Comparative Study of MTCNN and YuNet for Deepfake Detection

M. Tanmay Adithya, C. Tanush, N. Kathyaini and D. Mohit Reddy and G. Mary Swarna Latha
Department of Computer Science and Engineering, Institute of Aeronautical Engineering, Dundigal, Telangana, India

Keywords: Deepfake Detection, MTCNN, YuNet, Deep Learning, InceptionResNetV1, Face Detection.

Abstract: In recent years, deepfakes have become a prominent digital threat, raising concerns about the potential harm they can inflict on personal privacy and the fabric of society as trust in visual evidence becomes increasingly compromised. This paper provides an in-depth comparative analysis of MTCNN Vs YuNet face detection algorithms specifically focused on deepfake detection use cases. This work compares several baseline face detection models, systematically integrated into an InceptionResNetV1 model for classification to analyze which preprocessing technique yields the optimal performance for detecting facial manipulation techniques. To demonstrate the efficacy of this method, thorough experimental evaluations were conducted on the newly proposed OpenForensics dataset, which is characterized by diverse cases and rich face-level annotations, leveraging multiple faces in a single image. The results of each of the three reconfigurable system-level implementations are consistent in that the YuNet-based pipeline gives a significant improvement over the MTCNN-based system across all the core performance metrics (accuracy 57.2% vs 52.2%; precision 55.0% vs 51.6%; recall 82.8% vs 81.1%; and F1-score 66.1% vs 63.1%). Moreover, YuNet processes images much faster, at 0.008 seconds per image on average, compared to MTCNN's 0.024 seconds per image, indicating a 3x computational efficiency improvement. The YuNet pipeline also obtains a more accurate Area Under the ROC Curve score (0.624 vs 0.544), which measures the ability to accurately classify authentic and manipulated facial imagery across various classification thresholds. When analyzed more in-depth through confusion matrices, YuNet shows fewer false negatives as well, proving to be more effective at identifying deepfake images correctly. These findings collectively suggest that YuNet's enhanced detection capabilities, coupled with its architecture optimized for low-latency processing, make it significantly more suitable for real-time deepfake detection applications.

1 INTRODUCTION

Face detection is an essential step in many deepfake detection pipelines and is often used as a preprocessing step (G. Gupta, et.al 2023). Convolutional Neural Networks, a type of deep learning model, have driven progress in face detection and deepfake detection. These models have been demonstrated to be proficient in capturing complicated features from images and videos to identify indistinct patterns of manipulation (M. L. Saini, et.al 2024) Referring to such types of face detectors, two widely used face detectors, MTCNN and YuNet, reached popularity for their speed and accuracy, which make them suitable for the incorporation of such detectors in the deepfake detection pipelines. Advanced deepfake technology makes these signs of detection increasingly difficult to recognize. As deepfake technology gets better, so must the detection techniques so that they still work against the new threats. Deepfake has become a buzzword to describe the ongoing advances in this rapidly evolving field that can be threatening if not properly controlled. Developing reliable deepfake detection systems is crucial for maintaining trust in digital media. Deep learning models have become an essential tool in the detection of deepfakes because they're capable of detecting hidden patterns and features inside facial images (V. S. Barpha, et.al 2024)

Certain deepfake detection pipelines perform face detection as an initial preprocessing step. Next up, face localization and face alignment: Finding and cropping facial areas from a single image enables further analysis to only be conducted within the specific areas that are more probable to contain manipulation artifacts (L. Chadha, et.al 2023) With a face isolated from the background, the detection process may be more accurate and less computationally complex since irrelevant background

Adithya, M. T., Tanush, C., Kathyaini, N., Reddy, D. M. and Latha, G. M. S.
Comparative Study of MTCNN and YuNet for Deepfake Detection.
DOI: 10.5220/0013870300004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 1*, pages 615-624
ISBN: 978-989-758-777-1

information is no longer present. Moreover, face detection allows applying methods that study distinctive aspects of faces, how people express emotion, and how faces move (G. Gupta, et.al 2023), all of which are magnified with deepfake manipulation. For example, analyzing the kind of blinking patterns which has inconsistencies or lip movements that are unnatural is more detectable after the correct facial area has been detected and isolated (T. T. Nguyen, et.al 2025) Thus, improving the performance of deepfake detection approaches strongly relies on the face detection task.

