Volume 10, Issue 12, December – 2025 International Journal of Innovative Science and Research Technology
ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec111

Deepfake Detection in Manipulated Images/ Audio

Harish Chaudhary[1]; Nandeesh C. R[2]; Gagan T. N[3]; V. Tejas Aarya[4]; Dr. Shakunthala B. S[5]; Chethan Kumar T.[6]

[1,2,3,4]8[TH] Sem, Dept. of ISE Kit, Tiptur
[5,6]Associate Professor Dept. of ISE Kit, Tiptur

Publication Date: 2025/12/06

Abstract: The study presents a three-stage framework leveraging advanced deep learning techniques to enhance deepfake detection across multimedia datasets—image, audio, and video. The initial stage utilizes an Xception Net-based model achieving 95.56% accuracy for image detection via depth-wise separable convolutions on the CelebA dataset. The second stage employs a hybrid CNN and LSTM approach for audio analysis, achieving 98.5% accuracy on the DEEP-VOICE dataset. The final stage integrates XceptionNet and LSTM for video detection, yielding 97.574% accuracy across multiple datasets. To improve model robustness, class weighting addresses dataset imbalances. This research advances detection methodologies, crucial for maintaining digital integrity and combating misinformation.

Keywords: *Deepfake, Convolutional Neural Networks, Long Short-Term Memory, XceptionNet, Celeb Dataset.*

How to Cite: Harish Chaudhary; Nandeesh C. R; Gagan T. N; V. Tejas Aarya; Dr. Shakunthala B. S; Chethan Kumar T. (2025) Deepfake Detection in Manipulated Images/ Audio. *International Journal of Innovative Science and Research Technology,* 10(12), 51-61. https://doi.org/10.38124/ijisrt/25dec111

I. INTRODUCTION

The rise of advanced artificial intelligence (AI) technologies, particularly deep neural networks (DNNs), has led to a significant increase in the usage of manipulated images, videos, and audio files. Contemporary techniques allow for the creation of realistic counterfeit human faces, videos, and voice mimicry. DNN-based methods for face replacement in deepfakes, such as autoencoders (AEs), Variational Autoencoders (VAEs), and Generative Adversarial Networks (GANs), are commonly employed to overlay a source face onto a target image. Recent developments have also introduced real-time voice cloning, generating high-quality speech that mimics target speakers. Deepfakes, notably including realistic videos of former US Presidents Barack Obama, Donald Trump, and George W. Bush, pose significant technological, social, and ethical challenges. Issues surrounding their use, particularly the creation of misleading content, have garnered attention in social media and state news outlets. Research by Tariq et al. highlights the adverse effects of deepfake impersonation on facial recognition systems. The rapid advancement of deepfake technology has raised pressing ethical, security, and privacy concerns, prompting a focus on automated detection methods. A variety of publicly available datasets have been developed to aid in detecting deepfakes, which typically include both genuine and manipulated videos, as well as edited images of individuals. Multiple techniques are currently employed to generate deepfakes.

CelebA is a comprehensive facial image dataset with around eight million attribute labels, featuring varied poses and complex backgrounds. The Deepfake Detection Challenge (DFDC) integrates generated cloned audio and deepfake videos, with the challenge utilizing a dataset of 128,154 videos from eight deepfake generation techniques. FaceForensics++ (FF++) contributes with 5,000 deepfake videos produced from 1,000 authentic YouTube sources. Both datasets include additional resources like Deepfake Detection (DFD) and FaceShifter. The DFDC was developed collaboratively by researchers and major tech companies. Moreover, Celeb-DF, introduced in 2020, consists of 500 authentic videos of 59 celebrities. A deepfake voice dataset, DEEP-VOICE, is also available, containing REAL and FAKE audio files for analysis.The study analyzes deep learning methods for detecting deepfake content across images, audio, and video. Using balanced samples and random sampling, models such as XceptionNet are trained on various datasets including CelebA for images and FaceForensics++ for videos. It incorporates temporal modeling for video and employs CNN and LSTM networks on the DEEP-VOICE dataset for audio detection. The models showed strong accuracy in differentiating between authentic and manipulated media.

II. RELATED WORK

This section examines the latest developments in artificial intelligence techniques for manipulating photos, videos, and audio files. A face-NeSt detection architecture that best chooses multiscale features for final prediction is presented in this work [1]. To determine the ideal ratio of multiscale features, it uses an adaptively weighted multiscale attentional (AW-MSA) module. Face-NeSt highlights

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec111

important feature regions both locally and globally across spatial and channel dimensions. A. Face-NeSt is lightweight compared to the widely used modern computer vision models. With AUC scores of 0.9823 on CelebDF, 0.9947 on DFDC, 0.9945 on DeepFake (FF++), 0.9905 on Face2Face (FF++), 0.9978 on FaceShifter (FF++), 0.9948 on FaceSwap (FF++), and 0.9548 on neuronal textures (FF++), it performs exceptionally well on three publicly accessible benchmark datasets. AUFF-Net, a unified network for detecting FaceSwap (FS) and Face-Reenactment (FR) Deepfakes, was introduced [24]. This method detects FS and FR by using temporal and spatial information from video samples. While the Bi-LSTM evaluated temporal information, an Inception-Swish-ResNet-v2 model was employed as a feature extractor for spatial information. A discriminative feature-vector group was created by adding three dense layers. The average accuracy for FS and FR in FaceForensic++ experiments was 99.21% and 98.32%, respectively.To distinguish between real and fake audio recordings, a lightweight machine learning-based framework was created [2]. Spectral, temporal, chroma, and frequency domain features are among the manually created audio features used in this technique. The accuracy of the ASVSpoof2019, FakeAVCelebV2, and In-The-Wild databases was 89%, 94.5%, and 94.67%, respectively. Explainability techniques improve transparency, clarify decision-making procedures, and pinpoint important characteristics for audio deepfake identification.

Using a spotted hyena optimizer, a hybrid-optimized deep-feature fusion-based deepfake detection (HODFF-DD) framework for videos was presented [3]. HODFF-DD can identify deepfake films created using a variety of methods and is reliable across ethnic groups and lighting conditions. It is composed of two primary parts: bidirectional long short-term memory (BiLSTM) and a proprietary model with InceptionResNetV1 and InceptionResNetV2. The custom model was used to extract frame-level features from faces that were taken from films. The generated feature sequences were then used to train a BiLSTM for the binary classification of actual and fraudulent videos. The efficacy of the method is demonstrated by evaluations on datasets such as Kaggle's FaceForensics++ using techniques like DeepFakes, FaceSwap, Face2Face, FaceShifter, and NeuralTextures, as well as FakeAVCeleb, which achieve above 90% accuracy on subsets like DeepFakes, FaceSwap, and Face2Face. An improved technique for identifying deepfakes in movies has been created using a graph neural network (GNN) [4]. This method divides detection into two stages: a mini-batch graph convolution network stream and a four-block CNN stream. In two stages, three fusion networks—FuNet-A, FuNet-M, and FuNet-C—were combined. The model's accuracy for different datasets was 99.3% after 30 epochs.Several color spaces were used in this investigation to improve deepfake detection [5]. They employed two stages: a representative forgery learning stage using multicolor space reasoning and a color-space-based forgery detection network. A forgery highlighting network, color-space modifications, and a manipulation cue-boosting network were used in the forgery learning stage. The cue boosting network improved feature representation, the color spaces offered advantages over RGB, and the forgery highlighting network discovered high-level semantic forgery clues and textual irregularities. Using the FaceForensics++, DFDC, and CelebDF datasets, they examined the method and discovered that it was successful in detecting multimedia information that had been falsified in a variety of color representations. To improve the adaptability and resilience of deepfake picture identification, our study employed an adaptive blind watermarking technique [6]. By employing mixed modulation and a sign-altered mean value, this method embeds coefficients to guarantee high image quality while thwarting attacks. Furthermore, relative positions in slightly altered or deepfaked photos are adaptively detected using blind adaptive deepfake detection using a tamper detection mean value. Through parameter optimization and watermark detection, the denoising autoencoder and grey wolf optimizer further enhance the method's performance. By adaptively integrating watermark information while preserving the original facial image, this method validates face validity and authenticates the image owner.

