Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 10 Issue 02
2026

*Research Article*
https://doi.org/10.56979/1002/2026/1203

# DenseNet-Based Detection of AI-Generated Driving Scene Images

**Dhairya Vyas[1*], Milind Shah[2], Khushboo Trivedi[3], Bhasha Anjaria[3], Bhumi Shah[3], and Sachin Patel[4]**

[1]Managing Director, Shree Drashti Infotech LLP, Vadodara, Gujarat, India.
[2]Department of Computer Engineering, Sardar Vallabhbhai Patel Institute of Technology (S.V.I.T), Vasad, Gujarat, India.
[3]Computer Science and Engineering Department, Parul University, Vadodara, Gujarat, India.
[4]Livelyhood Expert, Entrepreneurship Development Institute of India, Ahmedabad, Gujarat, India.
[*]Corresponding Author: Dhairya Vyas. Email: dhairyavyas@live.com

**Received:** November 10, 2025 **Accepted:** January 26, 2026

**Abstract:** The proliferation of deepfake technologies poses a significant challenge to the integrity of image data used in autonomous driving systems, where the distinction between real and manipulated images is critical for safe and reliable operation. This study proposes a novel deepfake detection framework designed specifically for real and fake image classification in autonomous driving environments. The primary aim is to enhance the robustness of autonomous systems against adversarial manipulations by leveraging advanced deep learning techniques. The proposed model incorporates DenseNet blocks to efficiently extract hierarchical features from complex visual data, ensuring improved detection accuracy and computational efficiency. The methodology includes preprocessing the dataset, augmenting it to simulate real-world variations, and training the model on a diverse set of real and fake images. Experimental results demonstrate the efficacy of the proposed framework, achieving an impressive 98% classification accuracy, thereby underscoring its potential as a reliable solution for real-time deepfake detection in autonomous driving scenarios.

**Keywords:** Deepfake Detection; Autonomous Driving; Classification; DenseNet Block; Deep Learning Framework

## 1. Introduction

The distribution of fake videos and images using digital transformation has developed into an important societal problem. As a result of the rapid development of deep learning in recent years, models that have significant learning capabilities, such as CNN and GAN, have been increasingly accessible to the public. This is because deep learning has been growing at a quick pace. As a result of the development of automated and easily obtainable deep learning models and processing units (GPUs), the process of generating fake videos via the use of deep learning techniques has grown less complicated. For instance, there is a method that is referred to as Deepfakes that makes use of deep learning technology to replace (swap) the human faces that are present in videos with other faces. The capability of replacing the identity of a person in a video with that of another person brings with it a wide range of uses that were not anticipated, in addition to concerns over the safety of information. This is because the face is an essential component of an individual's personality. When associated with speech synthesis technology, which can lead to confusion regarding identity recognition, this face manipulation approach has resulted in major societal problems [1]. To identify incidents of face transformation in images and videos, several different automated algorithms have been developed. Nonetheless, the primary challenge with these systems lies with the fact that the detectors are trained on fake images generated by a well-recognized modification method, making them proficient at detecting images produced similarly. Consequently, the efficiency of this detector in detecting images produced by novel forging procedures is significantly reduced. This encompasses the capability to detect fake images [1].

For example, the detecting method that has been proposed by several researchers can train and detect the Face2Face dataset that is part of the FaceForensics++ collection with a level of accuracy that is 92.77 percent. However, when the same trained model is applied to detect the Deepfakes data set of FaceForensics++, the accuracy falls dramatically to 52.32. This is an enormous drop from the previous performance. As a rule, it is generally accepted that it is not possible to create a forgery detection system that can detect each new method of forgery. This is because the method of forgery can be refined to overcome the detection mechanisms that are currently in place [1]. It has seen significant advancements in the technologies that are used for image creation and manipulation because of the development of convolutional neural networks (CNNs). Even though synthetic face images are becoming increasingly realistic, they are technically capable of supporting misbehavior that is associated with the methodologies of facial image synthesis. On the Internet, for instance, there are some misleading and improper political speeches, such as Obama's complaints about Trump, as well as romantic movies, in which the faces contained inside the videos are replaced with the faces of celebrities. Another example is the application known as Zao, which allows you to modify your face and the faces of others in the video by just taking a selfie and doing so in a matter of seconds [2]. Numerous synthetic face image classification models have emerged to address these security concerns. These models can distinguish between genuine and fabricated images efficiently. However, they are only able to provide a legitimate or fabricated result and are unable to provide exact classification particulars [2].

Image forensics, which is a discipline of technology that focuses on confirming the authenticity of provided images, is often considered to be the typical field of technology that handles diverse image transformations. In the discipline of image forensics, the smallest signs of forgery that are still present in images under a variety of configurations (such as the kind of transformation, the algorithm, the parameters, and the compression quality) are analyzed and discovered. For example, it is feasible to validate the integrity of an image by utilizing forensic techniques to ascertain whether the image has been manipulated [3].

### 1.1. Contribution

**The key contributions of this work are:**

1. A DenseNet-based architecture reconfigured for small-scale AI-generated driving datasets.
2. A leakage-free augmentation protocol tailored for safety-critical applications.
3. Comprehensive benchmarking under standardized training conditions.
4. Evaluation of real-time feasibility through inference time analysis.

## 2. Background & Significance

### 2.1. Research Questions

This research paper addresses the following research questions:

Q1: What is the significance of real and fake image classification, particularly for autonomous driving scenarios?

Q2: What are the challenges and advancements in detecting AI-generated autonomous driving scenes compared to other AI-generated content?

Q3: How can deep learning models be effectively integrated into real-time systems for authenticity detection in autonomous driving?

Q4: What are the recent challenges and future directions in real and fake image classification for autonomous driving scenes?

### 2.2. Importance and Significance of Real and Fake Image Classification

Deepfakes are a specific kind of technology that makes use of deep learning to generate fake movies, images, texts, or events. Another name for this technology is "deepfakes”. Over the course of 2017, the word "deepfakes" was initially utilized on Reddit, and ever since then, the use of deepfakes has risen. It is common practice to build deep fakes using one of two methods. The procedure that comes first is known as the face-swapping method. The face of a person from the input image is replaced with another face, often from a vast collection of faces, to accomplish this procedure. A generative adversarial network, often known as a GAN, is the second approach that can be utilized. The GAN algorithm was initially proposed in a few papers. It is a representation of two artificial neural networks (ANN) that compete to provide the optimal answer, which in this instance is the fake face that is the most realistic. The first artificial neural network (ANN) is responsible for producing a new image by selecting random examples from the dataset. The second ANN is responsible for determining whether the image that was formed is genuine. Deepfakes are developed because of technological advancements, high-resolution videos and images, and the development of artificial intelligence algorithms. These deepfakes have a highly lifelike appearance, and it is highly difficult to detect whether they are real [4].

**Figure 1** below shows the basic steps of the machine learning model. It is essential to have the ability to differentiate between authentic imagery and that which is created by machine learning models to distinguish between the two types of imagery. Verification of the image's validity and originality is accomplished by its detection of real data. For example, a Stable Diffusion Model (SDM) that has been fine-tuned might be utilized to develop a synthetic image of an individual committing a crime or vice versa. This would provide fake proof of justification for a person who was involved in something else. Identification of real data also confirms that the image is authentic and original. It is a serious concern in today's world that misinformation and fake news are widespread, and it is possible that machine-generated imagery might be exploited to influence the general population [5].

