International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

Enhancing Image Manipulation Detection through Ensemble ELA and Transfer Learning Techniques

Musaddik Habib Shirsho[1], Md Masud Rana[1], Jesmin Akhter[2], Abu Sayed Md. Mostafizur Rahaman[3],

[1]Department of Information and Communication Technology, Bangladesh University of Professionals (BUP), Dhaka, Bangladesh.
[2]Institute of Information Technology, Jahangirnagar University, Dhaka, Bangladesh
[3]Department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh.

E-mail address: musaddik.habib47@gmail.com, masudrana5772@yahoo.com, jesmin@juniv.edu and asmmr@juniv.edu

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

Abstract: Image manipulation techniques, such as copy-move, splicing, and removal methods, have become increasingly sophisticated, challenging the credibility of digital media. These techniques manipulate images at the pixel level, often leaving traces of tampering that can be detected through pixel-by-pixel analysis. This research introduces an innovative ensemble methodology that merges Error Level Analysis (ELA) with transfer learning leveraging deep convolutional neural networks (CNNs) to enhance image manipulation detection. The study involves extensive experimentation with various deep learning architectures and classifiers, with a focus on utilizing the CASIA1 and CASIA2 datasets for evaluation. The findings highlight that the combination of ResNet50V2 and ResNet101V2 models with Random Forest as the classifier exhibits superior performance compared to alternative ensemble techniques. This optimal configuration demonstrates high accuracy in discriminating between manipulated and unaltered images. The research emphasizes the significance of ensemble strategies in the realm of image manipulation detection, underscoring their potential for boosting detection accuracy and ensuring robust generalizability. The outcomes of this investigation shed light on the effectiveness of combining ELA and transfer learning for improved image authenticity assessment, providing valuable insights for advancing detection methodologies in the field. Here we achieved a promising outcomes, particularly with the Random Forest classifier, which attained accuracies of 97.671% and 92.497% on deep learning for the CASIA1 and CASIA2 datasets, respectively

Keywords: Image Manipulation, Image Manipulation Detection, Error Level Analysis (ELA), Transfer Learning, CNNs, Ensemble Methods, CASIA Datasets

1. INTRODUCTION

Image manipulation refers to the process of altering or modifying digital images using various techniques and software. It involves making changes to the content, appearance, or composition of an image to achieve a desired result. Image manipulation can be done for various purposes, such as enhancing the visual appeal of a photograph, removing imperfections or unwanted elements, or creating artistic effects. With the advancement of technology and the availability of sophisticated software like Adobe Photoshop, image manipulation has become more common, especially in the entertainment industry [1].

Image manipulation techniques, such as copy-move, splicing, and removal methods, have become increasingly sophisticated, challenging the credibility of digital media. These techniques manipulate images at the pixel level, often leaving traces of tampering that can be detected through pixel-by-pixel analysis. ML-based approaches can be used to detect and verify fraudulent or tampered images, helping to overcome forgery attacks [2]. On the other hand, machine learning-based approaches in image forensics aim to enhance the robustness of manipulation detection. These approaches utilize machine learning frameworks to learn low-level image attributes and detect these attributes in other images [3]. Additionally, deep neural network methods have been proposed for complex image manipulations, allowing for general image changes by modifying the input representation [4].

The research introduces an ensembled method for digital image manipulation detection, integrating error

E-mail:author's email

level analysis (ELA) and pretrained deep learning models. The study's contributions include proposing a robust detection approach, utilizing ELA to identify manipulation areas, and leveraging pretrained models to enhance detection performance. Using two different datasets demonstrates the method's effectiveness and robustness. The paper is structured to provide a comprehensive review of related work, detailed the methodology, present results and analysis, and conclude with future research directions.

2. LITERATURE REVIEW

Image forgery detection using machine learning is an important area of research in digital forensics and cyber security. Different machine learning-based approaches, such as feature-based schemes with machine learning and methods based on deep learning, have been explored [5]. Machine learning algorithms, such as Support Vector Machine (SVM) and k-nearest neighbors (k-NN), have been used for forgery detection [6]. These algorithms analyze large datasets of images and learn to recognize patterns and features indicative of forgery [7].

The use of machine learning-based techniques has improved the accuracy and speed of forgery detection compared to conventional statistical methods [8]. SVM has been used for feature extraction and reduction, leading to quick results in forgery detection under various test conditions [9]. The ELA method can be enhanced by converting RGB format images to ELA and using them to train deep learning models. This approach has shown promising results, with high validation accuracy and outperforming cutting-edge methods in terms of speed [10].

Furthermore, the ELA method can be used in conjunction with convolutional neural networks (CNNs) to detect various types of image forgeries, including im-age splicing and copy-move [11] [12]. By incorporating these advancements, the ELA method becomes more effective and reliable in identifying hidden forgeries in images.

The proliferation of editing tools and online platforms has led to a surge in fake images, necessitating the development of robust forgery detection techniques. Addressing the prevalent Copy Move Forgery (CMF) challenge, a transfer learning-based approach utilizing Deep Convolutional Neural Networks (Deep CNNs) pretrained with GoogLeNet parameters is proposed. This method, augmented by a novel optimization algorithm called Fractional Leader Harris Hawks Optimization (FLHHO), achieves notable effectiveness, demonstrated by high testing accuracy (0.930), True Negative Rate (TNR) of 0.938, and True Positive Rate (TPR) of 0.941 [13].

