IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DEEPFAKE DETECTION ON FACE IMAGES & VIDEOS USING DEEP LEARNING

[1]B. Manish, [2]C. Manish, [3]B. Sai Kiran Reddy

[1]Student, [2]Student, [3]Student

[1]Artificial Intelligence & Machine Learning,

[1]Anurag University, Hyderabad, India

ABSTRACT: In recent months, unfastened deep studying-based totally software program tools have facilitated the creation of credible face exchanges in photographs that depart few traces of manipulation, in what they may be referred to as deep faux face detection. Manipulations of digital pics were validated for numerous decades through the coolest use of visual consequences, latest advances in deep studying have caused a drastic increase in the realism of fake content and the accessibility in which it could be created. Those so-called AI-synthesized media (popularly known as fake). Creating fake photographs using Artificially clever gear are easy undertaking. But, with regards to the detection of those fake pictures, it's far a first-rate task. Because education the set of rules to identify the faux pics isn't always simple. We've got taken a step forward in detecting the fake images using a Convolutional Neural network. The system uses a convolutional Neural network (CNN) to extract functions at the body level. These capabilities are used to teach a CNN model which learns to classify if a given photograph has been challenged to manipulation or no longer. Anticipated result against a massive set of faux pix collected from the same old records set. We display how our system can be competitive and bring about this undertaking effect by using an easy architecture. Deep faux and real picture classification. The use of Deep studying is an emerging discipline that specializes in distinguishing between laptop-generated (deep fake) and authentic pictures. The rapid advancement in the era has made it increasingly more difficult to detect these deep fakes, as they come to be greater practical. This undertaking employs deep learning techniques to address this challenge, aiming to increase an effective and efficient model for accurate classification. This topic is selected due to the increasing occurrence and dangers related to deep faux technology. As deep fakes grow to be extra commonplace and convincing, the potential for misuse in various sectors turns into a widespread situation. Growing powerful detection methods is crucial to mitigate these risks and ensure the integrity and trustworthiness of virtual media.

Index Terms **-** Deep Learning, Image Classification, Deep Fake Detection, Convolutional Neural, Networks (CNN), Image Processing, Xception, MobileNet, Face Forensics++.

I. INTRODUCTION:

Deepfakes are a form of synthetic media created using deep learning techniques, particularly generative adversarial networks (GANs) and autoencoders. They involve superimposing or replacing someone's face in a video or image with another person's likeness, often resulting in highly realistic and convincing visual content. This technology has raised significant concerns due to its potential for misuse, including the spread of disinformation, defamation, and privacy violations. **[8].**

The term “deep fake” is a merged word of “deep learning” and “fake,” reflecting the underlying technology’s reliance on deep neural networks to manipulate and generate content. These neural networks are trained on large or vast amounts of data, typically consisting of all images and videos of the target individual whose likeness is to be replicated. By analyzing and learning patterns from this data, the model can generate realistic facial movements and expressions that closely mimic those of the target person. **[10].**

While deepfakes have garnered attention for their use in creating entertaining or comedic content, such as replacing or inserting celebrities into movies or music videos, they also pose significant ethical and societal challenges **[].** One major concern is the potential for deepfakes to be used maliciously to spread false information or manipulate public opinion. For example, they could be employed to create convincing videos of politicians saying or doing things they never actually did, leading to misinformation and distrust. **[12].**

Efforts to address the risks associated with deepfakes include the development of detection methods to identify manipulated content and the implementation of regulations to curb their harmful effects. However, as deepfake technology continues to advance, staying ahead of its potential misuse remains an ongoing challenge for researchers, policymakers, and technology companies alike. Ultimately, the responsible and ethical development and deployment of deepfake technology will require a concerted effort from multiple stakeholders to balance innovation with safeguarding against its negative consequences. **[11].**

Fig 1: Difference between fake and real image

II. LITERATURE SURVEY:

The literature survey encompasses a diverse array of studies conducted by researchers across various domains, focusing on topics ranging from Deepfake Detection. In recent years, there has been a notable surge in research efforts aimed at enhancing public safety, leveraging advancements in deep learning, computer vision, and detection technologies.

Researchers like **Siwei Lyu [14]** have carried out an extensive analysis addressing obstacles and future directions for study in the field of deepfakes. Notably, information loss during encoding poses a major barrier to the production of high-quality details such as facial hair and skin texture, which is a major drawback of existing DeepFake generation algorithms. Head puppetry, face swapping, and lip-syncing are three popular techniques that Lyu covers. Each has a distinct function, such as mimicking behavior or changing speech. The most common detection techniques are frame-level binary classification, which can be further subdivided into signal-level artifacts, physical/physiological aspect inconsistencies, and data-driven methods using Deep Neural Networks (DNNs). Lyu does draw attention to several drawbacks, such as manipulated social media and poor dataset quality.

