IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DETECTING FAKE NEWS IN SOCIAL MEDIA

N.S Kavitha[1] , S. Rithika[2] , M. Samyuktha [3] , S. Shanmugapriya [4]

[1]Associate Professor, Department of CSE, Sri Ramakrisna Institute of Technology, Coimbatore, Tamil Nadu , India
[2 3 4]UG students, Department of CSE, Sri Ramakrisna Institute of Technology, Coimbatore, Tamil Nadu , India

Abstract— Image manipulation and forgery present significant challenges across various domains, including forensics, security, and media authentication. This is a novel approach that integrates Error Level Analysis (ELA) with Convolutional Neural Networks (CNNs) to differentiate between authentic and tampered images. Using the Cassia v2 dataset containing both real and fake instances, diverse data augmentation techniques, such as flattening, resizing, and converting images to ELA format, are applied to enhance model robustness. The dataset is partitioned into 80% training and 20% validation sets to facilitate comprehensive model training and evaluation. Utilizing Keras, a Sequential model is developed, incorporating Conv2D, MaxPooling2D, Dropout, Flatten, and Dense layers for effective feature extraction and classification. Training utilizes the Adam optimizer for parameter optimization. Evaluation metrics, including loss, accuracy, and a confusion matrix, are employed to assess model performance. Results demonstrate promising accuracy, with the model achieving 98.8% training and 92.8% validation accuracy, showcasing the efficacy of the proposed methodology in accurately distinguishing between real and fake images. This approach holds potential for applications in image forensics, security, and authentication domains.

Key Terms: Fake images, ELA, CNN, Sequential model, Adam optimizer

I. INTRODUCTION

In today's digital age, the proliferation of image manipulation tools has made it increasingly challenging to discern between authentic and tampered images. Whether for journalistic integrity, legal evidence, or personal security, the ability to verify the authenticity of digital images is paramount. This project addresses this pressing issue by leveraging advanced techniques from the fields of image forensics and deep learning.

(CNNs) to differentiate between authentic images and those that have undergone various forms of manipulation, including copy-move and splicing operations. The Cassia v2 dataset serves as the foundation for training and validating the model, providing labelled examples of both real and fake images.

By employing data augmentation techniques and training a Sequential model with CNN layers, the project aims to automate the process of image authentication. The Adam optimizer is utilized to optimize model parameters, while evaluation metrics such as loss, accuracy, and confusion matrices are employed to assess the model's performance comprehensively.

Ultimately, the goal of this project is to develop a robust and accurate system capable of predicting and classifying whether an image is authentic or manipulated. Such a system holds significant potential for applications in various domains, including journalism, law enforcement, and digital media authentication.

II. SCOPE OF THE PROJECT

The scope of this project encompasses the development and implementation of a system for distinguishing between authentic and tampered images using Error Level Analysis (ELA) and Convolutional Neural Networks (CNNs). The project focuses on detecting common forms of image manipulation, including copy-move and splicing operations.

Key aspects of the project include data collection and preparation, where a diverse dataset of labeled images will be gathered and preprocessed to ensure uniformity and quality. Data augmentation techniques will be applied to enhance the model's robustness.

The project involves constructing a Sequential model using Keras, featuring CNN layers designed to effectively extract features from images and classify them as authentic or tampered. The model will be trained using the Adam optimizer and evaluated using metrics such as loss, accuracy, and confusion matrices.

Testing the trained model on unseen data will assess its real-world performance and usability. Documentation of the development process and potential extensions or enhancements, such as integrating additional detection techniques or optimizing for real-time performance, will also

be part of the project scope.

III. EXISTING SYSTEM

Existing systems that differentiate between legitimate and manipulated photos include a wide range of approaches and methodologies. Some systems use conventional image analysis techniques, such as pixel-level analysis or statistical methods, to detect irregularities or anomalies in images that could indicate manipulation. Manual involvement and professional expertise are often necessary in these approaches, making them labor-intensive and subjective.

