IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 26, Issue 3, Ser. 1 (May. – June. 2024), PP 64-71
www.iosrjournals.org

Video Deception: An Analysis of the Object-Oriented Frame Identification Method

[1]Chikati Madhava Rao, [2]Dr Santosh Kumar Yadav
[1]Research scholar, Research scholar, Shri Jagdishprasad Jhabarmal Tibrewala University, Vidyanagri, Jhunjhunu, Rajasthan.
[2]Professor & Director Research, Shri Jagdishprasad Jhabarmal Tibrewala University, Vidyanagri, Jhunjhunu, Rajasthan

ABSTRACT: *Deep learning, a powerful tool for learning, has evolved significantly due to the development of recurrent neural networks and perceptrons. However, optimal training requires overcoming challenges such as data preparation, optimization schemes, regularization processes, and hardware acceleration. Academics are now exploring methods to apply transfer learning, a contemporary deep learning technique, to the analysis of structured data in healthcare and finance. Ethical issues such as privacy and data bias are being addressed through self-supervised and federated learning. Hybrid models combining deep learning with other artificial intelligence techniques, such as rule-based systems and symbolic reasoning, are being explored. Multi-modal learning and explainable AI (XAI) are also being developed to improve the usability of deep learning models. Deep learning models have revolutionized various fields, including chemical characteristics prediction, material discovery, medication formulation acceleration, and quantum mechanical equations. They are also used in fields like genetics, astronomy, visual art, music, and visual art. Deep learning models are essential in various domains, including real-time language translation, art and creativity, autonomous vehicle research, language competency, climate data analysis, and prediction and prior warnings. Natural language processing methods are used in training these models to identify inappropriate language, uphold community standards, and promote positive interactions in online spaces. Generative models like Generative Adversarial Networks and Variation Auto encoders are crucial for the production of unique material in transdisciplinary creativity. Deep learning algorithms can provide warnings and predictions, and transfer learning methods can enhance NLP jobs in languages with limited resources.*

KEYWORDS: *Deep Learning Models, GEN AI, Deep Fake video detection and CNN*

I. INTRODUCTION

Recently, popular social media networks have begun circulating deep-fake videos. One main way these manipulative videos are made is by replacing a person's face with another and then adding genuine expressions. Perhaps far more dangerous than first believed, deep-fakes target huge audiences while most people are unaware bluff, they spread. The distribution of extortion, pornography, misleading news, and surveillance video is only a few examples harmful intents that may be shown via deep-fakes.

Numerous programs and technologies that can create photo-realistic deep-fake videos and images are freely available to the public. When it comes to communicating and sharing information in the modern digital age, video is among the most crucial tools at our disposal. If, after seeing or listening to any video material, we found out that it was false and had some evil intentions, it would utterly destroy our trust in our modern, tech-based society. We need to fix this if we want people to trust technology again and have a good relationship with it. The development of an algorithm capable of distinguishing between real and fake videos is currently critical in order to reduce the potential spread of deception caused by these altered films. There are a limited number of methods and answers to this problem, despite numerous research that various groups and individuals have undertaken. One is the broad use of GAN, which is in addition to the deep and substantial use of neural networks in this sector.

II. STUDY ON DEEP FAKE

Deep fakes, which are entirely synthetic multimedia that provide a risky approach to conduct a variety of fraudulent activities, such as spreading false information and identity theft, have unfortunately become better because to recent advancements in deep learning. The term deep-fake, a portmanteau of deep learning and fake, refers to the use of AI and machine learning algorithms in the production or editing of deceptive media content. As their misuse and capacity to taint actual information becomes more apparent, we need to work towards developing automated systems that can detect deep-fakes effectively so that we can remove them from circulation before they do additional damage.

