A ROBUST APPROACH FOR DEEP FAKE DETECTION USING CNN, RANDOM FOREST AND LSTM ALGORITHMS

Bhanu Sridhar Mantravadi
Information Technology GVPCEW
Visakhapatnam, India

Jaya Satya Durga Kumpatla
Information Technology GVPCEW
Visakhapatnam, India

Kavya Gembali
Information Technology GVPCEW
Visakhapatnam, India

Mohini Devi Kadagala
Information Technology GVPCEW
Visakhapatnam, India

Serena Kandavalli
Information Technology GVPCEW
Visakhapatnam, India

ABSTRACT:

The rapid development of deep learning techniques, particularly in the field of generative models, has led to the widespread creation and dissemination of realistic deepfake content. Deepfakes, or modified media that regularly and convincingly swaps out one person's image for another in images, audio, or video, offer a serious threat to a number of issues affecting society, including misinformation, infringements of privacy, and potential misuse for malicious intentions.

This work seeks to develop an effective deep fake detection system using machine learning and deep learning techniques such as CNN, Random Forest, and LSTM. The proposed solution uses modern neural networks and advanced feature extraction techniques to distinguish between real content and deepfake contents. The primary goal is to increase the reliability of multimedia content by providing a strong tool for evaluating the impact of deepfakes. To train a deep neural network, a large dataset of artificial and real media, including images and videos, must be used. The purpose of the model is to identify intricate patterns and characteristics that suggest deepfake manipulation. Transfer learning techniques are used to optimize the model's performance across numerous deepfakes, offering flexibility.

Key Words: Machine learning, Convolution Neural Networks (CNN), Random Forest, Long Short-Term Memory (LSTM).

I.INTRODUCTION

Recent years have seen a tremendous advancement in deepfake technology, making it easier to create synthetic data that is challenging to compare with real data. Deepfake detection mostly involves analysing a given set of data using machine learning techniques to find any anomalies that might have been created intentionally.

Using an image dataset, deepfake detection is achieved through a tedious method. The first step is data collecting, where we gather both synthetic and real datasets that include lighting settings, backgrounds, and facial expressions. After that, preprocessing methods including cropping, sharpening, and normalization are applied to these photos to guarantee consistency throughout the dataset.

In training algorithms such as convolutional neural networks (CNN), long short-term memory (LSTM), and random forests on the pre-processed dataset, feature extraction plays a crucial role and serves as the fundamental part of the detection process.

These models are employed in both dataset training and fine-tuning, where performance is optimized through the use of augmentation and hyperparameter adjustment. The real-time optimization methods that are used to guarantee the effectiveness of deepfake image detection. By employing techniques such as CNN, LSTM, and Random Forest, one can obtain a strong solution for identifying deepfake threats.

II.RELATED WORK

According to related research, deepfake detection has gained popularity quickly because of how quickly it is developing. Technology is improving its ability to detect deepfake photos by employing intelligent algorithms that hunt for unusual features in the images. Authorities and social media platforms are equally concerned with establishing regulations and resources to prevent these fakes from causing issues. Before artificial intelligence (AI) gained popularity, people used a variety of methods to determine whether an image was real or fake. First, they would carefully examine the image, looking for any hidden details, and then they would use specialized tools to analyses the image's errors and look for patterns in order to find similar images online.

The hidden elements of a given image alter during manual image comparison and inspection, which is time-consuming and yields inaccurate results. The characteristic signs that indicate a fake image were sometimes not apparent. A variety of techniques, including CNN, LSTM, and Random Forest, are employed in detection techniques to identify fraudulent images.

III.EXISTING SYSTEM

The historical data is being collected through prediction analysis in the current system, "Fig-1," which has been started. Particular methods for testing and training data that have been much improved, leading to a quick shift in the deepfake images.

The best-fit algorithm employs a variety of instruments and methods to determine whether the image is real or fake. More accurate predictions and better accuracy can be generated by machine learning algorithms.

When some methods are applied to manually distinguish between fake images, the results are inaccurate. The current system selects the best strategy to provide highly accurate scores, however it primarily provides unclear figures. The current approach, as given, is primarily unreliable in providing accurate results when there are minor modifications in the following area.

