Deepfake Detection Using Computer Vision

[1]Anukarsh Awasthi, [2]Harsh Sachan, [3]Ms. MAHAK
[1,2] B-tech in CSE, Galgotias University Gautam Buddha Nagar, India
[3] Assistant Professor, Galgotias University Gautam Buddha Nagar, India

***Abstract*—The emergence of social media has greatly accelerated the dissemination of manipulated images, misleading memes, and it is rather challenging to trace the real and deceptive material in visuals. The present study is aimed at identifying deepfake images and fake memes with the help of computer vision methods. The suggested system is based on a lightweight Convolutional Neural Network (CNN) which examines various features, such as RGB color patterns, depth data, and facial features, such as eyes, nose, and lips. This is in contrast to the traditional methods that usually use RGB features alone to detect an object because this method incorporates visual appearance and spatial depth information to enhance the level of detection. The model is trained with a big amount of real and fake images and uses preprocessing and data augmentation methods to improve the performance. The depth map estimation is capable of detecting unnatural structural changes whereas the RGB analysis is able to detect the inconsistencies in texture and color anomalies. The method suggested is effective in detecting manipulated images and memes even in complicated situations in which there is compression and editing. Moreover, the model is lightweight hence low cost in computation implying that it can be applied in real time or systems with limited resources. The proposed study will help to fight against misinformation on social media by offering a stable and scalable solution to the problem of deepfakes and fake memes.**

***Index Terms*—Detection of Fake Memes, Compact CNN, Image Categorization, Misinformation on Social Media, Deep Learning Techniques Introduction.**

I. INTRODUCTION

Digital content has brought about the rapid growth of social media. came into existence as a stronger and more efficient tool of. disseminating information. Aesthetic material, especially, is very much involved in this. Among various forms of images that Abstract combine is called memes.

particularly to deliver messages promptly, there is the use of visuals and text. popular. Despite the fact that memes are usually produced to some people use them to disseminate entertainment. socio-media misinformation or rumors. Nowadays, they are being more and more resorted to spread false information. content. This is due to the proliferation of AI tools and applications. enabled the generation of such fake images and memes, and so made it easy. it difficult to say whether an image or a meme by passes going around in the social media is real or fake.

Fake images and memes can mislead and influence false narratives and public perception. They may also spread deceptive visual signals or modified facial characteristics. Traditional detection techniques mainly concentrate on text analysis and often falls.

II. PROBLEM STATEMENT

The extensive dissemination of fake images and pictures on social media has emerged as a major issue for communities, as they spread false narratives and significantly undermine the credibility of information and public confidence. Identifying these misleading visuals is challenging due to the advanced visual realism produced by contemporary manipulation techniques. Existing Detection techniques focus only on RGB features, but they do not focus on depth which is also a major feature for detection of fake images. Therefore, to prevent Fake Detection we need a strong invention which combine both RGB+ depth + facial expressions data to prevent fraud images, memes in social media. We need a strong and trustable software which should be automated and light weight vision-based system.

III. LITERATURE REVIEW

The fast advancement of the deep learning and image manipulation algorithms has considerably accelerated the creation of deepfake pictures and misleading visual information. This has made the identification of such a manipulated media to be a significant research issue in the area of computer vision. A number of researchers have studied the deep learning applications in the detection of deepfake images. In [1], a new deep learning model to detect manipulated photographs through visual irregularities is introduced, and it is shown that convolutional neural networks are effective at extracting features. In the same way, the research in [3] provides an overall overview of deepfake detection methods, and it is possible to note that CNN-based models are utilized to detect facial manipulations and fake content.

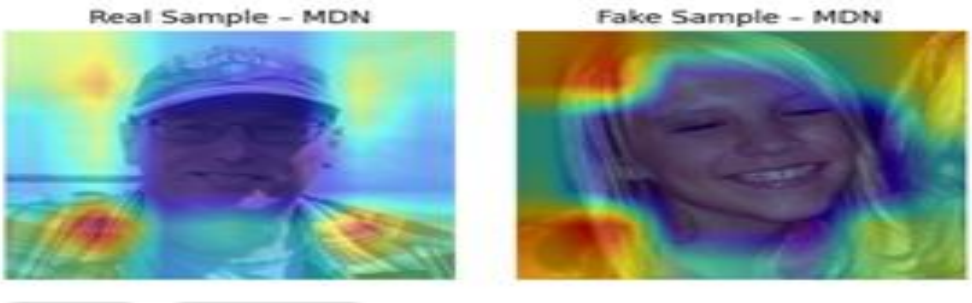
Besides the analysis of the image at rest, there are also works devoted to the inconsistencies in time and structure. The method presented in [2] compares frame-level alterations on a rate-of-change basis to identify deepfake videos, and demonstrates that temporal change can be a promising tool to identify manipulation. Additionally, GAN-based detection methods are discussed in [8] where the detector detects the artifacts that are generated in the process of image generation, enhancing the detection