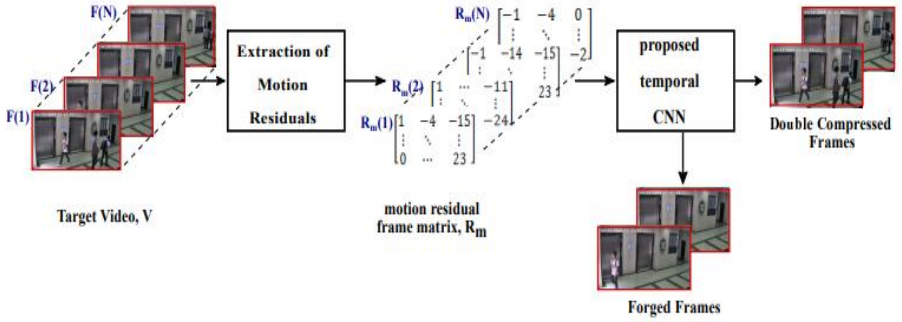
A dissertation including deceptive material in all four digital modalities—text, video, images, and audio is called Deep-fake. Deep-fakes that employ audio, and especially human voice, pose a greater threat because to the extensive biometric usage of speech in today's world. Security measures for speaker identification and verification systems, as well as other forms of phone-based and internet access control to financial and other portals, are enhanced by speech systems. One possible attack vector for such systems is audio spoofing. Converting voices, creating fake speech, playing back audio, etc., could all be part related approaches. The speaker's identity is concealed by impersonating the target or another person in several techniques, which might be seen as variants on vocal disguise. Wishing, efforts to breach speech authentication systems, fraudulent calls, and similar audio-based crimes often use these disguise-based strategies. Automated voice biometric systems are already vulnerable to voice masking, and forensic speech analysis is further complicated. Because deep-fakes make voice masking more effective, these concerns are made worse. The advancement of deep learning techniques, especially generative models such as generative adversarial networks and Wave Net models, is bringing synthetic speech closer and closer to real speech.

The modification of visual content to disseminate misinformation is a major issue in the modern digital era. Because faces are fundamental to human communication, controlling them is more desirable than manipulating other objects. Deep fakes is an amalgamation of Deep Learning and fake that describes the phenomenon of face swapping utilizing a deep learning technology. In the beginning, most Deep-fake videos were made in the film, entertainment, and advertising industries with legitimate aims. Making up faces is one excellent way to avoid utilizing real people's photos. This way, companies may stay out of legal hot water when it comes to royalties, copyrights, and other issues. However, concerns are rising that Deep fakes are being used for illicit or immoral purposes, such as in politics or pornography. An example of this is the near-death experience that Noelle Martin had when an explicit Deep fake film of hers was released online without her knowledge or permission. The method for detecting a modified video frame using artefact of object-based forgeries found in a frame's motion residuals is described in the preceding section. Here we provide a thorough explanation of motion residuals, and then we show a flowchart that illustrates the process of extracting motion residuals from a given frame. A video's most crucial component is its motion residuals. The artefact of object-based forgeries may be captured by these motion residuals. The six successive frames lift scenario from the 00055.mp4 video SYSU-OBJFORG datasets [20] are shown in figure 3.2. Everyone in all six frames is either standing or walking around on a static backdrop. Consequently, these frames' background information is comparable. Nevertheless, the orange-encircled individual in figures 3.2a–3.2f is in motion, and the brightness of its pixels are changing.

III. PROPOSED MODEL

By computing motion residuals, we may eliminate the unnecessary data that is included in the frame sequence. In order to calculate the motion residual, a reference frame is used. Using the I frame as a reference frame for motion residual extraction is not a good idea since, as mentioned earlier, the GOP size of advanced codes movies is adjustable. Accordingly, our study uses the same collusion strategy as [8] to extract motion residuals from movies encoded using sophisticated code's. The reference frame is calculated via a statistical procedure on a certain number of consecutive frames in the collusion technique. Making greyscale frames from a provided video is the first step in the extraction process. Also, before extracting the motion residual, choose the frame that corresponds to it. A temporal window of frames is generated by selecting an equal number of frames before and after the selected frame, with the selected frame being considered as the middle frame. For the purpose of motion residual extraction, a reference frame is obtained by performing a pixel-wise median operation on the selected window of frames. In order to get the motion residual that is specific to the present frame, the reference frame is removed from it. Each video frame is transformed into a motion residual frame by alliteratively repeating this technique.

