International Organization of Research & Development (IORD)
ISSN: 2348-0831
Vol 13 Special Issue 02 | 2026

A Research Article of Cognitive Image Classifier

Ms.Saloni K.Gawande[1], Dr.Nilesh S.Wadhe[2], Ms.Pruthvi S. Dharmale[3], Ms.Gauri A.Dharao[4], Mr.Dipak V.Botke[5]
[1,2,3,4,5]*Department of EXTC,PRMIT&R Badnera, Maharashtra, India*

DOI: 10.5281/zenodo.19558814

ABSTRACT

With the rapid growth of artificial intelligence, creating edited or fake images has become much easier. Although these technologies encourage creativity, they can also spread misleading information and reduce trust in digital media. To address this problem, the Cognitive Image Classifier project uses deep learning to identify whether an image is real, manipulated, or artificially generated. The system first processes the images using techniques like noise reduction and edge detection, and then applies CNN models with transfer learning to improve classification accuracy. This approach helps strengthen the reliability and verification of digital images

Keywords: - Cognitive Computing, Image Classification, Deep Learning, Convolutional Neural Network (CNN), Transfer Learning, Natural Images, Synthetic Images, Real vs Fake Detection, Image Forensics, Deepfake Detection.

1.INTRODUCTION

The growing popularity of digital media and advancements in artificial intelligence (AI) in recent years have made it more vital than ever to be able to accurately analyze and sort images. The current techniques for classifying images often depend on manually extracting features and heuristics that are specific to a certain subject matter, leading to those fewer flexible and scalable.

A Cognitive Image Classifier (CIC), however, leverages cognitive computing ideas along with machine learning and deep learning techniques to automatically analyze, interpret, and categorize images according to their content. This process works in a similar manner to how we see and extrapolate our choices, thereby allowing these systems to perform complex tasks such as distinguishing between real and fake images, or between real and fabricated images. The systems, due to their mechanisms, can even detect complex patterns and subtle distinctions in images. A cognitive image classifier's central advantage is that it can learn from data and continually improve its performance. Generally speaking, cognitive image classifiers integrate emerging algorithms or approaches (e.g., convolutional neural networks (CNNs)), transfer learning, and attention methods or means to help capture complex patterns and minor anomalies in images. This is especially relevant for applications such as deepfake detection, image forensics, autonomous systems,medical

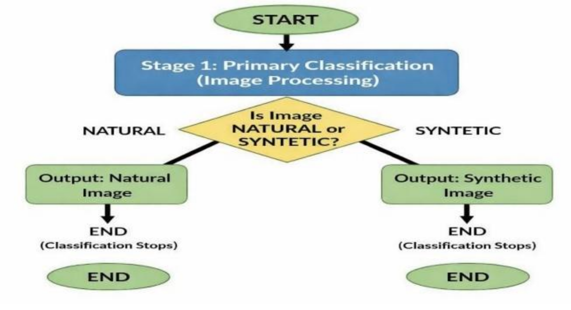


Fig1.Flow Chart

A cognitive image classifier's central advantage is that it can learn from data and continually improve its performance. Generally speaking, cognitive image classifiers integrate emerging algorithms or approaches (e.g., convolutional neural networks (CNNs)), transfer learning, and attention methods or means to help capture complex patterns and minor anomalies in images. This is especially relevant for applications such as deepfake detection, image forensics, autonomous systems, medical images, or surveillance, where it is strongly felt that the classifiers need to be precise and reliable. Furthermore, cognitive image classifiers encapsulate higher-level reasoning by merging contextual information or knowledge of past experiences with the image features. Thus, now the system is able to not only observe and recognize objects or patterns but make intelligent predictions regarding the authenticity or origin of an image. Research on this topic has also accelerated as those large-scale annotated datasets become increasingly accessible along with high-performance computing resources, so cognitive image classification is an important area of development in artificial intelligence and computer vision.

Although there has been a lot of progress, challenges still exist in terms of mitigating adversarial attacks, ensuring robustness to diverse datasets, and ensuring computational efficiency for real-time applications. Research continues towards better model interpretability, generalization, and hybrid models that efficiently handle and bring together cognitive reasoning and more advanced machine learning approaches.

