International Conference on Intelligent Computing and Explainable AI (ICICEA) – April 2026

RESEARCH ARTICLE OPEN ACCESS

Web Based Fake Image Detection Using Machine Learning Algorithm

Archana K**, Janani G**, Mangalambigai S**,Dr.Koteeshwari RS*,

* Associate professor Electronics and Communication Engineering,

**UG Students -Electronics and Communication Engineering,

E.G.S Pillay Engineering College,Nagapattinam,

India

Abstract- Fake images have become a serious issue in the modern digital era due to the widespread use of social media and the rapid sharing of misleading content. This paper proposes a web-based system for detecting fake images using machine learning techniques. The developed website allows users to upload images and instantly verify their authenticity. The system integrates image processing methods with machine learning algorithms such as Convolutional Neural Networks (CNN) to analyze visual patterns and detect manipulations. A dataset containing both real and fake images is used to train and evaluate the model. The backend processes the uploaded image, extracts relevant features, and classifies it as real or fake, while the frontend provides an easy-to-use interface for users. Experimental results show that the system achieves high accuracy and provides quick predictions. This web-based solution can be widely used in social media platforms, digital forensics, and online content verification systems to prevent the spread of false information and improve trust in digital media

I. Introduction

In recent years, the rapid spread of digital images online has changed how we communicate and share information. However, this technology has also made it easier to create and share manipulated and misleading visual content. With powerful editing tools and advanced models like Deep fakes and GANs (Generative Adversarial Networks), it has become harder to tell real images from fake ones. The rise of fake images threatens trust, cybersecurity, journalism, political discussions, and social stability, making it necessary to develop strong and automatic detection methods.

Machine learning (ML) has become a promising way to tackle the problem of detecting fake images because it can learn complex patterns and subtle details that people might miss. Traditional rule-based detection methods have difficulty scaling and adapting to new manipulation techniques. In contrast, data-driven ML methods can effectively extract important features from large image datasets. Recent progress in deep learning, particularly with convolutional neural networks (CNNs), has shown impressive results in recognizing visual patterns, making them well-suited for spotting artifacts from image tampering.

This paper looks at how machine learning frameworks can be used to detect altered images accurately and in real time. We suggest a hybrid detection model that combines handcrafted feature extraction with deep learning classifiers to enhance robustness against various manipulation types. Our research also examines how dataset diversity and pre-processing methods can affect model performance. Through thorough experiments and analysis, we aim to provide a scalable and reliable solution for detecting fake images that can be used in real-world situations.

II. LITERATURE SURVEY

S. Beram, et al. [1] proposed a way for identify Doctoring in digital images. Doctoring typically involves several steps, usually in the sequence of initial image-processing operations such as scaling, rotation, contrast shift, smoothing, and more. These methods used are dedicated to three categories of statistical features including binary similarity, image quality and wavelet Statistics. The three categories of forensic facilities are as follows:

1. Image quality measurements: They focus on the difference between the doctored image and its source. If the original image is unavailable, this test is simulated by a vague version of the image

2. Higher order wavelet statistics: They are derived from multi-level decomposition of the image.

3. Binary similarity measurements: These measures capture the correlation and texture characteristics between and within the Bit planes of lesser importance, which are more susceptible to manipulation. To influence the detection of doctoral effects, first, single tools are developed to identify the required image-processing functions. Then, these individual "weak" detectors assembled together to determine the presence of a doctorate in an expert fusion scheme.

· Swaminathan et al. [2] proposed a method for assessing camera and post-camera operation fingerprints for verifying the integrity of the photos. A new methodology is operations. It primarily focuses on forensic analysis of digital camera images and observations based on both internal and external acquisition devices. The Internal fingerprints of several in-camera processing functions are made to assess by detailed imaging model and analysis of its components. This is the primary function of Internal fingerprints. When further processing is enforced upon the camera, the image captured is set to design as a manipulation filter so as to achieve linear time lapse estimation and to evaluate the internal fingerprints associated to the operations performed later by the camera to blind deconvolution technique. In cases when the test image is not a camera output but created by any other image processes that is due to the lack of fingerprints placed by the camera. Any diversity or change between new types of fingerprints and old projected camera fingerprints show that the image underwent through some kind of process like stenographic embedding or image tampering after initial capture.

· H. Cao et al. [3] designed a new ensemble manipulation detector to detect a wide range of manipulation types on native image patches.

Fan et al. [4] proposed integrating statistical noise features with interchangeable image file format header features for manipulation detection.