Other systems use machine learning methods, particularly deep learning models such as convolutional neural networks (CNNs), to automatically learn and identify patterns related with picture alteration. These algorithms can achieve excellent accuracy in differentiating between legitimate and tampered photos, especially when trained on huge datasets of labelled samples.

Furthermore, the efficiency of existing systems can be influenced by factors such as the quality and diversity of the training data, the complexity of the detection algorithms, and the versatility of the performance evaluation measures. Furthermore, certain algorithms may struggle to generalize to new or unknown types of tampering, resulting in lower accuracy and dependability in real-world circumstances.

Overall, while existing image authentication systems have made great progress, there are still obstacles and limits that must be addressed to increase their effectiveness, robustness, and practical usefulness in a variety of fields. Continued research and development efforts are required to progress the state-of-the-art in image forensics and authentication.

IV. LITERATURE SURVEY

Ravi Shankar, et al. [1] presented a comprehensive approach to detecting image manipulation that includes three proposed methodologies: metadata analysis, error level analysis, and the use of a machine learning algorithm. The authors emphasize the growing prevalence of image manipulation and the importance of effective detection techniques, especially in the context of social media and digital communication. It discusses the importance of detecting image forgery in combating misinformation and false propaganda. It also describes the proposed algorithm's components, such as Error Level Analysis and the use of a Convolutional Neural Network for transfer learning. The results of applying transfer learning to the VGG16 model are presented, demonstrating the approach's effectiveness in detecting image manipulation. It also highlights the potential for future research in applying the proposed model to various multimedia and video content. Overall, it provides a comprehensive overview of the challenges and methodologies for detecting image manipulation using machine learning.

Y. Patel, et al. [3] presented an enhanced dense convolutional neural network (D-CNN) architecture for deepfake picture detection. It solves the issue of recognizing deepfake photos from a variety of sources and resolutions. The proposed model is trained on a dataset that includes 10,000 genuine photographs and 5,000 deepfake images, with the goal of creating a balanced set. The study assesses the model's performance using accuracy, precision, recall, and F1 score measures, yielding an accuracy of 97.2% on the test dataset. The suggested architecture reads input images with a height and width of 160 pixels and uses a variety of data augmentation techniques.

Y. K. Zamil, et al. [4] proposed a fusion method to combat fake news on social media, with a focus on enhanced detection and interpretability. It provides a model that combines text and picture features by utilizing pre-trained models such as Electra and XLnet for text feature learning, ELA for image feature extraction, and EfficientNetB0 for image learning. In addition, the study uses the Local Interpretable Model-agnostic Explanations (LIME) method to improve the proposed model's interpretability. The results show that the combination of text and picture features, as well as the usage of ELA and LIME, gives a more reliable approach for detecting fake news than previous strategies. The study experiments with three popular datasets, including Weibo, MediaEval, and CASIA, and highlights the importance of multi-transformers and multimodal fusion for improved performance in fake news detection. The proposed model outperforms single-modal models and emphasizes the significance of interpretability and confidence in the model's predictions.

V. PROPOSED SYSTEM

The suggested approach uses a mix of Convolutional Neural Networks (CNNs) and Error Level Analysis (ELA) to improve the detection of legitimate and manipulated photos. The Cassia v2 dataset, which is tagged with '0' for authentic photos and '1' for fraudulent ones, will be utilized by the system to undergo thorough preprocessing and augmentation of data.

The Sequential model from Keras that the system will use consists of CNN layers such as Conv2D, MaxPooling2D, Dropout, Flatten, and Dense layers. Accurate categorization is made possible by this architecture, which is designed to efficiently extract information from images. During training, the Adam optimizer will be employed to achieve effective parameter optimization.

Metrics including loss, accuracy, and confusion matrix will be used to assess the system's performance. For a thorough model evaluation, the dataset will be divided into 80% training and 20% validation sets. The system will be able to predict and categorize whether an image is real or fake using this approach, with an expected accuracy of 92.8% for validation and 98.8% for training.