The study suggests using ResNet-50 with 25 convolutional layers and ImageNet as a feature extractor to diagnose anomalies in images of bottles, spoons, and cartons, achieving high prediction accuracies of 99%, 95%, and 90% for the datasets, respectively [14]. A new image splicing detection method based on deep learning and transfer learning has been proposed to improve accuracy, reduce training time, and simplify model complexity. By leveraging a pre-trained MobileNetV2 model and transfer learning, the approach achieves state-of-the-art accuracy in detecting spliced images with minimal training data and time requirements [15].

Deep learning techniques offer improved image forgery detection by extracting complex features from images, surpassing traditional methods. This advancement addresses the challenges posed by technological advancements in image editing software, ensuring the authenticity of images in various communication mediums [16]. Detecting digital image tampering is vital due to the widespread use of manipulated images for deceptive purposes. This paper reviews various methods, including advanced deep learning techniques, to enhance image forgery detection and ensure the integrity of digital photos [17].

Comparative Study between Significant Relevant Research Works

Paper Title	Summarized Abstract	Methods Used	Limitations
An approach for copy-move image multiple forgery detection based on an optimized pre-trained deep learning model [13]	- The paper proposes a transfer learning-based method for detecting copy-move image forgery. - The method utilizes a pre-trained deep CNN model and an optimization algorithm.	- Transfer learning-based method utilizing a Deep Convolutional Neural Network (Deep CNN) Optimization algorithm called Fractional Leader Harris Hawks Optimization (FLHHO)	- Deep learning-based techniques suffer from generalization issues. - The proposed method may have limitations in detecting other types of image forgeries.
Image anomalies detection using transfer learning of resnet-50 convolutional neural network [14]	- Deep learning used for data-based fault diagnosis in smart manufacturing. - Proposed ResNet-50 model achieved high prediction accuracy for anomalous images.	- Deep learning with ResNet-50 Transfer learning using ImageNet as a feature extractor	- Deep learning models for fault diagnosis have shallow depths compared to other areas. - Limited accuracy of final prediction due to small, seeded test size.
Image Splicing Detection based on	- Image splicing detection using deep learning and transfer learning -	- Feature engineering and machine learning-based	- Deep learning model requires large

Paper Title	Summarized Abstract	Methods Used	Limitations
Deep Convolutional Neural Network and Transfer Learning [15]	Proposed model achieves high accuracy with less training data	detection - Deep learning-based detection with automatic feature extraction	training data - Deep learning model is time-consuming and costly
A Review on Deep Learning Techniques for Image Forgery Detection [16]	- Image forgery detection is crucial due to the widespread use of images in communication, but traditional methods relying on handcrafted features have limitations. - Deep learning techniques offer promising results for detecting image tampering by extracting complex features, leading to better performance compared to traditional methods.	- Traditional methods use handcrafted features for image forgery detection. - Deep learning techniques are used for image tampering detection.	- Traditional methods for image forgery detection have limitations in identifying specific types of tampering. - Deep learning techniques have better performance in extracting complex features from images.
Comprehensive analyses of image forgery detection methods from traditional to deep learning approaches: an evaluation [17]	- Digital images serve as crucial evidence in various fields, but they are vulnerable to tampering using readily available editing software. - Image tamper detection methods, including both traditional and advanced deep learning approaches, are evaluated in this study to address the challenge of distinguishing authentic images from manipulated ones.	- Evaluation of various image tamper detection methods. - Comparative study of image criminological (forensic) methods.	- Deep learning techniques have limitations in image forgery detection. - The paper addresses the limitations of recently developed deep learning techniques.

The study is guided by two main research areas:

Firstly, it explores the optimization of ensembled methods, which integrate Error Level Analysis (ELA) and transfer learning from pre-trained deep learning models, with the aim of improving the accuracy and robustness of digital image manipulation detection systems.

Secondly, it investigates the influence of various combinations of deep learning models and classifiers on the efficacy of the ensembled method in detecting both manipulated and authentic images.

3. METHODOLOGY

The algorithm begins by acquiring images from the CASIA1 and CASIA2 datasets. These datasets serve as the foundation for the subsequent analysis. The images are then preprocessed and converted into Error Level Analysis (ELA) formatted images. This preprocessing step is essential for enhancing the quality and consistency of the data before further processing.

Following the preprocessing stage, the data is divided into training (80%) and testing (20%) datasets. This division ensures that the models are trained on a substantial portion of the data while retaining a separate set for evaluation. The algorithm then proceeds to build deep learning models, including VGG16, VGG19, ResNet50V2, ResNet101V2, MobileNetV3Small, and MobileNetV3Large. These models are instrumental in extracting features from the images and capturing essential patterns.

After extracting image features using the deep learning models, the algorithm concatenates these features and prepares the corresponding labels. Subsequently, the features are preprocessed to optimize their compatibility with the classifier models. The algorithm then constructs classifier models such as Support Vector Machine (SVM), Random Forest, Gradient Boosting Classifier, K-Nearest Neighbors, and Naive Bayes. These classifiers are pivotal in categorizing and predicting image attributes based on the extracted features.

