International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-15 Issue-1, October 2025

Detection of Authenticity of Images

Virginia Peter Gonsalves, Tahira H. Shaikh, Muskan A. Pathan, Plasin Francis Dias

***Abstract**: The proliferation of AI-generated images, enabled by deep learning algorithms, has become a matter of concern about fake information, media manipulation, and the rapid decline in trust in visual content. In this study, we focus our efforts on developing a system that ensures the credibility of digital media through image verification. A custom Convolutional Neural Network (CNN) designed specifically for authentication detection, trained on a dataset comprising 5,392 authentic images and AI-generated images, was employed. The dataset was divided into three parts, including training (3,964), validation (714), and test (714). Data augmentation techniques were used to preprocess the dataset, which included rotation, flipping, and brightness adjustments, thereby creating a more versatile representation of various images. We also obtained the CNN architecture by using four convolutional blocks, which included batch normalisation, max pooling, and dropout layers, thereby preventing overfitting. Next, we followed these blocks with some dense layers that were correctly applied for binary classification. The model was successfully trained for 50 epochs using the Adam optimiser (with a learning rate of 0.0001) and binary cross-entropy loss. It also included callbacks for early stopping, model checkpointing, and reducing the learning rate. Interestingly, we successfully classified 93.56% of the test set as authentic, achieving a precision rate of 95.87%, a recall rate of 91.04%, and an AUC-ROC of 0.9754. This firmly states that the model achieves the discriminatory power we desire. A classification report reaffirmed the truth of balanced precision and recall for both authentic and fake classes. Furthermore, our study reveals the more detailed applications of data augmentation and batch normalisation as key features for achieving high success and accuracy. The result can not only be an answer to one of the threats of synthetic imagery but also can be executed in journalism, digital forensics, and social media moderation. This research addresses the issues related to synthetic imagery, thereby contributing to the further establishment of trust in visual media and reducing the risks associated with fake news.*

***Keywords**: AUC-ROC, Convolutional Neural Network, Data augmentation, Normalization.*

Abbreviations:
CNNs: Connected Neural Networks
GANs: Generative Adversarial Networks
AI: Artificial Intelligence
ViTs: Vision Transformers

Manuscript received on 24 May 2025 | First Revised Manuscript received on 31 May 2025 | Second Revised Manuscript received on 18 September 2025 | Manuscript Accepted on 15 October 2025 | Manuscript published on 30 October 2025.

*Correspondence Author(s)

Virginia Peter Gonsalves*, Student, Department of Electronics and Communication Engineering, KLS VDIT, Haliyal (Karnataka), India. Email ID: virginia204g@gmail.com, ORCID ID: 0009-0005-5089-8001.

Tahira H. Shaikh, Student, Department of Electronics and Communication Engineering, KLS VDIT, Haliyal (Karnataka), India. Email ID: shaikhtahira301@gmail.com, ORCID ID: 0009-0003-4462-3330.

Muskan A. Pathan, Student, Department of Electronics and Communication Engineering, KLS VDIT, Haliyal (Karnataka), India. Email ID pathanmuskan1015@gmail.com, ORCID ID: 0009-0006-4228-7531.

Prof. Plasin Francis Dias, Assistant Professor, Department of Electronics and Communication Engineering, KLS VDIT, Haliyal (Karnataka), India. Email ID: plasindias@gmail.com, ORCID ID: 0000-0001-9740-0809.

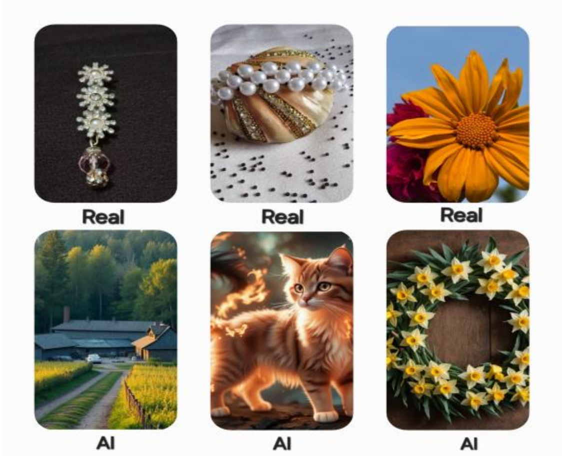
I. INTRODUCTION

The rapid increase and availability of Artificial Intelligence (AI) technologies, primarily through models such as Generative Adversarial Networks (GANs) and diffusion models, have led to the creation of synthetic images that are often highly indistinguishable from authentic images. While these technologies offer new ways to boost creativity, imagination, and save time, they also create critical problems, such as the spread of false or incorrect information, fake videos (deepfakes), and cheating in fields like law enforcement, news reporting, online shopping, and many others. As AI-generated visuals become increasingly sophisticated, the human eye alone can no longer reliably determine the authenticity of images, threatening trust in digital media and amplifying the need for an automated detection mechanism. To address the challenge of differentiating genuine images from synthetic ones, this paper presents a system that utilises a Convolutional Neural Network (CNN) to achieve high accuracy in image authenticity detection.

