

Deepfake Detection Using Artificial Intelligence and Convolutional Neural Networks.

Sintu Kumar, Ravi Kumar, Prof. (Dr.) Renu Bagoria

Jagannath University, Jaipur, Rajasthan, India

sintukumar852191@gmail.com

Abstract

Deepfake technology is getting better and better because of improvements in artificial intelligence and deep learning algorithms. Now it is possible to make videos and pictures that look very real so it is hard to tell if they are fake just by looking at them. People are using deepfakes in ways on social media in politics and in cybercrime, which is causing big problems with false information and privacy.

This research paper is about a system that uses a special kind of computer program called a convolutional neural network to detect deepfakes. The system looks at parts of a persons face like their eyes and nose to figure out if a video is fake or real. We use a different steps to make this work including taking frames from videos getting rid of extra information finding faces and using the computer program to learn about the faces.

We look at the face the eyes and the nose to make sure we get it right. We use videos from datasets like FaceForensics++ and DFDC to train and test our system. When we try out our system it works well and can tell if a video is fake or real. The system is good at detecting videos and does not make many mistakes.

This system can help us make sure videos and pictures are real which is important, for keeping us safe on the internet.

Keywords— Deepfake Detection, Convolutional Neural Network, Artificial Intelligence, Computer Vision, Face Recognition, Deep Learning.

Introduction

Digital media has changed a lot because of intelligence. One of the things about it is deepfake technology. This technology uses techniques to make fake videos, audio and images. It is so good that it can switch one persons face with another persons face in a video. These fake videos look really real. Can fool people.

Now that it is easier to make deepfakes people are getting worried about cybersecurity, politics, movies, news and social media. Fake videos can be used to spread information hurt peoples reputations change what people think and cheat people. As these fake videos get better the ways of checking if a video is real or not are not working anymore.

There are some computer methods, like Convolutional Neural Networks that are really good at looking at pictures and figuring out what is in them. These methods can automatically find things in pictures that're not easy to see. This can help us find out if a video is a deepfake or not.

This research is about making a system that can detect deepfakes. This system uses computer methods to look at different parts of a persons face like the eyes, nose and the whole face. It does a things to check if a video is real or not. The system is better, at detecting deepfakes because it looks at parts of the face not just the whole face. This makes it more reliable.

Literature Review

Y. Li, M. Chang, and S. Lyu (2018) proposed a way to spot deepfakes by looking at how eyes blink. They used computer programs called Convolutional Neural Networks (CNNs) and Long-term Recurrent Convolutional Networks (LRCN). Their method checked for eye blinking in fake videos. This is because early deepfake makers struggled to make eyes blink naturally. The approach worked well in finding inconsistencies in videos.. It mainly

looked at eye movements. It did not work well for very advanced deepfake videos. The system was good, at detecting inconsistencies.. Deepfake videos are getting better. So the system needs to be improved to keep up.

However this can be achieved by focusing on deepfakes. Their eye blinking patterns and Convolutional Neural Networks (CNNs) and Long-term Recurrent Convolutional Networks (LRCN) can help in deepfake detection.

A. Karandikar, V. Deshpande, and S. Singh (2020) developed a CNN-based deepfake detection framework using facial extraction and image classification techniques. Their research utilized VGG-16 architecture for feature extraction and classification of real and fake videos. The model successfully identified manipulated facial textures and visual inconsistencies. However, the system performance decreased when low-resolution or compressed videos were used.

D. Wodajo and S. Atnafu (2021) introduced a Convolutional Vision Transformer (CViT)-based deepfake detection model. Their approach combined CNN-based local feature extraction with transformer-based global attention mechanisms. The proposed hybrid architecture improved feature learning capability and achieved high accuracy on benchmark datasets such as FaceForensics++. Despite its strong performance, the model required high computational resources and GPU support during training.

Nikhil Sharma and Ram Kumar Sharma (2025) They came up with a deepfake detection system that uses a CNN. This system uses the DeepFake Detection Challenge dataset. To make this system work they did a things. They looked at each frame cleaned up the images made sure everything was the same and then used CNN to figure out if an image was real or fake. The DeepFake Detection Challenge dataset was really important for this. They used it to test their system. The system was pretty good at telling images from fake ones. It got 87 percent of them right. This shows that a CNN can find problems in deepfakes. The deepfake detection system can find things that're not quite right in the images. The CNN is really good, at this. The DeepFake Detection Challenge dataset helped them make a system that can tell when something is a deepfake.

