ISSN 2085-4579

Image Splicing Forgery Detection using Error Level Analysis and CNN

Irmawati[1], Naufal Ikhsan Rizki Prasetya[1]
[1] Information System, Universitas Multimedia Nusantara, Tangerang, Indonesia
irmawati@umn.ac.id, naufal.prasetya@student.umn.ac.id

Accepted on December 11th, 2023
Approved on December 18th, 2023

***Abstract*— The issue of image forgery through splicing has become increasingly relevant in the current digital era. Splicing involves the manipulation of images by combining parts of two or more different images to create a deceptive composite image. This technique can be employed for various purposes, including the dissemination of false information, damaging someone's reputation, or even creating confusion in specific contexts. Several techniques used to detect splicing involve statistical analysis, color analysis, and texture analysis. Additionally, artificial intelligence developments, such as deep learning, have been applied to enhance detection capabilities. In this study, we employed a Convolutional Neural Network (CNN) model to identify image deviations caused by splicing. Optimization was performed on the convolutional layers of the model to improve CNN performance. The integration of Error Level Analysis (ELA) was also implemented to aid in identifying splicing forgeries, where portions of one image are combined with parts of another. Areas that have undergone splicing may exhibit noticeable differences in error levels. The dataset utilized for this research was sourced from DVMM and CUISDE. The validation accuracy results for our CNN model before incorporating ELA were 61% for DVMM and 74% for CUISDE. After adding ELA, the CNN model demonstrated improved detection accuracy, achieving validation rates of 72% for DVMM and 71% for CUISDE.**

***Index Terms*—image forgery detection; splicing; error level analysis; cnn**

I. INTRODUCTION

In the contemporary digital era, the pivotal role of images in disseminating information across diverse platforms, including social media, healthcare, television, and various online applications, cannot be overstated. The ubiquitous availability of image editing tools and software on portable devices, such as smartphones and laptops, has facilitated the manipulation of images for various purposes, rendering them easily accessible. While images may undergo editing for benevolent purposes, intentional alterations with malicious intent are categorized as manipulation or forgery. Manipulative practices may involve concealing crucial information, such as obscuring individuals or objects within the image. Such manipulated images are occasionally employed as deceptive evidentiary material in legal proceedings, for financial gain through heightened engagement on social media, or for attaining popularity in the media sphere. Consequently, authenticating the integrity of images assumes paramount significance in thwarting the dissemination and endorsement of misinformation. Furthermore, validating the authenticity of images is imperative to deter the reliance on edited visual content as legal evidence [1].

Image forgery detection techniques can be broadly classified into two main categories: active and passive methods. Active techniques involve incorporating specific information into the image during its creation. This may encompass embedding watermarks or digital signatures as deliberate markers or signatures indicating the image's authenticity. The objective of active techniques is to furnish authentication or verification, thereby rendering it more challenging to forge or tamper with the image without modifying or eliminating these embedded features [2].

Passive detection techniques, such as those employed in identifying copy-move and splicing, are preferred for their efficiency. Unlike active techniques, passive methods do not require incorporating additional information into the image during its creation. Instead, they rely on analyzing the intrinsic characteristics and patterns within the image itself. This makes passive techniques quicker to execute, as they do not involve the processing overhead of adding extra information to the image. Passive methods are particularly effective in detecting certain types of image manipulations without the need for additional embedded features [3].

Image manipulation, particularly through techniques like copy-move or splicing, involves a straightforward procedure of copying and pasting elements within the image. When this pasting operation occurs, it introduces clear structural changes to the original image. The micro-texture pattern present inside the pasted area and along its boundaries undergoes modifications, creating differences and irregularities that become apparent along the edges of the altered region [4].

The Convolutional layer comprises a sequence of kernels or filters adaptable for extracting local features from the input. Each kernel performs calculations on a feature map. The Pooling layer also called the downsampling layer, reduces the resolution of the preceding feature maps. Pooling introduces invariance to minor transformations and distortions by dividing inputs into separate regions with a specified size, generating one output from each region [5]. The Fully Connected layer is typically deployed at the network's conclusion for classification purposes. Diverging from pooling and convolution, it constitutes a global operation, aggregating input from the feature extraction stages and comprehensively analyzing the output from all preceding layers [6].

Numerous studies have been conducted on image forgery detection, as referenced in [7][8][9]. The existing research indicates that the outcomes in detecting image splicing forgery remain suboptimal. Therefore, our objective is to investigate this phenomenon using the model we have developed.

This study proposes a deep learning algorithm incorporating a CNN and ELA to identify images subjected to splicing manipulation. The model is introduced and evaluated using image-splicing datasets available on the internet, including the DVMM and CUISDE datasets.