It is now well established that the state-of-the-art in deep learning practices has found applications with great success, primarily in computer vision, ever since. Neural Networks can learn hierarchical feature representations automatically, and their advanced techniques have revolutionised several tasks, such as image categorisation, object recognition, segmentation, and many more. Deep Learning is at the forefront of AI, and CNN is one of the essential architectures that have emerged from it. In contrast to conventional fully connected neural networks, CNNs exploit spatial hierarchies through convolutional operations, weight sharing, and pooling, which allows them to be very efficient for image-based tasks.

The CNN architecture was calibrated on a high-quality dataset comprising approximately 5,392 real and synthetic visuals, spanning a broad range of fields, from human portraits to landscapes and product photos. Fig. 1 displays some of the images from the dataset that are labelled as Real or AI. Unlike the classic methods that rely on manual extraction and heuristic rules, the CNN model automatically learns subtle details from images, such as inconsistencies in lighting, textures, patterns, or unnatural edge transitions, which is referred to as feature extraction.

The CNN model works by analysing images through multiple layers that detect patterns and details. These layers help identify subtle differences between authentic photos and AI-generated ones, such as unusual textures or edges. The CNN's ability to focus on local information and spatial patterns has made it the best choice. By advancing detection methods for AI-generated content, this



[Fig.1: Sample Images from the Dataset Displaying Real and AI-Generated Categories]

The system strengthens defences against misinformation, enables verification of image authenticity, and reduces opportunities for AI-driven manipulation. It advocates for the ethical development of AI by emphasising transparency and accountability in system design. In an era where digital trust is critical, this work safeguards visual authenticity, promotes responsible innovation, and empowers users to discern authentic content in an AI-driven landscape.

II. FOUNDATIONAL WORK

With the ever-increasing sophistication of AI-generated imagery, the need for detection systems has also increased. This has led to advancements in detection techniques, one of which is Vision Transformers (ViT). Using ViTs, Darshan Lamichhane [1] achieved superior detection accuracy by examining specific image features and comparing them to standard CNN-based methods—similarly, Md. Zahid Hossain, Farhad Uz Zaman, and Md. Rakibul Islam [3] introduced CNNs with ViTs and Explainable AI techniques, such as Grad-CAM, to detect artefacts introduced by image generators in the biophysics simulation pipeline, resulting in better classification results. Delong Liu and Yongping Lin [5] developed a hybrid method by incorporating convolutional modules into Vision Transformers (ViTs), thereby enhancing the precision of deepfake face detection through improved spatial feature extraction, which facilitates attention calculation.

The robustness and reliability of real-world operation are another crucial concern that comes before the hybrid model. For instance, Jose M. Badia, Ignacio Martin-Salinas, German Leon, Adrian Amor-Martin, Frias-Dominguez, Jose A. Belloch, and Almudena Lindoso [4] investigated the performance of ViT and CNN models under neutron radiation in edge AI environments. This is especially important for hardware reliability. The Dense ResNet-50 model has also drawn significant attention from the scientific community due to its practical applications. For example, Gurpreet Singh and Kalpna Guleria [14] targeted this model in their analysis of real fake image discrimination. Place the first layer as the input, hidden layers 2 through n + 1, then into major network feature points, and the last layer for output. In the meantime, Weilun Zhou, Jingxuan Liu, and Mingxian Wang [2] demonstrated a real-time CNN-based hazard detection system for self-driving vehicles. They emphasised the flexibility and low delay of CNNs in safety-critical fields. Meanwhile, although the end goal may differ, the principles of real-time detection and deploying lightweight models are applicable in a straightforward manner for fake image detection tasks on mobile or embedded devices.

With the rise of fake news, deep learning has spread to more areas in general. Srivanth Srinivasan, Nischitha P, Akshita Chavan, and Mohana [6] did not escape the trend either: They conducted a comparative evaluation between several different deep-learning techniques across this field on face/pose images and showed that, despite CNNs' lower levels of spatial information capture than ANNs, they score values higher for videos therein, resulting in far stronger ability to capture dependencies. They also introduced a special collaborative AI model based on multi-agent behaviour for scalable and fault-tolerant detection systems that handle synthetic media. Bambang Sugiantoro [15] chose a different path and rolled out a collaborative AI model utilising a multi-agent mechanism for networked detection of fake media, providing flexibility and fault tolerance. Dr. T. Srinivasa Ravi Kiran, Y. Sudha Madhuri, Venkata Bala Annapurna P., Dr A. Lakshmanarao, and B. S. N. Murthy [7] focus on skin cancer detection, demonstrating that hybrid CNN-ResNet models excel in fine-grained image classification, a feature also beneficial for recognising small changes in forged images. In conjunction, these surveys illustrate that transformer architecture convergence, CNN robustness, explainable AI, and deployment reliability are the main drivers of the next iteration of systems for detecting AI-generated content.