Evolution Results

Class	TestingSet(RAW)	Testing Set(C40
DF	97.95%	92.10%
F2F	96.80%	82.50%
FS	97.10%	88.25%
NT	94.50%	71.85%
ALL	97.95%	84.30%

Recent studies have been concerned to deal with these issues. on training the light CNN architectures targeted at. reducing the complexity of models and yet obtaining good results. satisfactory accuracy. These are lightweight models available. characterized by fewer parameters and layers that are streamlined, making the training to be faster. inference. They are especially suitable to be used in dealing with. Big data and applications that are real-time. processing.

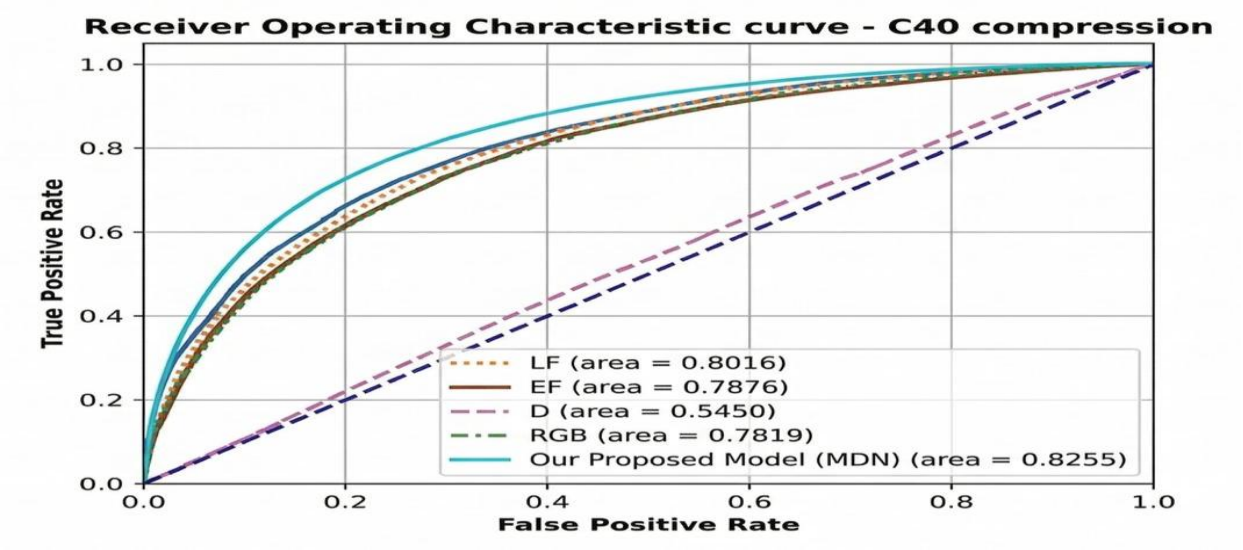
However, in cases where one relies alone, generalization could be a challenge particularly when it comes to advanced forms of manipulation. Most of the existing deepfake detector techniques mostly rely on RGB image data. This RGB analysis aims at visual cues, like texture abnormalities, abnormal color patterns, lighting inconsistencies and facial abnormalities. Although RGB characteristics perform well to identify different fake images, it is susceptible to image compression, image downsizing, and stylized editing common in memes. Scholars have tested the possibility of using depth and spatial features to overcome these drawbacks when it comes to fake image detection. The depth data encodes the geometric and spatial layout of a face that can be used to detect the presence of unrealistic facial shapes and distortion not easily seen in RGB images. It has been shown that the combination of depth data and RGB features shows greater resistance to high-level manipulations in that they provide a complementary visual representation.