Figure 1: The six consecutive frames of 055.mp4 video of SYSU-OBJFORG datasets representing the orange encircled persons movement in lift scene.



The video frames may be identified as faked or double compressed using the suggested approach, which identifies object-based forgery at the frame level. Two layers make up the proposed method: one for frame categorization and one for motion residual extraction. Forged frames and double compressed frames make up the object-based fabricated video. In order to identify frame-level forgeries, a conventional neural network (CNN) is developed and used to categories frames as either fake or double compressed. The suggested conventional neural network (CNN) is called a temporal-CNN since it uses the residual time information of a moving video.

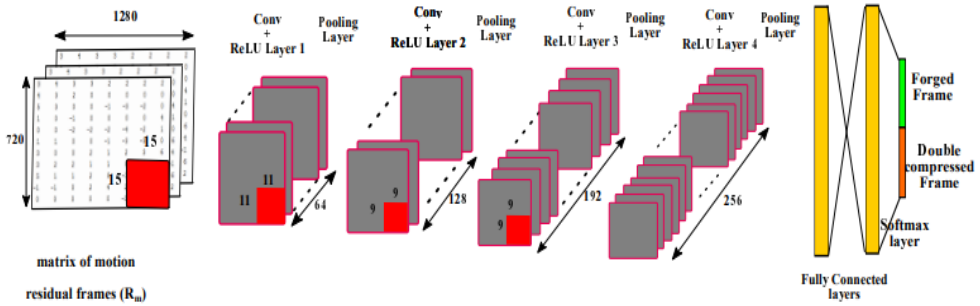


Figure 2: Representing the detailed architecture proposed temporal-CNN.

The proposed temporal-CNN is a deep learning model designed for high-definition movies, utilizing conventional layers with 15x15, 11x11, and 9x9 sized kernels. The ReLU activation layer imparts non-linearity onto the final feature maps. The feature maps generated by the maximum pooling method have less dimensions. The temporal-CNN is tested on Xenon PCs with 64 GB of RAM and running MATLAB 2018a.

Chen et al. developed the object-based fabricated video datasets SYSU-OBJFORG, which consist of 100 original 720p films and 100 synthetic movies compressed using the H.264 codex. The datasets are divided into two groups, forged and double compressed, and used to train and test the temporal-CNN.

To assess the proposed temporal-CNN's robustness, ten test videos are retrieved from the SYSU-OBJFORG datasets, resulting in SYSU-OBJFORG VFVL. Ten films are selected randomly from the SYSU-OBJ datasets, one being a fake and the other real. All fabricated videos have their frame size and running time changed, allowing for the removal of extraneous elements. The same logic applies to the preservation of fake frames during the chopping process of video frames. In summary, the proposed temporal-CNN is a deep learning model that uses conventional layers and the ReLU activation layer to extract intrinsic characteristics of high-definition movies.

Table 1: The detailed description of SYSU-OBJFORG VFVL datasets.

S.No	Video Name	Frame Size	Total Frames
1	00003_125_254.mp4	640 X 1024	151
2	00004_000_118.mp4	640 X 1024	186
3	00015_133_268.mp4	512 X 896	159
4	00018_158_267.mp4	512 X 1024	130
5	00026_000_125.mp4	640 X 1024	158
6	00039_000_409.mp4	512 X 1152	181

The suggested temporal-CNN is evaluated and discussed in this part under several test cases. We compare the outcomes suggested approach to those of Chen et al. Forged frame detection using stenography and the work of D'Aminao et al. Detection technique based on patch matches [3]. In addition, activation maps are used for research and analysis. As a last step, we test the suggested approach for post-processing attacks. The size temporal window is a critical element in the motion residual extraction approach. To determine the optimal temporal window size for the temporal-CNN that has been suggested, a battery of tests has been run. The SYSU-OBJFORG datasets [20] contains two sets of fake videos: one for training and one for testing. Eighty films make up the training group, while twenty make up the testing group. In order to identify motion residual frames that are both compressed and fabricated, the temporal-CNN is trained using 80 fabricated movies. The trained temporal-CNN is evaluated for the motion residuals with varying window sizes. For this purpose, we extract motion residuals from the 20 test movies using a variety of temporal window widths. For motion residual extraction, 11, 15, 19, 21, 23, and 25 are the window sizes that were used. Since the datasets contains at least one fabricated segment of 25 frames in length, any temporal window size greater than 25 is not chosen.