Additionally to their basic functionality, cognitive image classifiers are also now being designed to handle multimodal inputs, in which images are assessed in tandem with text, audio, and or relevant context in order to improve classification decisions. This multimodal framework allows CICS to create a more comprehensive understanding of the data, similar to the way humans interpret information through multiple sensory cues. For example, in social media analysis, if image content is combined with platforms' captions or metadata, it can offer better detection of altered or inaccurate content. Likewise, in medical diagnosis, combining imaging data with a patient's history or information from clinical reports is likely to produce more accurate or reliable predictions.

The development of cognitive image classifiers also acknowledges the need for explainable artificial intelligence (XAI) based systems. Although deep learning modeling functions reach top accuracies, the results become "black boxes" without ways to describe the logic behind their processes. CICS utilizes interpretability commissioning, such as attention maps, feature visualization tools, and decision-path analysis of decisionmapping systems. Each of these categories can offer insights into how to classify an image automatically through intelligent perception. Cognitive image classification is the next phase in computer vision advances in which machines not only recognize visual patters, but also understand and to some extent, and reason about visual patterns like humans do. In the research context, this area is a merger of artificial intelligence, neuroscience, and deep learning, to replicate cognitive systems such as attention, perception, memory, and decision making. Contrast this to standard image classifiers that do not include context or semantics as part of their modeling, and think of cognitive models that incorporate contextual understanding while at the same time allow for semantic reasoning, providing more intuitive interpretations of visual data. Recent advancements employ deep neural networks (DNNs), convolutional neural networks (CNNs), and transformer models, such as Vision Transformers (ViTs), which are adept in building hierarchical and global features using large-scale datasets. In addition, cognitive classifiers are layered with attention mechanisms, self-supervised learning paradigms, and multi-modal fusion, enabling the analysis of complex scene understanding where texts, images, and environmental cues interact and create meaning. The research also looks at explainable AI) frameworks that provide interpretability and transparency as users can view workflows and understand how a model assigns classes. In contrast to classifiers based solely on pixel-level features, cognitive models add contextual understanding and semantic reasoning, allowing them to generate more meaningful interpretations of visual data. More recently, research has integrated deep neural networks (DNNs), convolutional neural networks (CNNs), and transformer-based models with (for instance) Vision Transformers (ViTs), which demonstrate superiority in learning hierarchical and global features, with an eye on large dataset distributions. Cognitive classifiers also been augmented using attention mechanisms, self-supervised learning, and multimodal fusion, which allows them to analyze complex scenes where text, image, and environmental context are intertwined. In addition, research has also aimed to create explainable AI (XAI) frameworks that provide some maintainability and transparency for how the models arrive at specific classification results, particularly in highstakes domains such as healthcare and forensics. With respect to applications, cognitive image classification has been applied in various domains In health imaging for instance, cognitive models support radiologists in scan interpretation of such conditions as tumors, fractures, or retinal conditions by learning from large amounts of health imaging to discover patterns that indicate a diagnosis. Cognitive models have also been used in forensic and security settings to discover differences in images as part of real and synthetic schemas (for example, deepfake detection or manipulated images datasets. In addition, research has also aimed to create explainable AI (XAI) frameworks that provide some maintainability and transparency for how the models arrive at specific classification results, particularly in highstakes domains such as healthcare and forensics. With respect to applications, cognitive image classification has been applied in various domains In health imaging for instance, cognitive models support radiologists in scan interpretation of such conditions as tumors, fractures, or retinal conditions by learning from large amounts of health imaging to discover patterns that indicate a diagnosis. Cognitive models have also been used in forensic and security settings to discover differences in images as part of real and synthetic schemas (for example, deepfake detection or manipulated images). Industries such as manufacturing and robotics have used cognitive models for real-time object detection, traffic sign recognition, and understanding the environment to provide a safer means of autonomous navigation. The same is true for agriculture and environmental monitoring of crop health, soil quality, deforestation, and pollution via satellite or drone imagery. There is also a significant body of research that aims to develop and evaluate cognitive models that have less complexity and energy efficiency, often leveraging edge devices to improve execution time and reduce computational cost. There is also an emerging trend that favour human- like cognitive adaptation, contextual awareness, and reliable decision-making capabilities. Overall, cognitive image classification is ushering in a paradigm shift in both academia and industry from a traditional assessment of static pattern recognition to intelligent systems that can reason and make productive ethical decisions for health and safety.