M. C. Stamm and K. J. R. Liu, [5] proposed different methods not only for the detection of global and local contrast enhancement but also for identifying the use of histogram equalization and for the identification of global addition of noise to a previously JPEG- compressed image. The methods used are as follows.

i) Identification of globally applied contrast enhancement: contrast enhancement operations appear as non-linear pixel mappings that display artifacts in the image histogram. Non-linear mapping is divided into locally mapped regions. Contract mapping maps multiple unique input pixel values to a single output pixel value. Results in spontaneous peak in the image histogram.

ii) Improve the locally applied contrast image: Contrast enhancement operation can be implemented locally to hide the visual evidence of image tampering. Localization of these activities may find evidence of cut-and-paste type forgery. Forensic techniques can be extended in a manner to detect such cut-paste forgery.

iii) Detecting histogram equalization in the image: Similar to other contrast enhancement operation, the Histogram Equalization Operation introduces spontaneous peaks and gaps in the image histogram. Methods for the detection of histogram equalization in the image have been extended. iv). Image noise detection: Additive noise can be applied to a picture globally in an attempt to eradicate visual evidence of forgery as well as legally necessary indicators of other tampering tasks. The noise detection technique can detect whether the image is in speckled noise, Gaussian sound etc.

· M. Stamm and K. Liu [6] focuses on identifying the image change, focus on the functions used to retrieve and edit the potential information about the unaltered version of the image. Even though the detection of such activities is probably not related to malicious tampering yet they doubt the image's authenticity and content. An iterative method based on the likelihood model is proposed to estimate the contrast enhancement mapping used to unify the image. The likelihood model identifies the histogram entries that occur with the corresponding growth arti facts

· P. Ferrara, et al. [7] proposed a paper in which a comparison between two forensic methods for reverse engineering of series is generated by linear contrast enhancement by interleaved double JPEG compression. His first approach was based on peak to valley behaviour of histogram double-volume Discrete Cosine Transform coefficient, while the other one was based on the distribution of first-digit Discrete Cosine Transform Coefficient. Both the methods elaborate the study of processing chain considered for the identification of chain processing and estimation of its parameters. Particularly, the proposed method provides an estimation to the linear improvement and quality factor of previous JPEG compression.

III. PROPOSED SYSTEM

A. PRE- PROCESSING

Pre-processing is an important step in **fake image detection using Machine Learning**. It prepares the input images so that the machine learning or deep learning model can analyse them effectively. Raw images collected from different sources may contain noise, different sizes, lighting variations, and unnecessary information. Therefore, pre-processing techniques are applied to clean and standardize the images before training the model.

B. RESIZE

The first step in pre-processing is **image resizing**. Images in a dataset may have different resolutions, which can affect the training process. To maintain consistency, all images are resized to a fixed dimension such as **224×224 or 128×128 pixels**. This helps reduce computational complexity and ensures that the model processes all images in the same format.

C. NORMALIZATION

Another important step is **image normalization**. Normalization scales pixel values to a standard range, usually between **0 and 1** or **−1 and 1**. This improves the learning efficiency of machine learning algorithms and helps the model converge faster during training.

It also prevents certain features from dominating due to larger numerical values.

D. NOISE REMOVAL

Noise removal is also applied during pre-processing. Images collected from the internet or cameras may contain noise or distortions that can affect the detection accuracy. Techniques such as filtering or smoothing are used to remove unwanted noise while preserving important features of the image.

E. IMAGE AUGMENTATION

Image augmentation is often used to increase the diversity of the training dataset. This technique generates new images from existing ones by applying transformations such as rotation, flipping, cropping, zooming, and brightness adjustment. Augmentation helps the model learn more robust features and reduces overfitting, especially when the dataset size is limited.

G. COLOUR SPACE CONVERSION AND FEATURE ENCHANCEMENT

Another key step is **colour space conversion and feature enhancement**. Sometimes images are converted from RGB to grayscale or other colour spaces to highlight certain patterns and textures that may indicate manipulation. Feature enhancement techniques can also be applied to emphasize edges and structural details that help identify fake or manipulated regions.

H. NOISE REMOVAL

Finally, the processed images are **converted into numerical arrays or tensors**, which can be used as input for machine learning models such as **Convolutional Neural Networks (CNNs)**. Proper pre-processing improves the quality of the input data and significantly enhances the accuracy and performance of fake image detection systems.

