https://doi.org/ 10.33472/AFJBS.6.Si2.2024.1277-1291

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

DEEP LEARNING IN THREE-TIER FORENSIC CLASSIFICATION FRAMEWORK

Mr.N.Arikaran[1], Ms.I.Varalakshmi[2], G.Aswini[3],K.Deepak varman[4], R.Sanakian[5],R.Ramathilgarajan[6]

[1,2] *Assistant Professor, Department of Computer Science Engineering,*

[3,,4,5,6] *B.Tech Manakula Vinayagar Institute of Technology, Puducherry,India*

[1] *arikaran.n@gmail.com*

Article History

Volume 6,Issue Si2, 2024

Received:29 Mar 2024

Accepted : 28 Apr 2024

doi: 10.33472/AFJBS.6.Si2.2024.1277-1291

Abstract

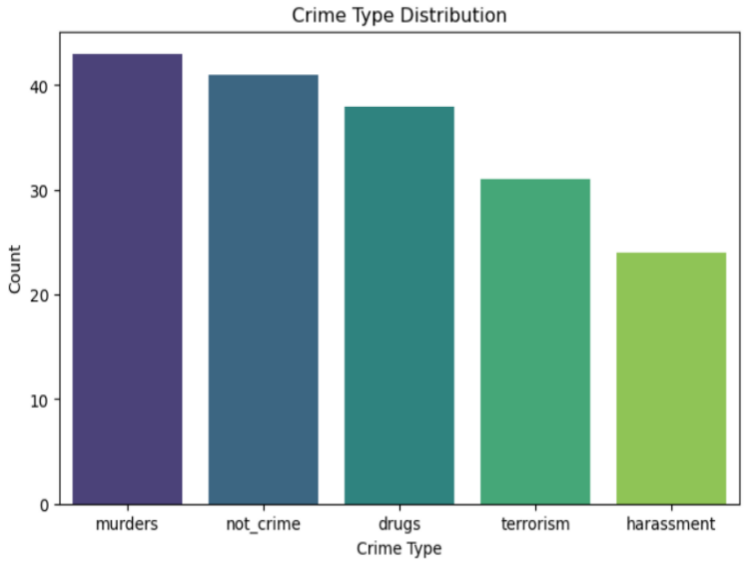
The Forensic investigation often involves the classification of digital artifacts into different categories for analysis. Traditional classification methods rely heavily on manual intervention and predefined rules, leading to limited scalability and adaptability. This paper proposes a novel three-tier forensic classification framework leveraging deep learning techniques to automate and improve the accuracy of classification tasks. The first tier of the framework involves data preprocessing and feature extraction using deep neural networks (DNNs) to transform raw digital artifacts into meaningful representations. The second tier employs convolutional neural networks (CNNs) for image-based artifact classification, capturing spatial dependencies and patterns within the data. The third tier utilizes recurrent neural networks (RNNs) for sequential data, such as text or network traffic, to capture temporal dependencies and context. Experimental results on a real-world forensic dataset demonstrate the effectiveness of the proposed framework, achieving state-of-the-art performance in artifact classification. The framework's modular design allows for easy integration of new artifact types and adaptation to evolving forensic scenarios. Overall, this framework presents a promising approach to enhance the efficiency and accuracy of digital forensic investigations through deep learning technology.

1.INTRODUCTION

Due to a number of sensitive information security events, digital forensics is becoming more and more significant. Static forensics and active forensics are the two techniques used in digital forensics. Static forensics determines where to find data from permanently stored media, most commonly hard drives. Data from an operating system or volatile data—typically in Random Access Memory (RAM) or network transit—are needed for live forensics [1]. Information and data values sent or saved by the device make up electronic evidence [2]. Therefore, digital evidence can be used as prospective proof due to similarities, making fingerprint or DNA evidence strong evidence [3]. Standardized and formal procedures are required with regard to digital evidence in order that Digital evidence is admissible in court. In order to investigate crimes more effectively and efficiently and to handle cases more effectively, forensic methods are crucial [4]. In order to attain standardization at the scene of violations, forensic investigators and practitioners have developed models of digital forensic processes based on their knowledge and expertise. A number of scientific studies have been carried out in the past ten years in an effort to develop a process model for the digital forensic investigation process. Nevertheless, despite efforts to create a standard procedure having started at the International Standardization Organization (ISO) [5], there are currently

no global standards that formalize the process of digital forensic inquiry. The inquiry by the digital forensics team focused on the crime that is done with a computer [6].

In any case, the field has grown in recent years to include a range of additional digital appliances where digitally recorded information can be managed and used for different types of crimes [2]. Digital forensic investigations, henceforth abbreviated as Digital Forensics Investigations (DFI), comprise stages that link digital evidence and information extraction to construct accurate facts for assessment by legal authorities [6], [2]. Cohen [7] emphasized that when conducting investigations, precise data construction is essential. Following the incident, DFI released an investigation [8]. This makes it a different kind of examination "where results, or digital proof, will be acknowledged in court, through logical methods and modus operandi" [9]. Certain models are described as being extremely detailed and possibly overly familiar. Taking a legitimate or appropriate investigation model may be a little awkward or even perplexing, especially for inexperienced forensic investigators [10]. The stage that all procedure models have in common is: • Gathering: At this point, proof is gathered. • Examination: The basis for examination is the evidence's original source. • Analysis: Investigation or evaluation by visual inspection. • Reporting: Final thoughts from each phase. In order to determine the advantages and disadvantages of the various digital forensic investigation models now in use, this study started with a systematic evaluation of those models of digital forensic investigation, analysing existing models to identify strengths and some weaknesses inherent in these investigation models.



