

A COMPREHENSIVE APPROACH TO PREVENTING DATA LEAKAGE AND STRENGTHENING CYBERSECURITY

TMP-2023-24-082

PROPOSAL PROJECT REPORT

LOGU J.J – IT20638818

B.Sc. (Hons) in Information Technology Specializing in Cyber Security

Department of Computer System and Engineering

Sri Lanka Institute of Information Technology

August 2023

A COMPREHENSIVE APPROACH TO PREVENTING DATA LEAKAGE AND STRENGTHENING CYBERSECURITY

TMP-2023-24-082

PROPOSAL PROJECT REPORT

(Malicious Image detection and classification using Deep Learning techniques)

LOGU J.J– IT20638818

Supervisor – Mr. Amila Senarathne

Co – Supervisor –

B.Sc. (Hons) in Information Technology Specializing in Cyber Security


Department of Computer System and Engineering

Sri Lanka Institute of Information Technology

August 2023

DECLARATION

We declare that this is our own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	Student ID	Signature	Date
LOGU J. J	IT20638818		25/08/2023

The above candidate is carrying out research for the undergraduate Dissertation under my supervision.

Signature of the Supervisor

Date

.....

.....

Signature of the Co-Supervisor

Date

.....

.....

ABSTRACT

Nowadays, the escalating threat of data breaches has become a paramount concern for businesses. The safeguarding of sensitive data is a top priority, necessitating the attention of top management, IT administrators, and experts alike. Traditional security measures like firewalls are proving inadequate in the face of evolving cyber threats. Data Loss Prevention (DLP) systems are a new desire in the struggle for data security. This research initiative comprises 4 interconnected subcomponents aimed at providing a comprehensive strategy to minimize data loss and enhance cybersecurity.

The first subcomponent, "Safeguarding Systems by Identifying Unusual User Patterns " uses a machine learning algorithm to proactively identify unusual user behavior and access patterns, enabling timely responses to potential threats. The second, "Utilizing NLP Techniques for Enhanced Data Protection" explores natural language processing algorithms to automatically identify sensitive information in text, boosting data safety by detecting insider risks. The third, "Unveiling Patterns and Anomalies to Mitigate Data Breach Risks" employs innovative data analysis tools and machine learning to uncover hidden patterns and minimize breach risks. lastly, "Malicious Image Detection and Classification Using Deep Learning Techniques" focuses on defending against image-based cyberattacks by utilizing convolutional neural networks to distinguish between legitimate and malicious images.

With a foundation in expertise and practicality, this research not only contributes to the theoretical advancements of cybersecurity but also equips businesses with effective tools to navigate the complex digital landscape and safeguard their critical data.

This study proposes a ground-breaking strategy that improves the efficiency of harmful picture identification and classification by using the capabilities of deep learning, notably convolutional neural networks (CNNs). The main goal of this work is to create cutting-edge models that can precisely locate and detect malware cloaked in picture files. This study addresses the critical challenge of improving cybersecurity by using the sophisticated capabilities of CNNs, known for their prowess in image processing.

his study also investigates ensemble approaches' potential as a spur for improved accuracy. The classification accuracy is significantly improved by using a fleet of CNN models, each of which is specially set up with unique architectural details, hyperparameters, or initializations. By combining distinct model predictions into a holistic anomaly score, the ensemble technique not only increases classification accuracy but also strengthens the system's stability. This study combines accuracy, flexibility, and resilience in an original way. The result is an advanced system ready to combat the growth of harmful image-based attacks. This study promotes a paradigm change in boosting cybersecurity by using the combination of deep learning and ensemble approaches, giving stakeholders the ability to proactively detect, isolate, and eliminate harmful entities encoded in picture files.