Peng Chen, et. al. **[15]** have introduced FSSPOTTER, a unified framework designed to simultaneously analyze spatial and temporal information within videos. Videos are divided into consecutive segments by the Spatial Feature Extractor (SFE), which then processes each clip to produce frame-level characteristics. SFE extracts intra-frame spatial characteristics by using convolution layers from the VGG16 network with batch normalization. To improve feature extraction, the framework also uses a superpixel-wise binary classification unit (SPBCU). A Bidirectional LSTM is employed by the Temporal Feature Aggregator (TFG) to detect temporal discrepancies across frames. The probabilities indicating whether the clip is real or fake are then computed by a fully connected layer, which is followed by a softmax layer. The FaceForensics++ dataset was used to assess the methods and show how well the system detected deepfakes.

Huy H. Nguyễn [19] Using capsule networks to detect forged images and videos uses a method that uses a capsule network to detect forged, manipulated images and videos in different scenarios, like replay attacks detection and computer-generated video detection.

Moreover, scholars such as **Shivangi Aneja** et al. **[17]** presented Profound Dissemination Exchange (DDT), an exchange learning approach tending to zero and few-shot exchange challenges in imitation location. Their strategy utilizes a distribution-based misfortune definition, outflanking baselines altogether in both scenarios. By utilizing an ImageNet-pre-trained ResNet-18 neural organize, DDT accomplishes 4.88% higher location proficiency for zero-shot and 8.38% for few-shot exchanges, broadening the scope of fraud location.

III. EXISTING SYSTEMS:

The existing system focuses on the analysis and utilization of previously deployed deep fake detection algorithms. It extensively explores classic detection methods and contemporary deep learning-based approaches, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM). Classic methods often rely on rule-based heuristics, while deep learning methods harness the hierarchical learning capabilities of neural networks to discern intricate patterns indicative of deep fakes. The study aims to provide a comprehensive understanding of the strengths and limitations of these methods in identifying manipulated content. In contrast, the proposed system takes a proactive stance by integrating cutting-edge deep learning algorithms. The proposed system aims to enhance detection accuracy, contributing to a deeper comprehension of deep fake production and distribution for more effective countermeasures.

3.1 DISADVANTAGES OF EXISTING SYSTEMS

1. The existing system relies on traditional and contemporary methods, potentially limiting adaptability to emerging deep fake techniques.
2. Classic methods' reliance on rule-based heuristics may struggle with the evolving sophistication of deep fake creation, leading to decreased accuracy.
3. While deep learning methods are employed, the existing system may lack a comprehensive understanding of the intricate patterns characterizing evolving deep fake technologies.
4. Depending on data complexities, the existing system may exhibit variable accuracy, potentially rendering it less reliable across diverse scenarios.

3.2 Proposed System

The proposed system aims to combat the escalating threat of deep fakes through an integrated approach leveraging cutting-edge deep learning algorithms. Utilizing a Kaggle deep fake dataset, our system employs InceptionResnetV2, VGG19, CNN, and Xception algorithms for comprehensive evaluation. The emphasis is on developing a robust detection mechanism to discern authentic content from manipulated ones. By scrutinizing the intricacies of deep fakes, the system aspires to contribute to a deeper understanding of their production and distribution. The integration of multiple algorithms ensures a nuanced analysis, while the dataset provides a diverse range of scenarios for effective training and testing. This holistic approach seeks to

enhance the overall accuracy and reliability of deep fake detection, mitigating the potential societal consequences of fake news, impersonations, and privacy violations. The proposed system stands as a pivotal step toward fortifying our digital landscape against the malevolent misuse of deep fake technology.

3.3 Advantages Of Proposed System

1. The proposed system integrates advanced algorithms like InceptionResnetV2 and Xception, enhancing its capability to detect sophisticated deep fake manipulations effectively.
2. Utilizing multiple algorithms facilitates a nuanced analysis, ensuring a robust detection mechanism capable of discerning authentic content from manipulated instances.
3. The system benefits from a Kaggle deep fake dataset, providing a diverse range of scenarios for training and testing, enhancing its adaptability to real-world situations.
4. The proposed system takes a holistic approach, aiming not only to enhance accuracy but also to deepen comprehension of deep fake production, distribution, and societal consequences, offering a more proactive defense strategy.