Fig.1.Existing System.

IV. PROPOSED SYSTEM

As per "Fig. 2" of the suggested system, the data is taken into consideration for analysing the fake images. Random Forest, Long Short-Term Memory (LSTM), and Convolutional Neural Network (CNN) are the three proposed system models. The goal of this project can be increased from generating a web-based platform to creating a browser extension that identifies deep fakes automatically. Our approach focuses on finding all varieties of deepfake photos with lighting, cropping, and resizing. refining the pictures. The suggested system's simple system model is depicted in "Fig. 2."

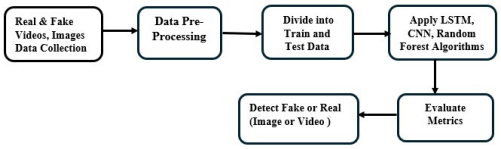

Fig-2. Proposed System

The information gathered over a number of years in order to identify fake photos. Specifically, two machine learning libraries were employed to solve the issue. The first is NumPy, which is widely used to clean, adopt, and employ data, as well as to convert data into the format required for analysis. The second is Pandas, which is used to clean data before combining it and making predictions. The primary approach to the supervised learning model is to learn patterns, relationships, and adjustments.
After processing the data from the training set, they were repeated for the testing data. Next, the data frames were adjusted, enabling users to get the data ready for feature extraction.

V.DATASET USED

The 'Faces224' dataset contains the integrated manipulated images created using the deep fake techniques which involves facial swaps, expressions, and other forms of facial manipulations.

A dataset named "Faces 224 Deepfake Dataset" might indicate that the images have a resolution of 224*224 pixels, this dataset contains 95000 images.

A. Data preparation

The process of making the raw data flexible so that it can be used for additional preprocessing and predictive analysis, also known as data preparation. The dataset, which has some 224*224 pixel-resolution images in it, both real and fake.

B. Data preprocessing

The dataset is gathered and preprocessed during this stage of data preprocessing in order to ensure efficient performance and accurate evaluation.

We preprocess an image by cropping it, adjusting its dimensions, by applying additional features. The dataset is then divided into train, test, and validation sets once the pixel values have been properly normalized. The provided data is normalized, with values ranging from 0 to 1.

C. Data Visualization

Data Visualization which facilitates the visual representation of data 'Fig-3'. The visualized data can be represented in the form of histograms, pie charts, bar plots etc.

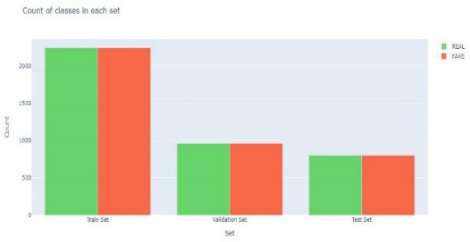

Fig-3 Bar Graph

VI.ALGORITHM USED

A. Convolutional Neural Network (CNN):

Convolutional neural networks (CNN) are a specific type of deep-learning design that has generated a lot of study in computer vision. Computationally, CNN is effective. CNN was utilized to discriminate between authentic and deepfake images by training three separate CNN models. The Dense layer, Max pooling layer, and Dropout layer of a customized CNN model are currently under implementation. In order to identify if the images are real or fake, this method comprises the feature extraction, data preprocessing, augmentation, and classification stages.

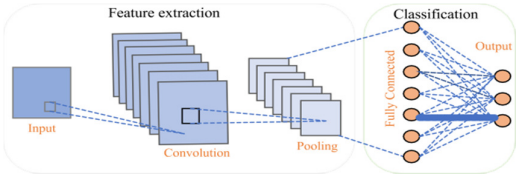

Fig-4: Convolutional layer[7]

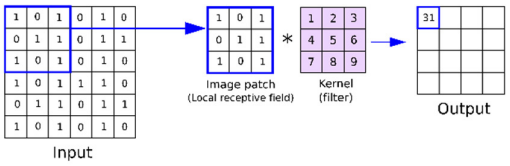

Fig-5: Structure of convolutional operation[13]

The convolutional layer works by taking the input, converting it to an image patch, then multiplying it by the filter path to produce the output kernel dimensions.

