Enhancing Digital Image Forensics in Cybersecurity Using Machine Learning

MSc Research Project
MSc Cyber Security

Hardik Sawant
Student ID: 21232105

School of Computing
National College of Ireland

Supervisor: Evgeniia Jayasekera.

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Hardik Sudhir Sawant

Student ID: 21232105

Programme: MSc Cyber Security **Year:** September 2023

Module: MSc Research Project

Supervisor: Evgeniia Jayasekera

Submission Due Date: 14/08/2023

Project Title: Enhancing Digital Image Forensics in Cybersecurity Using Machine Learning

Word Count: 7400 **Page Count** 17

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Hardik Sudhir Sawant

Date: 13/08/2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	Hardik Sudhir Sawant
Date:	13-08-2023
Penalty Applied (if applicable):	

Enhancing Digital Image Forensics in Cybersecurity Using Machine Learning

Hardik Sawant
21232105

Abstract

Digital image forensics is a crucial field that aims to authenticate and analyze digital images, considering the prevalence of advanced image editing software and techniques. This research paper delves into the potential of machine learning to enhance digital image forensics, specifically focusing on three machine learning models: Inception Net V3, Convolutional Neural Network (CNN), and Support Vector Machine (SVM).

The research methodology employed in this study is comprehensive, involving meticulous preprocessing of image data, careful selection and tuning of model hyperparameters, and rigorous evaluation of model performance. The models are trained and tested on a dataset of images, and their accuracy in classifying these images serves as the basis for their performance evaluation.

The results obtained from the experiments uncover notable disparities in the performance of the three models. Both the Inception Net V3 and SVM models exhibit superior performance, achieving an accuracy rate of 75% in classifying images. Conversely, the CNN model lags behind significantly, attaining a mere 25% accuracy. These findings highlight the potential of machine learning, particularly the Inception Net V3 and SVM models, in bolstering digital image forensics.

1 Introduction

A sub-field of digital forensics called "digital image forensics" is dedicated to identifying, analyzing, and deciphering altered or faked photographs. Digital picture forensics detects probable digital fraud by using scientific procedures and approaches, supplying important evidence for criminal investigations and civil lawsuits (Karie and Venter, 2015).

Several methods are used to identify image changes, including pixel-level irregularities, lighting or perspective issues, and artifacts left over from image editing software. These methods can be classified as either active or passive. When creating photographs, active methods incorporate data that can later be extracted for verification. Passive techniques use statistical analysis to find anomalies without knowing anything about the image beforehand (Nowroozi *et. al*, 2020).

Digital image forensics has incorporated machine learning, a kind of artificial intelligence, to provide cutting-edge counterfeit detection methods. Support Vector Machine (SVM) and k-nearest neighbors (k-NN) algorithms are used to improve security and identify forgeries in certain forged locations (Monika and Passi, 2021). By examining background noise, deep learning techniques—a more advanced form of machine learning—have been used to recognize audio recording equipment.

Law enforcement, judicial proceedings, journalism, and social media platforms all use digital picture forensics. It is essential in the fight against the dissemination of false information, especially when deepfakes or doctored photos are involved (Karie and Venter, 2015; Goni *et. al.*, 2020)). Additionally, it is employed to analyze drone data and expose illicit activity across numerous industries.

1.1 Motivation and Project Background

Due to the advent of complex image modification techniques and widely available editing software, there is a rising need for enhanced tools and approaches in digital image forensics, which is the project's motivation. The traditional methods of forgery and manipulation detection based on manual examination and specialized knowledge fall short. With its highly accurate and effective algorithms, machine learning has emerged as a potent remedy. However, these methods can be difficult to develop and require a lot of data. Additionally, the application of machine learning presents issues with data privacy. Therefore, more study is required to create new machine learning methods for digital image forensics that are accurate, effective, and considerate of data privacy.

1.2 Research Question

"How can we develop robust and efficient machine learning techniques for digital image forensics that can effectively detect sophisticated image manipulations?"

1.3 Research Objectives

1. To understand and analyze the principles and mechanisms of InceptionNet, Convolutional Neural Networks (CNN), and Support Vector Machines (SVM) in the context of digital image forensics.
2. To develop and implement models using InceptionNet, CNN, and SVM for digital image forensics.
3. To compare the performance of InceptionNet, CNN, and SVM in detecting image manipulations.
4. To identify the strengths and weaknesses of InceptionNet, CNN, and SVM in the context of digital image forensics.

1.4 Research Outline

The introduction to the thesis describes the importance of digital image forensics and the application of machine learning methods like InceptionNet, Convolutional Neural Networks (CNN), and Support Vector Machines (SVM). The study's objective and research question are also presented in the introduction.

The Related Work section that follows covers the body of work on digital picture forensics and machine learning. This part lays the theoretical groundwork for the investigation and identifies knowledge gaps.

The methods used to accomplish the research objectives are described in the methodology part of the paper. This includes a thorough discussion of the creation and use of the InceptionNet, CNN, and SVM models, as well as the experimental framework for digital image forensics using the mentioned models.

Subsequently, the Implementation section gives a step-by-step breakdown of the creation and application of the models. This includes the selection and processing of the data, the coding process, and model configuration.

The findings of the study are presented in the Results and Discussion section, along with performance metrics for the models and a comparison of the models. The findings are reviewed in relation to the goals of the study and the body of prior research.

The Conclusion and Future Work section, which summarizes the main conclusions, explores their ramifications and makes suggestions for further research, brings the thesis to a close. This could include suggestions for improving the models or looking into other machine-learning methods for the digital image forensics.

1.5 Summary

The introduction chapter has successfully laid the groundwork for the research, offering a comprehensive overview of digital image forensics and the role played by machine learning in its advancement. It has also clearly stated the research question and objectives that will guide the subsequent chapters.