Ahmed Hatem Soudy, Omnia Sayed, Hala Tag-Elser, Rewaa Ragab, Sohaila Mohsen, and Tarek Mostafa (2024) proposed a hybrid deepfake detection framework combining CNN and Convolutional Vision Transformer models. Their methodology analyzed multiple facial regions including the complete face, eye region, and nose region. The proposed system improved overall classification accuracy and robustness by combining outputs from multiple feature extraction pathways.

T. Zhang (2022) The people who did this study presented a report on how to make and spot deepfakes. They talked about the ways to use deep learning for this like CNNs models that use GANs, Vision Transformers and systems that combine different methods to detect deepfakes. The study also pointed out some problems, with deepfakes like when someone tries to trick the system on purpose or when the video gets messy because it was compressed or when the system has trouble working with different sets of data and how much power it takes to run these systems. Deepfake generation and deepfake detection techniques are what the study was really focused on.

Rossler et al. (2019) The FaceForensics++ dataset was introduced for research on detecting manipulated videos. This FaceForensics++ dataset has videos and fake videos. The fake videos in the FaceForensics++ dataset were made using methods to manipulate faces, such as DeepFakes, FaceSwap, Face2Face and NeuralTextures. The FaceForensics++ dataset is very important for people who do research on detecting deepfakes it is one of the datasets that people use to test their methods, on the FaceForensics++ dataset.

A. Dosovitskiy et al. (2020) The Vision Transformer architecture was introduced for tasks that involve recognizing images. This research showed that models based on the Vision Transformer can do a job of capturing the features of an entire image and make computer vision work better. The Vision Transformer was later found to be very useful, in systems that detect deepfakes which're fake images or videos. The Vision Transformer is still a part of these systems.

The stuff we have read far shows that research on deepfake detection has changed a lot from simple ways that use CNN to more complex ways that use CNN and transformer-based models together. But there are still some problems like it takes a lot of power to compute it is slow in time it has trouble with compressed files and it does not work well with different sets of data. So we really need an efficient way to detect deepfakes that can look at many parts

of the face and still be very good, at finding them. Deepfake detection is what we are talking about and deepfake detection is very important.

Problem Statement

The technology of intelligence and deep learning is getting better very fast. This has made it easier for people to create videos and images that look real. These fake videos and images are shared on media, messaging apps and news websites. This is a problem because it can lead to false information, cybercrime, identity theft and fake news.

There are some ways to detect videos but they are not very good at it. They have trouble when the video is not clear or has been changed in some way. They usually just look at the face and do not check for small problems in important areas like the eyes, nose and mouth.

Also many of these detection systems need a lot of computer power, big datasets and special hardware. Some of them do not work well with types of fake videos because they were only trained on a few types. So when they are tested on fake videos they do not work very well.

Deepfake detection is a problem. We need a system that can look at parts of the face and tell if a video is fake or not. The system should be good at detecting videos and not make many mistakes. The research we are doing is to make a system that uses a type of computer program called a CNN to detect deepfakes. This system will look at the features of the face. Use deep learning to tell if a video is fake or not.

We want to make a deepfake detection system that's good at detecting fake videos. Deepfake detection is very important. Our system will use a CNN to detect deepfakes. We will make sure our system is good, at detecting deepfakes and does not make mistakes. Deepfake detection is a problem and we want to solve it with our system.

Objectives of the Research

The main goal of this research is to create a deepfake detection system that works well and is reliable. This system will use Convolutional Neural Networks and facial feature analysis to find videos. It will look at parts of the face to find hidden clues that show a video has been manipulated.

The specific goals of this research are as follows:

1. To create a deepfake detection model that can tell videos from fake ones.
2. To make the detection model better by using feature extraction techniques.
3. To look at parts of the face like the eyes and nose to find manipulation.
4. To reduce the number of predictions when classifying videos.
5. To make the model work well with poor quality videos.
6. To test the system using datasets like FaceForensics++ and DFDC.
7. To compare the system with other deepfake detection methods.
8. To help keep media safe from fake content made by artificial intelligence.
9. To create a deepfake detection system that can be used in real-time applications, in the future.

