ISSN 2723-6471

Fake vs Real Image Detection Using Deep Learning Algorithm

Fatoni[1*], Tri Basuki Kurniawan[2], Deshinta Arrova Dewi[3], Mohd Zaki Zakaria[4], Abdul Muniif Mohd Muhayeddin[5]

[1]*Faculty of Science Technology, Universitas Bina Darma, Palembang, Indonesia*

[2]*Postgraduate Program, Universitas Bina Darma, Palembang, Indonesia*

[3]*Faculty of Data Science and Information Technology, INTI International University, Malaysia*

[4,5]*Faculty of Computer and Mathematics Sciences, University Technology Mara, Malaysia*

(Received: August 18, 2024; Revised: September 5, 2024; Accepted: October 22, 2024; Available online: December 29, 2024)

Abstract

The purpose of this research project is to address the growing issues presented by modified visual information by developing a deep learning model for identifying between real and fake images. To enhance accuracy, this project evaluates the effectiveness of deep learning algorithms such as Residual Neural Network (ResNet), Visual Geometry Group 16 (VGG16), and Convolutional Neural Network (CNN) together with Error Level Analysis (ELA) as preprocessing the dataset. The CASIA dataset contains 7,492 real images and 5,124 fake images. The images included are from a wide range of random subjects, including buildings, fruits, animals, and more, providing a comprehensive dataset for model training and validation. This research examined models' effectiveness through experiments, measuring their training and validation accuracies. It comes out with the best accuracy of each model, which is for Convolutional Neural Network (CNN), 94% for training accuracy, and validation accuracy of 92%. For VGG16, with both training and validation accuracy reaching 94%. Lastly, Residual Neural Network (ResNet) demonstrated optimal performance with 95% training accuracy and 93% validation accuracy. This project also constructs a system prototype for practical applications, offering an interface for real-world testing. When integrating into the system prototype, only Residual Neural Network (ResNet) shows consistency and effectiveness when predicting both fake and real images, and this led to the decision to choose ResNet for integration into the system. Furthermore, the project identified several areas for improvement. Firstly, expanding the model comparison for discovering more successful algorithms. Next, improving the dataset preprocessing phase by incorporating filtering or denoising techniques. Lastly, refining the system prototype for greater appeal and user-friendliness has the potential to attract a larger audience.

Keywords: Fake Image Detection, Deep Learning, Residual Neural Network (ResNet), Error Level Analysis (ELA), Process Innovation

1. Introduction

In recent years, the detection of fake images has become a critical research area due to the increasingly widespread of manipulated visual content in various contexts. Selva et al. [1] have proposed various methods for detecting image forgeries, ranging from passive techniques that rely on image features and statistical analysis to active techniques that involve image manipulation and reconstruction. Some of the notable approaches include Error Level Analysis (ELA), Local Phase Quantization, and Chromatic components, as shown by Qurat-Ul-Ain et al. in their research [2].

These methods have shown promising results in detecting various types of image forgeries, such as copy-move, splicing, and tampering, in addition. Various techniques, such as LBP and HOG, were used in combination with resizing and ELA to assess image authenticity [3]. ELA involves compressing both real and forged images at a specific level of quality and then taking the difference between these compressed images and the original images.

Furthermore, deep learning models have also emerged as powerful tools for image classification and recognition tasks, including the detection of fake images. Several studies have evaluated the performance of different deep learning models, such as CNN, Generative Adversarial Networks (GANs), and self-evaluation, on public datasets such as CASIA and COMOFOD [2]. However, the accuracy and precision of detection may vary depending on the quality and complexity of the images, the type of forgery, and the choice of features and models employed. The prevalence of fake images in today's digital landscape poses significant challenges in terms of trust, credibility, and authenticity. With the increasing accessibility of advanced image editing tools and techniques, it has become remarkably easy to produce fake

***Corresponding author: Fatoni (fatoni@binadarma.ac.id)**
DOI: https://doi.org/10.47738/jads.v6i1.490

 ISSN 2723-6471

photos that are virtually indistinguishable from genuine ones [3]. This has led to a growing skepticism surrounding the veracity of photos used in a wide range of domains, including media, online marketplaces, and authentication processes. While traditional image forgery detection methods have been effective in detecting common forms of manipulation such as splicing, copy-move, and retouching [4], they may fall short when it comes to identifying fake images generated using deep learning algorithms, particularly GANs [5].

Moreover, Deep learning models can generate highly realistic and convincing images, making it increasingly challenging for traditional detection methods to differentiate between real and fake. Using deep learning models' advantages, which are excellent at extracting complex characteristics and patterns from images, holds promise for spotting minute traces and artifacts that traditional techniques could overlook. Utilizing deep learning algorithms allows for the detection of complex patterns, hidden anomalies, and minute differences that are challenging to spot using more traditional fake detection methods.

These deep learning algorithms have gained significant recognition for their effectiveness in detecting fake images. Some research comparative studies of deep-learning models that can benefit deepfake detection. For example, VGG16, MobileNetV2, XceptionNet, and InceptionV3 [6]. But it's specifically for face detection and using fake face image datasets.