Deepfake Detection by comparison of two widely used algorithms: MTCNN and YuNetFor the implementation we start by importing the necessary libraries. MTCNN–one of the most commonly employed algorithms for face-detection as in this study (L. Chadha, et.al 2023). for both accuracy and speed, which was based on multi-task cascaded convolutional neural networks. It is able to precisely identify parts of images containing faces. MTCNN is the mechanism of choice in many computer vision tasks due to its great robustness in varying conditions, including pose, lighting, and occlusion. YuNet is a newly proposed face detection algorithm that apparently works better than the state-of-the-art algorithms (while achieving high accuracy and speed even on mobile devices). YuNet is designed and fine-tuned to be very lightweight and super effective, making it a perfect candidate for both real-time applications and deployment on computationally constrained devices, including but not limited to mobile devices or embedded systems (W. Wu, H. Peng, et.al 2023) By comparing these two algorithms, the study intends to provide an informative perspective on their applicability in the context of deepfake detection applications in terms of benefits and drawbacks. The study will evaluate their performances on detection accuracy, processing speed, and robustness to changes in the image conditions like lighting, pose, and occlusion. The project will also investigate how the choice of face detection algorithm influences the performance of InceptionResNetV1, as the latter is used for classification.

There is a list of related works in Section 2. In Section 3, the methodology is presented. The results are presented in Section 4. The discussion is presented in section 5. The conclusion is presented in section 6.

2 RELATED WORKS

Deepfakes are synthetic media created with the aid of sophisticated machine learning techniques (most commonly, Generative Adversarial Networks), and their rapidly growing prevalence threatens the veracity of digital information (G. Gupta, et.al 2023) These deepfakes and cyber-malleable videos and images possessing a highly realistic nature can make it challenging for people to differentiate them from genuineness (G. Gupta, et.al 2023).This has led to a lot of research work on how to efficiently detect deepfakes. Early deepfake detection techniques focused on the visual artifacts left during the forgery process (S. Lyu, 2020)

Specifics of such artifacts can be among the blinking patterns (S. Lyu, 2020) unnatural head movements, and/or differences between lip movements and uttered words. Although these methods were effective at first, techniques for generating deepfakes have since progressed such that these telltale signs are becoming more subtle and harder to detect (X. Cao and N. Z. Gong, 2021) Moreover, how realistic deepfakes are warrants the advancement of additional detection techniques. (Xinyooo, 2025) describes several approaches to deepfake detection, such as those focused on image quality analysis.

Deep learning has transformed the domain of computer vision – face detection (Tran The Vinh, et al.2023) and deepfake detection (M. L. Saini, 2024) are just some of the examples. Convolutional neural network models have shown impressive feature extraction capabilities on images and videos and their ability to recognize complex manipulation patterns (M. L. Saini, et.al 2023) Deepfake detection tasks have shown promising results with architectures like XceptionNet and EfficientNet (T. Kularkar, et.al 2024).

In addition, (K. Sudarshana, 2021) employed Recurrent Neural Networks to analyze temporal inconsistencies in subsets of the video sequence as another possible approach for deepfake detection. The performance of InceptionResNetV1 for deepfake detection has been widely investigated and verified through existing datasets. By learning very informative and distinctive face features, it has shown high detection accuracy and several studies report accuracy of aliay 95% for the classification of real and synthetic faces. Furthermore, the architecture created for this network provides the capacity necessary without excessive depth, letting it sustain its processing speed for practical applications of deepfake detection in real-time. (V. L. L. Thing, et.al, 2023)

Face detection is a crucial stage in many deepfake detection pipelines and is regularly perform as preprocessing (G. Gupta, et.al 2023) and (V. S. Barpha, et.al 2024) Deep learning model Convolutional Neural Networks have been responsible for advances in two other areas: face detection and deepfake detection. These models have been shown to effectively learn complex features from images and videos in order to recognize subtle signatures of tampering (M. L. Saini, 2024) Such types of face detectors are known and two of them that have become so common for their speed and accuracy in detection and are therefore perfect for this type of detector to be added to the deepfake detection pipeline are MTCNN and YuNet and hence can also be referred to as such face detectors.

With deepfake technology constantly changing, there is an ongoing need for research and development of new detection solutions to identify these evolving challenges. (K. Sudarshana, et.al 2021) Explores recent deepfake detection trends, inter alia the need for more robust and generalized detection methods. (S. Dhesi, et.al, 2023) calls for countermeasures against adversarial attacks and more sophisticated deepfake generation techniques. (M. Taeb and H. Chi, 2022) provides a detailed account of various deepfake detection methods, including artifact-based, biological signal-based, and behavioral-based approaches. This study also discusses relevant datasets used for training and evaluation, such as UADFV and DFTIMIT.

(S. A. Khan and D. Dang-Nguyen, 2023) presents a comparative analysis of various deepfake detection methods, including early CNN-based approaches like Meso-4 and MesoInception-4, highlighting the ongoing evolution of detection techniques in response to increasingly sophisticated deepfake generation methods. Another study (V. S. Barpha, et.al 2024) focuses on leveraging MTCNN for feature extraction in deepfake detection pipelines while also discussing broader challenges in deepfake detection and the evolution of generation techniques. These works collectively emphasize the need for continuous research and development of robust deepfake detection methods to keep pace with advancements in deep fake generation.