II. Research Materials

A. *Deep Learning*

Deep learning is a machine learning technique that leverages neural networks. It is characterized by multiple processing layers structured to extract progressively intricate features from the data. The hierarchical architecture of deep learning networks enables them to autonomously learn and represent complex patterns and features as they progress through the layers. This characteristic makes deep learning particularly effective for tasks such as image and speech recognition, where the data exhibits hierarchical and intricate structures [10].

B. *CNN*

CNN are a type of neural network widely recognized for their exceptional accuracy in image classification tasks. CNNs are organized into several layers, comprising three key components: the Convolutional layer, the Pooling layer, and the Fully Connected layer. Tailored for tasks such as image recognition and classification, CNNs derive their effectiveness from the architectural design of these layers—specifically, the Convolutional, Pooling, and Fully Connected layers. This design plays a pivotal role in achieving high accuracy in tasks of this nature [11].

C. *Error Level Analysis (ELA)*

ELA is acknowledged as a significant method for detecting image alterations. This method involves saving the image at specific compression levels and assessing the variance resulting from the compression. When an image is initially saved as a JPEG, compression takes place, facilitated by various editing software tools such as Adobe Lightroom, GIMP, and Adobe Photoshop. ELA serves as a technique to emphasize differences between authentic and manipulated images by analyzing the error levels introduced during compression [7]. ELA is a forensic method involving the recompression of an image with a predetermined error rate after the initial compression using lossy techniques. The fundamental principle is to measure the absolute difference between the original and recompressed images under controlled error conditions. These calculated differences can reveal inconsistencies or variations that may arise during image manipulation or forgery [12].

III. Related Work

Detection methods for image forgery are primarily designed to identify irregular patterns that should not be present in manipulated images. Two approaches exist for detecting image forgery: active and passive [13]. Several researchers have studied detecting copy-move and image splicing using convolutional neural network algorithms. In the paper by Mallick et al. [8]. CNN is employed with various models such as ELA, VGG16, and VGG19 to detect copy-move and splicing. The method was tested using CASIA V2 and NC2016 datasets, yielding accuracy values for the ELA model of 70%, VGG16 of 71%, and VGG19 of 72%.

In the following research conducted by Vijayalakshmi K et al. [7], ELA is employed to detect copy-paste images. The method was tested using the MICCF200 dataset. The MICCF200 dataset underwent an augmentation process in this study to maximize the image size. The dataset produced an accuracy value of 99% for detecting copy-paste images.

In a subsequent study by Pandey et al. [14], ELA was utilized to identify tampered images. The methodology was tested using the CASIA v2 dataset, and in this research, a localization process was applied to the CASIA v2 dataset. The outcomes of the dataset localization process resulted in an accuracy value of 88% for detecting tampered images.

The research conducted by Muniappan et al. [9] employed a CNN to identify copy-move and splicing occurrences. The proposed method underwent testing on the MICCF2000, CASIA V1, and CASIA V2 datasets. Specifically, on the MICCF2000 dataset, CNN was utilized to detect copy-move images, yielding an accuracy of 76%. For the CASIA V1 dataset, CNN was employed to identify splicing images, achieving an accuracy of 79%. Additionally, on the CASIA V2 dataset, CNN was utilized to detect copy-move and splicing images, demonstrating a notable accuracy of 89%.

IV. METHODOLOGY

CNN are advanced deep learning networks expertly trained for various computer vision applications. A notable advantage of CNNs lies in their adept utilization of local spatial coherence within input images, facilitating parameter sharing and reducing overall weight. Typically, a CNN consists of three key layers: the convolutional layer, the pooling layer, and the fully connected layer. Each of these layers performs a unique function in the overall processing of the network [15].

The convolutional layer consists of a sequence of kernels or filters that can be tailored to extract local features from the input. Each kernel is utilized to conduct computations on a feature map. The pooling layer, also known as the downsampling layer, diminishes the resolution of the preceding feature maps. Pooling introduces invariance to minor transformations and distortions by partitioning the inputs into distinct regions with a specified size, thereby generating one output from each region [5]. The fully connected layer is commonly employed after the network for classification purposes. Diverging from pooling and convolution, it constitutes a global operation. This layer aggregates input from the feature extraction stages and comprehensively analyzes the output from all preceding layers [6].

Building on prior research, we adopt ELA image processing [7]. This approach has exhibited a high accuracy score, prompting our interest in evaluating its performance on a model that we have developed.

This study's methodology comprises six processes: data collection, preprocessing, data splitting, modelling, model evaluation, and model optimization. Data collection involved accessing the Columbia University repository, specifically utilizing the datasets titled "Columbia Image Splicing Detection Evaluation Dataset (DVMM)" and "Columbia Uncompressed Image Splicing Detection Evaluation Dataset (CUISDE)." Subsequently, during the preprocessing stage, the dimensions of images across all datasets were adjusted to 224 x 224 pixels and converted to the .jpg format.

