International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

Image Classification And Explainable Identification Of AI – Generated Synthetic Images

Summaya[1], Samia Tabassum[2], Shaik Zubair Pasha[3], Mrs. T. Anitha[4]

[1,2,3] B.E. Students, Department of CSE, ISL Engineering College (OU), Hyderabad, India.
[4] Assistant Professor, Department of CSE, ISL Engineering College (OU), Hyderabad, India.

ABSTRACT:-

Recent advances in synthetic image generation, particularly through artificial intelligence, have led to the creation of images so realistic that they are virtually indistinguishable from real photographs. This presents significant challenges for data authenticity and reliability, especially in areas such as journalism, social media, and scientific research, where the integrity of images is critical. This study proposes an approach to effectively distinguish between real and AI-generated images using a deep learning model based on ResNet50. The classification task is framed as a binary problem, where images are categorized as either "real" or "AI-generated." While synthetic images can replicate complex visual details such as lighting, reflections, and textures, subtle visual imperfections often differentiate them from genuine photographs. The study investigates these differences, focusing on minor artifacts and inconsistencies that are typically present in AI-generated content, such as background distortions, lighting anomalies, and unnatural textures. These artifacts are not always perceptible to the human eye, but can be reliably detected by machine learning models. The ResNet50 model is employed to learn and classify these visual cues, enabling the system to achieve high accuracy in distinguishing real images from synthetic ones. By training on a large dataset of both real and AI-generated images, the model identifies key image features that serve as indicators of authenticity. The study also explores the interpretability of the model's decisions, shedding light on which aspects of the images are most informative for classificationInspection of structural cracks is critical for maintaining the safety and longevity of bridges and other infrastructure. Traditional methods for crack detection are often manual, labor-intensive, and prone to human error. Recent advances in deep learning and semantic segmentation provide a promising alternative, but obtaining high-quality annotated data remains a significant challenge. This paper introduces an enhanced approach to crack detection using deep learning, leveraging synthetic data generation and advanced semantic segmentation techniques. We propose the use of DeepLabV3 with a ResNet50 backbone, an extension of the DeepLabV3 architecture that incorporates a robust ResNet50 feature extractor to improve segmentation. Our approach involves generating synthetic crack images to address the data scarcity issue. This is achieved using the StyleGAN3 for realistic image synthesis. By integrating these synthetic datasets with the DeepLabV3+ model, we aim to boost segmentation performance beyond the capabilities of standard models. Hyperparameter tuning is performed to optimize the DeepLabV3 with ResNet50 configuration, achieving significant improvements in segmentation. We employ data augmentation techniques such as motion blur, zoom, and defocus to further refine model performance. The proposed method is evaluated against existing state-of-the-art techniques, demonstrating superior accuracy. The results indicate that our approach not only enhances the crack detection but also offers a novel application of synthetic data generation in deep learning for semantic segmentation. This research provides new insights into leveraging advanced neural networks and synthetic data for improved structural crack analysis.

INTRODUCTION:-

In recent years, artificial intelligence (AI) and generative models have made significant strides in generating synthetic images that closely mimic real-

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP77-85

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

world photographs. With advancements in deep learning techniques such as Generative Adversarial Networks (GANs) and other image synthesis methods, AI-generated content has become increasingly indistinguishable from authentic imagery. While these developments open up numerous possibilities across fields such as entertainment, marketing, and art, they also pose challenges, especially in areas requiring data integrity and authenticity, such as journalism, social media, and scientific research. As synthetic images become more realistic, it becomes critical to develop robust methods for detecting AI-generated content to ensure the reliability of digital information. However, distinguishing between real and synthetic images is no simple task, as AI-generated images can replicate intricate visual details such as textures, lighting, reflections, and even complex compositions. This project explores how to leverage advanced deep learning models, specifically ResNet50, to classify images as real or AI-generated based on subtle differences that may not be immediately visible to the human eye. These differences often manifest as small visual imperfections or inconsistencies in the background, textures, and lighting, which can be leveraged by machine learning models to reliably distinguish between synthetic and authentic images. This study aims to develop an effective system that automates the detection of AI-generated images, providing a valuable tool for image forensics, content verification, and maintaining the trustworthiness of digital media.

LITERATURE REVIEW

TITLE: Predicting image credibility in fake news over social media using a multimodal approach.