1.Activation Functions:

Convolutional processes are usually performed before using nonlinear activation functions such as ReLU (Rectified Linear Unit). ReLU adds nonlinearity and facilitates the network's ability to understand intricate input-to-output connections.

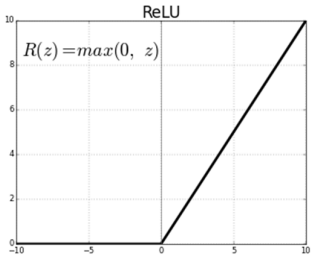

Fig-6 ReLU function

Other activation functions, such as the 'Sigmoid' shown in Figure 4, are also known as logistic activation functions. Any real value can be entered into this function, and the output values range from 0 to 1[4].

Fig-7. Sigmoid function [5].

Mathematically it can be represented as:

$$Sigmoid\ /\ Logistic$$

$$f(x) = \frac{1}{1 + e^{-x}}$$

Fig-7.Sigmoid Equation[5].

For estimating the probability as an output, this model is frequently employed. Given that everything has a probability that ranges from 0 to 1.

2.Pooling Layer:

This convolutional layer component controls overfitting through minimizing complexity and down sampling feature maps. The pooling layer consists of max pooling layers (which select the maximum value inside each window) and average pooling layers.

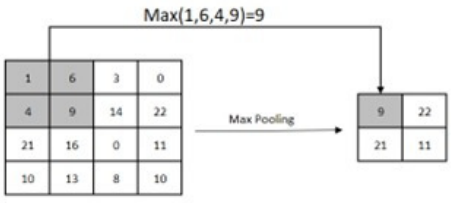

Fig-8 Pooling Layer

3.Fully Connected Layer:

In CNN architectures, the fully connected layers are frequently employed to carry out the classification and regression tasks. In fully connected layers, the neuron applies the linear transformation to the input vector through the weighted matrix [3].

$$y_{jk}(x) = f\left(\sum_{i=1}^{n_H} w_{jk}x_i + w_{j0}\right)$$

Equation for fully connected layer [4].

B. LONG SHORT-TERM MEMORY(LSTM):

While the LSTM network "Fig. 5" is useful for classifying, extracting, and generating predictions based on time series data, there may be a delay due to the time series' unknown duration. LSTM were developed primarily to address the vanishing and gradient descent issues that might be encountered in regular RNN. In order to discriminate between normal (real) and abnormal (fakes) sequences during training, the LSTM network first learns temporal patterns and dependencies between successive frames.
Additionally, the LSTM cell has a memory cell that retains data from earlier time steps and applies it to the current time step to modify the cell's output. The next cell in the network receives the output from each LSTM cell, which enables the LSTM to analyse the process of the sequential data in multiple phases [6].

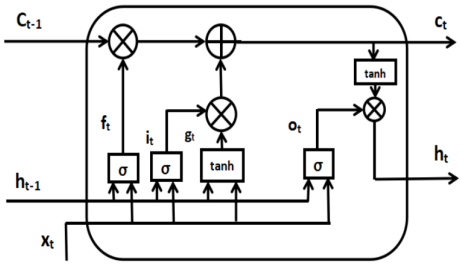

Fig-9. LSTM network

This learnt knowledge is used for classification, with the LSTM output feeding into a classifier to assess if a picture is real or fraudulent. As long-term dependencies between various time series of data can be found, LSTMs are mostly used to analyse the process and then validate the sequential data. Convolutional LSTM (CLSTM) and deep LSTM (DLSTM) are two examples of the different models. Well-defined necessary information can be gathered with a single, single-layered LSTM.

The LSTM has divided into various steps based on the performance:

Previous time step information is recognized by the input gate. The "tanh" function and input gate are used to seek for new inputs that can change the cell state. The output gate that provides the necessary data for improved integrity and performance.

