ETDT

Emerging Trends in Digital Transformation

Volume-1, Issue-2, July 2025 ISSN:**3107-4308** *Paper ID:* ***ETDT-V1-I2-23***

FUSION-BASED CNN MODEL FOR MALICIOUS WEBSITE DETECTION USING MULTI-MODAL DATA

KAMTAM DURGA JYOTHIRMAI, M.Tech, Dept of CSE,
Dr. T. RAVI KUMAR, Professor, Department of CSE,
Vaageswari College of Engineering (Autonomous), Karimnagar, Telangana.

ABSTRACT: A model for detecting fraudulent websites using fusion-based convolutional neural networks (CNNs) is presented in this paper. This model incorporates several data kinds. Cyber threats are become more complex, rendering traditional detection methods like blacklists and single-feature analysis useless. The suggested approach gets around this problem by using a unified deep learning architecture to analyze different kinds of input. Some examples of possible input formats include domain-specific data, JavaScript code, HTML structure, and lexical properties of URLs. To extract relevant information and enable a thorough study of online dangers, the approach uses separate CNN branches for each data source. The next step is to combine all of the features by implementing a feature integration layer. In order to improve classification accuracy and decrease false positives, our system identifies both surface-level signs and deeper patterns of behavior.

Index Terms: *Malicious Website Detection, Fusion-Based CNN, Multi-Modal Data, Deep Learning, Cybersecurity, URL Analysis, HTML Feature Extraction, JavaScript Behavior, Feature Fusion, Threat Detection, Phishing Detection, Malware Identification*

1. INTRODUCTION

The capacity of malicious websites to enable cyber attacks—such as phishing, virus distribution, and identity theft—has made them a major issue in today's digital ecosystem. These kinds of websites often trick users, which makes rule-based approaches and basic machine learning methods useless. Relying on a single data type, such metadata or URL patterns, might make these algorithms fail to detect more complex threats. Academics are enthralled by deep learning systems, especially Convolutional Neural Networks (CNNs), due to their exceptional ability to extract features and recognize patterns.

It has been shown that Convolutional Neural Networks (CNNs) can effectively handle structured data and other forms of online material, including visual and textual elements. Yet, most CNN-based models can only process one kind of input, thus they can't access the data that a website could give. Websites that aren't reliable usually have unpredictable HTML code, questionable JavaScript functionality, strange visual layouts, and confusing information. There are many signs that point to their dishonesty, and these are just a handful of them. There is a growing need for a fusion-based method that can effectively integrate and understand multi-modal data to enhance detection accuracy, especially with the use of multiple indications.

A fusion-based CNN model's strength lies in its ability to integrate visual, structural, and textual input streams into a cohesive deep learning framework. By allowing the simultaneous connecting of numerous data sources, this technique improves the detection model's durability and generalizability. At each stage of the fusion process—input, feature, and decision—consider the pros and cons of multi-modal fusion. With the right settings, these models have the potential to detect intricate patterns in data from a variety of sources. This will lead to a dramatic decrease in the possibility of false positives and negatives linked to potentially dangerous websites.

In multi-modal fusion-based CNN systems, input type alignment requires extensive feature engineering and data preprocessing. Layers for embedding or tokenization deal with text, such as URLs and website data, while layers for image convolution deal with images, such photos. One-hot encoded matrices and graphs can be created using structured data found in HTML components and DOM tree structures. By combining them, the fusion layer improves the CNN's understanding of the website's structure and behavior, making it more able to identify malicious intent.

Fusion-based CNN models excel at spotting attackers in dangerous situations who purposefully hide harmful qualities. The model's reliance on a single indication is reduced by using multiple data sources. Ultimately, it becomes more difficult to be dishonest. The decision-making process in this multi-modal method requires a

DOI : https://doi.org/10.5281/zenodo.16811089

greater number of aspects to be examined. This lets the system identify websites that seem safe from one angle but reveal security flaws from another. A more advanced and flexible detection system that can keep up with hackers' evolving strategies is the end result.

