International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Implementation of Machine Learning Techniques to Assess the Authenticity of Multimedia Data

Shruthi S.[1], Manjunath R.[2], Shivashankar[3]

Submitted: 12/12/2023 **Revised**: 16/01/2024 **Accepted**: 01/02/2024

Abstract: Deep learning algorithms have become so strong because of expanded registering power that it is currently moderately simple to deliver endeavoring to utilize man-made brainpower (simulated intelligence). The casing level highlights are removed by our framework utilizing a Res-Next Convolution human-like counterfeit recordings, some of the time known as "profound fakes." The utilization of these life like face traded deepfakes to incite political turmoil, stage fear monger assaults, or blackmail individuals is effectively possible. In this paper, we give a clever profound learning-based framework that can effectively recognize between computerized reasoning produced bogus recordings and genuine ones. Naturally spotting substitution and re-order deepfakes is conceivable with our strategy. To battle man-made consciousness (computer based intelligence), we are brain organization, and these elements are then used to prepare the Long Momentary Memory (LSTM) based Repetitive Brain Organization to decide if the video has been controlled in any capacity, i.e., whether it is a deepfake or a genuine video. We test our strategy on a sizable amount of adjusted and blended informational collections made by joining the different open informational indexes like Face-Forensic++, Deepfake discovery challenge, and Celeb-DF to reenact continuous situations and work on the model's presentation on constant information. We likewise exhibit how our framework might create serious outcomes with an extremely clear and dependable methodology. In the mixed media information handling, the sight and sound information handling innovation is clearly better than the information mining innovation and information pressure innovation. At long last, under the help of profound learning information, we reason that sight and sound information handling innovation is generally utilized and cited in different fields. Accordingly, with the advancement of media, how much interactive media information is expanding; in this way, we ought to enthusiastically foster sight and sound information handling innovation in an overall manner.

Keywords: *Convolution Neural Network, Generative adversarial networks, Repetitive Brain Organization, LSTM, Supervised learning.*

1. Introduction

Deepfakes, accumulate of the terms "deep learning" and "fake," are produced by employing techniques such as overlaying the face photos of an objective person onto the profiles of a source person in order to produce films in which the objective person performs or says anything that the source person says. This belongs in the face-trade subset of deepfakes. Deepfakes, as the term is more often used, are impressive performances given by artificial intelligence that can also be lip-matched. Deep lip-sync fakes are recordings that have been altered to sync the mouth of movements with the audio. Manikin ace deepfakes are recordings.

an impartial person energetically imitating the movements of their eyes, heads, and faces of another person, who is situated. Deep learning models, like auto encoders and generative adversarial networks (GANs), which have seen widespread use in the computer vision field, are currently the common underlying mechanism for producing deepfakes, even though some can be created using traditional visual effects or computer graphics techniques [1-2]. These models are used to assess the emotions and movements on the face and to generate synthetic pictures of many individuals with comparable facial expressions and movements. Deepfake approaches sometimes require large amounts of image and video data in order to train models to produce lifelike photographs and movies. Deepfakes sometimes originate with celebrities and politicians because they frequently have a large online presence through photographs and videos.

Deepfakes were used to change the bodies of actors playing politicians or other celebrities in fake photos and videos. The first deepfake video featuring an actor's face instead of a celebrity's was released in 2017. Global security is jeopardised by the use of very sophisticated forgeries to produce recordings of world leaders and the distribution of fake speeches. Deep fakes can therefore be used to spread false information, wreak havoc on the financial system, trick voters and sway election outcomes, or incite international

[1]*Assistant Professor & Research Scholar, Computer Science Engineering, R R Institute of Technology, Bangalore-560 090, Visvesvaraya Technological University, Belagavi-590 018, India.*
ORCID ID: 0009-0009-4441-5975
[2]*Professor & Head, Computer Science Engineering, R R Institute of Technology, Bangalore-560 090, Visvesvaraya Technological University, Belagavi-590 018, India.*
ORCID ID: 0000-0002-7335-5975
[3]*Professor, Electronics & Communication Engineering, R R Institute of Technology, Bangalore, Visvesvaraya Technological University, Belagavi-590 018, India.*
ORCID ID: 0000-0001-7694-4815
** Corresponding Author Email: shruthisrinivas.1618@gmail.com & drmanjunath.raj@gmail.com*

conflict over politics or religion.