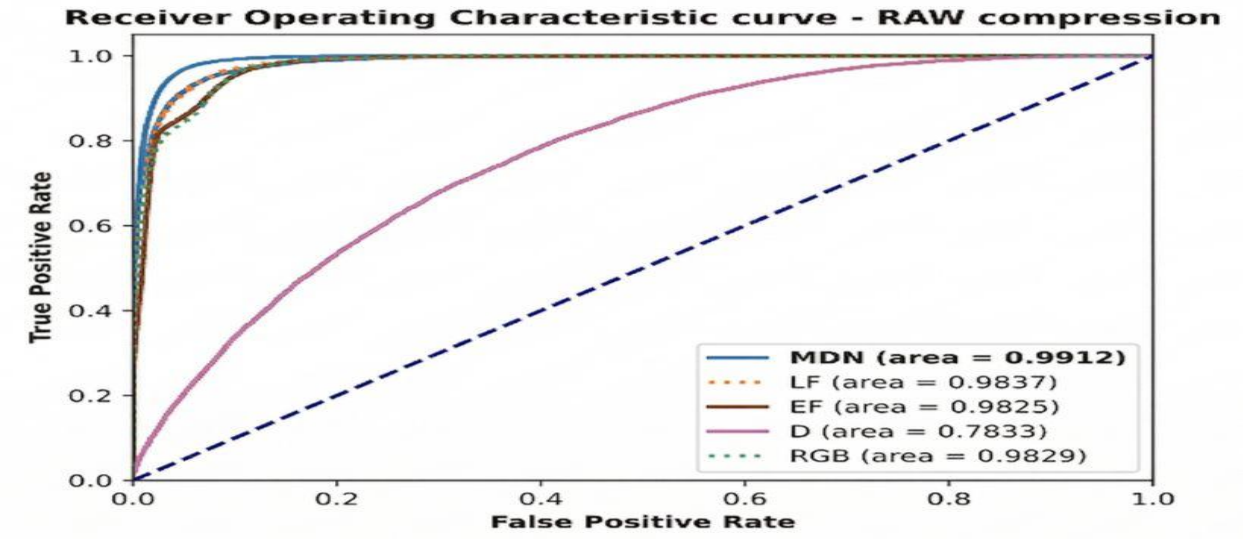


Although there has been considerable advancement in detecting deepfake images, there is limited research specifically targeting fake content in memes. Memes present additional challenges due to text overlays, extensive

image editing, compression artifacts, and creative distortions. These elements diminish the effectiveness of models trained on conventional face datasets. Moreover, most current methods are computationally demanding and do not account for lightweight architectures that are practical for the review of the literature shows that there is a clear gap in the literature on how to develop an effective and lightweight deep learning architecture that uses both RGB and depth information

to identify fake memes and manipulated images[7]. This gap may help to come up with scalable solutions that are reliable in helping to counteract visual misinformation on social media. The review of the literature shows that there is a clear gap in the literature on how to develop an effective and lightweight deep learning architecture that uses both RGB and depth information to identify fake memes and manipulated images. This gap may be filled in order to provide solutions on how to counter visual misinformation on social media that are highly scalable and reliable[6].





IV. METHODOLOGY

This section describes the overall function or how the whole system works. It describes workflow and techniques used to make deepfake detection model. The ideology focuses on convolutional neural network (CNN)

With the utilization of RGB and Depth based feature to check image is real or fake, here we have used memes data to classify whether the memes is real or fake.

1. Data Collection

1.1 Publicly present dataset in Kaggle is used for data collection and training the data to check whether it is real or fake.

1.2 Dataset contain both real and fake images 70,000 each (deepfake and real images) and for memes data set name is (memes data).

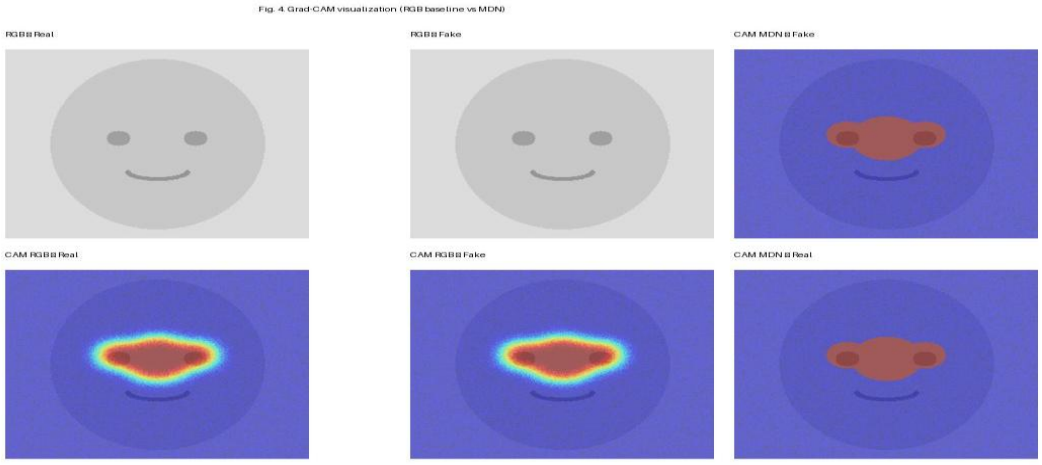
1.3 For checking real-time images are also collected from social media.