Current deepfake detection methods for plaintext faces were the main focus of this study [7]. However, for practical application, sensitive data must be computed safely. One potential solution is the Secure DeepFake Detection Network (SecDFDNet). We describe an additive secret-sharing method for safe detection of DeepFake faces. Additionally, it has been demonstrated that multi-party secure interaction protocols like SecReLU, SecSigm, SecSpatial, and SecChannel are secure with minimal space and communication complexity. The SecDFDNet model achieves the same accuracy as the plaintext DFDNet while outperforming numerous other models by merging these secure protocols with a trained plaintext DFDNet. High-quality and low-quality facial images generated using different algorithms may be distinguished by a unique deepfake detection network [8]. To deal with low-quality images, this architecture combines a standard spatial stream with a frequency stream. To distinguish between real and fake photos, hierarchical supervision was used. This study included multispa-ce channel-representation learning to build an MSCR-ADD [9]. For deepfake detection, this method combines encoders that are channel-specific, channel-differential, and channel-invariant. MSCR-ADD performs better than the state-of-the-art techniques, according to experimental results on four benchmark datasets. By highlighting the differences and similarities between the channels in binaural audio, feature representations in channel-differential and channel-invariant spaces enable efficient artifact identification in false audio. n order to improve the accuracy of deepfake identification, AVFakeNet integrates both visual and audio modalities [10]. With input, output, and feature extraction blocks, AVFakeNet is a Dense Swin Transformer Net (DST-Net). A specifically made swine transformer module with dense layers in the input and output bhads was employed in the feature extraction block. The effectiveness and generalizability of this unified architecture were shown through extensive experiments on five datasets, including audio, visual, and audio-visual deepfakes, as well as a cross-corpora analysis. The results show that by

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec111

examining both audio and visual streams, the suggested framework effectively detects deepfake videos.

III. METHODOLOGY

This research aims to address the challenges introduced by deepfake media in audio, video, and image formats. To reduce the dissemination of misinformation and preserve digital integrity, this framework pro poses comprehensive detection and mitigation strategies utilizing advancements in artificial intelligence, machine learning, and statistical learning. A multimodal framework for determining the authenticity of media content, including audio, videos, and images, is shown in Figure 1. The initial phase involves data collection, which encompasses the acquisition of raw data from three sources: audio, video, and images. These serve as inputs for the subsequent stages. Each data type undergoes distinct processing during the data preprocessing phase. To ensure consistency in size, scaling, and format, image data are first subjected to face detection, which isolates the facial regions, followed by normalization. Mel Frequency Cepstral Coefficients (MFCC) are extracted from audio data to capture relevant acoustic features. Video data are processed by extracting frames, identifying faces within the frames, and executing audio-video synchronization to ensure the alignment of the audio and visual streams. During the model training pha se, predictive models are constructed using preprocessed data. XceptionNet, a deep learning architecture, is employed for feature extraction from the image data, with a softmax layer utilized for classification. For audio data, feature extraction is followed by classification using a Recurrent Neural Network (RNN). Video data analysis involves XceptionNet for frame-level feature extraction and Bi-LSTM for temporal feature modeling. Dynamic Time Warping (DTW) is employed to assess audio-video synchronization and ensure temporal coherence of the media. In addition, a softmax layer is utilized for video feature classification. The trained model is subsequently employed to categorize input media as either authentic or fraudulent, ensuring a comprehensive and reliable detection process across diverse data modalities.

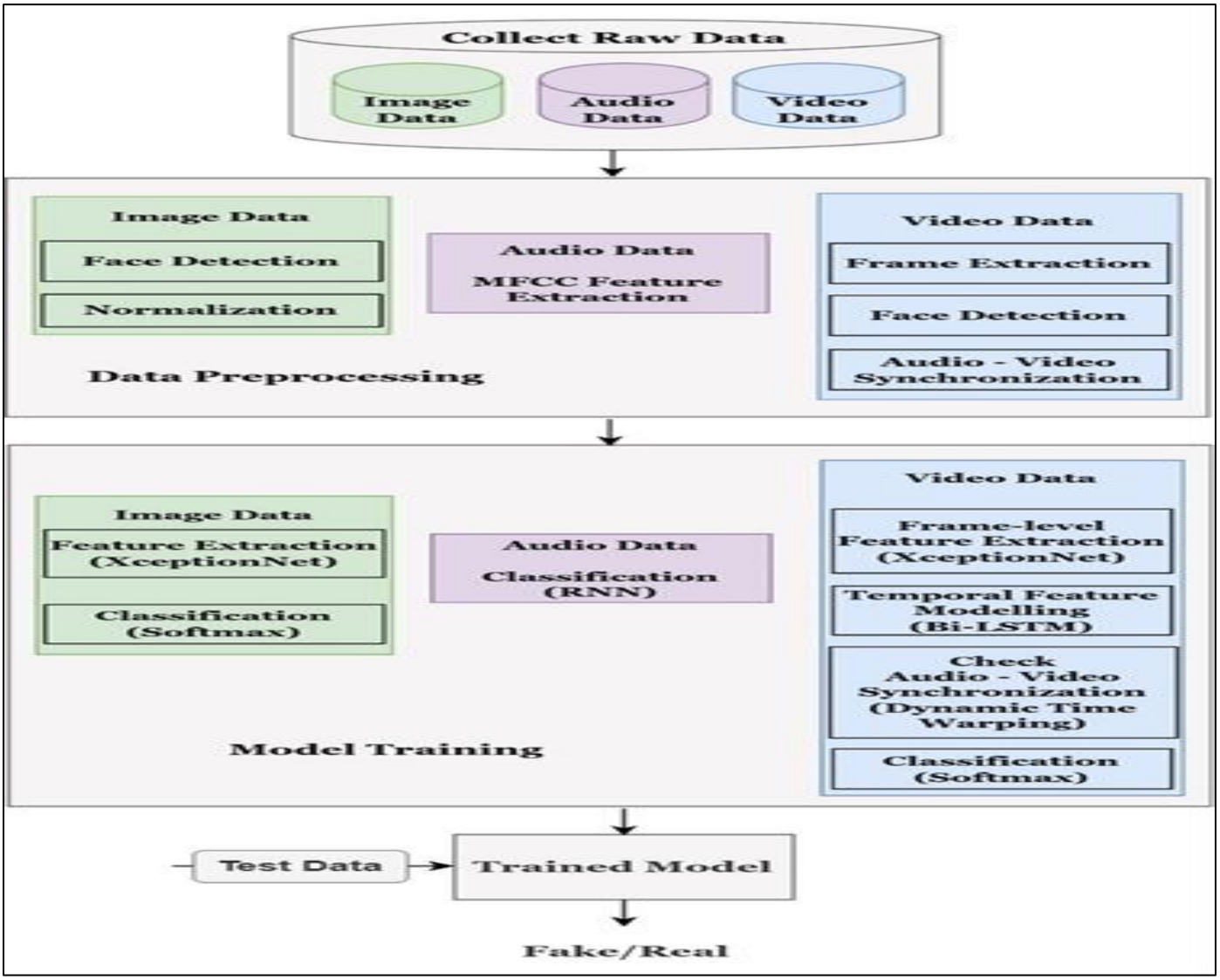

Fig 1 Proposed Methodology

- *Image Deepfake Detection*

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec111

Image deepfake detection focuses on identifying manipulations that are often created using generative models, such as Generative Adversarial Networks. The challenge lies in detecting subtle manipulations such as face swaps, altered expressions, or synthetic creations that are visually convincing. Techniques include the analysis of inconsistencies in lighting, facial landmarks, and texture patterns. Deep learning models are widely used to identify these anomalies and are trained on large datasets of real and fake images to enhance their ability to distinguish between them. This study develops an XceptionNet-based image deepfake detection model using the CelebA dataset.