The ease of creating and sharing digital videos has increased due to the growth of social media and media sharing platforms, as well as advancements in mobile camera technology. Deep learning has made it possible to achieve technologies that were thought to be unattainable just a few years ago[3]. One of these is the modern generative model, which can synthesise audio, video, images, and speech that are remarkably lifelike. The world has become more accessible thanks to text-to-speech, and these models have produced training data for medical imaging among many other uses.

Similar to any groundbreaking technology, this has created new challenges. Deep generative algorithms that can alter audio and video samples produce "deepfakes". Since the release of the first open-source deepfake producing tools and approaches in late 2017, an increasing number of synthesised media clips have been produced [4]. While the majority are presumably intended to be funny, some might be harmful to both individuals and society as a whole. Because editing tools are now widely available and subject matter experts are in high demand, there were more realistic phoney videos until recently. Deepfakes are being shared on social media more and more, which promotes spamming and the dissemination of false information. Imagine a plausible hoax where our country's president declares war on its neighbours or where a well-known celebrity targets their fan base. The public could be scared or misled by these crafty, malevolent fakes.

Face-swapping is the technique of artificially changing a subject's identity in a photo or video by automatically substituting a different face for the original one. Two pairs of encoders and decoders were used in the procedure; however, the two encoders shared training parameters while the decoders were trained independently. For instance, in face scenarios, face A enters the encoder first, and then face B replaces face A in the decoder by connecting to it[5]. This technique required a challenging instruction process and a significant amount of processing power. The quantity of target images and videos that were accessible for use as training materials determined how effective this strategy was. However, since this methodology led to the creation of the first deepfake technique, it has gained more popular. Faceswap-GAN is a popular face swap method. By extending the creative deepfakes method, Faceswap-GAN improves the output of the automatic coding system by incorporating antagonism and perceptual loss. By incorporating counter losses, the final image's reconstruction quality is enhanced. Eye orientation is enhanced and the output picture's face aligns with the input image when perceptual loss is taken into account. The technique of deepfakes is enhanced by this technology.

The technique of keeping one's own identity while assuming the posture and facial expression of a source character in a video is known as face-reenactment. Face2Face precisely re-renders the output video while animating the target actor's facial expressions in the target video using real-time face reconstruction. Numerous studies have been conducted to examine different aspects of deepfake video detection systems. A human, for instance, blinks once every two to ten seconds, with a maximum half-second interval between each blink. Deepfake films can be distinguished from authentic ones more easily since their subjects hardly blink at all.

The shade of the eyes can be used to identify medium changes in addition to blinking patterns. It was found that there was a lack of global uniformity as a result of the large range of hair and eye colour employed to create fictional representations. They proposed a method for identifying iris pixels to determine the characteristics of eye colour and for differentiating facial features in an image. This makes it possible to distinguish deepfake photos or videos from their authentic equivalents.

Understanding the difference between real and deepfake movies becomes crucial. We are using Artificial intelligence to counter AI. Deep fakes are made via apps like FaceApp and Face Swap, which employ pre-trained neural networks like GAN or Auto encoders. In our method, the chronological temporal investigation of the video frames is handled by an LSTM-based artificial neural network, and the frame-level features are extracted by a pre-trained Res-Next CNN. The ResNext Convolution neural network can be used to gather frame-level data in order to determine the authenticity or Deepfake state of a video. Fake detection technology needs to be developed in order to identify deepfakes and prevent them from spreading online. We offer a novel deep learning-based technique called Deepfake Video Detection that can reliably differentiate between real and synthetically generated bogus videos. While there is no shortage of tools for producing deepfakes, there are few for identifying them. Our technique for detecting deepfakes will greatly limit their online distribution. We will provide a web-based interface via which people may upload movies and mark them as real or fake. This concept can be further upon by developing an automated deepfake detection online platform or browser plugin. Popular apps like WhatsApp and Facebook's software may use this concept to make it simple to identify deepfakes before they are shared with other users.