3.4 Proposed System Design

In this project work, there are Five modules and each module has specific functions, they are:

1. Data Collection
2. Pre-Processing
3. Feature Extraction
4. Model Selection
5. Train-Test Split and Model FITTING
6. Model Evaluation and Deployment

3.4.1 Data Collection:

In this project, we use fake faces and real faces datasets collected from Kaggle. Pixel values from images are taken as input and labels are used as output each folder has 180 images, and videos which are used for training.

3.4.2 Pre-Processing:

Pre-processing is a process or technique used to improve image quality and boost visualization. In detecting fake face images, image processing is a crucial phase that helps to improve the images quality. This can be one of the most critical factors in achieving good results and accuracy in the next phases of the proposed methodology. Clean and preprocess the data to ensure consistency and remove noise. Poor or low-quality photos could produce outcomes that are not up to par. During the preprocessing phase, we performed background elimination, elimination of non-essential, image enhancement, and noise removal.

3.4.3 Feature Extraction:

Image Features: Extract relevant features from images such as facial landmarks, textures, and distortions.
Video Features: Extract temporal features from video sequences to capture motion patterns.

3.4.4 Model Selection:

Machine Learning Models: Choose appropriate ML models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), or their variants.
Deep Learning Models**:** Consider architectures like ResNet, VGG, or custom-designed networks for feature extraction and classification.
Ensemble Methods: Combine multiple models to improve overall performance and robustness of the model.

3.4.5 Train-Test Split and Model Fitting:

Now, we split our dataset into training and testing data. The purpose of this distribution is to evaluate the performance of our model with unknown data and to determine to what extent our model generalizes to training data. This is followed by model fitting**,** which is an important step in the model building process. Train the Xception model on the training dataset, allowing the custom layers to learn task-specific features for deepfake detection. Monitor the model's performance on the validation set and adjust hyperparameters as needed to prevent overfitting.

3.4.6 Model Evaluation and Deployment:

Evaluate the trained Xception model on the testing dataset to assess its performance. Calculate metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to measure the model's effectiveness in detecting deepfake images.

Metrics: Define evaluation metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC).
Cross-validation: Employ techniques like k-fold cross-validation to assess model performance robustly.
Confusion Matrix Analysis: Analyze the confusion matrix to understand model strengths and weaknesses.

3.4.7 Deployment:

Thresholding: Set appropriate decision thresholds to classify instances as real or fake.
Integration: Integrate the module into existing systems or frameworks for broader usage.

3.4.8 Architecture:

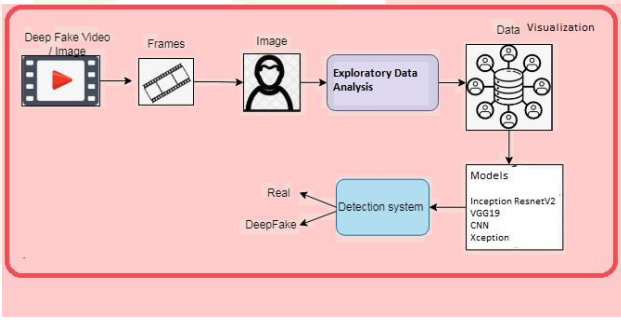

Fig 2: System Architecture

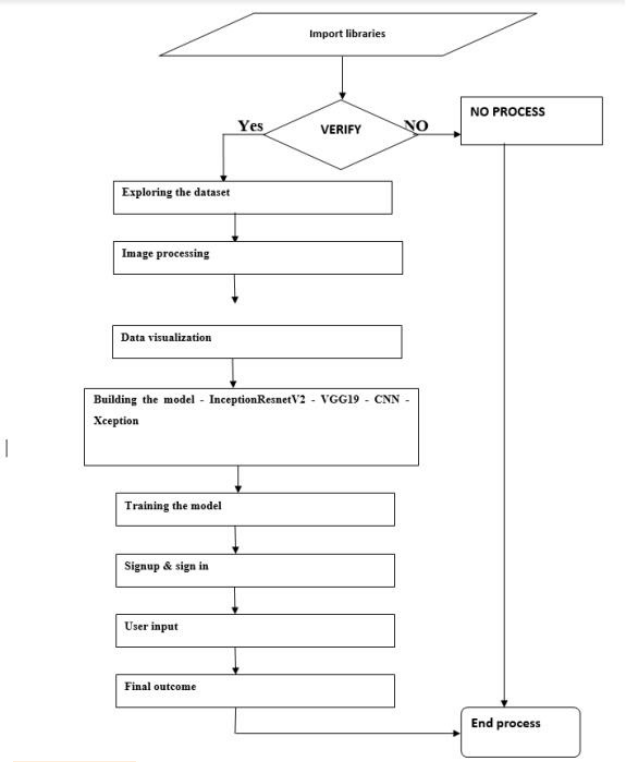

Fig 3: Model Architecture

IV. IMPLEMENTATION:

4.1 System Requirements:

1) Software: Anaconda
2) Primary Language: Python
3) Frontend Framework: Flask
4) Back-end Framework**:** Jupyter Notebook
5) Database: Sqlite3
6)Front-End Technologies: HTML, CSS, JavaScript and Bootstrap4.