➢ *XceptionNet Image Deepfake Detection Model*

XceptionNet is a sophisticated convolutional neural network architecture developed by Google resear- chers in 2016. XceptionNet is a modified version of the inception architecture that incorporates depth-wise separable convolutions to enhance the performance and reduce the model's parameter count. The Xcep- tionNet architecture comprises three stages: entry, middle, and exit stages. With 71 layers, including 36 convolutional layers, 3 fully connected layers, and additional auxiliary layers for regularization and training purposes, XceptionNet provides a robust framework for image classification tasks. The input to the XceptionNet model is an image with dimensions of 299 × 299 × 3, where 299 represents the width and height of the image, and 3 denotes the number of color channels (RGB). The input image undergoes normalization and is subsequently processed through convolution layers for feature extrac- tion. The architectural structure comprises three distinct stages: entry, middle, and exit, with the middle stage incorporating a series of depth-wise, separable convolutions. The entry stage focuses on reducing the spatial dimensions of the image while increasing the number of filters. This component comprises several convolution layers, followed by max-pooling for downsampling.

IV. RESULTS

Deepfake detection skills have been greatly enhanced by the use of sophisticated deep learning models, such XceptionNet, and its combination with Long Short-Term Memory networks. These models use cutting-edge techniques to accurately detect modified media content, including video and image modifications, and identify audio deepfakes. One of the best methods for spotting image-based alterations, especially in deepfake photos, is the convolutional neural network XceptionNet. Its capacity to concentrate on fine-grained features and subtle artifacts created during manipulation is improved by the use of depth-wise separable convolutions. When it comes to identifying face-swapped photos and spotting inconsistent editing, CelebA has demonstrated exceptional performance. In order to counteract deepfake material, its ability to extract fine-grained information is essential. The deepfake detection capabilities are extended to video-based media by combining XceptionNet and LSTM networks. LSTM networks record sequential temporal changes across video frames, whereas XceptionNet conducts spatial analysis by identifying frame-specific pixel-level anomalies. By identifying both temporal and spatial artifacts, this hybrid technique guarantees a thorough analysis of video modifications.

This combined approach produces accuracy rates of 97%, according to datasets like Celeb-DF and DeepFake Detection Challenge. A reliable method for detecting deepfakes in dynamic video content is the combination of XceptionNet spatial precision with LSTM temporal modeling. For the identification of audio deepfakes, Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have proven to perform well together. This architecture combines the sequential pattern recognition capabilities of LSTMs with the spectral feature detection capabilities of CNNs for audio data. This makes it possible to find temporal irregularities and audio distortions, which are essential for spotting deepfake audio. With a 98% accuracy rate, CNN+LSTM architectures are essential for identifying modified audio information, including voice cloning and artificial speech. The time-domain waveform is displayed in Figure 2, with the audio signal amplitude on the y-axis and the time in seconds on the x-axis. With higher peaks denoting greater intensity and lower troughs denoting decreased acoustic energy, the waveform depicts the temporal fluctuation in sound amplitude. The temporal dynamics of the audio signals are examined using this representation.

A mel-frequency cepstral coefficient (MFCC) visualization is shown in Figure 3, where the y-axis represents MFCC coefficients and the x-axis represents time. Color intensity variations show that each value represents the magnitude of a certain MFCC coefficient at a given period. The short-term power spectrum of a sound source is captured by this graphic, which is helpful for understanding temporal fluctuations in frequency content in speech and audio processing applications.

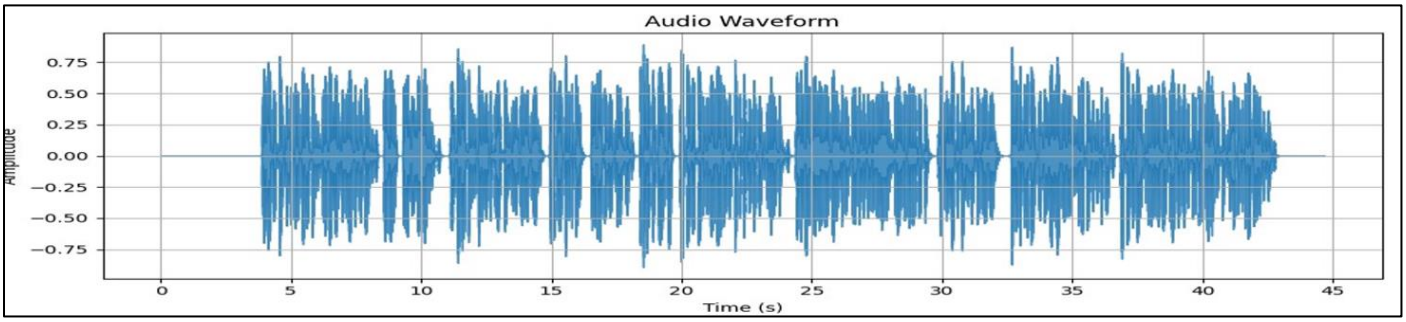

Fig 2 Time-Domain Waveform for Amplitude Variation Over Time

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec111

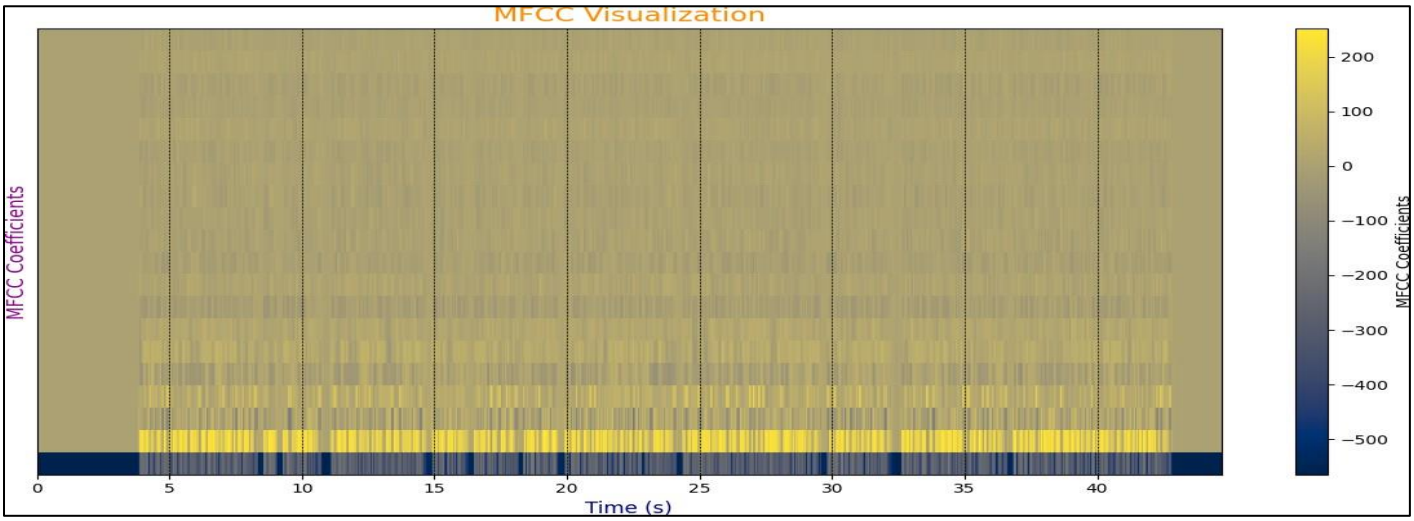

Fig 3 MFCC Visualization For Mel Frequency Cepstral Coefficients (MFCC) Over Time

A spectrogram is displayed in Figure 4, where the y-axis represents frequency on a mel scale and the x-axis represents time.The chroma feature visualization is shown in Figure 5, where the y-axis represents 12 different chroma values that correspond to 12 pitch classes in western music and the x-axis represents time. The intensity of a specific pitch class at a given moment is indicated by each color. This visual aid, which highlights the prominence of various notes or chords over time, is helpful for examining harmonic and melodic content in music processing