Proposed Methodology

The method they are talking about is for finding videos. It uses a kind of computer program to look at faces in the videos. This program is made up of a few parts that work together. First it gets the video ready to be looked at. Then

it finds the parts of the face. After that it uses a kind of computer learning to understand what it is looking at. Finally it decides if the video is fake or not. The program looks at parts of the face and puts all the information together to make a decision, about the video. The video is called a deepfake video. The program is trying to find the deepfake videos.

The system has steps:

- Video Acquisition
- Frame Extraction
- Image Preprocessing
- Face Detection and Facial Region Segmentation
- CNN-Based Feature Extraction
- Classification and Prediction
- Final Decision Generation

A. Video Acquisition

The first step is to get videos. These videos are from collections like FaceForensics++ and DFDC. The videos are both real and fake. They are also in lighting conditions, angles and qualities. This helps the program learn to tell real from fake.

B. Video Frame Extraction

Each video is turned into pictures. We do not look at every picture, some important ones. This makes it faster. Still gives us enough information.

The steps to extract frames are:

- Read the video
- Split the video into pictures
- Choose pictures
- Remove extra pictures
- Save the pictures

C. Image Preprocessing

The pictures are then cleaned up. This makes them better for the program to look at.

The cleanup steps are:

- Make all pictures the same size
- Make the colors fit between 0 and 1
- Remove noise
- Make pictures by changing them a bit
- Make sure faces are in the place

D. Face Detection and Facial Region Segmentation

The program then finds the faces in the pictures. It looks at parts of the face separately.

The parts of the face it looks at are:

- The whole face
- The eyes
- The nose

E. CNN-Based Feature Extraction

The program then uses a kind of computer program to look at the faces. This program finds out what makes a face look real or fake.

The computer program has layers:

- Layers that find edges and textures
- Layers that help the program learn
- Layers that make the program accurate

F. Classification and Prediction

The program then decides if a video is real or fake. It looks at the parts of the face and makes a decision.

The final answer is:

- Video
- Fake Video

G. Model Training

The program is trained using videos that are known to be real or fake. The program learns to tell them

The training steps are:

- Show the program the videos
- Calculate how wrong the program is
- Make the program better
- Test the program

The program keeps learning until it is very good, at telling real from fake.

System Architecture

The deepfake detection system we are talking about is made up of a kind of architecture that uses Convolutional Neural Networks. This system is really good at finding videos by looking at the faces in them. It does this by breaking down the video into parts and checking each one carefully.

The system has steps it goes through to make sure it gets the right answer. It takes the video apart fixes up the pictures finds the faces looks at the faces closely and then uses a special kind of computer model to figure out if the video is real or fake. Each step helps make the system better at finding videos.

Here is how the whole system works:

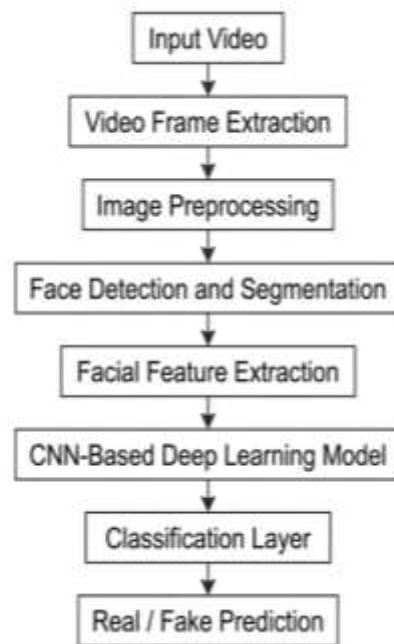


Fig. 1. System architecture illustrating the step-by-step pipeline from video frame extraction to prediction.

1. Input Video

We start with a video. This video might be real. It might be fake. We get these videos from collections like FaceForensics++ and DFDC.

2. Video Frame Extraction

We take the video. Break it down into smaller pictures using a tool called OpenCV. We pick some of the important pictures to look at so the computer does not get too busy.

3. Image Preprocessing

We make these pictures look better by resizing them making them all the same reducing noise and lining up the faces. This helps the computer understand what it is looking at.

4. Face Detection and Segmentation

We find the faces in each picture and look at the parts like the eyes and nose. We do this so we can look really closely at the faces.