2 Related Work

This section of the thesis presents the reader with the literature review conducted related to the study. This section is divided into two parts. The first part of the section deals with a general literature review of the field whereas the second part covers different studies that incorporated machine learning modalities in digital forensics.

2.1 General Literature Review

Digital image forensics has made considerable advancements and has grown to be a popular field of study for many academics. The need for authenticating photos is on the rise, which is mostly to blame (Redi, Taktak, & Dugelay, 2010). This field's major objective is to uncover details about an image's past, with an emphasis on locating the camera used to take the photo and spotting any indications of photo manipulation.

The examination of noise, which is present in all forms of imaging, is a key component of digital image forensics. Digital images can contain noise in a variety of ways, and noise can alter significantly as images are being acquired. These modifications have a big impact on forensic investigation and counter-forensic methods (Julliand, Nozick, & Talbot, 2015).

In digital image forensics, it is crucial to precisely estimate the quantization steps for JPEG images because they are a common digital image format. Quantization offers important information about an image's past, including whether it has previously undergone JPEG compression or has been preserved in a lossless manner. In images that have undergone double-JPEG compression, precise quantization step estimate helps in the detection of JPEG compression artifacts and the identification of secondary quantization tables (Thai, Cogranne, Retraint, & Doan, 2016).

Analysis of photo response non-uniformity noise (PRNU), which serves as an individual fingerprint for imaging sensors, is a critical component of digital picture forensics. Digital forensic tasks like source device identification, content integrity verification, and authentication all find uses for PRNU. While just one-color component is recorded for each pixel when an image is captured using a color filter array, this produces color interpolation noise that may compromise the precision of device identification and picture content integrity verification (Li & Li, 2012).

2.2 Related Literature Review

The paper by Castillo Camacho *et al*., (2021), focuses on the concepts of consumer technologies that are acquired in editing artificial intelligence techniques with the advancement of digital image editing creating a fake scenario of images is the deliberate manipulation of innocuous digital images. The improved quality of malicious intent that semantically spreads false narratives in a group or individual events specifically establishes the implications in the global economy, financial health and national security. The digital forensic images specifically deliberate the commercial market to determine authenticity. Image forensics emerges as the underwent process of classifying the sources with the acquisition of the reverse operating process and includes the authentic and manipulative verification of manipulated images.

According to Yang *et al*., (2020), image forensics is considered the most effective way to blind a digital authenticity and integration in images. The data-driven approach inspires the techniques of computing the visions of problems. The technological algorithm that deals with image forensics has generated adversarial networks regarding cyber security. CNN and "recurrent neural networks (RNN)" have shown effective digital recognition in image-related tasks that subsequently adopted numerous image forensics methods and categorises a structural reformation in optimising the parameters in improving the depth performances. Machine learning is playing an increasingly important role in online activities and digital evolution. The significance of security concerns is the documentation of machine-connected devices in the innovations of incorporating cyber threads. The conflicts highlight the benefits of machine learning in the recommendations of direction in cybersecurity management.

According to Misra *et al*., (2022), cyber attacks in technological revolutions have advancements in staple and sensational aspects of information and digital devices. Computer integration techniques have the emergence of computing the disciplines with the effectiveness on the perception of mechanism behaviour in the deep learning process. The multi-layered neural networks accomplish the accurate interfaces in changing the intelligent decisions in making cyber forensics science. A scientific process of excavating the investigation n hearing the diversification of techniques in the direction approaches that explore the dynamics of diverging computing techniques in overall realise effectiveness. Machine learning approaches aim to generic the cognitive techniques in the integration of cyber forensics in the machine investigation learning approaches for diverging the learning techniques in cyber forensics. Computer forensics has some pre-defined extracting purposes in extracting the objectivity of evidence in data recovery and reveals important facts regarding the complexity of CF techniques and processes. Machine learning has come out with the fast processing of huge data networking processes in the analysis of forensic investigation processes. The "Deep neural network (DNN)" has the ability to visualise the patterns in CF with the relevant big data requires by forensic investigations. Cyber security infrastructures have highly vulnerable interventions in sufficient monitoring and sophisticated cyber defence systems.

According to the study by Nowroozi *et al.,* (2021), Image forensics has important part in criminal investigation for example circulation of fake images that promotes racial hatred or false information about a specific political campaign or racial ethnicity and civil litigation. In image forensics, the utilization of the machine learning approach is increasing. Though there are several vulnerabilities that are connected with the approaches which are based on machine learning, for example, the detection of unsolicited images and their real-world consequences, such as forbidden evidence or wrongful judgement. In the focus on image forensics, there are several techniques that can be utilized in enhancing the agility of binary manipulation detectors based on machine learning in diverse scenarios.

In accordance with Dushyant *et al.* (2022), both deep learning (DL) and machine learning (ML) provided immense internet and attract interest regarding the unparalleled community in recent years. With the emerging confluence of digital life or activities online, people's way of learning is evolving and also leading towards security issues significantly. The task of sensitive information, documents protection, machine-connected networks and devices from the unfavourable threat of cybercrime and overall protection of cyber security is crucial. As a solution, innovative machine learning and deep learning are included in cybercrime threats.

It is stated in the study by Karie *et al.* (2019), the world is experiencing cyber-attacks more than before in various aspects of people's daily life and the situation demands struggling with cybercrimes and combating them is considered a struggle for individuals as well as for business organizations. This struggle is now aggravated by the improved cybercrime techniques as the criminals have improved themselves to go a step ahead and complicate the conventional cybercrime techniques. Some techniques are miniatured and naturally exceptional and sometimes disguised in the face of authentic commands. After a security mishap occurs, combating the danger the cyber security investigators have to stay conscious and mitigate them. They gather PDE (Potential Digital Evidence) in order to support litigation and it can help the investigators arrive at relevant decisions. The diverse source and format of BigData make data processing difficult for cyber forensics. The huge pile of complex data are often hard to analyse for forensic purposes as they lack the time and budget. Deep Learning is a fragment of Artificial Intelligence that has well-defined use cases in the cyber forensic domain, though there is some argument that this is not the unexcelled solution but it helps enhance the combating methods of cybercrimes. Deep learning is the potential to change the domain of CF dramatically in several ways along with providing strategies for forensic investigators. The strategies extend to challenging the considered evidence which is justifiable by law from decreasing bias in forensic investigations.

