Ilkogretim Online - Elementary Education Online, 2021; Vol 20 (Issue 5): pp. 7181-7189
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.05.809

Intelligent Morphed Image Identification using Error Level Analysis and Deep learning

Gaayan Kumar, Amity University Uttar Pradesh, Noida, India, gaayanverma@gmail.com
Sunil Kumar Chowdhary, Amity University Uttar Pradesh, Noida, India, skchowdhary@amity.edu
Abhishek Srivastava, Amity University Uttar Pradesh, Noida, India, abhishek.sri13@gmail.com

Abstract. Images are often tinkered with the intention to benefit one party. In reality, images are often considered as solid evidence to prove something concrete. Therefore, fake news or any form of information that has been manipulated in such a way it benefits the party in fault or results in misleading. One of the biggest sources of news or information manipulation is image falsification. To detect the image falsification, it takes a substantial amount of image data and robust model that can process every pixel but still provides efficiency and flexibility to support daily life use. In the era, where huge amount of data is generated every second, Deep learning is the right solution because deep learning thrives as the dataset increases in size. Therefore, Convolutional Neural Network (CNN) in combination with Error Level Analysis (ELA) to facilitate model computing, Detection of a forged image has achieved an accuracy of 91.33% and convergence with only 9 epochs.

Keywords: Image, Forgery, Error, Deep Learning, Analysis, Fake, News, Machine Learning

I. INTRODUCTION

In this technological age, many people have already become a victim of image forgery. Many people use techniques for manipulating images and using them as evidence misleading the court. Photos sharing via social media is often considered authentic which is completely false. Social media is an excellent platform for socializing, sharing, and disseminating knowledge, but if not careful, it may mislead people, even because of false promotion. Although dealing with most due to pixelation, photoshop images are visible. Some of the low-quality jobs provided by novices are indeed true. Especially on the political stage, manipulated images can destroy the credibility of politicians.

According to the EU High-Level Expert Group, false news is defined as any form of inaccurate, false, or misleading information that is presented, promoted, or designed. Behind the fake news, there are several reasons for the publication. One of them is to get economic benefits, whether it is through increasing the number of news clicks or making news that is not supposed to benefit one of the parties. In addition, fake news can also affect price shares, which can provide benefits to parties who released the report. Another reason is to get support or bring another party down social or political [1].

Based on statistics belonging to the Indonesian Telematics Society (MASTEL) in 2017, the types of false news that are often received are socio-political, SARA (ethnicity, religion, and race), health, food, and drink financial fraud, and science and technology. As many as 84.5% of all respondents stated that they felt disturbed by the existence of false news, and more than 70% agreed that fake news disturbed the harmony of society and hampered development [2]. Apart from writing, about 40% of respondents stated that the spread of fake news was also often accompanied by pictures. Images are used by humans to reproduce reality and are commonly used as evidence of news, publications, or facts. Fake news that has a supporting model tends to be accepted and trusted by the public.

In general, humans are more comfortable to remember the form of images than writing. According to the Social Science Research Network, as many as 65% of people are people who enjoy learning through visuals. In marketing and visual science, it is mentioned that the image has a massive effect on an article. People are more likely to respond when there are images rather than writing. According to themed infographics on the impact of images in the world of marketing, images can increase the number of respondents for an article by up to 94% [3]. Therefore, a picture is a strong element in spreading information.

To determine an authentic or fake image, it is challenging to see with the naked eye, special techniques and specific research are required to be able to know for sure that a picture is an original image or has been modified. For ordinary people, this might be difficult to do. For this reason, Image forgery detection technology was introduced. This technology requires a lot of images, and each model has many constitu-

ent pixels. With ordinary machine learning, this technology will be challenging to develop. Therefore, big data and deep learning are the correct solutions to the problem of false image detection.

