocessed spectrogram samples, with Fig 2 showcasing fake spectrograms and Fig 3 displaying real spectrograms. These preprocessed spectrogram samples were used in the training, testing, and validation phases of the model development.

4.2 Summary of Enhanced DGAN model architecture

The summary outlines the model's layers, output shapes, and parameters, summarizing the architecture of the sequential model used.

TABLE 3. DETAILED LAYER OF THE SEQUENTIAL MODEL

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 62, 62, 64)	1,792
max_pooling2d (MaxPooling2D)	(None, 31, 31, 64)	0
conv2d_1 (Conv2D)	(None, 29, 29, 128)	73,856
max_pooling2d_1 (MaxPooling2D)	(None, 14, 14, 128)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 128)	3,211,392
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 2)	258

The summary shows a sequential neural network architecture comprising multiple layers designed for processing input data, such as spectrograms, using convolutional and dense layers. The model begins with two Conv2D layers, each followed by MaxPooling2D layers, which reduce spatial dimensions and control overfitting. These convolutional layers extract features by capturing patterns like edges or textures. A Flatten layer then converts the 3D output into a 1D vector, preparing it for the fully connected Dense layers, where complex relationships in the data are learned. A Dropout layer is included to prevent overfitting by randomly dropping neurons during training. The final Dense layer, with 2 neurons, outputs the model's predictions for classification tasks (e.g., real vs. fake). The model has a total of 3,287,298 trainable parameters and achieved a test loss of 0.051 and a test accuracy of 98.4%, demonstrating its robustness and effectiveness in classification tasks.

4.3 Testing Accuracy and Loss Results

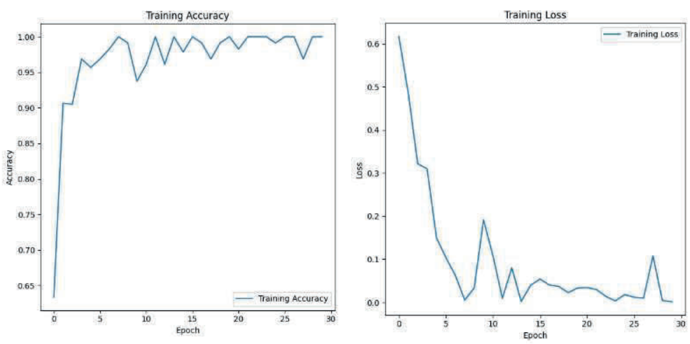

Fig. 4 Graph of Training Accuracy and Training Loss

The training data were processed over 30 epochs. During this period, Training Loss illustrates the model's performance, with an initial sharp decrease indicating rapid learning, followed by fluctuations but an overall downward trend, suggesting improvement over time. Similarly, Training Accuracy shows a sharp increase initially, indicating rapid improvement, and around epoch 5, the accuracy plateaus, maintaining high values above 0.95 for the remaining epochs, suggesting that the model achieves and sustains excellent accuracy early in the training process.

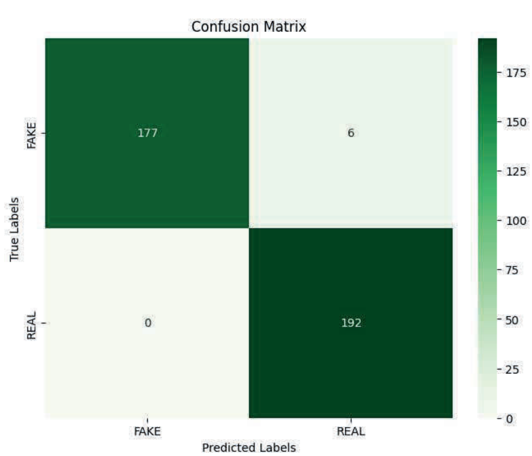

Fig. 5 Confusion Matrix

The confusion matrix shows the performance of a classification model distinguishing between 'fake' and 'real' labels. It correctly identified 177 'fake' instances and 192 'real' instances, with 6 'fake' instances misclassified as 'real'. There were no 'real' instances misclassified as 'fake'. The color intensity indicates the frequency of each classification, with darker shades representing higher counts.