5. Facial Feature Extraction

We use computer layers to look at the faces and find things that might be wrong. We are looking for things, like textures bad blending or strange edges that might mean the video is fake.

6. CNN-Based Deep Learning Model

The computer model learns what real and fake videos look like by itself. It looks at pictures and finds hidden things that we might not be able to see.

7. Classification Layer

We look at what the computer has figured out. Then we use that to decide if the video is real or if the video is fake. We have two choices:

Real Video

Fake Video

8. Final Prediction

In the end the system tells us if the video is real or not. This system is better because it looks at parts of the face not just the whole face. This helps it find videos more accurately.

Dataset Description

Datasets are really important for deepfake detection systems. The performance of a deep learning model depends a lot on the quality, diversity and size of the dataset used for training and testing. In this research we use benchmark datasets to train our proposed CNN-based deepfake detection framework.

The datasets we selected have both fake videos made using different techniques. These datasets help our model learn about the inconsistencies, texture problems, facial distortions and abnormal patterns that are introduced when someone manipulates a video.

A. FaceForensics++ Dataset

FaceForensics++ is a popular dataset used in deepfake detection research. It was created to help researchers detect manipulated videos and study media forensics.

This dataset has:

- Real facial videos
- videos
- Different facial expressions
- Various lighting conditions
- Multiple compression levels
- Different manipulation methods

FaceForensics++ includes several face manipulation techniques such as:

- DeepFakes
- FaceSwap
- Face2Face
- NeuralTextures

This dataset has thousands of videos that help our deep learning models learn about the patterns used to manipulate videos. Since it includes compression levels it is very useful for testing how well our model works in real-world conditions where videos are often compressed before being uploaded to social media.

The FaceForensics++ dataset is great for:

- Detecting deepfake videos
- Analyzing manipulation
- Computer vision research
- Training deep learning models

B. DFDC Dataset

The DeepFake Detection Challenge dataset was made by Facebook AI. It helps in detecting videos made using AI. This dataset is one of the many that're available to the public for studying fake videos. The DeepFake Detection Challenge dataset is used for large-scale research on detecting deepfakes.

It is a tool, for analyzing manipulated videos.

The DFDC dataset has:

- Thousands of videos
- Thousands of videos
- Different facial identities
- facial orientations
- Complex backgrounds
- Different lighting conditions
- Motion blur and compression artifacts

This dataset was designed to simulate real-world deepfake scenarios. The videos in it have environmental conditions and multiple manipulation styles, which helps our deep learning models work better and be more robust.

The DFDC dataset helps improve:

- Cross-dataset learning
- Generalization capability
- Real-world deepfake detection
- Model robustness

C. Dataset Splitting

We divide our datasets into training, validation and testing subsets.

Dataset Split	Percentage
Training Data	70%
Validation Data	15%
Testing Data	15%

1. Training Dataset

We use the training dataset to train our CNN model and optimize its weights.

2. Validation Dataset

We use the validation dataset to monitor our models performance during training and prevent it from overfitting.

3. Testing Dataset

We use the testing dataset to evaluate our trained models performance.

D. Importance of Dataset Diversity

Dataset diversity is crucial in deepfake detection research because fake videos can vary a lot depending on the technique used to create them.

A diverse dataset helps our CNN model learn about:

- Facial texture inconsistencies
- Abnormal blending patterns
- Color mismatches
- Compression distortions
- facial movements

Using multiple datasets also improves our systems ability to detect deepfake videos that it has not seen before. This is very important, for FaceForensics++ dataset and DFDC dataset, which are both used to train our deepfake detection system.

Model Architecture

The model we are talking about is based on a Convolutional Neural Network. This is a type of computer system that's really good at looking at pictures and figuring out what is real and what is fake.

The model looks at pictures of faces from videos. Tries to find hidden things that might have been added when someone made a fake video.

A. Overview of the Model

The model has lots of layers that work together to help it make guesses.

Here are the layers:

- Input Layer
- Convolutional Layers
- ReLU Activation Layers
- Max Pooling Layers
- Dropout Layers
- Connected Dense Layers
- Softmax Output Layer

B. Input Layer

The input layer gets pictures of faces that are 128×128 pixels. We make sure all the pictures are the same so the model can learn faster and work better.