Recent advances in AI-generated image detection have increasingly adopted CNN-based architectures due to their adeptness in high-level feature extraction and detection of manipulation artefacts. Jainullabdeen A, Kadhirvel M, and Ramya Kumari T [10] proposed a CNN-based solution, Imago Veritas, that emphasises lightweight and practical classification in governance likeness, where AI-generated content is distinguished from authentic images. Their approach yields substantial gains in accuracy and latency, particularly in constrained environments. They successfully demonstrated the application of CNNs for terrain analysis for UAV-based urban reforestation, showing the scalability of CNNs for environmental monitoring. Khushi Mittal, Kanwarpartap Singh Gill, Rahul Chauhan, Manish Sharma, and G. Sunil [11] advanced this work by fortifying CNNs against adversarial attacks, a growing concern in the domain. Their system meets the requirements of both real-time performance and robust deployment in secure environments. They also discussed the application of sequential CNN models to the class of playing cards, thus showcasing effective feature learning in structured visual datasets. On a different front, Santosh, Li Lin, Irene Amerini, Xin Wang, and Shu Hu [8] presented a CLIP-based technique tailored for capturing diffusion-generated images, employing

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

contrastive learning to detect semantic errors and showcasing strong robustness across generative models.

Muthaiah U, A. Divya, T.N. Swarnalaxmi, and Vidhyasagar BS [9] provided a solid foundation for their comparative study on classification strategies for AI-generated and authentic images. These authors reviewed the strength of CNNs in spatial pattern detection and the increasing foothold of transformers in contextual analysis.

Many application case scenarios exist at the periphery of AI-generated image detection from which learning can be extracted. Mahdyar Ravanbakhsh, Hossein Mousavi, Moin Nabi, Mohammad Rastegari, and Carlo Regazzoni [12] proposed a CNN-aware binary map for semantic segmentation, which facilitates achieving an object-level understanding of manipulated content. R. Deepika, Chaitanya Sai Kondoju, Praneeth Reddy Bokka, Adarsh Bonala, and Rithwik Reddy Banala [13] explored hybrid CNN-GNN models for gesture recognition, which represents a further step in considering multimodal fusion. Squeeze-and-Excitation ResNet is verified by Yash Mori and Nandini Modi [16] to enhance the classification of OCT images, providing a possible architectural insight into improving visual-feature attention. Abdulrahman Kariri and Khaled Elleithy [17] culminated in the development of an image recognition system for vision-impaired individuals, achieving high intersection-over-union accuracy and highlighting the broader applicability of CNNs in accessibility-oriented vision systems.

III. ARCHITECTURE AND IMPLEMENTATION

A custom Convolutional Neural Network is explicitly developed to distinguish between authentic and generated images, as shown in Fig. 2. The paper presents the intricacies of dataset preprocessing, model architecture design, training procedure evaluation metrics, and implementation details in alignment with image classification.

A. Dataset and Preprocessing

A successful image detection algorithm requires extensive development with a thorough dataset, which comprises 5,392 images divided into three sets: training images (3,964 images), validation images (714 images), and testing images (714 images). The training subset is used to train the model. Validation images are utilised to monitor and tune hyperparameters, thereby guarding against overfitting during training. Finally, the testing image set serves to evaluate the final output of the trained model. Every subset has subfolders corresponding to two classes: real and AI-generated.

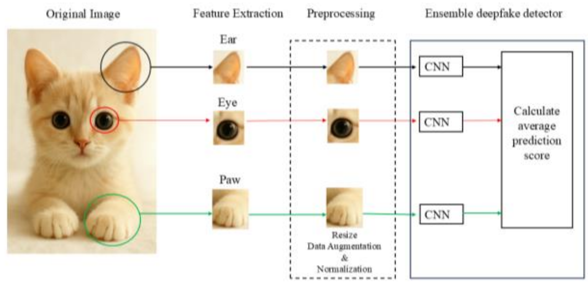


Fig.2: Workflow of Image Detection System Illustrating Feature Extraction, Preprocessing, Image Detection, and Predicting Whether the Image is Real or AI-Generated.

To ensure images that will feed the CNN model are preprocessed, the following kinds of preprocessing have been carried out:

1) Normalization: Pixel values normalized from [0, 255] to [0, 1] to keep the input scale homogeneous for the whole dataset.

2) Data Augmentation: To create diversity in the training subset, allowing the model to generalise more effectively, random transformations are applied, including gyrations (up to 15 degrees), zooms (up to 10%), horizontal flipping, contrast adjustments, and brightness variations (±10%). This derives many varied versions of the same data and thus reduces the chance of overfitting.

3) Resizing: All portraits in the subset are resized to a consistent resolution of 256 × 256 pixels, as required by the model's input specifications.

