electronics

MDPI

Article

An Approach to Deepfake Video Detection Based on ACO-PSO Features and Deep Learning

Hanan Saleh Alhaji [1], Yuksel Celik [2,*] and Sanjay Goel [2]

[1] Computer Engineering, Karabuk University, 78050 Karabuk, Turkey; 2038166088@ogrenci.karabuk.edu.tr
[2] Information Security and Digital Forensics, University at Albany, State University of New York, Albany, NY 12222, USA; goel@albany.edu
* Correspondence: ycelik@albany.edu; Tel.: +1-518-4664959

Abstract: The rapid advancement of deepfake technology presents significant challenges in detecting highly convincing fake videos, posing risks such as misinformation, identity theft, and privacy violations. In response, this paper proposes an innovative approach to deepfake video detection by integrating features derived from ant colony optimization–particle swarm optimization (ACO-PSO) and deep learning techniques. The proposed methodology leverages ACO-PSO features and deep learning models to enhance detection accuracy and robustness. Features from ACO-PSO are extracted from the spatial and temporal characteristics of video frames, capturing subtle patterns indicative of deepfake manipulation. These features are then used to train a deep learning classifier to automatically distinguish between authentic and deepfake videos. Extensive experiments using comparative datasets demonstrate the superiority of the proposed method in terms of detection accuracy, robustness to manipulation techniques, and generalization to unseen data. The computational efficiency of the approach is also analyzed, highlighting its practical feasibility for real-time applications. The findings revealed that the proposed method achieved an accuracy of 98.91% and an F1 score of 99.12%, indicating remarkable success in deepfake detection. The integration of ACO-PSO features and deep learning enables comprehensive analysis, bolstering precision and resilience in detecting deepfake content. This approach addresses the challenges involved in facial forgery detection and contributes to safeguarding digital media integrity amid misinformation and manipulation.

Keywords: deepfake detection; deep learning; ant colony optimization; particle swarm optimization

Citation: Alhaji, H.S.; Celik, Y.; Goel, S. An Approach to Deepfake Video Detection Based on ACO-PSO Features and Deep Learning. *Electronics* **2024**, *13*, 2398. https://doi.org/10.3390/electronics13122398

Academic Editor: Tahir Cetin Akinci

Received: 17 April 2024
Revised: 10 June 2024
Accepted: 12 June 2024
Published: 19 June 2024

1. Introduction

Deepfake videos, which leverage advanced deep learning techniques to manipulate facial expressions and actions in videos, have emerged as a significant concern in recent years [1]. The accessibility of the necessary technology has facilitated the creation and dissemination of deepfake content, posing serious threats to privacy, security, and trust in media [2]. Detecting deepfake videos has thus become a critical research area in combating misinformation and propaganda [3].

1.1. Motivations

There are rising concerns regarding the proliferation of deepfake videos and their potential impact on privacy, security, and trust in media. Thus, there is an urgent need to develop effective detection methods to identify and mitigate the spread of misinformation and propaganda. Various techniques, including facial expression analysis and machine learning algorithms, have been proposed for detecting deepfake videos [3–5]. However, inconsistencies in generated content and evolving manipulation techniques remain persistent challenges in this regard [6,7].

This paper proposes a novel approach to deepfake video detection, integrating features derived from ant colony optimization–particle swarm optimization (ACO-PSO) and deep

learning techniques. Our methodology leverages ACO-PSO features and deep learning models to enhance the detection accuracy and robustness [8]. Specifically, ACO-PSO features are extracted from the spatial and temporal characteristics of video frames, capturing subtle patterns indicative of deepfake manipulation [9].

1.2. Main Contributions

The main contributions of this paper are as follows:

- The integration of ACO-PSO features and deep learning for improved deepfake video detection.
- The utilization of transfer learning with pre-trained ACO-PSO models to address the limitations of the available training data.
- An emphasis on optical flow features to analyze temporal dynamics and enhance detection accuracy.
- A comprehensive analysis encompassing both the spatial and temporal dimensions of video data to strengthen deepfake detection capabilities.

Through extensive experiments using benchmark datasets, our proposed method demonstrates superior performance in terms of detection accuracy, robustness to manipulation techniques, and generalization to unseen data. Furthermore, the computational efficiency of our approach is analyzed, highlighting its practical feasibility for real-time applications.

2. Related Works

Deepfake video detection has recently become one of the most challenging strands of the scientific literature. Table 1 shows some deepfake video detection methods, as well as their advantages and disadvantages. This study, as outlined in [10], focuses on developing and evaluating two deep learning networks for automatic face manipulation detection in videos, targeting deepfake and Face2Face methods. Deep learning was chosen due to the limitations of traditional image forensics in videos, necessitating a low-layer approach to prioritize mesoscopic image properties. Networks are rigorously tested using established and newly collected datasets, revealing the significance of facial features, such as the eyes and mouth, in detecting deepfake manipulation. Vougioukas et al.'s [11] approach involved crafting a forensic methodology to tackle the growing danger posed by deepfake videos. Once confined to Hollywood and government circles, the widespread availability of deep learning tools has democratized their creation. Now, even individuals with limited resources can produce convincing fake videos that may have serious consequences. To counter this threat, the proposed method focuses on analyzing the facial expressions and movements unique to speech patterns in an attempt to authenticate videos and mitigate risks to democracy, security, and society. Yang et al. [12] introduced a temporal-aware pipeline for automated deepfake detection, utilizing a CNN to extract features and an RNN for manipulation classification. Despite its simplicity, the system performed competitively on diverse datasets, highlighting its efficacy in identifying manipulated content. This research underscores the methodology's robustness as an initial defense against fabricated media, emphasizing its potential significance in combating misinformation. Yi et al.'s [13] methodology focused on a new training strategy for generative adversarial networks (GANs), gradually expanding both the generator and discriminator networks to capture finer details as training progresses. This incremental approach accelerates training and enhances stability, enabling the creation of high-quality images, such a Large-scale CelebFaces Attributes (CelebA-HQ) images at a 1024^2 resolution. Additionally, the study introduced a simple method to increase diversity in generated images, achieving a remarkable inception score of 8.80 in the unsupervised Canadian Institute for Advanced Research, 10 classes (CIFAR-10) dataset. Mustak et al. [14] proposed a pioneering methodology, introducing an innovative generator architecture for generative adversarial networks that is inspired by insights from the style transfer literature. This novel architecture facilitates the automatic and unsupervised separation of high-level attributes, such as pose and identity

in human faces, and stochastic variations, such as freckles and hair, in the generated images, providing intuitive synthesis control across various scales. Furthermore, it enhances traditional distribution quality metrics, showcases improved interpolation properties, and adeptly disentangles latent factors of variation. The study introduced two novel automated methods to assess interpolation quality and disentanglement, along with presenting a new dataset of human faces distinguished by their high variation and quality, thereby enhancing the available resources for further exploration in this field. Traboulsi [15] focused on uncovering fake face videos generated by deep neural networks, utilizing eye-blinking detection as a key discriminator. This is crucial, as eye blinking is often absent in fake videos. The evaluation involves analyzing eye-blinking datasets, particularly for deepfake videos. Plans include exploring alternative neural network architectures for improved closed-eye detection and investigating dynamic blinking patterns for signs of tampering. Additionally, efforts will be made to identify other intrinsic physiological signals overlooked in AI synthesis to help develop more refined detection techniques. The approach outlined in [16] addressed the vulnerability of deep neural network (DNN)-based deepfake detection methods to adversarial techniques. With the advancement of video manipulation, the proliferation of fake videos complicates efforts to combat disinformation and maintain media integrity. This study delved into how current deepfake detectors can be evaded through adversarial modifications to synthetic fake videos, highlighting the resilience of these alterations against compression codecs. The research introduced two attack strategies, one in a white-box scenario and another in a black-box scenario, which is adept at misleading DNN-based deepfake detectors into misclassifying fake videos as authentic. The methodology outlined in [17] introduced a face-swapping (FS) generative adversarial network (GAN) for facial swapping and reenactment, featuring numerous technical enhancements. Unlike prior methods, FSGAN provides flexibility by enabling face swapping without requiring specific face training. It introduces a novel recurrent neural network (RNN)-based technique capable of handling pose and expression variations in both images and video sequences. For videos, it introduces continuous face-view interpolation, integrating reenactment, Delaunay triangulation, and barycentric coordinates. Occluded face regions are addressed through a face completion network, while a face-blending network ensures seamless blending while preserving skin color and lighting conditions by utilizing a unique type of Poisson blending loss. Comparative evaluations against existing methods confirmed the qualitative and quantitative superiority of this approach. The research in [18] focused on image animation, generating a video sequence to animate an object within a source image based on a driving video's motion. Unlike previous methods, this framework accomplished this task without requiring annotations or prior object knowledge. Trained on videos of similar object categories, this method applies a self-supervised approach to animate any object within the class. Using learned key points and local affine transformations, it handles complex motions. A generator network blends the appearance of the source image with motion from the driving video to address occlusions. Its superior performance across benchmarks and object categories showcased the framework's efficacy and versatility. The outlined methodology, neural voice puppetry, is a pioneering force in audio-driven facial video creation. Utilizing audio input from a source individual or digital assistant, the system crafts a convincing video of a chosen person synchronized with the input audio. This innovative process harnesses a deep neural network leveraging a latent 3D face model space, ensuring temporal stability learning and employing neural rendering to ensure authentic output frames. Notably adaptable, it accommodates diverse voices, including synthetic ones created using text-to-speech methods. Its applications range from audio-driven avatars to video dubbing and text-based talking-head synthesis, with its effectiveness being demonstrated through extensive examples and comparisons with cutting-edge techniques [19]. The authors of [20] outlined a methodology for speech-driven facial animation, synthesizing talking characters from speech signals using an end-to-end system that requires only a still image and an audio clip. By employing a temporal GAN with three discriminators, the method achieved synchronized lip movements, natural facial