After preprocessing and data splitting, the dataset was divided into training, testing, and validation data. The subsequent step involved modelling the dataset. Following the modelling phase, the model underwent evaluation using the test data, and subsequent optimization was conducted to enhance the model's outcomes.

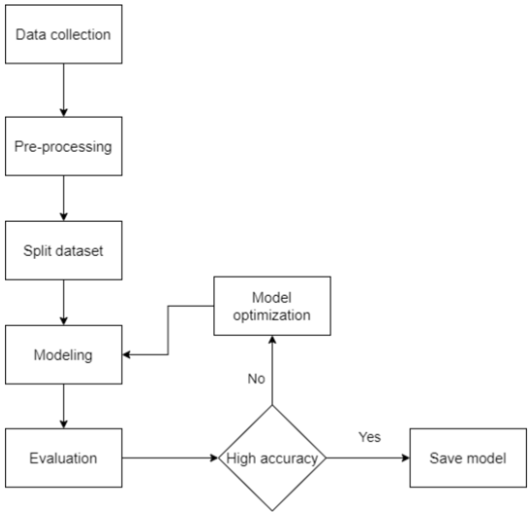


Fig. 1. Research Methodology

A. Data Collection

In this study, datasets were obtained from the GitHub website under the titles "Columbia Image Splicing Detection Evaluation Dataset (DVMM)" and "Columbia Uncompressed Image Splicing Detection Evaluation Dataset (CUISDE)." The dataset comprises images categorized into two groups: "Au", representing original images, and "Sp", representing images tampered with.

The DVMM Dataset includes a total of 1,845 images, with 933 being original images and 912 being tampered images. Similarly, the CUISDE Dataset comprises 365 images, with 184 original images and 181 tampered images.

B. Pre-Processing

The deep learning algorithm necessitates a consistent and standardized image format and size to ensure effective model training. Consequently, all datasets utilized in this research underwent a process of resizing and conversion. The images were resized to 224 × 224 pixels and converted to the .jpg format.

Augmentation was implemented for the CUISDE dataset. This decision was motivated by the limited size of the image data and the need to maximize accuracy. The augmentation technique employed involved adding horizontal flips to diversify the dataset. Before processing the data, pre-processing is essential to undergo a cleaning process to eliminate duplicate entries and rectify or remove inconsistent and incomplete data. Transformation data transformation converts or consolidates data into a predefined format [16].

C. Split Dataset

Additionally, the two datasets, featuring the categories 'Au' and 'Sp,' were partitioned into training, test, and validation data. Precisely, 10% of each category was assigned to the test data, while the remaining 90% was split in an 80:20 ratio for training data and validation data, respectively.

TABLE I. SPLIT DATASET

Dataset	Number of Dataset					
	Training		Testing		Validation	
	Au	Sp	Au	Sp	Au	Sp
DVMM	587	574	94	92	252	246
CUISDE	131	129	19	18	33	33

D. Modelling

The pre-processed dataset will be utilized to construct a model compatible with the CNN architecture. CNN provides flexibility in determining the desired number of convolutional layers. The activation function employed in the CNN is the Rectified Linear Unit (ReLU). The CNN model employed in this study consists of 5 layers. There are 32 filters in the initial convolutional layer, followed by 64 filters in the second and third convolutional layers. The subsequent layer employs 128 filters in the fourth and fifth convolutional layers. The kernel size for each convolutional layer is 3x3. Additionally, the pooling window on the pooling layer is set at 2x2

TABLE II. CNN STRUCTURE MODEL

Layer	Type	Activation Function	Output Shape	Kernel Size	Total Filter
1	input	ReLu			
2	2D Convolution	ReLu	224	3	32
3	2D Max Poling	ReLu	224	3	32
4	2D Max Poling	ReLu	109	3	64
5	2D Convolution	ReLu	54	3	64
6	2D Max Poling	ReLu	52	3	64
7	2D Convolution	ReLu	26	3	64
8	2D Max Poling	ReLu	24	3	128
9	2D Convolution	ReLu	12	3	128
10	2D Max Poling	ReLu	10	3	128
11	Flatten	-	5		
12	Dropout	-	3200		
13	Dense	-	3200		

E. Evaluation

After successfully creating the model, a performance evaluation is carried out to assess the accuracy value. This evaluation involves examining various metrics, including accuracy and F1 score. The metrics are calculated using a confusion matrix and ROC curve. If the accuracy results are suboptimal, the subsequent step optimises model performance.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$f1 - Score = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (2)$$