C.RANDOM FOREST

Random forest is a popular approach for regression and classification. The random forest algorithm is the most crucial machine learning algorithm in this case since regression and classification are the two most fundamental parts of machine learning. Here the supervised machine learning algorithm is also referred to as the random forest algorithm.
Overfitting, Speed, and Process are the three distinct criteria that determine which decision trees are included in the random forest [8]. Overfitting is eliminated using the random forest method, which is based on the majority vote or average. Every decision tree developed separately from the other algorithms.

In order to identify deepfake images, random forest functions as a kind of smart identify detector. It finds important details in the images, such as facial expressions, and it gathers patterns that are classified as real or fake. To identify the fake images, decision trees must be created, with each decision tree focusing on a distinct aspect of the images. Using a voting classifier to forecast unique features and determine if an image is real or fraudulent, the random forest method works. The random forest approach is explained in further detail in the section that follows "Fig. 6."

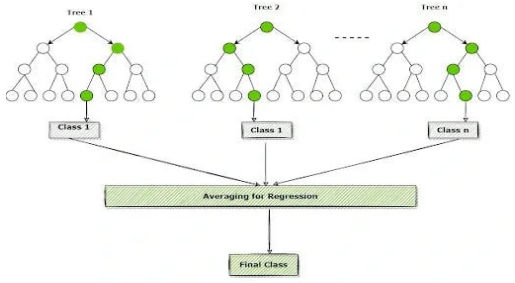

Fig-10. Random Forest [8].

VII.METHODOLOGY

The deepfake image prediction analysis using historical data from the "faces_224" dataset. The use of Random Forest, CNN, and LSTM models predicts the implementation of fraudulent images. Deep fake detection is made using both authentic and generated data. The precise model forecast in this case is shown in "Fig-11."

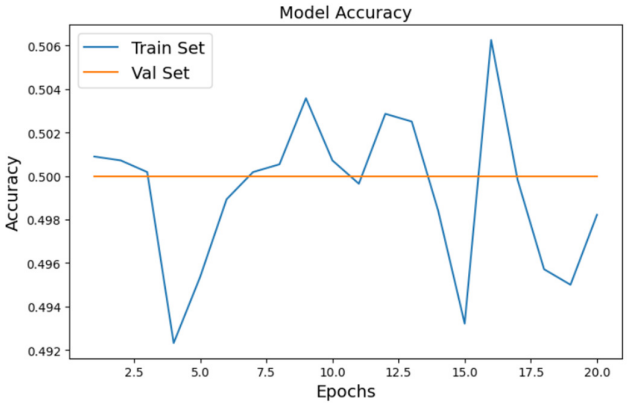

Fig-11 Model Prediction

Following the collection of raw data, the data is separated into train, test, and validation data during the pre-processing stage, which involves eliminating noise from the data. with a variety of techniques, including CNN, LSTM, and Random Forest. Algorithms can be tuned with respect to different parameters for finding the fake images.

A. Collection of Historical data

Collecting authentic and synthetic data is the first stage in identifying fake and real image sources of information. Historical information is derived from "Faces_224," which is composed of certain images that have been altered by cropping, resizing, and altering the size of an image in order to determine whether it is real or fake.

B. About the Algorithm

1.Convolutional Neural Network (CNN):

Using four convolutional layers, two Max pooling layers, and three dense layers—all of which operate in a step-by-step manner—this approach is utilized to distinguish between the fake and real images. The CNN model of a deepfake image and how it classifies the image are shown in "Fig-9." Sigmoid activation is commonly used in binary operations (real or fake) and classification tasks. In order to reduce the vanishing gradient issue during model training and facilitate the learning of increasingly complicated issues, the ReLu activation function is utilized to integrate nonlinearity in a neural network.

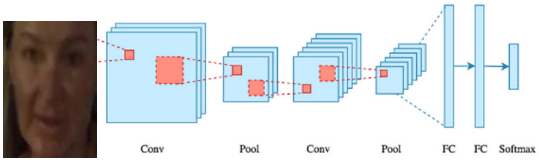

Fig-12 CNN Model Architecture [10]

2.Long Short-Term Memory (LSTM):

The pretrained model's knowledge is stored in the Long Short-Term Memory, and it uses this memory for predicting the model's output based on the requirements. The adaptive moment estimation optimizer, or "Adam" optimization function, is used to solve the learning rate continuously. In comparison to the other optimizers, it operates more quickly on larger datasets. The construction of the LSTM is shown in "Fig-10." The binary cross entropy loss function is primarily utilized for true or fake prediction. We have one dense layer, two LSTM layers, and a "Sigmoid" activation function for binary classification.

