Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Advanced Deepfake Detection Using Honey Badger Optimization and ELM Classifier

Munleef Quadir[1,2], Prateek Agarwal[3], Charu Gupta[4]

[1]*School of Computer Science & Engineering, Lovely Professional University, Punjab, India*

[2]*Department of Computer Science, Jazan University, Jazan, Kingdom of Saudi Arabia*

[4] *Department of Computer Science & Engineering, Bhagwan Parshuram Institute of Technology, Delhi, India*

[*]*Corresponding Author: Dr. Prateek Agrawal Email: prateek.agrawal@lpu.co.in*

ARTICLE INFO

Received: 18 Dec 2024

Revised: 10 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

Deepfake poses a significant threat in contemporary times as it has negative impact on society. It is incredibly difficult to distinguish manipulated faces from real ones, even when scrutinized closely. Deepfake often struggles to replicate natural human expressions and subtle facial movements accurately. Detection methods focusing on inconsistencies in facial expressions, unnatural movements, or mismatches between facial movements and the emotional context can identify manipulated content. The current approaches face challenges in handling post-processing effects such as compression, noise, and changes in lighting. There is a lack of extensive research addressing the detection of both audio and visual deepfake content. This paper introduces a novel model designed to identify deepfake content within video frames. Our model detects deepfake by splitting the videos into frames and extract the features. The spatio temporal features of the frames help us to identify multiple frames. To reduce resource utilization, the fused frames are given as input to the trained model. An optimization algorithm is used to find the optimal parameter and then final classification is done to identify real or fake using extreme learning classifier model. This model has distinguishing deepfake identification which shows an accuracy of 96.13% on Celeb-V1 dataset compared to existing methods such as MLP-CNN and Yolo InceptionResNetV2 XGBOOST.

Keywords: Celeb-V1, Classification, Deepfake, ELM, HBO, Optimization

1 INTRODUCTION

The popularity of smartphones and social platforms has leveraged the need to safeguard personal identity and be alerted to manipulated photos or videos of people. One type of synthetic audio-visual material created using artificial intelligence (AI) algorithms is called a "deepfake." Deepfake technology [Ismail et al., 2021] refers to AI-synthesized media that are hyper realistic and sometimes even indistinguishable from real media. They are created using deep

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

learning algorithms to morph the original video with someone else (effect). Deepfake creators obtain hundreds or thousands of images showing their target person smiling, scowling, or making various faces and body movements. These images are then fed into a computer learning system. The most sophisticated deepfake methods use this information to build a three-dimensional representation of the target's face. This requires a very large set of images to create a realistic-looking face, but once the model is created, it can be animated in many different ways. Advanced deep learning models and other machine learning techniques, which are trained on large datasets to precisely mimic the subtleties of human motions, voice patterns, and expressions, are the technological cornerstones of deepfake. This advancement in digital media technology has important ramifications for security, privacy, and information sharing, among other areas. As such, it is imperative to critically assess the ethical, legal, and societal consequences of these developments.

The rapid advancement and increasing sophistication of deepfake technology have raised a significant concern about the potential misuse, fraud, and violation of personal privacy. Deepfake detection is a challenging area in digital forensics [Siegel et al., 2021, Gaur, 2022]. Identifying real and fake videos is a serious concern that need to be addressed. Figure 1 represents a sample scenario of real and fake video frames. The deepfake detection technique involves a binary classification procedure where two distinct classes are identified: deepfake and original. These techniques are essential for distinguishing between real and manipulated photographs or videos, serving as a critical tool in identifying the authenticity of visual media content. Deepfake detection is not only crucial for identifying instances of manipulation but also plays a significant role in safeguarding against misinformation and deceptive practices. Deepfake often struggles to replicate natural human expressions and subtle facial movements accurately. Detection methods focusing on inconsistencies in facial expressions, unnatural movements, or mismatches between facial movements and the emotional context can identify manipulated content. The current approaches face challenges in handling post-processing effects such as compression, noise, and changes in lighting.

By adding multiclass, multi-label, and local classification/detection to the binary classification scheme, we can better handle the complexities of real-world classification procedure with two distinct classes: deepfake and original. Deepfake detection [Tolosana et al., 2020] involves the meticulous extraction of specific visual elements from an image or video to discern between authentic and artificially manipulated content. The categorization of deepfake detection techniques is based on the method used for feature extraction, which can be broadly classified into four groups: pixel-based, frequency domain, temporal analysis, and deep learning. By using these methods, professionals can increase the precision and dependability of their deepfake detection findings, giving them more confidence when examining films. Based on features that can be observed with the naked eye, such as blinking patterns, micro-expressions revealed through head position, and nuanced spacing between various facial features are meticulously analyzed. This allows the detection systems to accurately recognize and interpret the range of emotions and intentions expressed by individuals.

Current techniques use binary classification to detect deepfake videos. In addition, deepfake detection algorithms were tested in a controlled test environment and were not implemented in real-world scenarios. It takes more than

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

one label to detect a deepfake. To properly manage media forgeries that occur in the real world, binary classifications must be expanded to include multiclass, multi-label, and local classifications. Previous methods of identifying deepfakes are based on frame-by-frame binary classification, which determines how likely it is that a given frame is real or deepfake. This approach has two issues even though it's simple to comprehend. One major problem was that many Deepfake movies included temporal artifacts and real or deepfake frames usually appeared at regular intervals. The temporal consistency between frames was not explicitly taken into account. Secondly, there is a step that needs to be taken when a video-level integrity score is needed. We need to aggregate the scores over individual frames in order to generate this value. However, many data-driven deepfake detection methods, particularly those based on Deep Neural Networks (DNN), usually lack explainability since DNN models are opaque. As a result of their reliance on the signatures of already existing deepfake through the use of supervised and unsupervised machine learning (ML) approaches, current deepfake detectors are less likely to identify unknown deepfake. Every detection method, whether signature-based or anomaly-based, has advantages and disadvantages.