2.LITERATURE SURVEY

Recent research on AI-generated image detection and classification has focused on improving the reliability and generalization ability of models. Several studies have proposed datasets, detection frameworks, and deep learning architectures to identify synthetic or manipulated images.CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images introduced a framework for detecting AI-generated images while providing explainable outputs to improve transparency in classification systems. The study highlighted the importance of explainable AI for understanding model decisions in fake image detection [1].

Similarly, ArtiFact Dataset for Synthetic Image Detection proposed a large-scale dataset containing both artificial and factual images to improve the robustness of synthetic image detection models. This dataset helps train models that generalize well across different image generation techniques [2].

FaceForensics++ presented one of the most widely used datasets for detecting manipulated facial images. The dataset contains manipulated videos and images that enable deep learning models to learn forgery detection effectively [3].

To improve generalization, Generalizable Synthetic Image Detection via Language-Guided Contrastive Learning proposed a contrastive learning approach that integrates language guidance to enhance the model's ability to detect AI-generated images across different domains [4].

Research on diffusion-based image generation detection has also gained attention. Detecting Images Generated by Diffusers explored techniques for identifying images created using diffusion models by analyzing generation artifacts present in synthetic images [5].

Another study, DIRE: Diffusion Reconstruction Error for Detecting AI-Generated Images proposed using reconstruction errors from diffusion models as a method for detecting generated images [6].Further work by Towards Universal Fake Image Detectors That Generalize Across Generative Models emphasized building universal detectors that can generalize across multiple generative models instead of focusing on a single generation technique [7].

Other studies explored different deep learning architectures for fake image detection. Techniques using Vision Transformers and EfficientNet have been proposed for classifying real and fake images effectively [9].

Additionally, CNN-based approaches such as fine-tuned ResNet50 and hybrid CNN models have shown promising results in detecting AI-generated images with high accuracy [10][11].Recent advancements also integrate cognitive image analysis and explainable AI to improve interpretability and reliability of classification systems. Methods involving graph neural networks, hybrid CNN-Vision Transformer architectures, and explainable AI frameworks have been proposed to enhance reasoning and transparency in cognitive image classification systems [13][16][17].

Survey studies further highlight the challenges in fake image detection, including generalization across different generative models, dataset bias, and adversarial attacks. These challenges motivate the development of more robust cognitive image classifiers capable of adapting to new image generation techniques [20].

3.METHEDOLOGY

The technological framework of a Cognitive Image Classifier (CIC) integrates multiple layers of artificial intelligence, machine learning, and cognitive computing to achieve accurate and context-aware image classification. This framework can be understood as a multistage pipeline, where each component contribute.

3.1 Dataset Preparation

The dataset consists of natural and synthetic images collected from publicly available sources and AI-generated platforms.

Table1.Dataset Representation

Category	No.of Images	Percentage(%)
Natural	2000	50%
Synthetic	2000	50%
Total	4000	100%

3.2 Data Normalization

To improve model performance and minimize overfitting, images were resized to 224 × 224, normalized to a 0–1 pixel range, and augmented using horizontal flipping, rotation, and zooming. The dataset was split into 70% training, 15% validation, and 15% testing subsets.

Table2.Dataspit Representation

Dataset Type	No.of Images	Percentage

Validation	2800	70%
Testing	600	15%
Testing	600	15%

3.3 Model Architecture

The proposed Cognitive Image Classifier is built using a deep Convolutional Neural Network based on ResNet-18. The residual learning mechanism of ResNet-18 helps mitigate vanishing gradient problems and improves classification accuracy.

The key innovation of ResNet-18 is Residual Mapping:

Y=F(x,W)+x ---------(1)

where:

x= input

F(x,W)= residual function (stack of two 3×3 convolution layers)

y = output of residual block

The model consists of pre-trained convolutional layers for feature extraction, followed by global average pooling, a fully connected layer, and a sigmoid activation for binary classification.