With potential applications in the image forensics, security, and media authentication domains, the suggested system seeks to provide a robust solution for identifying legitimate and altered images by merging ELA analysis with CNNs and employing rigorous assessment procedures.

Error Level Analysis (ELA):

Error Level Analysis (ELA) is a forensic technique used to identify areas of an image that have been subjected to compression or manipulation. The algorithm compares the error levels of different regions within an image to detect inconsistencies caused by manipulation or compression artifacts.

Sequential Model

The Sequential model in keras is a core technique for quickly and easily building deep learning models. As a linear stack of layers, it makes it easier to design neural networks by allowing layers to be added sequentially. This easy technique makes it accessible to both new and seasoned practitioners, allowing for rapid development and testing.

Adam Optimizer:

The Adam optimizer is a widely used optimization algorithm in deep learning. Combining features from AdaGrad and RMSProp, it dynamically adjusts the learning rate for each parameter based on gradient magnitudes. This adaptability facilitates faster convergence and efficient training, particularly in scenarios involving sparse gradients or noisy data. With its versatility and effectiveness, Adam optimizer has become a preferred choice for optimizing neural network models.

VI. SYSTEM ARCHITECTURE

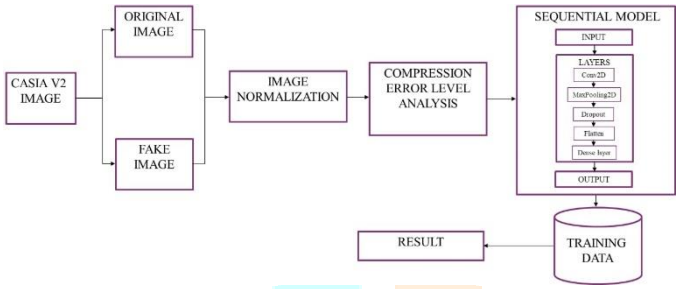

Figure 1: System architecture of the proposed system

Figure 1 depicts the steps to effectively distinguish between authentic and tampered images using Error Level Analysis (ELA) and Convolutional Neural Networks (CNNs).

1. Dataset Preparation:

- Begin with the unlabelled Cassia v2 dataset, consisting of various images portraying different scenes and subjects.

2. ELA Format Application:

- Apply Error Level Analysis (ELA) to the preprocessed images to highlight compression level discrepancies across regions.
- ELA aids in identifying potential areas of manipulation like copy-move or splicing, offering crucial insights into image authenticity.

3. Data Augmentation:

- Enhance dataset robustness through:
- Flattening: Ensuring consistent dimensions across all images.
- Resizing: Standardizing image sizes.
- Converting to ELA format: Highlighting compression discrepancies.

4. Model Construction:

- Utilize a Sequential model architecture from the Keras library.
- Incorporate Conv2D layers for 2D convolutional operations, extracting spatial features.
- Apply MaxPooling2D layers for downsampling and reducing spatial dimensions.
- Integrate Dropout layers for regularization, preventing overfitting by deactivating neurons.
- Include Flatten layers to transform the output into a one-dimensional vector.
- Implement Dense layers for learning high-level features and making predictions.
- Optionally add BatchNormalization layers for stabilizing and accelerating convergence.
- Utilize activation functions to introduce non-linearity into the model.