Videos are typically compressed and uploaded to save network bandwidth and protect user privacy. This practice, known as social media laundering, increases the likelihood of false positive detections—that is, the labelling of a legitimate video as deepfake—while also being harmful to the recovery of underlying manipulation traces. Thus far, social media laundering has had a major impact on the majority of data-driven, signal-level feature-based deepfake detection techniques. While there aren't as many simple, free, and open-source software tools available as there are for creating face-swapping videos, the situation will soon change because of the generation algorithms' increasing sophistication in detecting both audio and visual deepfakes [Chen et al., 2020]. The post-processing processes like as compression, noisy effects, light variations, and other characteristics are not well-suited to the current technologies. Furthermore, the majority of methods have concentrated on face-swap detection by taking advantage of its drawbacks, such as apparent artifacts. Apart from this, lip-synching and face re-enactment are two other forms of deepfake that are becoming more popular every day. The main contributions of this paper are summarized as follows:

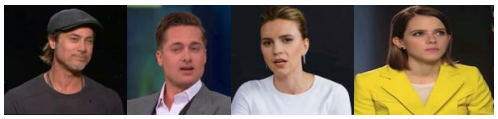
- To develop a deepfake detection model.
- To use Celeb-V1 dataset, to evaluate the effectiveness.
- To evaluate the performance of ELM classifier to classify deepfake videos with high speed and accuracy.
- To analyze and visualize the performance of classification result using a confusion matrix.

The remainder of this article is organized accordingly. In Section 2, we analyzed recent studies and reviewed them. Section 3 explains the proposed methodology in detail. Section 4 discusses classification model, followed by experimental setup and outcomes in Section 5. The conclusions of the study are stated in Section 6.

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Fake Videos



Real Videos

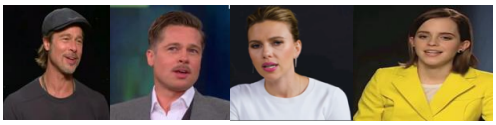


Figure 1: Samples from the dataset for real and fake video.

2 RELATED WORKS

Deep learning methods have been widely employed for deep fake detection. Most of the existing studies do not focus on extracting the feature of video frames. Numerous works has been reported on deep fake detection as discussed below:

2.1 Local Feature Based Deepfake Detection

When compared to visual features, the local feature-based technique have a better level of reliability. [Ramadhani et al., 2020] the authors compiled a small number of research that extract the properties of individual pixels using a pixel-based feature extraction technique. While some research employed steganalysis features and images, others used a smaller sample set and Photo Response Non-Uniformity (PRNU) analytic. Scale Invariant Feature Transform (SIFT) feature extraction is a better method for extracting important points from a picture than PRNU. This approach is thought to be more trustworthy than earlier ones. In order to distinguish between real and fake images or videos, deep-feature based deepfake detection use numerous layers to extract complicated properties while still performing pixel-level feature extraction. While many feature extraction techniques exist, including Local Binary Pattern (LBP) [Abdullah and Ali., 2023], Binary Gabor pattern (BGP), Binarized Statistical Image Features (BSIF), Local Phase Quantization (LPQ), Pyramid of Histogram of Oriented Gradients (PHOG), Speeded Up Robust Feature (SURF), and Image Quality Metric (IQM). The existing studies concluded that the local feature detection performs better in IQM.

2.2 Deep Feature Based Deepfake Detection

The deepfake detection method based on local features performs rather well in detecting deepfakes, but as the algorithm advances, deepfake detection gets harder. On the other hand, deep layer neural networks, which extract sophisticated features than local ones, form the foundation of deepfake detection.

According to the study discussed [Saikia et al., 2022], a new architecture for detection and classification of deep fake video has been proposed which utilized convolutional Long Short-Term Memory (LSTM) network on each block of deep fake image. This gave 97% accuracy for classifying 3 category video frames, namely deep fake, manipulated, original. After that, it used EfficientNet for video classification on the video sequence and gave state-of-the-art result of 98% accuracy on UCL dataset for original verification. In the study [Haiwei et al., 2022] adaptive over complete slice wavelet transform has been applied on 2D slices of video frames to capture the fine details in images. Furthermore, Convolutional neural network (CNN) has been applied to the slices of video frames to capture the fine details in images. Subsequently CNN has been applied to the coefficient of wavelet modified max pooling. Following a series of studies, it produced encouraging outcomes that are on par with any cutting-edge technique for detecting fakes.

2.3 Spatio-Temporal Feature Based Deepfake Detection

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Spatio-Temporal features driven deepfake identification involves gathering features from multiple consecutive frames to capture the temporal frames of the video. The study [Nguyen et al., 2021] discussed a learning model based on 3-D CNN which explores both spatial and temporal data from a frame sequence, whereas most of the existing works focus on either spatial or temporal data. Even though this deep learning method gives greater accuracy further research could be done on different facial reenactments. Face Forensics++ and VidTIMIT dataset are taken to evaluate the model and the experimental results shows the easiest detection is possible in other datasets compared to Face Forensics++.

Another approach [Haiwei et al., 2022] explores spatio-temporal features to identify the forged regions. The face micro expressions and geometric motion are explored using Spatial-Temporal Deepfake Detection and Localization (ST-DDL) and Anchor-Mesh Motion (AMM) for feature extraction. The computational cost in evaluating the entire face has made it easy to find triangular face by setting the anchor points. The Fusion Attention method integrated in ST-DDL improves the learning accuracy. A newly created non reproducible deepfake dataset namely ManualFake that hold the forged videos from different models in addition to the data from commercial software. Further studies referred to the use data augmentation along with ST-DDL to improve the robustness. However, the study showed that it couldn't detect interpolated frames and AMM extracts limited motion clue.