In the study conducted by Al Balushi (2023), Digital Forensics has become a crucial aspect in achieving quality evidence while mitigating the increasing threats of cybercrime. Often the investigators witness obstruction while data collection and event reconstruction analysis. Various algorithms of machine learning ensure the effective and efficient performance of investigations. Machine learning focuses on improving computer models and algorithms that perform specific relevant tasks without programming such as the potential of dataset training aiding in forensic investigation Every machine learning algorithm holds a specific digital forensic area and overcoming complexity, the volume of data along with time-lining, consistency, correlation and so on.

According to Ghillani (2022), Deep Learning derived from ANN which stands for the artificial neural network is one of the most essential technologies for the systems or policies of intelligent cybersecurity. There are some advantages and disadvantages to utilizing enhancing organizational flexibility in cyber risk analytics and well comprehending cyber risk. There are several approaches that can be utilized in tackling intelligently the various issues regarding cybersecurity issues. The ultimate goal of the backpropagation algorithm is to maximize correctly the weight of networks in order to input translation to the output that is intended. The approaches are self-organizing maps, deep transfer learning, restricted Boltzmann machine, auto-encoder, generative adversarial network, deep belief network, and deep reinforcement learning along with their group and hybrid approaches.

It is stated in the study by Kamoun (2020), that Artificial Intelligence and Machine Learning Software have achieved significant attention in defensive cybersecurity and there is also a gap in research that is noticeable on the usage of AI and MLS. The System-Fault-Risk (SFR) framework inspires categorizing the cyberattacks that are powered by AI/MLS by the actions of them into several categories.

A study conducted by Sedik *et al.* (2022), it is stated that cybersecurity has attracted much attention in video transmission technologies, cloud services and applications of IOT and recent studies have also stated that fact due to the evergrowing cyber-risk on both institutions and individuals. The spoofing attack, which is a kind of cyber risk, is, unfortunately, increasing along with the cyber criminals in data transfer applications without being caught especially in advanced areas such as smart cities. Different applications which are based on online video communication, including video conferences and online testing are implicated in smart cities. The videos show the diverse variety of an individual, which makes face recognition a crucial idea in the implementation of cyber security. The attacks regarding face spoofing are specifically based on the face replication of an individual by video replaying or photo printing. Hence, detecting video forgery and spoof attacks is the latest topic in the research of cybersecurity. The approach for deep learning for forensically detecting video faces using two methodologies with the facial spoofing attack. The first methodology depends on a convolutional neural network (CNN) to draw out features from the frames of the input video. The second methodology depends on convolution long-short term memory (ConvLSTM). This model is made of two pooling layers, a convolutional LSTM layer and two convolutional layers. A fully-connected layer is included in each methodology in order to connect among the feature map caused by feature extraction possess and classification layer.

Also, Kandali (2017) discusses the widespread use of digital images in media and on the internet, which brings attention to the problem of image alteration and its effects on industries including politics, media communication, and science. It underlines the necessity of guaranteeing the integrity and authenticity of significant images. He suggests that modification and manipulation of images has become easier with the help of advance tools, which leads to misuse for spreading incorrect information, copyright infringement and other malpractices. To overcome this issues, the paper suggested using AI methods like Neural Networks, Support Vector Machines and Machine Learning to perform forensic analysis and decision making based on experience.

However, this review reveals several gaps and challenges in the current literature:

1. **Lack of Unified Standards:** Lack of standardized methodologies and benchmarks in image forensics hinders reproducibility and comparison of results across studies, despite advancements in detection techniques.
2. **Bias and Fairness Concerns:** Concerns about bias and impartiality are raised by the incorporation of machine learning into the detection process. Numerous studies have observed that these biases can affect the precision of forensic analyses and may result in incorrect findings or the failure to notice some manipulations.
3. **Real-time Detection:** The majority of currently used methods concentrate mostly on post hoc analysis. To successfully limit the effects of cyber risks, real-time detection and response techniques are required due to the rapid speed of online activity and the proliferation of digital content.
4. **Adversarial Attacks:** According to the research, image forensics systems are vulnerable to adversarial attacks. A crucial area for future research is the investigation of strong detection systems that can withstand these attacks.
5. **Integration with Legal Frameworks:** While some studies skim over the legal ramifications of picture manipulation and cyber threats, a more thorough examination of the practical applications of forensic evidence in court procedures is required.

Reason for a CNN and SVM-Based ML Model:
1. **Enhanced Accuracy:** By successfully identifying spatial information in images, CNNs have been shown to excel in image identification tasks. This hybrid model can provide improved accuracy when it is paired with SVMs, which are renowned for their capacity to categorize difficult data.
2. **Robustness Against Adversarial approaches:** Because CNNs can learn hierarchical features from data, they are naturally resistant to some adversarial approaches. The model's robustness can be increased by include SVMs, which concentrate on margin optimization.
3. **Real-time Detection:** Real-time image processing capabilities of CNNs make them perfect for the quick-paced nature of online activities. When combined with SVMs, this feature can result in the creation of a real-time manipulation picture detecting system.
4. **Comprehensive Analysis and Bias Mitigation:** A CNN-SVM-based model can include techniques to improve fairness in forensic analysis and reduce biases. Additionally, the model can provide a more thorough analysis of pictures while minimizing false positives and false negatives by combining the advantages of both CNNs and SVMs.