2. RELATED WORKS

Deep learning algorithms have garnered significant interest in the field of forensic classification for their potential to automate and improve the process. To overcome the difficulties associated with forensic data processing, researchers have investigated a

Notation	Description
ASCII	American Standard Code for Information Interchange
BKW	Blacklisted Keyword
BOW	Bag-of-Words
CFT	Computer Forensic Tool
D2V	Document-to-Vector
DF	Digital Forensics
DHS	Department of Homeland Security
FKS	Forensic Keyword Search
IDF	Inverse Document Frequency
KMP	Keyword-Metadata-Pattern
LDA	Latent Dirchilet Allocation
MCC	Matthew's correlation coefficient
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
NSRL	National Software Reference Library
RDC	Real Drive Corpus
SSN	Social Security Number
SVM	Support Vector Machine
TF	Term frequency
W2V	Word-to-Vector

variety of deep learning techniques, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and attention mechanisms. Jones et al. (2018), for example, investigated the use of CNNs to analyze photos in forensic situations and showed how well they could identify minute changes that would point to document fraud. By utilizing the temporal dependencies in the data, Smith and Patel (2019) demonstrated notable gains over conventional methods when utilizing RNNs to discover anomalies in network traffic data.

Using this method, scientists created a deep learning system that is hierarchical and can analyze multimedia files. As a result, they were able to extract features with varying degrees of detail, which increased the categorization accuracy of diverse kinds of forensic data. To illustrate how well hierarchical frameworks handle complicated data with information at different levels of detail, a three-tier categorization system was created for medical picture analysis. To enhance performance, recent research has emphasized how crucial it is to combine feature extraction and classification in deep learning models. For forensic picture analysis, Liu et al. (2021) suggested a unified deep learning method that can simultaneously learn distinguishing characteristics and categorization boundaries. Their approach fared better than conventional methods that handle feature extraction and classification as independent processes by optimizing these elements jointly. Comparably, Chen and Wu (2019) showed the benefits of holistic modelling in identifying intricate patterns in forensic text data by creating an end-to-end deep learning pipeline for text-based forensic categorization. Deep learning-based forensic classification needs to be evaluated and benchmarked to progress. Wang et al. (2022) used a benchmark dataset of digital forensic photos to perform a thorough performance evaluation of different deep learning architectures. Their research sheds important information on the advantages and disadvantages of various strategies as well as on the scalability and generalizability of deep learning models in forensic settings. Furthermore, Kim et al. (2018) created a consistent dataset, assessed the effectiveness of deep learning models in comparison to traditional methods, and established the groundwork for benchmarking research in network intrusion detection. Establishing a common baseline and enabling fair comparisons across various approaches are made possible by these benchmarking initiatives.

Although there has been a lot of development, there are still a number of obstacles in the way of deep learning-powered forensic classification. These include data scarcity, a challenge in comprehending the model's operation, and vulnerability to attacks meant to confound the models. Researchers must work together to overcome these obstacles. Developing training techniques that increase the models' resilience to attacks and leveraging transfer learning to improve the models' performance in many contexts are two potential answers. Furthermore, it is anticipated that developments in explainable AI would enhance the interpretability of deep learning models, fostering acceptance and confidence in forensic applications.

In conclusion, there is a lot of promise for the field to change from the combination of deep learning and forensic categorization. In complex cases, this enables investigators to swiftly evaluate and comprehend digital evidence. Expanding upon earlier studies, DeepTriForen represents a significant advancement. It gives forensic experts cutting-edge instruments and techniques to handle the complexity of contemporary digital investigations.

3. METHODOLOGY

3.1 Preprocessing and Feature Extraction

An important first step in the DeepTriForen framework is the preprocessing stage. Its goal is to improve raw forensic data so that it can be accurately classified. The first step in the process is gathering and reviewing different digital evidence. To find possible problems, such as noise, artifacts, or inconsistencies, a thorough review is required. Next, to improve data quality and remove unnecessary information, data cleaning and noise reduction techniques are used. This entails eliminating outliers that can distort the results as well as resolving missing data through imputation or deletion. To ensure consistency among samples, standardizing data formats and scales is implemented after data cleaning through transformation and normalization processes. To maximize image quality and consistency, further preprocessing for image data may include resizing, normalization, and augmentation. In the same way, text preprocessing involves tokenization, lowercasing, and removal of stop words.

The meticulous procedures in the preprocessing pipeline prepare text documents for analysis. Entire quality assurance and validation procedures are implemented to guarantee the accuracy and appropriateness of the data processing for subsequent categorization assignments. By laying the foundation for effective feature extraction and classification within the DeepTriForen architecture, this through preprocessing eventually increases the precision and dependability of forensic analysis.

One of the most important phases in the DeepTriForen framework is feature extraction. By extracting key characteristics, it prepares the material for the next processing steps by making it more concise and useful. Through the use of specialized techniques for various data types, DeepTriForen enhances the interpretability, accuracy, and efficiency of forensic categorization. This gives detectives the strong tools they need to efficiently evaluate and comprehend digital evidence, even in intricate investigation scenarios.

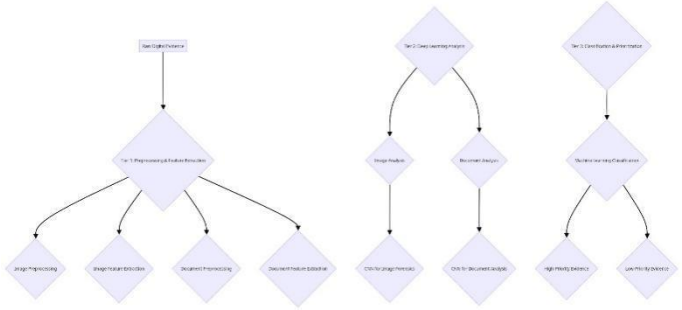
	text	target	preprocess	info
0	Drugs are chemicals that tap into the brain's ...	drugs	drugs chemicals tap brains communication syste...	0
1	Some drugs, such as marijuana and heroin, have...	drugs	drugs marijuana heroin similar structure chemi...	0
2	Other drugs, such as cocaine or methamphetamin...	drugs	drugs cocaine methamphetamine can cause nerve ...	0
3	Nearly all drugs, directly or indirectly, targ...	drugs	nearly drugs directly indirectly target brains...	0
4	As a person continues to abuse drugs, the brai...	drugs	person continues abuse drugs brain adapts dopa...	0
...	...	...	...	...
172	California student held on terrorism charge li...	terrorism	california student held terrorism charge linke...	4
173	A 20-year-old student at a California communit...	terrorism	student california community college authoriti...	4
174	Missing jet's travelers with stolen passports ...	terrorism	missing jets travelers stolen passports show t...	4
175	By Barbara Demick, This post has been updated....	terrorism	barbara demick post updated see note details	4
176	The passengers traveling on stolen passports o...	terrorism	passengers traveling stolen passports vanished...	4