A few researchers [Kolagati et al., 2022] explored combination of Multilayer Perceptron (MLP) and CNN to analyze the temporal features. The study used FaceForensics++ dataset and the Deepfake Detection Challenge (DFDC) dataset. The study relied on exploring inconsistency in facial features using MLP and feature extraction using CNN. The hybrid model provides a good accuracy and faster speed with limited computational resources. The research could be further explored to use the model in a more balanced dataset. Additionally, there is a need for further research to improve the detection of faces in dark environments and methods to improve face warping.

To reliably determine a video's authenticity, a hybrid deep learning technique that models both intra- and inter-frame information is being researched [Saikia et al., 2022]. To aid in the extraction of the temporal information, they also used optical flow, a conventional technique for temporal feature analysis. The technique utilizes potential inter-frame dissimilarities and is based on the characterization of the subject's face movements for the optical flow implementation. The study assessed a number of performance metrics, including F1-score, AUC, accuracy, recall, and precision. The model uses CNN and LSTM model to classify the videos based on larger set of frames. Table 1 shows the comparison of the previous research based on different features.

Table 1: State of art comparison

Author & Year	Local Feature	Deep Feature	Spatio temporal Feature
Ramadhani et al., 2020	√	x	X
Saikia et al., 2022	X	√	√
Abdullah and Ali., 2023	√	x	X
Haiwei et al., 2022	X	x	√
Nguyen et al., 2021	X	√	X
Kolagati et al., 2022	X	x	√

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Chen et al., 2020	√	x	X

Our research is made to determine fake content of a person that is generated using various generative model approaches without specifically focusing on each of the approaches. A general solution will be far more beneficial to prevent the spread of fake content. We focus on a specific generative model approach to generate our fake contents. By using a general solution method, we hope that our work can be a guideline for future prevention or mitigation of fake contents that are generated using various generative model approaches.

3 METHODOLOGY

The proposed methodology for the validation of deepfake detection is discussed in this section and is shown in Figure 2. It includes data collection, pre-processing, facial detection, feature extraction, and classification. The paramount objective of this model is to differentiate between real facial images and deepfake representations, a task of considerable significance amid the escalating emergence of deepfake technologies.

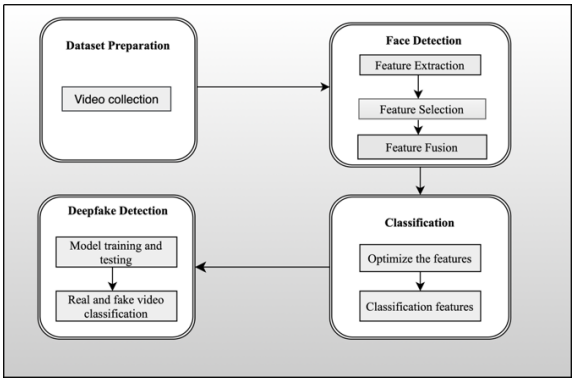


Figure 2: Phases of Deepfake detection

3.1 Dataset Preparation

The preliminary stage in any machine learning model is data preparation. The diversity of the training data is crucial for building robust deepfake detection models. Training datasets should include a wide range of demographics, lighting conditions, resolutions, and scenarios to ensure the model can generalize well to unseen deepfake.

The celebrity video data set (Celeb-DF v1) [Li et al., 2023] is an edited version of the Celeb-DF dataset and is the largest publicly available deepfake detection dataset. This dataset provides a higher resolution RGB frames, and compressed versions of the frames and post-processed videos. The dataset provides a total of 521 videos. It is split into frames and provides information on whether each frame is extracted from the original source, or if it is edited in some way. This allows for binary classification of each video, where label=0 denotes only real frames were extracted, and label=1 denotes some of the frames are edited. This dataset also provides a train-validation-test split configuration. Celeb-DF v1 has a train set of 400 videos (181 are fake, 219 are real), a validation set of 51 videos (26 fake, 25 real), and a test set of 70 videos (26 fake, 44 real) as in Table 2. The data is labeled as 0 and 1 to identify as real and fake videos. The dataset consisted of a diverse collection of deepfake and real face images, providing an ideal environment for assessing the capabilities of our model. This dataset is particularly challenging for deepfake detection methods, because it contains videos from a bigger variety of sources and contains compressed as well as full resolution videos. Furthermore, the higher number of real videos compared to fake videos in the train and test sets is representative of the real-world scenario and ideal for testing a developed deepfake detection method. However, further testing on more diverse datasets may be necessary to validate its generalizability.

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 2: Celeb-DF v1 Dataset

Celeb-DF v1 Dataset	Training Dataset	Validation set	Test Dataset
Number of videos	400	51	70
Real Videos	219	25	44
Fake Videos	181	26	26

3.2 Face Detection

3.2.1 Preprocessing

The input dataset needs to be prepared for subsequent analyses and processing tasks. To further reduce the noisy data, Bilateral Filtering [Zou et al., 2020] image processing technique is highly relevant to prepare the data. One important thing is this technique may be time consuming algorithm, but still used for a number of promising applications. The Bilateral Filtering (BF) equation can be expressed as follows,

$$BF = 1/Wp \sum\_q\epsilon s \; [\![G\sigma\_s \; (p-q) G\sigma\_r \quad (Ip-Iq) Iq]\!] \qquad (1)$$

From the (1), we infer the different variables as Wp is normalization factor, "G" "σ" _"s" is spatial Gaussian Gσ_r is range Gaussian, I for intensity, p and q represent pixels.