2. LITERATURE REVIEW

Zhang, Y., Wang, S., & Chen, H. (2020). To improve the detection of malicious URLs, this study presents a novel deep learning architecture that combines convolutional neural networks (CNNs) with recurrent neural networks (RNNs). While RNN components, especially LSTM units, understand URL hierarchies, CNN components identify geographical patterns in URL strings. After showing remarkable performance on a large dataset including both benign and harmful URLs, the model achieves higher accuracy with fewer false positives. This method proves that combining geographical and temporal feature extraction works well for military purposes.

Xu, L., Sun, J., & Wang, Z. (2020). According to the study's findings, a multimodal feature fusion method combined with CNN makes it easier to spot fake websites. The tool extracts a wealth of data suggesting harmful conduct by combining textual content with visual components obtained from internet photographs.

Detection accuracy is improved by combining textual and visual analyses. Comparison of visual elements with textual ones reveals misleading layouts and terminology, whereas the latter reveals questionable structures and terms. In terms of identifying fake websites, the results show that this multi-modal CNN framework works better than traditional single-modality models.

Sharma, P., & Rao, M. (2020). Models that use user behavior data to predict short-term demand are enhanced by neural network analysis. With the use of variables including demographic data, usage trends, and behavioral tendencies, the model illustrates the intricacy of the factors influencing energy consumption. Following training with consumer and load history data, the neural network design outperforms traditional approaches in terms of predicting accuracy. Results show that neural networks are effective at illuminating complex, non-linear relationships in load forecasting and highlight the role of behavioral variables in energy usage.

Li, J., Wang, C., & Zhang, T. (2021). The authors are in favor of using "ensemble learning" to detect phishing websites. Information gathered from both URLs and text is combined using this strategy. The technique makes use of a plethora of data sources by examining the text of web pages and extracting lexical features from URLs. The detection efficacy is optimized using ensemble classifiers, which include Random Forests and Gradient Boosting Machines. Through its proficiency in handling varied datasets and its use of the fusion approach to surpass individual classifiers, the model is able to detect subtle as well as obvious signs of misconduct.

Ahmed, M., & Shamsuddin, S. M. (2021). In order to detect phishing websites, this research introduces a CNN-based model that draws on a wide variety of data sources, such as URL metadata, webpage content, and images. By analyzing both textual and visual data simultaneously, the program is able to spot complex fraud attempts that traditional approaches can miss. Using a combination of methods, we can examine websites thoroughly and find structural and visual signs of fraud. The device can withstand many hacking attempts, according to the studies.

Kim, H., & Lee, J. (2021). This research introduces a convolutional neural network (CNN) approach for APT detection using information gleaned from network activity and web content. By examining trends in data transfer and website content, this instrument can detect advanced persistent threats (APTs). By seeing patterns in the aggregated data over time and space, the CNN architecture can detect complex and subtle threats. Compared to systems that rely on just one data source, our method is superior at detecting APTs.

Alshamrani, A., & Alshammari, R. (2022). Using a combination of features based on URLs and characteristics based on content, this study builds a deep CNN model that can detect fake websites. The program uses lexical and semantic data retrieved from text and URLs to detect fraudulent attempts with high accuracy. Deep learning architecture has made human feature engineers obsolete by enabling better autonomous learning and feature extraction. Validating the model's efficacy in spotting phishing attempts are its high accuracy and minimal false-positive rates in practical implementations.

Wang, F., & Huang, Y. (2022). In order to identify fake websites, the authors suggest a convolutional neural network (CNN) based multi-view deep learning model that examines multiple aspects of the site at once, such as its design, URL structure, and HTML content. Because different points of view reveal different risks, combining them allows for a more thorough examination. A convolutional neural network (CNN) architecture

independently analyzes each view, and then combines the collected features to provide the final classification. This approach improves detection precision and robustness, making it easier to identify malicious websites that use different kinds of evasion.