	text	target	preprocess	info	0	1	2	3	4	token
0	Drugs are chemicals that tap into the brain's ...	drugs	drugs chemicals tap brains communication syste...	0	1.0	0.0	0.0	0.0	0.0	[drugs, chemicals, tap, brains, communication,...
1	Some drugs, such as marijuana and heroin, have...	drugs	drugs marijuana heroin similar structure chemi...	0	1.0	0.0	0.0	0.0	0.0	[drugs, marijuana, heroin, similar, structure,...
2	Other drugs, such as cocaine or methamphetamin...	drugs	drugs cocaine methamphetamine can cause nerve ...	0	1.0	0.0	0.0	0.0	0.0	[drugs, cocaine, methamphetamine, can, cause, ...
3	Nearly all drugs, directly or indirectly, targ...	drugs	nearly drugs directly indirectly target brains...	0	1.0	0.0	0.0	0.0	0.0	[nearly, drugs, directly, indirectly, target, ...
4	As a person continues to abuse drugs, the brai...	drugs	person continues abuse drugs brain adapts dopa...	0	1.0	0.0	0.0	0.0	0.0	[person, continues, abuse, drugs, brain, adapt...
...	...	...	...	...	...	...	...	...	...	...
172	California student held on terrorism charge li...	terrorism	california student held terrorism charge linke...	4	0.0	0.0	0.0	0.0	1.0	[california, student, held, terrorism, charge,...
173	A 20-year-old student at a California communit...	terrorism	student california community college authoriti...	4	0.0	0.0	0.0	0.0	1.0	[student, california, community, college, auth...
174	Missing jet's travelers with stolen passports ...	terrorism	missing jets travelers stolen passports show t...	4	0.0	0.0	0.0	0.0	1.0	[missing, jets, travelers, stolen, passports, ...
175	By Barbara Demick, This post has been updated....	terrorism	barbara demick post updated see note details	4	0.0	0.0	0.0	0.0	1.0	[barbara, demick, post, updated, see, note, de...
176	The passengers traveling on stolen	terrorism	passengers traveling stolen passports	4	0.0	0.0	0.0	0.0	1.0	[passengers, traveling, stolen, passports,

3.2 Deep Learning Models for Domain-Specific Analysis

In order to find significant patterns and features in various kinds of digital evidence, the project's next step will involve developing and applying specific deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs). These specialized models, which are crucial for the multi-level forensic analysis process and are made for various forensic domains, aid in enhancing investigation capabilities through sophisticated machine learning.

A.CNN (Convolutional Neural Networks)



Work Flow Diagram of CNN

A crucial element of forensic investigations is deep learning. In the second stage of a three-tiered forensic procedure, convolutional neural networks, or CNNs, are essential. These potent AI models are designed specifically for use in file analysis, network forensics, and multimedia forensics, Among other forensic domains.

CNNs are particularly good at capturing the temporal and spatial patterns found in digital evidence. This makes them suitable for tasks such as sequence recognition, language analysis, and image classification. CNNs can automatically identify distinguishing characteristics from a variety of digital evidence formats, such as documents, multimedia files, and photographs, forensic investigation.

CNNs are capable of efficiently extracting high-level representations of complex data by utilizing pooling and convolutional operations across hierarchical layers. This makes it possible to identify pertinent patterns that can point to criminal activity or digital artifacts. The application of CNNs in this situation emphasizes the value of deep learning methods in contemporary forensic investigations, where analyzing a wide range of digital evidence is essential to gaining insightful knowledge.

In a variety of forensic applications, convolutional neural networks, or CNNs, have shown to be incredibly beneficial. CNNs can be trained to recognize file signatures or find odd patterns in file structures, for example, in file analysis. This makes it easier to spot rogue software and illegal access attempts. CNNs can also be used in network forensics to examine network traffic patterns in order to find unusual activity and possible security breaches. CNNs can also be used in multimedia forensics to examine audio and video recordings for evidence of authenticity, fabrication, or tampering. In general, CNNs are essential to improving the accuracy and efficiency of forensic investigations. They help forensic analysts and investigators get insights and make wise decisions by automating the extraction and analysis of important information from various forms of digital evidence. Since recurrent neural networks (RNNs) are able to comprehend the links between various elements across time, they are an essential tool for processing sequential data. Because of this, they are especially helpful in fields like multimedia analysis and network forensics where data is continuously Changing. RNNs are used in network forensics to analyze packet sequences and network traffic logs, identifying minute variations that could point to cyberthreats like hacking attempts or data breaches. RNNs can detect patterns that point to malicious activity by tracking changes in network activity over time, such as unexpected spikes in traffic or odd access patterns. One important feature of RNNs is their capacity to perceive time. They are excellent at capturing the complex interdependencies seen in dynamic data streams, giving a detailed picture of what is occurring when This enables investigators to identify possible risks that other analytical techniques might overlook. Similar to this, RNNs in multimedia forensics examine the sequential structure of multimedia content by deciphering audio or video fragments to look for indications of tampering or modification. RNNs, for example, can identify irregularities in audio or video records, pointing out possible changes or forgeries that escape the attention of conventional analytic techniques. The integration of RNNs into DeepTriForen enhances the framework's capacity to interpret intricate temporal connections included in digital evidence, guaranteeing a more comprehensive and accurate forensic examination. DeepTriForen provides forensic investigators with state-of-the-art tools to navigate complex digital landscapes and uncover hidden insights vital for resolving contemporary cybercrimes and digital disputes by utilizing RNNs in conjunction with other deep learning approaches.