B. Model Architecture

The Custom Conventional Neural Network (CNN) is used for dichotomous classification to distinguish authentic images from artificial images. CNN architecture is well structured. It consists of three convolutional blocks, each followed by a max-pooling layer, and then two dense layers. The description of architecture is as follows:

1) Convolutional Blocks: Four convolutional blocks extract hierarchical features from the uploaded images. Each one of those blocks has a pair of convolutional layers with an activation function of ReLU. The number of filters increases in the blocks: the first block accommodates 32 filters, the second block 64 filters, the third block 128 filters, and the fourth block 256 filters. Each block is then followed by the inclusion of batch normalisation, which serves to enhance the stability and efficiency of the model consistently. The last max-pooling layer uses a 2×2 pool size for down-sampling the spatial dimension in the feature maps. Here, there is regularization layer to prevent model overtraining.

2) Flatten Layer: The outcome of the fourth convolutional block will converge into a one-dimensional output vector, where it is supposed to pass from convolutional to dense layers.

3) Dense Layers: The architecture was designed to capture high-level feature information through two fully connected layers. The first dense layer consists of 512 neurons with ReLU activation, which is then subjected to batch normalisation and a dropout rate of 0.5. Next comes the second thick layer of 256 neurons, which again uses ReLU activation, followed by batch normalisation and dropout with a rate of 0.5. A final layer, consisting of a single unit with sigmoid activation, outputs probability scores to classify images into real and AI-generated. Model compilation was done using the Adam optimiser with binary cross-entropy as a loss function.

C. Training Procedure

The training procedure attempts to adjust the weights of the CNN model to counteract unforeseen disturbances, thereby

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

enhancing generalisation. Generally, these training and validation sets are used as a measure of training progress and to prevent overfitting. Hence,

The training process is segmented as follows:

1) Training Setup: The training setup involves training on the properly preprocessed training subset with a batch size of 32, aiming to strike a balance between efficiency and stability for approximately 50 epochs.

2) Overfitting Prevention: To combat overfitting, several mechanisms were implemented. Early stopping stops training once validation loss increases consecutively for over ten epochs and reverts the weights to the epoch with the least validation loss. Model checkpointing stores the best-performing validation accuracy model for later use under the name 'cnn.keras'. In addition, ReduceLROnPlateau incorporates methods to scale down the learning rate by a factor of 0.2 if the validation loss does not improve for five epochs, while enforcing a minimum learning rate of 10−7.

D. Evaluation Metrics

The performance of the trained model has been evaluated using multiple metrics on a test subset comprising images that were never presented during the training or validation process. The first metric assessed is accuracy, defined as the number of correctly classified images over the total pictures in the test subset. Precision is computed as true positives among predicted positives; recall as true positives among actual positives; F1 score, as a harmonic mean of both; therefore, AUC for ROC, giving the widest score of the model's classification performance.

E. Implementation Details

The methodology is implemented in TensorFlow with Keras acting as a high-level API, a highly popular deep learning framework. The Keras Image Data Generator utility handles data loading, preprocessing, and real-time data augmentation during training, thereby working well in batch mode while remaining highly integrated into the CNN model. This work utilizes NumPy for numerical computations, Matplotlib and Seaborn for data visualization, and Scikit-learn for evaluating performance metrics.

F. Prediction

To extend the practical utility of this trained custom CNN model, a special prediction script, known as the "predict file," was also developed. This script enables end-users to predict whether uploaded images are real or AI-generated using a pre-trained model. The prediction score is interpreted based on the class indices defined during training, i.e., {'AI': 0, 'Real': 1}. If the prediction score < 0.5, then such an image will be designated as AI-generated, which means higher chances of being synthetically produced. Whereas ≥ 0.5 is real, which indicates that it is from natural sources. The 0.5 threshold thus represents a perfect middle ground for decision-making.

IV. RESULT AND DISCUSSION

The model is optimised heavily through the preprocessing pipeline. Normalisation scales input values to a fixed, wide range, thereby propagating stable learning across batches, whereas techniques like controllable rotation, zooming, and horizontal flipping diversify and refine generalisation. These conditions prevent the model from overfitting and help it adapt to the inherent variability in nature's image data. Uniform resizing at all dataset entries to 128x128 pixels maintains structural consistency, enabling efficient training. The two folds thus form an underpinning system on which further enhancements can be built, enhancing feature extraction and maintaining classification robustness, even on relatively small datasets.

It is not just efficient; from a design point of view, the ensemble architecture would also be modular and scalable. Each region-specific CNN operates as a specialised detector, and the final prediction is obtained through a weighted aggregation of their outputs. This contributes to the precision of the process, simultaneously opening further developments—for example, in adding specific detection for background artefacts, lighting inconsistencies, or texture continuity to this approach, much as depending on the weightage of the prediction from multiple detectors. In contrast to transformer-heavy frameworks, which are computationally costly and require massive datasets for training, lightweight CNN-based ensemble architectures are ideally suited for real-time feasibility and are ready for deployment in edge devices. This further aligns with modern-day trends in explainable AI, where the decision-making process is made transparent through the traceable prediction paths of individual sub-networks.