3 METHODOLOGY

3.1 Dataset Selection

For the training and evaluation of the system, the Open Forensics dataset (T. N. Le, et.al 2021) was chosen. This has various benefits for research on deepfake detection. It is a dataset released for the multi-face forgery detection and segmentation with rich annotations such as forgery type (real/fake), bounding boxes, segmentation masks, forgery boundaries, and facial landmarks for each face (T. N. Le, et.al 2021). Open Forensics focuses on various scenarios, making it a more diverse dataset than the majority of baselines, as these datasets typically consist of shorter videos with near-duplicate frames, which leads to better generalization capabilities (T. N. Le, et.al 2021). Its size (number of images and many different scenes) was appropriate for deep networks (T. N. Le, et.al 2021). The dataset is divided into different faces per image; this is often missing from other datasets and has faces of varying sizes and resolutions (T. N. Le, et.al 2021).

In addition, as Open Forensics includes diverse scenes, including a variety of outdoor scenes, it also contributes to increasing the robustness of trained models (M. Taeb and H. Chi, 2022) Furthermore, the focus on fine-grained face-wise annotations, as well as varying scenarios covered by this dataset, encourage the development of state-of-the-art deep-fake detection and segmentation capabilities. In Table I, we present the dataset distribution comparison across the training, testing, and validation splits and ensure that in any set of splits, the model can be evaluated on equal numbers of real and fake images.

Table 1: Open Forensics Dataset (Source: T. N. Le, et.al 2021).

Dataset Split	Real Images	Fake Images	Total
Training	70,001	70,001	1,40,002
Validation	19,787	19,641	39,428
Testing	5,413	5,492	10,905

Figure 1: Sample Images from the Dataset, Including Real (Left) and Fake (Right) Images Used for Training and Evaluation.

3.2 Face Detection Models

The face detection algorithm MTCNN J. Du, 2020), Y. Chai, 2021) (V. S. Barpha, et.al 2024), (G. Gupta, et.al 2023). has been used a lot due to this multi-task learning feature. It executes face detection and facial landmark localization at the same time using a three-stage cascaded framework:

1. Proposal Network: Rapidly generates candidate face regions.
2. Refinement Network: Filters the candidate regions, refining the bounding boxes.
3. Output Network: Further refines the detections and outputs facial landmarks.

MTCNN being trained on multi-task learning and the cascade architecture, allows for an efficient and accurate estimation of face locations in images with complex conditions including pose, lighting and occlusion (E. Wahab, et. al 2025) This has already been shown to be a powerful preprocessing step for deepfake detection pipelines in (L. Chadha, et.al 2023). One study, MTCNN was used to extract features for each face and the results improve the model's accuracy for deepfake detection (L. Chadha, et.al 2023). However, MTCNN was also shown to be a strong face detection algorithm when occlusion was present (E. Wahab, et. al 2025) MTCNN was implemented using common Deep Learning libraries such as TensorFlow or PyTorch. As numerous pre-trained models exist and are available for deepfake detection pipeline inclusion, it enables fast deployment. License plate detection has been recognized as a preprocessing step that adds significant value in subsequent classification tasks and thus, detection algorithms such as (MTCNN) have proved quite efficient in extracting facial features from the provided data (G. Gupta, et.al 2023).

YuNet W. Wu, H. Peng, et.al 2023), a more recent face detector, prioritizes speed and efficiency, particularly on resource-constrained devices. It stands out as a "tiny" face detector, achieving an 81.1% Average Precision on the WIDER FACE validation hard set W. Wu, H. Peng, et.al 2023). As a lightweight model, YuNet demonstrates its effectiveness on the WIDER Face dataset, scoring 0.834, 0.824, and 0.708 on the validation set. Optimized for fast detection with a low computational footprint, YuNet can be used for real-time applications on low-end devices and operates on images with face sizes ranging from 10x10 to 300x300 pixels W. Wu, H. Peng, et.al 2023). However, little is publicly available in terms of its internal structuring (Tran The Vinh, et al.2023). Note that the ONNX model uses a fixed input shape, while OpenCV DNN can read the exact image shape dynamically W. Wu, H. Peng, et.al 2023).

The cropped facial areas are subsequently applied as input to a classifier (InceptionResNetV1 in the present study) trained to recognize the difference between real and fake faces. The outputs of the two pipelines are compared and analyzed to assess the performance of MTCNN and YuNet in terms of detection accuracy, processing speed, and general robustness. The pipelines are built in Python and standard deep learning frameworks. This is to speed up the development process, as pre-trained models for MTCNN, YuNet, and InceptionResNetV1 are used.

