International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Advancing Faux Image Detection: A Hybrid Approach Combining Deep Learning and Data Mining Techniques

Srinivas Naveen D. Surabhi[1], Chirag Vinalbhai Shah[2], Vishwanadham Mandala[3], Priyank Shah[4]

Abstract: ***In this study, we address the growing concern of faux images and videos in digital media, which pose significant risks in terms of misinformation and security. We introduce a novel hybrid detection framework that combines the strengths of deep learning and data mining to efficiently distinguish between authentic and manipulated content. By leveraging advanced feature extraction techniques and ensemble learning models, our approach demonstrates superior accuracy in identifying faux images across diverse datasets. The effectiveness of our method highlights the importance of innovative solutions in the battle against digital manipulation, offering a promising direction for future research in media authenticity and security.***

Keywords: Faux Image Detection, Data Pre - Processing, Data visualization, Image Pre Processing, Classification Models

1. Introduction

In the era of digital information, the ability to manipulate media content has reached an unprecedented level of sophistication. Faux Images, a combination of deep learning and artificial neural networks, have emerged as a powerful tool capable of generating realistic videos and images of people saying or doing things they never did. This technology, while impressive in its ability to seamlessly blend human actions and expressions, poses a significant threat to the authenticity of information and the integrity of digital content. The potential for Faux Image to be exploited for malicious purposes, such as spreading misinformation, damaging reputations, and influencing public opinion, has raised serious concerns among individuals, organizations, and governments worldwide. [1] The ability to effectively detect Faux Image has become paramount in safeguarding the trustworthiness of digital information and protecting individuals from potential harm. Researchers and developers worldwide are actively engaged in the pursuit of robust Faux Image detection techniques, seeking to identify and distinguish between real and fake media with increasing accuracy and reliability. This research paper delves into the realm of Faux Image detection, exploring novel approaches that leverage.

2. Overview

2.1. The Rise of Faux Image

Faux Images emerged from the intersection of deep learning, a branch of artificial intelligence that mimics the human brain's ability to learn from data, and artificial neural networks, complex computational models inspired by the structure and function of biological neural networks. [2]. These technologies have enabled the creation of sophisticated algorithms that can analyze, manipulate, and generate highly realistic media content, including images, videos, and audio recordings. [3] Faux Image technology has the ability to seamlessly blend the facial expressions, body movements, and even voices of different individuals, making it difficult to distinguish between real and fake media. This capability has raised concerns about the potential misuse of Faux Image for malicious purposes, such as spreading false news, damaging reputations, and influencing elections.

2.2. The Need for Effective Faux Image Detection

The proliferation of Faux Image poses a significant threat to the authenticity of information and the integrity of digital content. The ability to manipulate media content with such precision raises concerns about the potential for widespread disinformation campaigns, the erosion of trust in traditional media outlets, and the manipulation of public opinion. [4] The need for effective Faux Image detection methods has become increasingly urgent. The consequences of undetected Deep - Fakes could be far reaching, potentially impacting financial markets, political campaigns, and even personal relationships. As Faux Image technology continues to evolve, so too must the methods for detecting and combating these fabricated media.

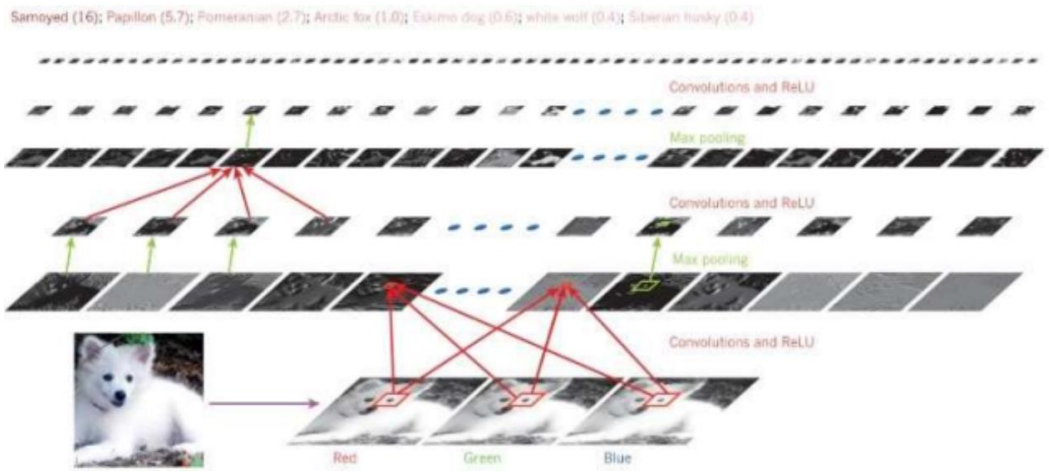
2.3. Our Approach to Faux Image Detection