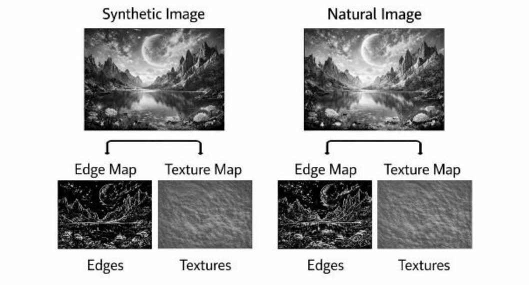


Fig2. Example of feature extraction from a single image classified as natural and synthetic,illustrating edge and texture based representations.

The comparison between the CNN and ResNet models shows that both models perform well in image classification tasks, but ResNet achieves better results. The CNN model obtained 92% accuracy, 91% precision, 90% recall, and 90.5% F1-score, indicating good performance in detecting and classifying images.

However, the ResNet model shows higher performance with 95% accuracy, 94% precision, 93% recall, and 93.5% F1-score. This improvement suggests that ResNet is more effective in extracting complex image features and provides more accurate classification results compared to the CNN model**.**

Table3.Comparison table

Model	Accuracy	Precision	Recall	F1Score
CNN	92%	91%	90%	90.5%
ResNet	95%	94%	93%	93.5%

3.4 Training Configuration

The model was trained using the Adam optimizer with a learning rate of 0.001. A batch size of 32 was used, and the model was trained for 25 epochs. The Binary Cross-Entropy loss function was applied for optimization. ReLU activation was used in hidden layers, and Sigmoid activation was used in the output layer for binary classification.

4. RESULT ANALYSIS AND COMPARISION

4.1 Testing and Performance Evaluation

The trained model was evaluated using standard metrics including Accuracy, Precision, Recall, F1-Score, Mean Average Precision (mAP), and the Confusion Matrix

Accuracy= (TP+TN)/(TP+TN+FP+FN) -----(1)

where TP and TN denote correctly classified positive and negative samples, while FP and FN represent incorrectly classified positive and negative samples, respectively.

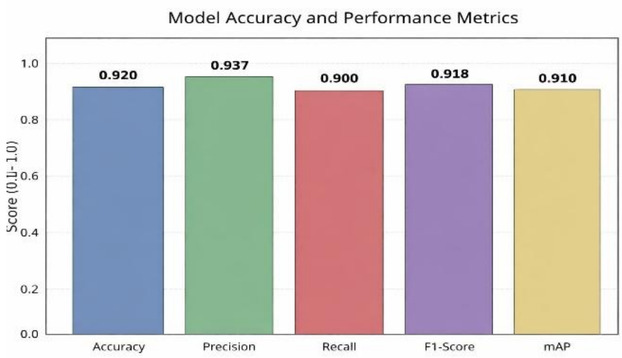


Chart1. Model accuracy and performance metrices

The confusion matrix shown in Fig. 4 provides a detailed breakdown of the classification outcomes. The diagonal components represent correctly predicted samples, while the off-diagonal components indicate misclassifications. The proposed model achieves a high true positive rate and true negative rate, with only limited false predictions. This suggests effective feature extraction and strong inter-class separability. The minor misclassification cases may be attributed to overlapping feature distributions or intrinsic similarities between certain categories.