AUTHOR: B. Singh and D. K. Sharma,

YEAR: 2022

DESCRIPTION: Infrared objects acquired from a long-distance have small sizes and are easily submerged by a complex and variable background. The existing deep network detection framework suffers greatly from the feature spatial resolution loss caused by the networks' depth and multiple downsampling operations, which is extremely detrimental for small object detection. So, a crucial and urgent goal is, how to trade-off network depth and feature spatial resolution, while learning feature context representation and interaction to distinguish from the background. To this end, we propose a deep interactive UNet architecture (short for DI-U-Net) with high feature learning and feature interaction ability. First, feature learning is first achieved through a multi-level and high-resolution network structure. This structure ensures feature resolution as the network depth increases, and also focuses on the object's global context information. Then, the feature interactive is further achieved by the dense feature encoder (DFI) module to learn object local context information. The proposed method yields strong object context representation and well discriminability, as well as a good fit for infrared small object detection. Extensive experiments are conducted on the SISRT dataset and Synthetic dataset.

TITLE: Writer-independent signature verification; evaluation of robotic and generative adversarial attacks.

AUTHOR : J. J. Bird, A. Naser, and A. Lotf.

YEAR: 2023

DESCRIPTION : Forgery of a signature with the aim of deception is a serious crime. Machine learning is often employed to detect real and forged signatures. In this study, we present results which argue that robotic arms and generative models can overcome these systems and mount false-acceptance attacks. Convolutional neural networks and data augmentation strategies are tuned, producing a model of 87.12% accuracy for the verification of 2,640 human signatures. Two approaches are used to successfully attack the model with false-acceptance of forgeries. Robotic arms (Line-us and iDraw) physically copy real signatures on paper, and a conditional Generative Adversarial Network (GAN) is trained to generate signatures based on the binary class of 'genuine' and 'forged'. The 87.12% error margin is overcome by all approaches; prevalence of successful attacks is 32% for iDraw 2.0, 24% for Line-us, and 40% for the GAN.

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP77-85

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

Fine Tuning with examples shows that false-acceptance is preventable. We find attack success reduced by 24% for iDraw, 12% for Line-us, and 36% for the GAN. Results show exclusive behaviours between human and robotic forgers, suggesting training wholly on human forgeries can be attacked by robots, thus we argue in favour of fine-tuning systems with robotic forgeries to reduce their prevalence.

TITLE: Photo Realistic text to-image diffusion models with deep language understanding .

AUTHOR: C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. Denton, S. K. S. Ghasemipour, B. K. Ayan, S. S. Mahdavi, R. G. Lopes, T. Salimans, J. Ho, D. J. Fleet, and M. Norouzi,

YEAR: 2022

DESCRIPTION: We present Imagen, a text-to-image diffusion model with an unprecedented degree of photorealism and a deep level of language understanding. Imagen builds on the power of large transformer language models in understanding text and hinges on the strength of diffusion models in high-fidelity image generation. Our key discovery is that generic large language models (e.g. T5), pretrained on text-only corpora, are surprisingly effective at encoding text for image synthesis: increasing the size of the language model in Imagen boosts both sample fidelity and image-text alignment much more than increasing the size of the image diffusion model. Imagen achieves a new state-of-the-art FID score of 7.27 on the COCO dataset, without ever training on COCO, and human raters find Imagen samples to be on par with the COCO data itself in image-text alignment. To assess text-to-image models in greater depth, we introduce DrawBench, a comprehensive and challenging benchmark for text-to-image models. With DrawBench, we compare Imagen with recent methods including VQ-GAN+CLIP, Latent Diffusion Models, and DALL-E 2, and find that human raters prefer Images over other models in side-by-side comparisons, both in terms of sample quality and image-text alignment.

TITLE: Text-to-music generation with long-context latent diffusion.

AUTHOR: F. Schneider, O. Kamal, Z. Jin, and B. Schölkopf, ‘‘Moûsai.