TABLE OF CONTENTS

DECLARATION.....	3
ABSTRACT.....	4
1. INTRODUCTION	6
1.1 Background & Literature Survey.....	6
1.2 Research Gap	9
1.2.1 Research Gap Comparison Chart	9
1.3 Research Problem	11
2. OBJECTIVES.....	11
2.1 Main Objective	12
2.2 Sub Objective.....	12
3. METHODOLOGY	12
3.1 System Diagram.....	14
3.2 Technologies	15
4. PROJECT REQUIREMENTS.....	16
4.1 Functional Requirements	16
4.2 Non-Functional Requirements.....	16
5. WORK BREAKDOWN STRUCTURE (WBS)	17
6. GANTT CHART.....	17
7. BUDGET	18
8. COMMERCIALIZATION	19
9. REFERENCE	19

1. INTRODUCTION

1.1 Background & Literature Survey

1. What is malicious Image?

Malicious images refer to digital images that are intentionally designed to carry out harmful or malicious activities. These images can be used to exploit vulnerabilities, execute attacks on computer systems and networks.

2. What is Malicious Image Classification?

Malicious image classification is the process of identifying images that contain harmful or inappropriate content, such as malware or explicit material. This differs from regular image classification, which aims to categorize images based on their visual features.

3. Importance of Malicious Image Classification?

Malicious images can pose a serious threat to individuals and organizations alike. For example, an image containing malware could infect a system when opened, potentially compromising sensitive data or causing damage to the system. Malicious images can also be used for phishing attacks or other forms of cybercrime, making it essential to be able to identify and mitigate these threats.

4. How can this support the overall output?

the system can support the overall output in various ways, improved cybersecurity measures and user safety in the digital environment. For example

- Enhanced Malicious Image Detection
- Early Threat Detection and Mitigation
- Robustness against Adversarial Attacks
- Improved User Safety and Trust

An enormous amount of research has been done to provide efficient methods for identifying and categorizing harmful picture sharing and distribution, as a result of the increasing increase of this kind of material. A thorough analysis of the current literature indicates several strategies and procedures used to achieve image security. Traditional picture analysis methods, such as content-based filtering and rule-based approaches, have proved essential in thwarting blatant harmful information. These techniques look for certain patterns or phrases connected to dangerous pictures using established rules or heuristics. However, due to their failure to adjust to changing threats and propensity to overlook increasingly subtle kinds of harmful information, they are often rendered ineffective.

Patil, V., Shetty, S., Tawte, A., & Wathare, S. (2023). Deep learning and binary representational image approach for malware detection. 2023 International Conference on Power, Instrumentation, Control and Computing (PICC), 1–7 [1].

A new departure from traditional approaches is the suggested malware classification technique using image-based representations. While convolutional neural networks (CNNs) and deep learning provide a new viewpoint on malware research, which has historically relied on code-level inspection. By converting binary representations into visual patterns using CNNs' pattern recognition skills, this developing area enables the precise categorization of malware families. This innovative method, which achieved an astonishing 92.73% accuracy, shows promise for increased effectiveness and interpretability. Future studies should focus on this strategy's interpretability, adversarial robustness, and scalability in order to counter growing malware threats. This study used the CNN sequential model to try to improve and simplify the categorization of malware. Malware is always evolving, making it challenging to recognize it.

Korkmaz, A., & Hanilci, C. (2022). Image forgery detection based on parallel convolutional neural networks. 2022 30th Signal Processing and Communications Applications Conference (SIU), 1–4 [2].

Robust picture fraud detection has become an important issue across industries including law, health, and communication in response to the increasing complexity of image alteration technologies. While earlier research focused on texture analysis and watermarking, the suggested technique is a pioneer in the simultaneous integration of three different deep neural network topologies. This deviation from standard uniform DNN techniques indicates an understanding of the complexity of picture fraud. The method's effectiveness is evaluated on a variety of datasets, producing promising classification accuracy. Future research should improve interpretability and address adversarial weaknesses. Overall, the effectiveness and potential applications of picture forgery detection are improved by this creative integration of parallel DNN architectures.

Das, D., & Naskar, R. (2022). Image splicing detection based on deep convolutional neural network and transfer learning. 2022 IEEE 19th India Council International Conference (INDICON), 1–6 [3].

Concerns regarding its improper usage have been raised because to the rise in easily accessible, inexpensive image editing software, which has increased the frequency of digital picture fraud. Notably, the practice of "image splicing," which combines bits from many sources to produce false composite images, has become more popular. Machine/deep learning-based detection and feature engineering are the two main kinds of detection techniques. Deep learning provides resilience, but the intricacy of its training demands significant resources. This paper develops an effective method for picture splicing detection by fusing deep learning with transfer learning. The model achieves state-of-the-art accuracy by using a Deep Convolution Neural Network (CNN) with MobileNetV2's pre-trained convolution layers. The foundation for this study is laid by existing research in deep learning for image modification, including transfer learning. For detecting spliced images, the creative combination of MobileNetV2 with transfer learning provides a novel approach and illustrates a viable route toward resource-efficient yet accurate counterfeit detection.