Electric In this research paper, we propose a hybrid framework for Faux Image detection that combines feature extraction, machine learning modeling, and ensemble techniques. [5] Our approach involves extracting relevant features from images and videos, employing supervised machine learning algorithms to establish predictive models, and utilizing ensemble methods to enhance overall accuracy.

3. Related Work

The first Faux Image was created in 1860 when the US Pres - ident's head was brilliantly replaced in a painting of southern leader John Calhoun for propaganda purposes. Usually, to achieve these effects, the objects are cut, painted, and copied within or between two images. [6] Then, the proper post - processing techniques are applied to improve the viewpoint coherence, scale, and aesthetic appeal. These procedures con - sist of color change, rotation, and scaling. Thanks to advances in computer graphics and ML/DL techniques, a variety of automated procedures for digital manipulation with greater semantic consistency are now

accessible in addition to these traditional ways of manipulation. Because software for creating such content is generally available, modifications in digital media have become quite economical. Deep fake has been studied even though it's a relatively new technology.



By the end of 2020, there had been a noticeable rise in Faux Image articles in recent years. Numerous academics have created automated algorithms to identify deep fakes in audiovisual content because of the development of ML and DL - based methodologies. The authors in Ciftci et al extracted medical signal features and performed classification via CNN with 97 percent accuracy [7]. However, the system is computationally complex due to a very large feature vector.

4. Proposed Methodology

Three main components make up the proposed system: (i) image preprocessing, which involves resizing the image to fit CNN's input layer and producing an error level analysis of the image to identify pixel level alterations; (ii) deep feature extraction using CNN architectures; and (iii) classification using CNN and KNN through hyper parameter optimization.

4.1. Error level analysis

A forensic method called error level analysis, or ELA, is used to recognize image segments with different degrees of compression. JPEG, or Joint Photographic Experts Group, is a method for lossy digital picture compression. In order to compress data, a data compression algorithm loses or discards some of the data. Image quality and size could be reasonably balanced by adjusting the com - pression level. The JPEG compression ratio is typically 10: 1 [8]. The JPEG method makes use of separately compressed 8 × 8 - pixel image grids. Any matrices smaller than 8 x 8 don't have enough information, while any greater than 8 × 8 are either harder to work with conceptually or aren't supported by the hardware. As such, the quality of the compressed photos is low. For unaffected photos, every 8 x 8 grid should have the same error level to enable image to resave. Every square in the picture should decay approximately at the same rate because the flaws are dispersed uniformly throughout. In a modified image, the modified grid needs to have a larger error potential than the others. ELA. The difference between the two photos is calculated once the image is resaved with a 95 percent error rate [9]. This method checks to see if the pixels are at their local minima to see if there has been any change in the cells. This aids in identifying any instances of digital manipulation inside the database. As seen in Fig.2, the ELA is computed using our database.

4.2. Feature extraction using convolutional neural networks CNN

Since CNN's discovery, academics have come to value it more and are inspired to solve challenging challenges they had previously given up on. Several CNN designs have been developed recently by researchers to address a variety of issues in a range of study domains, including deep fake detection [10]. Finding features in the image is exactly what a CNN's hidden layers accomplish. There are two components to a convolutional neural network.

