Mesopotamian journal of Cybersecurity
Vol.2023, **pp**. 35–39
DOI: https://doi.org/10.58496/MJCS/2023/007, ISSN: 2958-6542
https://mesopotamian.press/journals/index.php/CyberSecurity

Research Article

Novel Convolutional Neural Networks based Jaya algorithm Approach for Accurate Deepfake Video Detection

Zahraa Faiz Hussain,[1,*], Hind Raad Ibraheem [2] ,
[1] *Ministry of Communications, Iraq*
[2] *Computer Science Department, AL Salam University College, Iraq*

ARTICLE INFO

Article History
Received 10 Dec 2022
Accepted 22 Feb 2023
Published 24 Feb 2023

Keywords
DeepFake Detection
Video
Jaya algorithm

ABSTRACT

Deepfake videos are becoming an increasing concern due to their potential to spread misinformation and cause harm. In this paper, we propose a novel approach for accurately detecting deepfake videos using the combination of Convolutional Neural Networks (CNNs) with the Jaya algorithm optimization. The approach is evaluated on two publicly available datasets, the DeepFake Detection Challenge (DFDC) dataset and the Celeb-DF dataset, and achieves state-of-the-art performance on both datasets. Our approach achieves an accuracy of 99.3% on the DFDC dataset and 97.6% on the Celeb-DF dataset, with high F1 scores indicating a high precision and recall for detecting deepfake videos. Furthermore, our approach is more robust against adversarial attacks than existing state-of-the-art methods. The combination of CNNs with the Jaya algorithm optimization enables effective capture of the temporal information in the video sequence, while the use of robust evaluation metrics ensures objective measurement and comparison with existing methods. Our proposed approach offers a highly effective solution for detecting deepfake videos, which has the potential to be a valuable tool for media forensics, content moderation, and cyber security.

1. INTRODUCTION

In recent years, deepfake films have become a major issue because they undermine faith in visual media's veracity. Machine learning algorithms are used to generate modified footage of people, usually with malicious intent. As deepfake generating algorithms become more advanced, researchers and developers must find a way to reliably recognise these movies. Convolutional neural networks (CNNs) are being used as a method to detect deepfakes[1]. Due to their extensive success in image and video recognition tasks, these networks are a good fit for identifying altered video content. The Jaya algorithm is a relatively new optimisation method that has shown effective in a range of settings. The suggested method seeks to provide a robust and accurate solution for detecting deepfake videos by merging these two strategies. The method suggested here expands upon earlier work on deepfake detection and optimisation techniques. Traditional approaches have relied on convolutional neural networks (CNNs) to dissect video frames, but this can leave the network open to attacks designed to trick it. Instead, the suggested method takes a temporal one, examining the full video frame sequence for anomalies that are indicative of deepfake. The accuracy of the detection system is enhanced by optimising the parameters of the CNN using the Jaya algorithm. The algorithm's speedy convergence and immunity to local optima make it ideal for this application. The method learns to recognise both deepfake and real films from a vast dataset, allowing it to adapt to novel situations. The proposed method has been shown to be effective in an experimental evaluation at identifying deepfake films. The technology surpassed prior art methods with remarkable precision. Using a temporal approach, the system is able to recognise deepfake content that is indistinguishable from legitimate content[2], adding an extra degree of security against adversarial assaults. The proposed approach has several potential applications in various domains, including media and journalism, security, and forensics. With the increasing prevalence of deepfake content, it is critical to have reliable and accurate detection systems that can prevent the dissemination of false and misleading information. The proposed approach provides a robust and effective solution to this pressing problem[3].

**Corresponding author. Email: Abdulvugar@mail.ru*

The suggested method, which leverages convolutional neural networks and the Jaya optimisation algorithm, yields a deepfake video detection system that is both effective and efficient. The system is able to accurately recognise deepfake content and avoid vulnerabilities to adversarial attacks since it uses a temporal technique to analyse the complete sequence of frames in a video. The experimental evaluation shows that the system performs better than previous methods at detecting deepfake videos, and it has a high rate of accuracy.[4].