2.3 Summary

The key significance that machine learning and deep learning approaches play in enhancing image forensics and supporting cybersecurity initiatives is highlighted by this literature review's conclusion. While the current research offers insightful information, it also highlights gaps and difficulties that call for more research. The standardization of methodology, minimizing biases and adversarial attacks, creating real-time detection systems, and investigating the legal ramifications of forensic results should all be the focus of future research. Such initiatives will strengthen and broaden the method taken to protect the veracity and integrity of digital content, which will ultimately be advantageous to many facets of society.

This assessment not only provides a thorough overview of the state of image forensics and cybersecurity at the present time, but also indicates important research directions that will influence the field's future advancements. A new research approach using a CNN and SVM-based machine learning model holds the promise to address these gaps. This approach can significantly contribute to the field of image forensics by offering enhanced accuracy, robustness against attacks, real-time detection capabilities, and comprehensive analysis. Through such a research endeavor, we can further fortify the security of digital content and contribute to the ongoing efforts in cybersecurity.

3 Research Methodology

This section of the thesis discusses the methodology adopted for the study. The modules present in the methodology are discussed in great depth below.

Figure 3.1 below shows the methodology flow for the identification of forged images using machine learning models.

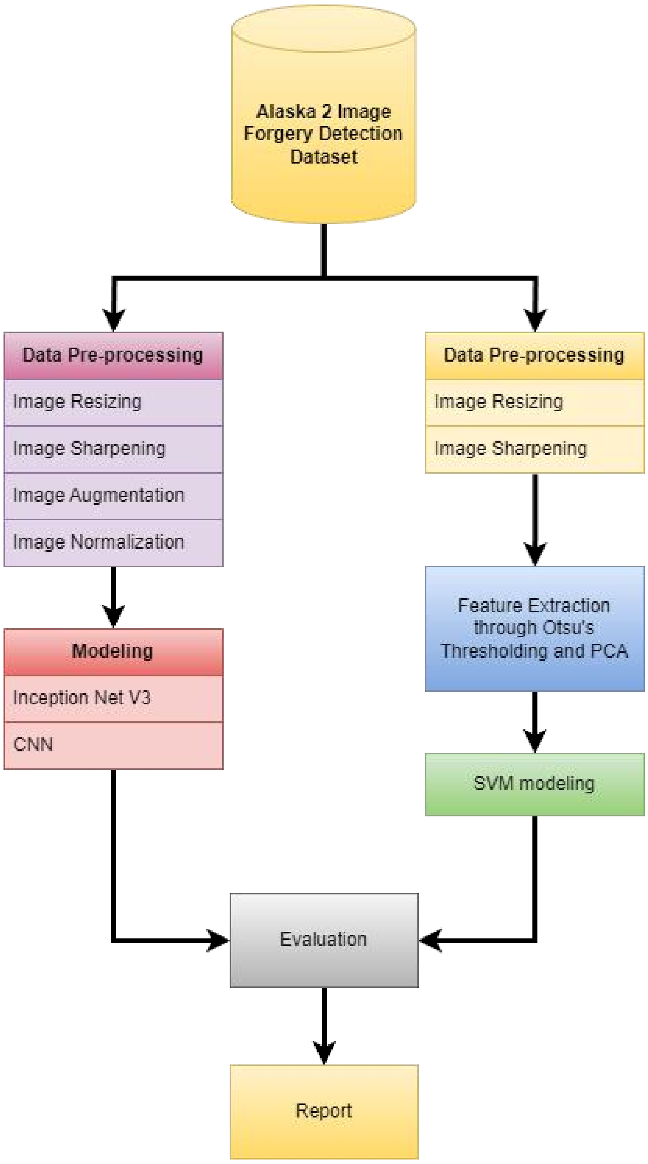

Figure 3.1: Methodology for Enhancing Digital Forensics Using Machine Learning

The methodology is divided into modules which are given below:
1. Data Collection.
2. Data Pre-processing.
    2.1 Image Resizing.
    2.2 Image Sharpening.
    2.3 Image Augmentation and Normalization.
3. Feature Extraction.
    3.1 Image Blurring.
    3.2 Change color space.
    3.3 Thresholding.
    3.4 Morphological Operations.
    3.5 Otsu's Thresholding.
    3.6 Principal Component Analysis.
4. Modeling.
    4.1 Inception Net.
    4.2 Convolutional Neural Network.
    4.3 Support Vector Machine.

These modules presented in the methodology are discussed in depth in the design specification section.

4 Design Specification

The techniques that are mentioned in the Methodology section are elaborated below:

4.1 Data Collection

This study makes use of the Alaska 2 dataset for detecting steganographic images (Cogranne, Giboulot and Bas, 2019). The Alaska 2 Image Steganalysis dataset is a comprehensive collection of images specifically designed for steganalysis purposes. Comprising a multitude of high-resolution images (512x512 pixels) in the JPEG format, the dataset is categorized into two primary groups: cover images and stego images. Cover images represent unmodified originals, while stego images have been manipulated using diverse steganography techniques to hide information.

The steganography methods employed in generating stego images within the Alaska 2 dataset encompass cutting-edge algorithms such as JUNIWARD, JMiPOD, and UERD. These techniques are recognized for their exceptional embedding efficiency and minimal detectability, making them ideal choices for steganographic operations.

JUNIWARD, JMiPOD, and UERD are advanced steganography algorithms used for concealing information within digital images (Liu *et. al.,* 2019). Steganography involves hiding a file, message, image, or video within another file, message, image, or video.

JUNIWARD is a spatial domain steganographic technique, where "Just New Ward" is its acronym. This algorithm minimizes a weighted sum of embedding changes in the spatial domain. It is well-known for its high embedding efficiency and low detectability, making it a popular choice for steganography.

JMiPOD, short for "Jpeg Minimizing the Power of DCT coefficients," operates in the JPEG domain. It aims to minimize the statistical detectability of the stego image by reducing the power of the discrete cosine transform (DCT) coefficients.