4.2 Hardware Requirements:

Processor: Multi-core processor (e.g., Intel Core i5 or higher) for efficient data processing and model training.
Memory (RAM): Minimum 8GB RAM for handling large datasets and running complex machine learning algorithms.
Storage: SSD (Solid State Drive) storage for fast data access and model loading.
Network Interface: Ethernet adapter for connecting to the network and capturing network traffic data.

4.3. Methodology:

Data Collection and Preprocessing:

Gather a dataset of labeled images or videos containing both real and deepfake content. Preprocess the data by resizing images to a standard resolution (e.g., 299x299 for Xception), normalizing pixel values, and augmenting the dataset if necessary to increase diversity.

Data Splitting:

Split the dataset into training, validation, and testing sets using a suitable ratio.

Model Initialization:

Initialize the Xception model pre-trained on ImageNet weights. This is typically available in deep learning libraries like TensorFlow or Keras.

Fine-tuning:

Fine-tune the pre-trained Xception model on the deepfake detection task using the training dataset.

Freeze the early layers of the Xception model (up to a certain depth) to retain the pre-trained weights and prevent them from being updated during training.

Add custom fully connected layers on top of the frozen layers to adapt the model to the specific deepfake detection task. Compile the model with an appropriate loss function (e.g., binary cross-entropy) and optimizer.

Training:

Train the Xception model on the training dataset, allowing the custom layers to learn task-specific features for deepfake detection. Monitor the model's performance on the validation set and adjust hyperparameters as needed to prevent overfitting.

4.4. Testing and Validation:

Test Case 1: Deepfake Image Detection - Genuine Image.

Input Validation:

Ensure that the attributes of the analyzed image are within the expected range and format, similar to the attributes in the training dataset.

Output Validation:

After processing the image, verify that the system confirms it as a "Genuine Image" and does not detect any synthetic or manipulated elements.

Functional Validation:

Confirm that the deepfake detection system correctly identifies real images and does not raise false alerts or flags when no deepfake elements are detected.

Test Case 2: Deepfake Image Detection - Fake Image

Input Validation:

Ensure that the attributes of the analyzed image are within the expected range and format, similar to the attributes in the training dataset.

Output Validation:

After analyzing the image, verify that the system correctly identifies it as a "Deepfake Image" and provides details about the synthetic elements detected.

Functional Validation:

Confirm that the deepfake detection system accurately identifies synthetic or manipulated images and raises appropriate alerts or flags when deepfake elements are detected.

4.5. Output Comparison:

	Accuracy	Recall	Precision	F1 Score	Sensitivity	\
InceptionResnet V2	0.918432	0.918415	0.918415	0.918415	0.918415	
VGG19	0.796416	0.796435	0.796435	0.796435	0.796435	
CNN	0.987048	0.987036	0.987036	0.987036	0.987036	
Xception	0.993900	0.993902	0.993902	0.993902	0.993902	

	Specificity	MAE	\
InceptionResnet V2	0.918415	<function mae at 0x00000204027E2288>	
VGG19	0.796435	0.32427	
CNN	0.987036	0.019514	
Xception	0.993902	0.009542	

	MSE
InceptionResnet V2	<function mse at 0x00000204027E23A8>
VGG19	0.162385
CNN	0.009877
Xception	0.004718

