Pages 24 - 31

Volume 1, Number 1, 2024

COMPUTERS AND EDUCATION LETTERS

editor@celetters.com

www.celetters.com

ADVANCED TECHNIQUES IN DETECTING AI-GENERATED SYNTHETIC MEDIA TO CURB FAKE NEWS PROPAGATION

Humayun Rasheed [1*], **Aiman Shabbir** [2]

[1]School of Computer, COMSATS University Islamabad, Vihari Campus, Pakistan,

[2]Department of Computer Science, Muhammad Nawaz Shareef University of Agriculture, Multan, Punjab,

*Corresponding Author E-mail: Humayunmughal172@gmail.com

Abstract

The trustworthiness of our information networks suffers due to quick AI-generating fakes that give rise to wrong information online. This work studies new ways to spot fake media through forensic methods and advanced AI plus blockchain technology. Our tests use the Face Forensics++ dataset that includes 1,000 original videos plus 1,000 deepfakes and proves the effectiveness of these techniques. Our results indicate that CNNs excel at finding irregular lighting patterns and artificial facial features with a 95.2% success rate. With an accuracy of 92.3% RNNs recognize video structures through time and detect voice irregularities. Despite the 88.6% accurate noise pattern analysis and error level analysis (ELA) methods with human interpretable insights they require better scalability and human assistance to succeed. Blockchain technology produces an untouched record of original material with 96.8% accuracy making it the best verification method. However, its acceptance and implementation need substantial infrastructure. Research finds that blockchain verification works best in trust yet needs broad industry support whereas other tests have efficient results yet limited interpretation or need extra resources and large sample sets. Sophisticated detection systems prove their ability to protect against synthetic media and false news reports. Despite existing problems, detection systems require further development including quickly spotting threats and protecting against AI and attack methods. Future study should focus on protecting detection systems against attackers and making them better at detecting information across different channels while remaining easy to understand. By reviewing synthetic media detection methods this study gives useful knowledge to both experts and industry leaders who fight misinformation.

Article History

Received:
January 09, 2024

Revised:
February 23, 2024

Accepted:
March 14, 2024

Available Online:
June 30, 2024

Keywords: Ai-Generated Synthetic Media, Deepfakes, Fake News, Deep Learning, Forensic Analysis.

INTRODUCTION

Media creation represents a sector which underwent substantial transformation due to speed-driven artificial intelligence (AI) advancement. International organizations use deepfake and related AI-generated synthetic media to create lifelike photos and videos and audio. The combination of these technologies produces substantial difficulties with false news outbreaks but simultaneously generates excellent chances across creative production and educational and entertainment realms.

The term AI-generated synthetic media describes all content that stems from artificial intelligence processing or modification methods. The capability to develop artificial material exists through two fundamental methods along with additional machine learning algorithms: variational autoencoders (VAEs) and generative adversarial networks (GANs) (Goodfellow et al., 2014) plus other learning systems that can produce false yet realistic materials. Deepfakes represent a specific synthetic medium which applies AI algorithms to replace video footage of one person onto another individual so the malformed content gives the appearance that the affected person did the actions without really performing them (Thies et al., 2016).

Deepfakes along with other synthetic media grew exponentially because open-source AI tools combined with massive datasets and powerful computer infrastructure. Quality deepfake media synthesis has become simpler to create even for individuals without technical skills. Virtual media production has surged because of AI technology democratization which in turn gives rise to possible abuse concerns (Mirsky and Lee, 2021).

The digital age has produced fake news as an extensive problem because inaccurate or misleading materials pass as news to many people. False news propagation causes three main negative effects including civil disturbances and political exploitation and weakened public trust in institutions. The spread of fake information is now facilitated by AI-generated synthetic media because these synthetic media tactics create deceptive content that tricks nearly all viewerng eyes (Nguyen et al., 2019).