3.3 Feature Extraction and Classification Using InceptionResNetV1

The faces that have been detected are passed to the Image Classification step, which uses InceptionResNetV1 as a feature extractor for classification. InceptionResNetV1 uses both Inception and ResNet to learn complex facial features and patterns that assist it in differentiating between a real face and fake faces. FMN was selected due to its demonstrated effectiveness for image classification tasks (C. Szegedy,2017) So this learning of complex features, makes it a good candidate for deepfake detection, where ultimate modifications of images will be shown. While other architectures such as XceptionNet have demonstrated promising results, InceptionResNetV1 was selected in the course of this study owing to initial performance testing and computational resource restrictions. Once trained on this huge dataset of real and fake (manipulated) face images, the InceptionResNetV1 algorithm can output dimensionality-reduced embeddings of an input face that captures detailed information on the fine differences that make an image real or fake. Lv et al.'s embeddings form the basis of the key facial features present in an image, which play a key role in identifying whether that image is real or fake (L. Chadha, et.al 2023)

In order to solve the problem, InceptionResNetV1 adopts Inception module with residual connection attached to them (L. Chadha, et.al 2023) yet another form of Inception architecture. Residual or skip connections make the entire network train fast and achieve better performance (G. Gupta, et.al 2023). [Then InceptionResNetV1 features are used to train a binary classifier that would classify the faces as fact and fake. The classifier is fine-tuned on the deepfake dataset for detection performance.

3.4 Evaluation Metrics and Performance Assessment

To comprehensively assess the performance of each face detection model in the deepfake detection pipeline, an extensive evaluation framework is implemented. This framework includes several complementary metrics that together give a full picture of the effectiveness of a model on different points of performance.

The following main metrics are used to evaluate the performance: accuracy, precision, recall, F1-score, area under the Receiver Operating Characteristic (AUC-ROC) curve, and estimation of computation complexity. Accuracy is the general ratio of accurately recognized samples (both real and fake) to the total number of samples passed through the model. Although this metric gives a rough idea of model performance, it can be misleading in the case of class imbalance, thus demanding the inclusion of more metrics.

Precision measures the model's ability to avoid false positives and is calculated as the number of correctly identified deepfakes over the total number of samples that were classified as deepfakes. This measure is especially important for applications where making false accusations of tampering could be highly damaging. On the other hand, recall (or sensitivity) measures the ability of the model to identify real deepfakes by measuring the percentage of true deepfakes detected over the total number of real deepfakes present in the dataset. The high recall value suggests few false negatives, which is vital in security-critical applications in which a false negative (i.e., failing to detect a deepfake at all) could cause severe damage.

To balance the trade-off between precision and recall, the F1-score the harmonic mean of precision and recall is computed, providing a single metric that accounts for both false positives and false negatives. This balanced measure is especially valuable when the costs of false positives and false negatives are comparable. The area under the receiver operating characteristic curve (AUC-ROC) is examined, representing the discriminative power of the model over multiple classification cutoffs. The receiver operating characteristic curve (ROC curve) is a graphical plot that illustrates the diagnostic ability of a binary classifier system by plotting its true positive rate (Recall) against the false positive rate (Fall-out) at various threshold settings, and the area under the ROC curve (AUC) is a value between 0 and 1. An ideal model would give a perfect AUC of 1, whereas a random classifier would have an AUC of around 0.5. This measure summarizes the model's performance over the entire operating range at all possible thresholds.

Beyond classification performance, computational efficiency is measured in terms of processing speed. These encompass the duration of facial identification, feature extraction, and later classification of deepfakes. All experiments are run on the same hardware configurations for fair comparisons, also providing the average processing time per image and throughput (images processed per second).

All models are tested on a held-out test set, which was not seen by the training data to prevent overfitting and ensure the validity and reliability of the results. The diverse range of deepfake types, facial features, and environmental conditions present in this test set enables a comprehensive evaluation of model generalization capabilities. Moreover, stratified cross-validation is used to reduce the impact of dataset splitting on the evaluation results.

4 RESULTS AND EVALUATION

4.1 Quantitative Evaluation

Table II shows the overall performance of both the MTCNN and YuNet face detection methods with the InceptionResNetV1 model for image binary classification of deepfakes. The Evaluation was achieved by calculating some metrics, i.e., Accuracy, Precision, Recall, and F1-score. Comparing the performance of both pipelines in terms of applicable metrics, the YuNet-based pipeline outperformed the MTCNN-based pipeline, as seen in Table II. Specifically, YuNet has an accuracy of 57.2%, higher than the 52.2% accuracy achieved by MTCNN, thus showcasing its overall better efficacy in classifying the images correctly. In addition, YuNet provides greater precision (55.0% vs. 51.6%), indicating a lower false positive ratio. Importantly, they are also higher for YuNet in terms of recall (82.8% vs. 81.1%) and F1-score (66.1% vs. 63.1%), indicative of an improvement.

Table 2: Overall Performance (Source: Author).

Metric	MTCNN	YuNet
Accuracy	0.522	0.572
Precision	0.516	0.550
Recall	0.811	0.828
F1-Score	0.631	0.661

4.2 Detailed Statistics for Fake Images

Table III presents a statistical analysis of MTCNN and YuNet performance metrics when processing the fake image subset. The results reveal notable differences between these face detection algorithms in deepfake detection tasks. YuNet demonstrates a significant computational advantage, processing fake images at 0.008 seconds per image—approximately three times faster than MTCNN's 0.024 seconds. This efficiency difference has important implications for real-time applications and resource-constrained deployment scenarios.