Fig 4: Accuracy of models

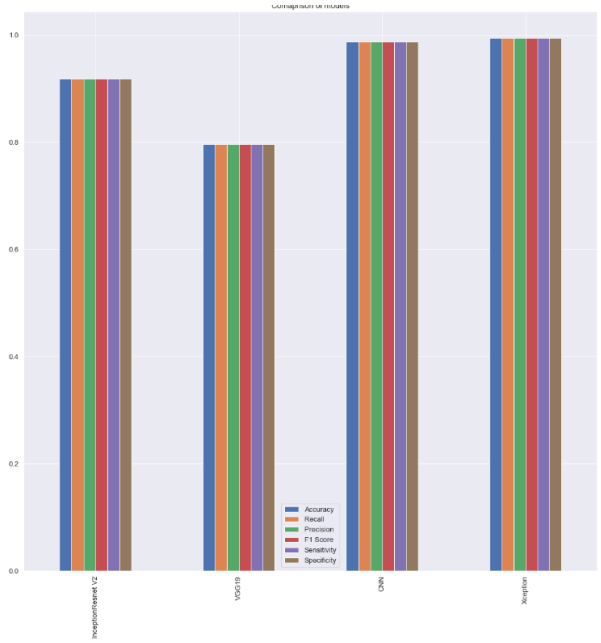

Fig 5: Comparison of Models

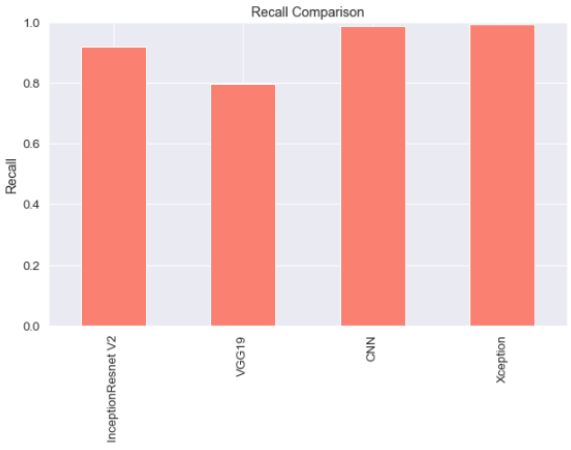

Fig 6: Recall Comparison

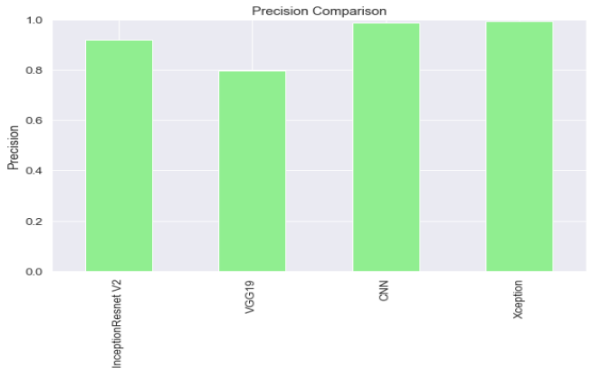

Fig 7: Precision Comparison

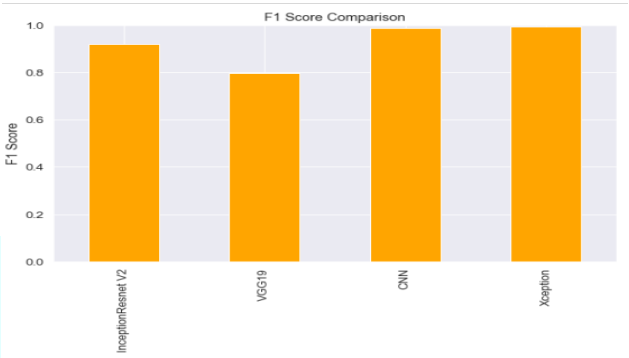

Fig 8: F1 Score Comparison

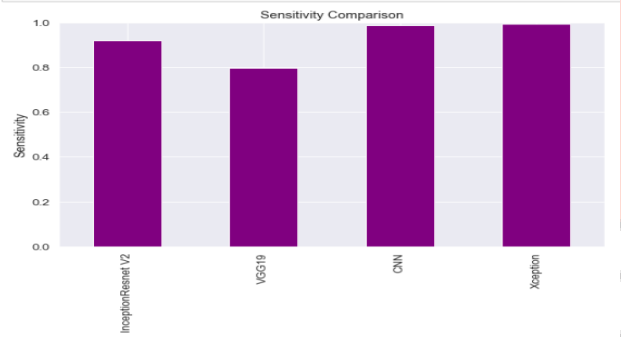

Fig 9: Sensitivity Comparison

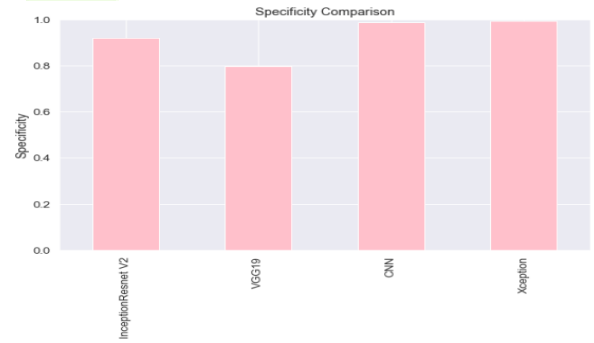

Fig 10: Specificity Comparison

4.6 Result Analysis:

Accuracy Comparison of Machine Learning Models:

Xception and CNN achieved the highest accuracy i.e, 99.3% and 98.7%, While VGG-19 and InceptionResnet V2 achieved 79.6% and 91.8%.