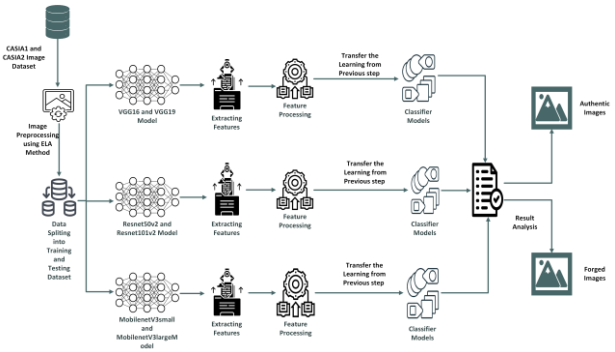

Figure 3.1: General Overview of the Research Model

Furthermore, the algorithm feeds the learnings from the feature extraction and preprocessing stages into the classifier models. To determine the best combination of classifiers, a voting classifier is employed. This approach

leverages the collective intelligence of multiple classifiers to enhance prediction accuracy. Finally, the algorithm compares and analyzes the outputs generated by the classifier models, providing valuable insights into the effectiveness of the image analysis process. Through this systematic approach, the algorithm facilitates comprehensive image analysis and enables informed decision-making based on the results.

3.1 Data Acquisition

The CASIA Version 1 Dataset consists of 921 authentic and 800 tampered images, created using Photoshop with various manipulations. CASIA1.0 for Testing contains 172 authentic and 288 tampered images, produced with software like GIMP and Paint.NET, lacking labels for tampering operations. Image forgery detection using the CASIA2.0 dataset has been explored, with methods including block processing and feature extraction using Convolutional Neural Networks (CNNs) [18].
The CASIA Version 2 dataset is a widely used benchmark for image tampering detection research. It contains 7,491 authentic images and 5,123 tampered images, covering various scenes and manipulation techniques. The tampered images are generated by two common operations: copy/pasting and image splicing. Another study presented a lightweight CNN model with four convolutional layers and four max-pooling layers, achieving high accuracy in detecting splicing forged images on the CASIA 2.0 dataset [19].

3.2 Error Level Analysis (ELA)

Error Level Analysis (ELA) is a technique crucial in detecting digital manipulations by resaving an image at a specific compression level and comparing it with the original to highlight areas of differing compression. Integrating ELA as a feature in deep learning models, such as Convolutional Neural Networks (CNNs), can enhance image forgery detection. Studies indicate that employing ELA can improve validation accuracy by approximately 2.7% and enhance test accuracy by aiding in identifying the true compression ratio of images and detecting fake images. However, the utilization of ELA may marginally slow down processing time by about 5.6%. [20]

In this research, the researcher used the ELA method for image preprocessing. The ELA method is much more popular compared to other traditional methods as it supports and converts the image in the Red Green Blue (RGB) format. The ELA algorithm is described below for better understanding the process.

Algorithm 1: Image Preprocessing using ELA

Image to ELA:

1. Check if the image format is supported (JPEG, PNG, BMP, TIFF).
2. Open the image and convert it to RGB mode.
3. Resave the image with the specified quality (90%).
4. Calculate the ELA (Error Level Analysis) image by taking the difference between the original and resaved image.
5. Get the minimum and maximum pixel values in the ELA image.
6. Scale the pixel values of the ELA image to the range [0, 255].
7. Save the ELA image with the same filename in the specified resave path.

Preprocess Data:

Loop through a directory and apply the *image*_to_ela function to each image in the directory.

This algorithm is used to generate ELA images from a set of original images. ELA is a technique that can be used to detect image tampering by identifying areas of the image where the pixel values have been modified. The algorithm first reserves the original images at a lower quality, and then calculates the difference between the original and resaved images. The resulting ELA image is then saved to a separate directory.

3.3 Deep Learning Models and Transfer Learning Overview

Transfer learning can be used when there is limited labeled data available for the target task, allowing the model to leverage knowledge learned from a related task with a larger amount of labeled data. This helps overcome the problem of overfitting and improves performance on the target task [21].
Transfer learning with the Deep learning model involves leveraging the pre-trained weights and architecture of the convolutional neural network (CNN) that has been trained on a large-scale dataset. Instead of training the model from scratch, transfer learning adapts the knowledge learned by the Deep learning model on a generic dataset to a specific task, such as image forgery detection in this context.

The process typically involves the following steps:

a) Dataset Selection: A comprehensive, well-labeled dataset that is broadly representative of the field of study is required for image classification to facilitate efficient model training. For this research, we have selected the CASIA dataset with both version one and two.

b) Image Preprocessing: Before training the models, the dataset undergoes preprocessing steps such as

resizing, normalization, and augmentation. These preprocessing techniques help enhance the model's ability to generalize across different inputs and improve robustness.

c) **Pre-trained Model Initialization:** The Deep learning model, pre-trained on a large dataset like ImageNet, is initialized with learned weights and architecture. This initialization captures general features and patterns from various images, enabling the model to recognize a wide range of visual concepts.

d) **Feature Extraction:** The pre-trained model serves as a feature extractor. Given a dataset of original and tampered images, the images are fed into the Deep Learning model, and the activations from one of the intermediate layers (often the last convolutional layer) are extracted as feature vectors. These feature vectors represent high-level semantic information about the input images.

e) **Fine-tuning or Feature Concatenation:** Depending on the specific task and dataset, there are two common approaches to transfer learning:

i. **Fine-tuning:** In this approach, the model undergoes fine-tuning on task-specific datasets to adapt its learned representations to the new dataset's characteristics.

ii. **Feature Concatenation:** Features extracted from pre-trained models are combined with handcrafted features or processed through additional layers to learn task-specific representations.

f) **Training Classifier:** A classifier (e.g., Support Vector Machine, Random Forest, etc.) is trained on the extracted features or the concatenated feature representation. This classifier learns to distinguish between authentic and tampered images based on the learned features.

g) **Evaluation and Fine-tuning:** The performance of the transfer learning approach is evaluated on a validation set. Depending on the results, further finetuning of hyperparameters or model architecture may be performed to im-prove performance.