TABLE 4. PERFORMANCE METRICS COMPARISON FOR DCGAN AND ENHANCED DCGAN

	DCGAN			Enhanced DCGAN		
	Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
Fake	1.00	0.75	0.86	1.00	0.97	0.98
Real	0.81	1.00	0.90	0.97	1.00	0.98

The table compares the performance metrics of a standard DCGAN and an Enhanced DCGAN across precision, recall, and F1-score for both fake and real images. The Enhanced DCGAN, excels with a precision of 0.97 and perfect recall, achieving an F1-score of 0.98. These results emphasize that the Enhanced DCGAN significantly outperforms the standard DCGAN in both distinguishing fake and real images, demonstrating superior overall effectiveness in performance metrics.

TABLE 5. COMPARISON OF AVERAGED VALIDATION METRICS OF THE MODELS

Model	**Acc**	**Pre**	**Rec**	**F1**
XGBoost (300)	.993	.995	.991	.993
Random Forest (310)	.989	.995	.983	.989
Quadratic Discriminant Analysis	.948	.969	.924	.946
Linear Discriminant Analysis	.889	.886	.893	.889
Ridge	.883	.884	.882	.883
Naïve Bayes (Gaussian)	.83	.864	.784	.822
KNN 1	.815	.808	.827	.817
SVM	.723	.815	.576	.675
Naïve Bayes (Bernoulli)	.692	.742	.587	.655
Stochastic Gradient Descent	.668	.732	.76	.681
Gaussian Process	.614	.97	.229	.372
DCGAN	.92	.93	.795	.675
Enhanced DCGAN	**.984**	**.98**	**.98**	**.98**

The table [34] presents a comparative analysis of various models across accuracy, precision, recall, and F1-score. Among the models listed, the Enhanced DCGAN stands out with an impressive accuracy of 0.984, precision of 0.980, recall of 0.980, and an F1-score of 0.980. This performance highlights its superior effectiveness in classification tasks, significantly outperforming models like XGBoost and Random Forest, which have slightly lower F1-scores of 0.993 and 0.989, respectively. In comparison, more basic models such as KNN, SVM, and Gaussian Process show notably lower metrics across all measures. The Enhanced DCGAN's consistent high performance underscores its exceptional capability in delivering reliable and accurate predictions, setting it apart from other algorithms listed in the table.

V. CONCLUSIONS

As deepfake audio technology advances, effective detection methods are increasingly critical. This study evaluates the performance of Enhanced DCGANs for detecting deepfake audio using the selected FoR datasets. Other deepfake detection techniques, including CNNs, hybrid models, and generative approaches, have struggled to keep pace with the rapid evolution of deepfake audio technologies. Enhanced DCGANs, with their advanced architecture and training methodologies, offer a promising solution. The study involves preprocessing audio data into spectrograms, training the Enhanced DCGAN on these representations, and evaluating its performance using metrics such as accuracy, precision, recall, and F1-score.

Results demonstrate that the Enhanced DCGAN outperforms other models significantly. It achieved an accuracy of 98%, surpassing other methods such as XGBoost and Random Forest. The Enhanced DCGAN's ability to maintain high performance across various FoR datasets highlights its robustness and effectiveness in detecting deepfake audio.

This study confirms that Enhanced DCGANs represent a significant advancement in the field of deepfake detection, offering superior performance and reliability compared to existing techniques. The findings underscore the potential of Enhanced DCGANs as a powerful tool for addressing the growing challenge of deepfake audio, suggesting their promising application in future detection systems.

ACKNOWLEDGMENT

The researchers would like to acknowledge Bukidnon State University for the scholarship grant and the faculty and staff of the Technological Institute of the Philippines-Quezon City for their invaluable support and guidance throughout the duration of this study.

REFERENCES

[1] Goodfellow I., Jean, PA., Mehdi, M., Bing, X., David, WF., Sherjil, O., Aaron, C. and Yoshua, B.: Generative adversarial networks., Commun. ACM, 63(11), pp. 139-144 (2020).

[2] J. Jenkins and K. Roy, "Exploring deep convolutional generative adversarial networks (DCGAN) in biometric systems: a survey study," *Discover Artificial Intelligence*, vol. 4, no. 1, May 2024, doi: https://doi.org/10.1007/s44163-024-00138-z

[3] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434. 2015.