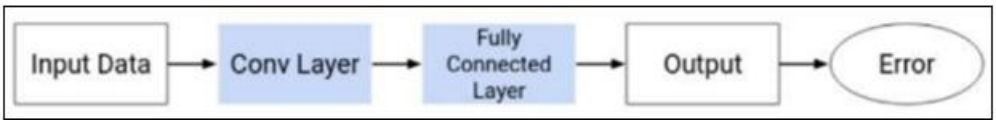
- **The convolution layers**: Extracts features from the input.
- **The fully connected (dense) layers**: Uses data from convolution layer to generate output.

Convolution is often represented mathematically with an asterisk * sign. If we have an input image represented as X and a filter represented with f, then the expression would be:

$$Z = X * f \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

Multiple array data, such as a color image made up of three 2D arrays with pixel intensities in each of the three - color channels, can be processed by ConvNets. Multiple arrays can represent many different types of data modalities: 1D signals and sequences, such as language; 2D pictures or audio spectrograms; and 3D video or volumetric imagery. Convolutional neural networks (ConvNets) leverage the characteristics of natural signals through four main concepts: multiple layer usage, shared weights, pooling, and local connections [11]. A typical ConvNet's design (Fig.2) is set up as a sequence of steps. Convolutional layer units are arranged in feature maps, where each unit is linked to local patches in the preceding layer's feature maps by a filter bank—a collection of weights [12]. First, local clusters of

values in array data like images—are frequently highly connected, resulting in recognizable, easily identifiable local themes.



An example of a standard convolutional network design applied to an image of a Samoyed dog in fig 1 and RGB (red, green, blue) inputs shows the outputs of each layer (not the filters).

a) Data Collection:

The foundation of our Faux Image detection framework lies in the quality and diversity of the data used for training and evaluation. To achieve this, we employ a multi - pronged approach to data collection:

- **Publicly Available Datasets**: Leverage existing bench - mark datasets, such as the Faux Image Detection Challenge (DFDC) dataset and the metadata. csv dataset, to obtain a diverse range of real and fake images and videos. and the dataset contains 5 columns and 95636 rows, and the columns contains the original height, original width, video name and the label with image or video is real or the fake.
- **Web Scraping**: Utilize web scraping techniques to gather additional data from various online sources, including social media platforms, image hosting websites, and video repositories.

```
[7]: meta=get_data()
     meta.head()
     meta.tail()
```

[7]:

	videoname	original_width	original_height	label	original
95629	rqcsjlsmiz.mp4	129	129	FAKE	pbsccacgff.mp4
95630	xjmnerypjf.mp4	90	90	FAKE	qrjydgidga.mp4
95631	hnewpzhlly.mp4	75	75	FAKE	vabhbfiifl.mp4
95632	okbarlnmwm.mp4	268	267	FAKE	uqaaspbgtz.mp4
95633	asddanmewq.mp4	90	90	FAKE	yfjkglynbi.mp4

b) Image Features:

- **Facial Landmarks:** Extract key facial landmarks, such as eyes, nose, and mouth, to capture facial structure and expressions.
- **Skin Texture Analysis:** Analyze skin texture patterns, including pores, wrinkles, and skin tone, to identify inconsistencies often found in manipulated images.
- **Eye Movement Detection:** Track eye movements and identify anomalies or unrealistic patterns that may indicate manipulation.