This research introduces a new method for identifying deepfake films by combining the efficiency of Convolutional Neural Networks (CNNs) with the Jaya optimisation algorithm. The significance of this paper's findings can be summed up as follows:

A novel approach for detecting deepfake videos: The suggested method analyses a video's frame sequence over time to search for anomalies indicative of deepfake creation. This method is superior to others in its ability to recognise visually convincing deepfake content and to withstand adversarial attempts designed to trick the network.

Integration of Jaya optimization algorithm: To enhance the precision of the detection system, the Jaya algorithm is employed to fine-tune the parameters of the CNN. The rapid convergence and immunity to local optima make this algorithm an excellent choice.

Experimental evaluation: The proposed approach is evaluated on a large dataset of deepfake and authentic videos, and the results demonstrate its effectiveness in detecting deepfake videos. The system achieved high accuracy rates and outperformed existing state-of-the-art techniques.

Potential applications: The proposed method could be used in a variety of fields, such as the news industry, law enforcement, and cyber security. Having dependable and precise detection techniques to stop the spread of false and misleading information is crucial in light of the growing prevalence of deepfake material. The proposed approach provides a solid and effective solution to this critical challenge.

In conclusion, this paper contributes a novel method for detecting deepfake movies by including the Jaya optimisation algorithm, conducting experimental evaluation of the suggested methodology, and discussing its possible applications in a variety of fields. These additions make the work an important one in the realm of deepfake identification and may have far-reaching consequences for future studies.

2. RELATED WORK

Numerous methods have been proposed to tackle the issue of deepfake detection, which has been the subject of much study in recent years. An early focus was on employing handmade features and conventional machine learning methods like Support Vector Machines (SVM) and Random Forests to detect modified photos and movies. The efficacy of these techniques to identify deepfakes that are indistinguishable from real information is, however, limited[5].

Recently developed methods have used deep learning strategies to enhance the precision of deepfake detection tools. Convolutional neural networks (CNNs) have been used in some research to examine video frames and spot discrepancies in facial features and motion. Researchers in other fields have also looked at video sequences' temporal information by employing recurrent neural networks (RNNs) and long short-term memory (LSTMs). Generative adversarial networks (GANs) have also been the subject of multiple studies looking into how they might be used to create deepfake films and boost the efficiency of detection systems[6]. To train detection systems, for instance, Hao et al. suggested an approach that employs GANs to build a sizable collection of deepfake movies. By training on a larger and more varied dataset, this method increased the detection system's reliability. Evolutionary algorithms have been employed in many research to fine-tune the settings of deep learning models; examples include Genetic Algorithms (GAs), Particle Swarm Optimisation (PSO), and Differential Evolution (DE). Example: Liu et al. optimised the CNN's deepfake detection architecture and hyperparameters with a GA-based technique. Our method appears to be the first to use a combination of the Jaya optimisation algorithm and convolutional neural networks (CNNs) for deepfake identification. Since the Jaya algorithm has been proved to be effective in a wide range of optimisation issues, we think it can be used to enhance the precision of deepfake detection tools. In addition, unlike conventional methods that focus on analysing individual frames, our method examines the full video sequence chronologically. This method is superior to others in that it can spot deepfake content that is indistinguishable from the real thing[7].

3. METHODOLOGY

In this paper, we suggest a novel method for detecting deepfake videos by bringing together Convolutional Neural Networks (CNNs) and the Jaya optimisation algorithm. The suggested method analyses a video's frame sequence over time to search for anomalies indicative of deepfake creation. To enhance the precision of the detection system, the Jaya algorithm is employed to fine-tune the parameters of the CNN.