YEAR: 2023

DESCRIPTION: The recent surge in popularity of diffusion models for image generation has brought new attention to the potential of these models in other areas of media synthesis. One area that has yet to be fully explored is the application of diffusion models to music generation. Music generation requires handling multiple aspects, including the temporal dimension, long-term structure, multiple layers of overlapping sounds, and nuances that only trained listeners can detect. In our work, we investigate the potential of diffusion models for text-conditional music generation. We develop a cascading latent diffusion approach that can generate multiple minutes of high-quality stereo music at 48kHz from textual descriptions. For each model, we make an effort to maintain reasonable inference speed, targeting real-time on a single consumer GPU. In addition to trained models, we provide a collection of open-source libraries with the hope of facilitating future work in the field. We open-source the following: - Music samples for this paper: https://bit.ly/anonymous-mousai.

TITLE: ArtVerse: A paradigm for parallel human–machine collaborative painting creation in Metaverses.

AUTHOR: C. Guo, Y. Dou, T. Bai, X. Dai, C. Wang, and Y. Wen,

YEAR: 2023

DESCRIPTION: Currently, the development of the foundation model, metaverse, non fungible token (NFT), and other emerging technologies has brought profound effects on the whole art field, including art creation, dissemination, transaction, etc. However, there is no research focusing on the framework, methodologies, and applications of the human–machine collaborative creation in the metaverse era. Based on parallel theory, this article proposes a novel human–machine collaborative creation paradigm called ArtVerse, in which machines take on the roles

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP77-85

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

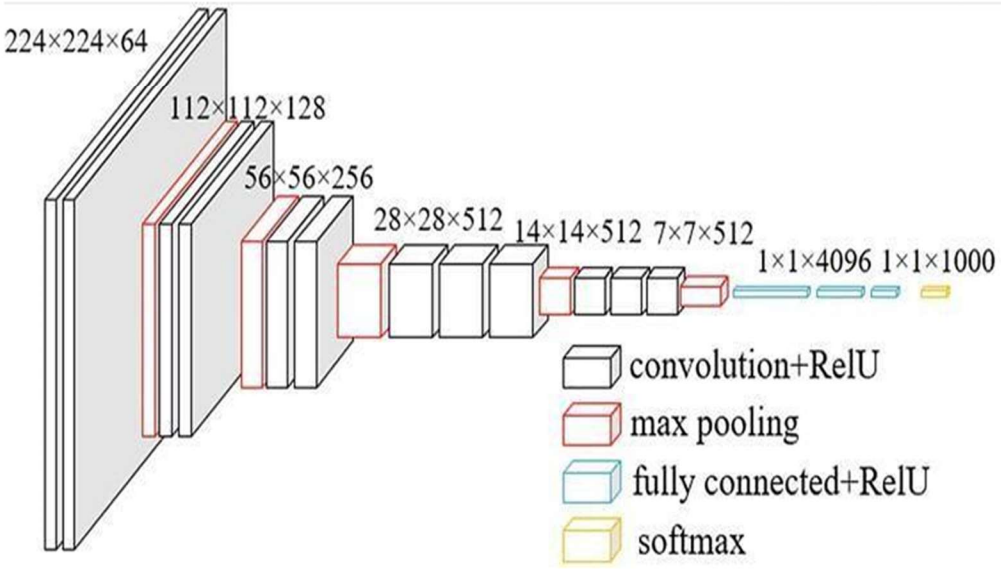
of humans to perform creation exploration and evolution and build decentralized art organizations.

DEVELOPING METHODOLOGY

The test process is initiated by developing a comprehensive plan to test the general functionality and special features on a variety of platform combinations. Strict quality control procedures are used. The process verifies that the application meets the requirements specified in the system requirements document and is bug free. The following are the considerations used to develop the framework from developing the testing methodologies.

ARCHITECTURE DIAGRAM:

SYSTEM ARCHITECTURE:



IMPLEMENTATION(ALGORITHM):

EXISTING TECHNIQUE:

Image processing and Neural Networks have been extensively used in the detection and classification of cancerous modules. Hence CNNs are more appropriate, for the task of nodule detection and classification. CNN's have more properties like multiple feature extraction.

When convolution layer, subsampling or pooling layer, fully connected layers such layers are combined, leading to Deep CNNs, it helps in increasing the accuracy of classification. The proposed CNN model

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP77-85

ISSN 2347–3657

Volume 13, Issue 2s, 2025

will be suitable for the early detection and classification.