Testing error probability Pe and temporal-CNN frame ac-curacies for various window widths of motion residuals are listed in table 3.2. With a 19-inch window, the Forged Frame Accuracy FFAC reaches a maximum of 97.36% and the Pe-value is 0.0519. The improvement in Frame Accuracy (FAC) for window size 23 compared to window size 19 is a meagre 0.12%. While 23-pane windows have a superior FFAC, 19-pane windows are almost 1.3% better. The main goal suggested approach is to accurately identify counterfeit frames, and it does this by using a temporal window that is 19 frames in size. Based on the data in table 2, we will be doing more experiments and testing with a temporal window size of 19. For various temporal window widths, Table 2 shows the testing probabilities proposed temporal-CNN's error Pe, DFAC, FFAC, and FAC.

To identify manipulated or duplicated frames in a video, the suggested temporal-CNN was built. Here, video frames from the SYSUOBJFORG datasets [10] are used to train and evaluate the developed temporal-CNN. We compare the outcomes of testing the suggested temporal-CNN with those analytical approach developed by [11].

Table 2: The confusion matrix discussing the classification accuracy for multi-class

	DCFAC	FFAC	FAC
CC-PEV	91.01	86.02	88.5
SPAM	94.03	92.11	93.14
CF*	92.39	85.39	88.81
CC_JRM	92.67	85.46	89.11
Proposed method	98.94	96.04	97.49

classification using proposed temporal-CNN. (where 'DC' represents Double Compressed).

We also display and study activation maps proposed temporal-CNN to understand how it classifies frames as either doubly compressed or forged. In order to evaluate the activation maps, we retrieved the video with the filename 00055 016-117.mp4 from the SYSU-OBJFORG datasets [3]. Fifty randomly selected frames from this video, which has undergone two compression and manipulations, are used. Then, at random, we choose the filters

last conventional layer and use them to extract the activation maps temporal CNN that was recommended. Also averaged are double-compressed activation maps with dimensions of 73 by 143 pixels and fifty frauds. These activation maps are shown as color maps. When looking at the color bar, a very high level of activity is indicated by a dark red at the very top and a very low level of activity by a dark blue at the very bottom. The average activation maps for manufactured frames, while the average activation maps for doubly compressed frames are shows that compared to double compressed frames, manufactured motion residual frames have larger, more irregular blobs. Fig 1,2 reveal significant activation in the large blob region, which is shown by the dark red colour. Training temporal features take on the properties associated with the artefact, as seen in the activation maps.-CNN

Table 3: The result of proposed temporal-CNN for the post-processing attack on SYSU-OBJFORG datasets [20] for 15 and 30 fps videos.