Moreover, this technique considers the spatial complexity and pixel value intensity and maintains the consistency even after noise reduction. Subsequently, log based contrast enhancement improves the clarity and visual quality of the image even in poor lighting conditions. The log based contrast enhancement can be expressed as represented,

$$S = c \log (1+r) \qquad (2)$$

where, c represents the scaling constant to adopt intensity value, r is the input intensity, s is output intensity, and the logarithmic base value is also considered. However, the log-based technique adjusts the contrast to increase the clarity of image.

3.2.2. Face Detection

For the face detection stage, we adopt the Multi-Task Cascade Neural Network (MTCNN) model. MTCNN is a state-of-the-art deep learning model that excels in accurately detecting and aligning faces in images. MTCNN is more efficient in terms of both outcome and CPU consumption since it can handle face identification and alignment problems simultaneously. MTCNN will detect the face and align it with a bounding box and 5 facial points (eye – left and right, nose, and mouth – left and right). This feature extraction result will be useful for aligning face images to be processed again later.

MTCNN follows a three-stage cascaded architecture as in Figure 3, such as a proposal network (P-Net), where an initial set of candidates bounding boxes is generated using a combination of convolutional and max-pooling layers; the refinement network(R-Net) uses convolutional layers to extract the features and adjust the coordinates to refine the coordinate accuracy, and finally alignment results (O-Net) to standardize the features in a consistent position. However, MTCNN does not perform well in the case of tiny faces because candidate bouncing is generated through a shallow CNN [Li, X., Yang, Z., & Wu, H., 2020]. The accuracy and localization of the images can be detected with MTCNN, enabling further analysis and classification to identify potential deepfake content. MTCNN detects face and uses a combination of deep learning models for feature extraction.

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

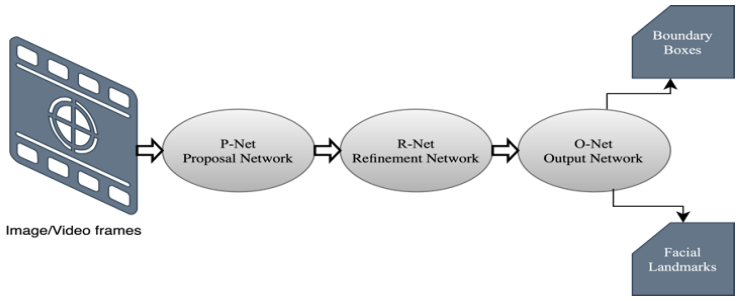
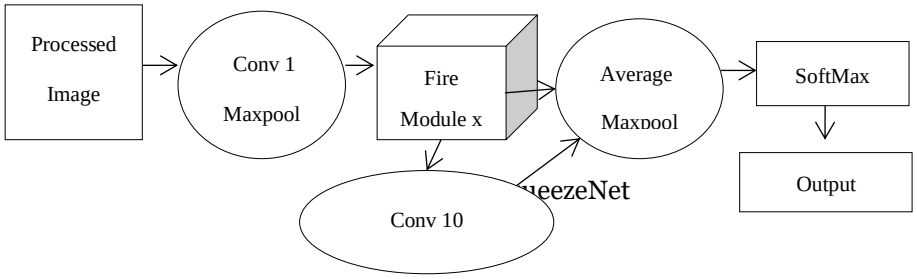


Figure 3: Please add an explanation for bold/italics/underline/color in the table footer

3.2.3 SqueezeNet

SqueezeNet is a small CNN architecture widely used for feature extraction as they are light weight architectures compared to traditional CNNs. They require less memory and can be trained easily, thus incurring less computational cost. The SqueezeNet architecture, as in figure 4, uses a squeeze layer with a 1 × 1 filter and an expanded layer. The building block of SqueezeNet comprises a combination of 1 × 1 and 3 × 3 filters, called fire modules. The image passes through a single convolutional layer and then to a combination of filters. Dropout layers are added after the Fire9 module to reduce over fitting. Due to the late use of down sampling, SqueezeNet features a "complex" bypass.



3.2.4 MixNet

MixNet extracts high-quality images from the low-light images. The low-level features are extracted with the respective dimensions, such as H representing Height, W for width, and C for channel. Subsequently, multiple stacked feature mixing blocks were used to generate high-quality images. In the final round, all extracted features are summed to restore the image, as shown in figure 5.

Both SqueezeNet and MixNet are chosen as they have a lower number of parameters and a simpler model, in turn saving computational resources as well as allocated memory.

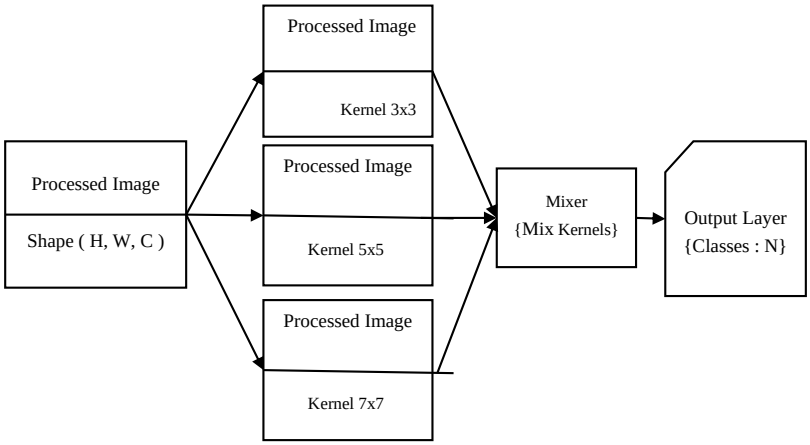


Figure 5: MixNet

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3.2.4 DTCWT

Dual Tree Complex Wavelet Transform (DTCWT) [Gao and Yang ,2021] functions as a merging process that selects the optimal inputs and transmits them to the subsequent stage. DTCWT is chosen as it has a better image fusion result compared to other methods. DTCWT can be applied to 1, 2, or 3-dimensional data. This is done by transforming the image into a tree. Each node in the tree corresponds to several low pass and high pass sub bands that can be complex. A forward and backward transform can be applied to the image using the tree data structure in order to obtain the transformed low pass and high pass sub bands at a user-defined level. This is known as a perfect reconstruction. Figure 6 demonstrates the workflow of DTCWT.