2. Literature Survey

FPGAN, a novel face de-identification technique built on the Generative Adversarial Network (GAN), was developed by the authors of this research. FPGAN is a complete solution designed to safeguard privacy related to face and vision [6]. In order to increase the FPGAN's feature extraction capability, the suggested approach uses two discriminators in addition to an upgraded U-Net generator with a seven-level network architecture. In order to ensure FPGAN performance, the authors also suggested a number of loss functions and an optimisation technique. In their experiment, they used FPGAN to prevent social robot anonymization and examined factors that may have an impact on model performance. Additionally, they suggested a whole new evaluation process for privacy protection and face de-identification. The authors evaluated their approach against four industry-standard methods that are already being used on a range of datasets.

Complex fake videos could previously only be produced by Hollywood studios or government actors. With a minimal quantity of data and computing power, anyone can now more easily produce engaging deepfake videos thanks to recent developments in deep learning. Negative outcomes from these recordings could include a world leader admitting to misconduct or a business tycoon manipulating stock prices around the world. The political process, national security, and society at large are all seriously threatened by these deepfakes. A forensic technique [7] that imitates the gestures and facial expressions connected to a person's speech pattern has been created to address this growing issue. When it comes to deepfake movies, these correlations can be broken, offering an authentication method.

A recent study has shown that convolutional neural networks may be trained to produce remarkably accurate models of human heads [8-9]. Nevertheless, a substantial dataset of photos of a single individual is needed in order to use this technique to create customized talking head models. Customized talking head models can only be trained on a small number of image views possibly even just one because real-world conditions could not make it possible to gather such a huge dataset. A few-shot capable system has been created in order to solve this problem. The system uses high-capacity generators and discriminators to do extensive meta-learning on a large dataset of movies. Next, it learns neural talking head models of individuals who have never been seen before, both few- and one-shot.

Deepfake technology has been the subject of a lot of attention lately. Deepfakes are developing swiftly, endangering the security and reputations of their victims. This technology automatically substitutes a fake face in photos and videos with a real one by using deep learning techniques [10]. This research by Luca Guarnera, Cristina Nastasi, Oliver Giudice, and Sebastiano Battiato details the method used to create deepfake facial photos. Additionally, the writers use normal procedures to do a forensic study on these photographs. Unfortunately, modern methods are not always reliable for verifying the legitimacy of these photos. The authors suggest a novel method that examines abnormalities in the frequency domain to pinpoint this problem.

In the era of fake news, large-scale public datasets and advances in deep learning techniques, particularly with regard to Generative Adversarial Networks (GANs), have made it possible to produce incredibly realistic fake content [11-12]. This article provides an in-depth analysis of ways for modifying facial photographs, such as Deepfake algorithms, and approaches for spotting these modifications. According to the research, there are four types of facial manipulations: total face synthesis (i), attribute manipulation (iii), expression switching (iv), and identity swapping (i.e., Deep Fakes) (ii). Important criteria for assessing false detection tactics for every manipulation group are included in the study, along with information on manipulation methods and publicly accessible databases. The study also presents the findings from several assessments that were carried out to ascertain the efficacy of the current false detection techniques. The study focuses on the improvements brought forth by the most recent generation of Deepfakes and the difficulties in detecting them. The report presents open topics and potential trends that should be taken into consideration to improve the profession, in addition to the survey results. The study demonstrates the increasing danger that remarkably realistic phoney content poses and the demand for sophisticated false detection methods. It is a useful resource for academics and professionals in photo editing and fake detection.

For many image and video synthesis tasks, the generative adversarial network (GAN) framework has proven to be a very useful tool [13-14]. It has made it feasible to produce high-quality, photorealistic photographs and films, something that was previously unfeasible or challenging to accomplish with conventional techniques. Additionally, it has created new opportunities for content creation across other industries. This article presents a thorough overview of GANs, emphasising the methods and software used in visual synthesis. The study discusses a number of significant methods created to stabilise the notoriously challenging GAN training. It also examines the various applications of GANs, such as image processing, neural rendering, video synthesis, and image translation.

Deepfakes, or Artificial Intelligence (AI) produced fake videos and audios, have been called into doubt by some as proof in a number of industries. This article addresses the issues raised by deepfakes as well as possible directions for further investigation. One of the main problems is that deepfakes are hard to spot as fakes because they may fool both humans and robots. Deepfakes have the ability to propagate false information, sway public opinion, and in rare circumstances, even be harmful. The author suggests looking into the development of comprehensible and reliable deepfake detection techniques in order to allay these worries and improve the reliability of audio and video evidence [15-16]. These techniques ought to provide an explanation for how the results were arrived at, and an independent verification of their dependability is required. In order to do this, the authors present a unique technique known as "dynamic prototypes," which entails teaching models to distinguish between authentic and fraudulent media by dynamically modifying their prototypes in reaction to incoming data.