expressions, and detailed frames. The ablation study results highlighted the importance of the frame discriminator, while the approach outperformed static GAN-based methods in terms of generating coherent sequences and precise mouth movements. Although limited to well-aligned frontal faces, the model shows promise in producing lifelike videos and suggests the need for future enhancements in terms of simulating real-world conditions and generating high-definition videos. The methodology outlined in [21] is intended to uncover AI-generated fake facial images or videos, termed deepfakes, by exploiting the integration of synthesized facial regions into authentic images, leading to discernible differences in estimated 3D head poses. Through experiments, this phenomenon was illustrated, and a classification method leveraging this cue was developed. Experimental evaluations on genuine face images and deepfakes were conducted, utilizing features derived from this observation and assessed using an SVM classifier. Masood et al. [22] fixed this with a deep learning model for realistic faces with head movement, emotions, and perfect lip-syncing. The system uses 3D reconstruction and a special module to ensure smooth transitions, all of which are personalized with a short video clip of the target person.

Table 1. Advantages, disadvantages, and accuracy of the related works.

Ref.	Method	Advantage	Disadvantage	Accuracy of Detection %
[1]	Two network architectures, one convolutional neural network (CNN) and one inception-based.	Provides lossless compressed videos.	Processing time is high.	98 (%)
[2]	A method used in forensics to simulate the facial expressions and body language that characterize a speaker.	High recognition rate.	A huge number of features.	99 (%)
[3]	CNN extracts frame-level information, which is subsequently used to train recurrent neural networks (RNNs) that can determine whether a video has been altered.	The number of features used is low.	Using both a CNN and RNN increases the process time.	97.1 (%)
[4]	They provide a fresh metric for assessing GAN outcomes.	As training advances, new layers that simulate ever-finer aspects are added.	Requires a high-resolution video stream.	8.80 (record inception score)
[5]	They suggested two new automated techniques that may be used with any generator design to measure interpolation quality and disentanglement.	Time training convergence is fast.	Requires high-resolution images and a style-based approach.	Not calculated
[6]	Because it is a physiological signal that is poorly represented in fake videos, they employed the artificial neural network (ANN) model to detect eye blinking in the videos.	Good results for identifying films created using deepfake.	Focus on specific aspects of videos, such as blinking patterns.	99.00 (%)
[7]	Using deepfake detectors based on DNNs, real videos are classified as fakes.	By adversarial altering fake movies created using deepfake's existing generation techniques, they showed that it is possible to evade such detectors.	Requires a high level of mathematical complexity and increases processing time.	97.49 (%)

Table 1. *Cont.*

Ref.	Method	Advantage	Disadvantage	Accuracy of Detection %
[8]	It provides a revolutionary face recreation method based on recurrent neural networks (RNNs) that account for differences in position and expression.	Utilizes a brand-new Poisson blending loss that fuses perceptual loss and Poisson optimization.	Requires combining many methods.	Structural similarity index (0.54)
[9]	Using a self-supervised formulation to separate the information about appearance and motion.	This framework performs best over a wide range of benchmarks and object classifications.	Occlusions that occur as a result of target motion and poor precision are modeled by a generator network.	80.6 (%)
[10]	They provide a photo-realistic output video of a target subject synchronized with the source input's audio. A deep neural network uses a latent 3D face model space to power this audio-driven facial recreation.	A set of audio- and text-based puppetry demonstrations demonstrate the method's capabilities.	This approach fails when there are numerous voices in the audio stream.	65 (%)
[11]	An end-to-end system that creates talking head films utilizing a person's still image and a speech-containing audio clip without the use of manually created intermediary features.	They shed light on the model's latent representation.	The recognition rate is low.	80 (%)
[12]	This approach is founded on the observation that deepfakes are produced by splicing a synthetic face region into the original image, creating faults in the process that may be seen when 3D head postures are calculated from the face photos.	Machine learning algorithm used for classification, which reduces the process time.	Low recognition rate and high training process time.	89 (%)
[13]	They are creating a deep neural network model that, given the inputs of a very brief video, V, of the target and an audio signal, A, from the source person, creates a talking-face video with a customized head position. Additionally, a novel memory-augmented GAN module is used.	A powerful technique that requires a minimal number of frames.	The recognition rate is low, and process time is high.	83.67 (%)

Kong et al. proposed an effective method to detect forged faces and locate manipulated regions in images [14]. Their approach uses a segmentation map for high-level semantic information and a noise map for low-level clues, combining these features for accurate detection.

Luo et al. introduced the Critical Forgery Mining (CFM) framework, enhancing generalization and robustness by adapting to various backbones [15]. CFM employs fine-grained triplets, knowledge-agnostic data augmentation, and fine-grained relation learning to mine critical forgery information.

Marwa et al. recommended a system crucial for e-commerce websites to increase product sales by helping users find items of interest based on their history or similar user profiles [16]. They utilized educational data mining to assess the effectiveness of e-learning courses [17], and examined various two-way join algorithms, proposing a new multi-way join algorithm, the hash semi-cascade join, which efficiently joins multiple datasets [18].

Mohamed et al. addressed the development of an emotion-based music recommendation system, aiming to recommend music that aligns with users' current emotions [19].

Their "Hybrid Emotion-Based Music Recommendation System" uses face recognition, color choice, and an arousal map to accurately assess the user's mood.

Sayed et al. discussed the development of an adaptive learning system designed to create personalized and engaging educational experiences for each learner [20], personalizing course materials and evaluation methods based on the unique characteristics of each student.

Another study introduced the Double-Weighted Truncated Nuclear Norm Minus Truncated Frobenius Norm Minimization (DtNFM) model for color image denoising. This model effectively addresses cross-channel differences and spatial noise variations, boasting advantages such as flexible treatment of rank components and an efficient, convergent ADMM-based algorithm. However, the model's implementation is complex and depends on manual parameter selection, which may limit its practicality. Despite these challenges, the DtNFM model shows superior performance in denoising various noise types. Future work will focus on enhancing its practicality and extending its application to low-rank tensor approximation [21].

A versatile dehazing framework was presented, specifically designed for nighttime hazy images, targeting various degradations, enhancing image contrast, and minimizing hidden noise. Key advantages include effective degradation handling, superior image quality, and adaptability to different degraded image scenarios. However, the framework's complexity and lack of real-time processing capabilities are noted limitations. Experimental results indicated that this approach outperforms current methods, with future research aimed at developing real-time solutions for outdoor computer vision applications [22].

The NightHazeFormer was introduced, a transformer-based framework designed for nighttime haze removal. It addresses various degradations and enhances generalization via a two-stage training process. Key benefits include comprehensive handling of degradations, effective two-stage training, utilization of a large-scale synthetic dataset (UNREAL-NH), and superior performance compared to current methods. However, potential drawbacks include implementation complexity and dependency on datasets. Experimental results illustrated NightHazeFormer's superior performance and effective generalization [23].

The proposed approach in [24] aims to create a single variational model addressing both haze removal and noise reduction in real-world images, effectively managing these challenges. It employs TV and GTV regularization to maintain structural edges and enhance image details, resulting in better visibility restoration and noise reduction than current methods. Extensive experiments validate its efficacy, yet it encounters challenges, such as implementation complexity, dataset reliance, and sensitivity to parameter adjustments. The study aimed to enhance visibility and visual quality in challenging atmospheric conditions, demonstrating notable advancements in both qualitative and quantitative assessments [24].

Sometimes, deepfake video frames are unclear, and to enhance their quality, we need to use the image haze removal method [25].

2.1. Leveraging Artificial Intelligence for Enhanced Deepfake Detection

The rapid evolution of artificial intelligence (AI) techniques has paved the way for innovative strategies to counter the proliferation of deepfake videos [26]. This section delves into incorporating AI methodologies into the detection pipeline, emphasizing their pivotal role in bolstering accuracy, resilience, and efficiency.