Most image files contain more than the number of images. They also contain details of the picture. Metadata or data of data provides information on the lineage of the image, including the various cameras used, information about the colour space, and application notes. Multiple image formats have different varieties of metadata. Formats like BMP, PPM, and PBM have less information and exceed the image size. In contrast, a camera's JPG contains various information, including the make and model of the camera, focal length information, aperture and time stamp. Unless the image is JPG converted or photoshopped, PNG files usually contain very little information. The data converted to PNG contains metadata in the source file.

To cater the above mentioned problem of image falsification, In this paper we have implemented a system that can use deep learning to determine whether an image is false, by which people can be aware about the authenticity of the source they are referring to . Coverage of this paper is divided into three parts, first part focused on introduction of problem while second part on implementation details, & the final section comprised of the experimental results.

II. LITERATURE REVIEW

In the paper detection of fake images via "The Ensemble of Deep Representations from Multicolor Spaces" by Peisong He, Haoliang Li, and Hongxia Wang, 8500 pairs of real and fake images were randomly divided into two subsets with the ratio 5:1. This data was used to train and validate with a batch size of 64. Shallow CCN was trained using ADAM (Adaptive Movement Estimation Method). Keras implemented the CNN, GTX 1080Ti was used for computation. Accuracy of 99.35% was obtained with the variation of the hyperparameter [4].

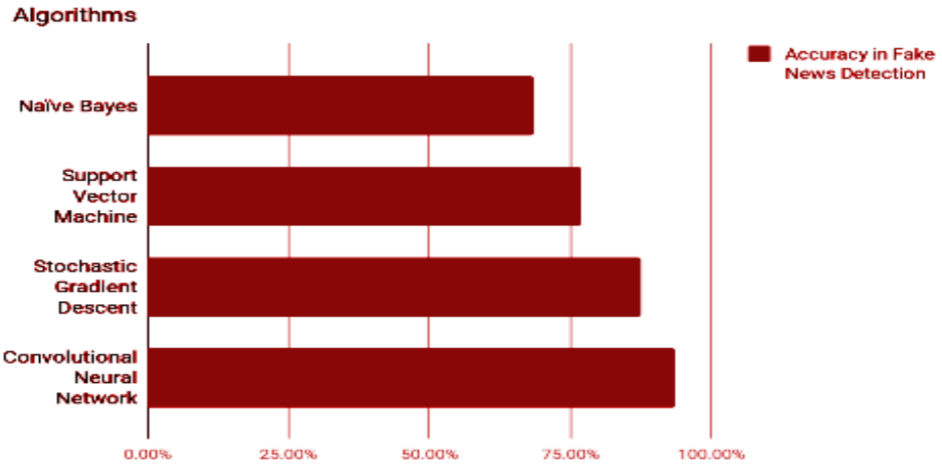

Fig. 1: Comparison of Accuracy in Detecting Fake News

The bar chart in figure 1, is computed using textual data from social media sources such as Twitter, Facebook, BuzzFeed, and also various internet news sources. As we can see, the best accuracy is obtained by Convolutional Neural Network (93% Accuracy), CNN is highly effective, but a lot of computation power is needed. Stochastic Gradient Descent (87.5% Accuracy), SGD has a faster learning rate, but the output contains noise. Support Vector Machine and Naïve Bayes have seen less use because of less accuracy [5].

The latest and the most refined process is to use pre-acquired knowledge to differentiate physical elements in images. Based on signal processing techniques, earlier approaches used indexes, such as unaligned JPG squares, compressed quantisation artefacts, resampling artefacts, colour filter tables [6], and photography. We were particularly inspired by the latest work of Agarwal and Farid [7], who used the expected differences between the imaging pipelines to detect tampered image areas. i.e. JPEG digital quantisation was truncated during different camera intercepts although these domain-specific methods have proven useful because of their ease of interpretation. Recently in the field of fake image detection, Trend has shifted from using previous knowledge to using end-to-end learning methods to use labelled training

data to solve specific forensic tasks. It is recommended to learn to detect seams by forming a CNN entirely on the labelled training data.