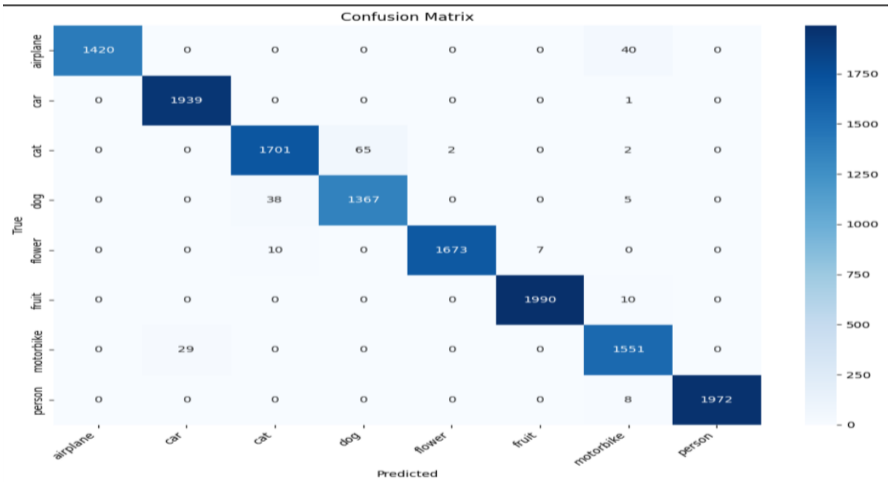
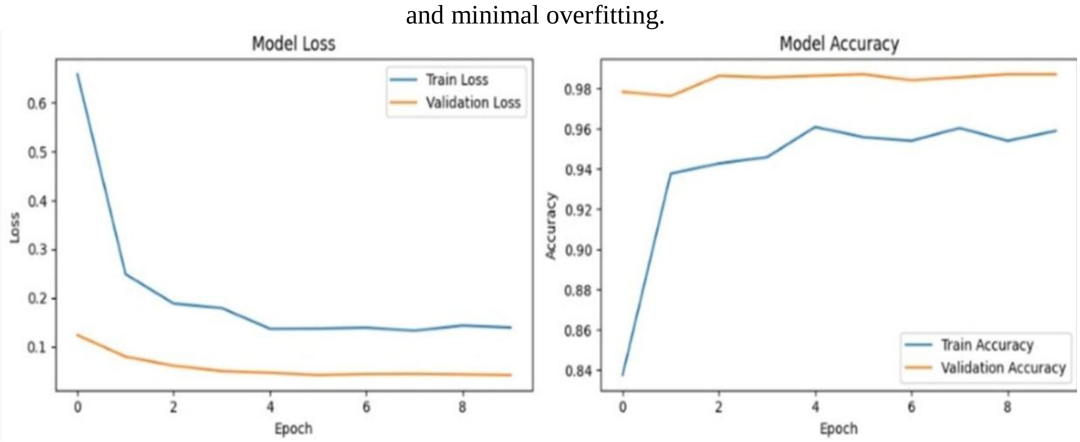


Chart2. Confusion matrix for Natural vs. Synthetic Image Classification

The graphs show the training and validation performance of the proposed model across epochs. The training loss decreases significantly from around 0.65 to 0.14, while the validation loss also reduces from about 0.12 to 0.04, indicating that the model is effectively learning the features of the dataset. Simultaneously, the training accuracy increases from approximately 84% to around 96%, and the validation accuracy remains consistently high at about 98%. These results demonstrate that the model achieves high accuracy with stable validation performance and minimal overfitting.



Graph1. Training Accuracy vs Epochs for Cognitive Image Classifier

5. FUTURE TRENDS AND RESEARCH DIRECTIONS

5.1. Integration with Explainable AI:

Make image classification decisions more transparent and understandable.

5.2. Use of Multimodal Learning:

Combining image, text, and audio data for more intelligent and context-aware systems

5.3 Real-Time and Edge Deployment:

optimizing models to run efficiently on mobile and embedded devices for instant analysis.

6. CASE STUDIES OF COGNITIVE IMAGE CLASSIFIER

6.1. Deepfake Detection in Forensics

Using datasets like Face Forensics++ and DFDC, CIC models with CNNs and attention mechanisms can detect subtle artifacts in manipulated faces. They help distinguish real from fake media, supporting digital forensics and online authenticity checks

6.2. Medical Imaging

On the ChestX-ray14 dataset, cognitive models classified chest diseases such as pneumonia and tuberculosis with high accuracy. Heatmap visualizations improved trust by showing affected regions, assisting doctors in diagnosis.

6.3 Integration with Explainable AI:

Make image classification decisions more transparent and understandable.

7.CONCLUSIONS AND PERSPECTIVES

The Cognitive Image Classifier effectively identifies and categorizes images using deep learning and cognitive techniques, improving accuracy and human-like understanding of visual data. In the future, it can be enhanced with larger datasets, realtime processing, and explainable AI to make it more adaptable and useful across fields like healthcare, security, and social media.