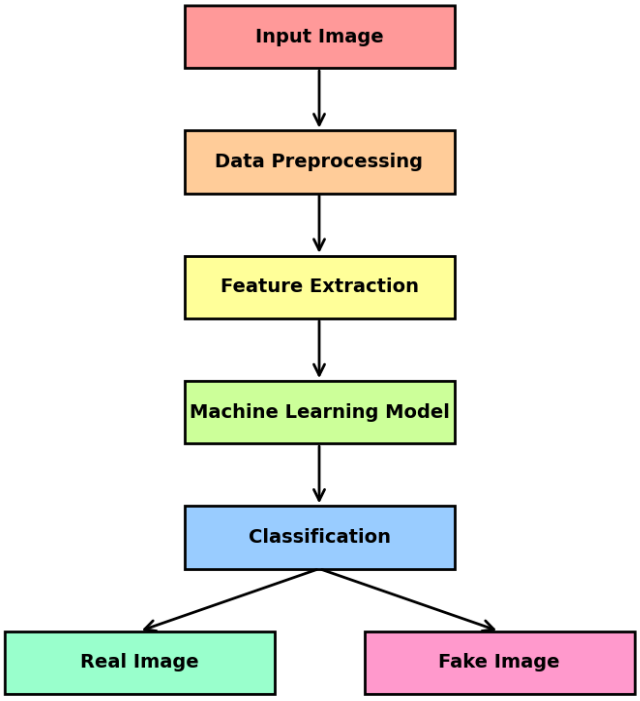


**Figure 1.** Basic flow of Real vs Fake Image Classification

Even if there has been a significant amount of research conducted around deepfake detection, there is always an opportunity for development in terms of both our efficiency and our effectiveness. It should be highlighted that the techniques to produce deepfakes are advancing at an impressive rate, which is leading to the creation of datasets that are becoming increasingly challenging, on which the approaches that were previously used would not perform effectively. Automated deep learning-based deepfake detection systems are being developed to mitigate the possible damage that might be caused by deepfake technology. Deepfake information has the potential to mislead and influence individuals, which can result in severe consequences, such as instability in politics, financial fraud, and harm to one's image. The development of such systems may have significant positive effects on a variety of fields and sectors worldwide. Furthermore, these technologies enhance the trustworthiness and dependability of material found in the media and online. As the technology behind deepfakes continues to advance and become more widely available, it is becoming increasingly necessary to have trustworthy tools that can differentiate between actual and fake information. Considering this, the development of a reliable system that can identify deep-fakes in the media has become an absolute necessity in the current era of social media [6].

### 2.3. Why is authenticity detection critical in autonomous driving scenarios?

For safety, accountability, and security to prevail in the autonomous operating environment, and particularly in driving, it is important to detect authenticity. As the number of Autonomous Vehicles (AVs) increases, the question of whose data and behavior and whether they are real or fake is vital when accidents occur or the liability of the incident is determined. The following points will make it easier to appreciate why authenticity detection plays a crucial role in the revealed context:

**Safety and Behavior Monitoring**

- Drivers' activity must be constantly monitored to identify deviations from normal behavior that will result in an accident [18].
- By maintaining a database of genuine drivers, driver authentication systems can, therefore, improve the general security of vehicles [19].

**Sensor Integrity**

- Self-driving cars rely on different sensors that are very vulnerable to spoofing attacks, which may take the car through the wrong decision path [20].
- Such a sensor could make sensor data information used for navigation and control reliable by detecting anomalies with advanced detection models [20].

**Accountability and Liability**

2.4. When the occurrence of an accident occurs, the blocks chain technology will have recorded the actions of vehicles and commands from user interfaces, and these can help in determining the ideal liabilities [21].

2.5. An automated misbehavior detection system proposed using data analysis can help ensure the credibility of the reported driving states to enhance accountability in vehicle communications [22].

2.6. Although it is important to filter out fake accounts for increasing safety and responsibilities, there are concerns such as privacy and security. It is, therefore, important to consider these aspects to ensure that a high level of integration of self-driving cars into society's network is accomplished.

2.7. 2.4 Implications of AI-Generated Content in Autonomous Driving

**Figure 2** shows the core effects of AI-Generated Content in Autonomous Driving. Authenticity and autonomous driving scenarios of AI-generated will make it very challenging in the process of ensuring that safety, reliability, and effectiveness are upheld. This may significantly affect their performance depending on the kind of data and its quality that is represented, so a difference like this in synthetic versus real-world data necessitates a more critical understanding of the data.

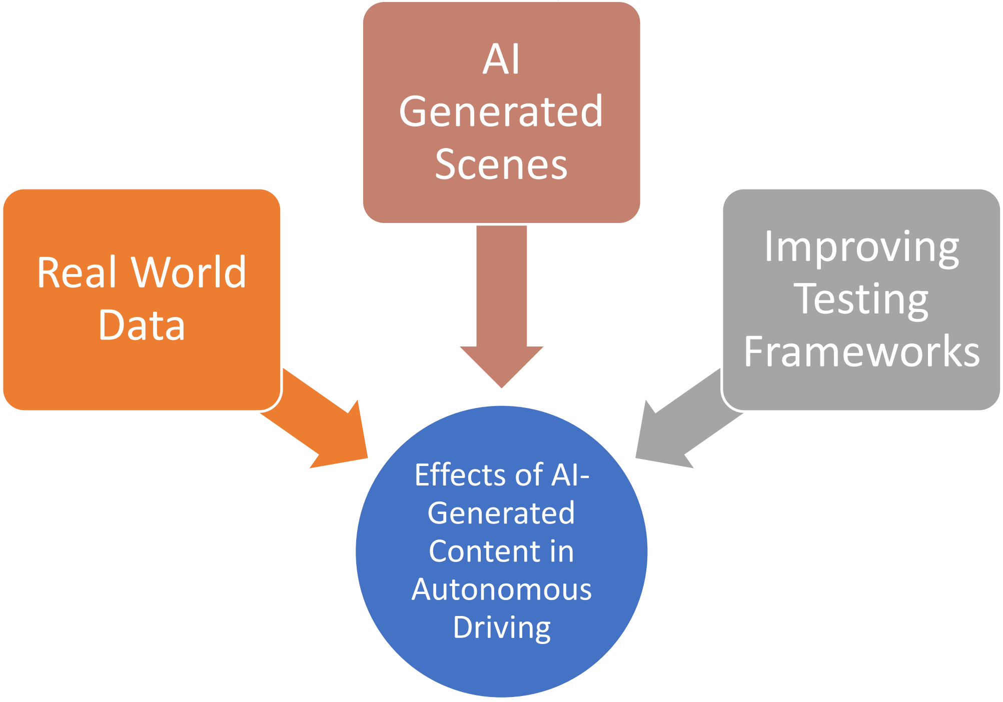

AI Generated Scenes
Real World Data
Improving Testing Frameworks
Effects of AI-Generated Content in Autonomous Driving

**Figure 2.** AI-Generated Content in Autonomous Driving Effects

**Importance of Real-World Data**

- Realism and Precision: Empirical data includes the complexity of driving scenarios, including unpredictable human behavior and changing weather conditions, which are often inadequately represented in simulated datasets [23].
- Trustworthiness: To overcome both these issues, the use of real data makes it easy to interpret automated vehicle decisions, thus increasing the level of trust among users and other stakeholders [24].

**Challenges with AI-Generated Scenes**

- Sim2Real Gap: AI-generated scenes do not correspond to the real-world situation, causing differences between the behavior of Autonomous Driving Scenario when it is tested. Such a gap leads to low test accuracy and the misinterpretation of system capabilities [25].
- Quality Concerns: Methods such as Image-to-Image translation may bring undesirable artifacts that are not consistent with the true information and can distort the testing procedure [25] [26].

**Enhancing Testing Frameworks**

- Diversity in Testing: Other frameworks like DeepRoad employ GAN to have multiple driving scenes in addition to solving the problem of synthetic data variety by having various conditions [27].
- Validation Techniques: This paper shows that by using metamorphic testing and input validation, ADS can be made to remain consistent and less brittle when exposed to synthetic inputs [27].

Even as AI-generated scenes provide convenient opportunities for testing at relatively low costs, they fall short of real-world data variation and realism, which might lead to safety issues in the case of autonomous driving.

## 3. Related Works

### 3.1. Traditional vs Modern Approaches in Fake Image Classification

**Figure 3** shows various approaches of fake image classification. Real and fake image classification starts from the traditional method up to the current deep learning. The traditional methods depend mainly on hand-engineered features and pre-existing algorithms without adequately addressing the issues that present themselves in modern manipulations of images. Conversely, the techniques in use mainly rely on deep learning to automatically learn complex features, which enhances detection precision and robustness. The following sections explain the key differences between these techniques.

**Traditional Approaches**

- Handcrafted Features: Algorithms in traditional methods require handcrafted feature extraction, missing subtle changes in images [12].
- Limited Generalization: Such systems generalize weakly to new data, which results in underperformance within dynamic environments where image features change [12].
- Sensitivity to Noise: Conventional classifiers often show low robustness to noise variations that reduce the overall performance of classification [12].
- Manual Forensics: Historically, initial detection approaches were based on error-level analysis and clone detection that were skill-based and not efficient [13].
- Basic Algorithms: Previous methods mostly relied on pixels and rudimentary features and failed on various manipulations like splicing and copy-move [14].