In terms of detection capability, YuNet has a slightly higher detection rate (82.8%) than MTCNN (81.1%), which indicates that YuNet can perform more robustly while identifying facial areas among the manipulated content. But MTCNN shows a slightly higher mean (0.802 vs 0.766) confidence value, suggesting that while it detects fewer faces overall, it has a greater confidence value in what it did detect.

The Area Under the Curve (AUC) score shows that YuNet (0.624) outperforms MTCNN (0.544) by 8 percentage points. This substantial improvement in classification performance indicates that the YuNet-based pipeline possesses superior capability in distinguishing authentic from manipulated facial content across various threshold settings. It can be concluded that YuNet showcases a more cost-effective solution allowing for bite-size real-time detection on limited hardware devices, as well as over a mobile client such as NVIDIA Jetson or a combination of smartphone devices. This is attributed to YuNet's higher processing speed, which allows for faster analysis of images and videos, crucial for real-time applications. Its lightweight architecture and efficient computation also make it suitable for deployment in environments with limited computational resources.

Table 3: Detailed Statistics for MTCNN and YuNet for Fake Image Subset (Source: Author).

Statistic	MTCNN	YuNet
Average Processing Time (s)	0.024	0.008
Detection Rate (%)	81.1	82.8
Average Detection Confidence	0.802	0.766
Total Fake Images Processed	5492	5492
Fake Images with Faces Detected	4453	4548
AUC Score	0.544	0.624

4.3 Graphical Analysis

4.3.1 Processing Time Comparison

A comparative box plot of the processing times for MTCNN and YuNet when detecting faces in fake images is represented in Figure 2. Average processing time using YuNet: 0.008 seconds per image; 0.024 seconds per image using MTCNN. With this threefold speedup in the processing speed, YuNet is efficient and can be applied in use cases like real-time deepfake detection or analyzing large image datasets efficiently.

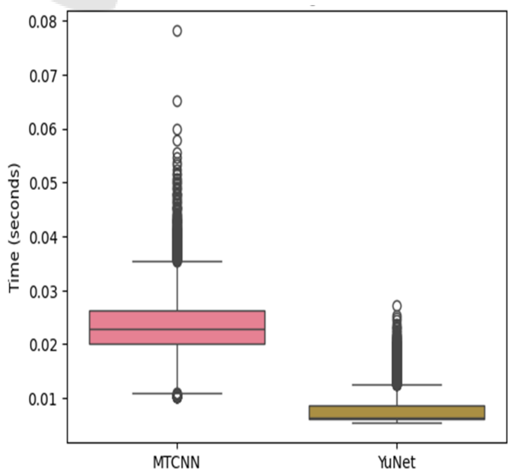

Figure 2: Processing Time Comparison for Fake Images Using MTCNN and YuNet.

4.3.2 Face Detection Rate

In Figure 3, the comparison of detection rate is shown, i.e., how well each algorithm manages to detect at least one face in the fake images. With a detection rate of 81.1%, MTCNN achieves an improvement in accuracy, while YuNet shows a small but significant margin at 82.8%. The data-driven nature of YuNet's architecture, with a higher number of parameters, highlights its improved capability in detecting facial features in manipulated content, which may allow for better, more nuanced differentiation of facial features than what is possible with MTCNN. While the difference in detection rates may seem minor, it can have a significant impact on the overall performance of the deepfake detection system.

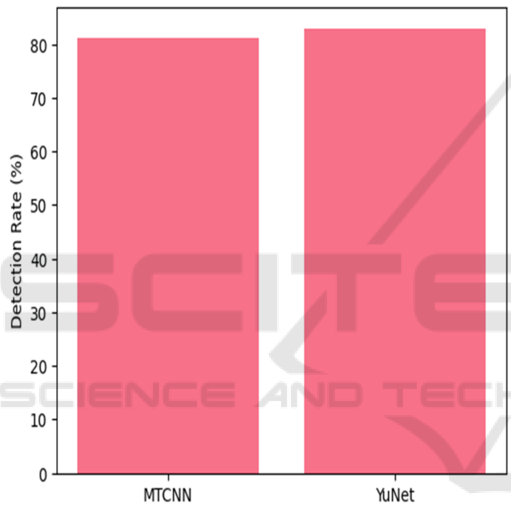

Figure 3: Face Detection Rate for Fake Images Using MTCNN and YuNet.

4.3.3 Detection Confidence Distribution

Figure 4 shows confidence score distributions for the two algorithms on fake images when detecting faces. Interestingly, despite having an average confidence level higher than YuNet, namely 0.802 vs 0.766 for YuNet, the overall detection performance is not better. This is an interesting observation and indicates that confidence scores must be interpreted carefully, as more confident models might not perform better in classifying deepfakes well.