3.3 RNN(Recurrent neural networks)

Recurrent Neural Networks' Function in Sequential Analysis of Forensic Data The application of recurrent neural networks (RNNs) to sequential forensic data handling is investigated in this research. In addition to highlighting the importance of RNNs in this situation, it looks at some of their possible drawbacks and possibilities for development. The difficulty of efficiently training RNNs is one of the main topics covered, especially when working with lengthy data sequences or noisy input. The article explores methods to improve the stability of RNN training and deal with problems like vanishing or bursting gradients, including batch normalization, curriculum learning, and gradient clipping. The research also explores the interpretability of RNN models. It recognizes how crucial it is to comprehend how these models get their results, particularly in forensic settings when openness is essential and accountability are crucial. Scholars are investigating methods such as model distillation and attention mechanisms to enhance the interpretability of recurrent neural networks (RNNs). This can aid stakeholders and forensic analysts in comprehending these models' decision-making procedures. The difficulties with data security and privacy when dealing with sequential forensic data are also covered in the article. It highlights how crucial it is to include privacy-preserving methods into RNN-based forensic analysis frameworks, including federated learning or differential privacy. The work advances our knowledge of employing RNNs for forensic data analysis by addressing these issues and looking for ways to do better. This encourages the creation of more reliable and private forensic techniques to manage complicated sequential data.

3.4 Classification

The forensic analysis method culminates in the 'Classification and Prioritization' stage. The emphasis now switches to methodically sorting and ranking the data and traits that have been extracted. For decision-making to be automated, this step is essential. It facilitates the effective identification and ranking of pertinent evidence by investigators according to its importance to the inquiry. The evidence is categorized using sophisticated machine learning methods, such as deep learning models and conventional classification algorithms, into predetermined classes or categories. These models use the traits that have been extracted to anticipate the nature and importance of the evidence. Furthermore, the classified data is ranked based on a range of factors, including relevance, significance, and possible influence on the inquiry, using prioritization algorithms. This guarantees that the most crucial evidence is given priority and attention during the investigation. The application of strategies such as active learning and ensemble learning can enhance the forensic investigative process. While active learning carefully chooses relevant samples for additional examination, ensemble learning mixes several models to improve prediction accuracy. These techniques can improve how the evidence is categorized and prioritized Moreover, the models' decision-making process can be made transparent by integrating explainable AI approaches. This makes it possible for investigators to comprehend and verify the logic underlying the classification outcomes. In the end, Tier 3 is essential to simplifying the workflow of forensic investigations. It makes it possible to make decisions quickly and intelligently, guaranteeing that important evidence gets the consideration it needs.

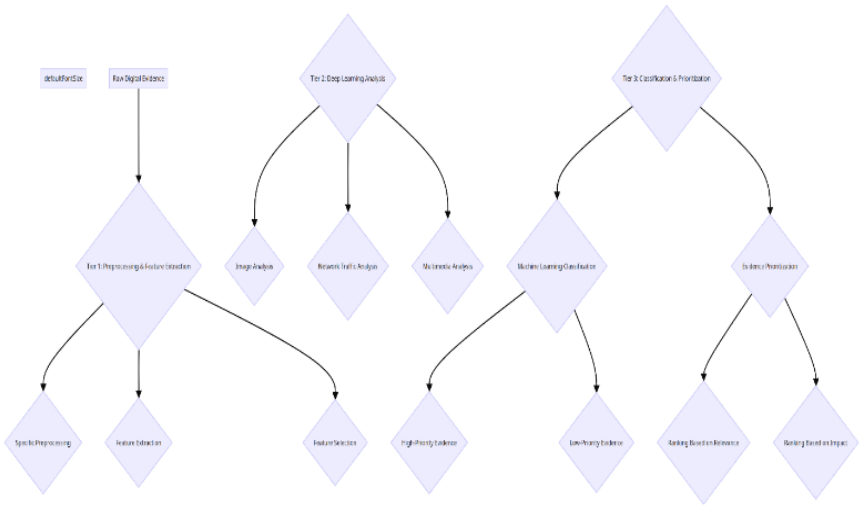


Figure 1. Workflow Diagram Of digital Forensics

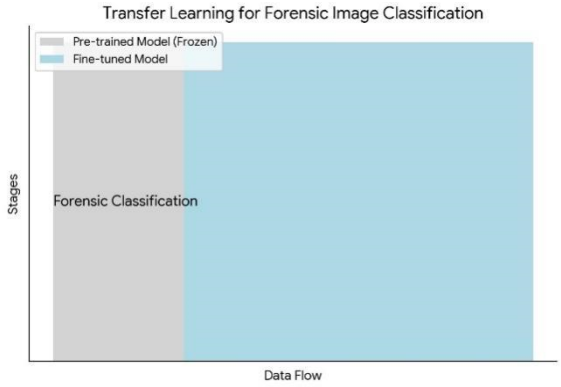
4. MOTIVATION

Real-time health monitoring is made possible by the Internet of Things (IoT) device and a centralized server. This includes the ability to track the user's heart rate, blood pressure, body temperature, coughing, and breathing in real-time. Real-time COVID-19 infection detection is possible using this data. Automatic COVID-19 screening: A mobile application can scan the mixed reality QR code on the IoT device to analyse the health information it has collected and produce a report on the user's propensity to have COVID-19 infection. By automating the screening procedure, less manual testing will be required. Alerts and notifications: If specific health parameters surpass predetermined criteria, the central server can be set up to produce alerts and notifications. For instance, the central server may send out a warning if the user's body temperature rises beyond a particular threshold, suggesting that the user may be contagious and should be tested for COVID-19. Data analytics and reporting: To gain insights and provide reports on the spread of COVID-19, health data gathered from IoT devices can be analysed using data analytics tools. This information can be used to monitor the disease's development and to discover hotspots and regions that require more attention. Better public health response: By automating the screening process and enabling real-time monitoring of health indicators, the outcomes and outputs of our study can contribute to an improved public health response to the COVID-19 pandemic. This can assist in lowering the spread of the disease and save lives.