2.1.1. Deep Learning in Deepfake Detection

- Deep learning, a subset of AI, has emerged as a cornerstone in the battle against deepfake manipulation [27,28]. By harnessing neural networks with intricate layers, deep learning models autonomously glean complex patterns and features from extensive datasets [29,30]. In the realm of deepfake detection, deep learning techniques offer several advantages.

- Feature representation: deep neural networks excel in automatically extracting hierarchical representations from raw data, facilitating the capture of nuanced cues indicative of deepfake alterations [31].
- End-to-end learning: these models can undergo end-to-end training, seamlessly integrating feature extraction and classification stages to streamline the detection process [32].
- Adaptability: deep learning models demonstrate remarkable adaptability to diverse manipulation techniques, rendering them adept at detecting evolving forms of deepfake content [33,34].

2.1.2. Convolutional Neural Networks for Facial Analysis

Convolutional neural networks have shown impressive efficacy in scrutinizing facial features and patterns, rendering them indispensable tools in the realm of deepfake detection [35,36]. By leveraging spatial hierarchies within facial images, CNNs can discern subtle disparities indicative of facial manipulation [37]. The key applications of CNNs in deepfake detection encompass the following:

- Facial feature extraction: CNN architectures adeptly extract discriminative facial features, such as textures, shapes, and expressions, facilitating precise discrimination between authentic and manipulated faces [36].
- Transfer learning: transfer learning, wherein pre-trained CNN models are fine-tuned on domain-specific datasets, proves instrumental in enhancing detection performance, particularly in scenarios with limited training data [38].

2.1.3. Ensemble Learning and Metaheuristic Optimization

Ensemble learning techniques, amalgamating predictions from multiple base classifiers, offer a potent strategy for use in augmenting detection accuracy and resilience against adversarial attacks [39]. By amalgamating diverse classifiers trained on varied subsets of data, ensemble methods mitigate the risk of overfitting and enhance generalization [40,41].

Moreover, metaheuristic optimization algorithms, such as genetic algorithms (GAs) and PSO, can be harnessed to fine-tune hyperparameters and fortify the robustness of deepfake detection models [42]. These algorithms furnish efficient search strategies for navigating the expansive parameter space, thereby refining model performance and convergence speed [41].

2.1.4. Integration of Artificial Intelligence (AI) Techniques for Comprehensive Detection

In practice, the most efficacious deepfake detection systems often amalgamate multiple AI techniques, capitalizing on the strengths of each approach to achieve exhaustive analysis and robust performance [43]. By amalgamating deep-learning-based feature extraction with ensemble learning for classification and metaheuristic optimization for model refinement, these integrated systems furnish state-of-the-art capabilities in terms of thwarting deepfake manipulation [44].

The integration of AI techniques, including deep learning, ensemble learning, and metaheuristic optimization, heralds a paradigm shift in combatting deepfake manipulation. By harnessing the collective potency of these methodologies, we can augment the detection accuracy, adaptability, and efficiency, thereby safeguarding the integrity of digital media and mitigating the perils associated with misinformation and manipulation [43].

2.2. Comparison between Deep Learning and AI Approaches

2.2.1. Deep-Learning-Based Detection

Deep learning methods for identifying deepfake videos have been scrutinized regarding both their effectiveness and shortcomings [45]. Using neural networks, CNNs, and RNNs, these approaches excel in capturing intricate nuances and features suggestive of manipulation [46].

2.2.2. Artificial-Intelligence-Based Detection

Expanding beyond the confines of deep learning, AI encompasses a broader array of techniques, such as ensemble learning, metaheuristic optimization, and traditional machine learning algorithms. These methodologies offer distinct advantages, including resilience against adversarial attacks and adaptability to the ever-evolving landscape of manipulation techniques [47].

The main objective of this paper is to determine the features of the optical flow based on ACO-PSO and then use these features in training machine learning (ML) to find the fake and non-fake faces in the video files. Three feature extraction methods based on optical flow have been used and tested in this study: ant colony optimization (ACO), PSO, and GA. This approach combines the three histogram-feature extraction methods and deep learning features and uses them in training the ML to recognize fake and non-fake faces. The optical flow method is used to demodulate the magnitude and orientation of the image. Magnitude is used in ACO, and orientation is used in PSO and GA. Combining the suggested method with any current feature descriptor can create a compact and expressive feature descriptor. The image quality of the proposed method is excellent, and it can handle images that cannot be read, such as fingerprints. In a variety of facial recognition circumstances, it outperforms existing state-of-the-art techniques. Image quality can affect the magnitude and orientation (it can cause one feature to be mistaken for another). This depends on a pixel's illumination, which can change the magnitude and orientation values. Handling unreadable images is more complex.

3. Material and Methods

Optical flow describes the movement of objects within an image. It is frequently applied to a series of images, such as video frames. Optical flow calculates the velocity of pixels within these images and predicts where these pixels may be in subsequent image series. The optical flow vector field demonstrates how shifting pixels in an object from the first image may result in an identical object in the second image. Since the optical flow field can be approximated if the corresponding pixels of an object are known, it is a type of correspondence learning. The flowchart of the optical flow algorithm used to extract features is shown in Figure 1.

ACO is inspired by the foraging behavior of ants, where they find the shortest path to a food source by laying down pheromones. The key steps in the ACO algorithm are as follows:

1. Initialization:

   Set the number of ants, m.
   Initialize the pheromone matrix, τ_{ij}, and heuristic information, η_{ij}.

2. Construct Ant Solutions:

Each ant constructs a solution by moving from node to node based on a probabilistic decision rule.

The probability, P_{ij}^{k}, of ant kk moving from node i to node j is given by:

$$P_{ij}^{k} = \frac{\left[\tau_{ij}\right]^{\alpha}\left[\eta_{ij}\right]^{\beta}}{\sum_{l \in \mathcal{N}_i^k}\left[\tau_{il}\right]^{\alpha}\left[\eta_{il}\right]^{\beta}} \quad (1)$$

where α and β are parameters that control the influence of the pheromone and heuristic information, respectively, and N_i^k is the set of feasible moves.

3. Update Pheromones:

After all ants have constructed their solutions, update the pheromone levels:

$$\tau_{ij} = (1-\rho)\tau_{ij} + \sum_{k=1}^{m} \Delta\tau_{ij}^{k} \quad (2)$$

where ρ is the evaporation rate and $\Delta\tau_{ij}^{k}$ is the pheromone deposited by ant k:

$$\Delta\tau_{ij}^{k} = \begin{cases} \frac{Q}{L_k} & \text{if ant } k \text{ use sedge } (i,j) \text{ in its solution} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where Q is a constant and L_k is the length of the solution constructed by ant k.

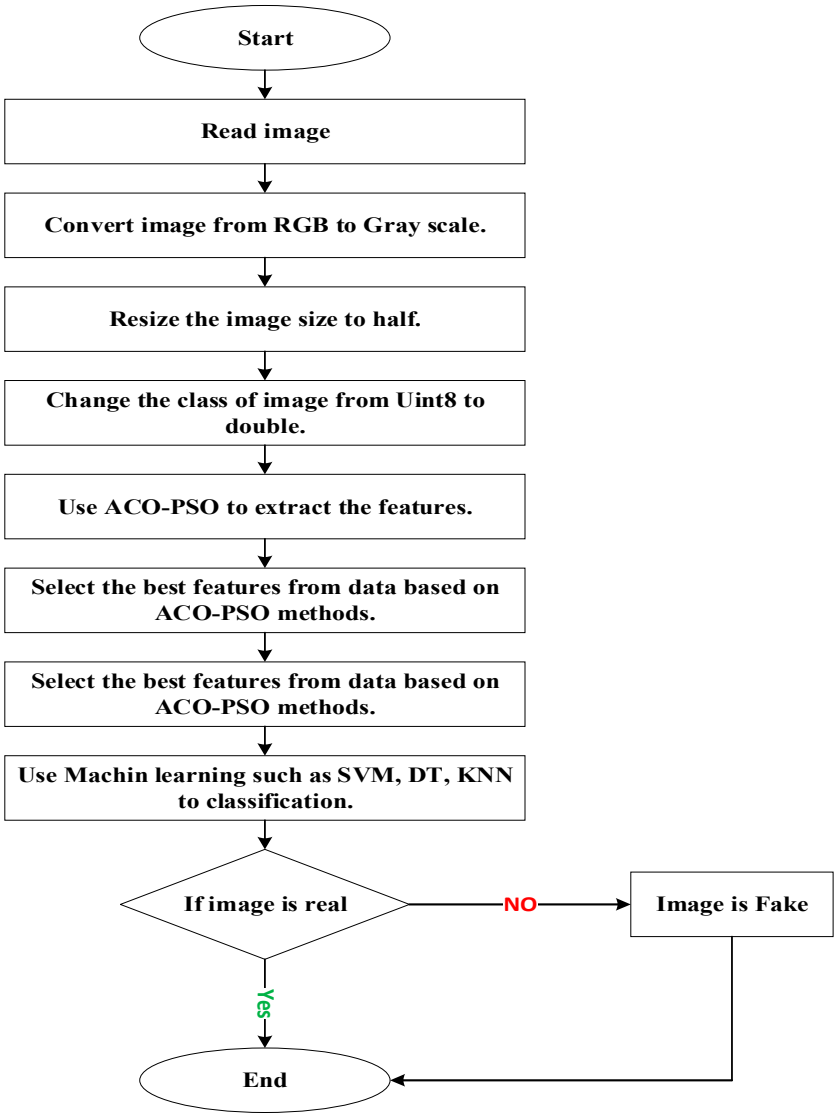


Figure 1. Proposed method.