Model Deployment Simulation:

The trained Xception and CNN models was deployed for deepfake analysis in a simulation environment. The simulation successfully differentiated between fake and real image/ video.

Overall Model Performance:

Xception and CNN exhibited exceptional accuracy and making them strong candidates for real-world deployment. While VGG19 acquired less accurate, still could befurther optimized or combined with other models for improved performance. Our model may not give proper results for the real-time videos and images because the new images and videos must be trained in order to be detected by the model.

v. RESULTS SCREENSHOTS:

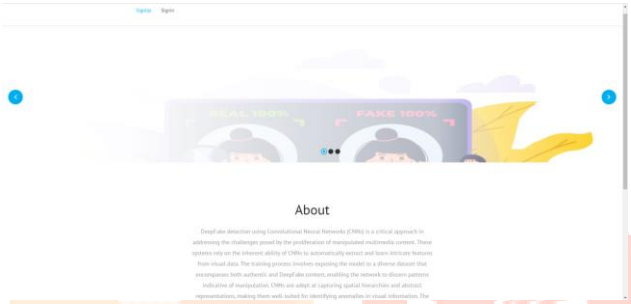

Fig 11: About Page

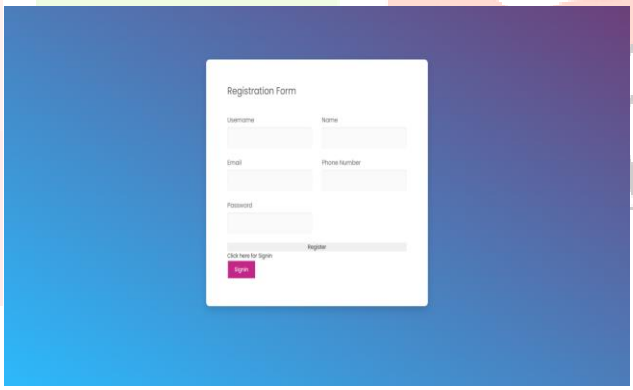

Fig 12: Registration Page

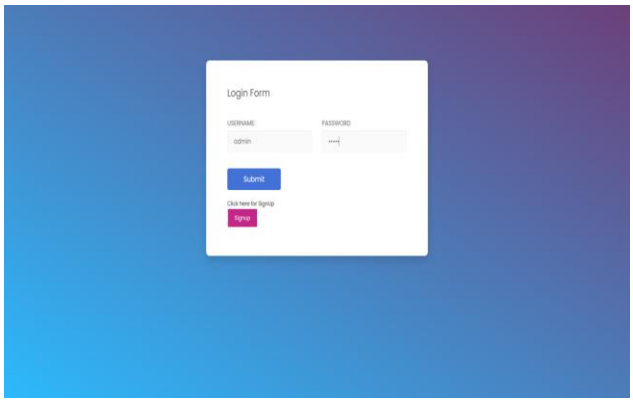

Fig 13: Login Page

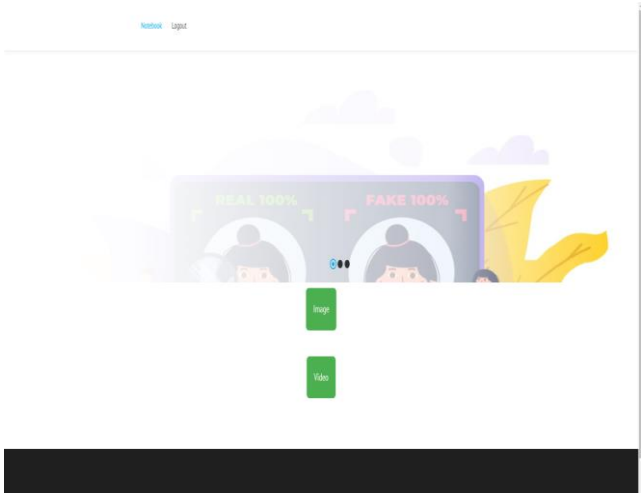

Fig 14: Home Page

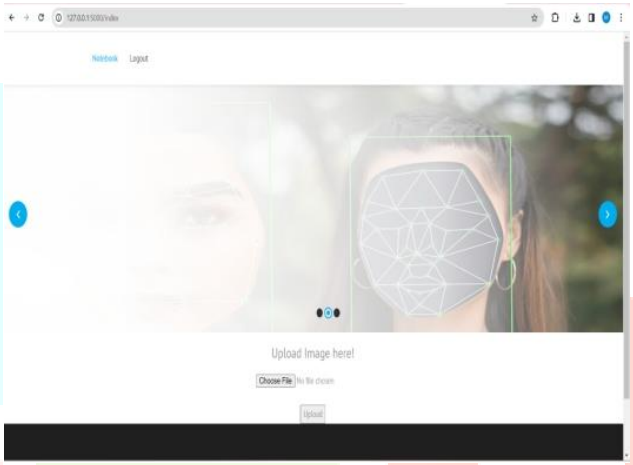

Fig 15: Uploading Page

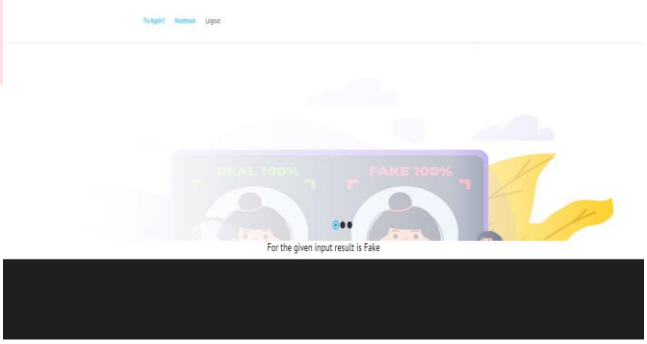

Fig: 16: Result Page

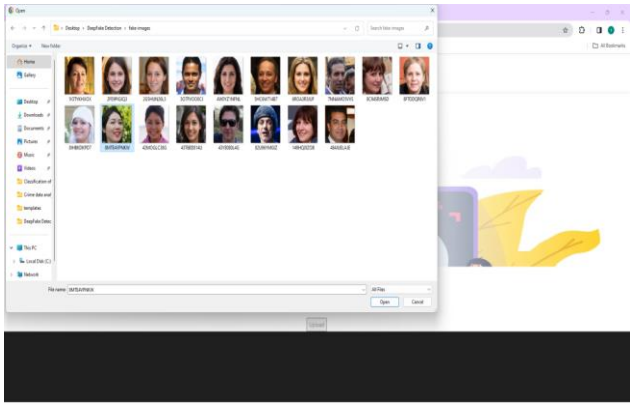

Fig 10: Images

VI. CONCLUSION:

The emergence of deep fake technology has presented both opportunities and challenges in the digital age. While deep fakes offer innovative avenues for creativity and entertainment, they also pose significant threats to truth, authenticity, and trust in digital media. As the capabilities of deep fake technology continue to evolve, so too must our efforts to detect and mitigate the spread of manipulated content.