3.3.1. Random Forest

Random Forest, an ensemble learning method, merges numerous decision trees to forecast outcomes, demonstrating proficiency in handling large and complex datasets. Renowned for its versatility, it excels in classification and regression tasks, mitigating challenges like overfitting and missing values while offering high predictive accuracy. [22]. They also have built-in protection against overfitting, reducing the risk of model performance deteriorating on new data [23].

3.3.2. KNeighbours

KNeighbours, a non-parametric classification algorithm, employs a majority voting scheme based on the K nearest neighbors to assign data points to classes. Renowned for simplicity and effectiveness, it finds utility across scientific fields, albeit encountering limited adoption in medical literature due to technical challenges. [24]. The kNN algorithm involves predicting outcomes based on the nearest neighbors in the dataset. Factors such as the choice of predictors, distance calculation, and the value of k can significantly impact the performance of the model [25].

3.3.3. GradientBoostingClassifier

GradientBoostingClassifier, a boosting algorithm, amalgamates numerous weak learners, typically decision trees, to construct a potent predictive model. Employing an iterative process, it progressively refines predictions by fitting new models to residuals, renowned for high accuracy in handling intricate datasets. However, the abstracts do mention the use of various classifiers for image forensics, such as blind forensic method [26], CNN-based multi-classifier [27], feature-based classifiers [28], and SVM-based forensic techniques [29]. These classifiers are used for tasks such as detecting adversarial images, source camera identification, document source printer identification, and identification of spliced im-ages.

3.3.4. SVM

Support Vector Machines (SVM) is a potent classifier known for its capability to delineate data points into distinct classes by identifying the optimal hyperplane. It prioritizes maximizing the margin between classes, rendering it resilient to outliers. SVM adeptly handles both linear and non-linear classification conundrums through varied kernel functions. Widely embraced in research for its adaptability and efficacy, SVM finds utility in machine learning image forensics for discerning genuine from counterfeit multimedia files, leveraging methods such as Discrete Fourier Transform (DFT) for feature extraction. [29].

3.3.5. Naive Bayes

Naive Bayes, a probabilistic classifier, employs Bayes' theorem assuming feature independence. Despite simplicity, it excels in various scenarios, especially text classification like sentiment analysis or spam detection, offering rapid training and low memory usage. [30]. It

is also used in the detection of melanoma skin cancer by analyzing dermoscopy images and extracting features [30]

4. Result and Discussion

The results obtained from the ensemble method applied to the CASIA1 dataset are quite promising. The accuracy scores achieved by the different classifiers indicate that the ensemble approach, combined with transfer learning from VGG16 and VGG19, is effective in detecting tampered images.

4.1 *VGG16 and VGG19 Model Results and Observations*

Table 4.1: Performance of VGG16 and VGG19 Models with Different Classifier Models for CASIA Version 1 Dataset

Dataset: CASIA1		
Aggregated Models	**Classifier Model**	**Accuracy**
VGG16 and VGG19	SVM	96.518%
	RandomForest	96.360%
	GradientBoostingClassifier	96.202%
	KNeighbours	95.727%
	Naive Bayes	95.253%

In this study, various classifiers were employed for image tampering detection, with SVM emerging as the top performer, achieving an accuracy score of 96.518%. SVM, renowned for its adeptness in handling complex datasets and high-dimensional feature spaces, demonstrated its suitability for this task. Following closely, the Random Forest classifier attained an accuracy score of 96.360%, leveraging ensemble learning to effectively capture intricate feature relationships.

The GradientBoostingClassifier, with an accuracy score of 96.202%, showcased its prowess in iterative model refinement for robust prediction. Additionally, KNeighbours achieved an accuracy score of 95.727%, while Naive Bayes attained 95.253%, both demonstrating commendable performance. KNeighbours, employing a simple yet effective majority voting scheme, and Naive Bayes, applying probabilistic classification with independence assumptions, proved their efficacy in image tampering detection.

4.2 Resnet50V2 and Resnet101V2 Model Results and Observations

Dataset: CASIA1		
Aggregated Models	**Classifier Model**	**Accuracy**
Resnet50V2 and Resnet101V2	RandomForest	97.671%
	KNeighbours	97.360%
	GradientBoostingClassifier	96.260%
	SVM	96.202%
	Naive Bayes	96.044%

Table 4.2: Performance of Resnet50V2 and Resnet101V2 Models with Different Classifier Models for CASIA Version 1 Dataset

The Random Forest classifier emerged as the most accurate among others, achieving an accuracy of 97.671%. Renowned for its proficiency in managing high-dimensional data and capturing intricate feature relationships, Random Forest's superior performance aligns with its inherent strengths. Similarly, the KNeighbours classifier achieved a notable accuracy of 97.360%, indicative of dis-tinct clusters within the dataset effectively separable by this non-parametric algorithm.

The GradientBoostingClassifier followed closely with an accuracy of 96.260%, leveraging ensemble learning to iteratively refine predictions. The SVM classifier achieved an accuracy of 96.202%, affirming its capability in handling high-dimensional data and effectively discerning between authentic and tampered images. Despite its simplicity, the Naive Bayes classifier attained an accuracy of 96.044%, underscoring the dataset's informative features conducive to discrimination.