Innovative approaches are needed to improve the precision and efficiency of image detection systems to solve this issue. In addition to harnessing the power of deep learning algorithms, integrating ELA further enhances the proposed strategy. ELA is a forensic technique that leverages variations in compression levels within an image to identify regions that have been manipulated. By incorporating ELA into the detection process, the proposed method gains an additional level of understanding and insight. The integration of ELA can significantly improve the precision and accuracy of identifying fake photos, thereby strengthening the overall effectiveness of the detection [7].

This research project will use deep learning algorithms for image processing, including CNN, Residual Neural Network (ResNet), and VGG16. These algorithms will be used to distinguish between fake and real photographs and to identify them. ELA will be added as a preprocessing step to these methods to improve performance by highlighting compression artifacts and assisting in differentiating between original and modified sections. The research will be conducted using a dataset consisting of randomly selected fake and real images obtained from Kaggle.

Proposing a method for detecting fraudulent photographs is extremely important due to its wide range of practical applications in fields such as social media, cybersecurity, and online markets. The capacity to reliably detect the validity of photographs is critical in various domains for decision-making, preserving user security and privacy, and sustaining confidence in online interactions. In the context of social media, where visual content is extensively disseminated, the presence of edited and manipulated photographs creates considerable issues. Detecting false photos is critical in combating the spread of misinformation, fake news, and online scams. Furthermore, in the field of cybersecurity, identifying fraudulent photos is critical for protecting against image-based threats such as identity theft or malicious content dissemination. Furthermore, confirming the validity of product photographs is critical in online markets for generating confidence and deterring fraudulent actions.

2. Fake and Real Image Detection Algorithms

2.1. Image Classification

Image classification is an important problem in computer vision that involves assigning a label to an image based on its content. It has many applications in various areas of image processing, including image and video retrieval, texture analysis, and the design of recognition or surveillance systems [8]. There are different types of image classification methodologies to be employed. However, it was Binary for my project because I could classify each image as either fake or real by using an either-or logic. By classifying an image into one of the two unique groups, this method enables me to determine the validity of an image [9]. My project is to create a model that determines if photographs are false or real, which is it's a binary classification technique on a random dataset. This model may be used to verify the validity of images in a variety of settings, including social media, cybersecurity, and online markets.

2.2. Error Level Analysis (ELA)

Error Level Analysis is a method for identifying picture manipulation that involves restoring the image to a specific quality level and comparing the ratio of compression levels. This ELA is regarded as the forensic method that looks at compression artifacts in images like JPEGs compressed using lossy methods [10]. You can spot parts of the image with

various levels of compression by using this tool. To make it easier to understand, compression takes place when a picture is first saved in JPEG format. Editing programs that support JPEG compression include Adobe Photoshop, GIMP, and Adobe Lightroom. A more round of compression is applied if the image is later edited in editing software and saved once again. The original digital camera photographs should have high ELA values. With each subsequent stage, the error rate is decreased [11].

ELA can be a valuable method for spotting potential image manipulations or alterations, according to research on its effectiveness in spotting fraudulent photographs. ELA can identify regions of a picture that might have been altered or tampered with by examining the differences in error levels created during compression. It is crucial to remember that ELA does not constitute absolute proof of picture modification. As a result, ELA can be a useful tool in a full toolkit for identifying fake or altered photos, but it should be used in concert with other methods and supporting data for more precise evaluations.

2.3. Deep Learning

Deep learning models have proven to have excellent abilities in a variety of image-related tasks by utilizing the strength of neural networks with several layers. In addition, denoising, style transfer, and image restoration have all benefited greatly from deep learning. Overall, deep learning has paved the way for breakthroughs in computer vision and visual intelligence by opening fascinating possibilities for image analysis, manipulation, and production.

CNNalso known as ConvNet, has an impressive ability to generalize in a greater way than networks with fully connected layers due to its deep feed-forward design [12]. These layers do local receptive field activities, recording and identifying local patterns that are later integrated hierarchically to identify complicated structures.

CNN have completely changed how image categorization tasks are performed. CNN can efficiently analyze and classify images because of their deep design, which contains convolutional layers for feature extraction and pooling layers for dimensionality reduction. CNN has outperformed conventional methods in image classification problems by utilizing hierarchical representations and learning from huge datasets.

A great generative model is GAN. However, the vanishing gradient, difficulty in training, and lack of diversity are just a few of the issues with the original GAN model. It has taken a lot of work to improve GAN using various optimization techniques. As a result, since 2014, a seemingly endless flood of hypotheses and publications about GANs has appeared, and numerous new GAN-based models have been suggested to increase the stability and quality of the results generated [13]. Actual GAN work began in 2017, using human faces to adopt picture augmentation that creates better illustrations at high intensity. The blog post by Olli Niemitalo from 2010 that served as the basis for adversarial networks is also known as Conditional GAN, which is a similar concept [14].