[4] Donahue, C., McAuley, J., & Puckette, M. (2018). Adversarial Audio Synthesis. *ArXiv*. /abs/1802.04208

[5] Aditi Govindu, P. Kale, Aamir Hullur, Atharva Gurav, and Parth Godse, "Deepfake audio detection and justification with Explainable Artificial Intelligence (XAI)," *Research Square (Research Square)*, Oct. 2023, doi: https://doi.org/10.21203/rs.3.rs-3444277/v1

[6] L. Passos *et al.*, "A Review of Deep Learning-based Approaches for Deepfake Content Detection," *arxiv.org*. Available: https://arxiv.org/html/2202.06095v3

[7] Preeti, M. Kumar, and H. K. Sharma, "A GAN-Based Model of Deepfake Detection in Social Media," *Procedia Computer Science*, vol. 218, pp. 2153–2162, Jan. 2023, doi: https://doi.org/10.1016/j.procs.2023.01.191. Available: https://www.sciencedirect.com/science/article/pii/S1877050923001916

[8] M. Anas Raza and K. Mahmood Malik, "Multimodaltrace: Deepfake Detection using Audiovisual Representation Learning," *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Vancouver, BC, Canada, 2023, pp. 993-1000, doi: 10.1109/CVPRW59228.2023.00106.

[9] Z. Lv, S. Zhang, K. Tang and P. Hu, "Fake Audio Detection Based On Unsupervised Pretraining Models," *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, Singapore, 2022, pp. 9231-9235, doi: 10.1109/ICASSP43922.2022.9747605.

[10] Wang, R., Huang, Y., Guo, Q., Xie, X., Ma, L., & Liu, Y. (2020). DeepSonar: Towards Effective and Robust Detection of AI-Synthesized Fake Voices. *ArXiv*. /abs/2005.13770

[11] Wenger, E., Bronckers, M., Cianfarani, C., Cryan, J., Sha, A., Zheng, H., & Zhao, B. Y. (2021). "Hello, It's Me": Deep Learning-based Speech Synthesis Attacks in the Real World. *ArXiv*. https://doi.org/10.1145/3460120.3484742

[12] Karla Schäfer, Jeong-Eun Choi, and Sascha Zmudzinski. 2024. Explore the World of Audio Deepfakes: A Guide to Detection Techniques for Non-Experts. In *3rd ACM International Workshop on Multimedia AI against Disinformation (MAD '24), June 10--14, 2024, Phuket, Thailand.* ACM, New York, NY, USA 10 Pages. https://doi.org/10.1145/3643491.3660289

[13] Frank, J., & Schönherr, L. (2021). WaveFake: A Data Set to Facilitate Audio Deepfake Detection. *ArXiv*. /abs/2111.02813

[14] Chakravarty, N., & Dua, M. (2024). A lightweight feature extraction technique for deepfake audio detection. *Multimedia Tools and Applications, 83*(3). https://doi.org/10.1007/s11042-024-18217-9

[15] Wei Gao, Hui Jin, Gengjie Yang (2022), Series arc fault diagnosis method of photovoltaic arrays based on GASF and improved DCGAN, Advanced Engineering Informatics, Volume 54, 101809, ISSN 1474-0346, https://doi.org/10.1016/j.aei.2022.101809.

[16] Haibo Tan, Yating Hu, Benxue Ma, Guowei Yu, Yujie Li, (2024). An improved DCGAN model: Data augmentation of hyperspectral image for identification pesticide residues of Hami melon, Food Control, Volume 157, 110168, ISSN 0956-7135, https://doi.org/10.1016/j.foodcont.2023.110168.

[17] Hongyu Zhong, Samson Yu, Hieu Trinh, Yong Lv, Rui Yuan, Yanan Wang, (2023). Fine-tuning transfer learning based on DCGAN integrated with self-attention and spectral normalization for bearing fault diagnosis, Measurement, Volume 210, 112421, ISSN 0263-2241, https://doi.org/10.1016/j.measurement.2022.112421.

[18] Sreedev Devakumar, Greeshma Sarath (2023), Forensic Sketch to Real Image Using DCGAN, Procedia Computer Science, Volume 218, Pages 1612-1620, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2023.01.139.

[19] Zhenhao Wang, Chong Wang, Long Cheng, Guoqing Li (2022), An approach for day-ahead interval forecasting of photovoltaic power: A novel DCGAN and LSTM based quantile regression modeling method, Energy Reports, Volume 8, Pages 14020-14033, ISSN 2352-4847, https://doi.org/10.1016/j.egyr.2022.10.309.