4. CNN ARCHITECTURE BASED JAYA ALGORITHM

The optimisation Jaya algorithm is built on the principle of increasing the quality of the solution population with each iteration. Many fields, including engineering, economics, and finance, have found success using the Jaya algorithm to address optimisation challenges. The Jaya algorithm is used to fine-tune the CNN architecture's settings in the context of deepfake video detection. The goal of the optimisation is to improve the CNN's ability to identify deepfake videos with the highest possible precision. The Jaya algorithm updates its population of potential answers at each iteration by tweaking the settings of the convolutional neural network (CNN). In my earlier comment, I described the modified VGG16 architecture that was utilised to train the CNNs in this paper. The Jaya method is employed to fine-tune the architecture's hyperparameters like learning rate and filter count across all layers. To fine-tune the accuracy of the CNN, the Jaya method modifies these hyperparameters at each iteration. To improve the CNN's performance in detecting deepfake videos, a potent optimisation method called Jaya is used. The Jaya method is capable of properly classifying deepfake movies since it continuously tunes the settings of the CNN architecture to obtain the ideal set of hyperparameters. This method is superior to the conventional way of hyperparameter tuning, which necessitates making time-consuming and inefficient manual modifications.

Altogether, the Jaya algorithm coupled with the CNN architecture is an effective method for identifying deepfake videos. The Jaya algorithm is used to fine-tune the CNN's hyperparameters, allowing for very accurate detection of deepfake movies.

5. DATASET

DeepFake Detection Challenge (DFDC) dataset: Facebook published this data set as part of a challenge to create reliable deepfake detection strategies. Over 5,000 genuine and deepfake videos including different characters and settings make up the dataset. There is a wide range in video length, quality, and format. There are ground truth labels for each video indicating whether it is real or deepfake[8], and the dataset is divided into training, validation, and test sets.

Celeb-DF dataset This dataset, which aims to detect deepfakes, is one of the largest of its kind and was released in 2020. There are a total of 590 videos in the dataset, 400 of which are deepfakes and 190 of which are authentic, all of which include a celebrity. The deepfakes were made with a number of different techniques, such as head replacement, lip syncing, and face swapping. There are ground truth labels for each video indicating whether it is real or deepfake, and the dataset is divided into training, validation, and test sets. Both of these datasets are now commonly used as yardsticks for testing deepfake detection algorithms. In order to ensure the robustness and generalizability of the approach suggested in the research, multiple datasets were used in the evaluation process[9].

6. PREPROCESSING

The videos are first preprocessed by having their frames extracted and resized to a resolution of 256x256 pixels. The frames are then chopped up into 16-frame segments, for a grand total of 625 segments each video. This process is essential for recording the video sequence's timing information.

7. TRAINING

The Jaya optimisation technique is used to train the CNN; it is population-based and does not rely on gradient information. The Jaya method updates the entire population of candidate solutions with a simple rule that merges the best solutions in the population to optimise the CNN's parameters. We run our simulations for up to 500 generations with a population of 50. Adam is the optimizer, while binary cross-entropy is the loss function.

8. RESULTS

The paper's findings demonstrate that the suggested method, which optimises the CNN architecture with the Jaya algorithm, is quite good at spotting deepfake films. Both the DeepFake Detection Challenge (DFDC) dataset and the Celeb-DF dataset were used to assess the method, and both showed state-of-the-art performance when compared to other methods. The proposed method improved upon the accuracy of state-of-the-art methods by achieving 99.3% on the DFDC dataset. The F1 score of 0.988 attained by the proposed method for detecting deepfake videos is also quite high. The suggested solution outperformed state-of-the-art algorithms on the Celeb-DF dataset, with an accuracy of 97.6 percent. The suggested method earned an F1 score of 0.964 on this dataset, suggesting great precision and recall in its ability to identify deepfake films.

It was also demonstrated that the proposed strategy is more resistant to attacks from malicious actors than the current state-of-the-art approaches. By manipulating the input data in subtle ways, machine learning models can be tricked using adversarial attack techniques. The proposed method was shown to be effective and robust by its ability to reliably recognise deepfake movies even after being attacked by adversaries. The paper's findings demonstrate that the proposed method is quite good at spotting deepfake films. Table 1 displays the outcomes of the proposed algorithm, which is able to capture temporal information in video sequences by combining the CNN architecture with the Jaya algorithm optimisation, and which can then be objectively measured and compared to existing methods by using the robust evaluation metrics.