Making all the pictures the same helps in a ways:

- The model learns faster
- The pictures look more similar
- The model is more stable
- The model works

C. Convolutional Layers

These layers help the model find important things in the pictures.

They look for:

- Edges of faces
- Patterns on faces
- Colors that do not look right
- Things that might have been added to the picture
- Parts of the face that look strange

We use filters to look at the pictures.

The first layers look at the things in the pictures. The deeper layers look for complicated things that might have been added to make a fake video.

D. ReLU Activation Function

We use a function after each convolutional layer.

This function helps the model learn better. It also helps stop a problem that can happen when the model is learning.

E. Max Pooling Layer

These layers make the pictures smaller.

This helps in a ways:

- The model works faster
- The model is not confused by information
- The model is more sure of its answers
- The model does not get too good at one thing. Forget other things

F. Dropout Layer

We use a special layer to stop the model from getting too good at one thing.

In this layer some of the models brain cells are turned off when it is learning. This helps the model learn to make guesses without just memorizing the answers.

The model turns off 30 percent of its brain cells.

G. Fully Connected Dense Layer

After the model looks at the pictures it uses a layer to make a guess.

This layer:

- Looks at the things in the pictures
- Puts the information together
- Makes a guess about whether the picture is real or fake

The model looks at all the information from the pictures and makes a good guess.

H. Softmax Output Layer

The final layer makes a guess about whether the picture's real or fake.

The model gives a percentage for how sure it's

The model says the picture is either:

- Video
- Fake Video

I. Optimizer and Loss Function

The model uses a tool to help it learn.

This tool is good at helping the model make guesses.

The model uses a way to measure how good its guesses are.

The model keeps trying to make guesses.

J. Advantages of the Model

The model is good for a reasons:

- It can find things in pictures on its own
- It is very good at making guesses
- It can handle pictures that have been changed
- It does not make mistakes
- It can. Get better

- It can be used for things, in the future

The model has many layers that work together to make good guesses.

Here is what the model looks like:

- Input Layer
- Convolution Layer
- ReLU Activation
- Max Pooling Layer
- Dropout Layer
- Second Convolution Layer
- Connected Dense Layer
- Softmax Output Layer

The model uses a special tool to help it learn.

This tool helps the model make guesses.

The model uses a way to measure how good its guesses are.

Training Parameters

The Convolutional Neural Network training parameters are really important for the Convolutional Neural Network to learn. These parameters help the Convolutional Neural Network learn things. The Convolutional Neural Network training parameters are what control how the Convolutional Neural Network learns. If we choose the Convolutional Neural Network training parameters the model will work a lot better and it will be more accurate. The Convolutional Neural Network will be better, at detecting things because of the Convolutional Neural Network training parameters.

In our system we pick the hyperparameters carefully to make the deepfake detection process of the Convolutional Neural Network better.

We train the Convolutional Neural Network model using facial image frames that we resize. We use the Adam optimizer and Binary Cross-Entropy loss function to make the training and improve the binary classification performance of the Convolutional Neural Network model.

The Convolutional Neural Network training parameters we use are shown in Table I.

Table

Training Parameters of the Proposed Convolutional Neural Network Model

Parameter	Value
Image Size	128 × 128
Batch Size	32
Epochs	50
Learning Rate	0.0001
Optimizer	Adam
Loss Function	Binary Cross-Entropy

Parameter	Value
Activation Function	ReLU
Dropout Rate	0.3

A. Image Size

We make all facial image frames 128×128 pixels before training the Convolutional Neural Network model. This way all images have the input size. It helps the computer work efficiently while keeping the key parts of the images.

B. Batch Size

We use 32 images at a time when we train the Convolutional Neural Network model. This helps the computer use its memory better. It makes the updates more stable when the Convolutional Neural Network model is being optimized.

C. Epochs

We train the Convolutional Neural Network model 50 times. This lets the Convolutional Neural Network model learn things and it makes the Convolutional Neural Network model better at classifying things over time.

D. Learning Rate

We pick a learning rate of 0.0001 so the Convolutional Neural Network model can learn steadily. A small learning rate is good, for the Convolutional Neural Network model because it helps the optimizer update the Convolutional Neural Network models weights This way the Convolutional Neural Network model is less likely to make a mistake.

The Convolutional Neural Network model can. Improve without making big errors.