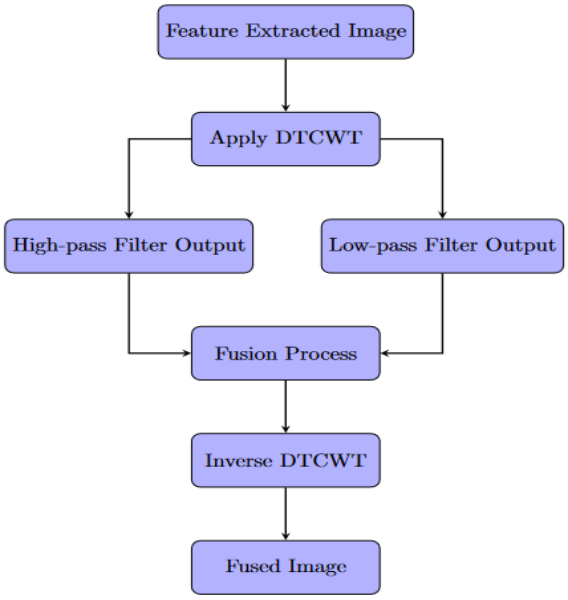


Figure 6: DTCWT

Following face detection, the feature extraction and fusion process is critical for recognizing and distinguishing real and deep fake faces. We employ a combination of three deep learning models, SqueezeNet, MixNet, and DTCWT. These models extract meaningful features from the detected faces and fuse them to create a comprehensive feature representation for each image.

4 CLASSIFICATION

To make the final decision on whether a given image is a real face or a deepfake, we utilize the extreme learning machine (ELM) model. ELM is chosen for its efficiency and effectiveness in classification tasks. The fused features obtained from the previous stage are input into the ELM model for classification.

ELM classifier [Nahiduzzaman et al .,2023] is a single layer feed forward network (SLFN) that have faster learning speed and smallest training error. The classifier gives good performance for functional and non-functional approximation. Most of the learning algorithms choose traditional methods to choose parameters to train the dataset. ELM classifier can handle complex patterned input samples and models that are difficult to identify using traditional methods.

4.1 Honey Badger Optimization HBO

HBO (Honey Badger Optimization) [Agarwal et al ., 2020] is a nature inspired algorithm that is based on meta heuristics optimization technique. The algorithm acts as an efficient search mechanism inspired by the behaviour of Honey Badger (HB). The search strategy follows exploration and exploitation. Honey badger lives in self-dug tunnels

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

and search for food in nest and beehives as interpreted by the bird's direction. HBO is used to optimize the features. The mathematical representation is as in (3)

$$P_j=Lowerlimit_j+r_1+(Upperlimit_j-Lowerlimit_j) \quad (3)$$

where pj is the position of honey badger, upperlimitj, lowerlimitj designates the upper and lower limit of search positions, and r is a random number between zero and 1.

Intensity Factor (IF) decreases with iteration and is the distance between the prey and HB and as in (4).

.

$$F=r_2\ (S/(4*PI*d*d)) \quad (4)$$

where S is source strength, S= (pi-pi+1)2, r2 is a random number between 0 and 1.

Density Factor (DF) is the transition from exploration to exploitation and ,decreases with iteration as in (5)

$$DF=C*exp((-t)/MaxT) \quad (5)$$

where C(constant)>=1,t-current iteration, T-maximum iterations

The distance between the target and badger is calculated as d2=pprey-pi, where pprey is the target position and pi is the honey badger.

There are two stages involved in updating an agent's position: the digging phase and the honey phase. In digging phase honey badger uses its smelling ability to locate the target in Cardioid motion and move around the target(F) and position according to the prey(pprey). The target search depends on the value of F as 1 or -1, if r6<0.5.

$$data=|\cos⟦(2\pi r_4)*[1-\cos(2\pi r_5)⟧ \quad |$$

$$P_New=P_Prey+F*\beta*I*P_Prey+F(r_3*DF*d_i*data) \quad (6)$$

The honey badger tracks the bird to find the beehives during the honey phase that is updated as new position (PNew) which is represented as,

$$P_New=P_Prey+F*r_7*DF*d_i \quad (7)$$

The honey badger would reach close to the prey and thus it saves the position. To augment the precision of classification accuracy, the HBO algorithm is applied for the refinement of the model's parameters. This optimization strategy is instrumental in the attainment of an optimal parameter configuration, thereby ensuring the ELM model's enhanced performance in the detection of deepfake images

4.2 Extreme Learning Classifier

ELM act as the base estimator, along with the defined parameter grid, accuracy scoring and other configuration parameters. Then create an instance of the HBO model. It uses the parameters control how HBO searches for the best ELM classifier. This specifically train and evaluate a model using ELM classifier and HBO to improve accuracy. HBO will explore different ELM models with varying neuron counts within this range to find the best performing one. More iteration may give better results but its time consuming.