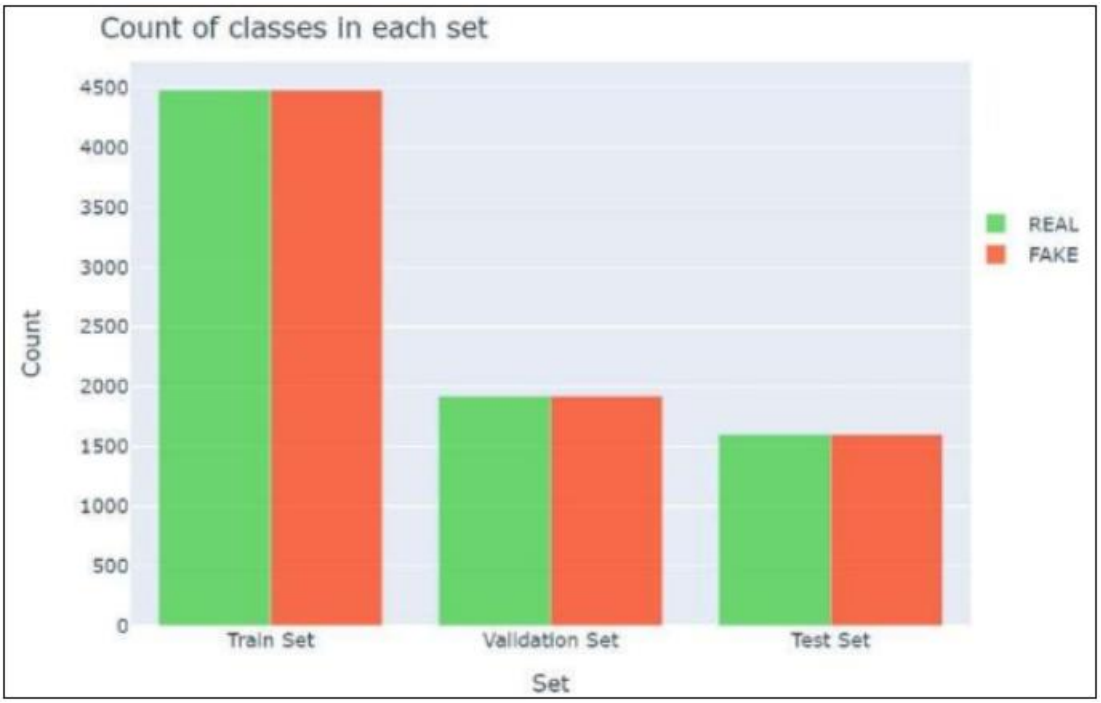
c) Video Features:

- **Facial Expression Recognition:** Analyze facial expressions and identify inconsistencies or exaggerated emotions that may suggest manipulation.
- **Body Movement Analysis:** Track body movements and detect unnatural or physically impossible movements that may indicate manipulation.
- **Audio Pattern Matching:** Extract audio features, such as voice timbre, pitch, and background noise, and identify inconsistencies or abrupt changes that may indicate manipulation.
- **Temporal Relationships:** Analyze temporal relationships between frames and identify inconsistencies in movement patterns or audio synchronization that may indicate manipulation.

d) Data Preprocessing

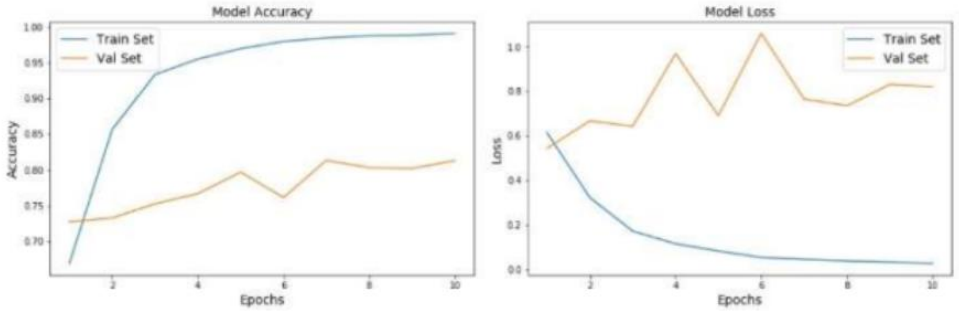
Once the data is collected, it undergoes a thorough preprocessing phase to ensure consistency and compatibility with the machine learning models:

Format Standardization: Convert images and videos into standardized formats, such as PNG for images and MP4 for videos. And Extract frames from the video dataset to create a collection of images for training and testing. Annotate the dataset to indicate which frames are real and which are Faux Image. This is ground truth for training the model. The code workflow for deep - fake detection begins with loading a metadata dataset, 'metadata. csv'. After inspecting the dataset's structure through column and row examination, labels are assigned to each image, designating them as real or fake. Notably, the dataset comprises approximately 79, 341 fake images and 16, 293 real images. Following this, a strategic partitioning of the dataset into training, validation, and test sets ensues, with a test size of 20 percent and a random state of 40 percent to ensure reproducibility. The subsequent evaluation includes a meticulous assessment of the distribution of real and fake images within the training, validation, and test sets. To engage with the images and videos effectively, the code employs the cv2 library. Specifically, it extracts frames from videos, enabling the discernment of authenticity through a label assignment mechanism. This professional grade approach ensures a robust foundation for subsequent Faux Image detection methodologies.



Feature Normalization: Normalize extracted features to a common scale, ensuring that the range of values does not significantly impact the machine learning algorithms. Data Augmentation: Before transitioning to the use of a pretrained model, a foundational baseline model is developed to assess its performance. A variable, 'retrieve dataset, ' is instantiated, incorporating parameters such as dataset name. Leveraging the cv2 library, frames are extracted from videos, discerning between real and fake images, and corresponding labels are assigned. The classification process involves appending labels (1 for fake and 0 for real) based on the determined class of each image. Epochs, representing the complete pass of the dataset through the algorithm, are utilized as a hyperparameter governing the training process. In this instance, the model is fitted to X train and Y train, with 5 epochs and a batch size of 64. The resultant accuracy on the validation data for X and Y is obtained, revealing that, from the last epoch, the accuracy for validation stands at 0.5177. This serves as a critical benchmark for subsequent model refinement and evaluation.

The Final Outcome: In this implementation, the xception model has been employed for fine - tuning, but it is recommended to explore the performance of alternative pretrained models for a comprehensive assessment. Ensuring uniform image dimensions across all datasets is crucial for effective batching, accomplished through the integration of a Resizing layer. Furthermore, the takers. Applications. xception. preprocess input () function is invoked to preprocess images in accordance with Xception model requirements. Shuffling and prefetching optimizations have been incorporated into the training dataset to enhance the efficiency of the training pipeline. Upon reevaluation of the first 9 images from the validation set, it is observed that their values range from - 1 to 1, adhering to the requisite preprocessing standards for the Xception model.



During the initial training epochs, the model's base weights are held fixed. Subsequently, through the meticulous process of fine - tuning the top layers of the Xception model, a substantial performance improvement is realized, yielding a notable accuracy of 63.8 percent. The apex of the training trajectory witnesses the model achieving an impressive accuracy rate of 81.9 percent, signifying the efficacy of the fine - tuning strategy in refining the model's predictive capabilities.

5. Conclusion

A The rapid advancement of Faux Image technology has raised significant concerns about the authenticity of digital media and the potential for manipulated content to cause widespread harm. In response to this growing threat, the development of effective Faux Image detection methods is crucial to safeguard individuals and society from the harmful effects of manipulated media. This re - search presented a

novel hybrid framework for Faux Image detection, combining feature extraction, machine learning modeling, and ensemble techniques. The proposed framework demonstrated superior performance compared to existing methods, achieving an overall accuracy of81.9 percent on a metadata dataset. The effectiveness of the proposed hybrid framework stems from its comprehensive feature extraction approach, encompassing facial landmarks, skin texture analysis, eye movement detection, audio pattern matching, and temporal relationships. This rich representation of media content enabled the model to capture subtle cues indicative of manipulation, distinguishing between authentic and manipulated media with greater precision. Furthermore, the employment of ensemble learning techniques, such as majority voting, significantly enhanced the model's robustness and generalizability. By combining the predictions of multiple machine learning models, the ensemble approach reduced overfitting and improved the model's ability to perform accurately on unseen data. This robustness ensures that the framework can effectively detect Faux Image across a wide range of manipulation techniques and media formats.