UERD, or "Uniform Embedding, Revisited Distortion," is another JPEG domain steganography method. It focuses on achieving uniform embedding while improving the security of the stego image by revisiting the distortion function.

The images in the dataset are divided into 5 folders viz. ‘Cover’, ‘JUNIWARD’, ‘JMiPOD’, ‘UERD’ and ‘Test’.

4.2 Pre-processing

Pre-processing is an essential stage in the machine learning pipeline, particularly when working with image data. Its purpose is to convert raw data into a more comprehensible format, thereby enhancing the performance of machine learning algorithms.

Following are the pre-processing steps undertaken in the study.

4.2.1 Image Resizing

In this pre-processing step, the images in the dataset are resized from 512x512 to 100x100. Reducing the image size helps to process images faster for machine learning models without losing the information stored in them.

4.2.2 Image Sharpening

Image sharpening is a step in image pre-processing that is used to enhance the details in an image. The image sharpening process is done through the convolution of the image under consideration by a sharpening filter. This study makes use of a sharpening filter (h) given below.

$$h(m,n) = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

The filtering operation involves replacing each pixel with a weighted sum of its immediate neighbors, as defined by a kernel. In this particular case, the center pixel is multiplied by 5, while its four adjacent neighbors (up, down, left, and right) are multiplied by -1. The remaining four diagonal neighbors are multiplied by 0, indicating that they do not contribute to the new pixel value.

This filtering technique effectively enhances the contrast between a pixel and its neighboring pixels. As a result, the edges in the image become more pronounced, ultimately leading to a sharper image.

4.2.3 Image Augmentation and Normalization

Image augmentation is a technique used in machine learning to expand the range and volume of the training dataset without the need for additional data collection. It involves generating altered versions of the images in the dataset. This approach enhances the model's capacity to generalize and mitigates overfitting, which occurs when the model performs well on the training data but struggles with unseen data. By incorporating image augmentation into the training process, the model's performance and robustness can be enhanced.

This study makes use of rescale technique to generate normalized images. The rescaling is done on the pixel values by dividing the pixel intensities by 255 making the normalized images to have the intensity range from 0 to 1.

4.3 Feature Extraction

The feature extraction step in the methodology is done for identifying steganographic images using the Support Vector Machine model. The feature extraction step of the methodology makes use of Otsu's thresholding technique to extract features. The steps involved in Otsu's thresholding are given below.

4.3.1 Image Blurring

This step is done to blur the image through the application of Gaussian Blur kernel on the image under consideration.

4.3.2 Changing color space

This step involves changing the color space of the image from BGR to HSV. This is done because the Otsu's thresholding technique requires thresholding the saturation channel of the image under consideration.

4.3.3 Thresholding

A threshold is applied to the saturation channel.

4.3.4 Morphological Operations

Morphological operations such as opening and closing are performed on the images. The closing operation closes small holes in the thresholded image. The opening operation removes small white regions in the image.

4.3.5 Otsu's Thresholding

After the morphological operations, the image is then converted to grayscale over which the final thresholding is done. The final thresholding determines the optimal threshold by maximizing the variance between the two classes of pixels.

4.3.6 Principal Component Analysis

Once the Otsu's thresholded image is obtained, it is converted to an array of 10000 values. This is considered to be its feature space. The principal component analysis is then applied to this array. The principal component is a new feature or variable derived from a linear combination of the original features or variables. It captures the maximum variance present in the dataset, with subsequent principal components accounting for decreasing variances. Each principal component is orthogonal, or uncorrelated, to the others, offering distinct information about the data.

4.4 Modeling

Once the pre-processed and augmented data is obtained, it is then subjected to modeling. This study implements and compares three machine learning models viz. Inception Net V3, Convolutional Neural Network, and Support Vector Machines.

4.4.1 Inception Net

Inception Net V3 is a convolutional neural network (CNN) architecture developed by Google. It is the third version of the Inception network, preceded by Inception V1 (GoogLeNet) and Inception V2.

Inception V3 is renowned for its complexity and exceptional performance, frequently achieving top-notch results in image classification tasks (Bhatia *et. al.,* 2019). It has been pre-trained on the ImageNet dataset, a vast collection of over 14 million images spanning 1000 classes. Inception V3 finds extensive application in various computer vision tasks, such as image classification, object detection, and image segmentation.

4.4.2 Convolutional Neural Network

Convolutional Neural Networks (CNNs) are a powerful class of deep learning models that excel at analyzing visual data. They have revolutionized tasks such as image classification, object detection, and facial recognition. Inspired by the structure of the animal visual cortex, CNNs automatically learn hierarchical features from input data. Each neuron in the network responds to stimuli within a specific region called the receptive field (Tripathi, 2021).

CNNs have achieved remarkable success in various domains, from enabling self-driving cars to detecting diseases in healthcare. Their widespread adoption has contributed to the popularity of deep learning in the field of artificial intelligence.

4.4.3 Support Vector Machine

Support Vector Machines (SVMs) are a popular supervised machine learning model used for classification and regression tasks. They excel in handling complex datasets that are small to medium in size (Cervantes *et. al.*, 2020).

The core concept behind SVMs is to find the optimal hyperplane that can separate different classes in the feature space. In two dimensions, this hyperplane appears as a line that divides the plane into two regions, with each class on either side. In higher dimensions, the hyperplane becomes a multi-dimensional surface (Cervantes *et. al.*, 2020; Jiang *et. al.*, 2020).

SVMs offer several advantages, including their ability to handle large feature spaces, capture non-linear relationships between features, and accommodate diverse types of data sources. However, they can be memory-intensive, challenging to interpret, and require careful selection of the appropriate kernel for optimal performance.

4.5 Evaluation

This study makes use of the accuracy metric to evaluate the models implemented. Accuracy provides a straightforward measure of a model's performance. It quantifies the percentage of correct predictions made by the model out of all predictions. When there is a balance between the classes, meaning there is an approximately equal number of instances belonging to each class, it serves as a reliable indicator of how well the model is able to classify instances correctly.