Pseudo code

Input Dataset: Celeb v1

Output: ELM HBO model Classification

a. Setup the following variables
b. n-particles: Number of ELM models evaluated simultaneously during optimization (default 5).

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

c. cv: Cross-validation method used for evaluating models within HBO (here, Stratified K-Fold with splits).

d. Verbose: Controls the level of logging information printed during optimization (default 1 for some messages).

e. iterations: Number of iterations for HBO to search for the best model (default 10).

f. Split the videos to frames

g. Performs face detection on each frame.

h. Extracts deep features using SqueezeNet and MixNet

i. Combine features using DTCWT

j. Create classifier model using ELM

k. Define a search space for the HBO and optimize the parameters

l. Iteratively search towards the best performing model

m. Classifies the extracted features as "Real" or "Fake" using HBO

This structured methodology underscores our comprehensive approach towards developing a robust model capable of effectively distinguishing between real and counterfeit facial representations, thereby contributing to the on-going efforts in mitigating the challenges posed by the advent of deepfake technology

4.3 Deepfake Detection

Figure 7 shows the architecture of the proposed model where initially a dataset containing a collection of video files is used to train and evaluate the deep learning models. The face detection stage utilizes a pre-trained MTCNN model to detect and isolate faces within video frames. Feature extraction stage employs two parallel deep learning networks, SqueezeNet and MixNet, to extract unique features from the detected faces. These features are crucial for differentiating real from fake images. The next level to extract features from both SqueezeNet and MixNet are combined using DTCWT. This fusion process creates a more comprehensive feature set that can potentially improve classification accuracy. The final stage involves classifying test dataset using ELM as real or fake and integrates HBO to optimize its performance.

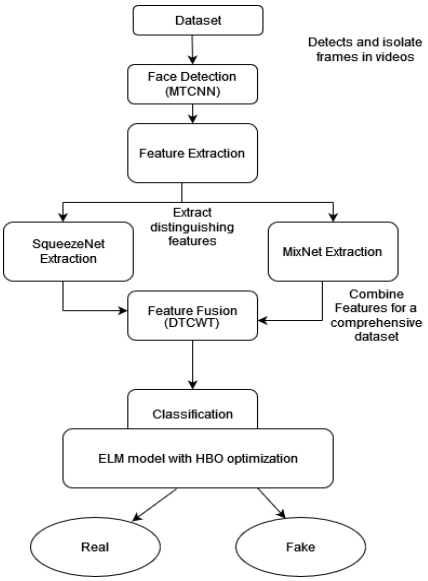


Figure 7: Proposed HBO ELM Model

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

5 RESULTS AND DISCUSSION

5.1 Extreme Learning Classifier

To analyze the performance of our proposed method we run the experiment on Jupyter notebook using python. The experiment is implemented using64 bit Intel® i7-9750 Hz CPU system with the internal memory of 8 GB RAM. In this section, we describe the evaluation of the DFFDR-HBODL model using the Celeb-DF v1 dataset. To evaluate the performance of our DFFDR-HBODL model, extensive experiments were conducted using the Celeb-DF v1 dataset, which is a well-established benchmark for deepfake detection.

Using a combination of our proposed Honey Badger algorithm and ELM, we demonstrated competitive accuracies in telling apart fake videos from real videos at about 96% correct classification using k-fold cross validation compared with current literature stating around that 96% accuracy is necessary to achieve real-time detection. Figure 8 shows some sample result generated from the experiment. Figure 9 shows a sample output generated after the implementation of the model.

Fake

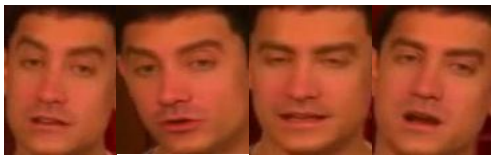


Figure 8: Please add an explanation for bold/italics/underline/color in the

Real

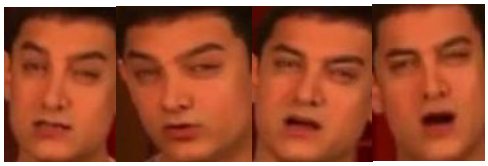


Figure 9: Samples results obtained based on proposed method

5.2 Performance Metrics

Evaluation metrics are used to predict the accuracy and how the learned data works well with the learning algorithms to predict the accuracy of the model. The following evaluation parameters are considered such as accuracy, precision, recall, F1-score, Specificity, and MCC. All the metrics are evaluated based on the following equations:

The ratio of the total number of guesses to the number of right predictions is known as accuracy

$$\text{Accuracy}=(TP+TN)/(TP+TN+FP+FN) \quad (8)$$

The ratio of accurate positive forecasts to total positive predictions, expressed in positive anticipated values, is known as precision.

$$\text{Precision}=(TP+TN)/(TP+FP) \quad (9)$$

Recall is a metric used to quantify model sensitivity. It can be defined as the proportion of accurately predicted outcomes to positive predictions.

$$\text{Recall}=TP/(TP+FN) \quad (10)$$

The F1-Score, which is computed as the harmonic mean of precision and recall, is a measure of the model's test

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

accuracy.

$$F\_1\text{-Score}=2*(\text{Precision}*\text{Recall})/(\text{Precision}+\text{Recall}) \quad (11)$$

Specificity of a model refers to its ability to correctly identify true negatives, which means that some actual negatives may be incorrectly classified as positives, known as false positives. This is also referred to as the True Negative Rate.

$$\text{Specificity}=TN/(TN+FP) \quad (12)$$

MCC is a single value classification statistic tool that is used to summarize confusion matrix. The value ranges from -1 to +1.

$$MCC=(TN*TP-FN*FP)/\sqrt{((TP+FP)(TP+FN)+(TN+FP)(TN+FN))} \quad (13)$$

Confusion matrix is used to compare the actual value with the predicted value from the data as in Figure 10. The matrix values, namely, true positive, false positives, and true negatives.