Test accuracy of 93.56% and an AUC of 0.98 show the powerful potential of the proposed CNN in establishing robust separation of authentic images from fake images on par with recent deepfake detection models [cite relevant paper if available]. A balanced F1-score of 0.94 and a confusion matrix provided good evidence for reliable separation across both classes. The significant variation in validation loss suggests some instability, which may be attributed to the relatively small size of the validation set. This leads to the recommendation for future work to further increase the size of the datasets tested or apply different regularisation strategies to stabilise training. While a 5,392-image dataset made training possible, larger datasets are expected to improve generalisation significantly.

The proposed ensembles of the CNN model are designed modularly such that they take an image as input and then split it into relevant semantic areas—ear, eye, and paw—each analyzed independently with convolutional subnetworks. This assembly provides room for fine localisation of regional anomalies, allowing for good interpretability. In contrast, an aggregation mechanism blends local predictions into a final classification, enhancing accuracy and facilitating real-time inference. Hybrid ViT-CNN architectures, in contrast, employ convolutional layers for spatial feature learning and transformer encoders

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

for modelling long-range dependencies to harness the multi-head self-attention mechanism for identifying subtle global irregularities. The CLIP-based semantic detectors introduce a contrastive learning step that aligns image feature representations with natural language embeddings for the detection of any occurring semantic mismatches as well as object-scene incoherences.

Those are especially effective against diffusion-based generative models. Traditional CNN classifiers, therefore, work by interfacing through stacked convolution and pooling layers that provide end-to-end feature learning, contributing to quick initiation and efficiency in deployment, but with regional insensitivity. Grad-CAM-enhanced models were proposed to provide a visual explanation that illuminates the discriminative regions of interest to a model for its classifying decisions, contributing to increased transparency in cases of forensic application, such as detector models for facial forgery. Finally, modular detection frameworks consist of decentralised classifiers trained on varied attributes or regions, thus enhancing scalability, fault tolerance, and ease of integration with emerging detectors as summarised in Table I. Altogether, these models portray a multifaceted view characterizing synthetic image forensics, with a balance among accuracy, explainability, and operational efficiency across diverse contexts.

The result presents the performance evaluation of the proposed Convolutional Neural Network (CNN) for distinguishing authentic images from fake images. The model was trained and evaluated on a dataset comprising 5392 images, split into training (3964 images), validation (714 images), and test (714 images) sets, following a 70%-15%-15% ratio. The model's effectiveness was assessed through various metrics, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC), along with a classification report, confusion matrix, and visual analyses of training dynamics and example predictions.

A. Performance Metrics on Test Set

The model's performance on the test set is summarised in Table 2. It achieved an accuracy of 0.9356, correctly classifying 93.56% of the test samples. The precision was 0.9587, indicating a high proportion of accurate optimistic predictions among all positive predictions, while the recall was 0.9104, showing the model identified 91.04% of actual positive cases. The F1-score, which balances precision and recall, was 0.94. Additionally, the AUC-ROC of 0.9754 highlights the model's excellent discriminative ability between authentic and AI-generated images.

B. Classification Report

Table III presents the detailed classification performance for both classes. For authentic images, the model achieved a precision of 0.91, a recall of 0.96, and an F1-score of 0.94. For AI-generated images, the precision was 0.96, the recall was 0.91, and the F1-score was 0.93. The macro and weighted averages for precision, recall, and F1-score were all 0.94, indicating balanced performance across classes. The overall accuracy on the test set, as derived from the classification report, was 0.94.

C. Confusion Matrix

The confusion matrix, shown in Fig. 3, provides insight into the model's prediction accuracy. Out of 357 authentic images, 343 were correctly classified, with 14 misclassified as AI-generated. For AI-generated images, 325 out of 357 were correctly identified, with 32 misclassified as authentic. This results in a balanced error distribution, confirming the model's robustness.

Table I: Detection Techniques Summary

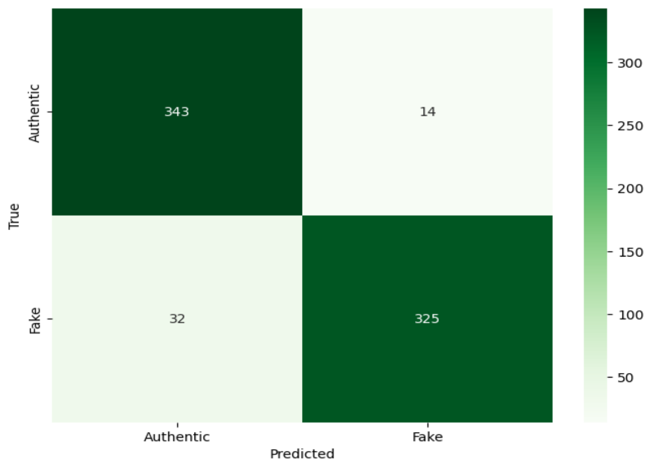
Model	Core Architecture	Detection Focus	Key Characteristics
Hybrid ViT-CNN	Transformer + Conv Layer	Global attention with spatial refinement	High accuracy, suitable for large datasets
CLIP-based Semantic Detector	Contrastive Learning	Diffusion model inconsistency	Generalizable and Semantic-Aware
Pure CNN Classifier	Standard CNN	Holistic image classification	Fast and efficient, limited feature depth
Grad-CAM Deepfake Detector	CNN + Class Activation Map	Facial artefacts and regions of interest	Explainable, adaptable to multi-modal data
Modular Detection Framework	Multi-agent CNN Ensemble	Distributed decision - making	Scalable, fault-tolerant, ideal for cloud integration