F. Optimization Model

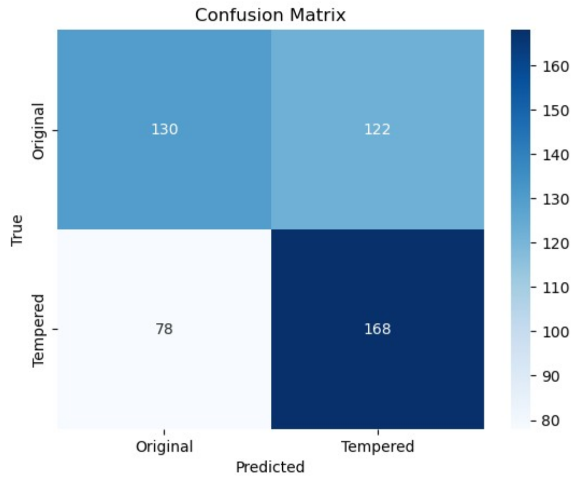
Optimization steps are implemented after creating and evaluating the model if the achieved accuracy is deemed unsatisfactory. This optimization involves making adjustments, such as adding or modifying convolutional layers, adjusting the optimizer, and tweaking optimizer parameters. Multiple simulations of various experiments in model creation are conducted to identify the optimal configuration leading to the highest accuracy value.

A low accuracy value was observed in the simulation conducted on the CNN model with four layers. In contrast, a notably improved accuracy was achieved when employing a CNN model with five layers. Consequently, the optimization strategy is focused on utilizing a CNN architecture with five layers.

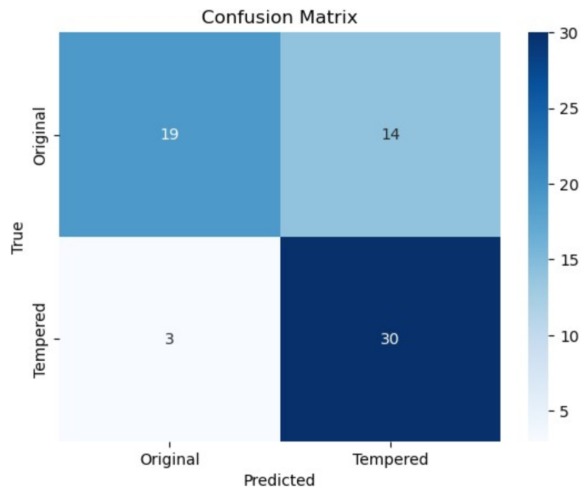
V. RESULT AND DISCUSSION

We conducted experiments using the CNN architecture to develop a model for detecting forgery in spliced images. Subsequently, we integrated CNN with ELA to assess their accuracy values following the compression of the dataset using ELA. The training of the CNN model involved using an early stopping feature, a batch size of 10, and an error level of 90%, which was applied to ELA.

The evaluation of the generated models included assessing their performance through a confusion matrix, providing a comprehensive breakdown of the performance of each model. The confusion matrix was generated in the evaluation phase using validation data specific to each dataset.



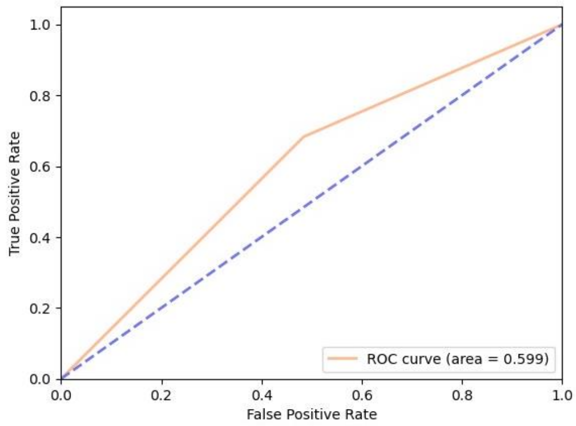
(A) DVMM



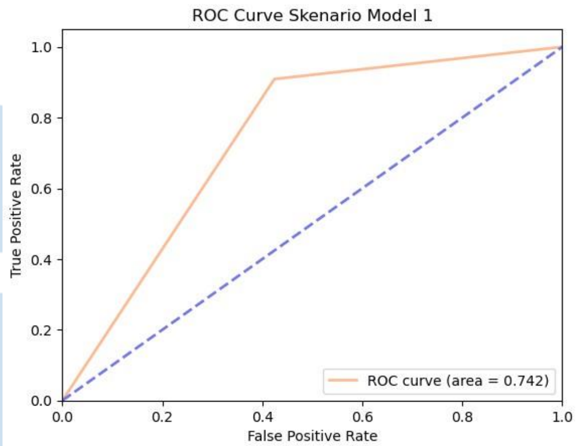
(B) CUISDE

Fig. 1 Confusion matrix CNN Models for validation data of DVMM and CUISDE dataset

Based on the provided confusion matrix, the validation accuracy value for our CNN model is 61% for DVMM and 74% for CUISDE. The F1 score values are 0.565 for DVMM and 0.690 for CUISDE. The performance for DVMM is lower than that of CUISDE, possibly due to the DVMM dataset containing grayscale or black-and-white images. Using a small batch size can impact the training process on this dataset, potentially affecting the attainment of a high accuracy value.