This article suggests a novel method for visualising the identification of phoney facial images using dynamic prototypes. Many deep fake detection algorithms now in use fail to take into account the temporal anomalies present in deep fake films because of their frame-by-frame and black-box architecture. These temporal artefacts are crucial for a human supervisor to recognise and understand deep fakes. An understandable and practical alternative is the Dynamic Prototype Network (DPNet), which use dynamic prototypes or representations to explain the temporal artefacts in deepfakes [17-18]. Through extensive testing, including on unidentified datasets like Celeb-DF, Deeper Forensics, and Google's Deepfake Detection, DPNet has proven its effectiveness in predicting deepfakes. Furthermore, deepfake dynamics are explained simply by DPNet. Moreover, the research formulates temporal logic specifications by utilising these dynamics to guarantee that the model conforms to intended temporal behaviours.

3. Design and Implementation of Machine Learning Techniques to Assess the Authenticity of Multimedia Data

We trained our PyTorch deepfake recognition model on an equivalent amount of real and fake movies in direct to avoid model bias in this method. The system architecture of the model is shown in the image. Deep network of profound learning has been effectively applied to solo component learning of single mode mixed media. In profound realizing, this paper proposes another utilization of profound organization learning highlights in various modes and a progression of errands for multimodal learning and tells the best way to prepare profound organization learning elements to tackle these undertakings.

3.1 Methodology for deep fake detection

After pre-processing, we were able to extract 1500 Real and 1500 Fake movies from the DFDC dataset. The FaceForensic++ (FF) dataset has 1000 Real and 1000 Fake videos, compared to 500 Real and 500 Fake movies in the Celeb-DF dataset. This indicates that 3000 of the 6000 films in our collection are authentic, while the other 6000 are fake. Figure displays the data sets' distribution.

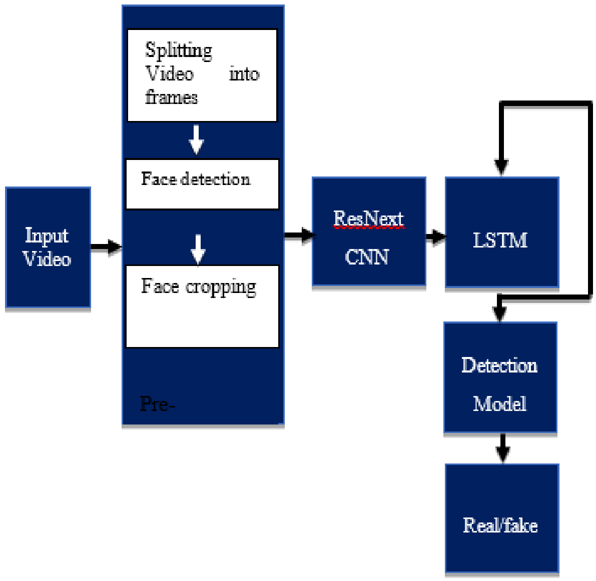

Fig. 1: Structure of the Machine Learning Techniques to Assess the Authenticity of Multimedia Data.

3.1.1 Compiling datasets

To augment the model's capacity for real-time prediction. The data was gathered using a variety of publicly available datasets, including FaceForensic++ (FF), Deepfake Detection Challenge, and Celeb-DF. Additionally, we merged and gathered datasets to create our own special dataset, which allowed us to quickly and effectively classify different movie genres. In order to reduce the algorithm's training bias, we looked at using a blend of 50% actual and 50% fake videos. Since audio deepfakes are outside the purview of this study, some audio-alarmed movies are included in the Deep Fake Detection Challenge (DFDC) dataset[6]. We pre-processed the DFDC dataset to remove the audio-altered films using a Python script.