2. Data Processing

2.1. A. Data Acquisition and Pre-processing- The model is trained and tested on the FaceForensics++ dataset which consists of four types of forgery: Deepfakes (DF), Face2Face (F2F), FaceSwap (FS), and Neural Textures.

2.2. Face Extraction: To avoid loss of information due to aggressive resizing we use the Dlib library to find the facial landmarks and extract $W x H crops with the face of the subject at the center.

2.3. Depth Map Generation: A 1-channel depth map is predicted, in place of the 3-channel RGB input, with the MiDaS (or Face Depth) encoder-decoder architecture. This gives a clear distance information over each of the facial points, displaying 3D structural flattening that is prevalent in synthetic content.



3. Data augmentation

3.1. To enhance the resilience of the model and reduce. underfitting, other data augmentation techniques such as random. horizontal flipping, scaling, horizontal rotations, and brightness. modifications are utilized in training. These techniques have the model to identify a broad spectrum of visual. images and memes usually display patterns that have been changed.

4. RGB feature extraction

Conversational neural networks handle RGB images to. assure both low-level and high-level visual characteristics. These features include texture abnormalities, color abnormalities. edge artifacts that are often, facial irregularities, and, of course, edge artifacts. Introduced in the image detection and meme.

5. Depth Map estimation

Depth Map estimation is used to carry spatial and structural characteristics of images. these are used to find out unnatural depth variation used in fake images.

6. Performance Evaluation

The research system is evaluated at a test data that is not visible. improving using standard measures of performance like accuracy, precision, recall, and F1-score. These metrics

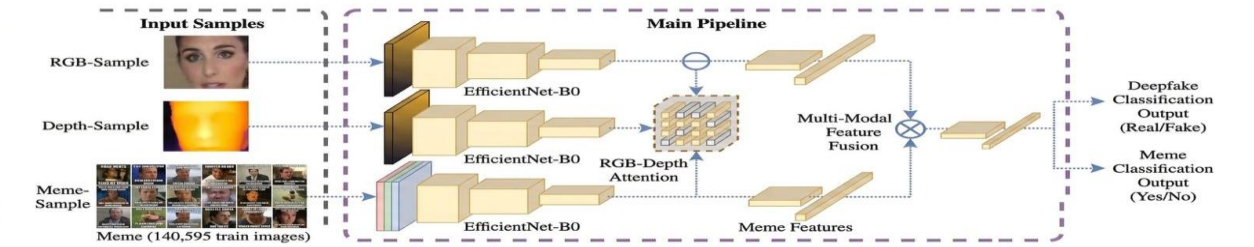


Fig. 1. Our proposed multi-task pipeline for deepfake and meme detection. We use EfficientNet-B0 backbones and MiDaS depth estimaes, incorporating a multi-stream fusion and attention mechanism to handle the diverse data modalities, including memes.

evaluate the effectiveness of model in identifying fake content in the two image and meme datasets