Visual Geometry Group, also known as VGG, is a typical deep CNN architecture with several layers [15]. The term "deep" describes the number of layers, with VGG-16 or VGG-19 having 16 or 19 convolutional layers, respectively. Further, in Vgg16 network construction, the network has over 138 million parameters, making it a sizable network. All of the convolution layers of the vgg16 network have the same configuration: the convolution core size is 3X3, and the step size is 1; there are a maximum of five pooling layers, all of which are 2x2 and have a step size of 2; There are three complete connection levels, the last layer having 1000 channels and 1000 label categories, the preceding two layers each having 4096 channels; The SoftMax layer is the lowest, and a ReLU nonlinear activation function comes after all hidden layers [16].

Kaiming et al. [17] introduced the residual network in their paper "Deep Residual Learning for Image Recognition," published in 2016. Adding more layers to Deep Neural Networks improves accuracy and performance and is typically done to handle complex problems. The idea behind adding additional layers is that they would gradually learn more complicated traits [18]. For instance, while detecting photos, the first layer may learn to recognize edges, the second layer may learn to recognize textures, the third layer can similarly learn to recognize things, and so on. The conventional Convolutional neural network model, however, has been found to have a maximum depth threshold. The errors in training and testing data for a 20-layer network and a 56-layer network are shown below [19], [20].

 ISSN 2723-6471

3. Method

This project is divided into three parts, each with its own set of tasks. The preparatory study, knowledge gathering and acquisition, data collection, and data preprocessing were all activities in phase 1. The activity engaged in phase 2 was system design. Finally, phase 3 tasks included system development, system testing, and evaluation, as well as final report documentation.

3.1. Phase 1

The first phase of the study focused on researching and learning about fake image identification using deep learning algorithms. The primary purpose of this phase was to learn about the challenges associated with detecting fake photos and to identify efficient ways for distinguishing real from edited photographs. Several significant actions were carried out during this phase, including preliminary study, knowledge acquisition and analysis, data collection, and data pre-processing.

The attention was to understand and gain knowledge on fake photo identification using deep learning algorithms. Discussions with the project supervisor and a thorough review of the existing literature on the issue took place during this activity. The goal was to find out the characteristics and problems associated with fake images, as well as effective methods for detecting them. This phase's results included the study's historical context, a clear problem description, research objectives, research scope, and the overall significance of the research.

Acquiring data and studying relevant articles related to fake image detection were the main objectives. This involved a detailed analysis of the literature from numerous sources, including internet sites, journal articles, research papers, and publications. This activity aimed to obtain a thorough understanding of the methods and models used to identify fake photos using deep learning algorithms, such as CNN, ResNet, and many more. It also included architectural designs, training techniques, evaluation metrics, and performance results.

The last activity in phase 1 was data collection. During this activity, a comprehensive dataset from Kaggle was successfully obtained. This dataset, known as the CASIA dataset, proved to be a valuable resource for the research project on using deep learning algorithms to detect fake images. The CASIA dataset comprised a total of 7,492 files of real images and 5,123 files of fake images, providing a substantial amount of data for analysis and model training.

3.2. Phase 2

In this phase, the focus was more concentrated on developing the model for the detection of fake images and real images using deep learning algorithms. It consisted of the models used for this project for data training to detect fake images and real images, plus the system architecture for this project.

In this activity for phase 2, the deep learning model's architecture was designed differently for each model., and the model uses advanced techniques which CNNs, Visual Geometry Group 16 (VGG16), and Residual Neural Network (ResNet). CNNs influenced the development of the deep learning model significantly. CNNs were an excellent choice for detecting fraudulent photographs due to their superior performance in image classification tasks. The model's performance and capabilities were then tested using the Visual Geometry Group 16 (VGG16) model, a well-known and extensively used CNN architecture that was explicitly used in the implementation approach.

Because of its deep layers and effective feature extraction capabilities, the VGG16 model was invaluable. Feature extraction capabilities of the VGG16 model were an invaluable basis for developing a strong and reliable false picture detection system. ResNet identifies itself by thoroughly analyzing the subtle patterns and unique properties present in an image. Its strength is in its capacity to comprehend minute details, allowing it to distinguish between legitimate and forged images.

Testing the models was crucial. The main goal was to evaluate their performance and identify the one with the highest accuracy in detecting fake and real photos. To test these models, various experiments were used to make sure the models had the best performance for the dataset used, with a focus on accuracy, precision-recall, and F1 score. The evaluation process included comparing each model's predictions to the actual labels provided to the photos. This comparison allowed the accuracy of each model in correctly identifying the photos to be determined.

In this project, a dataset of both actual and fake images was compiled. The photos were then pre-processed, which included tasks like scaling and normalization, and subjected to ELA, which enabled the identification of potential discrepancies in compression levels between actual and fake images to prepare them for future analysis. Following that, the dataset was then divided into training and testing sets to simplify the evaluation of various models, which used

 ISSN 2723-6471

80 training and 20 validations for default model training and 70: 20 for experiment to find the performance. Next, each model's performance and accuracy were shown in their graph and confusion matrix. Then, the model is integrated into the system prototype for user interactions.

3.3. Phase 3

This Phase is focused on the result of accuracy when combining ELA and, after the experiment, finding the best deep learning model to create documentation, which offers in-depth justifications, analyses, and graphic representations of the outcomes, was intended to be used for final year reports.