PROPOSED TECHNIQUE:

ResNet50 is a specific variant of the Residual Network (ResNet) architecture, originally introduced by Kaiming He and colleagues in their 2015 paper, Deep Residual Learning for Image Recognition. The "50" in ResNet50 refers to the number of layers in the network, making it a relatively deep and powerful model for image recognition tasks. ResNet50 has become a popular choice due to its ability to handle deeper networks while avoiding common issues like vanishing gradients and overfitting.

The key innovation of ResNet is its use of residual connections, which allow the network to learn residual functions instead of the original unreferenced functions. This architecture enables the model to retain information from earlier layers, making it easier for the network to learn complex mappings without degrading performance.

SOFTWARE TESTING:

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of tests. Each test type addresses a specific testing requirement.

Unit Testing:

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program input produces valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .It is done after the completion of an individual unit before integration. This is a structural testing that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Functional Unit:

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals. Functional testing is centered on the following items: Valid Input : identified classes of valid input must be accepted. Invalid Input : identified classes of invalid input must be rejected. Functions : identified functions must be exercised. Output : identified classes of application outputs must be exercised. Systems/Procedures: interfacing systems or procedures must be invoked.

System Unit:

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

Performance Test:

The Performance test ensures that the output be produced within the time limits,and the time taken by the system for compiling, giving response to the users and requests being sent to the system to retrieve the results.

Integration Testing:

Software integration testing is the incremental integration testing of two or more integrated software

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP77-85

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

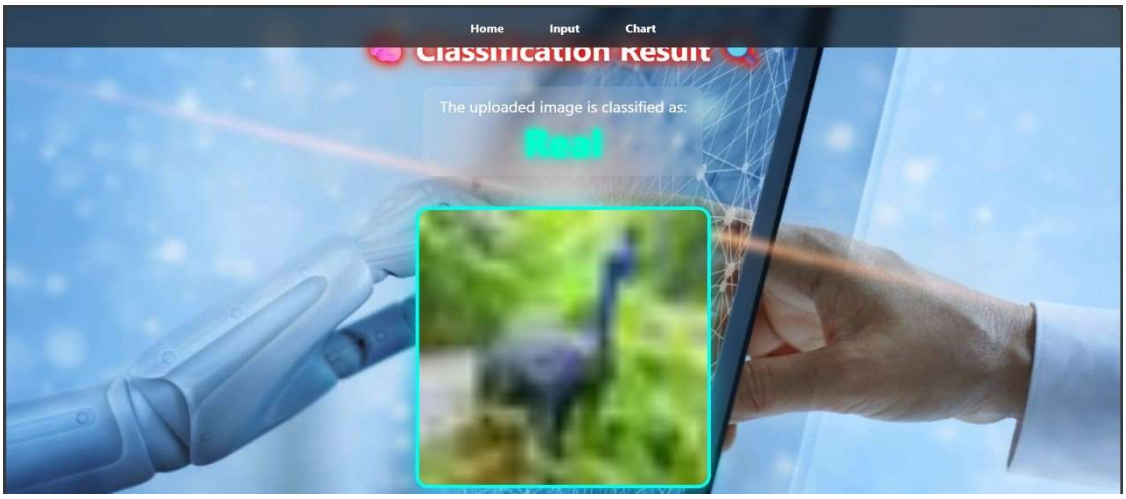
components on a single platform to produce failures caused by interface defects.

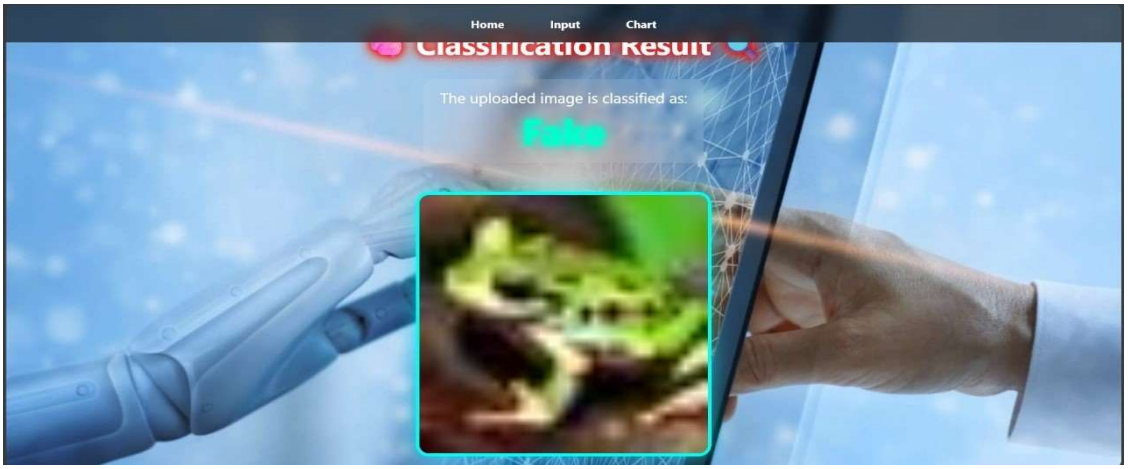
The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Acceptance Testing:

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

RESULTS / SCREENSHOTS





DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP77-85

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

CONCLUSION:

In conclusion, this study makes significant contributions to addressing the growing challenges posed by AI-generated images. By developing a deep learning-based system that effectively classifies images as either real or AI-generated, the proposed approach plays a crucial role in ensuring data authenticity and the reliability of digital content. The ability to identify subtle artifacts and inconsistencies in synthetic images not only enhances our understanding of AI-generated content but also offers practical solutions for applications in content verification and digital forensics. Moreover, the insights gained from this research provide a foundation for further exploration in explainable AI and real-time detection. The public release of the dataset, which includes both real and synthetic images, adds value to the broader research community, enabling future interdisciplinary studies aimed at advancing the detection and ethical use of AI-generated media.

FUTURE SCOPE:

Future work could focus on exploring advanced architectures, such as Vision Transformers and hybrid models, to improve classification accuracy and feature extraction. Attention-based mechanisms and explainable AI (XAI) methods could be incorporated to enhance the interpretability of the model. Additionally, updating the dataset to include images generated by newer AI models, such as StyleGAN and Stable Diffusion, will be essential as synthetic images become more realistic. Expanding the approach to specialized domains like medical imaging or deepfake detection could broaden its applicability. Moreover, real-time deployment for image classification on platforms like social media and mobile devices, as well as improving model generalization across diverse datasets, will be key for practical, scalable solutions. Finally, addressing ethical considerations, such as bias mitigation and fairness in the detection process, should be prioritized as AI-generated content continues to evolve.

REFERENCES:

1. K. Roose, ‘‘An AI-generated picture won an art prize. Artists aren’t happy,’’New York Times, vol. 2, p. 2022, Sep. 2022.
2. R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, ‘‘High resolution image synthesis with latent diffusion models,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2022, pp. 10684–10695.
3. G. Pennycook and D. G. Rand, ‘‘The psychology of fake news,’’ Trends Cogn. Sci., vol. 25, no. 5, pp. 388–402, May 2021.
4. B. Singh and D. K. Sharma, ‘‘Predicting image credibility in fake news over social

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP77-85

media using multi-modal approach," Neural Comput. Appl., vol. 34, no. 24, pp. 21503–21517, Dec. 2022.