Table 1: Comparison of Accuracies

Title of Paper	Author	Model Used	Accuracy
Detection of GAN-generated Fake Images over Social Networks	Francesco Marra, Diego Gragnaniello, Davide Cozzolino, Kuisa Verdoliva	Convolutional Neural Network (Uncompressed Dataset)	89.55%
Detection of GAN-generated Fake Images over Social Networks	Francesco Marra, Diego Gragnaniello, Davide Cozzolino, Kuisa Verdoliva	Convolutional Neural Network (Twitter Like Compressed Images)	81.51%
Fake Smile Detection Using Linear Support Vector Machine	Gede Aris Gunadi, Agus Harjoko, Retantyo Wardoyo, Neila Ramdhani	Linear Support Vector Machine	86%
Fake Face Detection Methods: Can they be Generalized	Ali Khodabakhsh, Raghavendra Ramachandra, Kiran Raja, Pankaj Wanik, Christoph Busch	CNN (Alex Based)	95.83%

These methods have also been applied to the problem of detecting specific signs of forgery, such as double JPEG saved images and improved contrast [8]. Perhaps the closest approach to our solution used in Bondi [9]. Another common forensic strategy is to train the model on a small class of automatically simulated operations, such as changing faces or using COCO segmentation masks for assembly. Besides, Zhou [10] proposed a solution to recognise the exchange of human faces by measuring the inconsistency and blur of the image introduced by the component. In concurrent work, Mayer [11] proposed to use the Siam network to detect whether a group of image patches have the same camera model (they suggest that this model can also be used for check; although promising, these results are very preliminary).

In our work, we are trying to reduce further the amount of information provided to the algorithm by teaching the algorithm to detect operations excluding actual annotations. To this end, we were inspired by the recent self-monitoring work [12], which formed a model by using only unlabelled data to solve defined tasks. Among these methods, the closest approach is the method of Doersch [13], by which they were able to form a model which predicts the respective position of plaque pairs in the image. Then, it was later realised that the more optimal path would be to use narrow as a shortcut to learning tasks. Although improving bad images is an annoying thing in their work, It was a useful signal for them in their self-monitoring algorithm as it aims to understand the properties of the imaging pipeline better, regardless of its semantics [14], which uses self-supervision to form a segmentation model to predict if pairs of patches appear simultaneously in space or time.

Our work is also related to detecting anomalies. Traditional visual anomaly detection mainly focuses on the discovery of abnormal semantic events, such as the availability everyday objects, actions that finds anomalies in the photos, and the content is designed to successfully deceive humans. The anomalous clues we are looking for must be indistinguishable to humans, and the semantics of the scene is unchanged.

III. METHODOLOGY

Our proposed methodology comprised of two main goals:

a. Propose a new method using deep learning to classify images as original images and images that have been modified, so that computing costs can be reduced.

b. Involves the use of ELA in machine learning as an effort to improve efficiency.

There are several motivations behind these two primary goals. Most of these studies require substantial computational costs (can be seen from the number of epochs and services needed), so that the flexibility of the proposed method is reduced and difficult to apply in everyday life because it is hampered by computing costs. There is a need for image detection methods to be able to adapt to the addition of image data and modification over time. Therefore, the purpose of detection of image forgery that is relatively more efficient and has an increase in scalability that is directly proportional to the rise in data.

3.1. Convolutional Neural Network (CNN)

The basic structure of CNN in figure 2 includes some essential parts, such as a convolutional layer, pooling layer and fully connected layer. One of the common uses of convolutional layers is the extraction of entities. Connect each input of the neuron to connect each layer to the local receiving field of the upper layer. The pooling layer can reduce the size of the data and maintain the characteristics of the network structure.