The process of developing the user interface for this system began with a thorough understanding of user requirements. The interface was designed with simplicity in consideration, allowing users to effortlessly upload photos in common formats such as JPG via drag-and-drop or browse functions. The interface and the backend detection module communicated efficiently and integrated with the deep learning model. The output showed a clear finding that indicated if the submitted image was real or false, along with a confidence score.

Final report documentation is the last activity, each phase's actions are reported or documented for the project report. It provides a clear overview of the entire project. It included an introduction to the topic, a literature review summarizing relevant research, and a detailed description of the methodology. Plus, the Presentation must be done after the report is completed.

4. Results and Discussion

During the project's data-gathering phase, a total of 7,492 fake photos and 5,123 real images were collected from Kaggle in JPEG and TIF type of images. This huge collection includes a different collection of photos, including animals, humans, architectural structures, fruits, and many other kinds of other categories.

The intentional inclusion of this diverse imagery ensures a comprehensive and different dataset, allowing for a thorough investigation of deep learning algorithms' effectiveness in detecting authenticity across a wide range of visual content. A subset of 5,123 fake and real photos was used throughout the model training and validation phase, making it a total of 10246 both fake and real images. This choice was made to ensure that all image categories were equally in the training and validation datasets by using an equal amount of false and genuine photos, allowing the model to learn and generalize effectively across both categories of images.

The obtained data goes through an important preprocessing stage to remove noise and ensure that the data is optimized and ready for use in the next phases of analysis and modeling. This phase includes procedures for enhancing and cleaning the dataset. The data preprocessing for this data collection is to resize the image to 128px of weight and 128px of height to improve data handling, enhance computing efficiency, and assist in network design. Next, ELA is being applied to the dataset as mentioned in methodology, using this method can improve feature representation for example, it highlights subtle modifications that are not easily visible to the human eye plus provide additional insights into how the model arrives at its predictions. Following the ELA, The next crucial step is to increase the brightness of the ELA image.

When the photos are subjected to ELA, a clear distinction between the original and processed versions emerges. The ELA procedure appears to introduce specific adjustments, emphasizing variances in image properties such as color intensity, pixelation, and contrast, potentially indicating manipulation or alterations within the photos, when applied to true and fake images, ELA behaves differently. Real photos have more white dots, showing parts of the original image with no other color of dots to make the image genuine. Fake photos, on the other hand, have a complicated pattern of colorful dots, implying numerous layers of adjustments or more advanced modifications. This distinction in ELA patterns assists in distinguishing between actual and false photos, providing useful information regarding the nature of changes within each category.

4.1. Experiment 1 Initial Model Configuration

The first experiment began with an investigation of model training using a specific configuration. This first configuration included a 30-epoch training typically, with each epoch representing a complete cycle of the dataset through the model. This setting lays the framework for later evaluations, with a batch size of 40 indicating the number of samples processed before updating the model's parameters, a learning rate of 0.001 regulating the step size in altering the model during training, and the ratio to split the dataset is 80% percent of training which contains 8196 and 20 % of validations which contain 2050 of fake and real images as shown in figure 1, figure 2, and figure 3.

ISSN 2723-6471

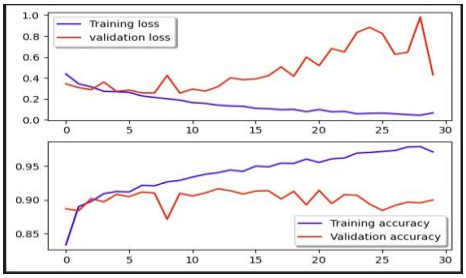

Figure 1. Cnn Experiment 1 Result

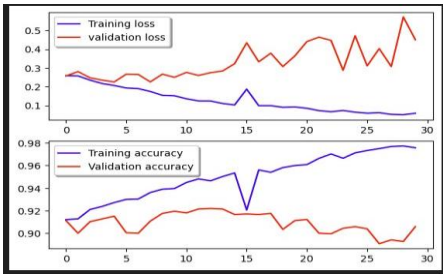

Figure 2. Vgg16 Experiment 1 Result

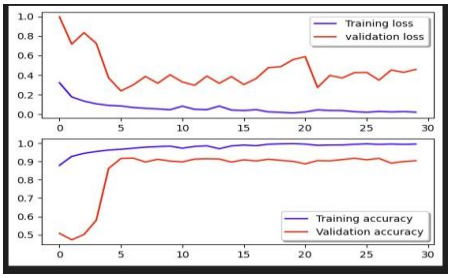

Figure 3. Resnet Experiment 1 Result

Specific parameters were fine-tuned when configuring the VGG16 and ResNet50 models to enable appropriate data augmentation. For VGG16, parameters such as rotation, width and height shift ranges, shear, zoom, and horizontal flipping were set to 30, 0.2, 0.2, 0.2, 0.2, and 'True,' allowing the model to effectively modify the dataset by rotating images up to 30 degrees, shifting them within certain proportions horizontally and vertically, applying shearing effects, zooming, and performing horizontal flips, all while using a 'nearest' fill mode.