FPS	DCFAC	FFAC	FAC
15	**89.98**	**94.40**	**91.73**
30	**72.06**	**97.84**	**82.01**

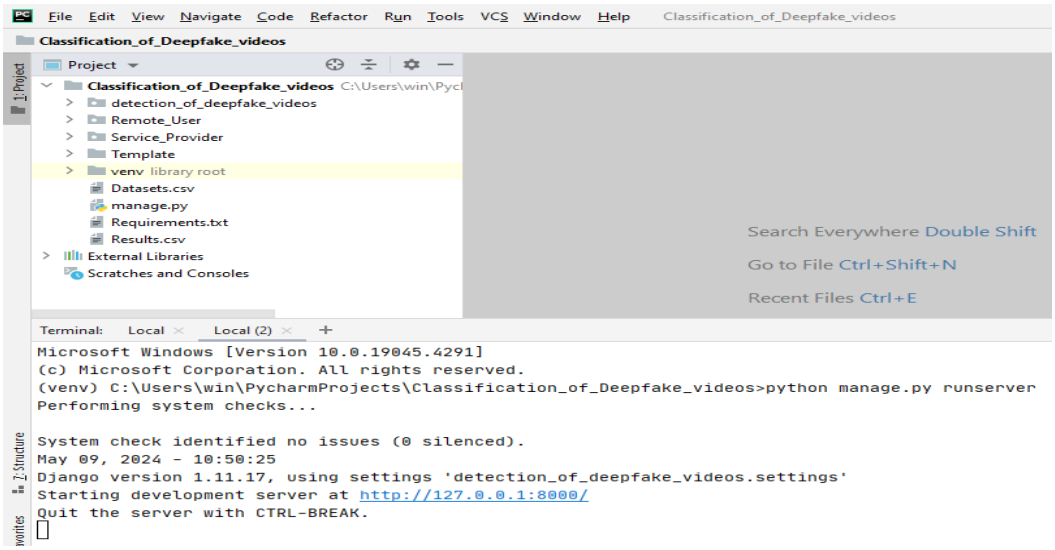
Fake detection is crucial in detecting digital forgeries, such as deep fakes in photographs, movies, and music. The study validates the construct model using statistical tools like Cronbach's Alpha and Pearson's Correlation and Confirmatory Factor Analysis. The construct has a high Cronbach's Alpha reliability rating of 0.81, suggesting strong internal consistency. The study examines various types of digital forgeries, providing context for future efforts to explain predictions and build detection methods for digital forgeries in general. The literature on image, video, and audio signal modification detection is also reviewed.

The Video Rewrite Program in 1997 was a significant work in face modification research, integrating computer vision, image processing, and speech processing to create driven visual speech. Facial animation and manipulation saw significant advances in the next two decades, with Face2Face, Synthesizing Obama, Head On, Face swap-GAN, Pro GAN, Style GAN2, Wave Net, and Mel Net. Deep fakes classification is a challenge with two possible outcomes: genuine or fake. CNN models have been developed to detect deep fakes, such as Meso Net, Exception Net, and Efficient Nets. U-Net and Eff-Y Net are well-known models for picture segmentation, while Eff-Y Net combines an Efficient Net encoder with a classification branch. Segmentation is a useful technique for training classifiers and creating segmentation masks, and trained attention maps display both changed and informative areas, helping with binary classification. Overall, deep fakes classification has evolved significantly over the years, with advancements in image processing, deep learning, and segmentation techniques. Researchers have developed an alternative network architectural pipeline for deep fake detection, incorporating LSTM and CNN capabilities for feature extraction and temporal sequence analysis. This approach helps differentiate between real and fake videos, reducing the risk of overfitting due to large datasets. Dynamic Face was proposed to reduce Log Loss by 15.2% to 35.3% across multiple datasets. The study also presents a method for detecting deep fake films using Transfer Learning, demonstrating the effectiveness of per-trained networks in identifying AI-generated counterfeit movies. The CNN-LSTM model outperforms the CNN-GRU model in detecting and processing temporal discontinuities, making it suitable for large datasets. The CNN-LSTM model achieved an accuracy rate of over 80% in uncovering hidden false information in test samples.

IV. RESULTS AND ANALYSIS

In these results analysed how Recurrent Neural Network-RNN give better performance when comparing to svm and Gradient boosting classifier, implemented based on python with Django where libraries are above mentioned algorithms and web framework implemented.