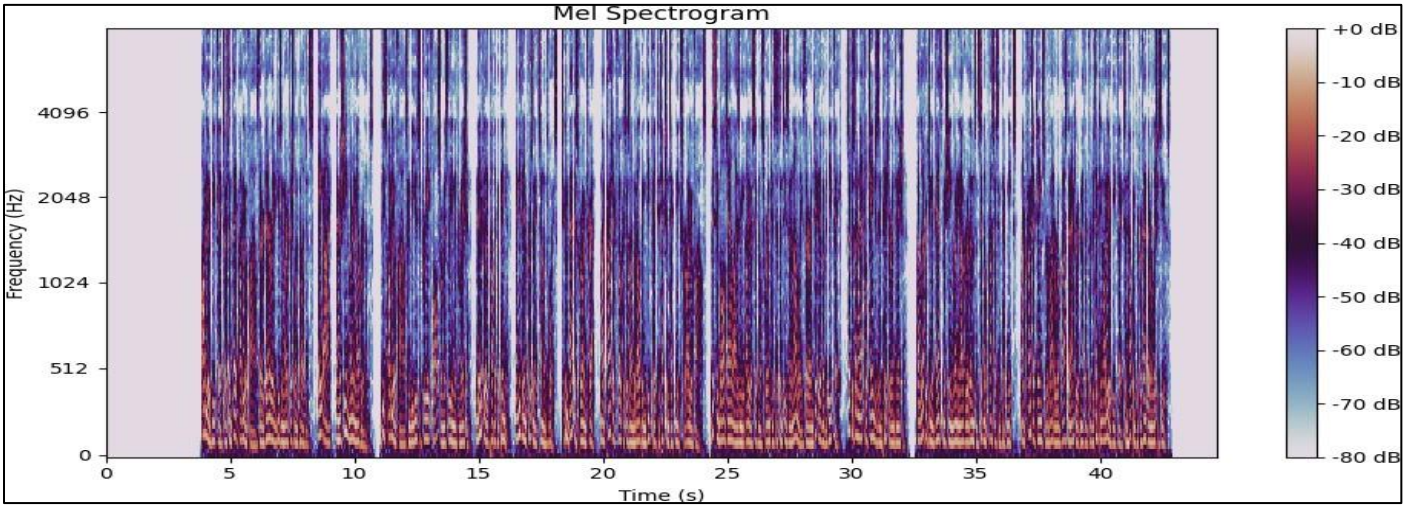

Fig 4 Melspectrogram for Frequency Distribution Mel Scale Over Time

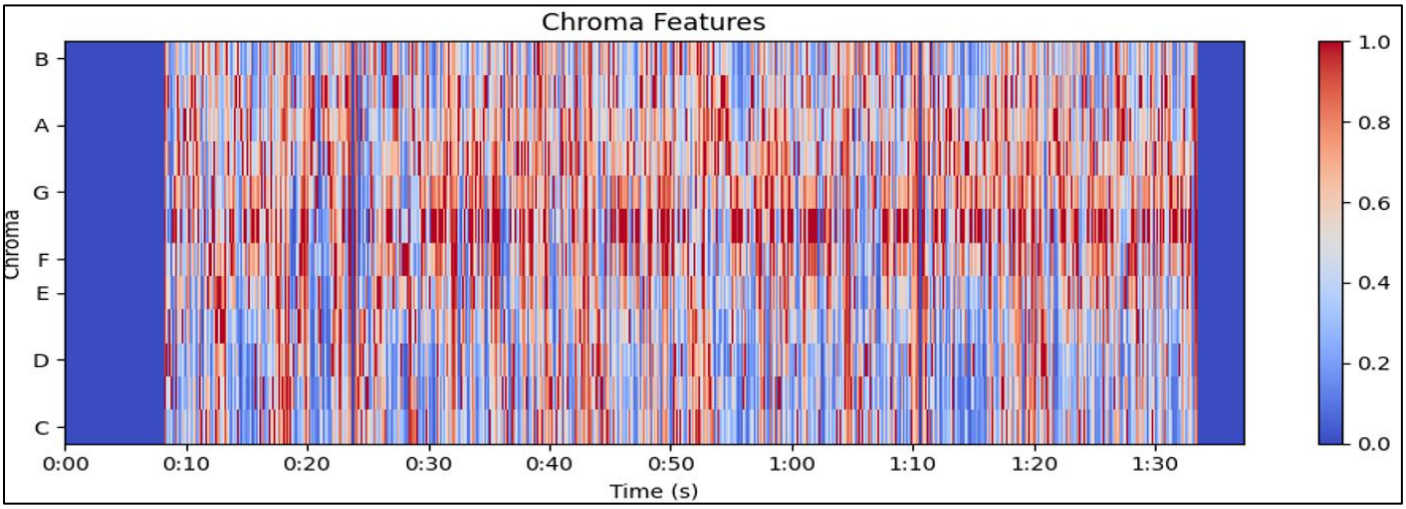

Fig 5 Chroma Feature Visualization for Pitch Class Intensity Over Time

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec111

The zero-crossing rate (ZCR) is displayed in Figure 6, where the y-axis represents the rate of zero crossings and the x-axis represents time. ZCR shows the frequency of sign changes in the audio stream over a certain period of time. Higher ZCR values indicate faster signal fluctuations, and this property is utilized in audio analysis to distinguish between different sound genres.

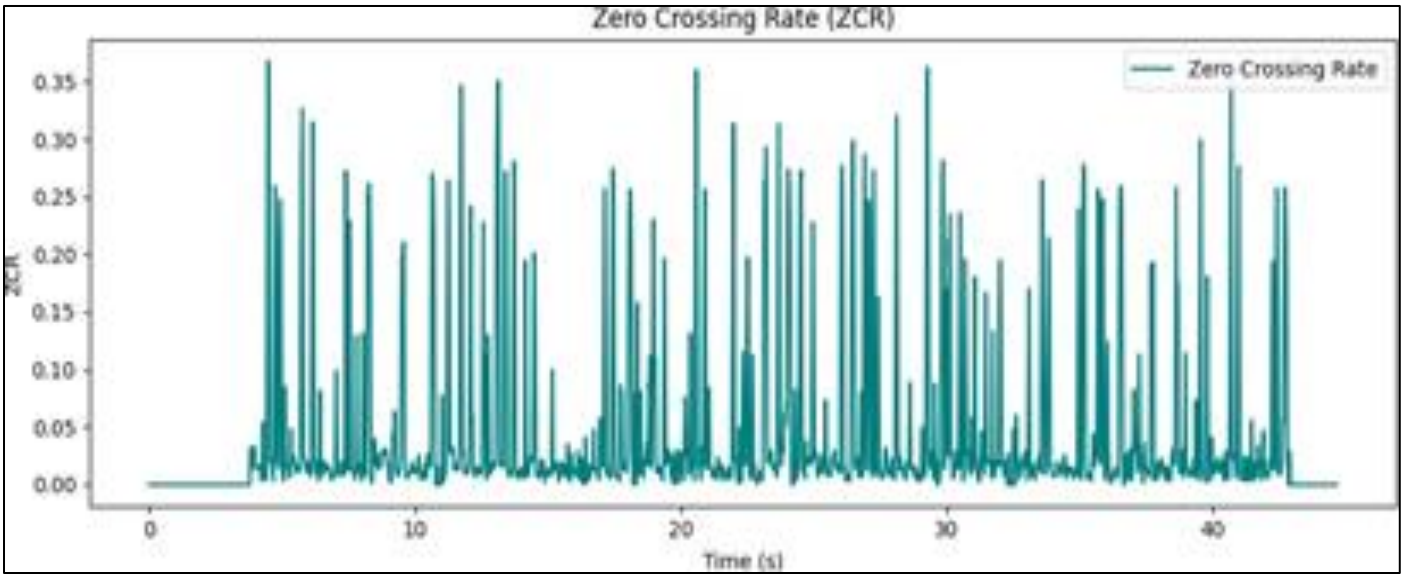

Fig 6 Zero Crossing Rate (ZCR) over time for rate of sign change

Fig 7 shows the Spectral Centroid over time, with the x-axis showing time and the y-axis depicting the spectral centroid in terms of frequency. The spectral centroid indicates the ¸center of mass.of the spectrum, and is often perceived as a measure of sound brightness. Higher values correspond to brighter sounds and lower values correspond to darker sounds. This feature is used to characterize the timbral quality of the sounds.