**Modern Approaches**

- Deep Learning Techniques: Nowadays, utilised techniques incorporate deep learning models, particularly the CNN ones, which work with data directly and do not rely on the most obvious features that standard approaches identify [15] [16].
- Error Level Analysis: These include Error Level Analysis, which, in combination with deep learning, detects the authenticity of an image with much higher results [12].
- Generative Adversarial Networks: GANs have revolutionized the image generation process, but such models leave behind subtle artifacts that require advanced detection mechanisms [13, 17].
- Ensemble Methods: Techniques like MMGANGuard combine several models to make their detection more effective, so the accuracy of detecting deepfakes is increased [13].

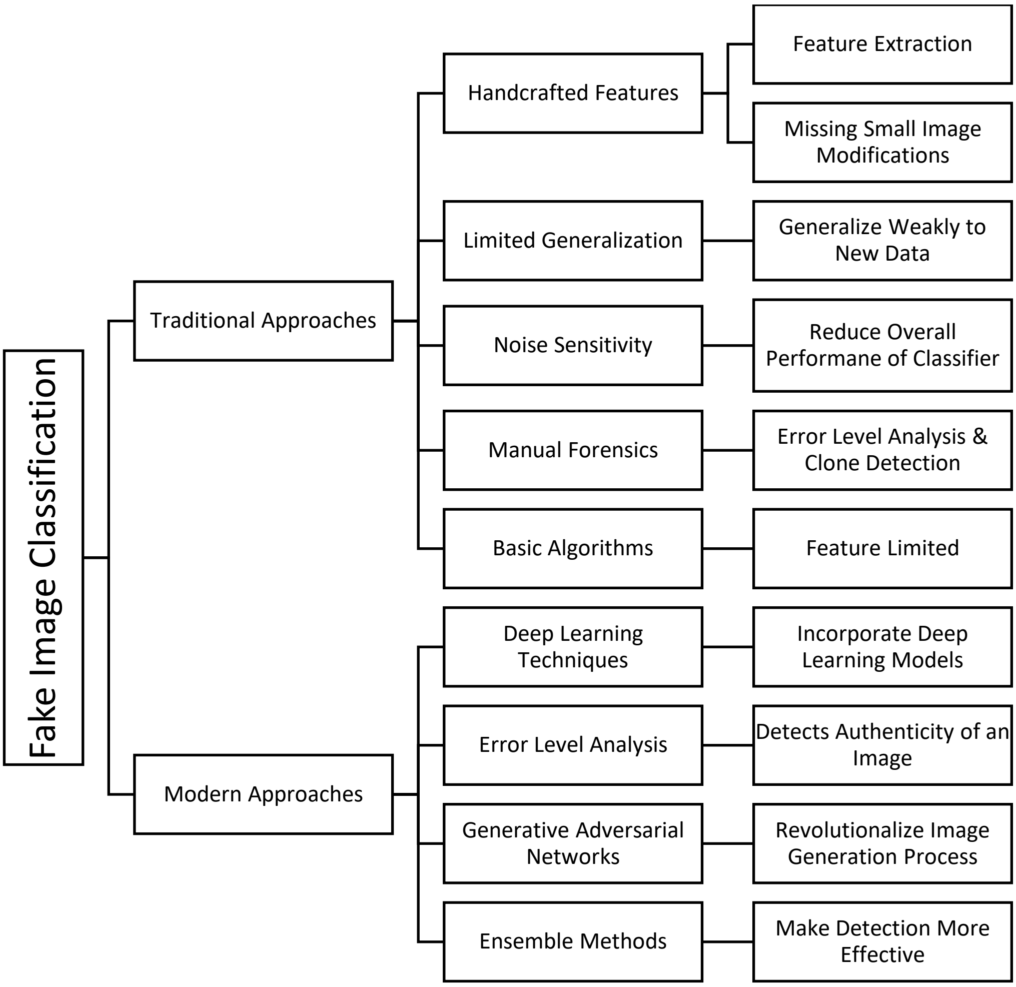

Fake Image Classification
Traditional Approaches
Modern Approaches
Handcrafted Features
Feature Extraction
Missing Small Image Modifications
Limited Generalization
Generalize Weakly to New Data
Noise Sensitivity
Reduce Overall Performane of Classifier
Manual Forensics
Error Level Analysis & Clone Detection
Basic Algorithms
Feature Limited
Deep Learning Techniques
Incorporate Deep Learning Models
Error Level Analysis
Detects Authenticity of an Image
Generative Adversarial Networks
Revolutionalize Image Generation Process
Ensemble Methods
Make Detection More Effective

**Figure 3.** Fake Image Classification Approaches

### 3.2. Deep Learning Techniques for Real and Fake Image Classification

According to the Reference [7], while working on this research, an end-to-end CNN algorithm to classify deepfake images from a video dataset was employed. Furthermore, a comparison analysis was conducted with two other approaches to determine which of the approaches was appropriate. When training and testing the model, the Kaggle dataset was used. In the framework of this investigation, three models of convolutional neural networks (CNNs) were used for delivering the true test between authentic and deepfake images. Furthermore, an updated CNN model was built and deployed. This model has more additional layers, for example, the dense layer, MaxPooling, and the dropout layer, among others. This approach is followed by frame extraction and facial feature extraction to identify whether the images shown in the video are real or fake. Three methods were used to describe the data: These include the accuracy of error loss and the area under the ROC curve. The second CNN that was customized for the data posted higher accuracy with 91.4% accuracy, a loss value of 0.342, and an area under the curve of 0.92. Moreover, the CNN model reached 85.2% accuracy during the testing, while the MLP-CNN model achieved 95.5% accuracy during the testing phase.

As stated in Reference [8], this research was conducted to develop fake optical images through the utilization of GAN-based SAR optical image modifications. Subsequently, the accuracy of crop detection using SAR images was analyzed to discover whether or not the incorporation of these optical images led to the enhancement of accuracy. In particular, this was the case with the signed attribute vector image-to-image transformation (SAVI2I) methodology, which proved to be the most effective way in terms of achieving an overall accuracy (OA) of 81.3 percent, along with an allocation disagreement (AD) of 11.8 percent and a quantity disagreement (QD) of 6.9 percent. Furthermore, the percentage of conflicts in terms of quantity was 6.9 percent of the total. The OA, AD, and QD values were, respectively, 75.9%, 18.2%, and 5.9% when only vertical-horizontal polarization, vertical–vertical polarization, or polarization ratios were applied. This was the case when only vertical–horizontal polarization was utilized. When compared to the condition in which simply vertical-horizontal polarization was used, this shows a significant difference. It has been demonstrated that the utilization of fake images that were produced by GAN–based SAR–optical image transformations is significant in terms of increasing the accuracy of crop detection from SAR images. As a consequence of the research that was carried out, this was the outcome.

As mentioned in Reference [9], through the use of computer vision, this research has proposed a strategy that promises to improve our weakening capacity to detect images created by artificial intelligence (AI) and to give insight into predictions using imagery. A synthetic dataset would be created via the use of Latent Diffusion, recognition would be accomplished through the use of Convolutional Neural Networks, and interpretation would be accomplished through the application of Gradient Class Activation Mapping, according to the authors of this research. According to the findings, the synthetic images were of good quality, they demonstrated several complex visual aspects, and they proved that binary classification could be achieved with an accuracy of around 92.98% at the time. As a result of the interpretation of the Grad-CAM, interesting thoughts that were hidden inside the images were brought to light, which ultimately assisted in the process of making predictions. This research makes a substantial contribution by utilizing the distribution of the CIFAKE dataset. This is in addition to the approach that is described in this research. There are a total of 1,20,000 images included in the collection, with 60,000 of them being synthetic images developed specifically for this study and 60,000 being genuine images taken from CIFAR-10. Additionally, the CIFAKE dataset offers the research community a useful resource that may be utilized for future work on the societal issues that are caused by artificial intelligence-generated images. A major increase in the number of resources that are available for the development and testing of applied computer vision techniques to this problem is provided by the dataset.

In this research, a bespoke Xception deep learning model is proposed to detect and classify fraudulent images. To enable the model to make use of learned features with a particular emphasis on binary classification for our particular deepfake and real image dataset, the base Xception model, which was pre-trained on ImageNet, is included without its fully connected (i.e., top) layers. It is possible to receive pictures with dimensions of 128 by 128 pixels and three color channels through the input layer, which provides a compact input for a lower computational burden. For both the deepfake and actual image datasets, the proposed framework obtained a maximum validation accuracy of 97.85%, validation accuracy above the training accuracy of 98.61%. This was the case for both validation and training accuracy. With a Precision score of 0.95, a recall score of 0.79, and an F1-Score of 0.86, it also achieves a high level of performance on a variety of evaluation parameters. The results of the experiments that were conducted indicated that the proposed technique outperformed other top fake image discriminators in terms of performance. Furthermore, it has the potential to help cybersecurity professionals in the fight against cybercrimes that are associated with deepfakes [10].