Screen 1

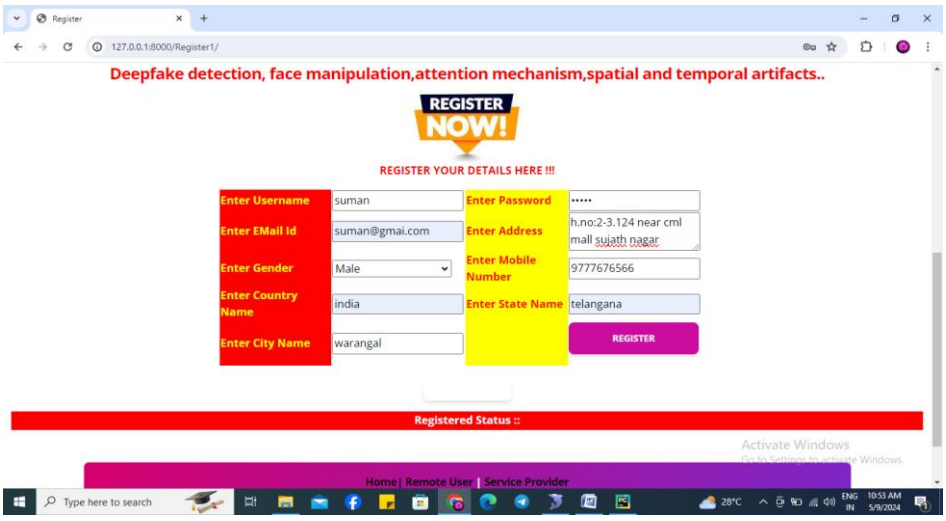


This is the homepage startup interface for adding additional libraries of applications, movie videos or any kind of videos, and datasets. If the user wishes to update any datasets or libraries, modifications can be made here.

Screen 2

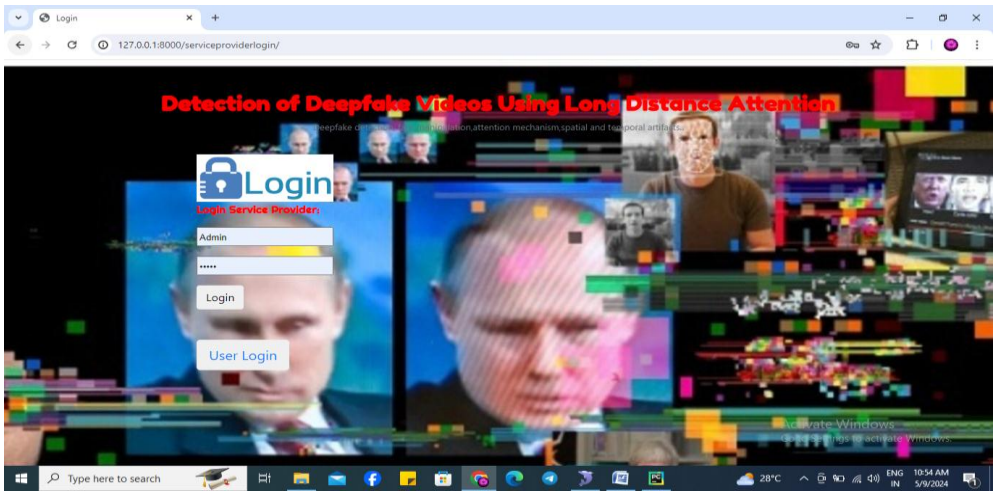
This application allows users and administrators to navigate to their own pages, and upon logging off, it redirects back to this page.

Screen 3



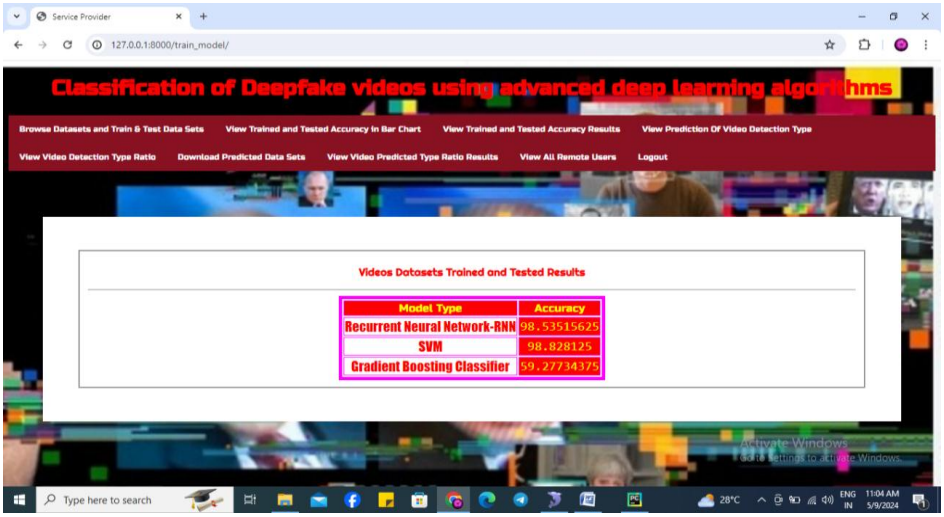
A new user can register with their demographic details at the top of the page; upon completion of registration, all information will be stored in the database.