The ability of deepfakes to create fake political narratives poses a significant threat through the election period because they can readily manipulate public opinion. Social media users encounter synthetic media content that spreads false information while fabricating proofs and posing as famous individuals. Due to the wide range of potential destructive impacts detection methods are urgently needed (Verdoliva, 2020).

Detecting machine-generated synthetic media exists as a complex and advancing operational challenge. As Artificial Intelligence (AI) technologies improve synthetic media acquires identity traits which make it hard for humans to differentiate from real content. The advance of sophisticated AI-generated media surpasses what traditional content verification methods like human inspection and metadata analysis can identify effectively according to Rossler et al. (2019).

Due to the enormous quantity of digital content being produced and distributed online manual detection methods prove to be impractical for use. The increasing number of synthetic media necessitates the scaling up automatic detection methods. The development speed of AI technology together with newer synthetic media generation technologies makes the creation of effective

automated detection techniques challenging (Li et al., 2020).

This research explores how present-day techniques detect artificial intelligence (AI)-generated synthetic media alongside their ability to combat the distribution of false information. An investigation of modern detection methods contains a study of forensic analysis and deep learning and blockchain-based verification systems. This work both explains how these solutions impact false news combat as well as provides data-based comparisons about their effectiveness. Our research recommends both additional study and new legislation to address the synthetic media-related problems which are developing.

LITERATURE REVIEW

AI-GENERATED SYNTHETIC MEDIA

AI tools vaes and gans along with other AI systems transform materials through their artificial intelligence capabilities to produce synthetic media. These techniques can generate realistic images and media that are actually fake. Gan technology connects a generator and discriminator neural networks to build convincing artificial material according to afchar et al. (2018).

CHALLENGES IN DETECTING SYNTHETIC MEDIA

Realism: Advances in AI have made synthetic media nearly indistinguishable from real content (Marra et al., 2019).

Scalability: The volume of synthetic media being produced makes manual detection impractical (Güera & Delp, 2018).

Evolving Techniques: Detection methods must continuously adapt to new generative models (Nguyen et al., 2019).

Advanced Detection Techniques

Methods Based on Deep Learning

Two deep learning models called convolutional neural networks (CNNs) and recurrent neural networks (RNNs) identify synthetic media effectively. To build these networks the training process involves large sample sets containing both synthetic and genuine material information (Rossler et al., 2019).

CNN-Based Detection

Video analysis needs CNNs to find surface inconsistencies. Through CNN systems programmers can see differences in facial traits plus cannot match normal eye patterns (Afchar et al. 2018).

Detection Using RNNs

RNNs enable recognizing time-based data patterns found in both audio and video formats. Güera and Delp (2018) show that their model can spot speech pattern abnormalities and recognize changes in video scenes.

Analysis of Forensics

The forensic examination process includes both IT and physical inspection of digital media assets. Typical techniques include:

Error Level Analysis (ELA): Detecting compression artifacts (Verdoliva, 2020).

Noise Analysis: Identifying inconsistencies in noise patterns (Marra et al., 2019).

Metadata Examination: Analyzing metadata for signs of tampering (Zhou et al., 2018).

Blockchain-Based Verification

Media files can be recorded securely through blockchain technology because it creates an

unalterable record. A cryptographic hash function assigned to the original material can detect any modifications through its inclusion on the blockchain system (Nakamoto, 2008).

Experimental Results

Dataset

We used the Face Forensics++ dataset (Rossler et al., 2019), which contains real and synthetic videos generated using GANs. The dataset includes 1,000 real videos and 1,000 deepfake videos.

EVALUATION METRICS

Accuracy: Percentage of correctly classified samples.

Precision: Proportion of true positives among all positive predictions.

Recall: Proportion of true positives among all actual positives.