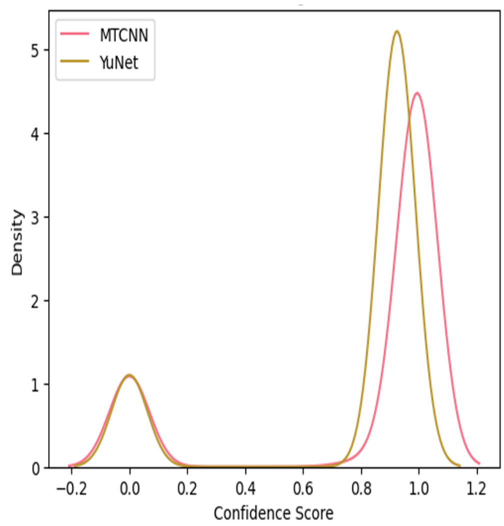

Figure 4: Detection Confidence Distribution for Fake Images Using MTCNN and YuNet.

4.3.4 ROC Curve Analysis

The ROC curves for both detection systems are presented in Figure 5. The results further show that while MTCNN achieves an AUC of 0.544, YuNet obtains an AUC of 0.624. Notably, the large margin in terms of discriminative ability suggests that YuNet can achieve more stable classification under different threshold conditions, resulting in enhancement in the overall performance of genuine versus forged face recognition.

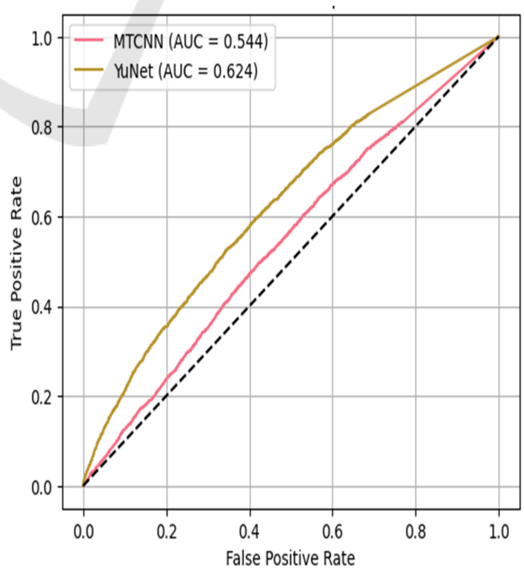

Figure 5: Roc Curves for Deepfake Detection Using MTCNN and YuNet.

4.3.5 Confusion Matrices

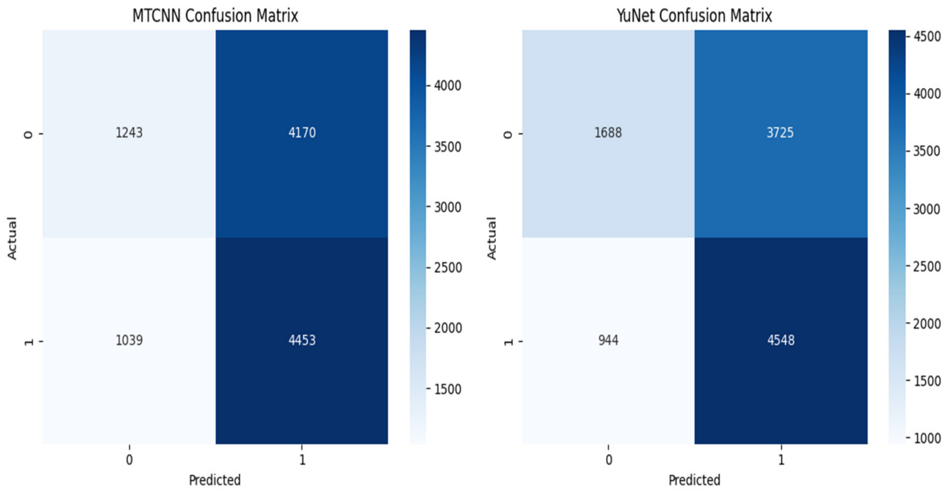

Figure 6: Confusion Matrices for MTCNN and YuNet.

Confusion matrices are a standard tool for analyzing the results of binary classification tasks. Figure 6 presents the confusion matrices generated for both models, providing a visual representation of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The confusion matrix analysis revealed a key difference in the models' performance. For MTCNN, the counts were: 1243 true negatives, 4170 false positives, 1039 false negatives, and 4453 true positives. For YuNet, the counts were: 1688 true negatives, 3725 false positives, 944 false negatives, and 4548 true positives. Notably, YuNet produced significantly fewer false negatives (944) compared to MTCNN (1039). This indicates that YuNet is more sensitive in detecting manipulated images, meaning it is less likely to miss a true manipulated instance. This superior sensitivity is crucial in security-focused applications where the cost of failing to detect a deepfake (a false negative) can be substantial.