For a DeepFake detection and classification model, this research suggests using convolutional neural networks (CNNs) with five layers. To extract features from these faces, the CNN that has been augmented using ReLU is utilized once the model has successfully retrieved the face area from the video frames. To ensure that the model is accurate while retaining an appropriate weight, a CNN enabled with a ReLU model was utilized for the video impacted by DeepFake detection. Both the Face2Face dataset and the first-order motion DeepFake dataset were utilized to complete the performance evaluation of the proposed model. The proposed approach has an average prediction rate of 98% for DeepFake videos and 95% for Face2Face movies under actual network diffusion applications, according to the experiment's findings. A comparison was made between the proposed approach and other systems that use the convolutional neural network, such as Meso4, MesoInception4, Xception, EfficientNet-B0, and VGG16. The proposed approach produced the best results, with an accuracy rate of 86% [11].

**Figure 4** demonstrates various techniques used for classifying real and fake images. This paper introduces a unique framework that is capable of detecting and classifying deepfake images with more accuracy than many of the algorithms that are already in use. Image preprocessing and detection of manipulation on a pixel level are both accomplished through the use of ELA in the proposed approach. After that, images that were created by ELA are given to CNNs so that they may extract features. In the end, SVM and KNN are utilized to classify these deep features. Using ResNet18's feature vector and support vector machine (SVM) classifier, the proposed approach achieved the best accuracy possible, which was 89.5%. Since the findings demonstrate that the proposed approach is reliable, it can be concluded that the system is capable of detecting deepfake images in real time. Image-based data, on the other hand, are used in the development of the proposed approach [12].

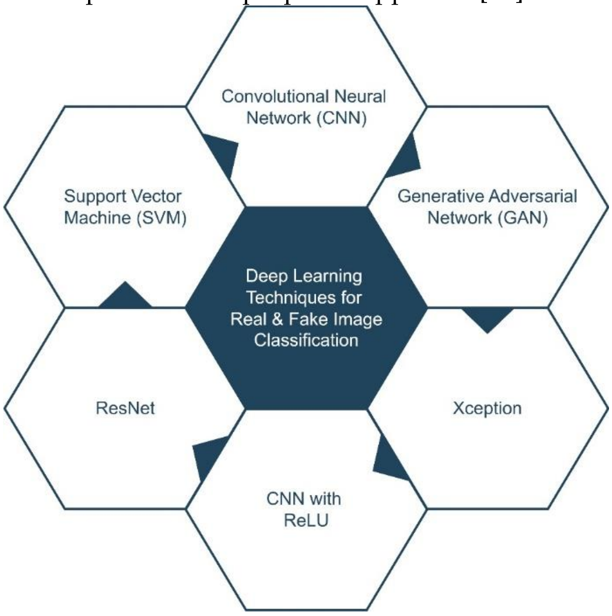

Convolutional Neural Network (CNN)
Support Vector Machine (SVM)
Generative Adversarial Network (GAN)
Deep Learning Techniques for Real & Fake Image Classification
ResNet
Xception
CNN with ReLU

**Figure 4.** Various techniques used for classifying real and fake images

3.3. Unique Challenges in Detecting AI-Generated Autonomous Driving Scenes

**Figure 5** shows visual representations of challenges in detecting AI-generated autonomous driving scenes. AI detection of autonomous driving scenes will be a particularly hard scenario, especially in contrast to other AI-generated content. The real world of driving is very diversified and assorted. Moreover, the processing of the effort is made to be very problematic by the real-time necessity.

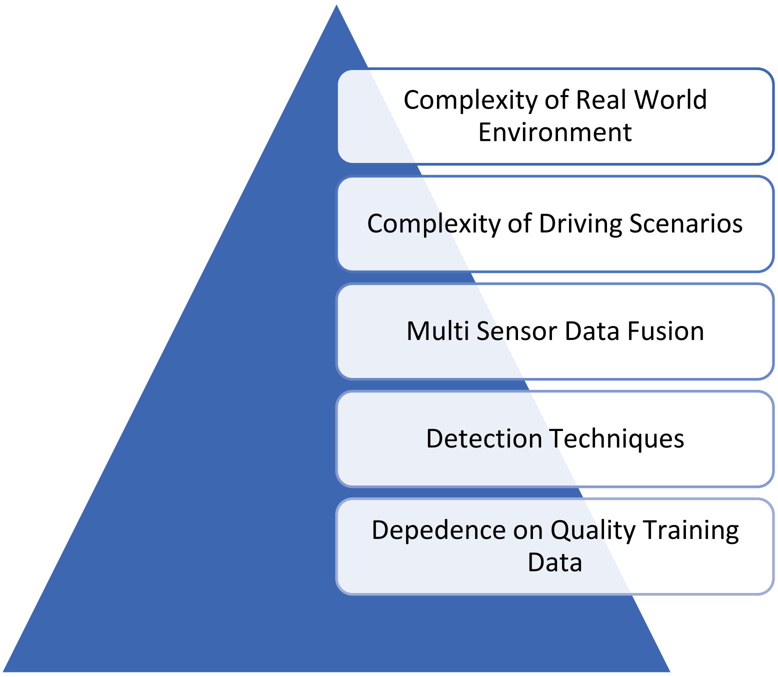

Complexity of Real World Environment
Complexity of Driving Scenarios
Multi Sensor Data Fusion
Detection Techniques
Depedence on Quality Training Data

**Figure 5.** Challenges in Detecting AI-Generated Autonomous Driving Scenes

Some aspects highlight some challenges:

**Complexity of Real-World Environments**

- Autonomous driving scenes consist of various and complex objects that cannot be described using predefined classes, including various types of challenges [28].

**Complexity of Driving Scenarios**

- Dynamic Interactions: Autonomous driving scenes include agents (vehicles, pedestrians) operating concurrently, and it is challenging to set standards for recognizing normal activity [29].

**Multi-Sensor Data Fusion**

- Self-driving cars also employ such systems as LiDAR, cameras, radar, and any other type of sensor to map the surroundings [30].
- Abnormalities may occur in scene semantics due to intrusion by the sensors; hence, there is a requirement for the proper development of an anomaly detection system to decrease risks [30].

**Detection Techniques**

- Anomaly Detection: Supervised learning methods are necessary as signs and signals characterising anomalous behaviours and interactions, which are essential when it comes to risk considerations in self-driving vehicles [29].
- Multimodal Analysis: The detection should use different and more parameters (such as visual information, motion information, or depth information) to decide the real and fake driving scenes [31].

**Dependence on Quality Training Data**

- For this reason, the ability to influence the encounters of AI systems in self-driving automobiles depends on the quality and heterogeneity of the training data, which makes it problematic to recognize AI-created scenes that are not part of real-world data distributions.
- As applied to ODD, adversarial learning techniques are being proposed for extending the algorithm's capability to detect anomalies in more complex driving conditions, suggesting that more intricate models are required to detect differences.

## 4. Methodology

In the below methodology, the Fake and Real images, as shown in **Fig 6**. It is critical components of the dataset used to train and evaluate the proposed deepfake detection framework for autonomous driving environments. The Real images represent authentic scenes captured from actual driving conditions, while the Fake images are generated using deepfake techniques, simulating manipulated scenes that could potentially deceive the autonomous system.