PSO is inspired by the social behavior of birds flocking or fish schooling. The algorithm maintains a population of particles, where each particle represents a potential solution. The key steps in the PSO algorithm are:

1. Initialization:

   Initialize the positions, x_i, and velocities, v_i, of n particles randomly.
   Set the personal best position, p_i, for each particle and the global best position, g.

2. Update Velocities:

Update the velocity, v_i, of each particle based on its personal best position and the global best position:

$$v_i(t+1) = \omega v_i(t) + c_1 r_1 [p_i - x_i(t)] + c_2 r_2 [g - x_i(t)] \quad (4)$$

where ω is the inertia weight, c_1 and c_2 are cognitive and social coefficients, and r_1 and r_2 are random values between 0 and 1.

3. Update Positions:

Update the position *xi* of each particle:

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (5)$$

4. Update Personal and Global Bests:

Update the personal best position, p_i, if the new position is better.

Update the global best position, g, if any personal best position is better than the current global best.

Hybrid ACO-PSO Algorithm

A hybrid ACO-PSO algorithm leverages the strengths of both ACO and PSO to achieve better performance in optimization tasks. One possible hybridization approach is outlined below:

1. Initialization:

Initialize the population of particles as in PSO.
Initialize the pheromone matrix as in ACO.

2. PSO Update:

For each particle, update its velocity and position using the PSO update rules.
Evaluate the new positions and update the personal and global bests.

3. ACO Update:

For each ant (which could be considered as a particle), construct solutions using the ACO probabilistic decision rule.

Evaluate the solutions and update the pheromone matrix as in ACO.

4. Information Sharing:

Integrate the information from PSO and ACO by influencing the pheromone update with the global best position found by PSO.

Use the pheromone levels to guide the particles in PSO by adjusting the velocities toward regions with higher pheromone concentrations.

5. Iteration:

Repeat the PSO and ACO updates for a fixed number of iterations or until convergence. The dataset used in this study was obtained from [48] and can be downloaded from the following link: (https://www.kaggle.com/competitions/deepfake-detection-challenge/data, 10 November 2023).

There are limited datasets available for deepfake detection. The Deepfake Detection Challenge (DFDC) dataset [49], one of the largest of these, can be used. It is a training dataset published for use in the DFDC Kaggle competition. Facebook created this dataset of more than 100,000 videos. The videos are shot by real people and contain a mix of real and fake videos. Various algorithms were used to create the deepfake videos. The size of this dataset is 471 GB, of which 4.5 GB were used in this paper. Some of the images are shown in Figure 2.

Figure 2. Real images and fake images.

4. Innovative Solutions for Overcoming Contemporary Challenges in Facial Forgery Detection

The method presented here provides a compelling resolution to the obstacles encountered in facial forgery detection, thereby playing a crucial role in preserving the integrity of digital media amid the escalating prevalence of misinformation and manipulation, as depicted in Figure 3.

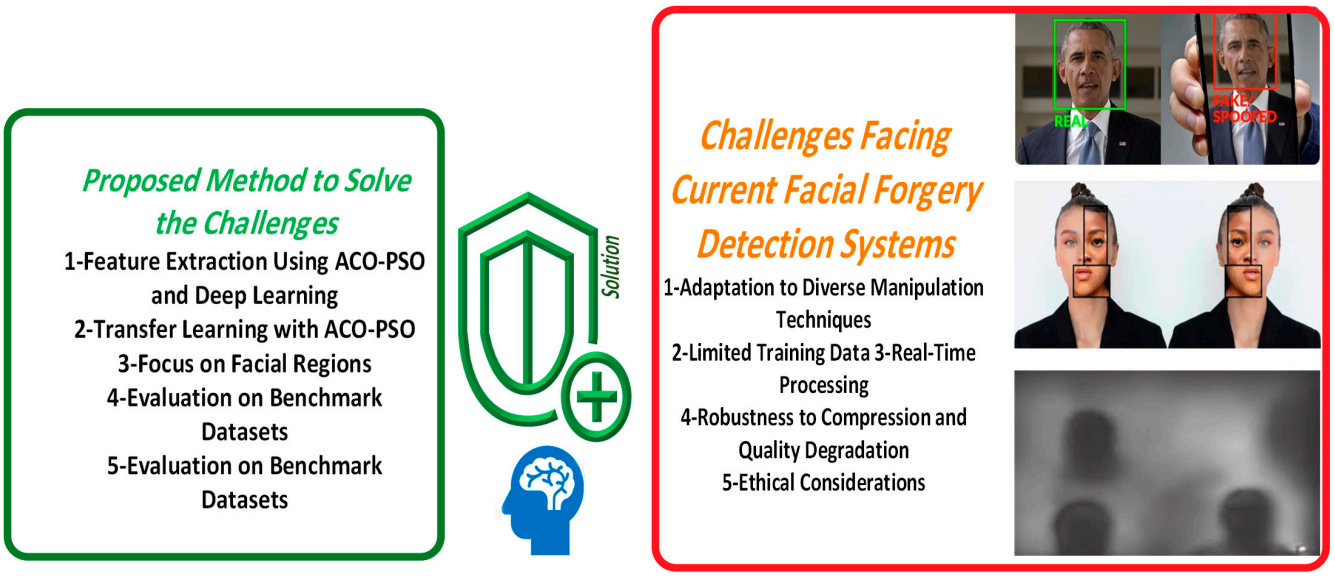


Figure 3. Proposed method to resolve the challenges facing current facial forgery detection systems.

A. Challenges facing contemporary facial forgery detection systems

1. Adapting to varied manipulation techniques: existing systems struggle with the ever-evolving manipulation techniques seen in deepfake creation, ranging from basic facial swaps to intricate alterations of expressions, lighting, and occlusions.

2. Limited training data: obtaining labeled data for training is hindered by privacy concerns and the expertise needed to label deepfake videos, impacting the system's ability to be generalized to new manipulation techniques.
3. Real-time processing: handling high-resolution video streams in real time, especially on platforms such as social media or video-streaming services, poses significant computational hurdles.
4. Robustness to compression and quality loss: deepfake videos undergo compression and quality degradation when shared online, challenging detection systems to maintain accuracy across various platforms and viewing conditions.
5. Ethical considerations: deploying these systems raises ethical concerns regarding privacy, consent, and potential misuse, necessitating the careful consideration of user consent, false positives, and privacy implications.

B. How our method addresses these challenges

1. Feature integration using ACO-PSO and deep learning: our approach combined the features of ACO-PSO and deep learning to enhance detection accuracy and robustness.
2. Transfer learning with ACO-PSO: we employed transfer learning with pre-trained ACO-PSO models to leverage knowledge from large datasets, reducing the need for extensive training data.
3. Focus on facial regions: by concentrating on facial features using the Viola–Jones face detection method, we streamlined detection and reduced computational overhead.
4. Evaluation of benchmark datasets: extensive experiments on benchmark datasets allowed us to assess our method's effectiveness in terms of detection accuracy, robustness, and generalization to new data.
5. Computational efficiency: we analyzed the computational efficiency of our approach, ensuring practical feasibility for real-time applications while maintaining accuracy.

As indicated in Table 2, these significant resources introduce innovative approaches to identifying deepfakes and tackling the various obstacles involved, including localization, adversarial learning, generalization, and the extraction of crucial cues. Researchers and practitioners have the opportunity to harness the insights and techniques provided in these resources to actively combat the widespread dissemination of deepfake content.

Table 2. Key resources for enhanced deepfake detection.

Resource Title	Description
Recognition and Localization	This resource focuses on recognizing and pinpointing deepfake manipulation by analyzing semantic markers and noise levels in facial images. By identifying specific features indicative of manipulation, it aims to elevate the accuracy and precision of deepfake detection and localization.
Dual Adversarial Learning	Employing dual adversarial learning techniques, this resource is intended to construct a robust face forgery detection framework capable of identifying a wide spectrum of facial alterations. Training two adversarial networks simultaneously enhances the model's ability to distinguish between authentic and doctored facial images.
Facial X-Ray	This resource proposes a methodical examination of subtle but pivotal clues within facial imagery by likening deepfake detection to an X-ray of the face. By scrutinizing these cues, it aims to achieve the more comprehensive detection of facial forgery, encompassing various deepfake manipulation techniques.
Beyond Conventional Wisdom	Breaking free from traditional approaches that rely solely on prior knowledge, this resource advocates for the extraction of crucial cues embedded within facial images to enhance forgery detection efficacy. By harnessing and leveraging these cues, it seeks to augment deepfake detection systems' ability to discern a broader array of manipulation techniques.