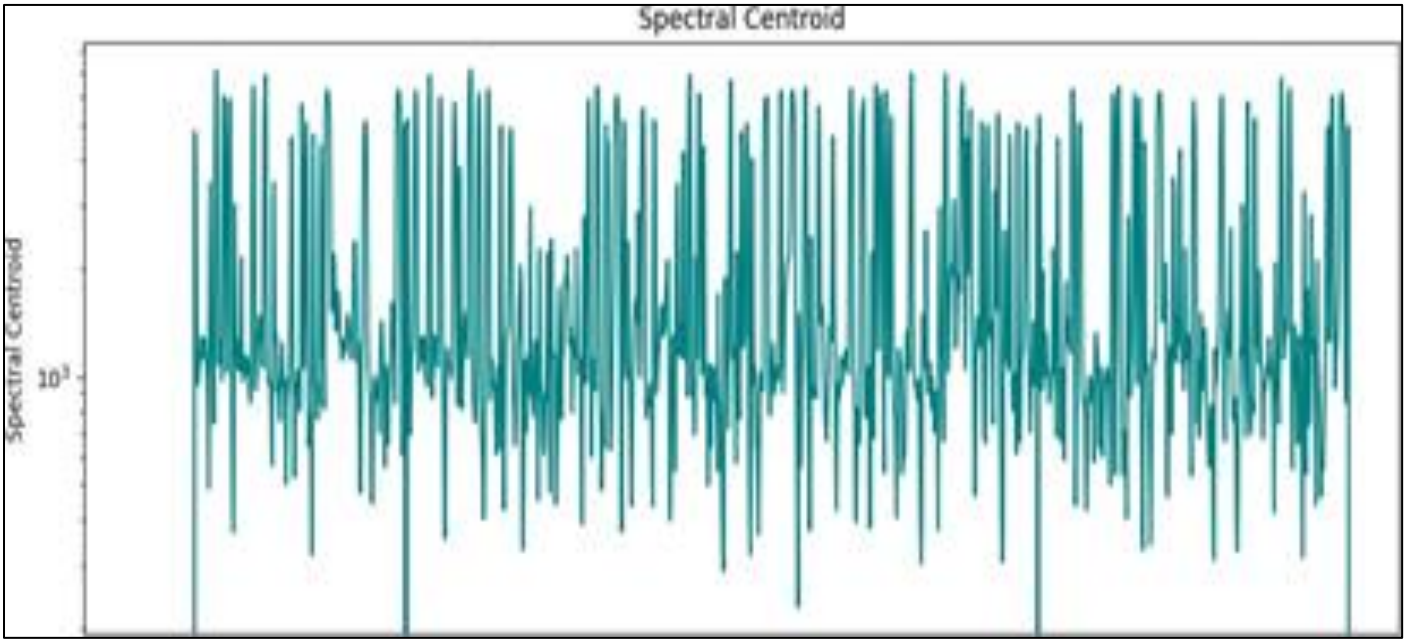

Fig 7 Spectral Centroid Over Time For Audio Spectrum Brightness

With the x-axis representing time and the y-axis representing the spectral flatness value, Figure 8 illustrates the spectral flatness with time. Spectral flatness measures how tonal or noise-like a sound is; values close to 0 indicate a more tonal signal, and values close to 1 indicate a flatter, more noise-like spectrum. This property is employed in a variety of audio signal processing activities and aids in differentiating between harmonic and noise-like sounds.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec111

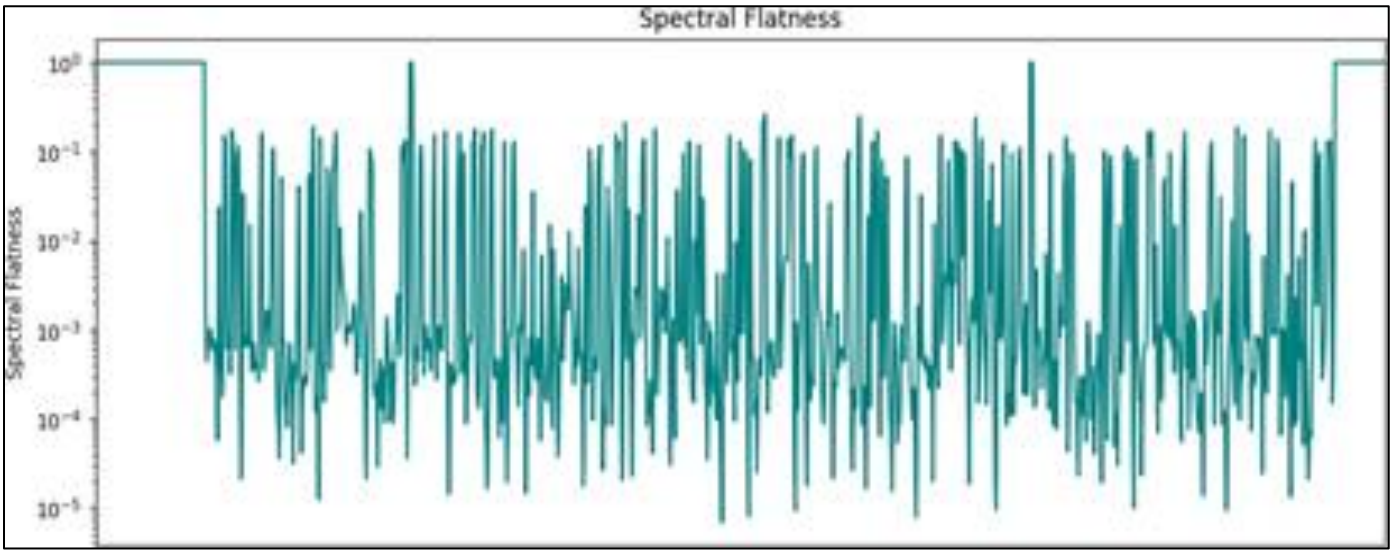

Fig 8 Spectral Flatness Over Time For The Tonal Nature Of Audio Signals

In conclusion, cutting-edge deepfake detection techniques for image and video media are represented by XceptionNet and its combination with LSTM networks. Accurate identification of altered content is ensured by the model's capacity to concentrate on minute temporal and spatial irregularities. Through CNN + LSTM architectures, these capabilities extend to audio-based deepfakes, providing a thorough method to handle the problems caused by deepfake media. These approaches continue to be at the forefront of precise and reliable media manipulation mitigation as datasets and models change.

Table 1: Performance of Different Models on Deepfake Detection across Media Types

Model	Accuracy (%)	Media Type	Dataset
XceptionNet	95.56	Image-based	CelebA
XceptionNet + LSTM	97.00	Video-based	FaceForensics++, DFDC, Celeb-DF
CNN + LSTM	98.00	Audio-based	DEEP-VOICE

The accuracy of several deepfake detection models for diverse media genres is shown in Table 1. On the CelebA dataset, the image-based XceptionNet model's accuracy was 95.56%. Using a mix of XceptionNet and LSTM, performance in video-based deepfake detection increased to 97% when evaluated on FaceForensics++, DFDC, and Celeb-DF datasets. CNN combined with LSTM demonstrated the highest accuracy of 98 % in detecting deepfake audio using the DEEP-VOICE dataset. These results imply that deep learning architectures are successful in detecting deepfake content, however their precision varies based on the type of media and dataset used.