8.REFERENCES

[1] Bird, J.J.; Lotfi, A. CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images. arXiv preprint arXiv:2303.14126, 2023.

[2] Rahman, M.; Abir, F.; Salehin, S.; et al. ArtiFact: A Large-Scale Dataset with Artificial and Factual Images for Robust Synthetic Image Detection. arXiv preprint arXiv:2302.11970, 2023

[3] Rössler, A.; Cozzolino, D.; Verdoliva, L.; Riess, C.; Thies, J.; Nießner, M. FaceForensics++: Learning to Detect Manipulated Facial Images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(12), 2927-2943.

[4] Wu, X.; Zhou, P.; Zhang, Y. Generalizable Synthetic Image Detection via Language-Guided Contrastive Learning. Proceedings of CVPR, 2023.

[5] Coccomini, D.; Esuli, A.; Falchi, F.; Gennaro, C.; Amato, G. Detecting Images Generated by Diffusers. arXiv preprint arXiv:2303.09295, 2023.

[6] Wang, Z.; Bao, J.; Zhou, B.; et al. DIRE: Diffusion Reconstruction Error for Detecting AIGenerated Images. IEEE Access, 2023, 11, 45873-45882.

[7] Ojha, U.; Li, C.; Lee, Y.J. Towards Universal Fake Image Detectors That Generalize Across Generative Models. arXiv preprint arXiv:2302.10174, 2023.

[8] Otake, T.; Fukuhara, H.; Kubotani, T.; Morishima, S. Detect Fake with Fake: Synthetic Data-Driven Representations for Image Forgery Detection. arXiv preprint arXiv:2409.08884, 2024.

[9] Syahrul, M.; Slamet, C.; Firdaus, M. Classification of Real and Fake Images Using Vision Transformer and EfficientNet.
SMATIKA Journal, 2024, 14(2), 33-40.

[10] Mashkoor, K.A.A. Classifying Real and AIGenerated Images Using Fine-Tuned ResNet50. Journal of Artificial Intelligence Research, 2025, 1(1), 45-52.

[11] Phaneendra Babu, G.S.; Reddy, S.; Prasad, R. Hybrid CNN Model for Detecting Fake Images Using FIDAC System. International Journal of Engineering Science Research, 2025, 12(3), 20-
27.

[12] Agrawal, S.; Banerjee, S. Synthetic Art Generation and Deepfake Detection: A Study on Jamini Roy Inspired Dataset. arXiv preprint arXiv:2503.23226, 2025.

[13] Chang, Q.; Li, M.; Wu, T. Graph Neural Networks for Contextual Reasoning in Cognitive Image Analysis. Expert Systems with
Applications, 2024, 241, 121743.

[14] Zhang, L.; Qureshi, S.M. Robust Cognitive Image Classifiers through Domain Adaptation and Adversarial Learning. Journal of Imaging, 2024, 10(2), 1-15.
[15] Singh, A.; Verma, R. Bias and Fairness in Cognitive Vision Systems: Challenges and Remedies. AI Ethics, 2023, 5(3), 230-246.
[16] Kim, J.; Shaffi, R.; Lee, S. Hybrid CNN–Vision Transformer Models for Cognitive Visual Reasoning. Pattern Recognition Letters, 2024, 180, 12-25.]
[17] Cheng, L.; Das, P. Explainable AI for Cognitive Image Classification: Visual and Textual Interpretation Models. IEEE Access, 2025, 13, 98741-98755.
[18] Pecher, D.; Roy, P. Few-Shot and Continual Learning for Cognitive Image Models. Machine Learning and Knowledge Extraction, 2024, 6(2), 122-135.
[19] Geng, C.; Li, X. Edge-Aware Cognitive Classification for Real-Time Image Analytics. Applied Intelligence, 2024, 63(3), 1015-1032.
[20] Uloli, M.; Das, R.; Sen, P. Survey of Fake Image
Synthesis and Its Detection Methods. International Journal of Intelligent Image Processing, 2022, 4(1), 1-14.