5. Dataset Description and Features

The dataset employed in this research served as the cornerstone for both training and assessing the efficacy of the novel deepfake detection approach proposed herein. Sourced from the DFDC database, generously provided by Meta, this dataset represents a rich tapestry of videos encompassing a broad spectrum of authenticity, ranging from genuine to synthetically manipulated content. Categorically organized, this dataset enables the systematic development and rigorous validation of robust deepfake detection algorithms.

Dataset Features

(1) Diverse video content: this dataset spans a wide gamut of video content, ranging from everyday scenarios to orchestrated performances, ensuring a comprehensive representation of real-world contexts.
(2) Manipulation techniques: videos within the dataset exhibit a plethora of manipulation techniques characteristic of deepfake creation, including facial swaps, expression alterations, lighting modifications, and occlusion effects. This diversity fosters the thorough training and evaluation of detection algorithms across various manipulation scenarios.
(3) Variation in resolution and quality: reflecting real-world conditions encountered across diverse platforms and recording devices, videos in the dataset exhibit variations in resolution and quality. This diversity fortifies the robustness of detection algorithms against discrepancies in video resolution and quality, which is vital for real-world deployment.
(4) Temporal dynamics: capturing the intrinsic temporal dynamics of video sequences, the dataset encapsulates motion patterns, temporal irregularities, and frame-level alterations. This temporal dimension enriches the training data, empowering algorithms to discern subtle temporal cues indicative of deepfake manipulation.
(5) Focus on facial regions: Considering the predominant focus on deepfake manipulation of facial features, the dataset emphasizes facial regions within videos. Leveraging the Viola–Jones face detection method, the dataset streamlines the extraction of facial frames, optimizing the detection process and enhancing computational efficiency.
(6) Annotation and labeling: Each video in the dataset undergoes meticulous annotation and labeling, indicating its authenticity status—whether it is real or a deepfake. These annotations serve as ground truth labels, ensuring the accuracy and reliability of performance metrics during algorithm training and evaluation.

By harnessing the diverse array of features embedded within the dataset, the proposed methodology undergoes rigorous training and evaluation, culminating in the development of robust deepfake detection capabilities. The incorporation of various manipulation techniques, resolution disparities, and temporal dynamics and an emphasis on facial regions enrich the training data, enabling algorithms to be effectively generalized to real-world scenarios and emerging manipulation techniques.

6. Results and Discussion

In this section, the results obtained from the proposed method are discussed. Ant colony optimization–particle swarm optimization yielded the best results. The main reason for this is the small size of the dataset. If we had used the entire DFDC dataset, we might have obtained better results with other networks. For frame extraction, we used one of MATLAB's libraries, the Viola–Jones face detection method. Viola–Jones face detection provides the frames from each video in the dataset. In this study, we set the number of frames to 10 for each video. With the mentioned face detection method, more than 4000 frames were extracted from the videos. The main reason for focusing on the faces in the videos is that deepfake videos are typically created by changing faces. Focusing on the face and ignoring other regions saved us a great deal of time and file size.

After all the steps performed on the dataset, the ACO-PSO model was built using the transfer learning method. There are several ways to realize that transfer learning pathways

are available, such as using this pre-trained model or creating a new one. There are two ways to employ a pre-trained model. One can train the model using weights after using pre-trained weights and biases as initial parameters. The alternative is to extract features using the pre-trained model. A classifier can be trained on the input image by using the parameters of the pre-trained model to extract features from it. If there is not enough data available, one can also train a new model using the weights of one that was made for a similar problem to the one that involved a great deal of data. The ACO-PSO family of models, which has already been trained, was employed in this study as a feature extractor, and a classifier was trained on top of it to produce the prediction output.

Ant colony optimization–particle swarm optimization was chosen as the initial model due to its proven effectiveness and suitability considering the size of our dataset. This decision considered the trade-off between model complexity and ensemble dimension, acknowledging that with larger datasets, alternative networks may yield superior results.

Transfer learning with a family of pre-trained ACO-PSOs was used to leverage the knowledge gained from a model trained on a larger dataset. This method contributed to reducing the need for extensive data for training and utilizing pre-existing features learned by the model.

By combining these diverse sets of machine learning algorithms, we intended to achieve the comprehensive and accurate detection of deepfake videos, considering the subtleties of facial changes and the overall dynamics of video sequences. Each algorithm was strategically chosen and contributed to the strength of our proposed method. The dataset used in this study was obtained from the DFDC database, provided by Meta. The dataset was primarily focused on deepfake videos, and it classified videos into either the deepfake or real categories. Each video underwent face extraction using the Viola–Jones face recognition method, which efficiently separated facial regions. The focus on the face was justified by the prevalent nature of deepfake videos, which primarily manipulate facial features. After the face extraction process, the dataset was further processed using the proposed method, specifically ACO-PSO, for feature extraction and deepfake detection. The colorectal dataset was specifically selected for evaluation, and the suggested strategy was applied to assess the approach's efficacy, with the number of frames being set to 10 per video. The division of the dataset was crucial for training and testing the proposed algorithm, allowing for a comprehensive evaluation of its performance in distinguishing between deepfake and real videos.

The decision to utilize ACO-PSO and extract optical flow features was a strategic move aimed at bolstering the precision and resilience of deepfake detection. In fact, ACO-PSO features were specifically selected for their capability to capture nuanced patterns and irregularities within the spatial and temporal dimensions of video frames, providing valuable cues indicative of deepfake manipulation. Leveraging the strengths inherent in ACO-PSO, our proposed methodology effectively discerned between authentic and deepfake videos, even considering the intricacies of sophisticated manipulation techniques.

Similarly, the inclusion of optical flow features assumed a pivotal role in deepfake detection by elucidating the movement of pixels within images over time. This dynamic insight proved particularly valuable in the realm of video analysis, where it can uncover disparities in motion suggestive of tampering. By integrating optical flow features into our methodology, we gained deep insights into the temporal dynamics of video sequences, thereby enhancing our capacity to accurately pinpoint deepfake alterations. In essence, the integration of ACO-PSO and optical flow features empowered our proposed method to provide a comprehensive analysis encompassing both the spatial and temporal dimensions of video data. This holistic approach fortified our ability to detect deepfake content and surmount the obstacles created by evolving manipulation techniques.

In this study, images were evaluated using accuracy, sensitivity, precision, and F1 scores, as determined by Equations (6)–(10), respectively [50–53]:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (6)$$

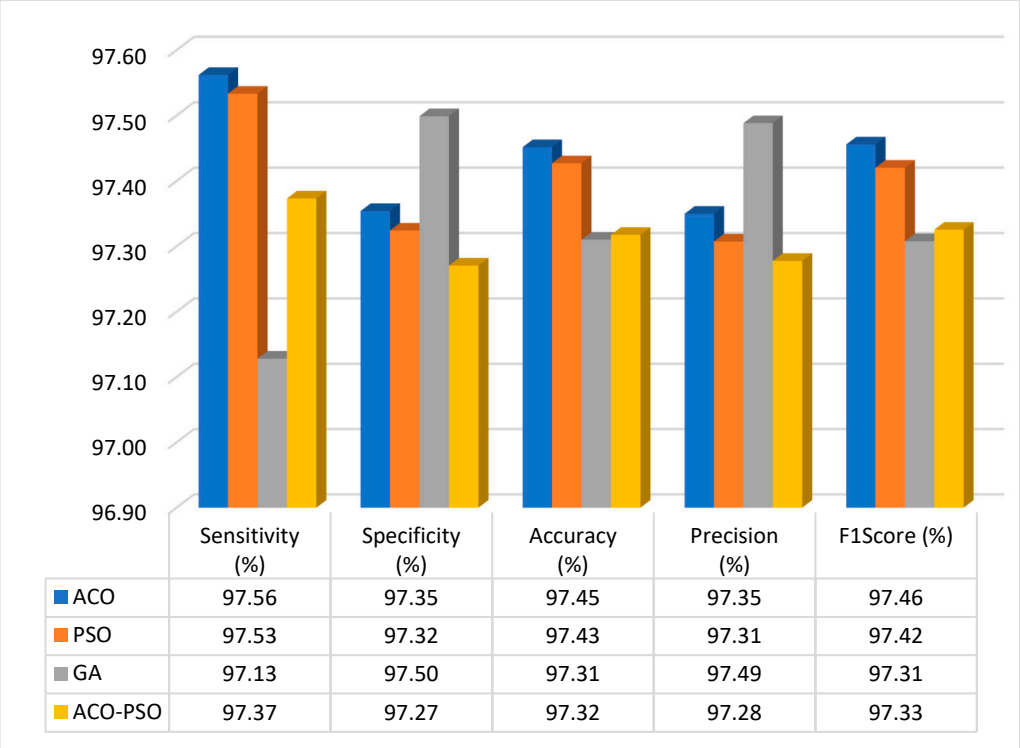
$$Sensitivity = \frac{TP}{TP + FN} \times 100 \tag{7}$$

$$Specificty = \frac{TN}{TN + FP} \times 100 \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \times 100 \tag{9}$$

$$F1 = 2 \times \frac{Recall \times Precision}{Recall + Precision} \times 100 \tag{10}$$

True positive (TP), false positive (FP), false negative (FN), and true negative (TN) were used to determine each of these indicators [53]. Accuracy (ACC) is the proportion of correct predictions out of all predictions. The decision tree and support vector machine were both used in this study's classification of deepfake detection. The new method was also contrasted with ambiguous varieties of traditional ACO, PSO, and GA methods. Regarding repeating the optical flow in an experiment, Figure 4 illustrates the precision values with which false and actual images can be identified. Four proposed methods were applied in this implementation to extract the features from deepfake detection images. These techniques included ACO-PSO.