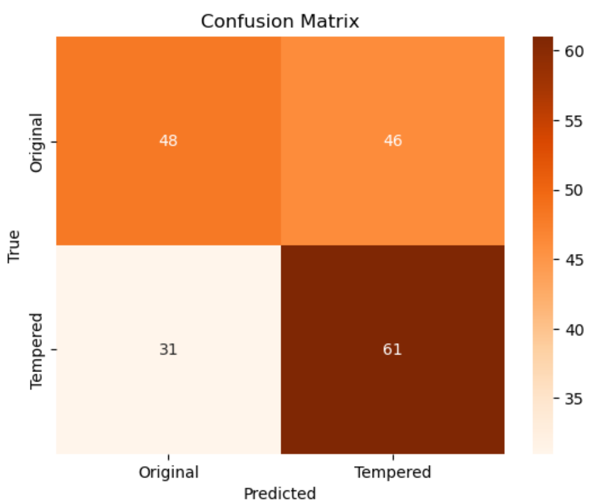
(A) DVMM

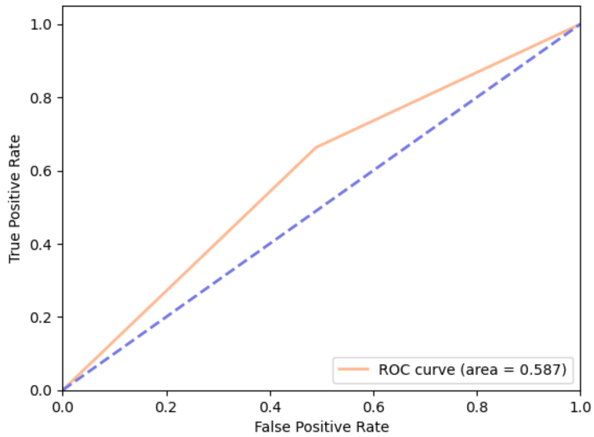


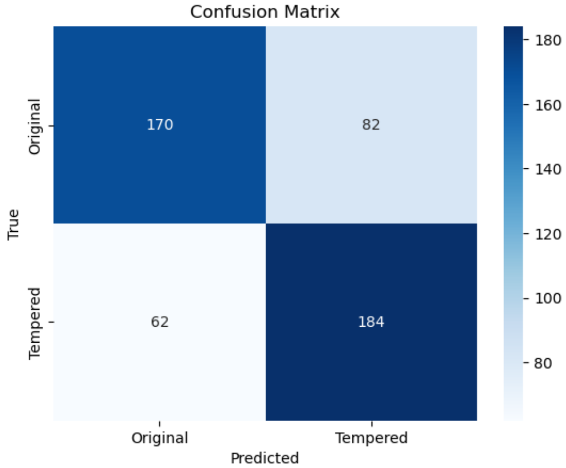
(B) CUISDE

Fig. 2 ROC Curves using CNN models validation dataset

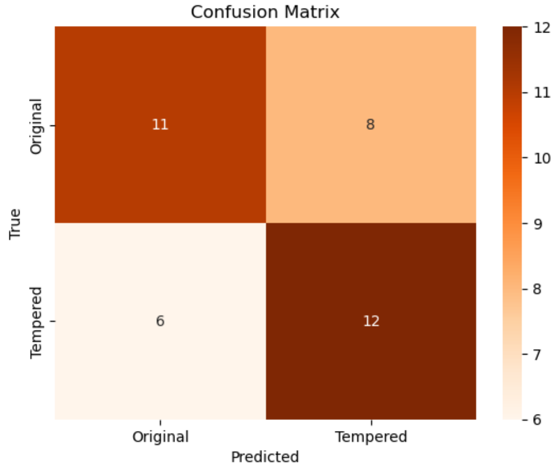
The performance of each model is also evident from the ROC curve. As depicted in the figure above, the ROC curve for CUISDE is superior to that of DVMM, consistent with the findings of the confusion matrix, which indicates higher performance for CUISDE compared to DVMM.