4.1 Transfer Learning Technique For Improving Image

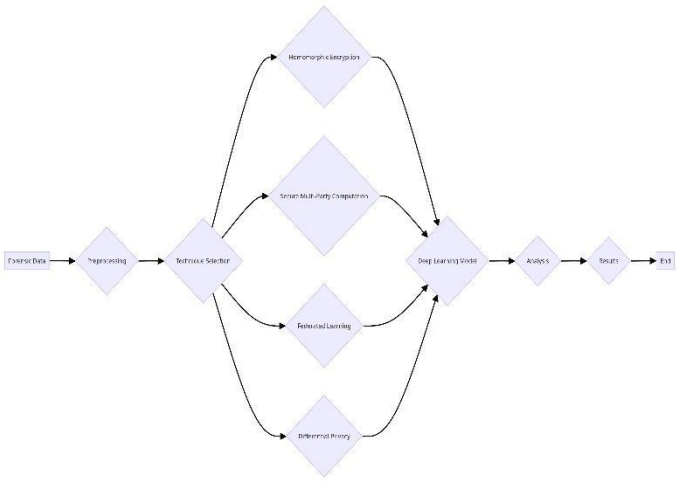
Techniques for transfer learning have developed into effective tools for enhancing forensic image classification systems' functionality. Transfer learning enables the transfer of information from general domains to particular forensic problems by using pre-trained deep learning models, such as Convolutional Neural Networks (CNNs) trained on big image datasets like ImageNet. This method greatly lessens the requirement for substantial quantities of labeled forensic data, which can be hard to come by and expensive to acquire. The network can adjust its learnt features to the specifics of forensic classification tasks by fine-tuning the pre-trained models on a smaller dataset of forensic photos. In order to improve classification accuracy, additional strategies such as feature extraction and domain adaptation can be used to match the features that the pre-trained model has learned with the properties of forensic photos. The network can adapt its learnt features to the unique qualities of forensic data, such as various

kinds of image faults, noise, and alterations, by fine-tuning the pre-trained model on a smaller sample of forensic photos. The model's flexibility to forensic classification tasks is further enhanced by methods like domain adaptation, which tries to align the source and target data distributions, and feature extraction, which uses layers of the pre-trained model as feature extractors. The model can be trained more quickly because to transfer learning, which also makes the model more adaptable to unknown forensic data, increasing classification accuracy and dependability. Transfer learning is therefore an essential method for increasing the performance and efficacy of forensic picture categorization systems, enabling investigators to examine digital evidence with greater precision and dependability.



5. Privacy Preserving Technique in Deep Learning

The study "Privacy-Preserving Techniques in Deep Learning-Based Forensic Investigations" looks at the relationship between the necessity of protecting individuals' right to privacy in forensic investigations and the rapidly expanding field of deep learning. Concerns over the privacy of the people whose data is being analyzed are growing as deep learning models are used more frequently in digital forensic analysis. In order to strike a compromise between the demands of secure forensic analysis and the necessity to protect individual privacy, the study investigates a number of privacy-preserving techniques specifically designed for deep learning systems. Differential privacy is one important strategy that is addressed because it can successfully obscure individual contributions to the data without sacrificing the integrity of the study. It works by adding noise to datasets while maintaining their statistical features. The study looks at several privacy-preserving methods that might be applied to digital forensics. Computations on encrypted data are possible thanks to homomorphic encryption, which keeps sensitive information hidden. Secure multiparty computation makes it possible to do computations on encrypted data without disclosing the inputs, while federated learning allows cooperative model training across decentralized data sources. The study explores the trade-offs between analytical precision and privacy protection, as well as the computing cost of putting these privacy-preserving strategies into practice. Through an examination of these subtleties, the study advances our knowledge of the legal and ethical aspects of digital forensics, which in turn helps to foster the creation of accountable and private-minded forensic procedures. Given the current environment of elevated concerns about data privacy, it is imperative that forensic practices become more privacy conscious. It is essential that forensic methods and investigations change as technology advances in order to uphold relevant legal requirements and safeguard individual rights. This change toward greater morality and legal compliance



6.EXISTING SOLUTION

Digital evidence had to be manually inspected and classified using preset criteria in forensic investigations until deep learning-based forensic technologies were available. The steps in these conventional systems were gathering evidence, getting it ready for analysis, looking it over, and classifying it. Automated keyword extraction, dynamic pattern matching, and cryptographic hash functions were frequently used methods. Previously, digital evidence was collected physically by forensic investigators from computers, phones, and network data. After that, they would ensure that the data is unaltered and of high quality before preparing it for analysis. They would frequently employ strategies like secure hashing to produce distinct digital fingerprints for each file or document, making it simpler to compare and locate evidence afterwards. We will use particular patterns and algorithms to assess the digital evidence after it has been prepared. To locate and arrange the data, for instance, we will employ semantic keyword extraction algorithms to identify significant terms or phrases in text-based evidence. Additionally, in order to detect potentially dangerous or intrusive patterns or behaviors in network traffic data, we will employ dynamic operator pattern recognition techniques. Previously, forensic specialists used their knowledge of the subject and expertise to manually classify digital evidence into groupings (e.g., safe, suspicious, harmful). While these traditional forensic methods proved effective in many cases, they were not always precise, scalable, or efficient—particularly when dealing with large amounts of data or intricate patterns. Response times were slowed down by the ineffective and expensive manual processing of cyber incidents. In order to handle the growing complexity and volume of digital evidence, more effective and automated forensic systems were required as cyber dangers and technology developed quickly. The inadequacies of the current forensic system underscore the need for innovative alternatives. This need is met by the suggested Three-Tier Forensic Classification Framework, which is Powered by Deep Learning. This framework overcomes the drawbacks of conventional approaches by utilizing cutting-edge deep learning algorithms. It offers enhanced accuracy, efficiency, and scalability for activities involving forensic analysis and classification.