Meanwhile, identical augmentation parameters were altered with different magnitudes for ResNet50; rotation was set to 40, and broader width and height shift, shear, and zoom ranges were set to 0.3, 0.3, 0.4, and 0.3, respectively. Both models were programmed to use horizontal flipping and the 'nearest' fill mode, with each model fine-tuned to meet its architecture and data requirements to improve adaptability and robustness when processing various image datasets.

4.2. Experiment 2 Result

Experiment 2 (see figure 4, figure 5, and figure 6) shows that the model learns slowly, with a learning rate of 0.0001. It uses early stopping to ensure that it does not overlearn or practice without significant improvement within 100 epochs. This method allows the model to learn more carefully and avoids wasting time training when it isn't improving much.

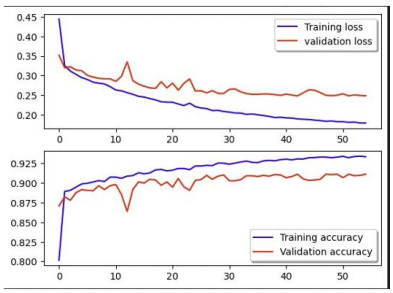

Figure 4. CNN Results in Experiment 2

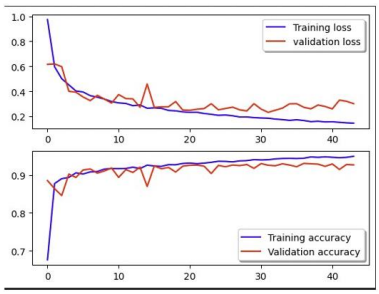

Figure 5. Vgg16 Results in Experiment 2

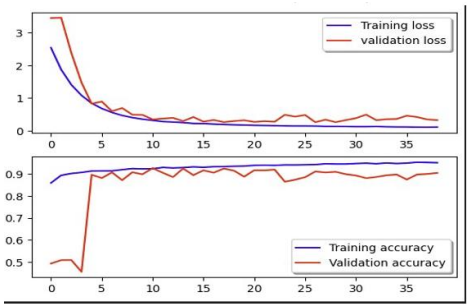

Figure 6. Resnet Results in Experiment 2

Across all aspects, the models performed well. The CNN showed effective early stopping at 55 epochs and achieved a peak training accuracy of 93% and a peak validation accuracy of 91%. VGG16 performed somewhat better, achieving 95% training accuracy and 93% validation accuracy while training in 44 epochs, as shown in figure 6. The ResNet model performed similarly to VGG16, achieving a matching accuracy for both training and validation and completing its training in 39 epochs. The continuous and significant reduction in training and validation loss for all three models, which remained below 1.0 throughout training, was impressive. This pattern shows that the models have strong learning and outstanding generalization ability.

4.3. Experiment 3 Result

For this experiment, the dataset split was changed from 80% training containing 8196 images and 20% validation of 2050 images to a revised split of 70% training, which equals 7172 images, and 30% validation, which equals 3074 images. The purpose of this was to allocate a bigger percentage to the validation set and to experiment on how the model reacted to the change, allowing the model to evaluate its performance on a larger subset of unseen data. This change was made to give a more comprehensive evaluation of the model's generalization capabilities.

In the outcome of this experiment, The CNN showed the same in training accuracy, recording 93%, and validation accuracy increased as experiment 2, recording 92%, with training ending after 63 epochs. The VGG16 model, on the other hand, had a lower training accuracy of 94% but a higher validation accuracy of 94% after 36 epochs. Lastly, Experiment 2's Residual Network (ResNet) maintained comparable training and validation accuracy levels, but the training ended after 33 epochs. Training and validation loss of those three models maintained its performance, which the loss is continuously decreasing across the epochs. Experiments 2 and 3 show comparable training and validation accuracies across all three models. However, a significant difference appeared in the graphical representation of the results. Experiment 2 demonstrated a more consistent and improved learning process by displaying a smoother training curve for all three models. Experiment 3 produced a somewhat less uniform graph, indicating a less smooth training process across the models.

4.4. Experiment 4 Result

In the final Experiment, a dataset split of 80% for training and 20% for validation was used based on the findings from Experiment 3, which exhibited improved accuracy and smoother model training with this split. Furthermore, the experiment configuration included a batch size reduction, changing from 32 to 28 batch sizes across the CNN, ResNet, and VGG16 models. This change attempted to take advantage of the benefits of smaller batch sizes, such as increased model convergence, greater model generalization, and a potentially finer-grained learning process. Smaller batch sizes frequently encourage a more extensive investigation of individual data instances per iteration, allowing for more precise adjustments to model parameters and encouraging a more meticulous learning pattern across models.