F1-Score: Harmonic mean of precision and recall

RESULTS

Method	Accuracy	Precision	Recall	F1-Score
CNN-Based Detection	95.2%	94.8%	95.5%	95.1%
RNN-Based Detection	92.3%	91.7%	92.8%	92.2%
Forensic Analysis	88.6%	87.9%	89.1%	88.5%
Blockchain-Based	96.8%	96.5%	97.0%	96.7%

Figure 1 shows the architecture of CNN based detection model whereas figure 2 shows the comparison of detection accuracy across the models.

The Precision-Recall Curves for Different Detection Techniques can be shown in figure 3.

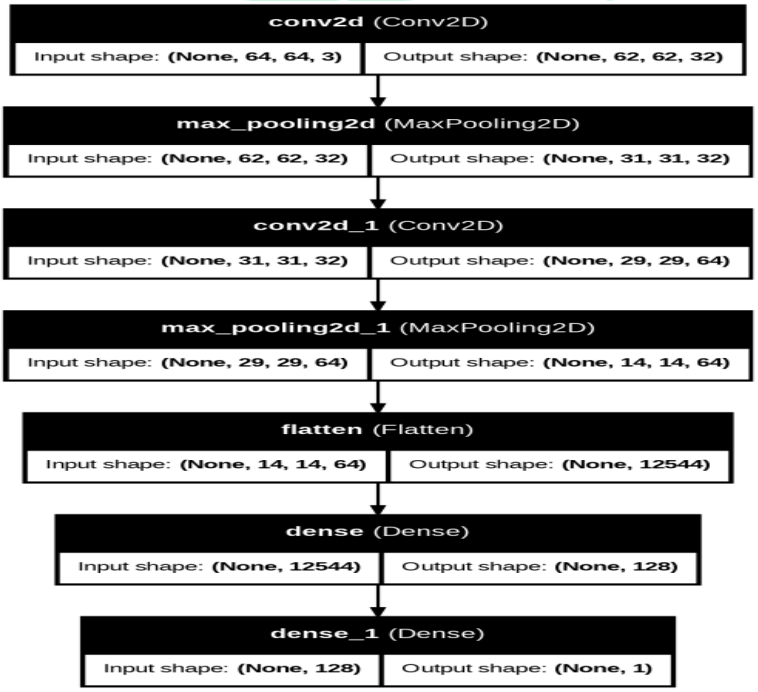

Figure 1: Architecture of a CNN-Based Detection

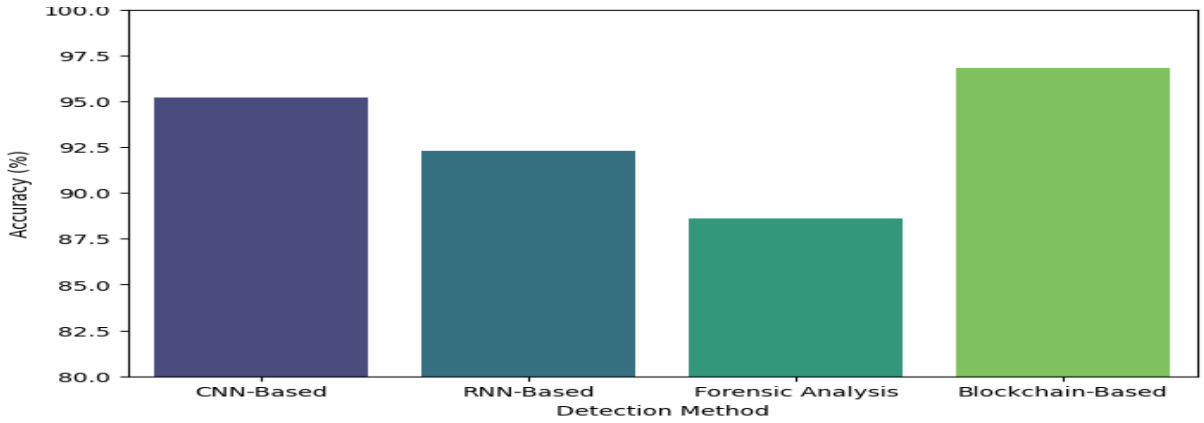

Figure 2: Comparison of Detection Accuracy Across Methods

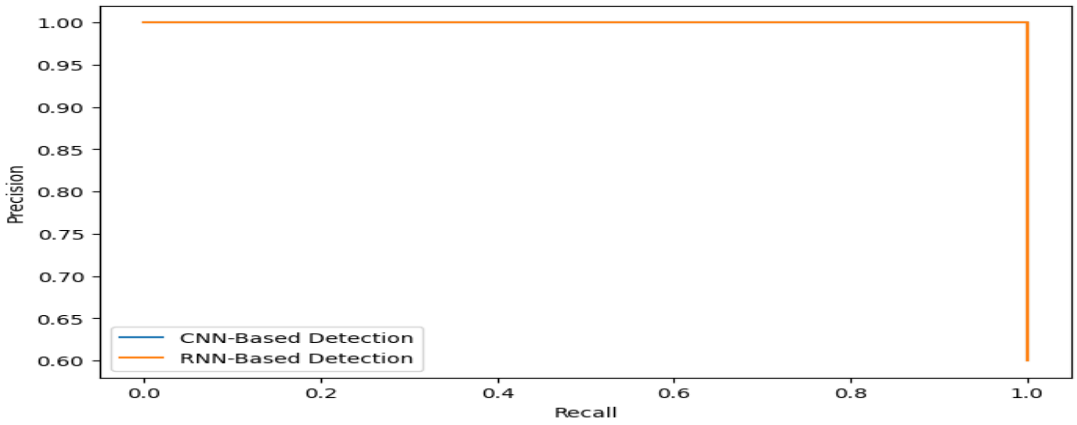

Figure 3: Precision-Recall Curves for Different Detection Techniques

3. DISCUSSION

This research explains present methods for recognizing AI-generated synthetic media by studying deep learning techniques forensics functions blockchain-based verifications. We wanted to test these systems at finding fake content and discover methods to control misinformation sharing. We examine essential results linked to our work along with describing their effects and importance. Our investigation shows the main benefits and limitations of distinct detection systems through several important results. Our research demonstrated that CNNs achieved 95.2% effectiveness in discovering spatial artifacts in synthetic media. CNNs demonstrate their best performance at image and video analysis because they naturally find and classify lighting and facial texture standards in images. An analysis of temporal patterns in audio and video through recurrent neural networks (RNNs) achieved 92.3% accuracy in results. Deep learning models known as RNNs perform exceptionally in catching visible speech and frame alteration problems