From the ROC curve (Figure 11) the area under curve (AUC) can be determined. The prosed model has an accuracy score of 0.99, which has 99% of efficiently classifying real and fake videos.

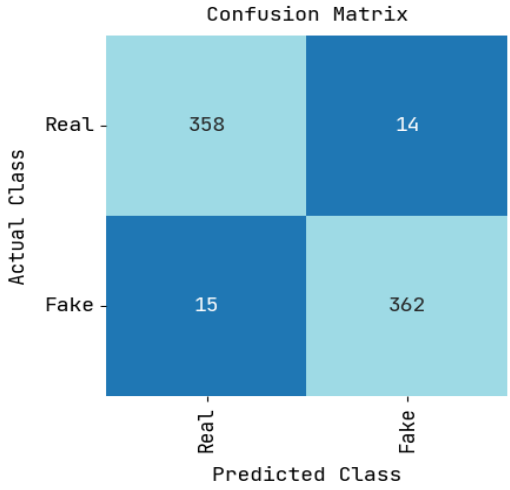


Figure 10: Confusion Matrix

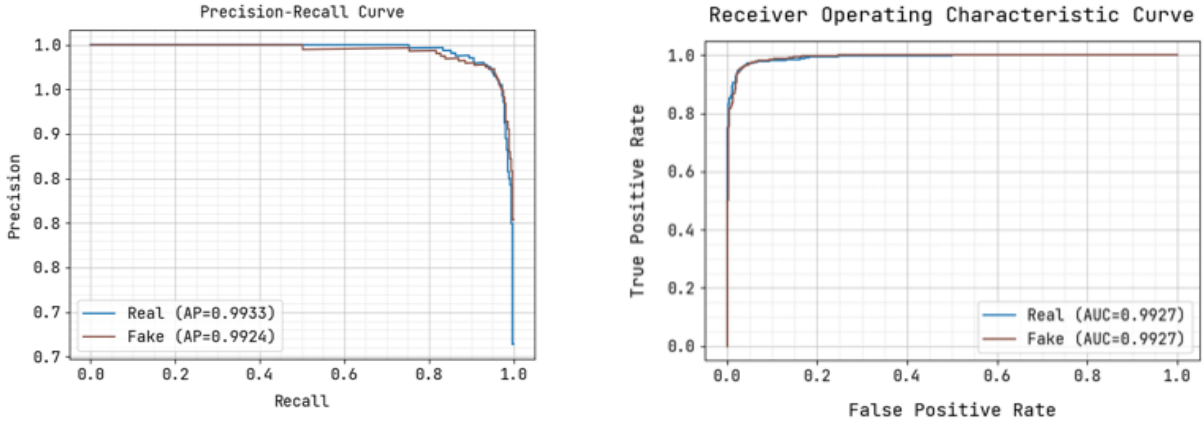


Figure 11: ROC Curve

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 3: Performance Metrics

Metrics	Values
Accuracy	0.9613
Precision	0.9613
Recall	0.9613
Specificity	0.9613
ROC AUC Score	0.9927
F-Score	0.9613
Mathews Correlation Coefficient	0.9226

The experimental results of our DFFDR-HBODL model are highly encouraging and demonstrate its effectiveness in deepfake face detection. The model's accuracy score, well above 0.9613, highlights its proficiency in distinguishing between real and deepfake faces. This level of accuracy is essential for real-world applications where the consequences of misidentification can be severe. Achieving a high precision score is vital to minimize false positives, especially in sensitive applications.

5.3 Comparison with existing works

To further verify the performance of proposed model, comparison with existing methods were performed. The two models for deepfake detection compared include MLP-CNN and Yolo InceptionResNetV2 XGBOOST. The detection accuracy is of 0.9613 shows improved performance in classifying deep fakes.

Table 4: Comparison with state of art models

Model	Accuracy	Precision	Recall	Specificity	AUC Score	F-Score
MLP-CNN [Kolagati et al., 2022]	0.83	0.87	0.84	0.84	0.87	0.87
Yolo InceptionResNetV2 XGBOOST [Ismail et al., 2021]	0.9073	0.8736	0.8539	0.9353	0.906	0.8636
HBO-ELM (proposed)	0.9613	0.9613	0.9613	0.9613	0.9927	0.9613

Our model's balanced precision and recall demonstrate its ability to maintain low false positives while detecting deepfake images effectively. The promising performance of the DFFDR-HBODL model underscores its potential for various practical applications. This technology can enhance security, safeguard online content, and protect against the proliferation of deepfake content on the internet. The use of the Celeb-DF v1 dataset, which includes a wide variety of deepfake and real face images, ensures that our model's performance is robust across different scenarios. However, further testing on more diverse datasets may be necessary to validate its generalizability. The application of HBO algorithm for parameter tuning in ELM has significantly improved the model's classification results. This optimization process is vital for achieving the highest level of accuracy.