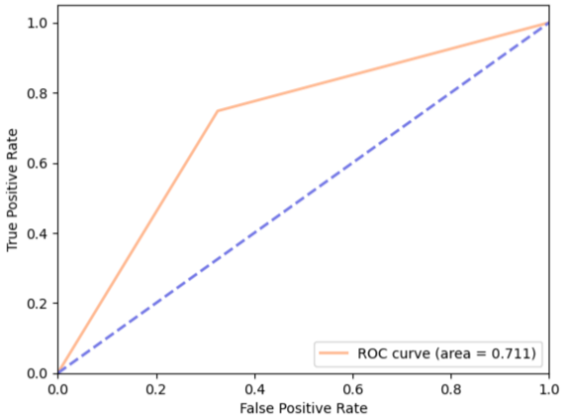




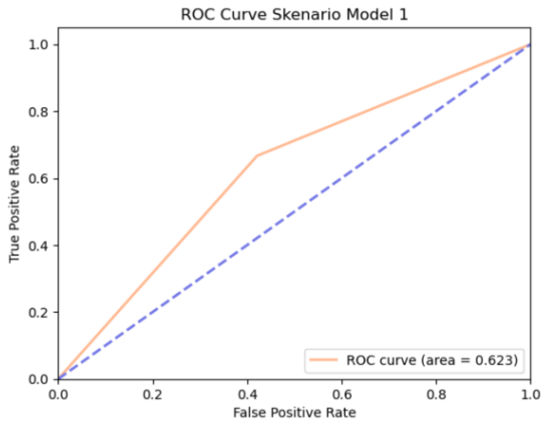


(A) DVMM





(A) DVMM+ELA



(B) CUISDE

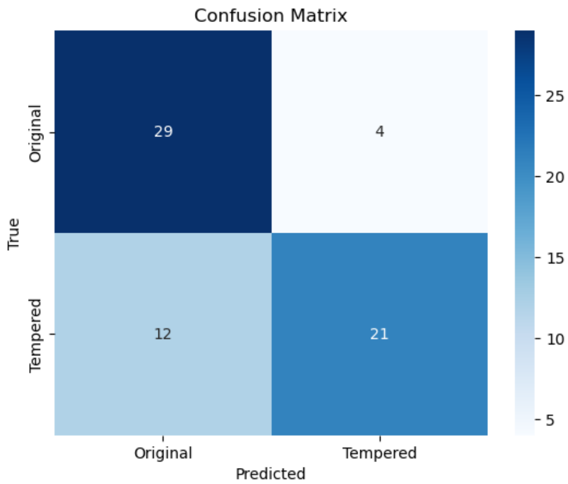
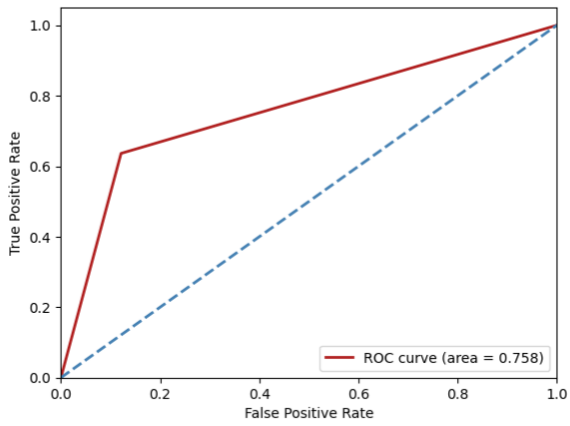


Fig. 3 Confusion matrix CNN Models for testing data of DVMM and CUISDE

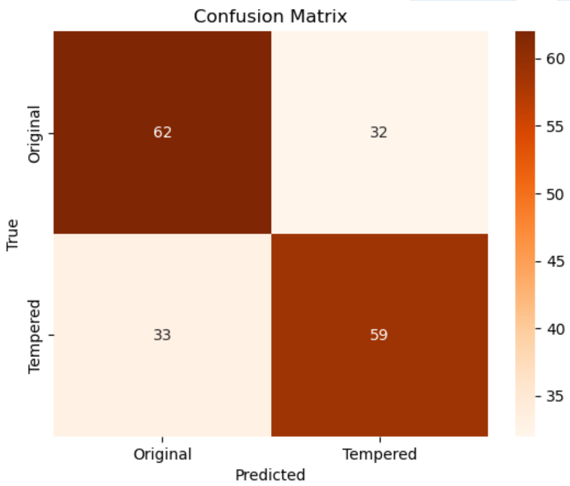
According to the presented confusion matrix, the testing accuracy values for our CNN model are 59% for DVMM and 64% for CUISDE.