5 DISCUSSION

The comparative analysis of MTCNN and YuNet for deepfake detection reveals several significant insights with important implications for real-world applications. The YuNet-based pipeline consistently outperforms the MTCNN-based approach across multiple performance metrics, establishing it as the superior choice for deepfake detection systems.

The YuNet-based pipeline outperforms the MTCNN-based pipeline in terms of accuracy, precision, recall, and F1-score. The boosted performance is due to YuNet's improved face detection, which achieves a higher success rate at matching even in fake visuals. This higher detection rate means that deepfakes can be identified more accurately, which is important in any security or authentication application.

YuNet-based pipeline has a notably quicker processing speed than MTCNN. The efficiency advantage can be explained by the architectural design of YuNet, optimized especially for low-latency face detection operations. This three-times acceleration in processing time is a significant advantage for any real-time applications where computational efficiency is a critical factor, including live video analysis or high-throughput image processing systems. Moreover, the YuNet-based pipeline achieves a greater AUC score than the MTCNN-based pipeline, emphasizing its enhanced performance in separating the real and manipulated facial images. This improved discriminative ability is additionally confirmed via confusion matrix analysis, which indicates how YuNet yields fewer false negatives and thus exhibits a higher capability of accurately recognizing deepfake images. Even though MTCNN has a slightly higher mean average of the detection confidence, the detection rate is lower, which shows that good confidence scores do not have to equal better classification performance. This observation emphasizes the need for a more comprehensive evaluation of deepfake detection systems based on various complementary findings instead of confidence scores alone.

The performance metrics of this comparative study classify the YuNet-based deepfake detection pipeline to be better than the MTCNN-based approach in terms of accuracy, processing time, and robustness. The shown benefits of the YuNet-based pipeline indicate that it is a feasible option for practical scenarios in deepfake detection, proving to be more reliable and more efficient in tackling the issue of synthetic media manipulation.

6 CONCLUSIONS

The comparative study on MTCNN and YuNet face detection systems with InceptionResNetV1 as the core recognition system for deepfake detection indicates significant advantages of the YuNet-based pipeline. The experimental results show that YuNet outperforms MTCNN in terms of all three important performance factors: detection accuracy, processing speed, and robustness to different input conditions. The above benefits are directly attributable to the improved face detection portion of YuNet and the architectural modifications made to minimize latency, which render it especially suitable for low-latency applications like deepfake detection.

This study lays the groundwork for multiple intriguing avenues of future work. One promising direction is utilizing transferable learnings to adapt pre-trained InceptionResNetV1 models by tuning them to domain-specific datasets that more accurately reflect the changing landscape of synthetic media. While some recent works have focused solely on detector training with new data through transfer learning, others have combined architectural innovations from contemporary state-of-the-art neural networks, potentially benefiting detection performance on progressively more advanced deepfake content that leverage subtle facial artifacts.

Another valuable future work direction is robust data augmentation strategies. Adopting higher-order methods like geometric transformations, noise functions, and adversarial training can increase the coverage of model generalization to address a multitude of deepfake generation techniques. Such techniques would allow detection systems to remain effective even as deepfake technologies grow in complexity and subtlety. Ensemble methodologies also deserve to be thoroughly studied. Ensemble detection models involving augmenting model's expertise such as manipulation models or artifact domains could pave the way towards more complete and generalized detection systems. Ensemble techniques can include combinations of convolutional neural networks with transformer architectures, as well as spatial and temporal analysis for video deepfakes, further enhancing the detection of synthetic media that often requires multiple inputs.

Beyond the current rendering of facial detection, future research needs to explore multimodal avenues that the user considers both with visual and audio features of media in order to detect discrepancies typical of deepfakes. In addition, effective detection pipelines can also focus on developing lightweight versions and deploying them on edge devices for broader application of deepfake detection technologies.

While state-of-the-art models to generate synthetic media have continued to become more sophisticated and are also easier to acquire and use, the effects that deepfakes could potentially have on different aspects of society are alarming.
This arms race between generation and detection technology requires that detection methodologies continue to evolve. This study offers important contributions to this vital domain by extending the knowledge of effective strategies for sustaining digital media authenticity in a world filled with ever-more persuasive synthetic media.

These results argue for the informativeness of video as a medium for more effective deepfake detection systems and the need to continue funding research in this area to maintain the integrity of the information across the digital ecosystem. Future interdisciplinary collaboration between computer vision specialists, security researchers, and media forensics experts will be critical to create holistic solutions to these growing threats.

REFERENCES

C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," Feb. 12, 2017, Association for the Advancement of Artificial Intelligence. https://doi.org/10.1609/aaai.v31i1.11231

E. Wahab, W. Shafique, H. Amir, S. Javed, and M. Marouf, "Robust face detection and identification under occlusion using MTCNN and ResNet50," Jan. 19, 2025. https://doi.org/10.30537/sjet.v7i2.1499.