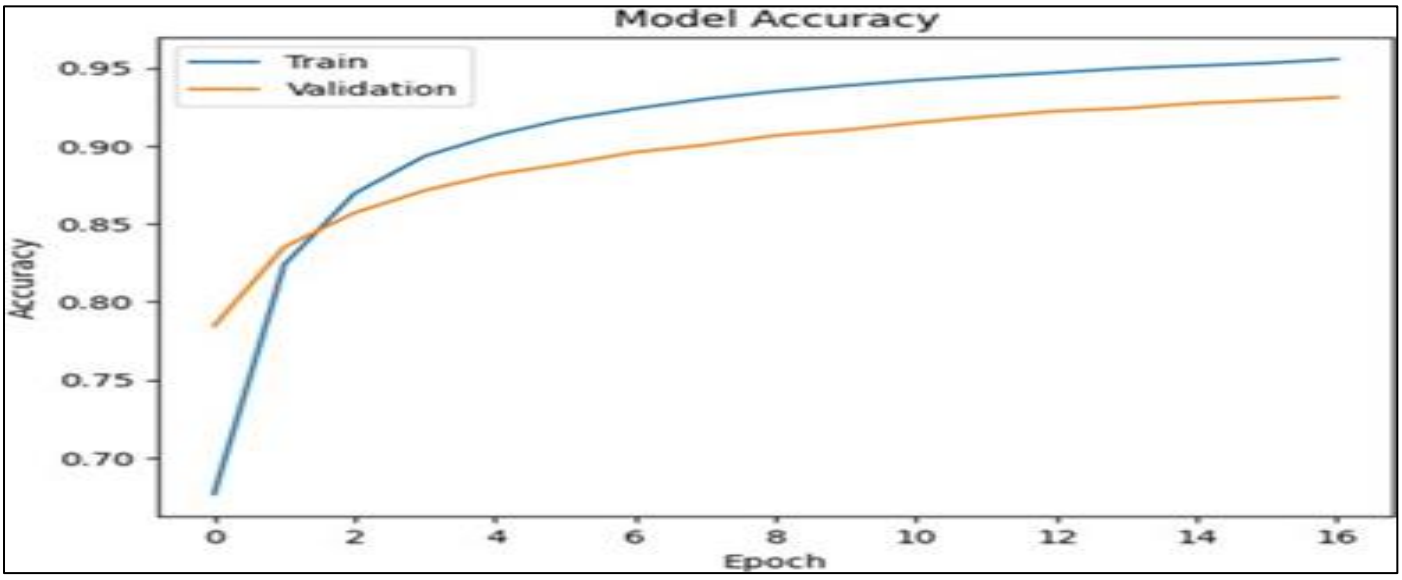

Fig 9 Training vs Validation Accuracy of XceptionNet Model

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec111

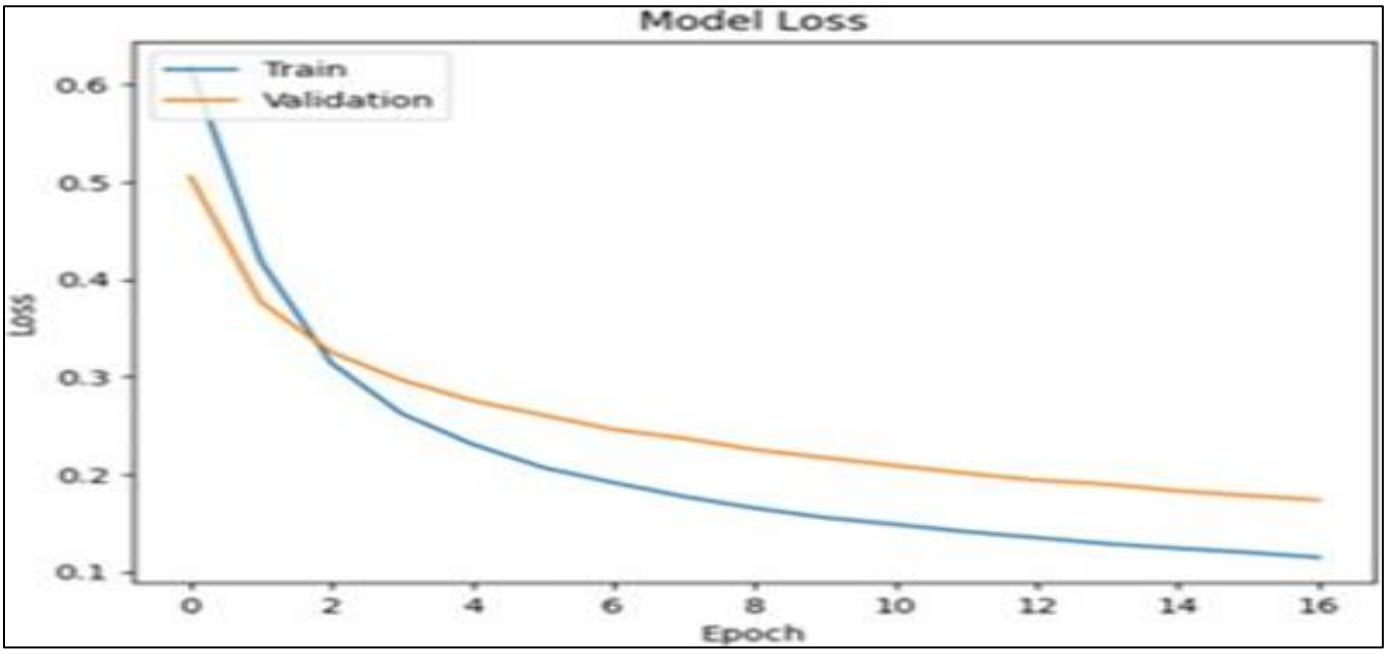

Fig 10 Training Vs Validation Loss of Xception Net Model

Fig 9 displays the model's accuracy during training and validation. The x-axis displays the epochs, while the y-axis displays the accuracy values. The training accuracy first declined before rapidly increasing as the model found patterns. The validation accuracy did, however, rise. As epochs go by, both accuracies stable, indicating that the model is convergent and has taken in the most important patterns. A possible modest overfitting is indicated by the difference between the lower validation accuracy and the greater training accuracy. This implies that while the model does well on training data, it does not perform well on untested data. The plateau at high accuracy levels shows that the model performed steadily. Figure 10 displays the model loss over training and validation epochs. The epochs are shown on the x-axis, while the loss values are shown on the y-axis. The training loss drops from the initial high level as the model learns. Additionally, the validation loss first declined before leveling out after a few epochs. At the end, the validation loss is still more than the training loss, suggesting a generalization gap and potential overfitting. While a slight discrepancy between the two loss curves is acceptable, a large divergence indicates that further training data or regularization techniques are required to enhance generalization.