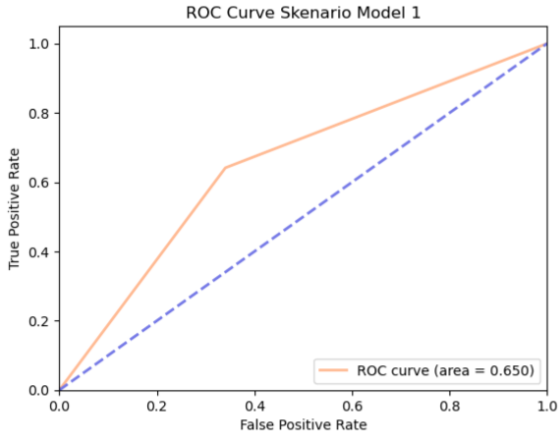


(B) CUISDE+ELA

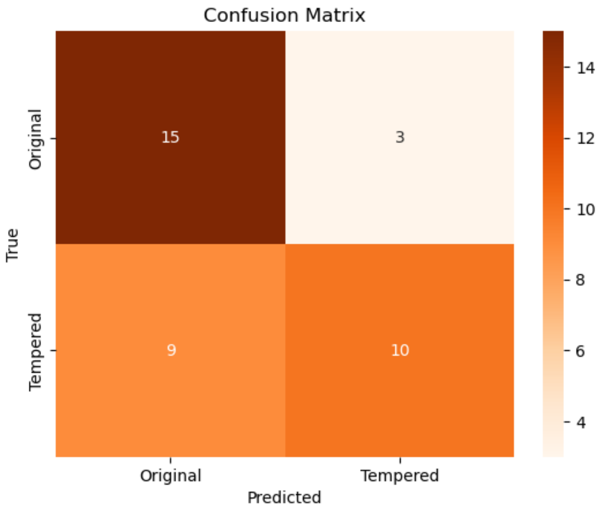
Fig. 4 Confusion matrix and ROC Curva CNN + ELA Models for validation data of DVMM and CUISDE

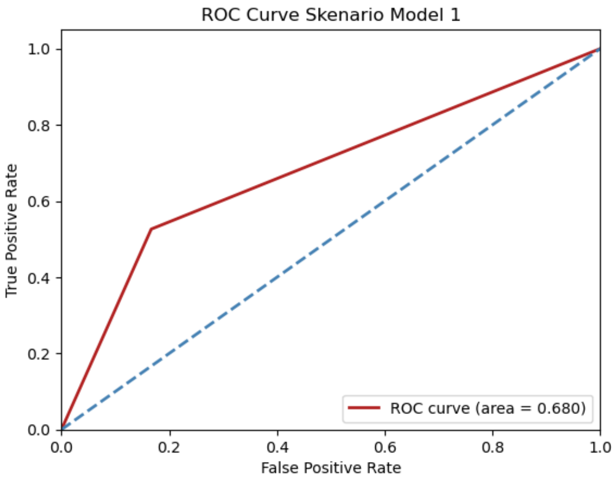
Based on the presented confusion matrix, the validation accuracy values for our CNN+ELA model are 72% for DVMM and 71% for CUISDE. The F1 score values are 0.702 for DVMM and 0.783 for CUISDE. In the CUISDE dataset, we implemented augmentation by enabling horizontal flip to improve the accuracy value.





(A) DVMM+ELA





(B) CUISDE+ELA

Fig. 5 Confusion matrix and ROC Curva CNN + ELA Models for testing data of DVMM and CUISDE

According to the provided confusion matrix, the testing accuracy values for our CNN+ELA model are 66% for DVMM and 72% for CUISDE. The resulting ROC curve values are 0.587 for DVMM and 0.623 for CUISDE.

TABLE III. THE PERFORMANCE RESULT OF OUR CNN MODEL

Dataset	Evaluation Metrics		
	Val Accuracy[%]	**F1-Score**	**ROC-AUC**
DVMM	61.24	0.565	0.599
CUISDE	74.24	0.690	0.742
DVMM +ELA	72.09	0.702	0.650
CUISDE +ELA	71.21	0.783	0.758

TABLE IV. THE PERFORMANCE COMPARISON OF OUR CNN MODEL AND RELATED WORK

Work	Dataset	Method	Accuracy (%)
Mallick[8]	CASIA V2 NC2016	ELA	70.00
		VGG 16	71.00
		VGG 19	72.00
Vijayalakshmi K et al[7]	MICC-F200	ELA	99.00
Pandey & Mitra [14]	CASIA V2	ELA+CNN	88.00
Proposed	**DVMM CUISDE**	**CNN**	**61.00**
		CNN	**74.00**
		CNN+ELA	**72.00**
		CNN+ELA	**71.00**

V. CONCLUSION

In this study, a CNN model augmented with ELA was developed to detect image splicing, resulting in an increased validation accuracy of 72% from the initial 61% for DVMM. For CUISDE, the validation accuracy changed slightly from 74% to 71%, with an improvement in testing results from 64% to 72%. In conclusion, this research successfully applied CNN with ELA to identify image splicing, enhancing accuracy in the DVMM and CUISDE datasets. These findings underscore the potential of deep learning models in addressing challenges associated with digital image manipulation. As research progresses, ongoing efforts will be directed towards refining and expanding the capabilities of the CNN model for more accurate and reliable image splicing detection.