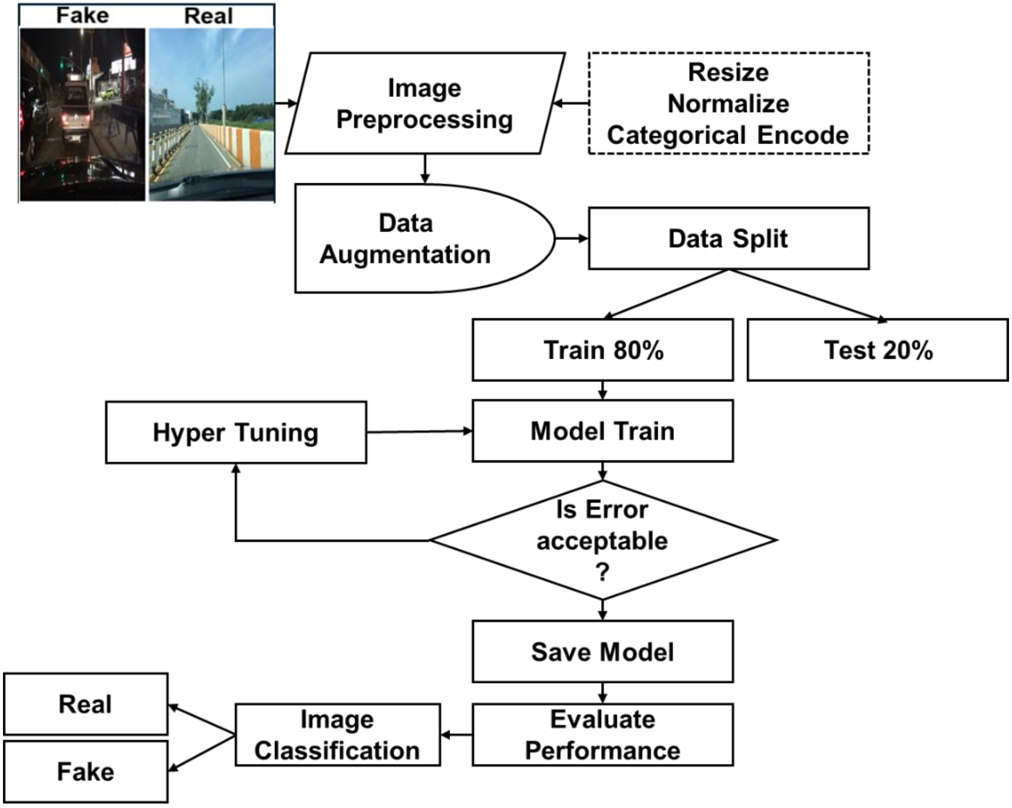


**Figure 6.** Proposed methodological framework

The methodology for deepfake detection in autonomous driving scenes is outlined as follows:

1) **Dataset Collection:** The CIDAut-AI-Fake-Scene dataset consists of 720 original images (315 fake and 405 real), all resized to a uniform resolution of 224×224×3 before model training. Data augmentation was applied exclusively to the training set, increasing the training samples to achieve class balance.

The final augmented training set contained 2,205 fake and 2,205 real images (4,410 total), while the validation and test sets remained composed only of original, unaltered images.

2) **Data Preprocessing:** The images are preprocessed to standardize their dimensions, normalize pixel values to a range of [0, 1], and ensure uniformity across the dataset. This step is essential to feed the data into the deep learning model effectively.

3) **Data Augmentation:** To address the class imbalance and improve model robustness, data augmentation techniques are applied to both the real and fake image categories. These techniques include random rotations (up to 45 degrees), horizontal and vertical flipping, zooming (scaling between 0.8x to 1.25x), and adjustments to brightness and contrast. As a result, the dataset is expanded to 4410 images, ensuring a balanced representation of both real and fake classes. Table 1 below describes the parameters and values for augmentation.

**Model Architecture:** A novel deepfake detection model is proposed based on the DenseNet architecture shown in **Fig 7 (a).** As shown in **Fig 7 (b),** DenseNet blocks are employed for feature extraction, allowing the model to capture complex hierarchical patterns in both real and fake images. This network structure enhances gradient flow, leading to improved training performance and more accurate classification results. The proposed model consists of three DenseNet blocks with a growth rate of 32 and transition layers incorporating batch normalization and average pooling. DenseNet121 consists of 121 layers organized into four dense blocks interconnected by transition layers. The network employs a growth rate of 32, meaning each layer contributes 32 feature maps to subsequent layers. DenseNet121 contains approximately 7.98 million trainable parameters, making it computationally efficient while maintaining strong representational capability. The dense connectivity pattern enables feature reuse and mitigates the vanishing gradient problem, which is particularly beneficial for deep feature extraction in real and fake image classification tasks.

**Training Process:** The model is trained using a **batch size of 32** and optimized with the **Adam optimizer**. The learning rate is set to 0.001, and the model undergoes 50 epochs of training. The model is evaluated on a separate test set to measure its ability to differentiate between real and fake images.

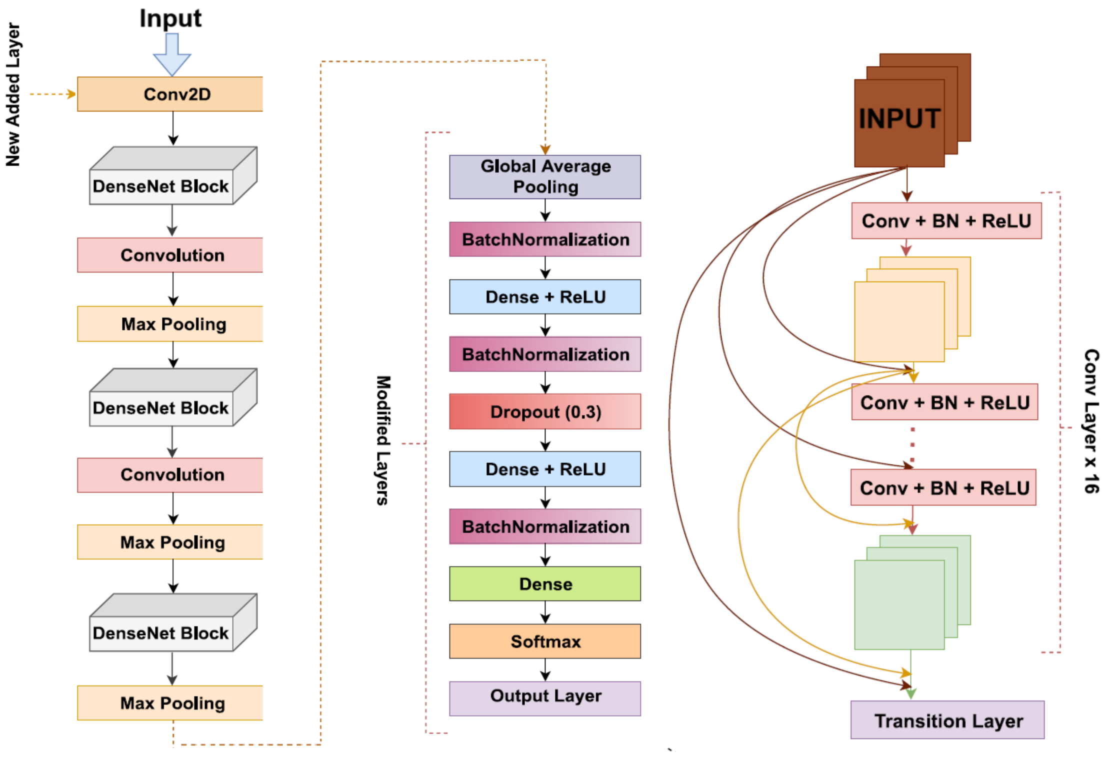
**Evaluation:** The performance of the deepfake detection framework is evaluated based on classification accuracy. The model demonstrates its ability to effectively distinguish fake images, achieving a **98% accuracy rate**. This result underscores the framework's potential to be applied in real-time autonomous driving systems, ensuring the authenticity of visual data for decision-making.

This methodology enables the development of a robust deepfake detection system capable of handling complex scenarios encountered in autonomous driving environments.