Table II: Performance Metrics on Test Set

Metric	Value
Accuracy	0.9356
Precision	0.9587
Recall	0.9104
F1-score	0.94
AUC-ROC	0.97

Table III: Classification Report

Class	Precision	Recall	F1-Score	Support
Authentic	0.91	0.96	0.94	357
Fake	0.96	0.91	0.93	357
Accuracy			0.94	714
Macro Avg	0.94	0.94	0.94	714
Weighted Avg	0.94	0.94	0.94	714



[Fig.3: Confusion Matrix]

D. Training and Validation Dynamics

The training and validation accuracy and

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

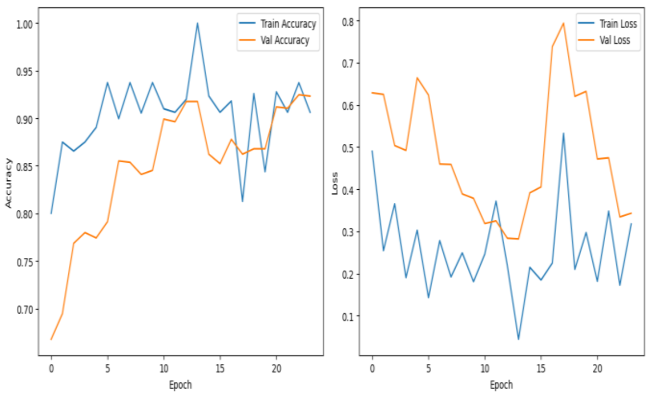
loss curves over 20 epochs are depicted in Fig. 4. The training accuracy (blue) stabilised around 0.90. In contrast, the validation accuracy (orange) peaked at approximately 0.92, showing consistent performance across both sets. The training loss (blue) decreased steadily to around 0.2, while the validation loss (orange) converged to approximately 0.3, indicating good generalization with minimal overfitting.

E. Receiver Operating Characteristic (ROC) Curve

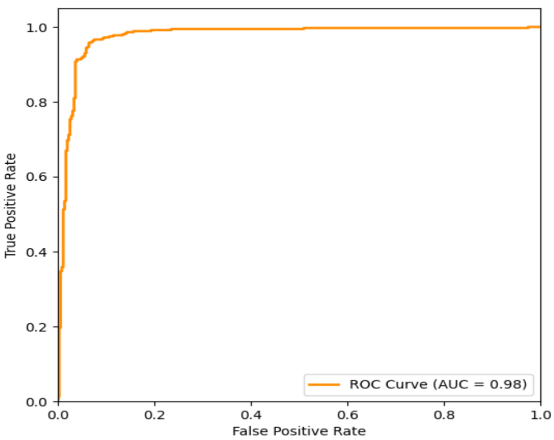
The ROC curve, presented in Fig. 5, illustrates the trade-off between actual positive rate and false positive rate. The AUC of 0.98 underscores the model's strong capability to distinguish between authentic and AI-generated images across various classification thresholds.

F. Example Classifications

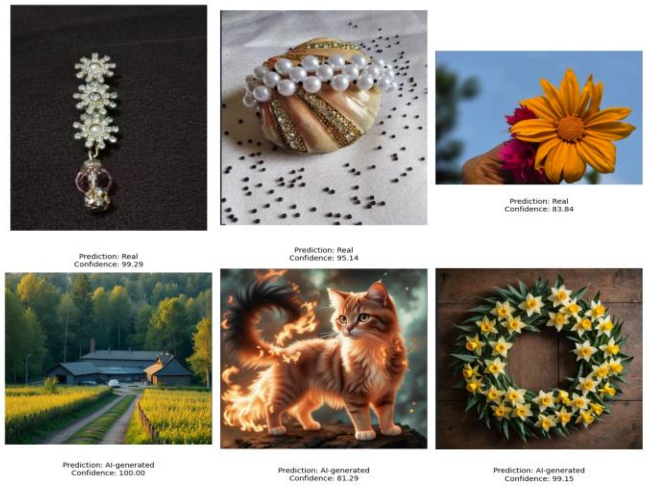
Fig. 6 showcases example images with their corresponding predictions and confidence scores. The model accurately identified authentic photos, such as a piece of jewellery (confidence: 99.29%), a decorative item (confidence: 95.14%), and a flower (confidence: 83.84%). It also correctly classified AI-generated images, including a landscape (confidence: 100.00%), a cat with flames (confidence: 81.29%), and a floral wreath (confidence: 99.15%). These examples demonstrate the model's reliability across diverse image types.