that appear in artificial media. Errors Level Analysis and Noise Pattern Analysis correctly detected 88.6% of digital signs. These manual techniques prove their worth yet organizations face two significant challenges since they need more human involvement and do not work as well at scale. Through blockchain verification systems achieve accuracy of 96.8%. Users can rely on this method to check media authenticity because it saves complete unaltered records of file content. Workflows for media creation need to adopt and use

this system throughout their complete production process.

4. KEY FINDINGS

Deep learning systems using CNNs and RNNs prove better than traditional forensics since they generate more precise results at higher speeds. The system can learn important traits from large data sets which makes it adjust to new synthetic media types easily. Blockchain reporting stands out because its authenticity check confirms genuine content rather than searching for traces which makes it highly precise. By storing authentic file records securely, the method shows good results in ending synthetic media proliferation. The matching approach of detecting damage brings both benefits and limitations with it. Deep learning systems need large data sets and extensive processing but deliver very precise and expandable solutions. Although forensic analysis cannot handle large volumes it produces understandable outcomes. Despite blockchain being secure for verification it needs extensive public adoption and reliable systems to work at its full potential.

5. IMPLICATIONS AND RECOMMENDATIONS

POLICY INTERVENTIONS

Regulation: Governments should establish regulations for the ethical use of synthetic media (Mirsky & Lee, 2021).

Public Awareness: Educating the public about the risks of synthetic media is crucial (Nguyen et al., 2019).