4.3 MobilenetV3small and MobilenetV3large Model Results and Observations

Table 4.3: Performance of Mobilenetv3small and Mobilenetv3large Models with Different Classifier Models for CASIA Version 1 Dataset

Dataset: CASIA1		
Aggregated Models	**Classifier Model**	**Accuracy**
MobilenetV3 small and MobilenetV3 large	SVM	90.174%
	RandomForest	89.732%
	GradientBoostingClassifier	89.052%
	KNeighbours	88.271%
	Naive Bayes	87.046%

In aggregating predictions, hard voting was employed to determine class labels based on the majority votes from individual classifiers, aiming to enhance overall classification accuracy. Results reveal the SVM classifier's highest accuracy of 90.174%, signifying effective image classification within the CASIA1 dataset. RandomForest and GradientBoostingClassifier models also performed commendably, achieving accuracies of 89.732% and 89.052% respectively, indicating successful utilization of learned features from ensemble models for precise predictions.

Conversely, KNeighbours and Naive Bayes classifiers exhibited slightly lower accuracies of 88.271% and

87.046% respectively. Nonetheless, the outcomes underscore the potential of ensembled methods in optimizing classification accuracy, warranting further exploration for refining model configurations and addressing classifier limitations.

4.4. Classification report of the models with CASIA1 dataset

Table 4.4: Performance of All Models with Different Classifier Models for CASIA Version 1 Dataset

Dataset: CASIA 1										
SL	Aggregated Models	Classifier	F1 Score		Precision		Recall		Support	
			0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0
1	VGG16 andVGG19	SVM	.96	.97	.92	.99	.99	.94	286	246
2	Resnet50v2and Resnet101v2	Random Forest	.96	.97	.99	.94	.93	.99	286	346
3	MobilenetV3small and MobilenetV3large	SVM	.95	.95	.93	.98	.97	.95	286	246

The experimental results from three ensemble models trained on the CASIA1 dataset provide valuable insights into their effectiveness for image forgery detection. Each model, comprising unique combinations of architectures and classifiers, demonstrates varied performance outcomes, highlighting nuanced strengths and capabilities.

Model 1, integrating VGG16 and VGG19 architectures with an SVM classifier, exhibits robust performance metrics. With balanced F1 Scores of 0.96 for authentic images and 0.97 for tampered images, the model demonstrates effective classification capabilities. Notably, its high precision values of 0.92 for authentic images and 0.99 for tampered images minimize false positives and accurately identify instances of image forgery. However, a slight trade-off is observed in recall values, with 0.99 for authentic images and a slightly lower 0.94 for tampered images.

Model 2, featuring ResNet50v2 and ResNet101v2 architectures with a Random Forest classifier, showcases strong performance with balanced F1 Scores of 0.96 for both authentic and tampered images. The model achieves impressive precision values of 0.99 for authentic images and 0.94 for tampered images, indicating minimal false positives and high accuracy in detecting image tampering instances. Notably, commendable recall values of 0.93 for authentic images and 0.99 for tampered images further enhance its effectiveness in capturing genuine instances of both classes.

Model 3, comprising MobileNetV3small and MobileNetV3large architectures with an SVM classifier demonstrates consistent performance across both classes. With balanced F1 Scores of 0.95 for authentic and tampered images, the model exhibits high precision values of 0.93 for authentic images and an impressive precision of 0.98 for tampered images. Noteworthy recall values of 0.97 for authentic images and 0.95 for tampered images validate its effectiveness in capturing genuine instances of both classes. Overall, these findings highlight the promising performance of ensemble models in detecting image forgery and underscore their potential for real-world applications in various domains.

4.5. VGG16 and VGG19 Experimental Results and Observations with CASIA2 Dataset

Table 4.5: Performance of VGG16 and VGG19 Models with Different Classifier Models for CASIA Version 2 Dataset

Dataset: CASIA2			
SL	Aggregated Models	Classifier Models	Accuracy
1	VGG16 and VGG19	SVM	89.409%
		RandomForest	88.799%
		GradientBoostingClassifier	88.775%
		Naive Bayes	88.701%
		KNeighbours	87.091%

The research investigated an ensembled approach for image forgery detection utilizing transfer learning from pre-trained VGG16 and VGG19 models. This approach addresses the challenge of limited labeled data commonly encountered in image forensics. To consolidate decisions, a hard voting scheme with equal weight was employed for the ensemble, incorporating SVM, Random Forest, GradientBoostingClassifier, KNeighbours, and Naive Bayes classifiers.

The proposed ensemble method achieved promising accuracy results on the CASIA2 dataset, with SVM reaching 89.409%, RandomForest at 88.799%, GradientBoostingClassifier at 88.775%, Naive Bayes at 88.701%, and KNeighbours at 87.091%. While accuracy remains a key metric, a comprehensive evaluation should include additional metrics like precision, recall, and F1 score to ensure robust detection of tampered images.

4.6. Resnet50V2 and Resnet101V2 Experimental Results and Observations with CASIA2 Dataset

Table 4.6: Performance of Resnet50V2 and Resnet101V2 Models with Different Classifier Models for CASIA Version 2 Dataset

This research evaluated the performance of various machine learning classifiers for image forgery detection using the CASIA2 dataset. Among the tested models, Random Forest achieved the highest accuracy of 92.497%, attributing its success to its ensemble learning approach and ability to capture complex relationships within the data. Following closely was the SVM classifier with an accuracy of 91.289%, demonstrating its strength in separating data points in high-dimensional spaces.

GradientBoostingClassifier achieved an accuracy of 89.755%, utilizing its iterative approach to refine predictions and identify subtle relationships within the dataset. KNeighbours and Naive Bayes also performed well, reaching accuracy levels of 88.721% and 88.171% respectively, demonstrating the effectiveness of their fundamental classification approaches in identifying patterns and similarities within the data.