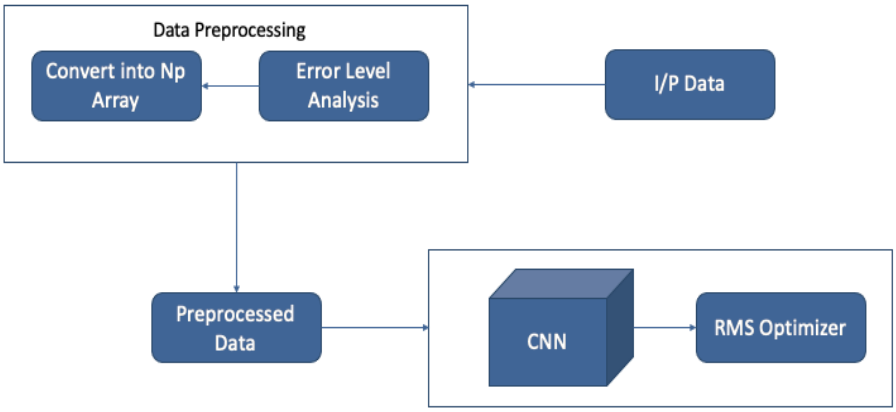

Fig. 2: General Architecture

CNN is a direct-acting network, and the information flow from the entrance to the exit is only one way. Although there are several types of CNN architectures, in general, CNNs have multiple convolutional and grouping layers. Then there are one or more fully connected layers. Generally, architectural design is divided into two main parts, namely, data preparation and model building. In the initial stage, the input data consisted of images in ".jpg" format with the following details: 1771 images with fake labels and 2940 models with original labels were captured during data preparation stage [4]. The data preparation step is a step of first converting each image as input data into an image obtained by analysing the error level. Then use the LA image to resize the banner ad to 128 x 128. Converting raw data to ELA results is a method for improving the efficiency of CNN model training. The ELA image generation function focuses on the part of the image where the error level exceeds the limit. Besides, compared to adjacent pixels, ELA pixel images tend to have similar or even very contrasting colours, so the CNN model formation becomes more active. After that, the image size changes. Normalization of image takes place by dividing each RGB value by the number 255. So that CNN converges faster (reaching the minimum overall loss value of the verification data) because the only the amount of each RGB value can be between 0 and 1. The data label is modified to where 1 corresponds to the counterfeit and 0 corresponds to the actual classification value. After that, the training data and the verification data are divided into 80% for the training data and 20% for the verification data.

Training data and validation data is used conduct in-depth learning model training using CNN. Optimisation implemented during exercise is the RMSPropoptimiser, which is one of the adaptive learning rate methods. The complete architecture used in the building model section can be seen in the image below or by using the link, the sheet is a perfect architectural image. In the deep learning model used, the first layer of CNN consists of convolutional layers in figure 3, the core size of the convolutional layer is 5x5, and the maximum number of filters is 32. The second layer of CNN consists of a convolutional layer with a core size of 5x5 and a total of 32 screens. The size of the MaxPooling layer is 2x2. These two convolutional lay-

ers select the neurons of the convolutional layer by using the unified unity of the kernel initialiser and the ReLU activation function so that they can receive useful data signals Entrance [16].

After that, the MaxPooling layer was added to the pressure drop of 0.25 to avoid overfitting. The layer followed by MaxPooling is a fully connected layer with 256 neurons and ReLU activation function. After a fully connected layer, a dropout of 0.5 will be added to prevent overfitting. The output layer used has a Soft-Max activation function.

In the architecture used, only two convolutional layers are needed because the results from the conversion process into an ELA image can highlight important features for knowing whether an image is original or has been modified. One of the biggest companies, Facebook uses ML to personalise the news feed of each customer and give them a unique experience. When a user stops scrolling to see or react to a particular friend post, the feed will display that friend's activity earlier that day. In the background, the software uses statistics and predictive analysis to identify patterns of users in user data and use them for news sources. They observe user activities such as comments, sharing, etc. According to various news and activities, the content of the news feed will be constantly adjusted.