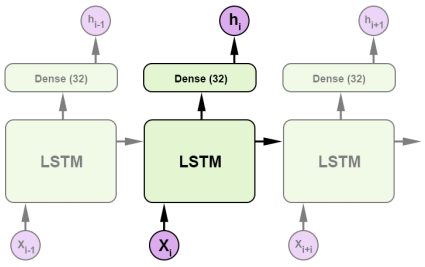

Fig-13: LSTM Model Architecture [11]

3.Random Forest:

In order to improve and predict accuracy and regulate the over-fitting mechanism, the Random Forest model uses a voting classifier with multiple decision tree classifiers on different data samples to classify a given model [11]. A meta estimator that fits the number of decision trees is called a random forest classifier. For more accurate regression and classification models, the random forest classifier is employed. The construction of a random forest is depicted in "Fig-11." where the ensemble learning process is employed by the model. We can lower the number of decision trees by utilizing n_estimators.

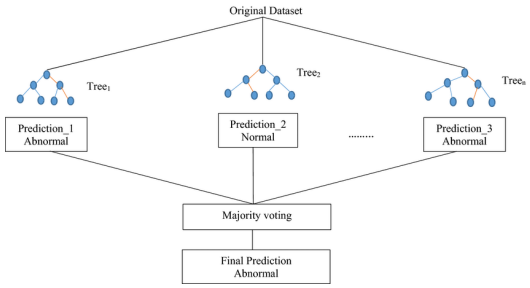

Fig-14: Random Forest Architecture [12]

4.Hyper Parameter Tuning:

A method for creating networks that improves the model's existing configuration is called hyperparameter tuning. Hyperparameter tuning is the process of determining which set of hyperparameters to optimize for a machine learning model in order to improve performance and process development. This selective technique allows the updating of the chosen layers throughout the training process by using a stochastic gradient descent optimizer (SGD) with a learning rate of 0.01 to determine the learning rate, which limits the step size during parameter updates. Based on the pre-trained layers' knowledge, this model modifies the weights to better fit newly discovered data.

VIII. RESULTS

Training and testing groups were created from the collected data. After that, the information that was taken out of the previous papers is examined and produced. via the use of random forest, LSTM, and CNN algorithms.

CNN scored 58% on accuracy. Following data collection and some data enhancement through the modification of specific unique qualities to identify the deepfake detection, the accuracy is then adjusted to 64%. The accuracy is changed to 80% once more using a step-by-step hyperparameter tuning technique. The accuracy of a CNN model is seen in this case in "Fig-15."

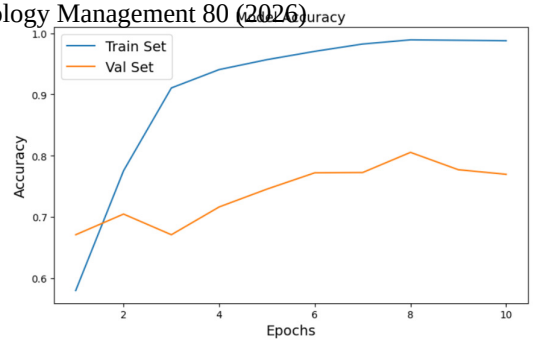

Fig-15 Accuracy CNN

Using an LSTM model with 20 epochs on the testing data, we were able to obtain a 50% accuracy rate in the model's ability to identify a deepfake image. Here, "Fig-16" illustrates an LSTM model's accuracy.