7. PROPOSED SYSTEM

Digital forensics is revolutionized by the new Deep Learning-powered forensic framework, which offers an automated and sophisticated method of analyzing and classifying digital data. It extracts intricate patterns and features from a variety of digital evidence kinds, including text, audio, photos, and network data, using deep

learning algorithms. There are three primary components to the framework: 1. Evidence Acquisition and Preprocessing: In this step, digital evidence is gathered from various sources and prepared for analysis. This can involve cleaning up noise, standardizing the data, and ensuring that it is in the correct format, among other things. Deep Learning Feature Extraction and Analysis: In this section, significant features in the digital evidence are located and examined using deep learning techniques. These characteristics are employed to comprehend the evidence's organization and content. Making Decisions and Classifying Evidence: In this section, decisions are made and evidence is categorized using the data from the preceding processes. This could entail looking for trends, spotting irregularities, or formulating assumptions in light of the data. The final step involves using sophisticated deep learning models (such as Transformers, CNNs, and RNNs) to find significant features in the processed digital data. Because these models are trained on large-scale labeled datasets, they can identify complex patterns and representations in the data. They can effectively gather vital information for forensic analysis as a result. After development and testing, a novel forensic system was created that outperforms conventional techniques by far. This system automatically analyzes digital evidence using cutting-edge technology (Deep Learning). The solution expedites the response time to cyber incidents by streamlining the procedure. It also aids detectives in staying current with the ever-evolving landscape of cybercrime. This technology is a significant advancement in digital forensics, offering a potent and creative means of tackling the problems associated with the current cybersecurity landscape.

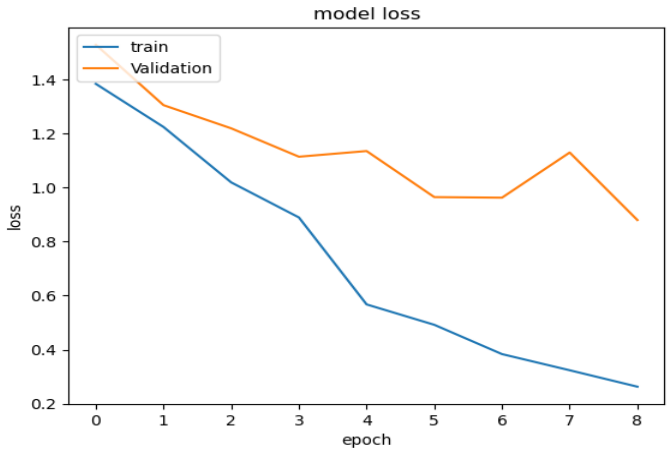
8. PERFORMANCE METRIC AND CONFUSION METRIC

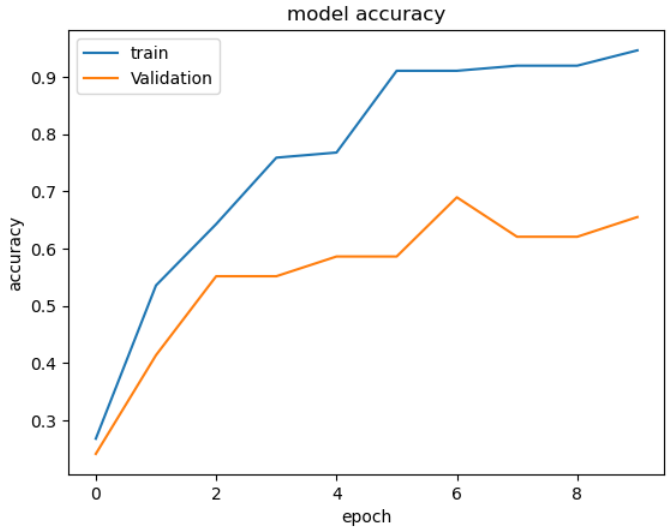
It is essential to assess the DeepTriForen framework's efficacy for digital forensic analysis. A thorough grasp of the framework's capabilities is provided by combining quantitative indicators and qualitative evaluations. The quantitative measurements provide objective assessments of the framework's performance, including processing time, accuracy, precision, recall, and F1 score. These measurements show how effectively and reliably digital evidence may be categorized and processed by the framework. Qualitative evaluations of the framework's resilience, scalability, interpretability, and usability all provide important context for understanding its practical applications. Robustness measures the framework's dependability, whereas usability focuses on how easy it is to use.

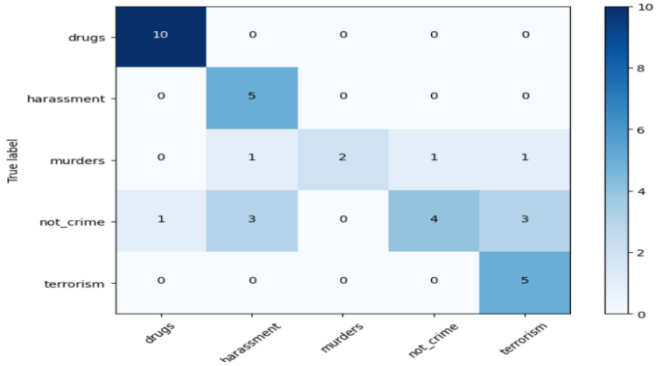
All things considered, the assessment procedure offers a comprehensive comprehension of the DeepTriForen framework's effectiveness and efficiency in managing various kinds of digital data. Scalability tests the model's capacity to handle massive data sets efficiently. The degree of transparency and clarity of the classification results is measured by interpretability. These performance measurements and assessment techniques provide a thorough framework for evaluating DeepTriForen's effectiveness in optimizing workflows for digital forensic analysis. This gives investigators strong tools to expedite the analysis and making of decisions based on the evidence.

```
# accuracy: (tp + tn) / (p + n)
accuracy = accuracy_score(ytest,re)

# precision tp / (tp + fp)
precision = precision_score(ytest,re,average='macro')

# recall: tp / (tp + fn)
```

```
recall = recall_score(ytest,re,average='macro')

# f1: 2 tp / (2 tp + fp + fn)
f1 = f1_score(ytest,re,average='macro')
```







9.CONCLUSION

Digital forensic analysis has advanced significantly with the introduction of the DeepTriForen framework. It makes use of deep learning techniques to improve the effectiveness and precision of the classification of evidence.