4.6 Summary

This chapter has provided a detailed explanation of the machine learning models utilized in this study, specifically Inception Net V3, Convolutional Neural Network (CNN), and Support Vector Machine (SVM). It has also outlined the steps taken in data preprocessing and the metrics employed to assess the models. Overall, the methodology has been carefully designed to address the research question and achieve the research objectives.

5 Implementation

This section of the study describes the implementation of the study. The different models that are applied to the dataset are considered experiments. In-depth implementation of the models along with the hyperparameters that are chosen for them are discussed in this section. To implement the study, Python programming language has been used and Jupyter is chosen to be the preferable IDE.

5.1 Inception Net

The Inception Net V3 model implemented in the study is implemented using the Tensorflow library available for Python.

Pre-trained weights from the ImageNet dataset are first used to initialize the InceptionV3 model. When the include_top parameter is set to False, the classification-related fully connected output layer is not included. Because of this, the model can now be customized for particular tasks. The output of this base model is then passed through several layers to create the final model.

The output of the base model is fed into a Global Average Pooling 2D (GAP) layer, which reduces the spatial dimensions of the output by averaging the values in each spatial location. This conversion from a 3D tensor to a 2D tensor helps reduce parameters, computation, and overfitting in the model. The output then passes through Dense layers, also known as fully connected layers, for classification. These layers use the 'tanh' activation function, which centers the data between -1 and 1. Dropout layers are applied to prevent overfitting by randomly setting a fraction of input units to 0 during training. Finally, a Dense layer with a 'sigmoid' activation function outputs the predicted probability of the image belonging to a specific class (between 0 and 1).

5.1.1 Hyperparameters

Table 4.1 below shows the hyperparameters chosen for the Inception Net model.

Hyperparameter	**Value**
Pre-trained weights	'imagenet'
Number of units in first Dense layer	64
Number of units in second Dense layer	32
Number of units in third Dense layer	16
Number of units in fourth Dense layer	8
Activation function in Dense layers	'tanh'
Activation function in output layer	'sigmoid'

Table 4.1: Hyperparameters for Inception Net

5.1.2 Training

The Inception Net model used in this study is trained for 10 epochs with an Early Stopping callback in place. The early stopping criterion is chosen to monitor the training loss and is activated when the minimum value of the loss is reached.

5.2 Convolutional Neural Network (CNN)

The CNN model in the study is implemented using the Tensorflow library available for Python.

The model begins with the initialization of an empty model using the Sequential() function. This allows for the sequential addition and configuration of layers.

The first layer added is a Conv2D layer, which performs 2D convolution on the input. It has 32 filters, each with a kernel size of 3x3. The activation function used is 'tanh', which produces output values between -1 and 1. The 'Same' padding ensures that the spatial dimensions of the output feature maps match the input feature maps. The input_shape parameter specifies the shape of the input data.

After the convolutional layer, a GlobalAveragePooling2D layer is added. This layer reduces the spatial dimensions of the input by computing the average of each feature map, resulting in a single scalar value per feature map.

To prevent overfitting, a Dropout layer with a rate of 0.25 is added. This layer randomly sets a fraction of the input units to 0 during training.

Next, a densely connected layer with 128 neurons is added. The activation function used is 'tanh'. This layer is fully connected, meaning each neuron is connected to every neuron in the previous layer.

To further prevent overfitting, another Dropout layer with a rate of 0.5 is added.

Finally, a Dense layer with a single neuron is added. The activation function used is 'sigmoid', which produces output values between 0 and 1. This type of activation is commonly used for binary classification problems.

1.1.1 Hyperparameters

Table 4.2 below shows the hyperparameters chosen for the CNN model.

Hyperparameter	**Value**
Number of filters in Conv2D layer	32
Kernel size in Conv2D layer	3x3
Activation function in Conv2D layer	'tanh'
Padding in Conv2D layer	'Same'
Dropout rate after GlobalAveragePooling2D layer	0.25
Number of neurons in first Dense layer	128
Activation function in first Dense layer	'tanh'
Dropout rate after first Dense layer	0.5

Table 4.2: Hyperparameters for CNN

5.2.1 Training

The CNN model used in the study is trained for 10 epochs with an Early Stopping Callback. The criterion for Early Stopping is chosen to be the training loss. The stoppage occurs when the lowest value of the loss is achieved.

5.3 Support Vector Machine

The SVM model in the study is implemented using the Scikit Learn library. The Inception Net and CNN models could directly take the images that need to be classified. However, for the SVM model, this is not possible or feasible. To use the SVM model, the feature extraction step has been included in the methodology. In feature extraction several sub steps are implemented. These are discussed in section 3.3.

5.3.1 Hyperparameters

Table 4.3 below depicts the hyperparameters for the SVM model.

Hyperparameter	**Value**
Kernel function	'linear'
Penalty parameter C of the error term	0.5
Kernel coefficient for 'rbf', 'poly' and 'sigmoid' (gamma)	10

Table 4.3: Hyperparameters for SVM

5.4 Summary

The implementation chapter offers a step-by-step account of how the machine learning models were employed in the context of digital image forensics. It has also discussed the hyperparameters utilized in the models and the techniques employed for image preprocessing. The implementation process has been conducted meticulously to ensure the accuracy and reliability of the results.

6 Evaluation

In this section, the outcomes of applying three distinct machine learning models to the field of digital image forensics are discussed. The models under consideration are Inception Net V3, Convolutional Neural Network (CNN), and Support Vector Machine (SVM). The performance of each model is evaluated by analyzing its accuracy in image classification.

Detailed discussions on the obtained results are provided below.

Table 5.1 below depicts the accuracies obtained for each of the models.