The CNN model obtained a peak training accuracy of 94% and a peak validation accuracy of 92% in the final experiment, with early stopping at 73 epochs. Moving on to VGG16, it had the best training accuracy of 95%, as well as the highest matching validation accuracy of 92%, stopping at 46 epochs. Meanwhile, RESNET maintained a high training accuracy of 95%, as well as an admirable validation accuracy of 93% with 32 epochs. Although the accuracy changes between experiments 2, 3, and 4 were minor, the graphical representations of both loss and accuracy revealed significant impacts on the models across the experiment, except for VGG16 in this Experiment, which shows the loss is going higher when at the end of the epochs. Experiment 3 highlighted VGG16 with the best graph display, demonstrating consistent performance. CNN's graph, on the other hand, did not show substantial variation between experiments, but RESNET presented its ideal graph in Experiment 4, indicating a development in model behavior throughout the experiment.

4.5. Best Result Confusion matrix

The Confusion Matrix visualizes the model's performance and overall effectiveness in classification tasks. The model with the best performance will be chosen for integration into the system prototype. Following a thorough examination that considered both loss and accuracy, the CNN model in Experiment 4 came out as the best performer.

It demonstrated the most promising performance among the iterations, with a stunning training accuracy of 94% and a validation accuracy of 92%. Next is VGG16, where the best Performance is from Experiment 3, with training and validation both having the same accuracy, which is 94% and lastly, ResNet has the best performance, which is 95% training and 93% validation accuracy, also in experiment 4.

Table 1 outlines the performance metrics for the CNN model in detecting real and fake images. For the real class, the precision is 0.97, indicating that 97% of the images classified as real are correctly identified. The recall of 0.86 shows that the model successfully detects 86% of all real images, while the F1 score of 0.92 demonstrates a balanced performance between precision and recall. For the fake class, the precision is 0.88, indicating that 88% of the images predicted as fake are accurate. The recall of 0.97 reflects the model's ability to identify 97% of all fake images, with an F1 score of 0.93 showcasing strong overall performance.

Table 1. CNN Performance Metrics

CNN Performance Metrics		
Class	**Metrics**	**Calculation**
Real	Precision	TP / (TP + FP) = 0.97
	Recall	TP / (TP + FN) = 0.86
	F1 Score	2 * (Precision * Recall) / (Precision + Recall) = 0.92

ISSN 2723-6471

CNN Performance Metrics		
Class	**Metrics**	**Calculation**
Fake	Precision	TN / (TN + FN) = 0.88
	Recall	TN / (TN + FP) = 0.97
	F1 Score	2 * (Precision * Recall) / (Precision + Recall) = 0.93

Table 2 presents the performance metrics for the VGG16 model. For the real class, the precision is 0.98, highlighting a very high accuracy in identifying real images, while the recall is 0.89, showing that 89% of all real images are detected. The F1 score of 0.93 indicates robust and balanced performance. In the fake class, the precision is 0.90, and the recall is 0.98, suggesting the model's nearly perfect ability to detect fake images. The F1 score of 0.94 underscores the superior performance of VGG16 in this class.

Table 2. Vgg16 Performance Metrics

CNN Performance Metrics		
Class	**Metrics**	**Calculation**
Real	Precision	TP / (TP + FP) = 0.98
	Recall	TP / (TP + FN) = 0.89
	F1 Score	2 * (Precision * Recall) / (Precision + Recall) = 0.93
Fake	Precision	TN / (TN + FN) = 0.9
	Recall	TN / (TN + FP) = 0.98
	F1 Score	2 * (Precision * Recall) / (Precision + Recall) = 0.94

Table 3 summarizes the performance metrics for the ResNet model. For the real class, the precision is 0.91, showing that most predictions for real images are correct. The recall is 0.93, indicating that the model captures 93% of all real images, while the F1 score of 0.92 reflects a good balance between precision and recall. For the fake class, the precision is 0.92, and the recall is 0.90, suggesting that 90% of all fake images are correctly identified. The F1 score of 0.91 demonstrates reliable performance.

Table 3. Resnet Performance Metrics

CNN Performance Metrics		
Class	**Metrics**	**Calculation**
Real	Precision	TP / (TP + FP) = 0.91
	Recall	TP / (TP + FN) = 0.93
	F1 Score	2 * (Precision * Recall) / (Precision + Recall) = 0.92
Fake	Precision	TN / (TN + FN) = 0.92
	Recall	TN / (TN + FP) = 0.9
	F1 Score	2 * (Precision * Recall) / (Precision + Recall) = 0.91

Overall, the VGG16 model achieves the highest performance among the three models, particularly in detecting fake images, as evidenced by its superior F1 scores in both classes. These findings suggest that VGG16 is the most effective model for this specific task. Figure 7 shows the Confusion Matrix, which displays the classification outputs and specific performance parameters of CNN, VGG16, and RESNEST optimal performance for this project.