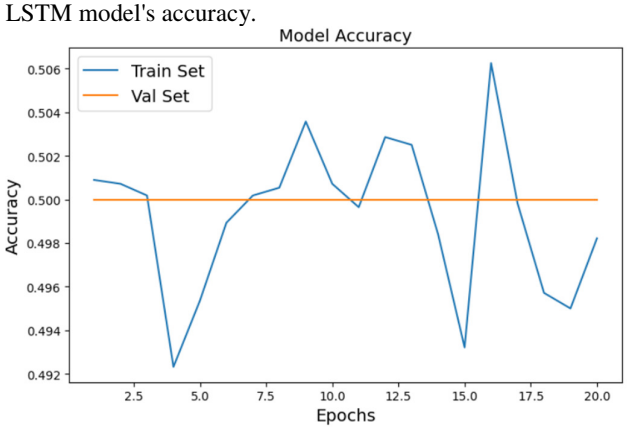

Fig-16: LSTM Accuracy

We obtained 58% accuracy by employing the random forest technique, which is utilized to forecast the accuracy score using the random forest classifier. The accuracy of a random forest model is seen in this instance in "Fig-17".

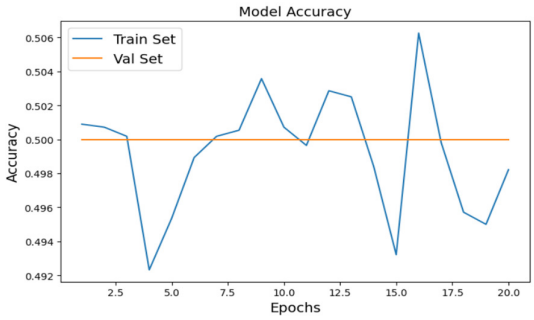

Fig-17: Accuracy Random Forest

IX. CONCLUSION

The suggested hybrid method, which makes use of LSTM networks, Random Forests, and CNNs, presents an effective way of effectively and accurately identifying deepfake photos. Through the investigation of the increasing problems brought about by deepfake technology, this study enhances the trustworthiness and reliability of digital media content.

X.REFERENCES

[1] Shweta Negi, Mydhili Jayachandran, and Shikha Upadhyay, "Deep fake : An Understanding of Fake Images and Videos". *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* Vol 7,2021.

[2] Md Shohel Rana , Mohammad Nur Nobi, Beddhu Murali, And Andrew H. Sung, "Deepfake Detection: A Systematic Literature Review". *IEEE access,* Volume 10, 2022.

[3] Asad Malik , Minoru Kuribayashi,
Sani M. Abdullahi, And Ahmad Neyaz Khan, "Deep Fake Detection for Human Face Images and Videos: A Survey" *IEEE access*, Volume 10, 2022.

[4] Diego Unzeta "Built-in fully connected layer vs convolutional layer explained",*built-in*, oct18,2022.

[5] Pragati Baheti "Activation functions in neural networks" ,*V7labs.com/blog*,may 27.2021.

[6] Manyank Banoula, "Introduction to Long Short Term Memory(LSTM)",*simplilearn*, Apr 27,2023.

[7] Asad Malik ,Minoru Kuribayashi,
Sani M. Abdullahi, And Ahmad Neyaz Khan, "Deep Fake Detection for Human Face Images and Videos: A Survey" *IEEE access*, Volume 10, 2022.

[8] Mohit Chaudhary' " Random Forest Algorithm-How it Works & Why it is so effective",*Turing.com/kb.*

[9] Ankit Chaunan, "Random Forest Classifiers and its hyperparameters",*medium.com/analytics-vidhya,* Feb 23,2021.

[10] Arden Dertat, "Medium:Applied Deep Learning-part 4-Convolutional neural networks*",towardsdatascience.com* Nov8,2017.

[11] Boris Shishov, " Mental Workload Estimation on Facial video using LSTM Network", *automated Control Systems dept., Gubkin Russian State University of Oil and Gas, Moscow, Russia*, june 2017.

[12] A. S. M. Shafi, Julakha jahan jui, M M Imran Molla and Mohammad Motiur Rahman," Detection of colon cancer based on microarray dataset using machine learning as a feature selection and classification techniques", July 2020.

[13] Anh H. Reynolds, "Convolutional Neural Networks(CNN), *anhreynolds.com/blogsPhysical/Analytical Chemist.*2019.