ACKNOWLEDGMENT

The authors would like to thank the support provided by Universitas Multimedia Nusantara during this study.

REFERENCES

[1] N. K. Hebbar and A. S. Kunte, "TRANSFER LEARNING APPROACH FOR SPLICING AND COPY-MOVE IMAGE TAMPERING DETECTION.," *ICTACT J. Image Video Process.*, vol. 11, no. 4, 2021.

[2] B. Ahmed, T. A. Gulliver, and S. alZahir, "Image splicing detection using mask-RCNN," *Signal, Image Video Process.*, vol. 14, no. 5, pp. 1035–1042, 2020, doi: 10.1007/s11760-020-01636-0.

[3] P. M. Raju and M. S. Nair, "Copy-move forgery detection using binary discriminant features," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 2, pp. 165–178, 2022, doi: https://doi.org/10.1016/j.jksuci.2018.11.004.

[4] A. Alahmadi, M. Hussain, H. Aboalsamh, G. Muhammad, G. Bebis, and H. Mathkour, "Passive detection of image forgery using DCT and local binary pattern," *Signal, Image and Video Processing*, vol. 11, no. 1. pp. 81–88, 2017. doi: 10.1007/s11760-016-0899-0.

[5] S. Albelwi and A. Mahmood, "A framework for designing the architectures of deep Convolutional Neural Networks," *Entropy*, vol. 19, no. 6. 2017. doi: 10.3390/e19060242.

[6] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5455–5516, 2020, doi: 10.1007/s10462-020-09825-6.

[7] N. V. S. K. Vijayalakshmi K, J. Sasikala, and S. Chandrasekaran, "Copy-paste forgery detection using deep learning with error level analysis," *Multimed. Tools Appl.*, pp. 1–25, May 2023, doi: 10.1007/s11042-023-15594-5.

[8] D. Mallick, M. Shaikh, A. Gulhane, and T. Maktum, "Copy Move and Splicing Image Forgery Detection using CNN," *ITM Web Conf.*, vol. 44, p. 03052, 2022, doi: 10.1051/itmconf/20224403052.

[9] T. Muniappan, N. B. A. Warif, A. Ismail, and N. A. M. Abir, "An Evaluation of Convolutional Neural Network (CNN) Model for Copy-Move and Splicing Forgery Detection," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 2, pp. 730–740, 2023.

[10] M. Mahyoub, F. Natalia, S. Sudirman, and J. Mustafina, "Sign Language Recognition using Deep Learning," in *2023 15th International Conference on Developments in eSystems Engineering (DeSE)*, 2023, pp. 184–189. doi: 10.1109/DeSE58274.2023.10100055.

[11] A. Anton, N. F. Nissa, A. Janiati, N. Cahya, and P. Astuti, "Application of Deep Learning Using Convolutional Neural Network (CNN) Method for Women's Skin Classification," *Sci. J. Informatics*, vol. 8, no. 1, pp. 144–153, 2021.

[12] P. Kubal, V. Mane, and N. Pulgam, "Image Manipulation Detection Using Error Level Analysis and Deep Learning," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 4 SE-Research Article, pp. 91–99, Sep. 2023.

[13] V. Sharma, S. Jha, and R. Kumar Bharti, "Image Forgery and it's Detection Technique: A Review," *International Research Journal of Engineering and Technology*. 2016.

[14] A. Pandey and A. Mitra, *Detecting and Localizing Copy-Move and Image-Splicing Forgery*. 2022.

[15] F. Z. Mehrjardi, A. M. Latif, M. S. Zarchi, and R. Sheikhpour, "A survey on deep learning-based image forgery detection," *Pattern Recognit.*, vol. 144, p. 109778, 2023, doi: https://doi.org/10.1016/j.patcog.2023.109778.

[16] J. F. Andry, H. Tannady, G. Dwinoor Rembulan, and D. Freggy Dinata, "Analysis of the Omicron virus cases using data mining methods in rapid miner applications," Microbes Infect. Dis., vol. 4, no. 2, pp. 323–334, 2023, doi: 10.21608/mid.2023.194619.1469.