The overall summary of the model is depicted in the figure 2.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 124, 124, 32)      2432

 conv2d_1 (Conv2D)           (None, 120, 120, 32)      25632

 max_pooling2d (MaxPooling2D  (None, 60, 60, 32)       0
 )

 dropout (Dropout)           (None, 60, 60, 32)        0

 flatten (Flatten)           (None, 115200)            0

 dense (Dense)               (None, 256)               29491456

 dropout_1 (Dropout)         (None, 256)               0

 dense_1 (Dense)             (None, 2)                 514

=================================================================
Total params: 29,520,034
Trainable params: 29,520,034
Non-trainable params: 0
_________________________________________________________________
```

Figure 2 Sequential model summary

5. Training with Adam Optimizer:

- Train the model using the Adam optimizer, which iteratively updates parameters to minimize loss.
- During training, the model learns to classify images as real or fake based on extracted features.

6. Evaluation:

- Assess the model's performance using metrics like loss, accuracy, and confusion matrix.
- Split the dataset into 80:20 training and validation sets to evaluate generalization ability.
- The confusion matrix identifies classification discrepancies, providing insights into model performance.

7. Prediction:

- Once trained and evaluated, the model predicts and classifies input images as real or fake.
- By analyzing extracted features and comparing them to learned patterns, the model achieves high accuracy in image classification.

VII. RESULT AND ANALYSIS

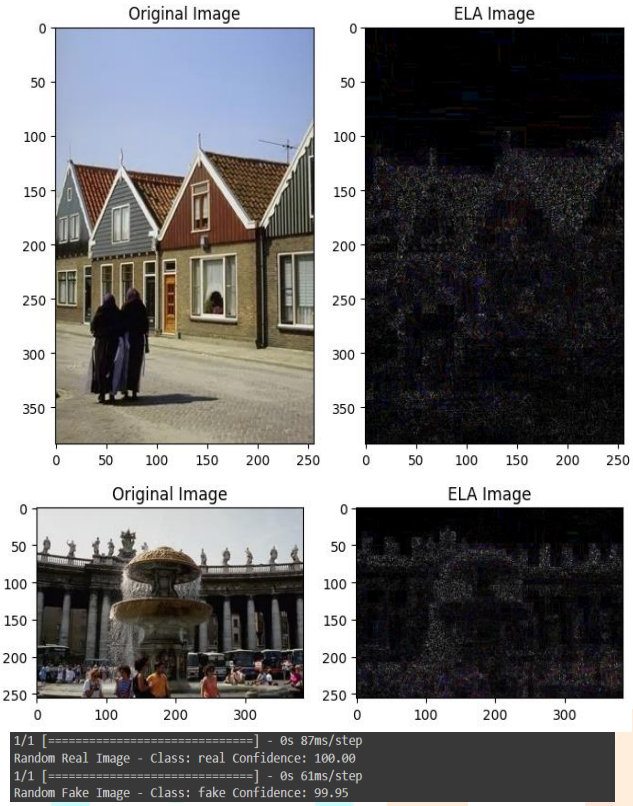

Figure 3 Random real and fake image class prediction

The figure 3 displays one real and one fake image randomly selected from the CASIA2 dataset and then displays their Error Level Analysis (ELA) representations. ELA is a technique used to highlight areas of potential manipulation or tampering in digital images by comparing the error levels between original and recompressed versions. The ELA images provide a visual indication of potential areas of interest for further analysis.

After displaying the ELA images, the code tests a model's predictions on the selected images. The model classifies the images into two categories: real or fake. For the real image, the model assigns a class of "real" with 100% confidence, indicating high certainty in its classification. Similarly, for the fake image, the model assigns a class of "fake" with a confidence of 99.95%, suggesting strong confidence in its prediction as well.

Overall, the output demonstrates the effectiveness of the model in accurately classifying both real and fake images based on their features. The high confidence scores suggest that the model has learned meaningful patterns and features from the training data, enabling it to make reliable predictions on unseen images. This capability holds significant promise for applications in image authentication and forensics, where the ability to distinguish between authentic and tampered images is crucial.

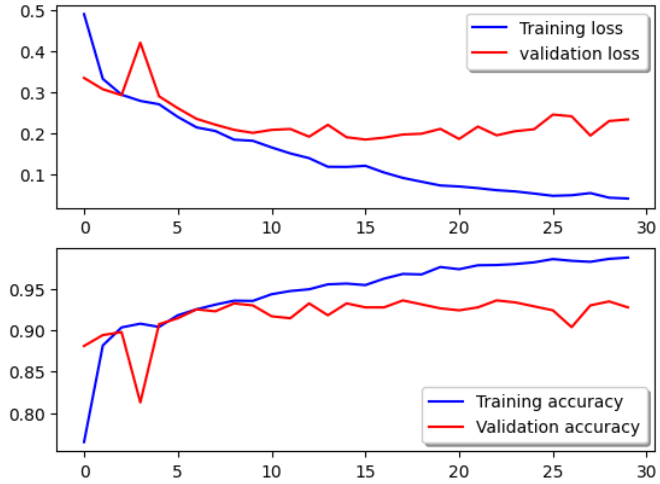

Figure 4 Loss and accuracy curves for training and validation