Screen 4



This is a login page from which every user or administrator will access their own privileged pages. Prior to utilizing the deepfake program, users must authenticate their credentials.

Screen 5



1. Users may add datasets, analyse the proportions of various categories of deepfake movies, and assess the outcomes. This is the operational framework of such a website:
2. Upload Datasets: Users are permitted to upload their own films or datasets comprising various deepfake videos.
3. Ratio Analysis: The platform presumably offers instruments to evaluate the ratios or proportions of several categories of deepfake movies.
4. Video Assessment: Users may examine deepfake videos, potentially employing metrics such as realism, detection scores, or other criteria to evaluate the persuasiveness of the deepfake.
5. Result Visualisation: The platform may provide visual representations of the results, like charts or graphs, to assist users in comprehending the analysis of deepfake categories and their efficacy.

V. CONCLUSION

The detection of altered movies has become a crucial endeavour due to the increasing misuse of sophisticated synthetic media in several fields, including government, journalism, and personal safety. A method of video manipulation involves altering an individual's appearance or speech, often using advanced algorithms such as Generative Adversarial Networks (GANs). The difficulty is in distinguishing these false modifications from the genuine original. Deep learning is a powerful tool that can be integrated into a Django application to

identify and forecast deepfake videos. This process can be achieved by combining video processing tools like OpenCV with strong machine learning frameworks like TensorFlow or PyTorch. By integrating this detection model into a Django-based web application, users can easily classify videos as legitimate or deepfake. The Django framework allows for easy-to-understand web interfaces for users to perform detection, upload videos, and view findings. Machine learning models are used to efficiently analyze videos frame-by-frame to detect modifications that could be signs of deepfakes. The system can be personalized with advanced visualizations, real-time detection, and dashboards that are tailored to each user's video history. This approach simplifies the process of detecting deepfake videos and provides a scalable and user-friendly platform for analysis.

REFERENCES

[1]. Andrews, E. L. (2019). How fake news spreads like a real virus | Stanford School of Engineering. Stanford ENGINEERING.
[2]. Apuke, O. D., & Omar, B. (2020). Fake news and COVID-19: modelling the predictors of fake news sharing among social media users. In Telematics and Informatics.
[3]. Durall, R., Keuper, M., Pfreundt, F. J., & Keuper, J. (2020). Unmasking DeepFakes with Simple Features. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.
[4]. Kumar, S., Singh, S., & Kumar, J. (2019, January). Gender classification using machine learning with multi-feature method. In 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0648-0653). IEEE.
[5]. Prasadi Peddi and Dr. Akash Saxena (2015), The Adoption of a Big Data and Extensive Multi-Labled Gradient Boosting System for Student Activity Analysis, International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211, Volume 3, Issue 7, pp:68-73.
[6]. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
[7]. Sun, Y., Wang, X., & Tang, X. (2013). Deep convolutional network cascade for facial point detection. In Proceedings IEEE conference on computer vision and pattern recognition (pp. 3476-3483).
[8]. Yang, H., Cho, K. C., Kim, J. J., Kim, J. H., Kim, Y. B., & Oh, J. H. (2023). Rupture risk prediction of cerebral aneurysms using a novel convolutional neural network-based deep learning model. Journal of NeuroInterventional Surgery, 15(2), 200-204.
[9]. Yang, H., Zhu, K., Huang, D., Li, H., Wang, Y., & Chen, L. (2021). Intensity enhancement via GAN for multimodal face expression recognition. Neurocomputing, 454, 124-134.
[10]. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In Proceedings IEEE conference on computer vision and pattern recognition (pp. 2921-2929).
[11]. Zhou, P., Han, X., Morariu, V. I., & Davis, L. S. (2017). Two-Stream Neural Networks for Tampered Face Detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)