Deepa, K., Adithyakumar, K. S., & Vinod, P. (2022). Malware Image Classification using VGG16. 2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS), 1–6 [4].

Static and dynamic analysis are used to identify harmful apps since standard signature-based techniques can't keep up with newly developed malware. Using the 'Maling' dataset, which contains images of malware families, this work develops a unique strategy for malware classification by integrating deep learning methods with image-based attributes. Although deep learning in malware detection has been studied before, the combination of varied classifiers including SVM, XGBoost, DNN, and Random Forest with a pre-trained VGG16 model for feature extraction is novel. This method for locating malware families shows promise in terms of improving classification precision and flexibility. The methodology's effectiveness in correctly detecting malware families is supported by experimental assessments carried out on a significant dataset, which has contributed to the development of reliable and adaptable malware detection systems. This strategy may be used in future studies to meet new risks and widen categorization methods.

1.2 Research Gap

Convolutional neural networks (CNNs), in particular, show tremendous promise for the categorization and detection of malicious images via the convergence of deep learning approaches. However, there are significant research gaps and difficulties that need consideration and investigation. gap is caused by the dearth of different datasets that correctly depict the wide range of possible harmful picture formats. Deep learning algorithms must rely on extensive, representative datasets in order to be effective. To close this gap, concentrated efforts must be made to collect and disseminate datasets that include a wider range of harmful material, including different picture sources and image alteration methods.

A substantial difficulty is also presented by the sensitivity of deep learning models to adversarial assaults. The models must be fine-tuned to be resilient against such assaults, and this is a crucial study area. Investigating strategies such as adversarial training and robust optimization might increase the models' robustness and dependability in real-world circumstances. While the proposed study has a focus on model accuracy, there is a critical need to address the explainability and interpretability of deep learning models. The transparency and reliability of these models may be increased by offering insights into model choices using strategies like attention visualization or attribution approaches.

The suggested method for classifying harmful images provides numerous noteworthy improvements over current techniques. The current systems often depend on established approaches like content-based filtering or rule-based procedures, which are restricted in their capacity to respond to changing threats and spot subtle harmful information. The suggested method, in contrast, uses deep learning, specifically CNNs, to automatically recognize the detailed characteristics and patterns that distinguish fraudulent photos.

Due to their dependence on predetermined criteria, current systems may not be able to reliably identify new types of dangerous information. By fine-tuning CNN models on a particular dataset of harmful and benign pictures, the proposed approach closes this gap by being able to recognize even minute variations between malware and typical image content. The use of ensemble approaches, with numerous CNN models trained on various configurations, further improves classification resilience and accuracy.

1.2.1 Research Gap Comparison Chart

Existing system	Our approach
Limited Model Performance	fine-tuning each CNN models
Limited use of ensemble methods	In-depth implementation of ensemble methods for improved accuracy and robustness.
Standard CNN architectures used for image classification.	Exploration of advanced CNN architectures optimized for malicious image classification.

Table-1 Existing system vs our approach system.

1.3 Research Problem

Malicious image assaults have become a prevalent and worrying kind of cyber danger in today's digital environment. These attacks take advantage of users' propensity to engage with visual media by embedding dangerous material inside what seem to be innocent graphics. Malicious pictures may refer to a range of risks, including as phishing attempts, the spread of malware, and the sharing of unsuitable or dangerous information. According to recent statistics, up to September 2021, harmful image assaults increased alarmingly and accounted for over 25% of all recorded cyber incidents. Given that attackers continue to take advantage of the flaws present in our growing dependence on digital imaging, this percentage emphasizes the significance of comprehending and managing this kind of danger. This pattern highlights the need for creative research that is in line with the current danger environment.