Figure 4 displays the training and validation performance of a machine learning model throughout numerous epochs. The blue line in the top plot represents training loss, which steadily decreases as the model learns from the training data. Concurrently, the red line indicates the validation loss, which falls but not substantially, showing effective generalization to previously unseen data. The reasonable gap between the training and validation loss curves shows that the model is not overfitting, which is a desirable outcome in model training.

The blue line in the bottom plot represents training accuracy, which gradually increases as the model generates increasingly accurate predictions on the training data. Similarly, the red line shows the validation accuracy improving, albeit less prominently than the training accuracy. Again, the reasonable gap between the training and validation accuracy curves indicates satisfactory generalization of the model.

Overall, the plots demonstrate that the model is training effectively and generalizing well to new data. However, it's worth considering experimenting with additional epochs and exploring different hyperparameters to potentially further enhance the model's performance. These actions could help ensure that the model converges optimally and achieves even better accuracy and loss results.

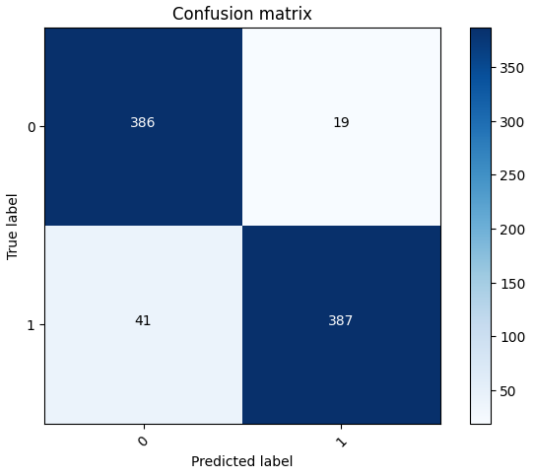

Figure 5 Confusion Matrix

The confusion matrix in figure 5 highlights the exceptional performance of the image classification model on the validation dataset, boasting an impressive accuracy of 98.8%. The model demonstrates remarkable proficiency in accurately classifying

images, with 386 real and 387 fake images correctly identified. Such high accuracy underscores the effectiveness of the model in distinguishing between authentic and tampered images with remarkable precision.

Moreover, the model's precision and recall values, exceeding 90% for both classes, further affirm its robustness and reliability. Despite the presence of a few false negatives and false positives, the overall diagonal alignment of the confusion matrix indicates that the model predominantly makes correct predictions. This alignment emphasizes the model's consistency and its ability to reliably discern between real and fake images across various manipulation types.

In essence, the confusion matrix showcases the exceptional capabilities of the image classification model, highlighting its potential for real-world deployment in image forensics, security, and authentication domains. With such high accuracy and precision, the model offers a promising solution for accurately identifying tampered images and preserving the integrity of digital content.

VIII. CONCLUSION

The Error Level Analysis (ELA) algorithm has demonstrated exceptional performance in predicting the authenticity of images, achieving impressive accuracies of 98.8% for training and 92.8% for validation datasets. This underscores ELA's effectiveness in discerning between real and fake images, making it a compelling choice for image authentication tasks.

Moreover, the versatility of ELA is evident in its capability to handle images of various sizes and types. Through the process of flattening and resizing, images can be seamlessly converted to the ELA format, enabling the algorithm to generalize effectively to unseen data. This adaptability ensures robust performance across diverse image datasets, enhancing the reliability of the authentication process.