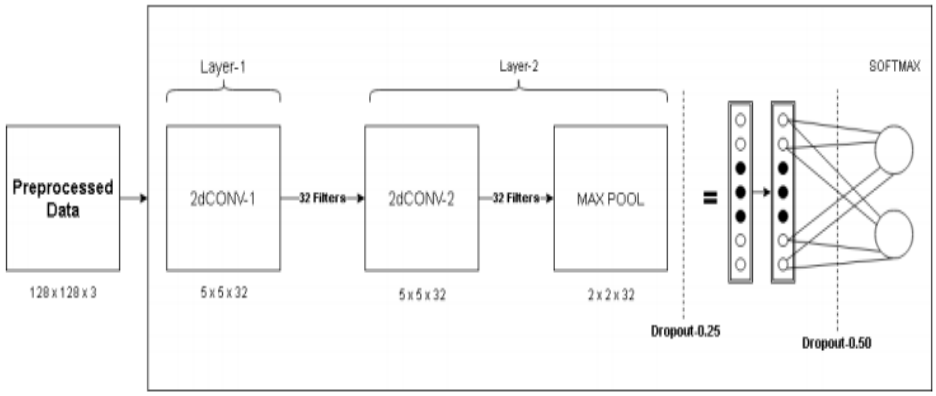

Fig. 3: CNN Architecture

In the input layer, multi-layered perceptron which has 3 hidden layers and 1 output layer present. After choosing the image which has to be evaluated, it will be converted from the compression and error level analysis steps to the ELA representation, 100% and 90% of the image are used to construct the ELA image.

After ELA, the image is pre-processed to convert to a length and breadth of 128*128 px. After pre-processing, the images will be serialized in the table. Since 1 RGB (red, green and blue) pixel occupies 3 values. Therefore, an array which has to represent 10,000 pixels will have 30,000 integer values. The current state data set cannot form a model. It must be converted to a state that is perfectly suitable for the task at hand, that is, detecting abnormalities at the pixel level introduced by the forging operation. Based on the ideas here, we designed the following method to create related images based on the data provided. For each false image, we have a corresponding mask. We use this mask to sample false images along the edges of the sewing area to ensure that the false and non-false parts of the image contribute at least 25%. These samples will have unique limitations that only exist in erroneous images. The CNN web design must understand these limitations. Since the three channels of the mask contain the same information (false parts of the image in different pixels), we only need one channel to extract the samples.

Almost all these technologies use functions based on the content of the image (that is, the visual information present in the image). CNN is inspired by the visual cortex. Technically, these networks are designed to extract functions that are important for classification, that is, these functions minimize the loss function. Network parameters - the weight of the atomic nucleus is learned by gradient descent to produce the most obvious features of the image transmitted to the network. These entities are then passed to the fully connected layer that performs the final classification task. After looking at some forged images, it is clear that forged areas of the human visual cortex can be found. Therefore, CNN is an ideal deep learning model for this work. If the human visual cortex can detect it, then it will design more functions for this

task in the network. During training, the network entered the multi-layer perceptron network and also defined output neurons. MLP is a fully connected neural network. The result obtained is in the form of multiple neurons. The first neuron is used to represent a fake image, and the second neuron is used to represent a real image. If the given image is false, set the false neuron to 1 and the real number to zero. Otherwise, set false to zero and real to 1.

3.2. Error Level Analysis (ELA)

Error level analysis is one of the techniques for detecting image operations by recording images at a certain quality level and calculating the comparison between compression levels [17]. Generally, this technique is performed on frames in a lossy format (lossy compression). The image type used to extract this data is JPEG. In JPEG images, compression is performed independently for each 8*8 pixel image. If the image is not processed, all 8*8 pixels of the image must have the same error status [18].