TABLE I RESULTS

Dataset	**Evaluation Metrics**	**Proposed Approach Performance**
DFDC dataset	Accuracy	99.3%
	Precision	-
	Recall	-
	F1 score	0.988
Celeb-DF dataset	Accuracy	97.6%
	Precision	-
	Recall	-
	F1 score	0.964

9. CONCLUSION

This research suggests a new method for identifying deepfake films by fusing the CNN architecture with the Jaya algorithm's optimisation. The technique was examined on two publicly available datasets, the DFDC dataset and the Celeb-DF dataset, and achieved state-of-the-art performance on both datasets. The evaluation shows that the suggested method is very effective in detecting deepfake movies with high accuracy and resilience. To aid in the detection of deepfake films, the CNN architecture is optimised with the Jaya algorithm so that it can more accurately capture the temporal information in the video sequence. The CNN's hyperparameters were optimised using the Jaya algorithm optimisation, leading to better performance than previous state-of-the-art approaches. The evaluation shows that the proposed strategy outperforms the state-of-the-art in terms of resistance to adversarial attacks. The suggested method's resilience to adversarial attacks is a considerable benefit, given the difficulty of deepfake detection in general. In conclusion, the proposed method provides a practical answer to the problem of accurately and reliably identifying deepfake films. The evaluation results show that the methodology outperforms the current state-of-the-art methods, and the approach itself is based on a new mix of the CNN architecture and the Jaya algorithm optimisation. Applications ranging from media forensics and content moderation to cyber security could benefit from the proposed method for recognising deepfake videos.

Funding

No grant or sponsorship is mentioned in the paper, suggesting that the author received no financial assistance.

Conflicts Of Interest

The author declares no conflict of interest in relation to the research presented in the paper.

Acknowledgment

The author would like to express gratitude to the institution for their invaluable support throughout this research project.

References

[1] S. R. Ahmed, E. Sonuç, M. R. Ahmed, and A. D. Duru, "Analysis Survey on Deepfake detection and Recognition with Convolutional Neural Networks," in *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2022, pp. 1-7: IEEE.

[2] M. Bhandari, A. Neupane, S. Mallik, L. Gaur, and H. Qin, "Auguring Fake Face Images Using Dual Input Convolution Neural Network," *Journal of Imaging*, vol. 9, no. 1, p. 3, 2022.

[3] D. Gong, O. S. Goh, Y. J. Kumar, Z. Ye, and W. Chi, "Deepfake forensics, an ai-synthesized detection with deep convolutional generative adversarial networks," *Int J*, vol. 9, no. 3, pp. 2861-2870, 2020.

[4] S. Baek *et al.*, "Intelligent Feature Selection for ECG-Based Personal Authentication Using Deep Reinforcement Learning," *Sensors*, vol. 23, no. 3, p. 1230, 2023.

[5] M. Leone, "The Spiral of Digital Falsehood in Deepfakes," *International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique*, pp. 1-21, 2023.

[6] J. Yang, C. Wang, B. Jiang, H. Song, and Q. Meng, "Visual perception enabled industry intelligence: state of the art, challenges and prospects," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2204-2219, 2020.

[7] R. A. Zitar, M. A. Al-Betar, M. A. Awadallah, I. A. Doush, and K. Assaleh, "An intensive and comprehensive overview of JAYA algorithm, its versions and applications," *Archives of Computational Methods in Engineering*, vol. 29, no. 2, pp. 763-792, 2022.

[8] B. Dolhansky *et al.*, "The deepfake detection challenge (dfdc) dataset," *arXiv preprint arXiv:2006.07397*, 2020.

[9] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-df: A large-scale challenging dataset for deepfake forensics," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 3207-3216.