5. N. Bonettini, P. Bestagini, S. Milani, and S. Tubaro, "On the use of Benford's law to detect GAN-generated images," in Proc. 25th Int. Conf. Pattern Recognit. (ICPR), Jan. 2021, pp. 5495–5502.
6. D. Deb, J. Zhang, and A. K. Jain, "AdvFaces: Adversarial face synthesis," in Proc. IEEE Int. Joint Conf. Biometrics (IJCB), Sep. 2020, pp. 1–10.
7. M. Khosravy, K. Nakamura, Y. Hirose, N. Nitta, and N. Babaguchi, "Model inversion attack: Analysis under gray-box scenario on deep learning based face recognition system," KSII Trans. Internet Inf. Syst., vol. 15, no. 3, pp. 1100–1118, Mar. 2021.
8. J. J. Bird, A. Naser, and A. Lotfi, "Writer-independent signature verification; evaluation of robotic and generative adversarial attacks," Inf. Sci., vol. 633, pp. 170–181, Jul. 2023.
9. A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-shot text-to-image generation," in Proc. Int. Conf. Mach. Learn., 2021, pp. 8821–8831.
10. C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. Denton, S. K. S. Ghasemipour, B. K. Ayan, S. S. Mahdavi, R. G. Lopes, T. Salimans, J. Ho, D. J. Fleet, and M. Norouzi, "Photo Realistic text to-image diffusion models with deep language understanding," 2022, arXiv:2205.11487.
11. P. Chambon, C. Bluethgen, C. P. Langlotz, and A. Chaudhari, "Adapting pretrained vision-language foundational models to medical imaging domains," 2022, arXiv:2210.04133.
12. F Schneider, O. Kamal, Z. Jin, and B. Schölkopf, "Moûsai: Text-to-music generation with long-context latent diffusion," 2023, arXiv:2301.11757.
13. F. Schneider, "ArchiSound: Audio generation with diffusion," M.S. thesis, ETH Zurich, Zürich, Switzerland, 2023.
14. D. Yi, C. Guo, and T. Bai, "Exploring painting synthesis with diffusion models," in Proc. IEEE 1st Int. Conf. Digit. Twins Parallel Intell. (DTPI), Jul. 2021, pp. 332–335.
15. C. Guo, Y. Dou, T. Bai, X. Dai, C. Wang, and Y. Wen, "ArtVerse: A paradigm for parallel human–machine collaborative painting creation in Metaverses," IEEE Trans. Syst., Man, Cybern., Syst., vol. 53, no. 4, pp. 2200–2208, Apr. 2023.
16. Z. Sha, Z. Li, N. Yu, and Y. Zhang, "DE-FAKE: Detection and attribution of fake images generated by text-to-image generation models," 2022, arXiv:2210.06998.
17. R. Corvi, D. Cozzolino, G. Zingarini, G. Poggi, K. Nagano, and L. Verdoliva, "On the detection of synthetic images generated by diffusion models," 2022, arXiv:2211.00680.
18. I. Amerini, L. Galteri, R. Caldelli, and A. Del Bimbo, "Deepfake video detection through optical flow based CNN," in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW), Oct. 2019, pp. 1205–1207.
19. D. Güera and E. J. Delp, "Deepfake video detection using recurrent neural networks," in Proc. 15th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS), Nov. 2018, pp. 1–6.
20. J. Wang, Z. Wu, W. Ouyang, X. Han, J. Chen, Y.-G. Jiang, and S.-N. Li, "M2TR: Multi-modal multi-scale transformers for Deepfake detection," in Proc. Int. Conf. Multimedia Retr., Jun. 2022, pp. 615–623.
21. P. Saikia, D. Dholaria, P. Yadav, V. Patel, and M. Roy, "A hybrid CNNLSTM model for ideo Deepfake detection by leveraging optical flow features," in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2022, pp. 1–7.
22. H. Li, B. Li, S. Tan, and J. Huang, "Identification of deep network generated images using disparities in color components," Signal Process., vol. 174, Sep. 2020, Art. no. 107616.

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP77-85

ISSN 2347–3657

Volume 13, Issue 2s, 2025

23. S. J. Nightingale, K. A. Wade, and D. G. Watson, ‘‘Can people identify original and manipulated photos of real-world scenes?’’ Cognit. Res., Princ. Implications, vol. 2, no. 1, pp. 1–21, Dec. 2017.
24. A. Krizhevsky and G. Hinton, ‘‘Learning multiple layers of features from tiny images,’’ 2009.
25. Mohammed Abdul Bari, Shahanawaj Ahamad, Mohammed Rahmat Ali,” Smartphone Security and Protection Practices”, International Journal of Engineering and Applied Computer Science (IJEACS) ; ISBN: 9798799755577 Volume: 03, Issue: 01, December 2021 (International Journal,U K) Pages 1-6
26. M.A.Bari, Sunjay Kalkal, Shahanawaj Ahamad," A Comparative Study and Performance Analysis of Routing Algorithms”, in 3rd International Conference ICCIDM, Springer - 978- 981-10-3874-7_3 Dec (2016) ; Impact Factor :4.18
27. Mohammed Shoeb, Mohammed Akram Ali, Mohammed Shadeel, Dr. Mohammed Abdul Bari, “Self-Driving Car: Using Opencv2 and Machine Learning”, The International journal of analytical and experimental modal analysis (IJAEMA), ISSN NO: 0886-9367, Volume XIV, Issue V, May/2022

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP77-85