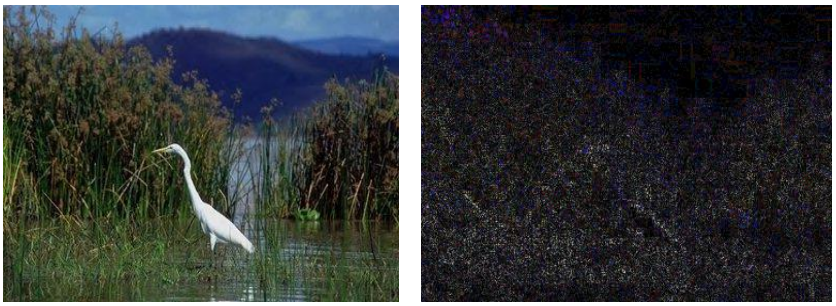

Fig. 4: Real Image to ELA

JPEG is a lossy format, but the amount of error introduced by each re-recording is not one line. Any modification of the image will change the image and make the area stable (without other errors) unstable. For example: if a modified image is saved again, it will retain 75% of its original quality & then copying of some information like, copying the book on the shelf and added the toy dinosaur to the shelf. 95% of ELA can recognize the change because these areas are no longer at the lowest error level. Because Photoshop merges information from multiple layers and changes many pixels, other areas of the image are slightly more volatile. Saving the image from 90% to 90% is equivalent to 81% of the backup. Similarly, saving the image at 75% and then re-recording it at 90% (75% to 90%) will produce almost identical (90% to 90%), which is a backup of 67.5%. and after about 64 re-recordings, there is almost no change. However, when the image is modified, the 8*8 unit which is modified does not have the same error level as the previous unmodified image. Error level analysis (ELA) works by re-recording images with a known error rate and then calculating the difference between the images.

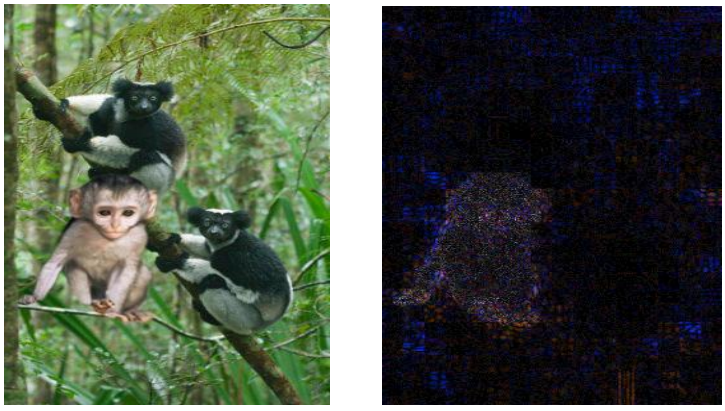

Fig. 5: Fake Image to ELA

If the change is not large, it indicates that the device has reached the minimum local error value of this quality level. However, if the variation is large, the pixel is not at its local minimum, but the ”original” pixel. Use the ImageJ library to perform error level analysis. The system first records 100% quality images. Then use ImageJ to convert the same image to a 90% quality image. The difference between the two is found through the difference method as shown if figure 4 & figure 5. The generated image is the ELA im-

age required for the input image. The image will be saved as a buffer image and sent to the neural network for further processing.