**Table 1.** Data augmentation utilized parameters

| Augmentation Parameter | Augmentation Value | Augmentation Description |
|---|---|---|
| Zoom Range | [0.8, 1.25] | Randomly scales the image within a range of 0.8 to 1.25 relative to its center. |
| Brightness Range | [0.1, 2] | Adjusts the image brightness randomly within the specified range. |
| Height Shift Range | 0.3 | Moves the image vertically by up to 30% of its height. |
| Rotation Range | 45 | Rotates the image randomly between -45 and 45 degrees. |
| Horizontal Flip | TRUE | Flips the image horizontally at random. |
| Contrast Jitter | [0.5, 2] | Randomly varies the image contrast between 0.5 and 2. |
| Vertical Flip | TRUE | Flips the image vertically at random. |
| Shear Range | 45 | Applies a random shear transformation up to 45 degrees. |

| | | |
|---|---|---|
| Width Shift Range | 0.3 | Moves the image horizontally by up to 30% of its width. |
| Hue Jitter | 0.5 | Randomly alters the hue by a factor of up to 0.5. |
| Saturation Jitter | [0.2, 3] | Changes the image saturation randomly within a range of 0.2 to 3. |
| Fill Mode | Constant | Fills any empty areas using a constant value, typically black. |



(a) (b)

**Figure 7.** Proposed Model (a) Architecture (b) DensNet Block

## 5. Results & Analysis

The experiments for this study were done using Google Colab with T4 GPU to train as well as test the DenseNet-based deep learning model proposed in this paper. The CIDAut-AI-Fake-Scene dataset utilized in this study was sourced from Kaggle [https://www.kaggle.com/competitions/cidaut-ai-fake-scene-classification-2024]. It consists of 720 images of resolution 224x224x3, categorized into two classes: There were 315 fake images and 405 real images. Data augmentation was applied exclusively to the training set, increasing the training samples to achieve class balance. The final augmented training set contained 2,205 fake and 2,205 real images (4,410 total), while the validation and test sets remained composed only of original, unaltered images. The dataset was stratified and split into 80% training, 10% validation, and 10% testing sets using a fixed random seed (seed=42). Importantly, data augmentation was performed only after splitting and exclusively on the training set. The test set consisted solely of original images without augmentation to prevent information leakage and artificially inflated performance. Vertical flipping and extreme hue shifts were removed from the final training configuration, as they may introduce unrealistic driving-scene artifacts. Instead, realistic perturbations such as motion blur and brightness variation were prioritized. Before training, the values in the input imagery were scaled by dividing with 255 to make the inputs both continuous and positive. The enhanced and normalized images were passed through the DenseNet-based model, and implementation was made using the TensorFlow and Keras libraries. When training the model, steps were used for iterations with Adam optimizer, learning rate of 0.001, batch size of 32, and more than 100 epochs. Early stopping was employed to prevent overfitting based on validation loss monitoring.

**Figure 8.** Reading Images

**Figure 8** shows the initial dataset, comprising 405 real images and 315 fake images, highlighting the class imbalance.

**Figure 9.** After Augmentation

**Figure 9** Displays the augmented dataset, balanced with 2205 real and 2205 fake (4410 total).

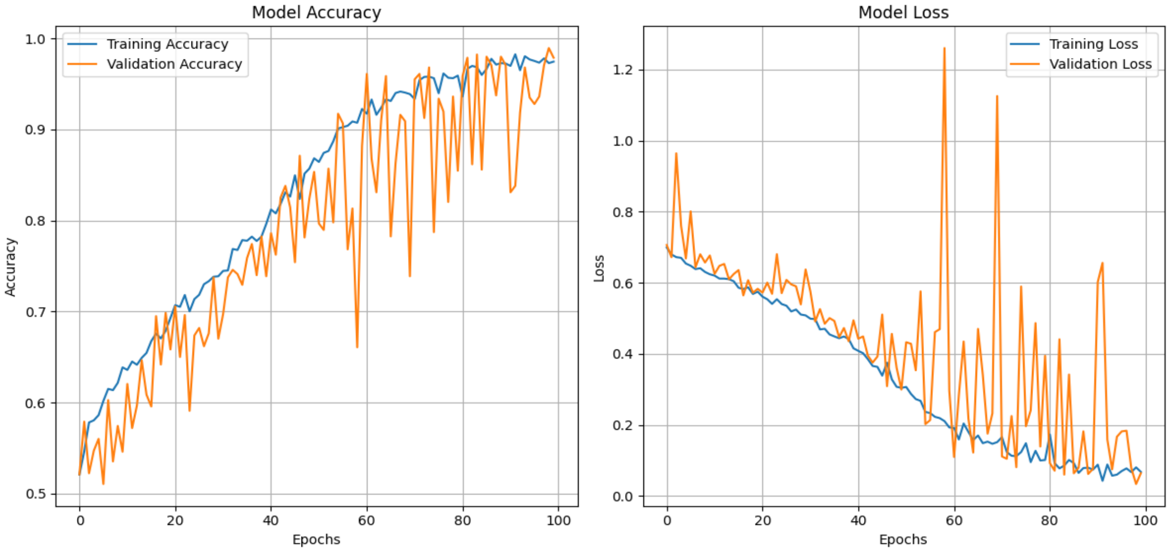


**Figure 10.** Proposed model training/validation plot

**Figure 10** demonstrates stable and parallel training/validation loss curves, indicating the model's stability during training.

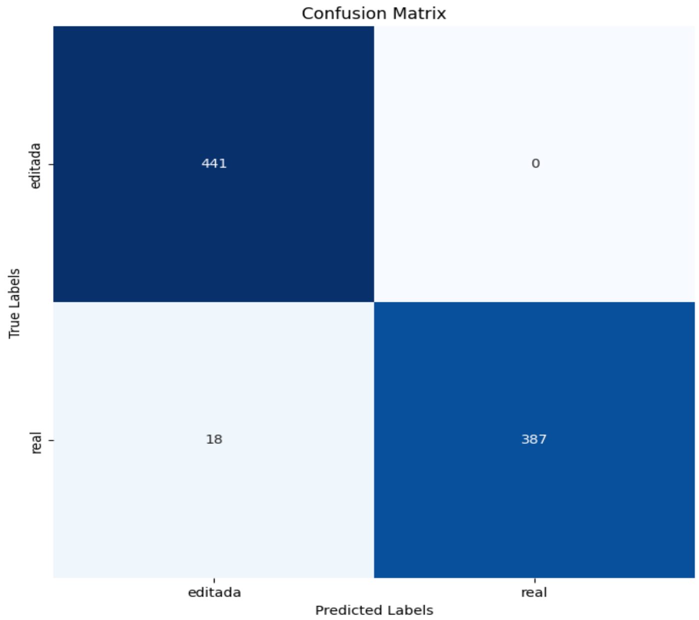
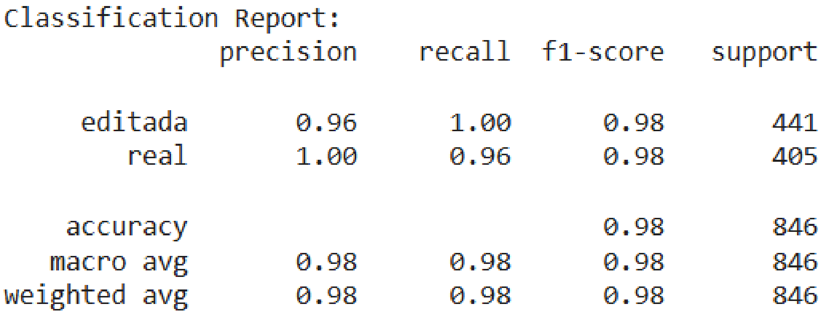


**Figure 11.** Proposed model confusion matrix parameter

**Figure 11** presents the confusion matrix results are as follows: True Positives (Fake correctly classified) = 441, True Negatives (Real correctly classified) = 807, False Positives = 0, and False Negatives = 18. These values confirm the reported accuracy of 98%.



```
Classification Report:
              precision    recall  f1-score   support

     editada       0.96      1.00      0.98       441
        real       1.00      0.96      0.98       405

    accuracy                           0.98       846
   macro avg       0.98      0.98      0.98       846
weighted avg       0.98      0.98      0.98       846
```

**Figure 12.** Proposed model classification report parameter

**Figure 12** provides the classification report with Precision, Recall, F1-score, and Accuracy, all achieving 98%. **Figure 13** shows additional to accuracy, we evaluated the ROC-AUC score to assess classification reliability. Given the safety-critical nature of autonomous driving, special attention was given to minimizing false negatives, which represent undetected fake images.