However, deep fake detection systems are not without their limitations. Challenges such as adversarial adaptation, limited training data, and ethical considerations underscore the complexity of combating digital deception. Addressing these challenges requires ongoing research, collaboration, and interdisciplinary efforts to enhance the effectiveness, robustness, and ethical integrity of detection systems.

Moreover, the deployment of deep fake detection systems must be accompanied by thoughtful consideration of legal and regulatory frameworks, ensuring that detection technologies are used responsibly and by ethical principles. Public awareness and education initiatives are also essential to empower individuals to recognize and critically evaluate manipulated media content. In navigating the evolving landscape of deep fake technology, stakeholders must work together to foster innovation, uphold integrity, and protect the fundamental values of truth and trust communication.

REFERENCES:

[1] Manoj Kumar Das, Manav Kumar, Ishank Kumar kapil, Dr Rajesh Kumar yadav, “Deepfake Creation Using Gans and Autoencoder and Deepfake detection”, in 2023 nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN). IEEE, (2023).

[2] Harsh Chotaliya, Mohammed Adil Khatri, Shubham Kanojiya, Mandar Bivalkar, “Review: DeepFake Detection Techniques using Deep Neural Networks (DNN)”, in 2023 6th International Conference on Advances in Science and Technology (ICAST). IEEE, (2023).

[3] Kartik Bansal, Shubhi Agarwal, Narayan Vyas, “Deepfake Detection Using CNN And DCGANS To Drop-Out Fake Multimedia Content: A Hybrid Approach”, in 2023 International Conference on IoT, Communication and Automation Technology (ICICAT). IEEE, (2023).

[4] Ruby Chauhan, RenuPopli, Isha Kansal, “A Systematic Review on Fake Image Creation Techniques”, in 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom). IEEE, (2023).