Overall, this evaluation highlights the potential of various machine learning techniques for image forgery detection. The diverse strengths showcased by each classifier offer valuable insights into the field and pave the way for further investigation into optimizing these techniques for even more robust and accurate results.

4.7. MobilenetV3small and MobilenetV3large Experimental Results and Observations with CASIA2 Dataset

Table 4.7: Performance of Mobilenetv3small and Mobilenetv3large Models with Different Classifier Models for CASIA Version 2 Dataset

Dataset: CASIA2			
SL	**Aggregated Models**	**Classifier Models**	**Accuracy**
1	MobilenetV3small and MobilenetV3large	RandomForest	89.824%
		KNeighbours	89.632%
		GradientBoostingClassifier	88.440%
		SVM	88.255%
		Naive Bayes	87.476%

The ensembled method for image forgery detection, which combines MobilenetV3small and MobilenetV3large models through transfer learning, demonstrates promising results by leveraging their collective knowledge. Incorporating multiple classifiers like SVM, RandomForest, GradientBoostingClassifier, KNeighbours, and Naive Bayes further enhances system performance. Through hard voting, the system ensures robust decision-making by aggregating the majority consensus from these classifiers, reducing the risk of erroneous classifications.

Dataset: CASIA2			
SL	**Aggregated Models**	**Classifier Models**	**Accuracy**
1	Resnet50V2 and Resnet101V2	RandomForest	92.497%
		SVM	91.289%
		GradientBoostingClassifier	89.755%
		KNeighbours	88.721%
		Naive Bayes	88.171%

Accuracies achieved, particularly with RandomForest and KNeighbours leading at 89.824% and 89.632% respectively, highlight the efficacy of the ensembled approach. While accuracy serves as a primary metric, evaluating precision, recall, and F1-score provides a comprehensive understanding of the system's effectiveness. Leveraging the widely recognized CASIA2 dataset ensures the reliability and generalizability of the findings, thus advancing the field of image forgery detection.

4.8. Classification Report of the models with CASIA2 Dataset

Table 4.8: Performance of All Models with Different Classifier Models for CASIA Version 2 Dataset

Dataset: CASIA2									
SL	**Aggregated Models**	**Classifier**	**F1 Score**		**Precision**		**Recall**		**Support**	
			0.0	**1.0**	**0.0**	**1.0**	**0.0**	**1.0**	**0.0**	**1.0**
1	VGG16 and VGG19	SVM	.83	.92	.74	.97	.93	.87	1138	2960
2	Resnet50v2 and Resnet101v2	Random Forest	.85	.93	.87	.98	.95	.91	1158	2940
3	MobilenetV3small and MobilenetV3large	Random Forest	.81	.85	.88	.92	.93	.87	1138	2940

The ensemble models analyzed in the study demonstrate varying levels of performance, as reflected in their F1 Scores, which range from moderate to relatively high. For instance, the VGG16-VGG19 ensemble with SVM achieves moderate F1 Scores of 0.83 for authentic images and 0.92 for tampered images. In contrast, the ResNet50v2-ResNet101v2 ensemble with Random Forest achieves relatively high F1 Scores of 0.85 for authentic images and 0.93 for tampered images, indicating superior performance compared to other models.

Additionally, precision values for tampered images consistently remain high across all models, suggesting minimal occurrences of false positives. This consistent trend underscores the effectiveness of the ensemble methods in accurately identifying tampered images while maintaining a low rate of false alarms.

While precision for authentic images differs across models, high recall values for both authentic and tampered images indicate effective capture of instances from both classes. Additionally, support values

demonstrate the models' ability to handle imbalanced datasets, with greater support for tampered images, mirroring re-al-world scenarios.

5. Conclusions

In conclusion, this research presents a novel approach to digital image manipulation detection by combining Error Level Analysis (ELA) with transfer learning from deep convolutional neural networks (CNNs). Through experimentation with various deep learning models and classifiers on the CASIA1 and CASIA2 datasets, we have demonstrated the effectiveness of ensemble methods in achieving accurate and reliable detection results.

The key success points of our research include:

- Proposing an ensemble method that integrates ELA and pre-trained deep learning models for image manipulation detection.
- Conducting comprehensive experimental evaluations to assess the performance of different ensemble architectures and classifier algorithms.
- Achieving promising outcomes, particularly with the Random Forest classifier, which attained accuracies of 97.671% and 92.497% for the CASIA1 and CASIA2 datasets, respectively.
- Demonstrating minimal false positives and high accuracy in identifying image tampering instances, highlighting the effectiveness of the ensemble model.

Our findings contribute significantly to the existing body of knowledge in computer vision, digital forensics, and cybersecurity. By systematically evaluating ensemble architectures and classifier algorithms, we provide valuable insights into the performance and effectiveness of various approaches in detecting image manipulation.

Moving forward, there are several research gaps and future works to consider:

- Diverse Datasets: Future research should explore larger and more diverse datasets to better represent real-world image manipulation scenarios.
- Model Optimization: Further optimization and fine-tuning of ensemble models are necessary to achieve even higher performance levels.
- Alternative Architectures: Exploring alternative ensemble architectures and incorporating advanced preprocessing techniques could enhance detection accuracy and efficiency.

In summary, our research underscores the efficacy of ensemble methods in digital image manipulation detection and offers insights for enhancing accuracy and generalizability. By addressing research gaps and pursuing future works, we aim to advance the field and contribute to the development of robust and reliable detection techniques.