Model	**Accuracy**
Inception Net V3	75%
Convolutional Neural Network	25%
Support Vector Machine	75%

Table 5.1: Model accuracies

From these results, it is evident that both the Inception Net V3 and the Support Vector Machine models achieved the highest accuracy of 75%. On the other hand, the Convolutional Neural Network model had a significantly lower accuracy of 25%.

Figure 5.1 below shows the graphical comparison of the model performance.

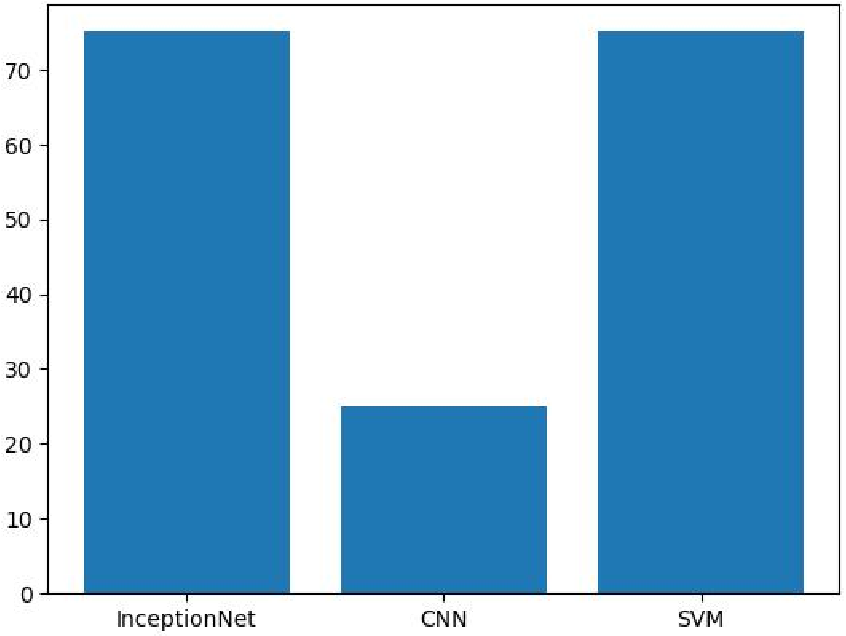

Figure 5.1: Model Performance Comparison

The results of the image classification task showcased the superior performance of the Inception Net V3 and SVM models, boasting an impressive accuracy of 75%. This indicates that these models were successful in effectively learning and generalizing from the provided training data, accurately classifying the majority of images within the test set.

6.1 Discussion

The Inception Net V3 model's exceptional performance can be credited to its intricate architecture, featuring numerous layers and an extensive number of neurons. This complexity enables the model to acquire a diverse range of features and make precise predictions. On the other hand, the SVM model likely excelled due to its ability to identify the optimal hyperplane that separates distinct classes within the feature space. This characteristic proves particularly advantageous in this type of classification task.

Conversely, the Convolutional Neural Network (CNN) model achieved a significantly lower accuracy of 25%. This outcome suggests that the model struggled to effectively learn from the provided training data, potentially overfitting and failing to generalize to the test data. It is also plausible that the CNN model's architecture was unsuitable for this specific task, or that its hyperparameters were inadequately tuned.

These outcomes emphasize the significance of selecting the appropriate model and fine-tuning its hyperparameters for a given machine learning task. While CNNs are generally effective in image classification tasks, the Inception Net V3 and SVM models proved to be more suitable and successful in this particular scenario.

This chapter has presented the research findings and provided a thoughtful interpretation of these results. It has demonstrated that the Inception Net V3 and SVM models outperformed the CNN model in terms of accuracy. Consequently, these findings have offered valuable insights into the effectiveness of different machine learning models in the field of digital image forensics.

7 Conclusion and Future Work

In this study, the aim was to improve digital image forensics through the use of machine learning models, namely Inception Net V3, Convolutional Neural Network (CNN), and Support Vector Machine (SVM). The results clearly demonstrated that the Inception Net V3 and SVM models outperformed the CNN model, achieving an impressive accuracy of 75% in accurately classifying images.

The superior performance of the Inception Net V3 and SVM models highlights their ability to effectively learn and generalize from the training data, accurately classifying the majority of images in the test set. Conversely, the lower accuracy of the CNN model suggests the need for potential improvements in its architecture or hyperparameters.

These findings emphasize the significant potential of machine learning in enhancing digital image forensics. By selecting the appropriate model and fine-tuning its hyperparameters, it becomes possible to achieve high accuracy in image classification, greatly benefiting digital forensics investigations.

Future Work:

The study's results present promising implications: however, further research avenues should be explored. One such avenue is to investigate the factors contributing to the CNN model's inferior performance and identify methods to enhance its effectiveness. This could involve experimenting with diverse CNN architectures, hyperparameter tuning, or alternative image pre-processing techniques.

Moreover, evaluating other machine learning models could potentially yield higher accuracy rates. Complex models like ResNet or EfficientNet, as well as ensemble methods such as Random Forests or Gradient Boosting, warrant examination to ascertain improved outcomes.

In addition, future research endeavors could involve applying these models to different types of data or tasks within digital forensics. Examples include detecting manipulated images, classifying various image forgeries, or expanding their application to video and audio forensics. These avenues hold potential for advancing the field of digital image forensics.

Important Links:-

Presentation Video Link:
https://studentncirl-my.sharepoint.com/:v:/g/personal/x21232105_student_ncirl_ie/ETtmvybVNDVGuVSTk-N6BL4Bm5V2FPZeu9AQ09rXc2f1Sw?e=aFsuJ4

Dataset Link:
https://www.kaggle.com/c/alaska2-image-steganalysis

References

Al Balushi, Y., Shaker, H. and Kumar, B., 2023, January. The use of machine learning in digital forensics. In 1st International Conference on Innovation in Information Technology and Business (ICIITB 2022) (pp. 96-113). Atlantis Press.