[Fig.4: Training and Validation Curves Showing Accuracy (left) and Loss (Right) Over Epochs]



[Fig.5: ROC Curve]



[Fig.6: Obtained Output of Sample Images]

V. CONCLUSION

This paper describes an innovative approach to image forgery detection that combines machine learning with computer vision, utilising image processing techniques. The system accurately classified images as real, fake, or AI-generated using Convolutional Neural Networks (CNNs). The experiments conducted on standard datasets proved the effectiveness of the proposed approach. This study developed a comprehensive method for authenticating digital images. Based on the experimental results, the model is effective in accurately detecting both manipulated images and images captured by the camera.

DECLARATION STATEMENT

After aggregating input from all authors, I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This article has not been funded by any organizations or agencies. This independence ensures that the research is conducted with objectivity and without any external influence.
- **Ethical Approval and Consent to Participate:** The content of this article does not necessitate ethical approval or consent to participate with supporting documentation.
- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Author's Contributions:** The authorship of this article is contributed equally to all participating individuals.

REFERENCES

1. Darshan Lamichhane, "Advanced Detection of AI-Generated Images Through Vision Transformers," IEEE publisher, vol. 1, no. 1, pp. 3644 - 3652, Dec. 2024. DOI: https://doi.org/10.1109/ACCESS.2024.3522759
2. Weilun Zhou, Jingxuan Liu, and Mingxian Wang, "A Real-Time Hazard Detection AI Accelerator for Autonomous Vehicles Based on Convolutional Neural Networks," 2025

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Asia-Europe Conference on Cybersecurity, Internet of Things and Soft Computing (CITSC), vol. 1, no. 1, pp 68-72, Mar. 2025. DOI: https://doi.org/10.1109/CITSC64390.2025.00019

3. Md. Zahid Hossain, Farhad Uz Zaman, and Md. Rakibul Islam, "Advancing AI-Generated Image Detection: Enhanced Accuracy through CNN and Vision Transformer Models with Explainable AI Insights," 2023 26th International Conference on Computer and Information Technology (ICCIT) 13-15 December, Cox's Bazar, Bangladesh, vol. 1, no. 1, pp. 1-6, Feb. 2024. DOI: https://doi.org/10.1109/ICCIT60459.2023.10440990
4. Jose M. Badia, Ignacio Martin-Salinas, German Leon, Adrian Amor-Martin, Frias-Dominguez, Jose A. Belloch, Almudena Lindoso, "Reliability of Vision Transformers and CNNs on Edge AI systems under neutron radiation," IEEE Transactions on Nuclear Science, vol. 1, No. 1, pp. 1-11, Jan. 2025. DOI: https://doi.org/10.1109/TNS.2025.3536519
5. Delong Liu and Yongping Lin, "A DeepFake Face Detection Method Using Vision Transformer with a Convolutional Module," The 17th International Conference on Advanced Computer Theory and Engineering, vol. 1, no. 1, pp. 1-5, July. 2024. DOI: https://doi.org/10.1109/ICACTE62428.2024.10871782
6. Srivanth Srinivasan, Nischitha P, Akshita Chavan, and Mohana, "Deepfake Image and Video Detection using Deep Learning Algorithms," Proceedings of the International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI-2025), vol. 1, no. 1, pp. 1-5, Feb. 2025. DOI: https://doi.org/10.1109/ICMSCI62561.2025.10893930
7. Dr.T.Srinivasa Ravi Kiran, Y.Sudha madhuri, Venkata Bala Annapurna P, Dr. A.Lakshmanarao, and B.S.N.Murthy, "A Novel Approach to Skin Cancer Detection using CNN, ResNet, and Hybrid Feature Extraction," Proceedings of the Second International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS-2024), vol.1, no.1, pp. 217-222, Oct. 2024. DOI: https://doi.org/10.1109/ICSSAS64001.2024.10760465
8. Santosh, Li Lin, Irene Amerini, Xin Wang, and Shu Hu, "Robust CLIP-Based Detector for Exposing Diffusion Model-Generated Images," 2024 IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), vol. 1, no. 1, pp. 1-7, July. 2024. DOI: https://doi.org/10.1109/AVSS61716.2024.10672612
9. Muthaiah U, A. Divya, T. N. Swarnalaxmi, and Vidhyasagar BS, "A Comparative Review of AI-Generated vs Real Images and Classification Techniques," Proceedings of the Fourth International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS-2024), vol. 1, no. 1, pp. 141-147, Dec. 2024. DOI: https://doi.org/10.1109/ICUIS64676.2024.10866220
10. Jainullabdeen A, Kadhirvel M, and Ramya Kumari T, "CNN-Driven Terrain Analysis and Infrared Soil Hydration Control in Autonomous UAV-Based Urban Reforestation," 2024 International Conference on System, Computation, Automation and Networking (ICSCAN), vol. 1, no. 1, pp. 1-5, Dec. 2024. DOI: https://doi.org/10.1109/ICSCAN62807.2024.10894668
11. Khushi Mittal, Kanwarpartap Singh Gill, Rahul Chauhan, Manish Sharma, and G Sunil, "Playing Cards Classification and Detection Using Sequential CNN Model Through Machine Learning Techniques Using Artificial Intelligence," 2024 International Conference on E-mobility, Power Control and Smart Systems (ICEMPS), vol. 1, no. 1, pp. 1-4, Apr. 2024. DOI: https://doi.org/10.1109/ICEMPS60684.2024.10559365
12. Mahdyar Ravanbakhsh, Hossein Mousavi, Moin Nabi, Mohammad Rastegari, and Carlo Regazzoni, "CNN-Aware Binary Map for General Semantic Segmentation," IEEE, vol. 1, no. 1, pp. 1923-1927, Sept. 2016. DOI: https://doi.org/10.1109/ICIP.2016.7532693
13. R Deepika, Chaitanya Sai Kondoju, Praneeth Reddy Bokka, Adarsh Bonala, and Rithwik Reddy Banala, "Hand Gesture Recognition using CNN-GNN," Proceedings of the Third International Conference on Automation, Computing and Renewable Systems (ICACRS-2024), vol. 1, no. 1, pp. 1482-1487, Dec. 2024. DOI: https://doi.org/10.1109/ICACRS62842.2024.10841662
14. Gurpreet Singh and Kalpna Guleria, "A Dense Residual Network-50 Model for Identification of Real and Fake Images," 2024 Second International Conference Computational and Characterization Techniques in Engineering & Sciences (IC3TES), vol. 1, no. 1, pp. 1-5, Nov. 2024. DOI: https://doi.org/10.1109/IC3TES62412.2024.10877597
15. Bambang Sugiantoro, "Deepfake face images: Explainable detection using deep neural networks and class activation mapping," 2024 IEEE International Symposium on Consumer Technology (ISCT), vol. 1, no.1, pp. 86-90, Aug. 2024. DOI: https://doi.org/10.1109/ISCT62336.2024.10791156
16. Yash Mori and Nandini Modi, "Enhancing OCT Image Classification for Retinal Disease Diagnosis: A Novel Approach Using Squeeze-and-Excitation ResNet," 2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC), vol. 1, no. 1, pp. 940-945, Jul. 2024. DOI: https://doi.org/10.1109/AIC61668.2024.10730829
17. Abdulrahman Kariri and Khaled Elleithy, "Astute Support System for Visually Impaired and Blind with Highest Intersection over Union for Object Detection and Recognition with Voice Feedback," vol. 2, no. 1, pp. 1-7, Nov. 2024. DOI: https://doi.org/10.1109/LISAT63094.2024.10807856