IV. FLOWCHART

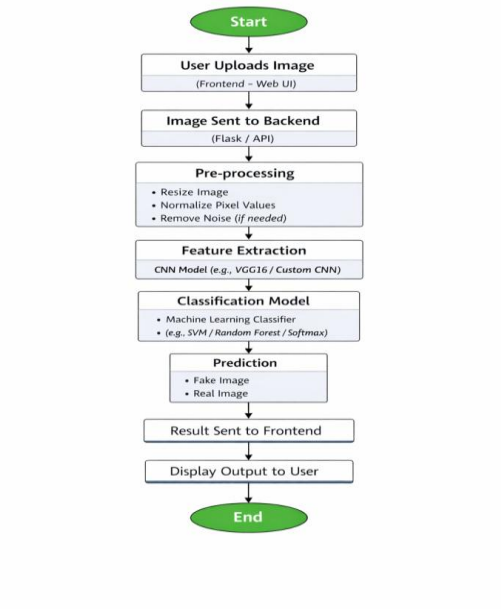


Fig.4.1 Flowchart

V. CONVOLUTIONAL NEURAL NETWORK

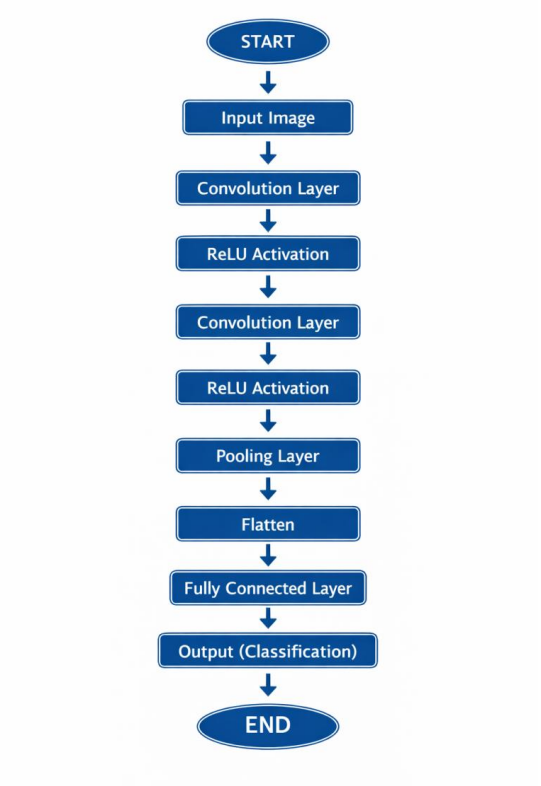


Fig.5.1 CNN Algorithm

Convolutional Neural Networks (CNNs) play a crucial role in fake image detection using machine learning by automatically learning and extracting complex visual features from images. Unlike traditional methods that rely on manually designed features, CNNs can identify subtle patterns such as noise inconsistencies, unnatural textures, pixel-level distortions, and arti facts introduced by image manipulation techniques like GANs or deep fake generation.

In a typical fake image detection system, the input image is first pre-processed (e.g., resizing, normalization, or frequency domain analysis) and then passed through multiple CNN layers. The convolutional layers extract low-level features (edges, colours) and high-level features (shapes, textures), while pooling layers reduce dimensionality and improve efficiency. The final fully connected layers classify the image as real or fake based on the learned features.

CNNs are highly effective because they can generalize across different types of manipulations and datasets. In many advanced systems, CNNs are combined with other machine learning classifiers such as Support Vector Machine (SVM) or Random Forest to further enhance accuracy and robustness. Due to their high performance and ability to detect hidden arti facts, CNN-based approaches are widely used in real-time fake image detection applications and digital forensics.

Common preprocessing techniques include:

Filtering (low-pass, high-pass, band-pass filters)
Normalization to maintain uniform signal amplitude
Noise removal using techniques like wavelet transfer
This step ensures that the signal is smooth and ready for accurate feature extraction.

The CNN model consists of:

- Convolution layers to extract patterns
- Pooling layers to reduce dimensionality
- Fully connected layers for classification

Convolutional Neural Network:

Convolutional neural networks (CNNs) are a specialised class of artificial neural networks carefully created for handling grid-like input structures. Time series data, which may be thought of as a 1D grid, or photographs, which are effectively 2D grids made up of pixels, might both be examples of these data grids. In essence, the input layer, one or more hidden layers, and the output layer are the three basic building blocks of CNNs and other feed forward neural networks.