IV. RESULTS

The results obtained from the proposed method have obtained an accuracy of 91.33% that is reflected in form of accuracy curve and the loss curve shown in figure 6 & 7. It can be seen in the picture above that the best accuracy is obtained on the 9th epoch.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 124, 124, 32)      2432
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 120, 120, 32)      25632
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 60, 60, 32)        0
_________________________________________________________________
dropout_1 (Dropout)          (None, 60, 60, 32)        0
_________________________________________________________________
flatten_1 (Flatten)          (None, 115200)            0
_________________________________________________________________
dense_1 (Dense)              (None, 256)               29491456
_________________________________________________________________
dropout_2 (Dropout)          (None, 256)               0
_________________________________________________________________
dense_2 (Dense)              (None, 2)                 514
=================================================================
Total params: 29,520,034
Trainable params: 29,520,034
Non-trainable params: 0
_________________________________________________________________
```

Fig. 6: Neural Network Overview

Validation loss value after the 9th epoch starts to flat and finally increase, which is a sign of overfitting. The method of identifying the number of epochs that is good for use during training is early stopping. With this method, training will be stopped when the accuracy of validation starts to decrease or the value of validation loss starts to increase. Since the ELA conversion result image function is used to make the learning model more effective and the RGB values are normalized, the number of learning periods required is small and convergence cannot be achieved. Each pixel can also accelerate the convergence of the CNN model. The accuracy obtained by the model is doing the classification can be said to be quite high. This is an indication that the feature is in the form of an ELA image successfully used to classify whether the image is an original image or has been modified.

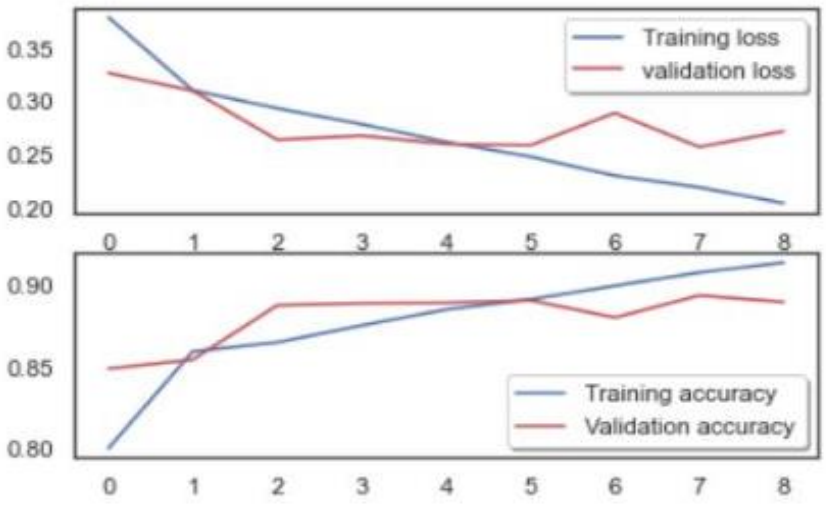

Fig. 7: Graph on Accuracy versus Loss

V. CONCLUSIONS

In this study, several conclusions can be drawn from the results of machine learning using ELA and CNN.

a. CNN uses two convolutional layers (2dCONV), one MaxPooling layer, Dropout of 0.25, one fully connected layer, dropout of 0.5 and the output layer with softmax can achieve 91.33% accuracy as shown in figure 9.

b. Using ELA can increase efficiency and reduce training costs. This can be seen from the reduction in the previous layers and the number of required periods. In our proposed model, the time required to reach convergence is only 9.