 ISSN 2723-6471

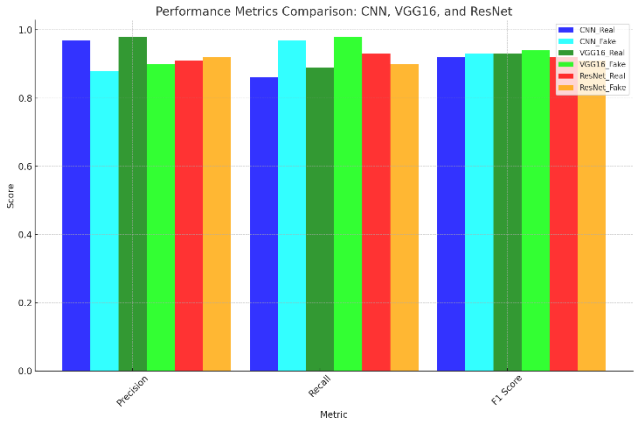

Figure 7. Performance Metrics Comparison: CNN, VGG16, and ResNet

The True Positives (TP) are cases in which the model correctly predicts the positive class, this case correctly classifying real photos as real. True Negatives (TN): These are scenarios where the negative class predicted correctly, correctly detecting forged images as fake. False Positives (FP) occur when the model predicts the positive class inaccurately. In this case, it means that the model misidentifies a fake image as a real one. Lastly, False Negatives (FN) are occasions where the model wrongly predicts the negative class, which means it incorrectly labels a real image as fake.

4.6. System Prototype

This prototype system is a Python-based image analysis tool that makes use of a variety of technologies. The system, which was created with Tkinter for the graphical user interface, includes PIL (Pillow) for image processing tasks such as scaling and enhancement. Keras incorporates a pre-trained ResNet model for image classification, while NumPy aids with numerical computations. As a first step, the system performs preprocessing on user input images. It prepares images by standardizing their sizes using techniques such as ELA to ensure consistent model input. Furthermore, the system normalizes pixel values to align them with the model's processing requirements. The system begins with the preprocessing of user-input images by using ELA to improve image quality and detect suspected tampering. Images undergo critical changes using ELA, such as standardizing sizes for consistent model input and improving visual contrast.

ResNet's capacity to understand small details enables it to discover complicated patterns in data, potentially improving its ability to generalize to real-world images. Furthermore, despite almost worse numerical performance in certain evaluation criteria, characteristics such as processing efficiency and deployment practicality benefit ResNet, making it more suitable for practical applications.

5. Conclusion

This project is set out to be an important tool for analyzing fake and authentic photos to prevent scams and other deceptive tactics. The ability to accurately recognize the authenticity of photos is crucial in many domains, including decision-making, user security and privacy, and maintaining trust in online interactions. For example, the large number of modified and manipulated images creates significant issues, particularly in the fields of social media, where visual content is widely shared. Detecting fake pictures has become critical in the fight against misinformation, fake news, and online scams. Within the system prototype, users must choose photographs suspected of being fake. The method predicts the legitimacy of the images once they have been selected. While the predictions may not always match the real picture authentication completely, this technique acts as a beginning step, providing insights into the possible authenticity of the selected images.

The focus on implementing Deep Learning approaches comes from their outstanding performance in image detection tasks. The strength of Deep Learning resides in its ability to understand subtle patterns and features inside images, allowing it to distinguish between genuine and modified images. This second more focus on model development and comparison, this objective is to find the best performance between the models chosen. After going through a process of literature review and examination of existing research, methodologies, and advancements of Deep Learning applied to image authentication, The final model to use in this project came out with three deep learning models, which is CNN, VGG16 and ResNet, these models have their unique and for images detection. The comparison is for finding the best model performance overall in this project. The third objective was to develop a system prototype to check whether

the image is real or fake, the primary focus is on the system's functionality related to user interactions for determining if provided images are fake or authentic.

This project has several improvements that need to be applied to the works. Firstly. The model used, expanding the model comparison for detecting fake and real photos by including more models such as Xception and Densely Connected Convolutional Network (DenseNet), might widen the scope of evaluation and potentially lead to the discovery of more successful algorithms for this purpose. Second, by improving the dataset preprocessing phase by adding highly effective approaches such as filtering or denoising techniques, as well as Exploratory Data Analysis in change with Error Level Analysis, model performance might be greatly improved. Finally, making the system prototype more appealing and user-friendly may attract and engage more people, hence increasing its utility and adoption.

References

[1] S. Selva Birunda, P. Nagaraj, S. Krishna Narayanan, K. Muthamil Sudar, V. Muneeswaran, and R. Ramana, "Fake Image Detection in Twitter using Flood Fill Algorithm and Deep Neural Networks," in *Proceedings of the Confluence 2022 - 12th International Conference on Cloud Computing, Data Science and Engineering*, vol. 2022, no. 1, pp. 285–290, 2022. doi: 10.1109/Confluence52989.2022.9734208.

[2] Q.-U.-Ain, N. Nida, A. Irtaza, and N. Ilyas, "Forged Face Detection using ELA and Deep Learning Techniques," in *Proceedings of 18th International Bhurban Conference on Applied Sciences and Technologies, IBCAST 2021*, vol. 2021, no. 1, pp. 271–275, 2021. doi: 10.1109/IBCAST51254.2021.9393234.

[3] S. Pawar, G. Pradhan, B. Goswami, and S. Bhutad, "Identifying Fake Images Through CNN Based Classification Using FIDAC," in *2022 International Conference on Intelligent Controller and Computing for Smart Power, ICICCSP 2022*, vol. 2022, no. 1, pp. 1–5, 2022. doi: 10.1109/ICICCSP53532.2022.9862034.