[20] Zhenhao Wang, Chong Wang, Long Cheng, Guoqing Li (2022), An approach for day-ahead interval forecasting of photovoltaic power: A novel DCGAN and LSTM based quantile regression modeling method, Energy Reports, Volume 8, Pages 14020-14033, ISSN 2352-4847, https://doi.org/10.1016/j.egyr.2022.10.309.

[21] Mengfei Chen, Weihong "Grace" Guo (2023), DCGAN-CNN with physical constraints for porosity prediction in laser metal deposition with unbalanced data, Manufacturing Letters, Volume 35, Supplement, Pages 1146-1154, ISSN 2213-8463, https://doi.org/10.1016/j.mfglet.2023.08.116.

[22] B. Sandhiya, S. Kanaga Suba Raja, K. Shruthi, S. Praveena Rachel Kamala (2024), Brain tumour segmentation and classification with reconstructed MRI using DCGAN, Biomedical Signal Processing and Control, Volume 92, 106005, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2024.106005.

[23] Gao, Yang (2023). Audio Deepfake Detection Based on Differences in Human and Machine Generated Speech. Carnegie Mellon University. Thesis. https://doi.org/10.1184/R1/21842454.v1

[24] Deep Voice - Deepfake Voice Recognition [Dataset] [Online]. Available: Kaggle. https://www.kaggle.com/datasets/birdy654/deep-voice-deepfake-voice-recognition

[25] B. Dolhansky *et al.*, "The DeepFake Detection Challenge Dataset," *arXiv (Cornell University)*, Jun. 2020.

[26] J. C. Dheeraj, K. Nandakumar, A. V. Aditya, B. S. Chethan, and G. C. R. Kartheek, "Detecting Deepfakes Using Deep Learning," *2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*, Aug. 2021, doi: https://doi.org/10.1109/rteict52294.2021.9573740

[27] A. Jain, Pavel Korshunov, and S. Marcel, "Improving Generalization of Deepfake Detection by Training for Attribution," Oct. 2021, doi: https://doi.org/10.1109/mmsp53017.2021.9733468

[28] D. Gong, "Deepfake Forensics, an AI-synthesized Detection with Deep Convolutional Generative Adversarial Networks," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 3, pp. 2861–2870, Jun. 2020, doi: https://doi.org/10.30534/ijatcse/2020/58932020

[29] Hengnian Qi, Zihong Huang, Baichuan Jin, Qizhe Tang, Liangquan Jia, Guangwu Zhao, Dongdong Cao, Zeyu Sun, Chu Zhang (2024), SAM-GAN: An improved DCGAN for rice seed viability determination using near-infrared hyperspectral imaging, Computers and Electronics in Agriculture, Volume 216, 108473, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2023.108473.

[30] Noboru Babaguchi *et al.*, "Preventing Fake Information Generation Against Media Clone Attacks," *IEICE Transactions on Information and Systems*, vol. E104.D, no. 1, pp. 2–11, Jan. 2021, doi: https://doi.org/10.1587/transinf.2020mui0001

[31] Sudarshana Kerenalli, Vamsidhar Yendapalli, and M.C. Chinnaiah, "Classification of Deepfake Images Using a Novel Explanatory Hybrid Model," *Commit Journal*, vol. 17, no. 2, pp. 151–168, Sep. 2023, doi: https://doi.org/10.21512/commit.v17i2.8761

[32] D. Feng, Kuang Rensheng, Z. Yue, S. Long, Zhu Xiaogang, and Zhu Guopu, "A survey of Deepfake and related digital forensics," *Zhongguo tuxiang tuxing xuebao*, vol. 29, no. 2, pp. 295–317, Jan. 2024, doi: https://doi.org/10.11834/jig.230088

[33] Došilović, F.K., Brčić, M., and Hlupić, N.: Explainable artificial intelligence: A survey. In: Proc. Of 41st International convention on information and communication technology, electronics, and microelectronics (MIPRO), Opatija, Croatia, pp. 0210-0215 (2018).

[34] Bird, J. J., & Lotfi, A. (2023). Real-time Detection of AI-Generated Speech for DeepFake Voice Conversion. arXiv:2308.12734 [cs.SD]. https://doi.org/10.48550/arXiv.2308.12734