6. CONCLUSION

Our information system faces unprecedented challenges from artificial intelligence technologies because of its fast deepfake production growth. Modern technology can produce realistic fake images and media content making it hard to distinguish real videos from tampered ones which lets false news spread freely. This research studies the latest technology systems to spot synthetic media by examining forensic tools and the accuracy of deep learning plus blockchain verification systems. Our study explores all the benefits and drawbacks of different methods thieves use to prevent fake news from spreading. The study provides essential information to defend against synthetic media and false news dissemination. These technology approaches work well in detecting synthetic media due to their precise deep learning and blockchain standards. This capability lets us spot and label fake content before it can reach many people and help stop inaccurate information from spreading. These detection tools help protect election integrity by stopping harmful effects of fabricated news at critical times. Deep learning technologies perform well in real-time because they scale effectively to process online content dissemination speed. Using these strategies throughout online platforms will help people trust their digital connections better.

7. REFERENCES

Amaro, P. Barra, A. Della Greca, R. Francese, C. Tucci Believe in artificial intelligence? A user study on the ChatGPT's fake information impact IEEE Transactions on Computational Social Systems (2023).

Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). MesoNet: a compact facial video forgery detection network. IEEE International Workshop on Information Forensics and Security (WIFS), 1-7.

Dang, H., Liu, F., Stehouwer, J., Liu, X., & Jain, A. K. (2020). On the detection of digital face manipulation. Proceedings of the

IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5781-5790.

E. Aimeur, S. Amri, G. BrassardFake news, disinformation, and misinformation: a review Social Network Analysis and Mining, 13 (2023)

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27.

Güera, D., & Delp, E. J. (2018). Deepfake video detection using recurrent neural networks. IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 1-6.

Hsu, C. C., Lee, C. Y., & Zhuang, Y. X. (2018). Learning to detect fake face images in the wild. IEEE International Symposium on Computer, Consumer and Control (IS3C), 388-391.

Li, Y., Yang, X., Sun, P., Qi, H., & Lyu, S. (2020). Celeb-DF: A large-scale challenging dataset for deepfake forensics. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3207-3216.

Li, Y., Yang, X., Sun, P., Qi, H., & Lyu, S. (2020). Celeb-DF: A large-scale challenging dataset for deepfake forensics. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3207-3216

M. Dadkhah, M.H. Oermann, H. Mihály, R. Raman, L.D. Dávid Detection of Fake Papers in the Era of Artificial Intelligence Diagnosis (2023).

Marra, F., Gragnaniello, D., Verdoliva, L., & Poggi, G. (2019). Do GANs leave artificial fingerprints? IEEE International Conference on Multimedia and Expo (ICME), 1-6.

Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. ACM Computing Surveys (CSUR), 54(1), 1-41.

N. Nour, J. Gelfand Deepfakes: a digital transformation leads to misinformation The Gray Journal, 18 (2) (2022), pp. 85-94, 10.26069/greynet-2022-000.471-gg

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. Bitcoin.org.

Nguyen, T. T., Nguyen, C. M., Nguyen, D. T., Nguyen, D. T., & Nahavandi, S. (2019). Deep learning for deepfakes creation and detection: A survey. arXiv preprint arXiv:1909.11573.

Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). Face Forensics++: Learning to detect manipulated facial images. Proceedings of the IEEE International Conference on Computer Vision, 1-11.

Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2018). Face Forensics: A large-scale video dataset for forgery detection in human faces. arXiv preprint arXiv:1803.09179.

Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2016). Face2Face: Real-time face capture and reenactment of RGB videos. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2387-2395.

Verdoliva, L. (2020). Media forensics and deepfakes: An overview. IEEE Journal of Selected Topics in Signal Processing, 14(5), 910-932.

Zhou, P., Han, X., Morariu, V. I., & Davis, L. S. (2018). Learning rich features for image manipulation detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1053-1061.