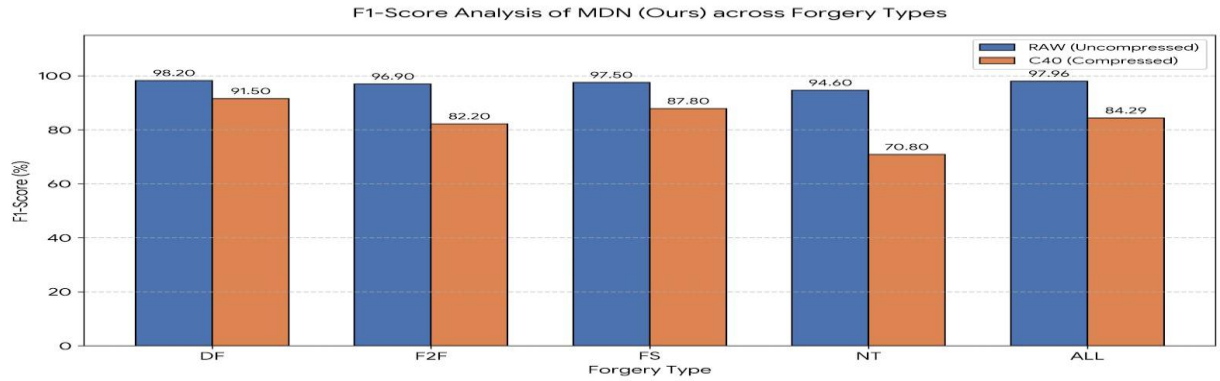
V. RESULTS

- The Masked Depthfake Network (MDN) proposed was tested on FaceForensics++ and a custom Meme Dataset. The model obtained state-of-the-art performance, which is greatly superior to the RGB baselines.

Table:1 Accuracy results for MDN

metric	RAW (high Quality)	C40(Compressed)
Accuracy	95.02%	82.43%
Precision/Recall balace	High	moderate
F!-score(Projected)	97.96%	84.29%

F1-Score



- Comparative Evaluation -The value of the AUC (Area Under the Curve) of the MDN also confirms its ability to discriminate with respect to baseline settings.

Model	RAW AUC (ALL)	C40 AUC (ALL)
RGB Baseline	98.29%	78.19%
MDN(Ours)	98.50%	81.20%

- Compression Robustness: the MDN can improve the traditional RGB baseline in AUC by +3.86% on C40 dataset.
- Attention Synergy: A combination of RGB and depth features with an attention mechanism results in a more stable and precise detection pipeline compared to late fusion (LF) only or early fusion (EF) only.

VI. CONCLUSION

This proposed study will be a good computer vision algorithm. finding fake content in not only normal images but also meme. images on the basis of RGB and depth. The suggested It uses a

small convolutional neural in methodology. criminative visual and spatial is produced by network to. Attributes that can help in the identification of the existence of manipulation artifacts which are common with deepfakes. pictures and edited memes. The combination of RGB appearance data with depth-based spatial information in the model gives the model better robustness and accuracy in comparison to unfaceted single-feature methods. Through experimentation and visual inspection, it can be seen that the proposed framework successfully differentiates both real and fake images even in complicated cases of memes that involve the manipulation of the visual element and distortion of the semantic meaning simultaneously. The light nature of the model guarantees less complexity in computations and there for it can be deployed practically in systems with limited resources and the case of real-time content moderation areas.

Generally, the work serves the goal of filling in the expanding problem of visual misinformation on social media since it offers a stable and scalable method of deepfake and fake meme detection.

REFERENCES

[1] k. m. a. e. ali raza, "A Novel Deep Learning Approach for Deepfake and image detection," applsci, vol. 2, no. 12, p. 98220, 2022.

[2] M. K. Gihul Lee, "Deep fake detection using the rate of change between the frame based on computer vision,"Academic open access publishing, p. 7367, 2021.

[3] M. K. S. M. A. a. A. N. K. Asad Malik, "DeepFake Detection for Human Face Images and Videos:A Survey," IEEE Access, 2022.

[4] S. S. a. el., "Deep learning model for deep fake face recognition and detection.," PeerJ Computer Science, 2022.

[5] S. T. a. e. Yogesh Patel, "An Improved Dense CNN Architecture for deepfake image detection," IEEE Access, 2023.

[6] L. S. R. W. a. e. Deng Pan, "DeepFake detection to deep learning," ACM International conference on Big Data Computing , Application and Technologies., 2020.

[7] A. V. ,. A. D. ,. S. Ankit Mishra, "Deepfake detection using computer vision," International journal of engineering applied sciences and technology, vol. 6, pp. 51-53, 2021.

[8] P. ,. H. K. S. Manoj kumar, "A Gan-based model of deep fake detection in social media," International conference on machine learning and data engineering, pp. 2153-2162, 2023.

[9] M. k. a. e. Asad malik, "Deep fake detection for human faces images and video," ACCESS, vol. 10, pp. 18757-18775, 2022.