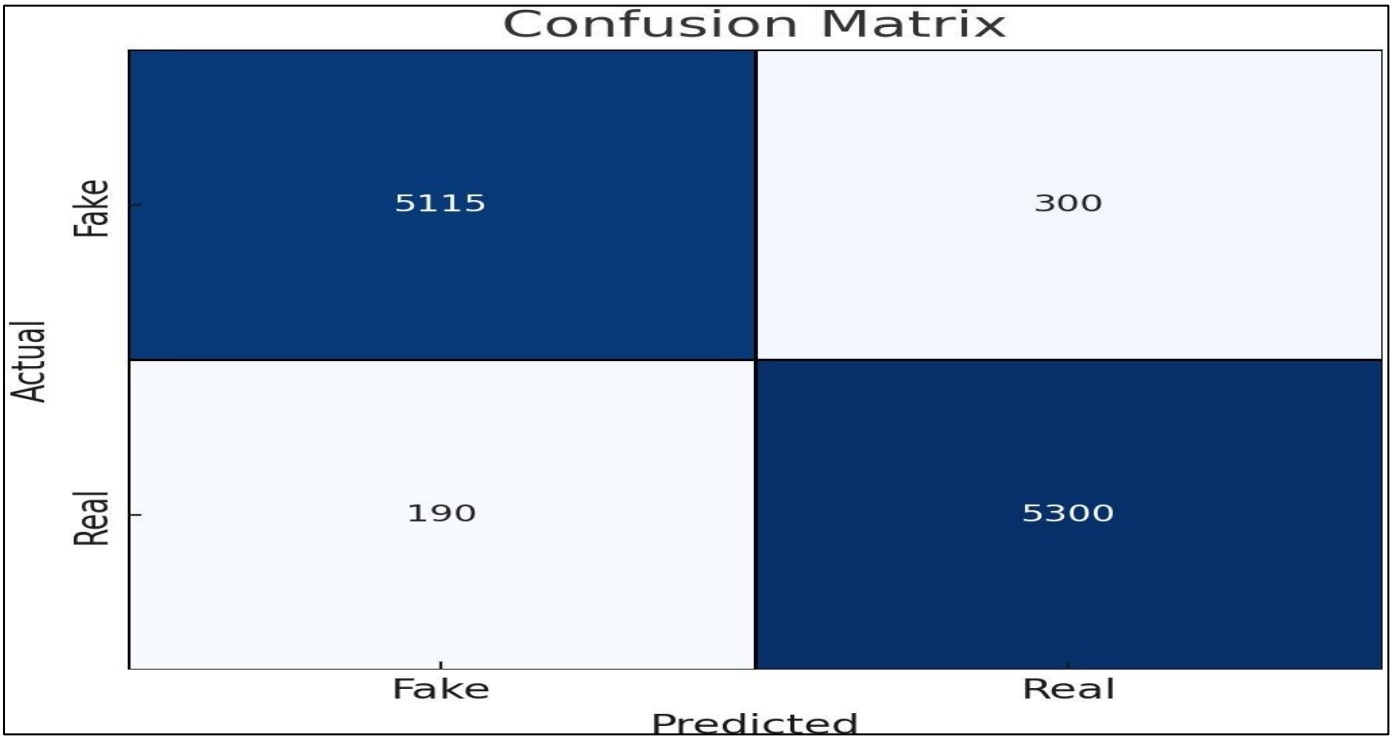

Fig 11 Confusion Matrix for Deepfake Image Detection Model

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec111

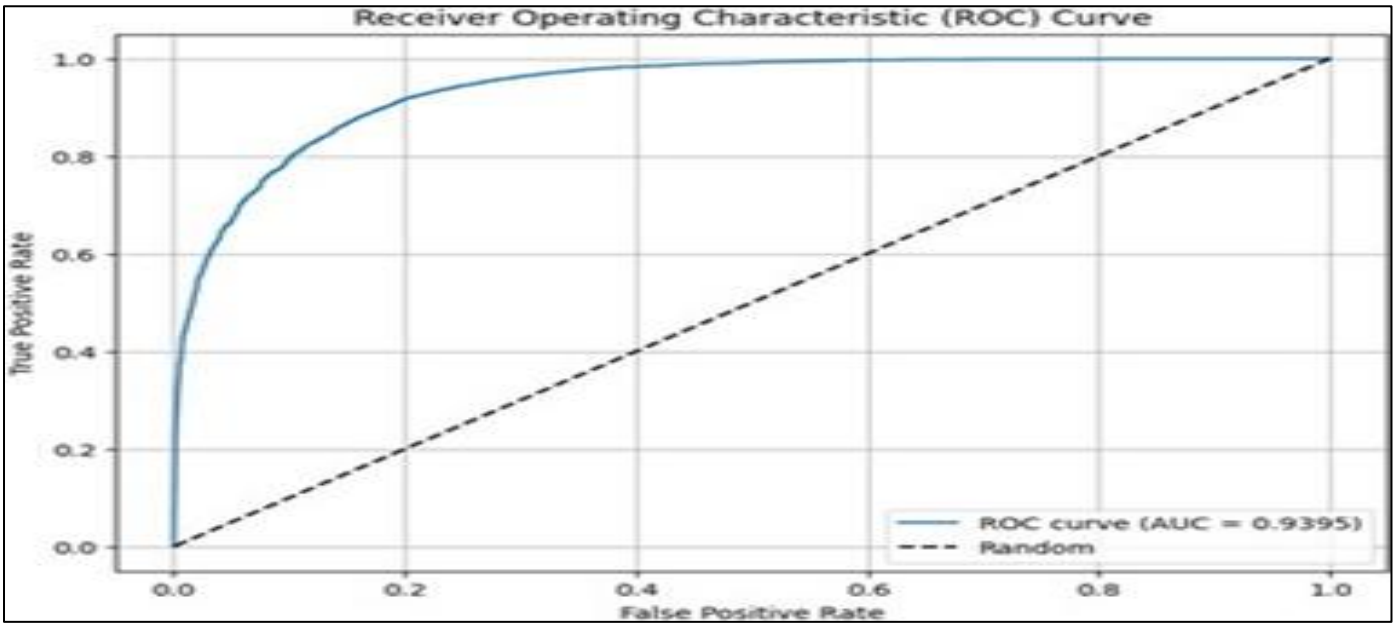

Fig 12: Xception Net Model's Discriminative Capability

Fig 11 displays the confusion matrix of the XceptionNet-based deep fake picture recognition model for fake and real classes. Real labels are displayed in rows, while fake labels are displayed in columns. 4882 bogus instances were properly recognized by the model (top-left value: 5115). The top-right values (300) are false positives. False negatives, in which actual cases are mistakenly classified as fakes, are represented by the bottom-left value (190). The bottom-right value (5300) represents true positives. The majority of the predictions in this matrix are accurate, indicating excellent accuracy. There is potential for improvement given the model's moderate percentage of false positives and negatives. The model did well overall, although it could perform more evenly if there were fewer misclassifications. The capacity of the model to differentiate between classes at various thresholds is assessed using the Receiver Operating Characteristic (ROC) curve, as shown in figure 12. The true positive rate (TPR) and false positive rate (FPR) are displayed on the y- and x-axis, respectively. The curve illustrates the effectiveness of the model in distinguishing between classes, as the decision threshold varies. A curve hugging in the upper-left corner indicates a perfectly discriminative model.