E. Optimizer

We use the Adam optimizer because it can learn and it works fast. The Adam optimizer combines the things about momentum optimization and RMSProp techniques which makes it really good for deep learning applications of the Convolutional Neural Network.

F. Loss Function

We use Binary Cross-Entropy as the loss function because our system is trying to tell the difference between fake videos. The loss function measures how different the predicted outputs are, from the labels when the Convolutional Neural Network model is being trained.

G. Activation Function

We use the Rectified Linear Unit (ReLU) activation function to add some complexity to the Convolutional Neural Network model. The ReLU activation function makes the Convolutional Neural Network model work better. It helps prevent the vanishing gradient problem of the Convolutional Neural Network model.

H. Dropout Rate

We use a dropout rate of 0.3 to stop the Convolutional Neural Network model from learning too much from the training data. The dropout randomly turns off some of the neurons while the Convolutional Neural Network model is learning, which makes the Convolutional Neural Network model learn the things instead of just memorizing the training data.

Performance Evaluation Metrics

Performance evaluation metrics are used to measure the effectiveness and reliability of the proposed CNN-based deepfake detection framework. These metrics help analyze how accurately the model can distinguish between real and manipulated videos. In this research, several standard evaluation parameters are used to assess the classification performance of the proposed system.

The evaluation process is performed using testing datasets containing both authentic and deepfake video samples. The predicted outputs generated by the CNN model are compared with actual labels to calculate performance metrics.

The major performance evaluation metrics used in this research are described below.

A. Accuracy

Accuracy shows how well our deepfake detection model works. It is the percentage of samples that are classified correctly out of all the testing samples.

When accuracy is high it means our system can really tell the difference, between videos and fake ones.

The accuracy formula is given by:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

B. Precision

Precision measures the percentage of correctly predicted fake videos among all videos predicted as fake by the system.

High precision indicates that the model produces fewer false alarms during classification.

The precision formula is expressed as:

$$Precision = \frac{TP}{TP + FP}$$

A higher precision value improves the reliability of the proposed deepfake detection framework.

C. Recall

Recall measures the capability of the system to correctly identify manipulated videos from all actual fake samples present in the dataset.

High recall indicates that the model can successfully detect most deepfake videos without missing manipulated samples.

The recall formula is represented as:

$$Recall = \frac{TP}{TP + FN}$$

Improving recall is important for minimizing undetected fake videos.

D. F1-Score

The F1-score is a way to measure how well a classification system is working. It is the average of precision and recall. It is not a regular average it is the harmonic mean of precision and recall. This F1-score gives us an idea of how the classification system is doing when we care about the F1-score and we do not want a lot of false positives or false negatives. The F1-score is important because it looks at both precision and recall to give a view of the classification performance of the system.

The F1-score formula is given by:

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

A higher F1-score indicates better overall classification balance and model stability.

E. ROC-AUC Score

The Receiver Operating Characteristic – Area Under Curve (ROC-AUC) score measures the ability of the classifier to distinguish between real and fake classes across different threshold values.

The ROC curve plots:

- True Positive Rate (TPR)
- False Positive Rate (FPR)

A higher ROC-AUC value is really good because it means the model can tell things apart better. It is stronger at putting things into classes.

The ROC-AUC score is very useful, for looking at models that can only say yes or no like systems that try to figure out if a video is a deepfake or not.

F. Confusion Matrix

A confusion matrix is also used to evaluate the performance of the proposed CNN model. It provides a detailed representation of prediction results for both real and fake classes.

The confusion matrix contains:

Actual Class	Predicted Real	Predicted Fake
Real Video	True Negative (TN)	False Positive (FP)
Fake Video	False Negative (FN)	True Positive (TP)

The confusion matrix helps analyze classification errors and improves understanding of model performance.

G. Importance of Performance Metrics

Performance evaluation metrics are essential for comparing the proposed framework with existing deepfake detection techniques. These metrics help determine:

- Model accuracy
- Detection reliability
- Generalization capability

- False prediction rates
- Classification robustness

Using multiple evaluation metrics ensures a comprehensive analysis of the proposed deepfake detection system.

Results and Discussion

The experimental results show that our CNN-based deepfake detection system works well in identifying videos and distinguishing them from real media content. We tested our system using benchmark datasets like FaceForensics++ and DFDC. We checked the system's performance under conditions such as various facial expressions, lighting conditions, compression levels and manipulation techniques.