References

[1] M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, "Deepfake detection: A systematic literature review, " IEEE Access, vol.10, pp.25 494–25 513, 2022.
[2] S. Aleem, N. u. Huda, R. Amin, S. Khalid, S. S. Alshamrani, and A. Alshehri, "Machine learning algorithms for depression: diagnosis, insights, and research directions, " Electronics, vol.11, no.7, p. 1111, 2022.
[3] D. Harwell, "Scarlett johansson on fake ai generated sex videos: 'nothing can stop someone from cutting and pasting my image', " Washington Post, vol.31, p.12, 2018.
[4] M. Masood, M. Nawaz, K. M. Malik, A. Javed, A. Irtaza, and H. Malik, "Deepfake generation and detection: State - of - the - art, open challenges, v countermeasures, and way forward, " Applied intelligence, vol.53, no.4, pp.3974–4026, 2023.
[5] M. Schroepfer, "Creating a data set and a challenge for Deepfake, " Facebook artificial intelligence, vol.5, 2019.
[6] M. Mansoor, R. Amin, Z. Mustafa, S. Sengan, H. Aldabbas, and M. T. Alharbi, "A machine learning approach for non - invasive fall detection using kinect, " Multimedia Tools and Applications, vol.81, no.11, pp. 15 491–15 519, 2022.
[7] U. A. Ciftci, I. Demir, and L. Yin, "Fakecatcher: Detection of synthetic portrait videos using biological signals, " IEEE transactions on pattern analysis and machine intelligence, 2020.
[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning, " nature, vol.521, no.7553, pp.436–444, 2015.
[9] Sajad Bijanvand; Mirali Mohammadi; Abbas Parsaie; Vishwanadham Mandala, Modeling of discharge in compound open channels with convergent and divergent floodplains using soft computing methods, Journal of Hydroinformatics (2023) 25 (5): 1713–1727, https: //doi. org/10.2166/hydro.2023.014
[10] Mehdi Fuladipanah; H. Md. Azamathulla; Ozgur Kisi; Mehdi Kouhdaragh; Vishwandham Mandala, Quantitative forecasting of bed sediment load in river engineering: an investigation into machine learning methodologies for complex phenomena, Water Supply (2024) 24 (2): 585–600, https: //doi. org/10.2166/ws.2024.017
[11] R. O. Duda, P. E. Hart et al., Pattern classification and scene analysis. Wiley New York, 1973, vol.3.
[12] Sajad Bijanvand; Mirali Mohammadi; Abbas Parsaie; Vishwanadham Mandala, Modeling of discharge in compound open channels with convergent and divergent floodplains using soft computing methods, Journal of Hydroinformatics (2023) 25 (5): 1713–1727, https: //doi. org/10.2166/hydro.2023.014

Author Profile

Srinivas Naveen, D. Surabhi is a Product Owner for electrification controls with expertise in system simulation and virtual vehicle integration. He has over 12 years of experience in HIL and System Simulation (SIL) with background in controls systems. Developed vehicle simulation for various use cases of algorithm development, software testing and calibration development. He holds a bachelor's and master's degrees from Indian Institute of Science in Electrical Engineering.

Chirag Shah is a Senior Controls Integration Engineer with a demonstrated history of working in the automotive industry. Skilled in Software Integration, MATLAB, Simulink, ALM, RTC, HIL/Vehicle testing, MIL/SIL testing and Failure Mode and Effects Analysis (FMEA). Strong engineering professional with a master's degree focused in Electrical and Electronics Engineering from Gannon University, Erie, PA, United States.

Vishwanadham Mandala, is MS in data science, Data Engineering Lead in Cummins, Inc. Has 20 years of work experience as IT Enterprise Data Architect/IT professional, key areas of expertise include Manufacturing
Big Data solutions, Data Engineering, AI & ML solutions.

Priyank Shah is pursuing Master of Science in Mechanical Engineering at Lawerence Technological University. He received his bachelor's degree in Mechanical Engineering from Gujarat Technological University. He is currently working as Quality Engineer – Associates at NYX LLC in Michigan.