References

[1] A. Rao, "PHOTOGRAPHY IN HOLLYWOOD: IMAGE MANIPULATION IN MODERN ENTERTAINMENT," Researchers World: Journal of Arts, Science and Commerce, vol. 8, no. 1, pp. 167-172, 2016.

[2] E. Nowroozi, A. Dehghantanha, R. M. Parizi, and K.-K. R. Choo, "A Survey of Machine Learning Techniques in Adversarial Image Forensics," arXiv.org, Oct. 19, 2020. https://arxiv.org/abs/2010.09680v1

[3] "US20190043210A1 - Machine learning based image processing techniques - Google Patents," US20190043210A1 - Machine learning based image processing techniques - Google Patents, Aug. 04, 2017. https://patents.google.com/patent/US20190043210A1/en

[4] Y. Vinker, E. Horwitz, N. Zabari, and Y. Hoshen, "Image Shape Manipulation from a Single Augmented Training Sample," arXiv.org, Jul. 02, 2020. https://arxiv.org/abs/2007.01289v2

[5] D. Das and R. Naskar, "Image Splicing Detection Using Feature Based Machine Learning Methods and Deep Learning Mechanisms," *Image Splicing Detection Using Feature Based Machine Learning Methods and Deep Learning Mechanisms | SpringerLink*, Jun. 21, 2022. https://link.springer.com/chapter/10.1007/978-981-19-3089-8_22

[6] M. Monika, A. Passi, and S. Passi, "Digital Image Forensic Based on Machine Learning Approach for Forgery Detection and Localization," Journal of Physics: Conference Series, vol. 1950, no. 1, p. 012035, 2021, doi: 10.1088/1742-6596/1950/1/012035.

[7] S. Mehta and P. Shukla, "A Review on Machine Learning-Based Approaches for Image Forgery Detection," *A Review on Machine Learning-Based Approaches for Image Forgery Detection | SpringerLink*, Jun. 16, 2023. https://link.springer.com/chapter/10.1007/978-981-99-1435-7_8

[8] "Comprehensive study on image forgery techniques using deep learning," *Comprehensive study on image forgery techniques using deep learning | IEEE Conference Publication | IEEE Xplore*. https://ieeexplore.ieee.org/document/10151540

[9] "Image Forgery Detection Using Machine Learning," *Image Forgery Detection Using Machine Learning | IEEE Conference Publication | IEEE Xplore*. https://ieeexplore.ieee.org/document/10046422

[10] "Image Forgery Localization and Detection using Multiple Deep Learning Algorithm with ELA," *Image Forgery Localization and Detection using Multiple Deep Learning Algorithm with ELA | IEEE Conference Publication | IEEE Xplore*. https://ieeexplore.ieee.org/document/10059408

[11] "Image Tampering Detection Using Error Level Analysis and Metadata Analysis," *Image Tampering Detection Using Error Level Analysis and Metadata Analysis | IEEE Conference Publication | IEEE Xplore*. https://ieeexplore.ieee.org/document/10169948

[12] "Performance Analysis of ELA-CNN model for Image Forgery Detection," *Performance Analysis of ELA-CNN model for Image Forgery Detection | IEEE Conference Publication | IEEE Xplore*. https://ieeexplore.ieee.org/document/10170007

[13] "An approach for copy-move image multiple forgery detection based on an optimized pre-trained deep learning model," *An approach for copy-move image multiple forgery detection based on an optimized pre-trained deep learning model - ScienceDirect*, Mar. 27, 2023. https://www.sciencedirect.com/science/article/abs/pii/S0950705123002587

[14] Z. T. Omer and A. H. Abbas, "Image anomalies detection using transfer learning of ResNet-50 convolutional neural network | Omer | Indonesian Journal of Electrical Engineering and Computer Science," *Image anomalies detection using transfer learning of ResNet-50 convolutional neural network | Omer | Indonesian Journal of Electrical Engineering and Computer Science*, Jul. 01, 2022. https://ijeecs.iaescore.com/index.php/IJEECS/article/view/28444

[15] "Image Splicing Detection based on Deep Convolutional Neural Network and Transfer Learning," *Image Splicing Detection based on Deep Convolutional Neural Network and Transfer Learning | IEEE Conference Publication | IEEE Xplore*. https://ieeexplore.ieee.org/document/10039789

[16] "A Review on Deep Learning Techniques for Image Forgery Detection," *A Review on Deep Learning Techniques For Image Forgery Detection | IEEE Conference Publication | IEEE Xplore*. https://ieeexplore.ieee.org/document/10037746

[17] P. Sharma, M. Kumar, and H. Sharma, "Comprehensive analyses of image forgery detection methods from traditional to deep learning approaches: an evaluation - Multimedia Tools and Applications," *SpringerLink*, Oct. 01, 2022. https://link.springer.com/article/10.1007/s11042-022-13808-w

[18] "Image Forgery Detection," *Image Forgery Detection | IEEE Conference Publication | IEEE Xplore*. https://ieeexplore.ieee.org/document/10151341

[19] K. M. Hosny, A. M. Mortda, N. A. Lashin, and M. M. Fouda, "A New Method to Detect Splicing Image Forgery Using Convolutional Neural Network," *MDPI*, Jan. 18, 2023. https://www.mdpi.com/2076-3417/13/3/1272

[20] "Error Level Analysis and Deep Learning for Detecting Image Forgeries," *Error Level Analysis and Deep Learning for Detecting Image Forgeries | IEEE Conference Publication | IEEE Xplore*. https://ieeexplore.ieee.org/document/10084286