The Sequential model from Keras, comprising Conv2D, MaxPooling2D, Dropout, Flatten, and Dense layers, serves as a powerful tool for feature extraction and categorization. Leveraging the strengths of this model, we achieved precise categorization of images, further boosting the accuracy of the authentication system.

By integrating the nuanced detection capabilities of ELA with the efficiency and adaptability of the Sequential Model, we have developed a highly reliable tool for detecting real and fake images. With applications extending beyond image forensics, this tool holds promise for addressing various challenges in fields such as security, media authentication, and beyond. Overall, our approach represents a significant advancement in image authentication technology, providing a robust solution for identifying and combating image manipulation and forgery.

IX. REFERENCES

[1]Ravi Shankar, Akshat Srivastava, Gurunath Gupta, Rohini Jadhav, Umesh Thorate , “Fake Image Detection Using Machine Learning” in IJCRT Volume 8, Issue 5 May 2020.

[2] A. H. Khalil, A. Z. Ghalwash, H. A. -G. Elsayed, G. I. Salama and H. A. Ghalwash, "Enhancing Digital Image Forgery Detection Using Transfer Learning," in IEEE Access,vol.11,pp.91583-91594,2023,doi:10.1109/ ACCESS.2023.3307357.

[3] Y. Patel et al., "An Improved Dense CNN Architecture for Deepfake Image Detection," in IEEE Access, vol. 11, pp. 22081-22095, 2023, doi: 10.1109/ACCESS.2023.3251417.

[4] Y. K. Zamil and N. M. Charkari, "Combating Fake News on Social Media: A Fusion Approach for Improved Detection and Interpretability," in IEEE Access, vol. 12, pp. 2074-2085, 2024, doi: 10.1109/ACCESS.2023.3342843.

[5] L. Ying, H. Yu, J. Wang, Y. Ji and S. Qian, "Fake News Detection via Multi-Modal Topic Memory Network," in IEEEAccess,vol.9,pp.132818-132829,2021,doi:10.1109/ ACCESS.2021.3113981.

[6] L. Ying, H. Yu, J. Wang, Y. Ji and S. Qian, "Multi-Level Multi-Modal Cross-Attention Network for Fake News Detection," in IEEE Access, vol. 9, pp. 132363-132373, 2021, doi: 10.1109/ACCESS.2021.3114093.

[7] J. Kang, S. -K. Ji, S. Lee, D. Jang and J. -U. Hou, "Detection Enhancement for Various Deepfake Types Based on Residual Noise and Manipulation Traces," in IEEE Access, vol. 10,pp.69031-69040,2022,doi:10.1109/ACCESS.2022. 3185121.

[8] K. Zhang, Y. Liang, J. Zhang, Z. Wang and X. Li, "No One Can Escape: A General Approach to Detect Tampered and Generated Image," in IEEE Access, vol. 7, pp. 129494-129503, 2019, doi: 10.1109/ACCESS.2019.2939812.

[9] Z. Wang and J. Jing, "Pixel-Wise Fabric Defect Detection by CNNs Without Labeled Training Data," in IEEE Access, vol.8,pp.161317-161325,2020,doi: 10.1109/ACCESS.2020.3021189.

[10] M. A. Hoque, M. S. Ferdous, M. Khan and S. Tarkoma, "Real, Forged or Deep Fake? Enabling the Ground Truth on the Internet," in IEEE Access, vol. 9, pp. 160471-160484, 2021, doi: 10.1109/ACCESS.2021.3131517.

[11] D. Li, H. Guo, Z. Wang and Z. Zheng, "Unsupervised Fake News Detection Based on Autoencoder," in IEEE Access, vol.9,pp.29356-29365,2021,doi:10.1109/ACCESS.2021. 3058809.