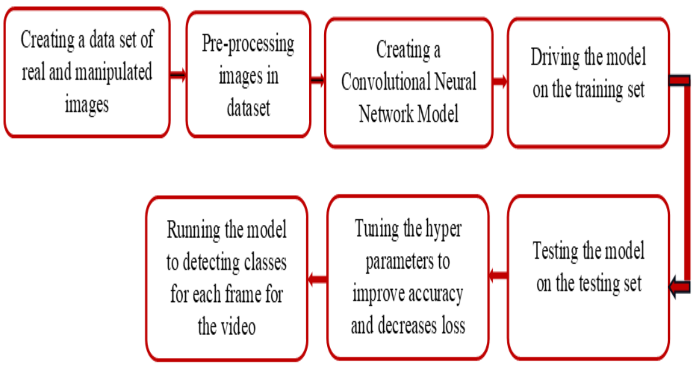

Fig. 2: Dataset and Pre-processing steps.

3.1.2 Pre-processing

In this stage, any superfluous noise is removed from the videos. To maintain the movie inside the allotted running time, facial recognition and cropping are employed. The first stage in creating the video is to divide it into frames. The face of each frame is detected once the movie has been split up into individual frames, and the frame is then crop along the face. The clipped frame is then used to reconstruct the video's frames into a new one. Every video is subjected to the algorithm, which produces a processed dataset of face-only videos. The faceless frames are not taken into consideration during pre-processing.

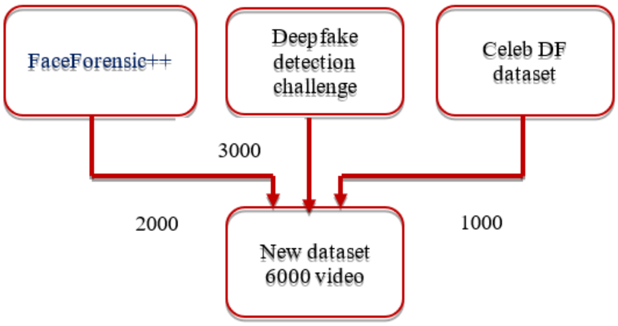

Fig.3: Processing of Dataset Gathering.

We selected a porch value based on the standard of all the frames in each video in order to maintain the same number of frames across all videos. Computer capacity restrictions are taken into consideration when using a threshold value. Although 300 frames will be produced in a 10-second film at 30 frames per second (fps), processing 300 frames at once in the experimental scenario is computationally demanding. As a result, we determined the threshold quantity to be 150 frames by considering the processing power of our Graphic Processor Unit (GPU) under test conditions. The new movie only includes the first 150 frames from the original film; the rest frames are in the new dataset. Instead of using a random approach, we showed how to best utilise Long Short-Term Memory (LSTM) by considering the first 150 frames in a sequential fashion. A frame rate of 30 frames per second and a resolution of 112 × 112 are saved for the newly created video.

3.1.3 Dataset Split

Train videos make up 70% of the videos in the dataset, while test films make up 30% of the total number of videos. To maintain balance, each split of the train and test has 50% genuine videos and 50% fake videos, films make up 30% of the total number of videos. To maintain balance, each split of the train and test has 50% genuine videos and 50% fake videos.

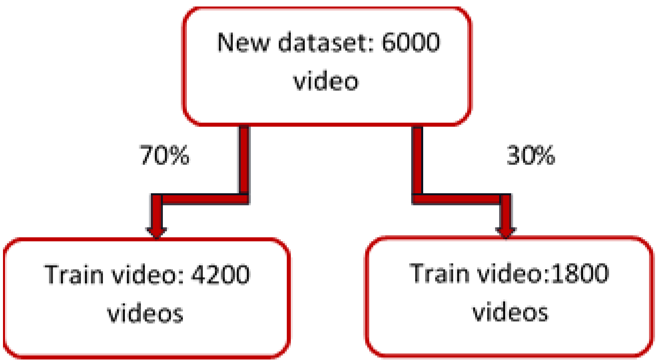

Fig.4: Train Trest Split.

3.1.4 Model Architecture

In our model, we use both CNN and RNN. Using the previously trained ResNext CNN model, an LSTM network is trained to identify the video as either pristine or deepfake based on the attributes that were recovered at the frame level. Using the Data Loader on the training split of the movies, the labels from the videos are loaded and fitted into the model for training.