The application of a mathematical procedure known as convolution, at least once within each of their layers, is the distinguishing feature that gives CNNs their name. Similar to how neurons in the human brain react when they come into contact with a certain stimulus, this convolution activity has a purpose. CNNs are particularly well-suited for tasks involving grid-like structures because they are highly effective at recognising patterns and features within the input data by applying convolution. In a convolution layer has been depicted for better understanding.

Every CNN also includes a layer known as pooling in addition to convolution. There are several different pooling techniques, with “max pooling” being one of the most popular. In max pooling, the network retains the maximum value while condensing the data within a rectangle neighbourhood. In order to improve the computational efficiency of the model, this pooling layer is crucial in lowering the spatial dimensions of the data. In essence, pooling aids in gathering the most important data while omitting irrelevant details, which is very useful in image processing tasks. The depicts the behaviour of a max pooling layer.

Max pooling layer:

The completely connected layer, which is a feature shared by all CNN models, is another essential element. These completely linked layers, which are frequently placed before the output layer, act as a link between the retrieved features and the ultimate prediction or classification goal. They give the network the ability to understand complex connections and relationships between the data acquired by convolution and pooling layers, enabling more precise and complex predictions.

CNNs, a potent and crucial tool in the field of deep learning, are created expressly to succeed in problems involving structured data grids, such as those involving images and time series data. In a variety of applications, from image recognition to natural language processing and beyond, they are able to extract meaningful features, reduce computational complexity, and provide exceptional performance because of their ability to perform convolutions, pooling operations, and use fully connected layers.

VI. FUTURE WORK

i)Use a **larger dataset** of real and fake images to improve the accuracy of the model.

ii)Apply **advanced deep learning algorithms** to enhance detection performance.

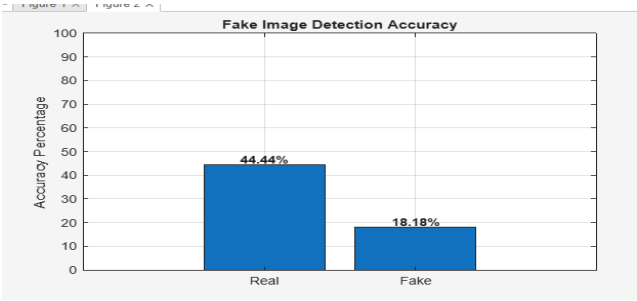
iii)Develop the system into a **real-time application** for faster image verification.

iv)Integrate the model with **web applications or mobile applications**.

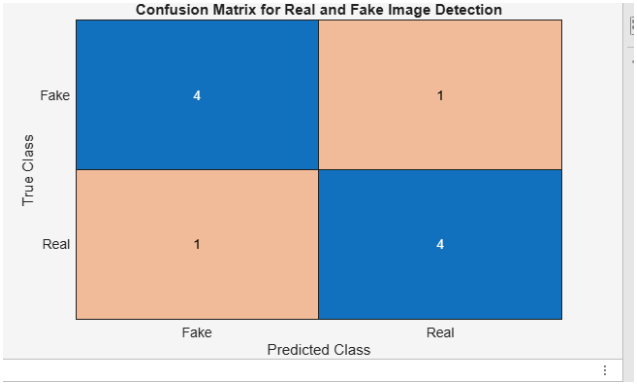
v)Improve the system to detect **different types of image manipulations** such as deep fakes and edited images.

vi)Enhance the system to provide **faster and more reliable results** for users.

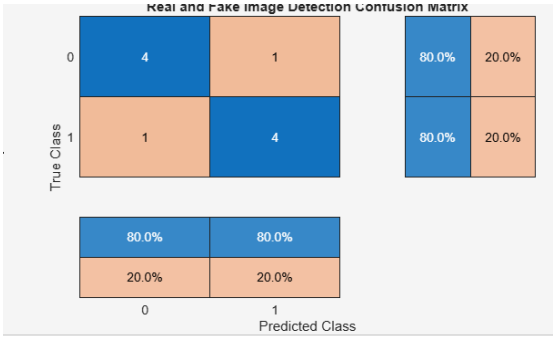
VI .RESULT AND DISCUSSION



a. Fake Image Detection Accuracy



b. Confusion martix for real and fake image detection



c. Real and fake image detection confusion matrix

VII . CONCLUSION

In conclusion, the **Fake Image Detection using Machine Learning** project aims to identify and classify images as real or fake by analyzing their visual patterns and features. With the rapid growth of digital media and image editing tools, manipulated images are becoming more common, making it important to develop reliable detection systems. Machine learning techniques, especially deep learning models such as **Convolutional Neural Networks (CNNs)**, play a significant role in automatically learning features that help distinguish fake images from genuine ones.In this project, various stages such as **data collection, pre-processing, feature extraction, model training, and testing** are implemented to build an efficient detection system. Pre-processing techniques improve the quality and consistency of the dataset, while the machine learning model analyzes patterns, textures, and inconsistencies present in manipulated images. The trained model can then

predict whether a given image is real or fake with good accuracy.