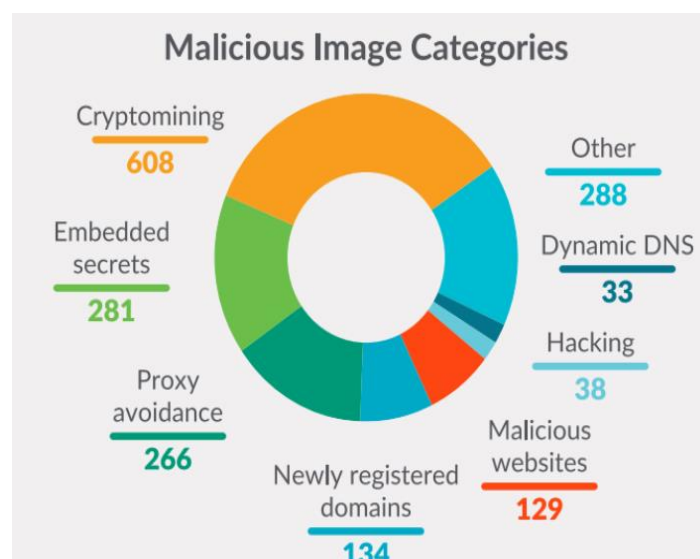


Figure 1.1 Malicious Image Categories

Organizations encounter a broad range of difficulties in their efforts to combat harmful image assaults that go beyond technological issues. These difficulties are influenced by a variety of elements, such as organizational dynamics, resource allocation, and communication tactics. Implementing effective security systems against this expanding cyber threat requires overcoming these obstacles. Modern harmful image assaults are sophisticated; thus, study must take an interdisciplinary approach. Understanding the complex techniques employed by attackers to conceal dangerous payloads inside photos requires a combination of skills in computer vision, encryption, and cybersecurity.

2. OBJECTIVES

2.1 Main Objective

The main objective of this project is to build a powerful deep learning model with the capacity to recognize fraudulent photos efficiently and reliably. The model is created to automatically understand detailed visual patterns and attributes linked to dangerous material by using the power of deep neural networks. The study comprises thorough model training utilizing several datasets that include a broad variety of harmful picture kinds. The goal is to achieve a high degree of accuracy in recognizing and categorizing harmful pictures by using cutting-edge architectures like convolutional neural networks (CNNs), their derivatives. If this goal is accomplished successfully, image security might be considerably improved, providing better defense against the spread of hazardous material across a variety of digital domains, including online platforms, communication channels, and beyond.

2.2 Sub Objectives

- **Identifying Effective Deep Learning Techniques for Malicious Image Classification: -**

This research project explores the discovery of the most efficient deep learning methods in the quest of strong malicious picture categorization. The research attempts to identify the approaches most suitable for precisely identifying hazardous material inside photos by using the broad landscape of deep neural networks, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and their variations. The study aims to shed light on the strategies that demonstrate higher classification accuracy and resilience in the face of emerging threats by analyzing the performance of these techniques over a varied variety of datasets and harmful picture types.

- **Comparing and Creating Multiple CNN Models: -**

This sub-objective entails a thorough investigation of convolutional neural networks (CNNs) via the comparison and development of several models specifically suited for the categorization of harmful images. The goal of the study is to thoroughly assess several CNN designs, including varied depths, kernel sizes, and feature extraction techniques. The research seeks to identify the CNN configurations that provide the best outcomes in terms of accuracy, sensitivity, and specificity via methodical testing. This investigation will shed light on the architectural and design decisions that most significantly aid in identifying and categorizing fraudulent pictures.

- **Fine-Tuning Each CNN Model: -**

This phase, which follows the development of several CNN models, is concerned with optimizing each model for improved performance in the categorization of harmful images. To improve a model's capacity to identify minor harmful patterns, fine-tuning entails tweaking model parameters including learning rates, dropout rates, and regularization methods. The goal of the study is to determine the best hyperparameters for each CNN architecture via iterative changes and cross-validation. This phase helps the models' ability to generalize and be as accurate as possible across different picture datasets and probable threat situations.

- **Implementing the Ensemble Model: -**

This study goal entails constructing an ensemble model to further improve the accuracy and resilience of harmful picture categorization, building on the insights acquired from individual CNN models. The ensemble model integrates predictions from many CNN models, using their various advantages to enhance classification performance as a whole. The ensemble method tries to reduce the limits of individual models and obtain greater accuracy rates by using approaches like majority voting or weighted averaging, eventually resulting in a more dependable and efficient solution for harmful picture identification.