3.1.5 ResNext

Rather than utilising the method from scratch, we extracted features using ResNext's pre-trained model. ResNext, an improved version of Residual CNN[8], performs better on deeper neural networks. The resnext50 32x4d model was utilised to carry out the experiment. A ResNext measuring 32 x 4 and having 50 layers has been used. The network will then be adjusted to correctly converge the gradient descent of the model by totalling any extra layers that are essential and select an acceptable learning rate. The 2048-dimensional attribute vectors that remain after ResNext's final pooling layers make up the chronological LSTM input.The LSTM uses fitted 2048-dimensional feature vectors as input to process sequences.

The model uses a linear layer with 2048 input features and 2 output features to learn the average rate of correlation between the input and output. An adaptive average polling layer with an output parameter of one is used by the model. Consequently, an image with an output size target of H x W is produced. The sequential processing of the frames is done via a sequential layer.The combined output feature vectors from the CNN and RNN are sent to a entirely connected film for cataloguing. The output of the fully coupled layer is a single probability value that specifies whether the input is a true or false video.

4. Result and Discussion of Machine Learning Techniques to Assess the Authenticity of Multimedia Data

Numerous media record information are valuable data; thus, it is important to examine this data for future exploration and usage, which is an issue that necessities top to bottom review. In any case, with the top to bottom utilization of learning techniques, the conventional strategies for acquiring viable data of sight and sound information can't be gotten precisely and rapidly, which carries challenges to information obtaining. Particularly, media information have a lot of data, and there are a lot of significant data concealed in struggle data, which makes it hard to mine. Under this foundation, how to understand the quick handling of enormous sight and sound information is a significant subject. The figure 5 shows the project's home page. where the video prediction's upload and file selection options are shown. The navigation bar contains the home button, which will direct you there.

Deep learning is a piece of AI, and it likewise acknowledges further learning in the field of AI. The profound learning strategy can consequently learn and get highlights as per input information, without tedious and work concentrated manual choice, which incredibly speeds up the culmination of undertakings. which makes it hard to mine. Under this foundation, how to understand the quick handling of enormous sight and sound information is a significant subject.

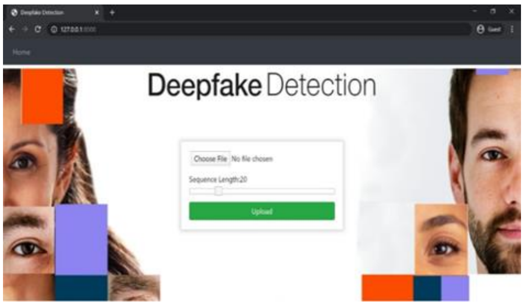

Fig.5: Home Page.

The figure 6 presents the output of a real video. As can be seen below, the uploaded video is split up into frames, and each frame's faces are clipped before being sent to the detection model. After that, training and test data are used to classify the outcome as real.

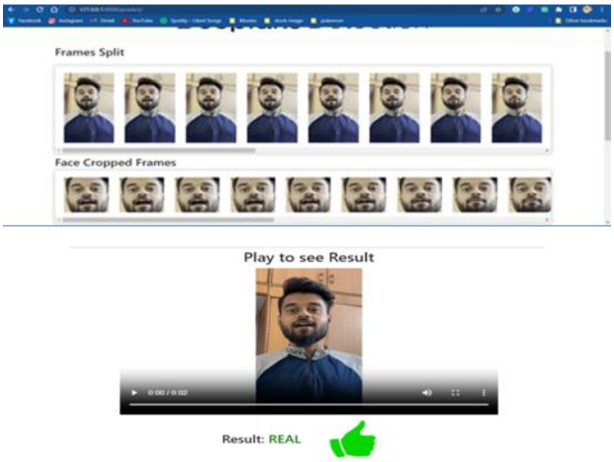

Fig. 6: Real Video Output.

Deep perusing learning has numerous application stacking layers, or at least, the contribution of the following layer is identical to the information yield portrayal of the upper layer. This organization has similar qualities as the customary brain organization, or at least, the progressive construction of the two organizations, that is to say, the design of the information yield organization; there is no association between neighboring hubs and there is no connection between the designs. The figure 7 shows what happens when a phoney video is created. As can be seen below, the uploaded video is split up into frames, and each frame's faces are clipped before being sent to the detection model. Then, utilising training and testing data, the output is flagged as fraudulent.