[21] R. B. Burns and C. B. Dobson, "Transfer of learning (training)," *Transfer of learning (training) | SpringerLink*. https://link.springer.com/chapter/10.1007/978-94-011-6279-1_9

[22] M. L. Wallace *et al.*, "Use and misuse of random forest variable importance metrics in medicine: demonstrations through incident stroke prediction - BMC Medical Research Methodology," *BioMed Central*, Jun. 19, 2023. https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-023-01965-x

[23] L. Tian, W. Wu, and T. Yu, "Graph Random Forest: A Graph Embedded Algorithm for Identifying Highly Connected Important Features," *MDPI*, Jul. 20, 2023. https://www.mdpi.com/2218-273X/13/7/1153

[24] Z. Zhang, "Introduction to machine learning: k-nearest neighbors," *Introduction to machine learning: k-nearest neighbors - Zhang - Annals of Translational Medicine*, Apr. 20, 2016. https://atm.amegroups.org/article/view/10170/html

[25] A. Lindholm, N. Wahlström, F. Lindsten, and T. B. Schön, "Machine Learning | Higher Education from Cambridge University Press," *Higher Education from Cambridge University Press*, May 27, 2022. https://www.cambridge.org/highereducation/books/machine-learning/30AC30764CCF1ACBF86188BECD1B00AE

[26] A. Peng, K. Deng, J. Zhang, S. Luo, H. Zeng, and W. Yu, "Gradient-Based Adversarial Image Forensics," *Gradient-Based Adversarial Image Forensics | SpringerLink*, Nov. 20, 2020. https://link.springer.com/chapter/10.1007/978-3-030-63833-7_35

[27] "Robust Multi-Classifier for Camera Model Identification Based on Convolution Neural Network," *Robust Multi-Classifier for Camera Model Identification Based on Convolution Neural Network | IEEE Journals & Magazine | IEEE Xplore*. https://ieeexplore.ieee.org/document/8353219

[28] S. Gupta and M. Kumar, "Forensic document examination system using boosting and bagging methodologies - Soft Computing," *SpringerLink*, Aug. 14, 2019. https://link.springer.com/article/10.1007/s00500-019-04297-5

[29] "Convolutional Neural Network based Digital Image Forensics using Random Forest and SVM Classifier," *Convolutional Neural Network based Digital Image Forensics using Random Forest and SVM Classifier | IEEE Conference Publication | IEEE Xplore*. https://ieeexplore.ieee.org/document/10053434

[30] "Bayesian Tools for Reliable Multimedia Forensics," *Bayesian Tools for Reliable Multimedia Forensics |*

IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/9848571

Musaddik Habib Shirsho is pursuing his M.Sc. degree from Department of Information and Communication Technology in Bangladesh University of Professionals, Mirpur, Dhaka-1216, and Bangladesh. He has received his B.Sc. degree in Information and Communication Engineering from Bangladesh University of Professionals in 2021. He is working as System and Cyber Security Executive at Smart Technologies (BD) Ltd. (largest IT Distributor Company in Bangladesh) currently, his research focuses on Application security, Cyber Defense, Cloud security.

Md Masud Rana, a dedicated academic researcher, holds a PhD in Nuclear Reactor Physics from Jahangirnagar University. With over 25 years of work experience and 14 years of instructional expertise, he has made significant contributions to the field. His research interests span Information Security, Nuclear Physics, Reactor Physics, and Modern Physics. He has published extensively in reputable journals and has presented valuable insights on topics like reactor safety parameters and education planning in the Bangladesh Army. As the Chairman of Information & Communication Technology at Bangladesh University of Professionals, he continues to inspire and educate future generations of scholars. His commitment to excellence and passion for knowledge make him a respected figure in the academic community.

Jesmin Akhter has received PhD degree in 2019 in the field of 4G wireless networks. from Department of Computer Science and Engineering of Jahangirnagar University, Savar, Dhaka, Bangladesh and obtained M.Sc Engineering degree in Computer Science and Engineering from Jahangirnagar University, Savar, Dhaka, Bangladesh in 2012. She also received her B.Sc. Engineering degree in Computer Science and Engineering from Jahangirnagar University, Savar, Dhaka, Bangladesh in 2004. Since 2008, she is a faculty member having current Designation "Professor" at the Institute of Information Technology in Jahangirnagar University, Savar, Dhaka, Bangladesh. Currently her research focuses are on IoT, network traffic, complexity and algorithms and software engineering. Being a dynamic and versatile person who is capable of merging innovative ideas, technology, knowledge, and experience for positive contribution towards the system development in the rapidly changing scenario of Information Technology and become a good teacher in the field of software and telecommunication security.

Abu Sayed Md. Mostafizur Rahaman has received PhD degree in 2014 from Department of Computer Science and Engineering of Jahangirnagar University, Savar, Dhaka, Bangladesh and obtained M.Sc. degree from Stuttgart University at Stuttgart, Germany in Information Technology (INFOTECH) in the branch of Embedded System Engineering in 2009. He received his B.Sc. degree in Electronics and Computer Science, from Jahangirnagar University, Savar, Dhaka, Bangladesh in 2003. Since 2004, he is a faculty member having current Designation "Professor" in the Department of Computer Science and Engineering of Jahangirnagar University, Savar, Dhaka, Bangladesh. During his graduation, he worked at BOSCH (biggest automobile company in Germany) as Trainee engineer (Industrial internship) as part of his graduate degree in embedded systems. Currently, his research focuses on Digital Forensics, Cryptography, IoT, Web Security and S/W Systems.