AUTHOR'S PROFILE

Virginia Peter Gonsalves, a final-year undergraduate student pursuing a degree in Electronics and Communication Engineering, brings a dynamic and multifaceted skill set to the field of technology. Proficient in image processing and deep learning, she demonstrates a strong aptitude for tackling complex technical challenges. Her expertise in CNNs is pivotal in her contribution. She excels in problem-solving, team collaboration, and project management, enabling her to work effectively in interdisciplinary settings. Passionate about innovation, she is driven to advance electronics and software development by leveraging her technical and collaborative spirit to deliver cutting-edge solutions.

Tahira Hafizurrehaman Shaikh is a final-year B.E. student in the the Electronics and Communication Engineering Department. She possesses strong technical proficiency in image processing and deep learning, with a particular interest in Convolutional Neural Networks (CNNs). Her strength lies in analytical problem-solving, effective teamwork, and interdisciplinary cooperation. Inspired by a passion for innovation, she actively discovers progress in electronics and software systems. She is committed to contributing to technological progress through dedicated research and desires to make meaningful changes in the field of study.

Muskan Ameerkhan Pathan is a final-year student in Electronics and Communication Engineering. She combines a firm academic grounding with hands-on experience in embedded systems, IoT architectures, and AI tools. She has demonstrated effective communication and problem-solving skills while working on complex, team-based technical challenges. Her ability to bridge hardware and intelligent software systems enables her to make meaningful contributions to smart automation and system optimisation. Passionate about image processing and AI innovation, she is focused on developing scalable, real-world solutions that integrate emerging technologies. Her dedication and interdisciplinary approach mark her as a promising contributor to the field.

Prof. Plasin Francis Dias is currently working as an Assistant Professor in the Department of Electronics and Communication, KLS VDIT, Haliyal, India. She has a strong academic background in VLSI Design and Embedded Systems. Her areas of interest include Signal processing, radar, image processing, and Machine learning. She has presented papers at international conferences and national conferences, published papers in international Journals, and more. She is passionate about teaching and research, with a focus on guiding students and contributing to advancements in her field through scholarly work and innovative practices for nearly 25 years.