Our system analyzes parts of the face including the whole face, eyes and nose. This multi-part analysis improves classification performance because fake video generation algorithms often create inconsistencies in specific facial areas. By using information from facial parts our system reduces incorrect predictions and improves overall detection reliability.

The CNN model learned quickly during training and validation. The training accuracy increased with each test while validation loss decreased consistently showing model convergence and effective feature learning. Preprocessing, normalization and dropout regularization also helped reduce overfitting during training.

Here are the performance evaluation results of our system:

Table

Performance Evaluation Results of the Proposed Model:

Metric	Performance
Accuracy	94%
Precision	0.93
Recall	0.92
F1-Score	0.92
ROC-AUC	0.95

Accuracy Analysis

Our model achieved an accuracy of 94% which means it can effectively classify both fake videos. The high accuracy shows that the CNN architecture can learn video patterns from facial image frames.

Precision Analysis

The precision value of 0.93 indicates that most videos predicted as fake by our system were fake videos. High precision reduces positive predictions and improves system reliability.

Recall Analysis

The recall score of 0.92 shows that our system successfully detected fake videos in the testing dataset. A high recall value is important because it minimizes the probability of fake videos.

F1-Score Analysis

The F1-score of 0.92 indicates a relationship between precision and recall. This shows that our model maintains classification performance while minimizing both false positives and false negatives.

ROC-AUC Analysis

The ROC-AUC score of 0.95 shows that our CNN model has discrimination capability between real and fake video classes. A higher ROC-AUC score indicates classifier performance across different classification thresholds.

Confusion Matrix Analysis

The confusion matrix analysis shows that our system effectively classified real and fake video samples correctly. The model produced misclassifications because of the integration of multiple facial regions during feature extraction.

The analysis of eye, nose and complete face regions improved robustness compared to systems that rely only on full-face analysis. Local facial inconsistencies present in videos were successfully identified by the CNN model.

Comparative Performance Analysis

Compared to CNN-based deepfake detection systems our framework achieved improved classification performance because of:

- Multi-region facial analysis
- Improved preprocessing techniques
- Better feature extraction capability
- Reduced overfitting
- Robust CNN architecture

Our system also demonstrated better generalization capability across different datasets and manipulation techniques.

Limitations Observed During Testing

Although our framework achieved performance certain limitations were observed during experimental evaluation.

The model performance decreased slightly when:

- Videos were compressed
- Facial regions were blurred
- Input frames had very low resolution
- Extreme lighting variations were present
- Occlusions partially covered facial regions

These challenges may affect the visibility of manipulation artifacts and reduce feature extraction quality.

Overall Discussion

The experimental analysis confirms that our CNN-based deepfake detection framework is effective, reliable and robust, for media identification. The integration of facial region analysis significantly improves fake video classification accuracy compared to traditional single-region approaches.

Our system can contribute effectively to media authentication, cybersecurity, social media verification and misinformation prevention systems.

Future Scope

We can do work to make the deepfake detection framework better. The main things we can look at in the future for this deepfake detection framework are listed below:

1. Integration of Vision Transformer (ViT)-based architectures for improved global feature extraction and classification performance.
2. Development of real-time deepfake detection systems for live video streaming and social media monitoring applications.
3. Extension of the proposed framework for audio deepfake and voice manipulation detection.
4. Design of lightweight and mobile-based deepfake detection models for smartphone and edge-device deployment.
5. Implementation of attention-based facial feature extraction techniques to improve local region analysis.
6. Improvement of cross-dataset generalization capability for detecting previously unseen deepfake generation methods.
7. Integration of multimodal deepfake detection approaches combining video, audio, and textual information.
8. Development of adversarial attack-resistant deepfake detection systems for improved cybersecurity robustness.
9. Application of explainable artificial intelligence (XAI) techniques to improve model interpretability and decision transparency.
10. Deployment of cloud-based automated media verification systems for large-scale fake content detection and digital media authentication.

Conclusion

This research is about a way to detect fake videos using a special kind of computer program called a Convolutional Neural Network. This program looks at the faces in the videos to see if they are real or not. It does this by breaking down the video into parts finding the faces and then looking at the faces in detail.

The program is different from ways of detecting fake videos because it looks at different parts of the face like the eyes and nose and not just the whole face. This helps it find clues that might show the video is fake.