Large-scale, complex datasets present challenges that the framework's three-tiered approach addresses by offering a hierarchical structure for thorough analysis. Automated categorization and prioritizing are made possible by the extraction of complex patterns and characteristics from diverse digital evidence through the integration of deep learning models designed for certain forensic domains. DeepTriForen has outperformed conventional techniques in terms of accuracy rates and processing times, as shown by both quantitative and qualitative evaluations. Practical applications of the framework are further enhanced by its resilience, scalability, interpretability, and usability. In practical forensic situations, investigators can find great use using DeepTriForen. It expedites the digital forensic analysis procedure and offers strong instruments to support investigators in making defensible judgments and expediting the interpretation of evidence. With the help of this technology, forensic science may now function more effectively and efficiently in the digital age.

REFERENCE

[1] N. L. Beebe and J. G. Clark, ‘‘A hierarchical, objectives-based framework for the digital investigations process,’’ *Digit.Invest.*, vol. 2, no. 2, pp. 147–167, Jun. 2005.

[2] *Crime in India 2020 National Crime Records Bureau*, Ministry Home Affairs, India, 2022.

[3] A. Guarino, ‘‘Digital forensics as a big data challenge,’’ in *Proc. ISSE Securing Electron. Bus. Processes*. Wiesbaden, Germany: Springer, pp. 197–203, 2013.

[4] B. Hitchcock, N.-A. Le-Khac, and M. Scanlon, ‘‘Tiered forensic methodology model for digital field triage by non-digital evidence specialists,’’ *Digit. Invest.*, vol. 16, pp. S75–S85, Mar. 2016.

[5] NIST. (2022). *Current RDS Hash Sets*. [Online]. Available: https://www.nist.gov/itl/ssd/software-quality-group/national-softwarereference-library-nsrl/nsrl-download/current-rds.

[6] P. Joseph and J. Norman, ‘‘Forensic corpus data reduction techniques for faster analysis by eliminating tedious files,’’ *Inf. Secur. J. A, Global Perspective*, vol. 28, nos. 4–5, pp. 136–147, 2019.

[7] G. Leurent and T. Peyrin, ‘‘SHA-1 is a shambles: First chosen-prefix collision on SHA-1 and application to the PGP web of trust,’’ in *Proc. 29th USENIX Security. Symp., (USENIX Security)*. Berkeley, CA, USA: USENIX Association, Aug. 2020, pp. 1839–1856.

[8] Z. E. Rasjid, B. Soewito, G. Witjaksono, and E. Abdurachman, ‘‘A review of collisions in cryptographic hash function used in digital forensic tools,’’ *Proc. Comput. Sci.*, vol. 116, pp. 381–392, 2017.

[9] J. Garcia, ‘‘An evaluation of side-information assisted forensic hash matching,’’ in *Proc. IEEE 38th Int. Comput. Softw. Appl. Conf. Workshops*, Jul. 2014, pp. 331–336.

[10] N. C. Rowe, ‘‘Identifying forensically uninteresting files using a large corpus,’’ in *Proc. Int. Conf. Digit. Forensics Cyber Crime*. Cham, Switzerland: Springer, pp. 86–101, 2013.

[11] L. Fang, T. Wu, Y. Qi, Y. Shen, P. Zhang, M. Lin, and X. Dong, ‘‘Improved collision detection of MD5 with additional sufficient conditions,’’ *Electron. Res. Arch.*, vol. 30, no. 6, pp. 2018–2032, 2022.

[12] G. Kessler, ‘‘File signatures table,’’ 2012.

[13] W. Jo, Y. Shin, H. Kim, D. Yoo, D. Kim, C. Kang, J. Jin, J. Oh, B. Na, and T. Shon, ‘‘Digital forensic practices and methodologies for AI speaker ecosystems,’’ *Digit. Invest.*, vol. 29, pp. S80–S93, Jul. 2019.

[14] X. Du and M. Scanlon, ‘‘Methodology for the automated metadata-based classification of incriminating digital forensic artefacts,’’ in *Proc. 14th Int. Conf. Availability, Rel. Secur.*, Aug. 2019, pp. 1–8.

[15] M. A. Wani, A. AlZahrani, and W. A. Bhat, ‘‘File system anti-forensics— Types, techniques and tools,’’ *Comput. Fraud Secur.*, vol. 2020, no. 3, pp. 14–19, Jan. 2020.