3. METHODOLOGY

Convolutional neural networks (CNNs) are used extensively in the approach for harmful picture identification and classification using deep learning methods to support cybersecurity operations. The main goal is to build sophisticated algorithms that can identify and categorize malware cloaked inside picture files, improving data breach prevention. Acquiring and curating a large collection of both malicious and benign photos is the first stage. This dataset serves as the basis for developing, testing, and assessing the performance of deep learning models. The dataset includes a wide range of harmful material as well as typical visual data to guarantee accuracy and generalization.

One of the most important stages of the process is fine-tuning. The models may be improved to accurately differentiate between hazardous malware and benign picture material by adapting current CNN architectures, like as VGG, ResNet, or Inception, to the particular goal of malicious image classification. To increase classification accuracy, this fine-tuning procedure involves modifying hyperparameters, dropout rates, and learning rates. An important development in the approach is the inclusion of ensemble methods. There are many trained CNN models, each with a different architecture, set of hyperparameters, or initialization. This ensemble technique makes use of the variety of these models since each one's particular capabilities enhance the performance of the group as a whole. The ensemble increases the resilience and accuracy of anomaly detection by combining the multiple model results to provide a final anomaly score.

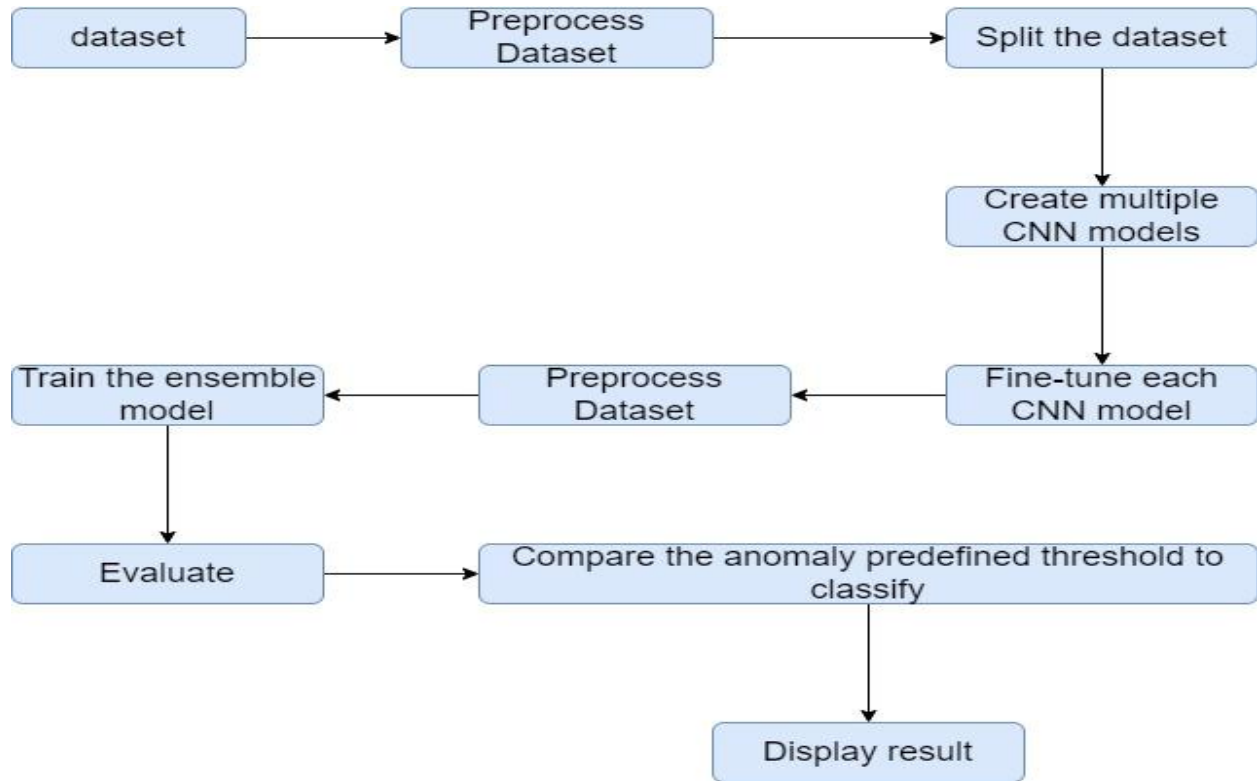
The methodology's iterative training and validation is a key component. The models go through several training cycles, and with each cycle, their capacity to distinguish between harmful and benign pictures is improved. The performance of the models is further guaranteed by cross-validation methods across several dataset subsets. The methodology's completion comprises meticulous testing and assessment on hypothetical data. The models are put through realistic situations to see how well they can identify harmful photos while reducing false positives.