```
# Predict the values from the validation dataset
Y_pred = model.predict(X_val)
# Convert predictions classes to one hot vectors
Y_pred_classes = np.argmax(Y_pred,axis = 1)
# Convert validation observations to one hot vectors
Y_true = np.argmax(Y_val,axis = 1)
# compute the confusion matrix
confusion_mtx = confusion_matrix(Y_true, Y_pred_classes)
# plot the confusion matrix
plot_confusion_matrix(confusion_mtx, classes = range(2))
```

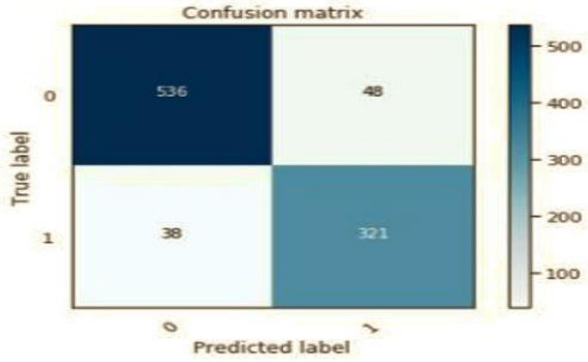

Fig. 8: Confusion Matrix

c. A confusion matrix as shown in figure 8, represents the performance of validation data where 1 symbolizes tampered images, 0 symbolizes the original images. The presented model predicted 536 out of 574 original images and 321 out of 369 tampered images accurately.

REFERENCES

1. Ozgobek, Ozlem, and Jon Atle Gulla. "Towards an understanding of fake news," CEUR Workshop Proceedings. 2017.
2. Kshetri, Nir, and Jeffrey Voas. "The economics of fake news", IT Professional 8-12, 19.6 (2017).
3. Bullas, Jeff. "6 Powerful Reasons Why You Should Include Images in Your Marketing." 2012.
4. Diambil Dari. "CASIA Image Tampering Detection Evaluation Database (CAISA TIDE) V2.0", Chinese Academy of Science. http://forensics.idealtest.org.
5. Faraz Ahmad and Lokesh kumar. "A Comparison of Machine Learning Algorithms in Fake News Detection", International Journal on Emerging Technologies 10(4): 177-183, 2019.
6. Popescu, Alin C., and Hany Farid. "Exposing digital forgeries by detecting traces of resampling." IEEE Transactions on signal processing 758-767, 2005.
7. Wen, Longyin, Honggang Qi, and Siwei Lyu. "Contrast enhancement estimation for digital image forensics." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM): 1-21, 14.2, 2018.
8. Rao, Yuan, and Jiangqun Ni. "A deep learning approach to detection of splicing and copy-move forgeries in images." 2016 IEEE International Workshop on Information Forensics and Security (WIFS). IEEE, 2016.
9. Bondi, Luca. "First steps toward camera model identification with convolutional neural networks." IEEE Signal Processing Letters: 259-263 24.3, 2016.

10. Zhou, Peng. “Two-stream neural networks for tampered face detection.” 2017 IEEE Conference on Computer Vision and Pattern recognition Workshops (CVPRW), IEEE, 2017.
11. Owen Mayer, M.C.S. “Learned forensic source similarity for unknown camera models.” IEEE International Conference on Acoustics, Speech and Signal Processing 2018.
12. Agarwal, Shruti, and Hany Farid, “Photo forensics from JPEG dimples.” IEEE Workshop on Information Forensics and Security (WIFS), IEEE, 2017.
13. Zhang, Richard, Phillip Isola, and Alexei A. Efros. “Split-brain autoencoders: Unsupervised learning by cross-channel prediction.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
14. Doersch, Carl, Abhinav Gupta, and Alexei A. Efros. “Unsupervised visual representation learning by context prediction.” Proceedings of the IEEE International Conference on Computer Vision, 2015.
15. Rawat, Waseem, and Zenghui Wang. “Deep convolutional neural networks for image classification: A comprehensive review.” Neural computation: 2352-2449, 29.9, 2017.
16. Nair, Vinod, and Geoffrey E. Hinton. “Rectified linear units improve restricted boltzmann machines.” Proceedings of the 27th international conference on machine learning (ICML-10), 2010
17. Krawetz, Neal, and Hacker Factor Solutions. “A pictures worth.” Hacker Factor Solutions 2007.
18. Gunawan, Teddy Surya, “Development of photo forensics algorithm by detecting photoshop manipulation using error level analysis.” Indonesian Journal of Electrical Engineering and Computer Science (IJEECS), 131-137, 2017.
19. Isola, Phillip, “Learning visual groups from co-occurrences in space and time.” Workshop track - ICLR 2016.
20. He, Peisong, Haoliang Li, and Hongxia Wang. “Detection of fake images via the ensemble of deep representations from multi color spaces.” IEEE International Conference on Image Processing (ICIP). IEEE, 2019.