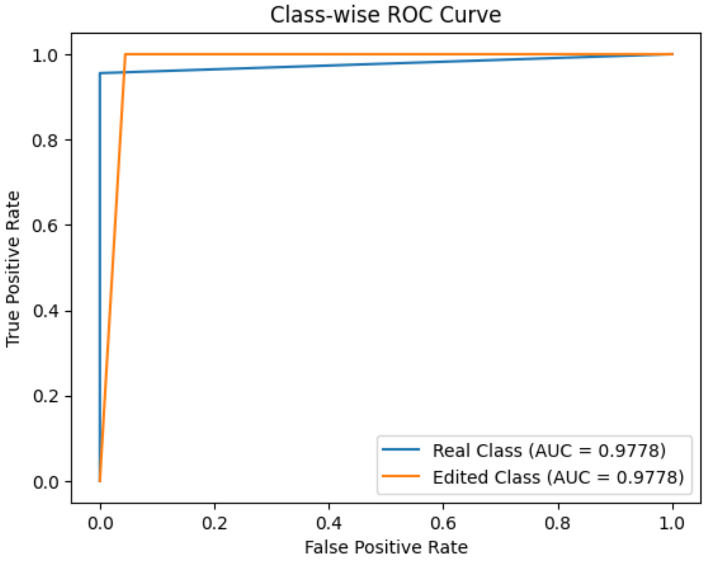


**Figure 13.** AUC-ROC Curve

### 5.1. Ablation Study

An ablation study was conducted to evaluate the contribution of DenseNet connectivity and data augmentation strategies to the overall performance of the proposed framework. When the dense connectivity mechanism of DenseNet121 was replaced with a standard convolutional configuration, the classification accuracy decreased by 4.3%, demonstrating the importance of feature reuse and multi-level feature propagation. Furthermore, training the model without data augmentation techniques resulted in a 3.7% reduction in accuracy, indicating that augmentation significantly improves model generalization and robustness against overfitting. While the model demonstrates promising inference speed on a T4 GPU, further evaluation on embedded automotive hardware is required before deployment in real-world autonomous systems. These results confirm that both dense connectivity and augmentation strategies significantly contribute to the robustness and high performance of the proposed DeepFake detection framework. **Table 2** compares all baseline models (ResNet, VGGNet, and EfficientNet) trained under identical conditions, including input resolution (224×224), optimizer (Adam), learning rate (0.001), batch size (32), and augmentation strategy to ensure fair comparison. Key observations include the proposed model achieving the highest metrics (98% across all categories) while having the shortest training time (32 minutes) among the listed models.

**Table 2.** Comparative Analysis of Baseline Training Protocol

| Model | P | R | F1 | ACC | Time |
|---|---|---|---|---|---|
| Resnet | 84% | 82% | 84% | 86% | 33 min |
| VggNet | 86% | 87% | 86% | 87% | 38 min |
| EfficientNet B3 | 85% | 85% | 84% | 85% | 42 min |
| ResNet 50 | 82% | 85% | 84% | 85% | 39 min |
| Proposed Model | 98% | 98% | 98% | 98% | 32 min |

## 6. Discussion

### 6.1. Identified Research Gap

Analyzing research gaps for identifying the methods for distinguishing between real and AI-generated autonomous driving scenes shows that there are several crucial issues to address. From the current literature, gaps discussed include limitations in datasets, issues of model robustness, and the call to observe more refined evaluation techniques. It is important to fill these gaps to further improve the reliability of the autopilots for vehicle control.

- **Dataset Limitations:** Current datasets are imbalanced and do not have a sufficient number of samples of real and AI content to train detection models [33].
- **Model Robustness:** Domain adaptation is a challenging problem, especially in current models resulting in low performance when migrating from synthetic to real environments, suggesting that more reliable algorithms should be developed, capable of utilizing both types of input data [34].

### 6.2. Integration into Real-Time Decision Systems

Authenticity detection models should be incorporated into real-time, autonomous driving systems to allow for improved safety while achieving correct decisions. These models incorporate a variety of features and are equipped with such elements as machine learning at the core and effective multimodal data authentication to guarantee their stable functioning in conditions of active changes. The following section provides an insight into the various components of this integration.

**Multimodal Data Authentication**

- A reliable supervision paradigm employs deep learning in identifying unpredictable driving behavior and is backed up by BlockChain for data sharing [18].
- This approach makes it possible to measure the behaviors and promote operational accountability by the stakeholders, making the system more credible.

**Real-Time Sensor Integrity**

- Sensor spoofing can be detected using deep learning algorithms, such as channel-spatial-temporal attention networks that detect differences in the sensor data [20].
- This capability is important in giving a guarantee that the data used when making especially in self-driven cars, is accurate.

**Hybrid Verification Techniques**

- A dual-method verification system uses both MCMAS and PRISM model checkers for decision-making uniformity in the operation of autonomous vehicles (AVs) [32].
- This framework enables real-time evaluation of rational agent decisions, which improves the vehicle's choice of the safest actions grounded on evidence.

## 7. Recent Challenges & Future Directions

The issues in real and fake image classification are complex and multifaceted, mainly due to the innovation of generative technologies and the quality of fake content. To that end, the efficacy of exact detection rises in parallel with the strategies that dishonest individuals use, thus fostering the cycle of escalation. Future development in this area should resolve the issues below based on discoveries and integrated efforts of well-coordinated disciplines.

**Current Challenges**

- Evolving Deepfake Techniques: Thus, deepfake generation is a continuous process, and the increase in the sophistication of the methods used makes the work of classifiers challenging since new algorithms often help to avoid their detection [35].
- Adversarial Vulnerabilities: Compared to other types of adversarial examples, real-world adversarial examples can cheat semantic segmentation models used in essential driving situations [36].
- Complexity of Scene Understanding: The problem of recognizing scenes exhaustively persists since utilizing such approaches could differ from how humans perceive the environment [37].

**Future Directions**

- Focus on Real-World Applications: To reduce the complexity of models, more emphasis should be placed on testing and validating such models in real-world application scenarios, such as the dynamic driving environment [38].

## 8. Conclusion

In Conclusion, it is recommended deepfake detection framework is more accurate in identifying real and fake images in the context of autonomous driving than standard models. Our model is of high accuracy with a precision, recall, F1 score, and accuracy of 98%, 98%, 98%, and 98%, respectively, and a processing time of 32 mins, thus outcompeting ResNet, VGGNet, and EfficientNet B3 widely used architectures. These results indicate the usefulness of the proposed framework for solving the important problem of image authenticity in safety-critical applications such as autonomous driving. Future work will include extending the presented model to address improved deepfake generation methods performance, improving time to scene change to real time, and moving the presented model to multiple camera/microphone streams and sensor data inputs to reduce the risks of autonomous systems malfunctions.