Step for implementation

comprises of various key components intended to create and assess efficient models for improved cybersecurity in a methodical manner.

1. Gathering and Preparing Data
2. Choice of Model Architecture
3. Training and fine-tuning
4. Ensemble Model construction
5. Training that is iterative and validation
6. Evaluation of Performance Metrics
7. Testing and evaluation in the real world
8. Comparing and choosing
9. Deploying and integrating models

3.1 System Diagram



3.2 Technologies

- Deep Learning
- Convolutional Neural Networks (CNNs)
- CNN architecture

4. PROJECT REQUIREMENTS

4.1 Functional Requirements

Image Classification Model: The system ought to make it easier to develop an image classification model that can correctly classify photos into predetermined groups. Convolutional neural networks (CNNs) are built by setting up the required architecture, configuring layers, and specifying activation functions.

CNN Model Creation: The system needs to let users choose the CNN model's design. This involves deciding on the quantity of fully linked layers, convolutional layers, and activation processes. Input dimensions, output categories, and any extra layers, such as batch normalization or dropout for regularization, should all be definable by users.

Fine-Tuning mechanism: The system must let the modification of already-built CNN models. Model parameters like learning rates, weight decay, and optimization techniques should be user adjustable. The model is modified for a particular job while preserving the pre-trained features by optimizing its weights and biases depending on a particular dataset.

4.2 Non-Functional Requirements

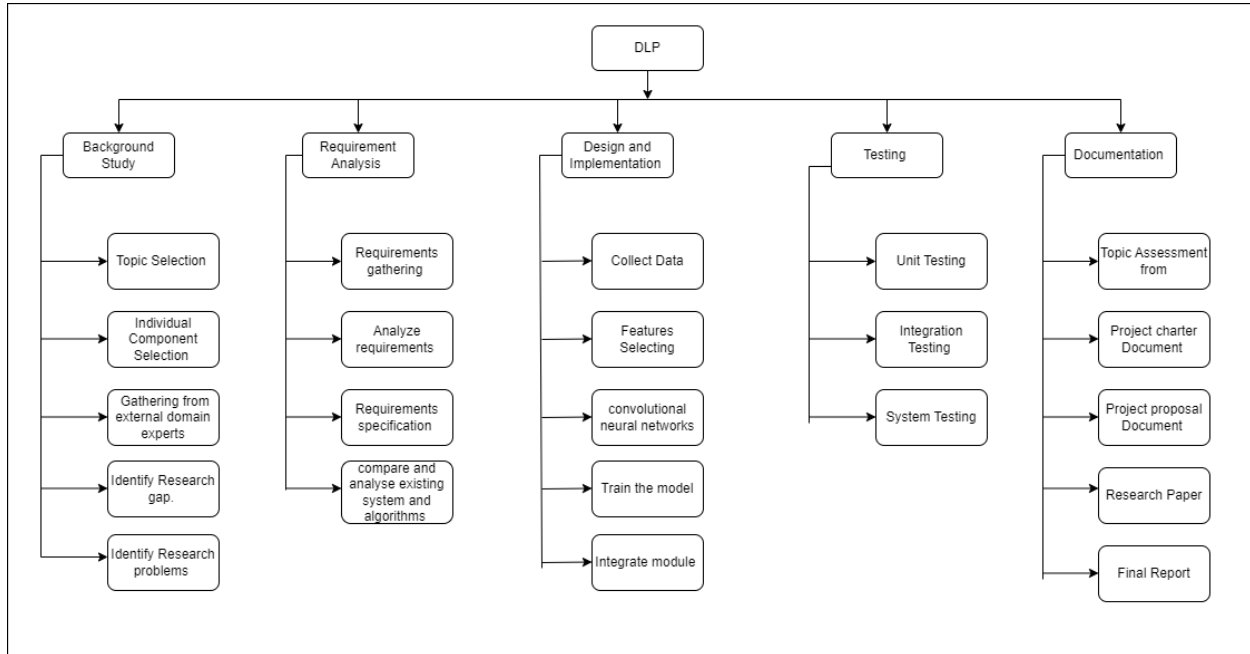
Performance: The system must show effective performance in terms of accuracy and processing speed. It needs to guarantee that picture classification activities are completed quickly and without sacrificing the accuracy of the outcomes.