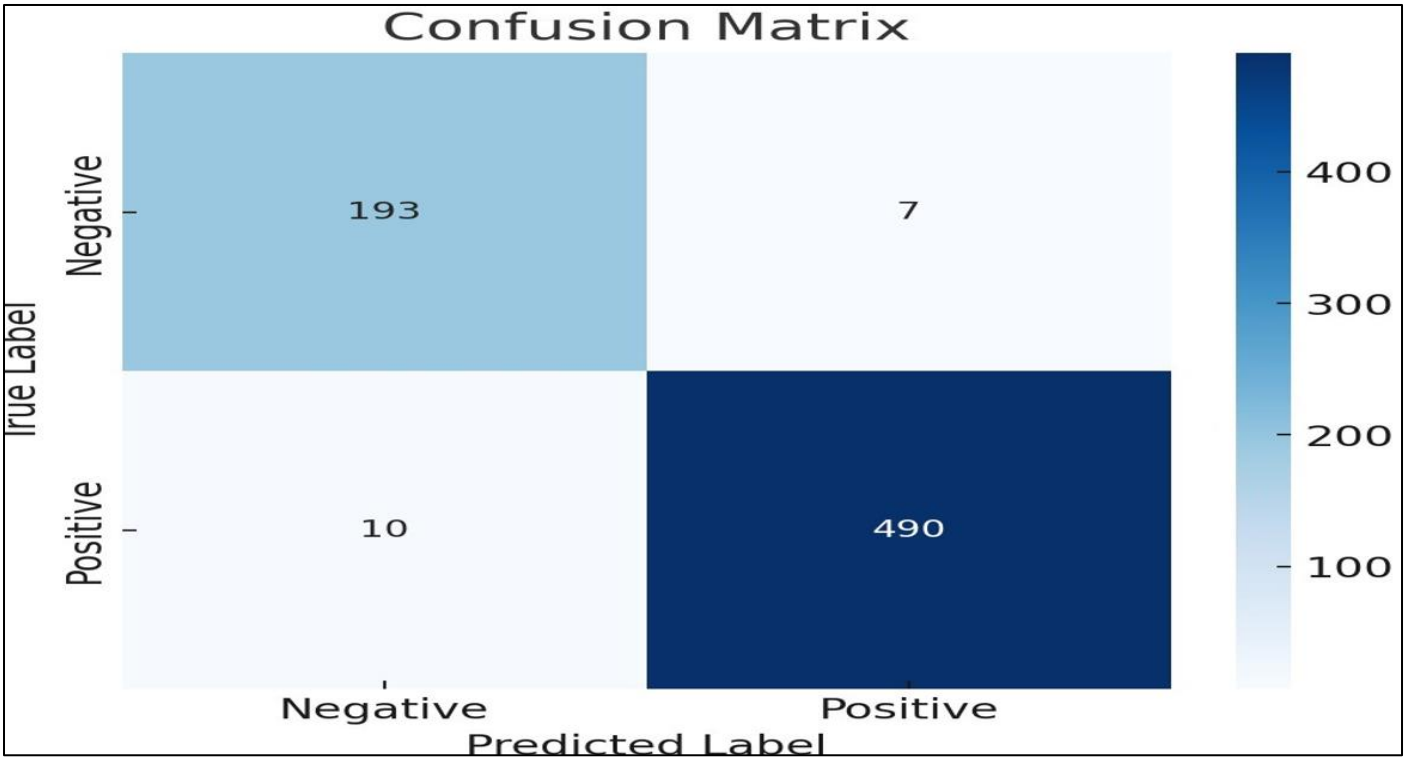

Fig 13 Confusion Matrix for Deepfake Video Detection Model

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec111

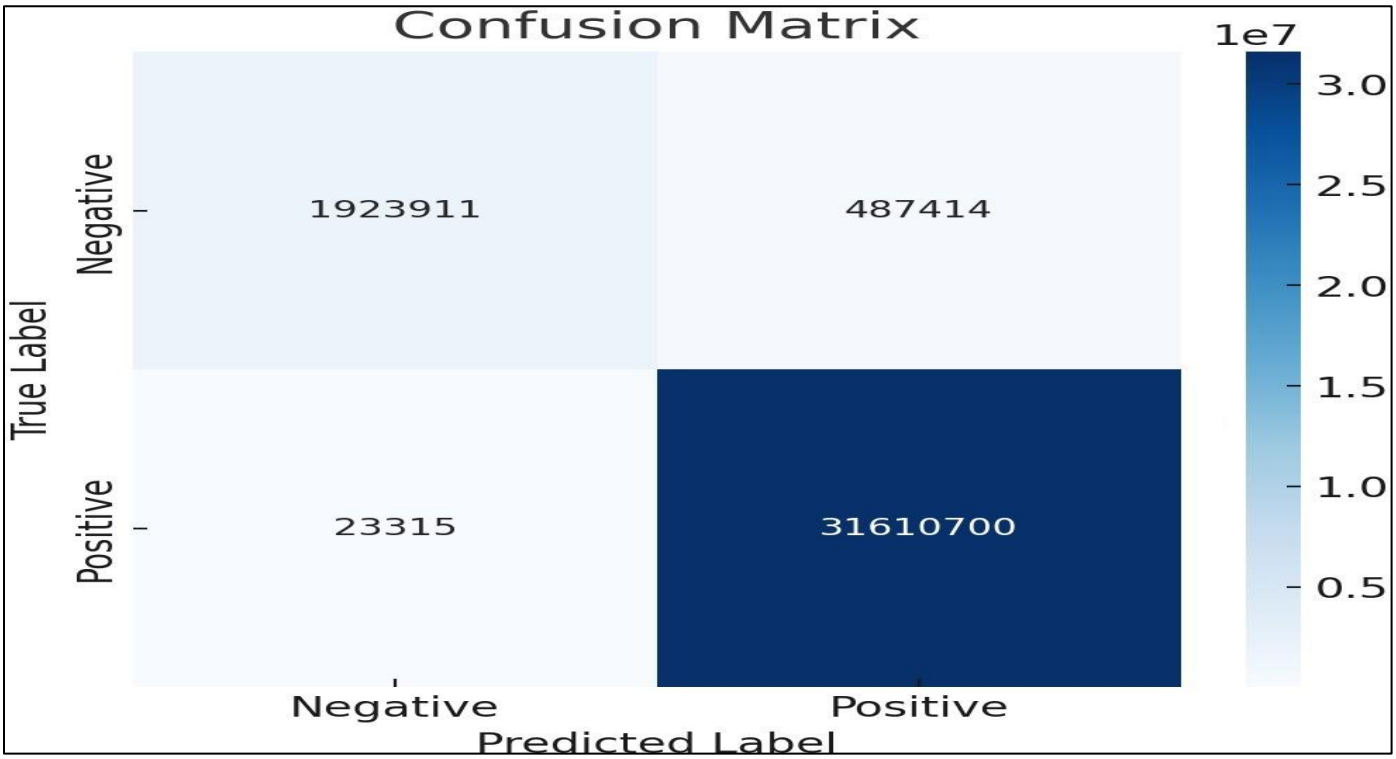

Fig 14 Confusion Matrix for Deepfake Audio Detection Model

This effectively demonstrates the performance of advanced deep learning models, such as XceptionNet, when used with LSTM networks for detecting deepfake content in any form of media. The extraction of spatial and temporal features is the reason for their accuracy in recognizing fake images, videos, and audio streams. The experimental results showed a very high accuracy for the developed models. Figure 13 and 14 show the confusion matrices for the deepfake video and audio detection models. The detection accuracies of the video- and audio-based systems are97 %, and 98 %, respectively, using XceptionNet + LSTM and CNN + LSTM

V. CONCLUSION

To solve the problem of identifying faked material in photos, videos, and audio, a novel three-stage deepfake detection framework is created employing cutting-edge deep learning techniques. With an accuracy of 95.56% on the Celeb dataset, Xception-Net shows promise for image-based media. A novel approach that combines CNN and LSTM networks achieves 98.5% accuracy on the DEEP-VOICE dataset in the field of audio deepfakes. With an accuracy of 97.574% on the Forensic++, DFDC, and Celeb-DF datasets, XceptionNet and LSTM networks function well together for video-based deepfake detection. This shows a great deal of progress toward a complete deepfake detection system by detecting several classes of deepfakes with excellent accuracy. These findings have practical consequences for government agencies, social media firms, and media verification platforms in the fight against false information. By integrating these models into real-time systems, stakeholders will be able to significantly improve their capacity to identify and stop deepfakes, protecting digital content integrity in modern information ecosystems. However, when the model is subjected to new or unseen manipulations, the performance changes, requiring additional research to improve generalizability. Additionally, these models struggle with scalability and real-time applicability, especially in contexts with limited resources.

REFERENCES

[1]. Agarwal, S., Farid, H., El-Gaaly, T. and Lim, S. N.: Detecting deep-fake videos from appearance and behavior, In: Proc. Of IEEE international workshop on information forensics and security (WIFS), pp. 1-6 (2020).

[2]. Müller, N. M., Czempin, P., Dieckmann, F., Froghyar, A., and Böttinger, K.: Does audio deepfake detection generalize? arXiv preprint, arXiv:2203.16263 (2022).

[3]. Lyu, S., Deepfake detection: Current challenges and next steps. In: Proc. Of 2020 IEEE international conference on multimedia & expo workshops (ICMEW), London, UK, pp. 1-6, (2020).

[4]. Goodfellow I., Jean, PA., Mehdi, M., Bing, X., David, WF., Sherjil, O., Aaron, C. and Yoshua, B.: Generative adversarial networks., Commun. ACM, 63(11), pp. 139-144 (2020).

[5]. Eiter, T. and Mannila, H.: Computing discrete Fréchet distance, technical report CD- TR 94/64, Technische Universitat Wien (1994).

[6]. Došilović, F.K., Brčić, M., and Hlupić, N.: Explainable artificial intelligence: A survey. In: Proc. Of 41st International convention on information and communication technology, electronics, and microelectronics (MIPRO), Opatija, Croatia, pp. 0210-0215 (2018).

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25dec111

[7]. Guo, W.: Explainable artificial intelligence for 6G: Improving trust between human and machine, IEEE Communications Magazine, 58(6), pp. 39-45 (2020).
[8]. Salih, A., Raisi-Estabragh, Z., Galazzo, I. B., Radeva, P., Petersen, S.E., Menegaz G. and Lekadir, K.: Commentary on explainable artificial intelligence methods: SHAP and LIME, arXiv preprint arXiv:2305.02012 (2023).
[9]. Lujain I., Mesinovic, M., Yang, K. W., and Eid, M. A.: Explainable prediction of acute myocardial infarction using machine learning and shapley values, IEEE Access, (8), pp. 210410-210417, 2020.
[10]. Ramprasaath, S. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. Grad cam: Visual explanations from deep networks via gradient-based localization. In: Proc. of IEEE international conference on computer vision (ICCV), Venice, Italy, pp. 618-626 (2017).