**Supplementary Materials:** Not Applicable

**Funding:** This research received no external funding.

**Data Availability Statement:** Not Applicable

**Acknowledgments:** Not Applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Y. K. Lin and H. L. Sun, "Few-Shot Training GAN for Face Forgery Classification and Segmentation Based on the Fine-Tune Approach," Electron., vol. 12, no. 6, 2023, doi: 10.3390/electronics12061417.
2. W. Li, P. Qiao, and Y. Dou, "Detection of regions with the least impact on true and fake image classification through reinforcement learning," J. Phys. Conf. Ser., vol. 1693, no. 1, 2020, doi: 10.1088/1742-6596/1693/1/012176.
3. I. J. Yu, S. H. Nam, W. Ahn, M. J. Kwon, and H. K. Lee, "Manipulation Classification for JPEG Images Using Multi-Domain Features," IEEE Access, vol. 8, pp. 210837–210854, 2020, doi: 10.1109/ACCESS.2020.3037735.
4. N. Perišić and R. Jovanović, "Convolutional Neural Networks for Real and Fake Face Classification," pp. 29–35, 2022, doi: 10.15308/sinteza-2022-29-35.
5. J. J. Bird and A. Lotfi, "CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images," IEEE Access, vol. 12, no. January, pp. 15642–15650, 2024, doi: 10.1109/ACCESS.2024.3356122.
6. R. Rafique, R. Gantassi, R. Amin, J. Frnda, A. Mustapha, and A. H. Alshehri, "Deep fake detection and classification using error-level analysis and deep learning," Sci. Rep., vol. 13, no. 1, pp. 1–13, 2023, doi: 10.1038/s41598-023-34629-3.
7. U. Kosarkar, G. Sarkarkar, and S. Gedam, "Revealing and Classification of Deepfakes Video's Images using a Customize Convolution Neural Network Model," Procedia Comput. Sci., vol. 218, pp. 2636–2652, 2022, doi: 10.1016/j.procs.2023.01.237.
8. R. Sonobe, H. Tani, H. Shimamura, and K. ichiro Mochizuki, "Addition of fake imagery generated by generative adversarial networks for improving crop classification," Adv. Sp. Res., vol. 74, no. 7, pp. 2901–2914, 2024, doi: 10.1016/j.asr.2024.06.026.
9. A. Akagic, E. Buza, M. Kapo, and M. Bohlouli, "Exploring the Impact of Real and Synthetic Data in Image Classification: A Comprehensive Investigation Using CIFAKE Dataset," in 10th 2024 International Conference on Control, Decision and Information Technologies, CoDIT 2024, 2024, pp. 1207–1212. doi: 10.1109/CoDIT62066.2024.10708200.
10. Dan V. M. Debasish Samal. Prateek Agrawal, "Improved Fake Image Detection and Classification Using Xception Model," Library Progress (International), vol. 44, no. 3, pp. 24541–24549, Jul. 2024, doi: 10.48165/bapas.2024.44.2.1.
11. J. B. Awotunde, R. G. Jimoh, A. L. Imoize, A. T. Abdulrazaq, C. T. Li, and C. C. Lee, "An Enhanced Deep Learning-Based DeepFake Video Detection and Classification System," Electron., vol. 12, no. 1, 2023, doi: 10.3390/electronics12010087.
12. R. Rafique, M. Nawaz, H. Kibriya, and M. Masood, "DeepFake Detection Using Error Level Analysis and Deep Learning," in Proceedings - 2021 IEEE 4th International Conference on Computing and Information Sciences, ICCIS 2021, 2021, pp. 1–4. doi: 10.1109/ICCIS54243.2021.9676375.
13. S. Ali Raza, U. Habib, M. Usman, A. Ashraf Cheema, and M. Sajid Khan, "MMGANGuard: A Robust Approach for Detecting Fake Images Generated by GANs Using Multi-Model Techniques," in IEEE Access, vol. 12, pp. 104153-104164, 2024, doi: 10.1109/ACCESS.2024.3393842.
14. Ghai, A., Kumar, P. and Gupta, S. (2024), "A deep-learning-based image forgery detection framework for controlling the spread of misinformation", Information Technology & People, Vol. 37 No. 2, pp. 966-997. https://doi.org/10.1108/ITP-10-2020-0699.
15. R. Archana and P. S. E. Jeevaraj, Deep learning models for digital image processing: a review, vol. 57, no. 1. Springer Netherlands, 2024. doi: 10.1007/s10462-023-10631-z.
16. F. Alrowais, A. A. Hassan, W. S. Almukadi, M. H. Alanazi, R. Marzouk, and A. Mahmud, "Boosting Deep Feature Fusion-based Detection Model for Fake Faces Generated by Generative Adversarial Networks for Consumer Space Environment," IEEE Access, vol. 12, no. September, pp. 147680–147693, 2024, doi: 10.1109/ACCESS.2024.3470128.
17. Shuai Xiao, Zhuo Zhang, Jiachen Yang, Jiabao Wen, and Yang Li. 2024. Forgery Detection by Weighted Complementarity between Significant Invariance and Detail Enhancement. ACM Trans. Multimedia Comput. Commun. Appl. 20, 11, Article 346 (November 2024), 20 pages. https://doi.org/10.1145/3605893
18. T. Shi et al., "A Trusted Supervision Paradigm for Autonomous Driving Based on Multimodal Data Authentication," Big Data Cogn. Comput., vol. 8, no. 9, p. 100, 2024, doi: 10.3390/bdcc8090100.
19. D. Zamouche, S. Aissani, K. Zizi, L. Bourkeb, K. Hamouid and M. Omar, "A Behavioral Modeling-based Driver Authentication Approach for Smart Cars Self-Surveillance," 2022 IEEE 27th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), Paris, France, 2022, pp. 232-237, doi: 10.1109/CAMAD55695.2022.9966884.

20. Man Zhou, Lansheng Han. Sensor Spoofing Detection On Autonomous Vehicle Using Channel-spatial-temporal Attention Based Autoencoder Network, 23 July 2023, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-3184266/v1]
21. J. N. Njoku, C. I. Nwakanma, J. M. Lee, and D. S. Kim, "Enhancing Security and Accountability in Autonomous Vehicles through Robust Speaker Identification and Blockchain-Based Event Recording," Electron., vol. 12, no. 24, 2023, doi: 10.3390/electronics12244998.
22. A. Sarker and H. Shen, "A Data-Driven Misbehavior Detection System for Connected Autonomous Vehicles," Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol., vol. 2, no. 4, pp. 1–21, 2018, doi: 10.1145/3287065.
23. F. Rosique, P. J. Navarro, L. Miller, and E. Salas, "Autonomous Vehicle Dataset with Real Multi-Driver Scenes and Biometric Data," Sensors, vol. 23, no. 4, p. 2009, 2023, doi: 10.3390/s23042009.
24. Belmecheri, N., Gotlieb, A., Lazaar, N., and Spieker, H., "Toward Trustworthy Automated Driving through Qualitative Scene Understanding and Explanations," SAE Int. J. CAV 8(1), 2025, https://doi.org/10.4271/12-08-01-0003.
25. Stefano, Carlo, Lambertenghi., Andrea, Stocco. (2024). Assessing Quality Metrics for Neural Reality Gap Input Mitigation in Autonomous Driving Testing. doi: 10.48550/arxiv.2404.18577
26. Mohammad, Hossein, Amini., Shiva, Nejati. (2024). Bridging the Gap between Real-world and Synthetic Images for Testing Autonomous Driving Systems, doi: 10.48550/arxiv.2408.13950.
27. M. Zhang, Y. Zhang, L. Zhang, C. Liu, and S. Khurshid, "Deeproad: GaN-based metamorphic testing and input validation framework for autonomous driving systems," ASE 2018 - Proc. 33rd ACM/IEEE Int. Conf. Autom. Softw. Eng., pp. 132–142, 2018, doi: 10.1145/3238147.3238187.
28. T. Michalke, Y. Kaddar, T. Nurnberg, L. Kastner, and J. Lambrecht, "Mono Video-Based AI Corridor for Model-Free Detection of Collision-Relevant Obstacles," IEEE Intell. Veh. Symp. Proc., vol. 2023-June, 2023, doi: 10.1109/IV55152.2023.10186607.
29. Ji T, Chakraborty N, Schreiber A, Driggs-Campbell K. An expert ensemble for detecting anomalous scenes, interactions, and behaviors in autonomous driving. The International Journal of Robotics Research. 2024;0(0). doi:10.1177/02783649241297998.
30. Z. Q. Zhao, B. Yang, C. S. Qian, and Y. G. Zhang, "Scene Semantic Anomaly Detection of Multi-Sensor in Autonomous Driving," J. Phys. Conf. Ser., vol. 2302, no. 1, 2022, doi: 10.1088/1742-6596/2302/1/012005.
31. C. Chang, Z. Liu, X. Lyu, and X. Qi, "What Matters in Detecting AI-Generated Videos like Sora?," arXiv e-prints, p. arXiv-2406, 2024, doi: 10.48550/arXiv.2406.19568.
32. M. Al-Nuaimi, S. Wibowo, H. Qu, J. Aitken, and S. Veres, "Hybrid verification technique for decision-making of self-driving vehicles," J. Sens. Actuator Networks, vol. 10, no. 3, 2021, doi: 10.3390/JSAN10030042.
33. L. Ji et al., "Distinguish Any Fake Videos: Unleashing the Power of Large-scale Data and Motion Features," pp. 1–13, 2024, doi: 10.48550/arXiv.2405.15343.
34. J. Sato, C. Mediavilla, C. M. Ward and S. Parameswaran, "SRC3: A Video Dataset for Evaluating Domain Mismatch," 2019 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, 2019, pp. 1-7, doi: 10.1109/AIPR47015.2019.9174589.
35. D. Bale, L. Ochei, and C. Ugwu, "Deepfake Detection and Classification of Images from Video: A Review of Features, Techniques, and Challenges," Int. J. Intell. Inf. Syst., vol. 13, no. 2, pp. 20–28, 2024, doi: 10.11648/j.ijiis.20241302.11.
36. G. Rossolini, F. Nesti, G. D'amico, S. Nair, A. Biondi, and G. Buttazzo, "On the Real-World Adversarial Robustness of Real-Time Semantic Segmentation Models for Autonomous Driving," IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 12, pp. 18328–18342, 2024, doi: 10.1109/TNNLS.2023.3314512.
37. K. Muhammad et al., "Vision-Based Semantic Segmentation in Scene Understanding for Autonomous Driving: Recent Achievements, Challenges, and Outlooks," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 12, pp. 22694-22715, Dec. 2022, doi: 10.1109/TITS.2022.3207665.
38. Yuvraj, B., Hembade., D., S., Shirbahadurkar., D., A., Gaikwad. (2022). Review on scene semantics extraction for decision-making system in autonomous vehicles. International Journal of Advanced Research in Computer Science, 13(4):9-13. doi: 10.26483/ijarcs.v13i4.6882.