	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)	F1Score (%)
ACO	97.56	97.35	97.45	97.35	97.46
PSO	97.53	97.32	97.43	97.31	97.42
GA	97.13	97.50	97.31	97.49	97.31
ACO-PSO	97.37	97.27	97.32	97.28	97.33

Figure 4. Results of the various methods.

The ACO-PSO technique demonstrated effectiveness in classifying image correctness according to analysis and evaluation. The accuracy was enhanced by carefully selecting the feature vector in the optical flow method. Additionally, the ACO-PSO method reduced server-to-server traffic during distributed training. For machine learning and training, the feature selection method identified the most significant features from both real and synthetic images. In this experiment, the accuracy of the ACO-PSO was 98.25%. The accuracy of the ACO-PSO on the colorectal dataset increased to an average of 98.88% as the population of feature vectors grew due to 20 experiments.

First, the colorectal dataset was selected, the suggested strategy was implemented, and the images were used to evaluate the approach's efficacy. Table 2 shows the results

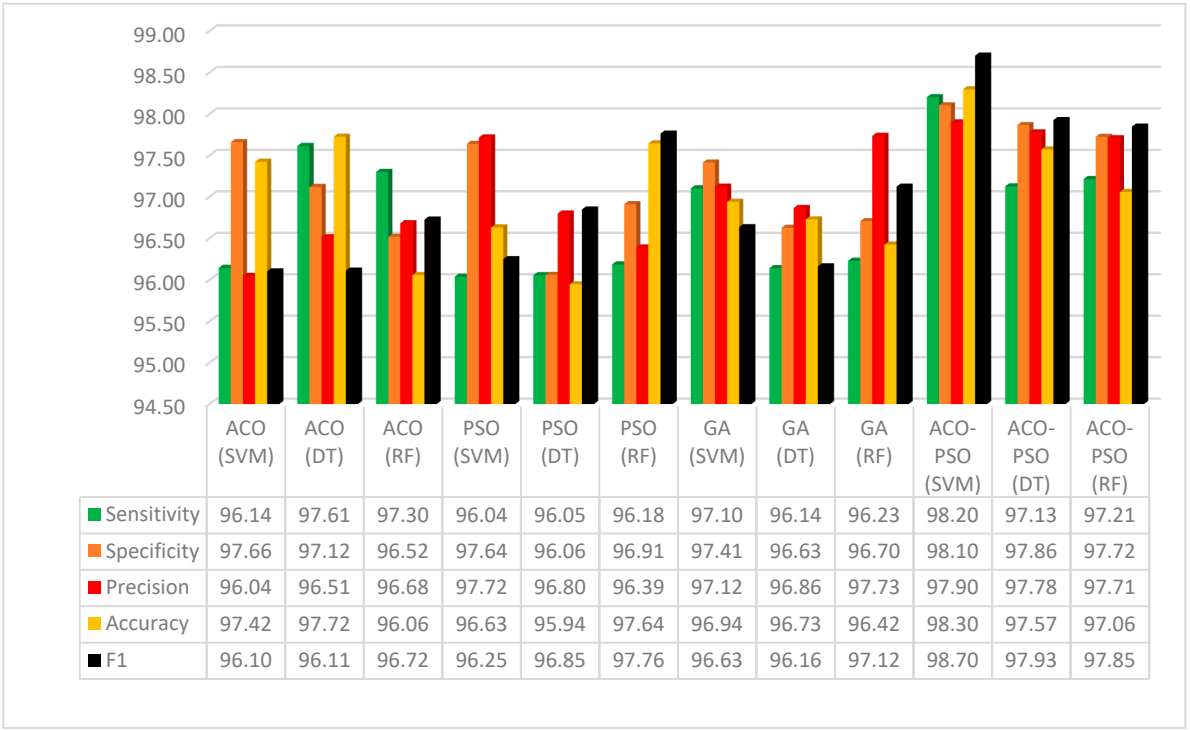
regarding the sensitivity, specificity, and accuracy index comparisons between the proposed method, such as ACO, PSO, GA, and optical flow algorithms based on ACO-PSO approaches. For two classes, the suggested method's mean sensitivity, specificity, and accuracy values were compared with those of competing methods. Table 3 shows the relationship between the sensitivity, specificity, accuracy, and F1 scores.

Table 3. Comparison of the proposed method with various metaheuristic methods.

Method	Sensitivity	Specificity	Accuracy	F1 Score
ACO	97.56	97.35	97.45	97.35
PSO	97.53	97.32	97.43	97.31
GA	97.13	97.50	97.31	97.49
ACO-PSO	97.37	97.27	97.32	97.28

According to the analysis, the suggested method's average sensitivity, specificity, accuracy, and F1 score index were 8.21%, 98.56%, 98.25%, and 97.56%, respectively. Compared to other methods, the suggested method performed better in terms of sensitivity, specificity, accuracy, and F1 score in the analysis and classification of false and real images. The ACO technique had an accuracy index of 93.90%. The proposed method's index was equivalent to 99.12%. The ACO algorithm yielded the worst results, as can be seen in Table 2. Sensitivity, specificity, accuracy, and F1 score were measured using the ACO algorithm, and the results were 97.18%, 95.79%, 93.90%, and 96.01%, respectively. The DFDC dataset was also used to further assess the suggested approach.

According to an analysis of the proposed method, as well as other ways of classifying fake and real images, the suggested method was based on four indicators (accuracy, precision, sensitivity, and F1 scores) for methods such as ACO, PSO, and GA. These indicators were derived through an examination of the proposed method and other methods. Classifying fake and real images was more successfully accomplished using optical flow. The comparison between the proposed method and other methods is shown in Figure 5.



	ACO (SVM)	ACO (DT)	ACO (RF)	PSO (SVM)	PSO (DT)	PSO (RF)	GA (SVM)	GA (DT)	GA (RF)	ACO-PSO (SVM)	ACO-PSO (DT)	ACO-PSO (RF)
Sensitivity	96.14	97.61	97.30	96.04	96.05	96.18	97.10	96.14	96.23	98.20	97.13	97.21
Specificity	97.66	97.12	96.52	97.64	96.06	96.91	97.41	96.63	96.70	98.10	97.86	97.72
Precision	96.04	96.51	96.68	97.72	96.80	96.39	97.12	96.86	97.73	97.90	97.78	97.71
Accuracy	97.42	97.72	96.06	96.63	95.94	97.64	96.94	96.73	96.42	98.30	97.57	97.06
F1	96.10	96.11	96.72	96.25	96.85	97.76	96.63	96.16	97.12	98.70	97.93	97.85

Figure 5. The proposed method's mean index of sensitivity, accuracy, and precision in comparison to rival techniques.

The proposed method successfully classified images from the DFDC dataset with a sensitivity, specificity, precision, accuracy, and F1 score of 99.34%, 99.08%, 98.91%, 99.12%, and 98.94%, respectively. Evaluations revealed that the optical flow approach had the lowest classification accuracy for deepfake detection, at 91.88%. The suggested method achieved the highest sensitivity score, while optical flow showed the lowest performance in this metric. In terms of precision, the proposed strategy performed the best with a score of 98.91%, whereas the optical flow method had the lowest precision index at 61.86%. The proposed strategy also outperformed other methods in the F1 index, while optical flow showed the worst performance.