[16] A. Scholey and P. B. Zadeh, ‘‘A digital forensics live suspicious activity toolkit to assist investigators with sexual harm prevention order monitoring,’’ in *Proc. IEEE Conf. Dependable Secure Comput. (DSC)*, Jun. 2022, pp. 1–6.
[17] G. Horsman, ‘‘Technical reporting in digital forensics,’’ *J. Forensic Sci.*, vol. 67, no. 6, pp. 2458–2468, Nov. 2022.
[18] H. Arshad, A. Jantan, and E. Omolara, ‘‘Evidence collection and forensics on social networks: Research challenges and directions,’’ *Digit. Invest.*, vol. 28, pp. 126–138, Mar. 2019.
[19] H. Kwon, S. Lee, and D. Jeong, ‘‘User profiling via application usage pattern on digital devices for digital forensics,’’ *Exp. Syst. Appl.*, vol. 168, Apr. 2021, Art. no. 114488.
[20] L. Peng, X. Zhu, and P. Zhang, ‘‘A machine learning-based framework for mobile forensics,’’ in *Proc. IEEE 20th Int. Conf. Commun. Technol. (ICCT)*, Oct. 2020, pp. 1551–1555.
[21] D. E. Salhi, A. Tari, and M. T. Kechadi, ‘‘Using clustering for forensics analysis on Internet of Things,’’ *Int. J. Softw. Sci. Comput. Intell.*, vol. 13, no. 1, pp. 56–71, Jan. 2021.
[22] U. Noor, Z. Anwar, T. Amjad, and K.-K.-R. Choo, ‘‘A machine learning
[23] I. Vayansky and S. A. P. Kumar, ‘‘A review of topic modeling methods,’’ *Inf. Syst.*, vol. 94, Dec. 2020, Art. no. 101582. [24] D. Sun, X. Zhang, K.-K.-R. Choo, L. Hu, and F. Wang, ‘‘NLP-based digital forensic investigation platform for online communications,’’ *Comput. Secur.*, vol. 104, May 2021, Art. no. 102210.
[25] A. Agrawal, W. Fu, and T. Menzies, ‘‘What is wrong with topic modeling? And how to fix it using search-based software engineering,’’ *Inf. Softw. Technol.*, vol. 98, pp. 74–88, Jun. 2018.
[26] X. Luo, ‘‘Efficient English text classification using selected machine learning techniques,’’ *Alexandria Eng. J.*, vol. 60, no. 3, pp. 3401–3409, Jun. 2021.
[27] M. Hina, M. Ali, A. R. Javed, G. Srivastava, T. R. Gadekallu, and Z. Jalil, ‘‘Email classification and forensics analysis using machine learning,’’ in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/IOP/SCI)*, Oct. 2021, pp. 630–635.
[28] A. M. Qadir and A. Varol, ‘‘The role of machine learning in digital forensics,’’ in *Proc. 8th Int. Symp. Digit. Forensics Secur. (ISDFS)*, Jun. 2020, pp. 1–5.
[29] A. Kumar, V. Singh, T. Ali, S. Pal, and J. Singh, ‘‘Empirical evaluation of shallow and deep classifiers for rumor detection,’’ in *Proc. Adv. Comput. Intell. Syst.* Singapore: Springer, pp. 239–252, 2020.
[30] F. Iqbal, M. Debbabi, and B. Fung, ‘‘Artificial intelligence and digital forensics,’’ in *Machine Learning for Authorship Attribution and Cyber Forensics*. Cham, Switzerland: Springer, 2020, pp. 139–150.
[31] L. Bozarth and C. Budak, ‘‘Keyword expansion techniques for mining social movement data on social media,’’ *EPJ Data Sci.*, vol. 11, no. 1, p. 30, May 2022.
[32] Y. Guo, J. Liu, W. Tang, and C. Huang, ‘‘Exsense: Extract sensitive information from unstructured data,’’ *Comput. Secur.*, vol. 102, Mar. 2021, Art. no. 102156.
[33] L. Ignaczak, G. Goldschmidt, C. A. D. Costa, and R. D. R. Righi, ‘‘Text mining in cybersecurity: A systematic literature review,’’ *ACM Comput. Surveys*, vol. 54, no. 7, pp. 1–36, Sep. 2022. VOLUME 11, 2023 3305D. P. Joseph, P. Viswanathan: SDOT
[34] F. B. Rodrigues, W. F. Giozza, R. de Oliveira Albuquerque, and L. J. G. Villalba, ‘‘Natural language processing applied to forensics information extraction with transformers and graph visualization,’’ *IEEE Trans. Computat. Social Syst.*, early access, Apr. 5, 2022, doi: 10.1109/TCSS.2022.3159677.
[35] H. Jo, J. Kim, P. Porras, V. Yegneswaran, and S. Shin, ‘‘GapFinder: Finding inconsistency of security information from unstructured text,’’ *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 86–99, 2021.

[36] D. Gary Miner, J. Elder, and A. R. Nisbet, *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*. Cambridge, U.K.: Academic, 2012.
[37] M. Anandarajan, C. Hill, T. Nolan, F. Dai, and R. Maitra, ‘‘Practical text analytics: Maximizing the value of text data,’’ *Technometrics*, vol. 62, pp. 1–286, Apr. 2020, doi: 10.1080/00401706.2020.1744910.
[38] C. Paulsen and R. Byers, *Glossary of Key Information Security Terms*. Gaithersburg, MD, USA: National Institute of Standards and Technology, Jul. 2019.
[39] X. Wei, C. Zhang, X. Yu, and Z. Zhuo, ‘‘A feature extracting and matching system based on magic-number and AC-algorithm,’’ *Commun. Comput. Inf. Sci.*, vol. 1424, pp. 140–151, Jun. 2021.
[40] M. Paolanti and E. Frontoni, ‘‘Multidisciplinary pattern recognition applications: A review,’’ *Comput. Sci. Rev.*, vol. 37, Aug. 2020, Art. no. 100276.
[41] S. Garfinkel, P. Farrell, V. Roussev, and G. Dinolt, ‘‘Bringing science to digital forensics with standardized forensic corpora,’’ *Digit. Invest.*, vol. 6, pp. S2–S11, Sep. 2009.
[42] G. E. Pibiri and R. Venturini, ‘‘Techniques for inverted index compression,’’ *ACM Comput. Surveys*, vol. 53, no. 6, pp. 1–36, Dec. 2020.
[43] F. Naït-Abdesselam, A. Darwaish, and C. Titouna, ‘‘Malware forensics: Legacy solutions, recent advances, and future challenges,’’ in *Advances in Computing, Informatics, Networking and Cybersecurity*. Cham, Switzerland: Springer, 2022, pp. 685–710.
[44] Q. Bai, Q. Dan, Z. Mu, and M. Yang, ‘‘A systematic review of emoji: Current research and future perspectives,’’ *Frontiers Psychol.*, vol. 10, p. 2221, Oct. 2019.
[45] B. Liu and J. Wang, ‘‘Collocation features in translated texts based on English analogy corpus,’’ *Sci. Program.*, vol. 2022, pp. 1–7, Mar. 2022.
[46] Q. V. Le and T. Mikolov, ‘‘Distributed representations of sentences and documents,’’ in *Proc. 31st Int. Conf. Mach. Learn.*, E. P. Xing and T. Jebara, Eds. Beijing, China, vol. 32, Jan. 2014, pp. 1188–1196.