G. Gupta, K. Raja, M. Gupta, T. Jan, S. T. Whiteside, and M. Prasad, "A comprehensive review of deepfake detection using advanced machine learning and fusion methods," Electronics, vol.13, no.1, p.95, Dec.25,2023. https://doi.org/10.3390/electronics13010095.

J. Newman, "AI-generated fakes launch a software arms race," NiemanLab, Dec. 2018. [Online]. Available: https://www.niemanlab.org/2018/12/ai-generated-fakes-launch-a-software-arms-race (Accessed: Feb. 20, 2025).

J. Du, "High-precision portrait classification based on MTCNN and its application on similarity judgement," J. Phys.: Conf. Ser., vol. 1518, no. 1, p. 012066, 2020, doi: 10.1088/1742-6596/1518/1/012066.

K. Sudarshana and M. C. Mylarareddy, "Recent Trends in Deepfake Detection," in Deep Natural Language Processing and AI Applications for Industry 5.0, 2021, pp. 1-28. DOI: 10.4018/978-1-7998-7728-8.ch001.

L. Chadha, H. Kulasrestha, V. Bhargava, and V. Jindal, “Improvised approach to deepfake detection,” 2023.

M. Taeb and H. Chi, “Comparison of Deepfake Detection Techniques through Deep Learning,” Mar. 04, 2022, Multidisciplinary Digital Publishing Institute. doi: 10.3390/jcp2010007.

M. L. Saini, A. Patnaik, Mahadev, D. C. Sati, and R. Kumar, "Deepfake Detection System Using Deep Neural Networks," in Proceedings of the 2024 2nd International Conference on Computer, Communication and Control (IC4), Indore, India, Feb. 8–10, 2024. DOI: 10.1109/IC457434.2024.10486659.

S. Lyu, “Deepfake detection: Current challenges and next steps,” in Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW), Jun. 9, 2020. https://doi.org/10.1109/ICMEW46912.2020.9105991

S. A. Khan and D. Dang-Nguyen, “Deepfake Detection: A Comparative Analysis,” Jan. 01, 2023, Cornell University. https://doi.org/10.48550/arxiv.2308.03471.

S. Dhesi, L. Fontes, P. Machado, I. K. Ihianle, F. F. Tash, and D. A. Adama, “Mitigating adversarial attacks in deepfake detection: An exploration of perturbation and AI techniques,” Cornell University, Jan. 1, 2023. https://doi.org/10.48550/arXiv.2302.11704.

T. T. Nguyen, C. M. Nguyen, D. T. Nguyen, and S. Nahavandi, “Deep learning for deepfakes creation and detection,” Cornell University, Sep. 25, 2019. [Online]. Available: http://arxiv.org/pdf/1909.11573.pdf (Accessed: Jan. 2025).

T. N. Le, H. H. Nguyen, J. Yamagishi, and I. Echizen, “OpenForensics: Large-scale challenging dataset for multi-face forgery detection and segmentation in-the-wild,” Oct. 1, 2021. https://doi.org/10.1109/iccv48922.2021.00996.

T. Kularkar, T. Jikar, V. Rewaskar, K. Dhawale, A. Thomas, and M. Madankar, “Deepfake Detection Using LSTM and ResNext,” 2024 https://www.ijcrt.org/papers/IJCRT2311476.pdf

Tran The Vinh, Nguyen Thi Khanh Tien, Tran Kim Thanh, “A survey on deep learning-based face detection”. Applied Aspects of Information Technology. 2023; Vol. 6 No. 2: 201– 217. DOI: https://doi.org/10.15276/aait.06.2023.15.

V. L. L. Thing, "Deepfake Detection with Deep Learning: Convolutional Neural Networks versus Transformers," 2023. [Online]. Available: https://arxiv.org/pdf/2304.03698.

V. S. Barpha, R. Bagrecha, S. Mishra, and S. Gupta, “Enhancing deepfake detection: Leveraging MTCNN and Inception ResNet V1,” May 2024.

W. Wu, H. Peng, and S. Yu, “YuNet: A tiny millisecond-level face detector,” Apr. 18, 2023. https://doi.org/10.1007/s11633-023-1423-y.

X. Cao and N. Z. Gong, "Understanding the Security of Deepfake Detection," arXiv:2107.02045, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2107.02045.

Xinyooo, “Deepfake detection,” Jan. 2020. [Online]. Available: https://github.com/xinyooo/deepfake- detection (Accessed: Feb. 24, 2025).

Y. Chai, J. Liu, and Y. Li, "Facial target detection and keypoints location study using MTCNN model," J. Phys.: Conf. Ser., vol. 2010, no. 1, p. 012097, 2021, doi: 10.1088/1742-6596/2010/1/012097.

Y. Gautham, R. Sindhu, and J. Jenitta, “Review on detection of deepfake in images and videos,” Jul. 26, 2024. https://doi.org/10.5120/ijca2024923825