[5] Dr. R. R. Rajalaxmi, Sudharsana P P, Rithani A M, Preethika S, Dhivakar P and Gothai E, “Deepfake Detection using Inception-ResNet-V2 Network”, in Proceedings of the 7th International Conference on Computing Methodologies and Communication (ICCMC-2023). IEEE, (2023).

[6] Sonia Salman and Jawwad Ahmed Shamsi, "Comparison of Deepfakes Detection Techniques", in 2023 3rd International Conference on Artificial Intelligence (ICAI). IEEE, (2023).

[7] Norah M. Alnaim, Zaynab M. Almutairi, Manal S. Alsuwat, Hana H. Alalawi, Aljowrah Alshobaili and Fayadh S. Alenezi, "DFFMD: A Deepfake Face Mask Dataset for Infectious Disease Era With Deepfake Detection Algorithms". IEEE, (2023).

[8] Rui Zhang; Zixuan Jiang; Changxu Sun, " Two-Branch Deepfake Detection Network Based on Improved Xception", Published in: 2023 IEEE International Conference on Electrical, Automation and Computer Engineering (ICEACE). IEEE, (2023).

[9] Jerry John and Ms. Bismin V. Sherif, "Comparative Analysis on Different DeepFake Detection Methods and Semi Supervised GAN Architecture for DeepFake Detection" in Proceedings of the Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC-2022). IEEE, (2022).

[10] Kalicharan Jalui, Aditya Jagtap, Saloni Sharma, Gilofer Mary, Reba Fernandes and Megha Kolhekar, "Synthetic Content Detection in Deepfake Video using Deep Learning" in 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT). IEEE, (2022).

[11] Niteesh Kumar, Pranav P, Vishal Nirney, Geetha V, "Deepfake Image Detection using CNNs and Transfer Learning", in 2021 International Conference on Computing, Communication and Green Engineering (CCGE) JSPM's RSCOE, Pune, India. Sep 23-25, 2021. IEEE, (2021).

[12] Tianchen Zhao, Xiang Xu, Mingze Xu, Hui Ding, Yuanjun Xiong, Wei Xia, "Learning Self-Consistency for Deepfake Detection", 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, (2021).

[13] Apurva Gandhi, Shomik Jain, "Adversarial Perturbations Fool Deepfake Detectors", in IEEE (2020).

[14] Siwei Lyu, "DEEPFAKE DETECTION: CURRENT CHALLENGES AND NEXT STEPS", IEEE International Conference on Multimedia & Expo Workshops (ICMEW) (2020).

[15] Peng Chen, Jin Liu, Tao Liang, Guangzhi Zhou, Hongchao Gao, Jiao Dai1, Jizhong Han, "FSSPOTTER: SPOTTING FACESWAPPED VIDEO BY SPATIAL AND TEMPORAL CLUES", IEEE International Conference on Multimedia and Expo (ICME) (2020).

[16] Mohammed A. Younus, Taha M. Hasan, "Abbreviated View of Deepfake Videos Detection Techniques", 6th International Engineering Conference "Sustainable Technology and Development" (IEC) (2020).

[17] Shivangi Aneja, Matthias Nießner, "Generalized Zero and Few-Shot Transfer for Facial Forgery Detection", arXiv:2006.11863v1 [cs.CV] (2020).

[18] Umur Aybars Ciftci,˙Ilke Demir, and Lijun Yin, Senior Member, IEEE, "FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. X, No. X, July (2020).

[19] Huy H. Nguyễn, Junichi Yamagishi, Isao Echizen, "CAPSULE-FORENSICS: USING CAPSULE NETWORKS TO DETECT FORGED IMAGES AND VIDEOS". IEEE, (2019).

[20] Z. Chen and H. Yang, "Attentive semantic exploring for manipulated face detection," in ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021, pp. 1985–1989.

[21] D. Guera and E. J. Delp, "Deepfake video detection using recurrent ¨ neural networks," in 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS). IEEE, 2018, pp. 1–6.

[22] Y. Li, M.-C. Chang, and S. Lyu, "In ictu oculi: Exposing ai created fake videos by detecting eye blinking," in 2018 IEEE International workshop on information forensics and security (WIFS). IEEE, 2018, pp. 1–7.

[23] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 2625–2634.

[24] Y. Li and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts," arXiv preprint arXiv:1811.00656, 2018.

[25] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "Mesonet: a compact facial video forgery detection network," in 2018 IEEE international workshop on information forensics and security (WIFS). IEEE, 2018, pp. 1–7.