Through the examination on the perplexing data worth of profound learning mixed media information, the exploration on sight and sound data administration assumes a vital part later on improvement. The figure 8 presents the model comparison graph at the top shows the differences in accuracy between the models. The diagram shows that compared to the other two models (Xception model and MobileNet model), CNN and RNN have superior accuracy.

Contrasted and conventional characters, media information have its own attributes in information type. Huge measure of information is a significant component of sight and sound information, and it additionally makes mixed media information handling innovation very troublesome. Sight and sound information are generally made out of various sorts of unstructured single media information.

Fig. 7: Fake Video Output.

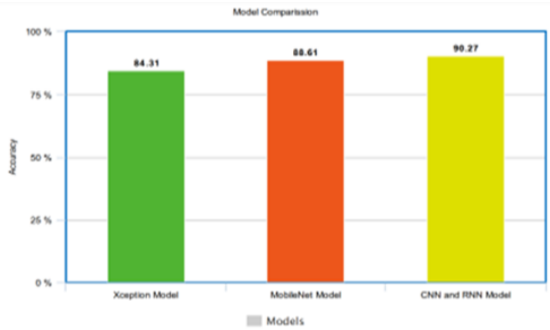

Fig. 8: Graph of Model Compression.

Contrasted and their remarkable single media, media has the attributes of consolidating sound, text, and picture, and the information measure of mixed media is a few times that of single media. Figure 9 represents the datasets are compared and their differences in accuracy are shown in the graph at the top. The graphic makes it evident that the faceForensic++, celeb-DF, and DFDC combined dataset is more accurate than the other three datasets.

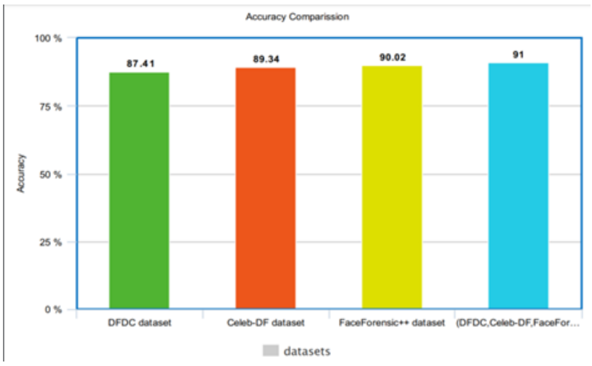

Fig. 9: Graph of Dataset Comparison.

5. Conclusion

People's trust in computerised media, augmented reality, robots, education, and many other organisations is starting to crumble due to deepfakes because recognising them isn't currently commensurate with placing confidence in them. They may intensify the annoyance and negative property felt by those beleaguered spread misinformation and hate speech, and even incite political unrest, violence, or war. This is particularly important these days because deepfakes are becoming easier to create and distribute swiftly thanks to online organising platforms. Deepfakes occasionally don't need to be distributed to enormous numbers in order to have a negative effect. Individuals who create deepfakes for malevolent purposes only require to distribute them to explicit audiences as ingredient of their harm system—they don't even require to use online entertainment. People need to learn to recognise when something is fake or real and should be more watchful of what they see. Police value and assessment cases have made considerable use of accounts and images as affirmations. Experts in electronic media criminal science with backgrounds in PC or rule execution and experience in social occasion, screening, and information separation may present them as affirmations in an authority court. Since even specialists are unable to see controlled drugs, it is possible that the advancements in machine learning and artificial intelligence have been employed to alter this mechanised substance. As a result, the conjectures of experts may not be sufficient to support these assertions. In any case, the presentation of mixed media information handling innovation is costly, so we want to further develop the handling productivity of interactive media information handling innovation. Also, an inside and out investigation of interactive media information handling innovation to further develop its weaknesses is required for the strong sight and sound to give better information handling innovation.

Author Contributions

Prof. Shruthi S: Methodology, implementation and project administration, **Dr. Manjunath R:** Problem definition and idea, writing—original draft preparation, **Dr. Shivashankar:** Data collection, corrections and report.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Yang Jing, Ma Haowei, Arshiya S. Ansari and G. Sucharitha, "Soft Computing Techniques for Detecting Cyberbullying in Social Multimedia Data", ACM Journals, Journal of Data and Information Quality, Vol. 15, No. 3, pp 1–14, 2023.