Overall, the proposed system helps in improving of manipulated or misleading images. This project demonstrates how machine learning can be effectively applied to solve real-world problems related to image verification and digital security. Future improvements can include using larger datasets, advanced deep learning architectures, and real-time detection systems to further enhance the performance and reliability of fake image detection.

VIII. REFERENCES

[1] H. Farid,
"Image Forgery Detection,"
IEEE Signal Processing Magazine, vol. 26, no. 2, pp. 16–25, 2009.

[2] A. Krizhevsky, I. Sutskever and G. E. Hinton,
"ImageNet Classification with Deep Convolutional Neural Networks,"
in Advances in Neural Information Processing Systems, 2012.

[3] I. Goodfellow et al.,
"Generative Adversarial Nets,"
in Advances in Neural Information Processing Systems, 2014.

[4] K. Simonyan and A. Zisserman,
"Very Deep Convolutional Networks for Large-Scale Image Recognition,"
arXiv preprint arXiv:1409.1556, 2014.

[5] C. Szegedy et al.,
"Going Deeper with Convolutions,"
in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[6] D. Kingma and J. Ba,
"Adam: A Method for Stochastic Optimization,"
in International Conference on Learning Representations (ICLR), 2015.

[7] J. Redmon, S. Divvala, R. Girshick and A. Farhadi,
"You Only Look Once: Unified, Real-Time Object Detection,"
in Proceedings of CVPR, 2016.

[8] K. He, X. Zhang, S. Ren and J. Sun,
"Deep Residual Learning for Image Recognition,"
in Proceedings of CVPR, 2016.

[9] M. Agarwal and H. Farid,
"Detecting Deep-Fake Videos from Facial Expressions,"
in IEEE International Workshop on Information Forensics and Security (WIFS), 2018.

[10] N. Nguyen, T. Nguyen and D. Nahavandi,
"Deep Learning for Deepfakes Creation and Detection: A Survey,"
IEEE Access, vol. 7, pp. 157–173, 2019.

[11] Y. Li and S. Lyu,
"Exposing DeepFake Videos by Detecting Face Warping Artifacts,"
in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2019.

[12] B. Dolhansky, R. Howes, B. Pflaum, N. Baram and C. Canton-Ferrer,
"The DeepFake Detection Challenge Dataset,"
in arXiv preprint arXiv:2006.07397, 2020.

[13] A. Rossler et al.,
"FaceForensics++: Learning to Detect Manipulated Facial Images,"
in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2019.

[14] H. Khalid, S. Tariq, M. Kim and S. S. Woo,
"Fake Image Detection using Frequency Domain Analysis,"
in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[15] S. Wang, O. Wang, R. Zhang, A. Owens and A. A. Efros,
"CNN-generated Images Are Surprisingly Easy to Spot… for Now,"
in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[16] T. Karras, S. Laine and T. Aila,
"A Style-Based Generator Architecture for
Generative Adversarial Networks,"
in Proceedings of the IEEE Conference on
Computer Vision and Pattern Recognition (CVPR),
2020.

[17] X. Yang, Y. Li and S. Lyu,
"Exposing Deep Fakes Using Inconsistent Head
Poses,"
in Proceedings of the IEEE International
Conference on Acoustics, Speech and Signal
Processing (ICASSP), 2020.

[18] P. Korshunov and S. Marcel,
"DeepFake Detection: A Survey,"
in Proceedings of the IEEE International
Conference on Biometrics (ICB), 2021.

[19] Z. Guo, G. Yang, J. Chen and X. Sun,
"Multi-Modal Deepfake Detection Framework,"
in Proceedings of the IEEE International
Conference on Multimedia and Expo (ICME),
2022.

[20] J. Zhao, L. Xie and X. Zhang,
"Vision Transformer for Deepfake Detection,"
in Proceedings of the IEEE Conference on
Computer Vision and Pattern Recognition (CVPR),
2023.