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

To sum up, the DFFDR-HBODL model represents a robust and efficient solution for deep fake face detection. The combination of advanced pre-processing, cutting edge deep learning models, and optimization provided by the HBO algorithm results in a highly accurate and reliable system. Our test findings using the Celeb-DF v1 dataset show that the model can successfully discriminate between faces that are real and those that are deepfake. We believe that this work contributes significantly to the field of deep fake detection and holds great potential for applications in various domains, including cyber security and online content verification. The model's performance was assessed using the binary classification standard assessment metrics of accuracy, precision, recall, and F1-score. Because it doesn't require patches or extra data from AI-generated media, the system may be used as a detection tool for real-world media and achieves excellent detection accuracy

6 CONCLUSION

In this study, the HBO algorithm along with the ELM classifier technique was implemented for the deepfake detection of deepfake images. The suggested approach performed more quickly and effectively in differentiating between authentic and fraudulent videos. SqueezeNet and MixNet were used to investigate deepfake video alterations. Pre-processing was performed using MTCNN. The proposed method shows an accuracy of 96.13% for the Celeb-DF dataset. Although the DFFDR-HBODL model has shown considerable promise, we recognize that the field of deepfake technology is continuously evolving. . In the future, we will extend this to several classifiers and use different distant metric measures to detect deepfake videos. To stay ahead of emerging challenges, we plan to conduct further research and development to enhance our model's performance and adapt it to evolving deepfake techniques. The possibility of expanding the methodology to other datasets for comprehensive validation is open for further research.

As deepfake generation and detection is emerging, so too will the methods used by malevolent actors to alter media content. Therefore, a pertinent concern moving forward is how to design detection methods in a way which accounts for the unknown possibilities of future deepfake generation methods. This would be done through developing software which can detect specific characteristics of a given deepfake. With the development of specific deepfake detection tools, it will be much more realistic to keep up with future deepfake generation methods.

REFERENCES

[1] Ismail, A., Elpeltagy, M., S. Zaki, M., & Eldahshan, K. (2021). A new deep learning-based methodology for video deepfake detection using XGBoost. Sensors, 21(16), 5413.

[2] Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. Neurocomputing, 70(1-3), 489-501.

[3] Hashim, F. A., Houssein, E. H., Hussain, K., Mabrouk, M. S., & Al-Atabany, W. (2022). Honey Badger Algorithm: New metaheuristic algorithm for solving optimization problems. Mathematics and Computers in Simulation, 192, 84-110.

[4] Agarwal, H. Farid, T. El-Gaaly, and S.-N. Lim, 'Detecting Deep-Fake Videos from Appearance and Behavior', in 2020 IEEE International Workshop on Information Forensics and Security (WIFS), New York, NY, USA: IEEE, Dec. 2020, pp. 1–6. doi: 10.1109/WIFS49906.2020.9 360904.

[5] Ranjan, P., Patil, S., & Kazi, F. (2020, March). Improved generalizability of deep-fakes detection using transfer learning based CNN framework. In 2020 3rd international conference on information and computer technologies (ICICT) (pp. 86-90). IEEE.

[6] Ramadhani, K. N., & Munir, R. (2020, November). A comparative study of deepfake video detection method. In 2020 3rd International Conference on Information and Communications Technology (ICOIACT) (pp. 394-399). IEEE.

[7] Nguyen, X. H., Tran, T. S., Nguyen, K. D., & Truong, D. T. (2021). Learning spatio-temporal features to detect manipulated facial videos created by the deepfake techniques. Forensic Science International: Digital Investigation,

36, 301108.

[8] Haiwei, W., Jiantao, Z., Shile, Z., & Jinyu, T. (2022). Exploring spatial-temporal features for deepfake detection and localization. arXiv preprint arXiv:2210.15872.

[9] Nahiduzzaman, M., Islam, M. R., Goni, M. O. F., Anower, M. S., Ahsan, M., Haider, J., & Kowalski, M. (2023). Diabetic retinopathy identification using parallel convolutional neural network based feature extractor and ELM classifier. Expert Systems with Applications, 217, 119557.

[10] Kolagati, S., Priyadharshini, T., & Rajam, V. M. A. (2022). Exposing deepfakes using a deep multilayer perceptron–convolutional neural network model. International Journal of Information Management Data Insights, 2(1), 100054.

[11] Saikia, P., Dholaria, D., Yadav, P., Patel, V., & Roy, M. (2022, July). A hybrid CNN-LSTM model for video deepfake detection by leveraging optical flow features. In 2022 international joint conference on neural networks (IJCNN) (pp. 1-7). IEEE.

[12] Zou, B., Qiu, H., & Lu, Y. (2020). Point cloud reduction and denoising based on optimized downsampling and bilateral filtering. Ieee Access, 8, 136316-136326.

[13] Siegel, D., Kraetzer, C., Seidlitz, S., & Dittmann, J. (2021). Media forensics considerations on deepfake detection with hand-crafted features. Journal of Imaging, 7(7), 108.

[14] Pokroy, A. A., & Egorov, A. D. (2021, January). EfficientNets for deepfake detection: Comparison of pretrained models. In 2021 IEEE conference of russian young researchers in electrical and electronic engineering (ElConRus) (pp. 598-600). IEEE.

[15] Kaddar, B., Fezza, S. A., Hamidouche, W., Akhtar, Z., & Hadid, A. (2023). On the effectiveness of handcrafted features for deepfake video detection. Journal of Electronic Imaging, 32(5), 053033-053033.

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

[17] Chen, T., Kumar, A., Nagarsheth, P., Sivaraman, G., & Khoury, E. (2020, November). Generalization of Audio Deepfake Detection. In Odyssey (pp. 132-137).

[18] Gaur, L. (Ed.). (2022). DeepFakes: Creation, Detection, and Impact. CRC Press.

[19] Abdullah, M. T., & Ali, N. H. M. (2023). DeepFake Detection Improvement for Images Based on a Proposed Method for Local Binary Pattern of the Multiple-Channel Color Space. International Journal of Intelligent Engineering & Systems, 16(3).

[20] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. Information Fusion, 64, 131-148.

[21] Gao, S., Xia, M., & Yang, G. (2021). Dual-Tree Complex Wavelet Transform-Based Direction Correlation for Face Forgery Detection. Security and Communication Networks, 2021, 1-10.