[2] Available online: https://theconversation.com/3-2-billion-images-and-720-000-hours-of-video-are-shared-online-daily-can-you-sort-real-from-fake-148630, 2023.

[3] Xu Z, Hong Z, Ding C, Zhu Z, Han J, Liu J et.Al, " MobileFaceSwap: A Lightweight Framework for Video Face Swapping". Proceedings of the AAAI Conference on Artificial Intelligence, 28;36(3):2973–81, 2022.

[4] Ehtiram Raza Khan, Saman Aisha, Deepak Kumar and Tabish Mufti, "A Systematic Review on Deepfake Technology", In book: Proceedings of Data Analytics and Management, DOI:10.1007/978-981-19-7615-5_55, pp.669-685, 2023.

[5] Jiacheng Lin, Yang Li and Guanci Yang, "Face de-identification method with generative adversarial networks for social robots", Neural Networks, doi.org/10.1016/j.neunet.2020.09.001Vol. 133, pp.132-147, 2021.

[6] Loc Trinh, Michael Tsang, Sirisha Rambhatla and Yan Liu, "Interpretable and Trustworthy Deep fake Detection via Dynamic Prototypes", 2021 IEEE Winter Conference on Applications of Computer Vision, DOI:10.1109/WACV48630.2021.00202, 2021.

[7] Yuming Gu, Mingming He, Koki Nagano, and Hao Li, "Protecting World Leaders Against Deep Fakes", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 38-45, 2019.

[8] Luca Guarnera, Oliver Giudice and Cristina Nastasi Sebastiano Battiato, "Preliminary Forensics Analysis of Deep Fake Images", 2020 AEIT International Annual Conference (AEIT), doi.org/10.23919/AEIT50178.2020.9241108, 2020.

[9] Ruben Tolosana, Ruben Vera-Rodriguez and Julian Fierrez, "Deep Fakes and Beyond:A Survey of Face Manipulation and Fake Detection", Computer Vision and Pattern Recognition, doi.org/10.48550/arXiv.2001.00179, Information Fusion, 2020.

[10] Ming-Yu Liu, Xun Huang, Jiahui Yu, Ting-Chun Wang and Arun Mallya, "Generative Adversarial Networks for Image and Video Synthesis: Algorithms and Applications", Proceedings of the IEEE, Vol.109, No.5, Pp.839 – 862, 2021.

[11] SiweiLyu.(2020) Deepfake Detection, "Current Challenges And Next Steps", IEEE International Conference on Multimedia & Expo Workshops, ISBN:978-1-7281-1486-6, 2020.

[12] Deng Pan, Lixian Sun, Rui Wang, Xingjian Zhang and Richard O, "Deep fake Detection through Deep Learning", IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, ISBN:978-1-6654-1567-5, 2020.

[13] Dr.Manjunath R, S.Balaji, "Review & analysis of multimedia data miming tasks and models", International Journal of Innovative Research in Computer and Communication Engineering, 2019.

[14] Egor Zakharov, Aliaksandra Shysheya, Egor Burkov and Victor Lempitsky, "Few- Shot Adversarial Learning of Realistic Neural Talking Head Models", IEEE/CVF International Conference on Computer Vision, ISBN:978-1-7281-4804-5, 2019.

[15] Dr.Manjunath R, S.Balaji, "An Adapative Fuzzy C-means clustering method for retrieval of videos in multi media", International Journal of Applied Engineering Research, 11 (9), 6680-6688, 2016.

[16] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," IEEE Transactions on Cognitive Communications and Networking, vol. 11, no. 4, pp. 2715–2614, 2021.

[17] Manjunath R, "Wireless Sensor Network Routing Protocols, Challenging Issues and Performance Comparison", Journal of Theoretical and Applied Information Technology, ISSN: 1992-8645, . Vol.101, No 19 , 2023.

[18] H. Frank, "Multimedia data maintaining and processing device," Alexandria Engineering Journal, vol. 21, no. 2, pp. 1761–1524, 2021.

[19] Dr. Manjunath R, "A Smart Biomedical Healthcare System to Detect Stress using Internet of Medical Things, Machine Learning and Artificial Intelligence", International Journal of Intelligent Systems and Applications in Engineering, ISSN:2147-67992147, 2023.