resilience: The system should display resilience by continuing to function consistently under a range of conditions. Regardless of variables such different lighting conditions, picture resolutions, or possible noise in the input data, it must reliably categorize photos.

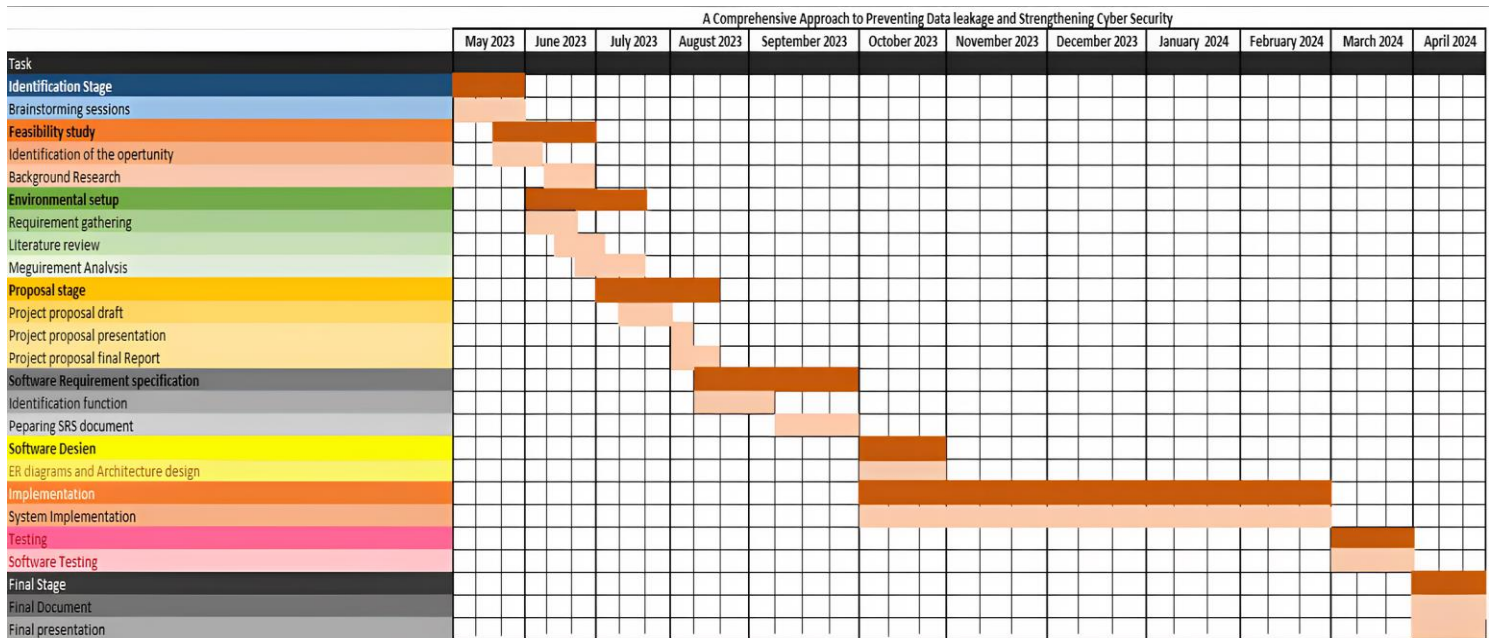
Scalability: The system must be scalable, meaning it must be able to handle growing workloads without noticeably degrading performance. The system should smoothly adapt increasing processing needs by effectively leveraging the available computer resources as the amount of pictures for categorization increases.