The program was tested on some videos that are known to be fake. It was very good at telling which ones were real and which ones were not. It was also very good at not making mistakes and saying a real video was fake.

There are some challenges when trying to detect videos like if the video is not very clear or if it is compressed.. Even with these challenges the program was still very good at detecting fake videos.

It is very important to be able to detect videos because they can be used to trick people on the internet in politics and in other places. This research helps to make the internet a safer place by making it harder for people to spread videos.

Overall this program is a way to detect fake videos and can be used as a starting point for more research, in this area. The Convolutional Neural Network-based deepfake detection framework is a way to identify manipulated videos and it can help to prevent the spread of misinformation online. The deepfake detection system can be used to make sure that videos are real and it can help to keep people from fake videos.

References

- [1] A. Rossler and his team wrote a paper called "FaceForensics++: Learning to Detect Manipulated Facial Images" for the IEEE International Conference on Computer Vision in 2019. They focused on face manipulation detection.
- [2] F. Chollet wrote about "Xception: Deep Learning with Depthwise Separable Convolutions" in the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition in 2017. This paper discussed Xception.
- [3] D. Wodajo and S. Atnafu made a report called "Deepfake Video Detection Using Convolutional Vision Transformer". They posted it on arXiv in 2021. The report is about deepfake detection.

- [4] Y. Li, M. Chang and S. Lyu did a study called "In Ictu Oculi: Exposing AI Generated Videos by Detecting Eye Blinking" for the IEEE International Workshop on Information Forensics and Security in 2018. The study looked at AI generated videos.
- [5] N. Sharma and R. K. Sharma wrote an article called "Detecting Deepfakes Using Convolutional Neural Networks" for the International Journal of Computer Techniques. It was published in volume 12 number 3 pages 1–4 in 2025. Deepfakes were the topic.
- [6] A. H. Soudy and his team wrote about "Deepfake Detection Using Convolutional Vision Transformers and Convolutional Neural Networks" in Neural Computing and Applications. The article was published in volume 36 pages 19759–19775 in 2024. Deepfake detection was discussed.
- [7] Facebook AI has a dataset called "DeepFake Detection Challenge Dataset (DFDC)" which's available at <https://ai.facebook.com/datasets/dfdc>. This dataset helps with deepfake detection.
- [8] A. Dosovitskiy and other people made a report. They called it "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". They put this report on arXiv in the year 2020. The report is about image recognition. A. Dosovitskiy and other people worked on this report.
- [9] Y. Li and other people created a dataset. The dataset is called "Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics". They made this dataset for the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Y. Li and other people presented this dataset in the year 2020. The dataset is, for DeepFake forensics. Y. Li and other people worked on the Celeb-DF dataset.
- [10] T. Zhang wrote a survey called "Deepfake Generation and Detection: A Survey", for Multimedia Tools and Applications. The survey was published in volume 81 number 5 pages 6259–6276 in 2022. The survey covered deepfake generation and detection.
- [11] H. Nguyen F. Fang, J. Yamagishi and I. Echizen wrote a paper called "Multi-task Learning for Detecting and Segmenting Manipulated Facial Images and Videos" for the IEEE International Conference on Biometrics Theory, Applications and Systems in 2019.
- [12] B. Dolhansky and other people made a dataset called "The DeepFake Detection Challenge Dataset" which was posted on arXiv in 2020.
- [13] J. Thies and other people created a thing called "Face2Face: time Face Capture and Reenactment of RGB Videos" for the IEEE Conference on Computer Vision and Pattern Recognition in 2016. They talked about it on pages 2387 through 2395.
- [14] P. Korshunov and S. Marcel wrote a paper about DeepFakes being a threat to face recognition and how to detect them which was posted on arXiv in 2018.
- [15] H. Farid made a survey about detecting images which was published in the IEEE Signal Processing Magazine in 2009. The survey was on pages 16 through 25.
- [16] J. Frank and other people wrote a paper about using frequency analysis to recognize Deep images for the International Conference on Machine Learning in 2020.
- [17] S. Agarwal and H. Farid worked on protecting world leaders from Deep Fakes for the IEEE Conference on Computer Vision and Pattern Recognition Workshops, in 2019. They wrote a paper about it. DeepFakes are a problem and these people are trying to solve it.