7. Limitations of the Proposed Method

1. Dataset size: Acknowledging the influence of dataset size on method effectiveness is crucial. While the ACO-PSO model exhibited proficiency with the current dataset, it is important to recognize that outcomes may differ with larger datasets. This constraint could impact the adaptability of the approach across diverse contexts with varying dataset sizes.
2. Computational complexity: The integration of deep learning techniques with optimization algorithms such as ACO-PSO can lead to computational intensiveness. This complexity may pose challenges for real-time applications, especially on platforms with limited computational resources.
3. Robustness to new manipulation techniques: The dynamic nature of deepfake creation constantly introduces novel manipulation techniques. Ensuring the method's resilience to these evolving techniques necessitates ongoing adaptation and updates, presenting a hurdle in terms of maintaining detection accuracy over time.
4. Ethical considerations: Deploying deepfake detection systems raises ethical concerns surrounding privacy, consent, and potential misuse. Addressing these ethical implications requires the development of comprehensive mechanisms to safeguard against misuse and ensure ethical usage.

Potential Solutions

1. Dataset expansion: expanding the dataset to include a wider array of manipulation techniques, resolutions, and quality levels can enhance method generalizability and facilitate a more thorough performance evaluation.
2. Model optimization: Exploring techniques to enhance the computational efficiency of the model without compromising detection accuracy is essential. This may involve strategies such as model compression, pruning, or the development of lightweight architectures tailored for real-time deployment.
3. Adaptability mechanisms: Developing mechanisms to dynamically monitor and adapt the detection system to emerging manipulation techniques in real time is critical. Continuous model retraining using incoming data streams or integrating anomaly detection algorithms can aid in identifying and addressing suspicious content.
4. Ethical framework development: Collaborating with ethicists, policymakers, and stakeholders to formulate comprehensive ethical frameworks is imperative. These frameworks should address issues of consent, user privacy, and potential societal impacts, ensuring responsible deployment and minimizing harm.

8. Conclusions

This study introduced a novel method for identifying counterfeit videos by integrating ACO-PSO features with deep learning techniques. Our findings underscored the effectiveness of this approach, achieving an impressive accuracy of 98.91% and an F1 score of 99.12% in distinguishing between genuine and manipulated videos. By strategically integrating ACO-PSO features, which capture spatial and temporal patterns indicative of deepfake manipulation, with deep learning classifiers, our approach offers superior precision and resilience in detection.

Furthermore, we leveraged transfer learning with pre-trained ACO-PSO models, tapping into knowledge gained from models trained on larger datasets. This strategy reduced the need for extensive training data while leveraging pre-existing features learned by the model. Consequently, it addressed challenges posed by limited datasets available for deepfake detection, yielding promising results even with smaller datasets.

Moreover, incorporating optical flow features enhanced our method's ability to analyze temporal dynamics, providing insights into motion disparities suggestive of tampering. This comprehensive approach, encompassing both spatial and temporal analyses, strengthened our ability to detect deepfake content amid evolving manipulation techniques.

Looking forward, it will be imperative to develop more robust models and further reduce error rates. Emphasizing advancements in the deepfake creation process and deploying deep-learning-based ensemble methods will be essential in staying ahead of emerging threats. By actively pursuing innovative techniques and drawing insights from ongoing research, we can continue to enhance the effectiveness and efficiency of deepfake detection, safeguarding digital media integrity and mitigating the risks associated with misinformation and manipulation.

9. Future Work

1. Enhanced ensemble methods: Delve into the untapped potential of refining ensemble learning techniques by integrating a broader spectrum of machine learning algorithms and architectures. Explore innovative combinations to fine-tune deepfake detection accuracy and resilience.
2. Dynamic feature selection: Investigate dynamic feature selection methods aimed at intelligently adapting to select the most pertinent features for deepfake detection. This approach has the potential to bolster efficiency and efficacy, particularly in real-time applications
3. Adversarial defense mechanisms: Forge robust adversarial defense mechanisms to counter the impact of adversarial attacks on deepfake detection systems. Explore advanced techniques, such as adversarial training and GANs, to fortify model resilience.
4. Continual learning strategies: Explore continual learning strategies to empower deepfake detection models to evolve and improve over time. This is especially crucial in the face of evolving manipulation techniques and emerging variants of deepfake technology.
5. Ethical and legal implications: Scrutinize the ethical and legal dimensions of deploying deepfake detection systems, delving into issues surrounding privacy, consent, and potential misuse. Develop comprehensive guidelines and frameworks to ensure the responsible deployment and ethical usage of deepfake detection technology.
6. User education and awareness: Place a spotlight on educating users about the existence and potential risks associated with deepfake technology. Empower users to critically assess media content and equip them with the necessary tools to safeguard against misinformation and manipulation.
7. Real-world deployment and integration: Initiate pilot studies and forge collaborations with pertinent stakeholders to seamlessly deploy deepfake detection systems and integrate them into real-world scenarios. This includes integration into platforms such as social media, video-streaming services, and forensic analysis tools.
8. Multi-modal fusion: Explore the fusion of diverse modalities, such as audio, text, and contextual information, with visual data to create a more comprehensive deepfake detection framework. Uncover synergies between various modalities to amplify detection accuracy and robustness.
9. Benchmarking and evaluation: Establish standardized benchmarks and evaluation protocols for deepfake detection systems to facilitate fair comparison and reproducibility across various research endeavors. Continuously update benchmarks to reflect evolving challenges and advancements in deepfake technology.

Author Contributions: Methodology, H.S.A.; Writing—original draft, Y.C.; Project administration, S.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets generated and analyzed during the current study are available from the corresponding authors upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Afchar, D.; Nozick, V.; Yamagishi, J.; Echizen, I. Mesonet: A compact facial video forgery detection network. In Proceedings of the 2018 IEEE International Workshop on Information Forensics and Security (WIFS), Hong Kong, China, 11–13 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–7.
2. Agarwal, S.; Farid, H.; Gu, Y.; He, M.; Nagano, K.; Li, H. Protecting World Leaders Against Deep Fakes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 15–20 June 2019; p. 38.
3. Güera, D.; Delp, E.J. Deepfake video detection using recurrent neural networks. In Proceedings of the 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zealand, 27–30 November 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–6.
4. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. *arXiv* **2017**, arXiv:1710.10196.
5. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4401–4410.
6. Li, Y.; Chang, M.-C.; Lyu, S. In ictu oculi: Exposing ai created fake videos by detecting eye blinking. In Proceedings of the 2018 IEEE International Workshop on Information Forensics and Security (WIFS), Hong Kong, China, 11–13 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–7.
7. Hussain, S.; Neekhara, P.; Jere, M.; Koushanfar, F.; McAuley, J. Adversarial deepfakes: Evaluating vulnerability of deepfake detectors to adversarial examples. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Virtual, 5–9 January 2021; pp. 3348–3357.
8. Nirkin, Y.; Keller, Y.; Hassner, T. Fsgan: Subject agnostic face swapping and reenactment. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 7184–7193.
9. Siarohin, A.; Lathuilière, S.; Tulyakov, S.; Ricci, E.; Sebe, N. First order motion model for image animation. In Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada, 8–14 December 2019.
10. Thies, J.; Elgharib, M.; Tewari, A.; Theobalt, C.; Nießner, M. Neural voice puppetry: Audio-driven facial reenactment. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 716–731.
11. Vougioukas, K.; Petridis, S.; Pantic, M. Realistic speech-driven facial animation with gans. *Int. J. Comput. Vis.* **2020**, *128*, 1398–1413. [CrossRef]
12. Yang, X.; Li, Y.; Lyu, S. Exposing deep fakes using inconsistent head poses. In Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 8261–8265.
13. Yi, R.; Ye, Z.; Zhang, J.; Bao, H.; Liu, Y.-J. Audio-driven talking face video generation with learning-based personalized head pose. *arXiv* **2020**, arXiv:2002.10137.
14. Kong, C.; Chen, B.; Li, H.; Wang, S.; Rocha, A.; Kwong, S. Detect and locate: Exposing face manipulation by semantic-and noise-level telltales. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 1741–1756. [CrossRef]
15. Luo, A.; Kong, C.; Huang, J.; Hu, Y.; Kang, X.; Kot, A.C. Beyond the prior forgery knowledge: Mining critical clues for general face forgery detection. *IEEE Trans. Inf. Forensics Secur.* **2023**, *19*, 1168–1182. [CrossRef]
16. Mohamed, M.H.; Khafagy, M.H.; Elbeh, H.; Abdalla, A.M. Sparsity and cold start recommendation system challenges solved by hybrid feedback. *Int. J. Eng. Res. Technol.* **2019**, *12*, 2734–2741.
17. Mohamed, M.H.; Ibrahim, L.F.; Elmenshawy, K.; Fadlallah, H.R. Adaptive learning systems based on ILOs of courses. *WSEAS Trans. Syst. Control* **2023**, *18*, 1–17. [CrossRef]
18. Mohamed, M.H.; Khafagy, M.H. Hash semi cascade join for joining multi-way map reduce. In Proceedings of the 2015 SAI Intelligent Systems Conference (IntelliSys), London, UK, 10–11 November 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 355–361.
19. Mohamed, M.H.; Khafagy, M.H.; Hasan, M. Music Recommendation System Used Emotions to Track and Change Negative Users' mood. *J. Theor. Appl. Inf. Technol.* **2021**, *99*, 4358–4376.
20. Sayed, A.R.; Khafagy, M.H.; Ali, M.; Mohamed, M.H. Predict student learning styles and suitable assessment methods using click stream. *Egypt. Inform. J.* **2024**, *26*, 100469. [CrossRef]
21. Shan, Y.; Hu, D.; Wang, Z. A Novel Truncated Norm Regularization Method for Multi-channel Color Image Denoising. *IEEE Trans. Circuits Syst. Video Technol.* **2024**. [CrossRef]