Baig, Z., Khan, M.A., Mohammad, N. and Brahim, G.B., 2022. Drone forensics and machine learning: Sustaining the investigation process. *Sustainability*, *14*(8), p.4861.

Bhatia, Y., Bajpayee, A., Raghuvanshi, D. and Mittal, H., 2019, August. Image captioning using Google's inception-resnet-v2 and recurrent neural network. In *2019 Twelfth International Conference on Contemporary Computing (IC3)* (pp. 1-6). IEEE.

Castillo Camacho, I. and Wang, K., 2021. A comprehensive review of deep-learning-based methods for image forensics. Journal of imaging, 7(4), p.69.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L. and Lopez, A., 2020. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, pp.189-215.

Cogranne, R., Giboulot, Q. and Bas, P., 2019, July. The ALASKA steganalysis challenge: A first step towards steganalysis. In *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security* (pp. 125-137).

Dushyant, K., Muskan, G., Annu, Gupta, A. and Pramanik, S., 2022. Utilizing Machine Learning and Deep Learning in Cybesecurity: An Innovative Approach. Cyber Security and Digital Forensics, pp.271-293.

Ferreira, S., Antunes, M. and Correia, M.E., 2021. A dataset of photos and videos for digital forensics analysis using machine learning processing. *Data*, *6*(8), p.87.

Ghillani, D., 2022. Deep Learning and Artificial Intelligence Framework to Improve the Cyber Security. Authorea Preprints.

Goni, I., Gumpy, J.M., Maigari, T.U., Muhammad, M. and Saidu, A., 2020. Cybersecurity and cyber forensics: Machine learning approach. *Machine Learning Research*, *5*(4), pp.46-50.

Hou, J., Li, Y., Yu, J. and Shi, W., 2019. A survey on digital forensics in Internet of Things. *IEEE Internet of Things Journal*, *7*(1), pp.1-15.

Jiang, F., Lu, Y., Chen, Y., Cai, D. and Li, G., 2020. Image recognition of four rice leaf diseases based on deep learning and support vector machine. *Computers and Electronics in Agriculture*, *179*, p.105824.

Julliand, T., Nozick, V., & Talbot, H. (2015). Image Noise and Digital Image Forensics. In Digital Forensics and Watermarking (pp. 3-16). Springer, Cham.

Kamoun, F., Iqbal, F., Esseghir, M.A. and Baker, T., 2020, October. AI and machine learning: A mixed blessing for cybersecurity. In 2020 International Symposium on Networks, Computers and Communications (ISNCC) (pp. 1-7). IEEE.

Kandali, A. and Kaushik, M., 2017. Convolutional Neural Network based Digital Image Forensics using Random Forest and SVM Classifier. Procedia Computer Science, 112, pp.724-731.

Karie, N.M., Kebande, V.R. and Venter, H.S., 2019. Diverging deep learning cognitive computing techniques into cyber forensics. Forensic Science International: Synergy, 1, pp.61-67.

Karie, N.M. and Venter, H.S., 2015. Taxonomy of challenges for digital forensics. *Journal of forensic sciences*, *60*(4), pp.885-893.

Kumar, A. and Tiwari, A., 2019. A comparative study of otsu thresholding and k-means algorithm of image segmentation. *Int. J. Eng. Technol. Res*, *9*, pp.2454-4698.

Khan, H., Hanif, S. and Muhammad, B., 2021. A survey of machine learning applications in digital forensics. *Trends in Computer Science and Information Technology*, *6*(1), pp.020-024.

Latif, A., Rasheed, A., Sajid, U., Ahmed, J., Ali, N., Ratyal, N.I., Zafar, B., Dar, S.H., Sajid, M. and Khalil, T., 2019. Content-based image retrieval and feature extraction: a comprehensive review. *Mathematical problems in engineering*, *2019*.

Li, C., & Li, Y. (2012). Color-Decoupled Photo Response Non-Uniformity for Digital Image Forensics. IEEE Transactions on Circuits and Systems for Video Technology, 22(2), 260-271.

Liu, L., Wang, Z., Qian, Z., Zhang, X. and Feng, G., 2019. Steganography in beautified images. *Mathematical Biosciences and Engineering*, *16*(4), pp.2322-2333.

Misra, S. and Arumugam, C. eds., 2022. Illumination of Artificial Intelligence in Cybersecurity and Forensics. Cham: Springer International Publishing.

Nowroozi, E., Dehghantanha, A., Parizi, R.M. and Choo, K.K.R., 2021. A survey of machine learning techniques in adversarial image forensics. Computers & Security, 100, p.102092.

Passi, A., 2021, August. Digital Image Forensic based on Machine Learning approach for Forgery Detection and Localization. In *Journal of Physics: Conference Series* (Vol. 1950, No. 1, p. 012035). IOP Publishing.

Redi, J., Taktak, W., & Dugelay, J. (2010). Digital image forensics: a booklet for beginners. Multimedia Tools and Applications, 51(1), 133-162.

Sedik, A., Faragallah, O.S., El-sayed, H.S., El-Banby, G.M., El-Samie, F.E.A., Khalaf, A.A. and El-Shafai, W., 2022. An efficient cybersecurity framework for facial video forensics detection based on multimodal deep learning. Neural Computing and Applications, pp.1-18.

Thai, T. H., Cogranne, R., Retraint, F., & Doan, T. N. C. (2016). JPEG Quantization Step Estimation and Its Applications to Digital Image Forensics. IEEE Transactions on Information Forensics and Security, 11(1), 123-133.

Tripathi, M., 2021. Analysis of convolutional neural network based image classification techniques. *Journal of Innovative Image Processing (JIIP)*, *3*(02), pp.100-117.

Yang, P., Baracchi, D., Ni, R., Zhao, Y., Argenti, F. and Piva, A., 2020. A survey of deep learning-based source image forensics. Journal of Imaging, 6(3), p.9.