Security: security is essential for maintaining the integrity of secret photographs and preventing unwanted access. User identification, role-based access control, and encryption of sensitive data are some of the techniques the system must provide.

5. WORK BREAKDOWN STRUCTURE (WBS)



6. GANTT CHART



7. BUDGET

Resources	Price(LKR)
Internet	2000.00
Stationary Materials	1000.00
Electricity	2000.00
Hardware Equipment	3000.00
Paper publish cost	5000.00
Training & Testing cost	4000.00
Total	17000.00

Table-2 Estimated budget details.

8. COMMERCIALIZATION

- **Create a sales plan for the industry.**

Industry analysis: Conduct thorough market research to identify industries with the highest demand for comprehensive cybersecurity solutions. Understand their pain points, regulatory requirements, and specific challenges associated with data leakage and cybersecurity.

Tailored fee Proposition: Craft a compelling value proposition that directly addresses the unique needs of each targeted industry. Highlight how our system's capabilities align with its requirements, showcasing its effectiveness in preventing data leakage and strengthening overall cybersecurity.

Strategic Partnerships: Forge partnerships with influential industry associations, organizations, and thought leaders. Collaborate to co-host webinars, workshops, or events that position our system as an innovative solution, gaining credibility and expanding our reach within the industry.

- **Design customer subscription plans.**

Tiered Plans: Develop various subscription levels, each catering to different business sizes and needs. Offer options like basic, standard, and premium plans, each with a distinct set of features and capabilities.

Scalability: Ensure that your subscription plans are designed to accommodate the growth and changing requirements of businesses. Provide flexibility for clients to upgrade or adjust their plans as their needs evolve.

Customization Flexibility: Integrate customization options within subscription plans, allowing clients to tailor features based on their specific data protection needs. This ensures that they only pay for the functionalities they require.

- **Provide excellent customer support.**

Dedicated support team: Assign a team of knowledgeable support representatives to promptly address inquiries and issues.

Multi-Channel assist: Provide support through various channels such as email, live chat, and phone.

24/7 Availability: offer round-the-clock assistance for critical concerns and urgent inquiries.

Knowledge Base: Develop an online resource with FAQs, tutorials, and troubleshooting guides.

Continuous training: Offer training sessions to help clients maximize the benefits of the system.

9. REFERENCES

- [1] V. S. S. T. A. & W. S. Patil, "Deep learning and binary representational image approach for malware detection.," International Conference on Power, Instrumentation, Control and Computing (PICC), 1–7., 2023.
- [2] A. & H. C. Korkmaz, "Image forgery detection based on parallel convolutional neural networks.," Signal Processing and Communications Applications Conference (SIU), 1–4., 2022.
- [3] D. & N. R. Das, "Image splicing detection based on deep convolutional neural network and transfer learning.," IEEE 19th India Council International Conference (INDICON), 1–6., 2022.
- [4] K. A. K. S. & V. P. Deepa, "Malware Image Classification using VGG16.," International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS), 1–6., 2022.
- [5] B. Hauer, "Data and information leakage prevention within the scope of information security," IEEE Access: Practical Innovations, Open Solutions, 3, 2554–2565, 2015.
- [6] E. G. & G. D. S. J. Agrawal, "Survey on data leakage prevention through machine learning algorithms.," International Mobile and Embedded Technology Conference (MECON), 121–123., 2022.

APPENDICES

ORIGINALITY REPORT			
7%	6%	1%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS
PRIMARY SOURCES			
1	www.coursehero.com Internet Source	3%	
2	Submitted to Sri Lanka Institute of Information Technology Student Paper	1%	
3	www.ijraset.com Internet Source	1%	
4	www.aidic.it Internet Source	<1%	
5	Debjit Das, Ruchira Naskar. "Image Splicing Detection based on Deep Convolutional Neural Network and Transfer Learning", 2022 IEEE 19th India Council International Conference (INDICON), 2022 Publication	<1%	
6	K Deepa, K S Adithyakumar, P Vinod. "Malware Image Classification using VGG16", 2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS), 2022 Publication	<1%	