22. Liu, Y.; Yan, Z.; Tan, J.; Li, Y. Multi-purpose oriented single nighttime image haze removal based on unified variational retinex model. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *33*, 1643–1657. [CrossRef]
23. Liu, Y.; Yan, Z.; Chen, S.; Ye, T.; Ren, W.; Chen, E. Nighthazeformer: Single nighttime haze removal using prior query transformer. In Proceedings of the 31st ACM International Conference on Multimedia, Ottawa, ON, Canada, 29 October–3 November 2023; pp. 4119–4128.
24. Li, C.; Hu, E.; Zhang, X.; Zhou, H.; Xiong, H.; Liu, Y. Visibility restoration for real-world hazy images via improved physical model and Gaussian total variation. *Front. Comput. Sci.* **2024**, *18*, 1–3. [CrossRef]
25. Xiao, J.; Zhou, J.; Lei, J.; Xu, C.; Sui, H. Image hazing algorithm based on generative adversarial networks. *IEEE Access* **2019**, *8*, 15883–15894. [CrossRef]
26. Mustak, M.; Salminen, J.; Mäntymäki, M.; Rahman, A.; Dwivedi, Y.K. Deepfakes: Deceptions, mitigations, and opportunities. *J. Bus. Res.* **2023**, *154*, 113368. [CrossRef]
27. Traboulsi, N. *Deepfakes: Analysis of Threats and Countermeasures*; California State University: Fullerton, CA, USA, 2020.
28. Don, L. Advanced Cybersecurity Strategies: Leveraging Machine Learning for Deepfake and Malware Defense. Cameroon. 2024. Available online: https://easychair.org/publications/preprint/pN7Q (accessed on 16 April 2024).
29. Owaid, M.A.; Hammoodi, A.S. Evaluating Machine Learning and Deep Learning Models for Enhanced DDoS Attack Detection. *Math. Model. Eng. Probl.* **2024**, *11*, 493–499. [CrossRef]
30. Wazirali, R.; Yaghoubi, E.; Abujazar, M.S.S.; Ahmad, R.; Vakili, A.H. State-of-the-art review on energy and load forecasting in microgrids using artificial neural networks, machine learning, and deep learning techniques. *Electr. Power Syst. Res.* **2023**, *225*, 109792. [CrossRef]
31. Thippanna, D.G.; Priya, M.D.; Srinivas, T.A.S. An Effective Analysis of Image Processing with Deep Learning Algorithms. *Int. J. Comput. Appl.* **2023**, *975*, 8887. [CrossRef]
32. Hassini, K.; Khalis, S.; Habibi, O.; Chemmakha, M.; Lazaar, M. An end-to-end learning approach for enhancing intrusion detection in Industrial-Internet of Things. *Knowl. Based Syst.* **2024**, *294*, 111785. [CrossRef]
33. George, A.S.; George, A.S.H. Deepfakes: The Evolution of Hyper realistic Media Manipulation. *Partn. Univers. Innov. Res. Publ.* **2023**, *1*, 58–74.
34. Masood, M.; Nawaz, M.; Malik, K.M.; Javed, A.; Irtaza, A.; Malik, H. Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward. *Appl. Intell.* **2023**, *53*, 3974–4026. [CrossRef]
35. Namazli, P. Face Spoof Detection Using Convolutional Neural Networks. *Probl. Inf. Soc.* **2023**, *14*, 40–46.
36. Alkishri, W.; Widyarto, S.; Yousif, J.H.; Al-Bahri, M. Fake Face Detection Based on Colour Textual Analysis Using Deep Convolutional Neural Network. *J. Internet Serv. Inf. Secur.* **2023**, *13*, 143–155. [CrossRef]
37. Gupta, A.; Pandey, D. Unmasking the Illusion: Deepfake Detection through MesoNet. In Proceedings of the 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT), Greater Noida, India, 9–10 February 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1934–1938.
38. Sajith, S.; Pooja, A.; Ramesh, T.; Rajpal, P.; Roshna, A.R.; Ahammad, J. Anemia Identification from Blood Smear Images Using Deep Learning: An XAI Approach. In Proceedings of the 2023 International Conference on Recent Advances in Information Technology for Sustainable Development (ICRAIS), Manipal, India, 6–7 November 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 147–152.
39. Zhang, C.; Costa-Perez, X.; Patras, P. Adversarial attacks against deep learning-based network intrusion detection systems and defense mechanisms. *IEEE/ACM Trans. Netw.* **2022**, *30*, 1294–1311. [CrossRef]
40. Mohammed, A.; Kora, R. A comprehensive review on ensemble deep learning: Opportunities and challenges. *J. King Saud Univ.-Comput. Inf. Sci.* **2023**, *35*, 757–774. [CrossRef]
41. Khaleel, M.; Yaghoubi, E.; Yaghoubi, E.; Jahromi, M.Z. The role of mechanical energy storage systems based on artificial intelligence techniques in future sustainable energy systems. *Int. J. Electr. Eng. Sustain. (IJEES)* **2023**, *1*, 1–31.
42. Zhang, L.; Zhao, D.; Lim, C.P.; Asadi, H.; Huang, H.; Yu, Y.; Gao, R. Video Deepfake Classification Using Particle Swarm Optimization-based Evolving Ensemble Models. *Knowl. Based Syst.* **2024**, *289*, 111461. [CrossRef]
43. Nailwal, S.; Singhal, S.; Singh, N.T.; Raza, A. Deepfake Detection: A Multi-Algorithmic and Multi-Modal Approach for Robust Detection and Analysis. In Proceedings of the 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), Chennai, India, 1–2 November 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–8.
44. Taji, K.; Sohail, A.; Shahzad, T.; Khan, B.S.; Khan, M.A.; Ouahada, K. An Ensemble Hybrid Framework: A Comparative Analysis of Metaheuristic Algorithms for Ensemble Hybrid CNN features for Plants Disease Classification. *IEEE Access* **2024**, *12*, 61886–61906. [CrossRef]
45. Passos, L.A.; Jodas, D.; Costa, K.A.P.; Júnior, L.A.S.; Rodrigues, D.; Del Ser, J.; Camacho, D.; Papa, J.P. A review of deep learning-based approaches for deepfake content detection. *Expert Syst.* **2022**. [CrossRef]
46. Bappy, J.H.; Roy-Chowdhury, A.K.; Bunk, J.; Nataraj, L.; Manjunath, B.S. Exploiting spatial structure for localizing manipulated image regions. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4970–4979.
47. Zhang, W.; Gu, X.; Tang, L.; Yin, Y.; Liu, D.; Zhang, Y. Application of machine learning, deep learning and optimization algorithms in geoengineering and geoscience: Comprehensive review and future challenge. *Gondwana Res.* **2022**, *109*, 1–17. [CrossRef]

48. Dolhansky, B.; Bitton, J.; Pflaum, B.; Lu, J.; Howes, R.; Wang, M.; Ferrer, C.C. The deepfake detection challenge (dfdc) dataset. *arXiv* **2020**, arXiv:2006.07397.
49. Zi, B.; Chang, M.; Chen, J.; Ma, X.; Jiang, Y.G. Wilddeepfake: A challenging real-world dataset for deepfake detection. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020.
50. Al Shalchi, N.F.A.; Rahebi, J. Human retinal optic disc detection with grasshopper optimization algorithm. *Multimed. Tools Appl.* **2022**, *81*, 24937–24955. [CrossRef]
51. Al-Safi, H.; Munilla, J.; Rahebi, J. Patient privacy in smart cities by blockchain technology and feature selection with Harris Hawks Optimization (HHO) algorithm and machine learning. *Multimed. Tools Appl.* **2022**, *81*, 8719–8743. [CrossRef]
52. Al-Safi, H.; Munilla, J.; Rahebi, J. Harris Hawks Optimization (HHO) Algorithm based on Artificial Neural Network for Heart Disease Diagnosis. In Proceedings of the 2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC), Tumkur, India, 3–4 December 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–5.
53. Al-Rahlawee, A.T.H.; Rahebi, J. Multilevel thresholding of images with improved Otsu thresholding by black widow optimization algorithm. *Multimed. Tools Appl.* **2021**, *80*, 28217–28243. [CrossRef]