[4] Y. Wang, V. Zarghami, and S. Cui, "Fake Face Detection Using Local Binary Pattern and Ensemble Modeling," in *Proceedings - International Conference on Image Processing, ICIP*, vol. 2021, no. 9, pp. 3917–3921, 2021. doi: 10.1109/ICIP42928.2021.9506460.

[5] P. He and H. L. H. W., "Detection of Fake Images via the Ensemble of Deep Representations from Multi-Color Spaces," in *Proceedings - IEEE Conference on Image Processing*, vol. 2019, no. 1, pp. 1–7, 2019.

[6] N. Khatri, V. Borar, and R. Garg, "A Comparative Study: Deepfake Detection Using Deep-learning," in *Proceedings of the 13th International Conference on Cloud Computing, Data Science and Engineering, Confluence 2023*, vol. 2023, no. 1, pp. 1–5, 2023. doi: 10.1109/Confluence56041.2023.10048888.

[7] R. Rafique, M. Nawaz, H. Kibriya, and M. Masood, "DeepFake Detection Using Error Level Analysis and Deep Learning," in *Proceedings - 2021 IEEE 4th International Conference on Computing and Information Sciences, ICCIS 2021*, vol. 2021, no. 1, pp. 1–6, 2021. doi: 10.1109/ICCIS54243.2021.9676375.

[8] N. Rasiwasia and N. V., "A Systematic Study of the Role of Context on Image Classification," in *Proceedings - IEEE Conference*, vol. 2019, no. 1, pp. 1–8, 2019.

[9] R. Kumari and S. Kr., "Machine Learning: A Review on Binary Classification," in *International Journal of Computer Applications*, vol. 160, no. 7, pp. 11–15, 2020. doi: 10.5120/ijca2017913083.

[10] A. Kumar Singh, C. Sharma, and B. Kumar Singh, "Image Forgery Localization and Detection using Multiple Deep Learning Algorithm with ELA," in *Proceedings - IEEE Conference*, vol. 2023, no. 1, pp. 123–128, 2023. doi: 10.1109/icfirtp56122.2022.10059408.

[11] D. Agrawal, H. Makwana, S. S. Dave, S. Degadwala, and V. Desai, "Error Level Analysis and Deep Learning For Detecting Image Forgeries," in *Proceedings - 7th International Conference on Computing Methodologies and Communication, ICCMC 2023*, vol. 2023, no. 1, pp. 114–117, 2023. doi: 10.1109/ICCMC56507.2023.10084286.

[12] S. Indolia, A. K. Goswami, S. P. Mishra, and P. Asopa, "Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach," in *Procedia Computer Science*, vol. 132, no. 5, pp. 679–688, 2018. doi: 10.1016/j.procs.2018.05.069.

[13] L. Lan et al., "Generative Adversarial Networks and Its Applications in Biomedical Informatics," in *Frontiers in Public Health*, vol. 8, no. 1, pp. 1–20, 2020. doi: 10.3389/fpubh.2020.00164.

[14] F. Marra, D. Gragnaniello, D. Cozzolino, and L. Verdoliva, "Detection of GAN-Generated Fake Images over Social Networks," in *Proceedings - IEEE 1st Conference on Multimedia Information Processing and Retrieval, MIPR 2018*, vol. 2018, no. 1, pp. 384–389, 2018. doi: 10.1109/MIPR.2018.00084.

 ISSN 2723-6471

[15] J. Raja, P. Shanmugam, and R. Pitchai, "An Automated Early Detection of Glaucoma using Support Vector Machine Based Visual Geometry Group 19 (VGG-19) Convolutional Neural Network," in *Wireless Personal Communications*, vol. 118, no. 1, pp. 523–534, 2021. doi: 10.1007/s11277-020-08029-z.

[16] Y. Pan, G. Zhang, and L. Zhang, "A spatial-channel hierarchical deep learning network for pixel-level automated crack detection," in *Automation in Construction*, vol. 119, no. 1, pp. 103357, 2020. doi: 10.1016/j.autcon.2020.103357.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2016, no. 1, pp. 770–778, 2016. doi: 10.1109/CVPR.2016.90.

[18] Z. Zhu, W. Zhai, H. Liu, J. Geng, M. Zhou, C. Ji, and G. Jia, "Juggler-ResNet: A Flexible and High-Speed ResNet Optimization Method for Intrusion Detection System in Software-Defined Industrial Networks," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 1, pp. 4224–4233, 2022.

[19] H. M. D. Kabir et al., "SpinalNet: Deep Neural Network With Gradual Input," in *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 5, pp. 1165–1177, 2023. doi: 10.1109/TAI.2022.3185179.

[20] S. Ajibade, A. Zaidi, S. Maidin, W. Ishak, and A. Adetunla, "A Quantitative Based Research on the Production of Image Captioning," in *International Journal of Intelligent Systems and Applications in Engineering*, vol. 2023, no. 4, pp. 1–10, 2023.