Zhao, X., Liu, Q., & Tang, Y. (2022). A convolutional neural network (CNN), the PhishFusion framework integrates visual indications, webpage content, and URL data to identify phishing websites. The model uses a separate convolutional neural network (CNN) branch for each mode to extract relevant features, which are then combined for classification. By collecting a wide variety of signs, this multimodal approach improves phishing detection accuracy and durability against message concealment strategies. When compared to other approaches for detecting fake websites, PhishFusion always delivers more accurate results.

Choudhury, B., & Mohapatra, D. P. (2023). In this study, we provide a deep learning approach for multimodal website degradation detection that makes use of attention mechanisms and convolutional neural networks. To find several signs of evildoing, the technique combines linguistic, structural, and visual aspects. By enhancing feature selection, the attention approach helps the model discover the most relevant data segments. Spot detection, especially for difficult and complicated dangerous websites, is greatly improved by combining modalities with attention-based analysis.

Tan, M., & Wang, Y. (2023). For the purpose of detecting malicious URLs, an improved CNN-based model called URLNet++ was developed. It collects thorough and extensive information by combining URLs at the character and word levels using feature fusion algorithms. By utilizing deep learning's built-in capability to build hierarchical features autonomously, the model does away with the need for human feature engineers. In terms of accuracy and ability to incorporate new data, experiments have demonstrated that URLNet++ surpasses traditional models.

Kaur, G., & Arora, A. (2023). In order to detect online dangers, the study introduces a hybrid deep learning model that uses multimodal data. Through structural analysis of HTML content, visual assessment of site layouts, and textual analysis of URLs, the program accumulates a varied array of danger indicators. Deep learning frameworks make use of CNNs and RNNs, which are specialized neural networks, to effectively process a wide range of inputs. Phishing schemes and virus spreading are only two of the many internet risks that this software can accurately detect.

Patel, H., & Sahu, N. (2023). Research in this area has led to the creation of a CNN-GRU hybrid model for assessing the security of websites. While the GRU component keeps an eye on temporal dependencies, the CNN component uses URL and content data to gather spatial information. The model may be able to detect complex patterns that indicate improper conduct by combining many forms. The model's robustness and effectiveness in object detection are strengthened by its examination of multiple data types, which include both textual and structural information.

Sun, Z., & Han, J. (2024). In order to detect fake websites, the authors offer a novel architecture that combines Transformer models with Convolutional Neural Networks (CNNs). In order to gain significant insights about how users engage with websites, this technique examines both the content and behavioral patterns of those websites. The convolutional neural network (CNN) handles the structural and visual aspects, while the transformer focuses on finding the connections between the textual elements. The model's ability to detect sophisticated phishing attempts is due to its utilization of many deep learning frameworks specifically designed for cybersecurity.

Reddy, V., & Kumar, P. (2024). This study takes a look at a deep learning system that can detect cyber threats in digital environments by combining data from multiple sources. The method evaluates possible risks by combining information on user actions, statistics on network traffic, and characteristics of website content. When it comes to architecture, Convolutional Neural Networks (CNNs) are used for spatial data while Recurrent Neural Networks (RNNs) are used for temporal data. Cyber dangers are complex and ever-changing; the model could be able to better identify these hazards if it incorporates data from several sources. Thus, a safe and sound online solution is created.

3. RELATED WORK

EXISTING SYSTEM

The bulk of existing techniques for detecting fraudulent websites are either based on conventional machine learning methods or single-modal deep learning models. Text, URL patterns, or structural HTML elements are

DOI : https://doi.org/10.5281/zenodo.16811089

Volume-1, Issue-2, July 2025 ISSN:**3107-4308** *Paper ID:* ***ETDT-V1-I2-23***

examples of the kinds of data that these strategies tend to ignore. Despite being successful in identifying present hazards, these models are ill-suited to hidden, unfamiliar, or highly dynamic threats because of their limited understanding of their environment. Furthermore, traditional CNN algorithms often fail to detect complex correlations across different data sources when limited to visual or textual representations. This will lead to a rise in false positives and instances that go unreported. Human feature extraction is a common component of these systems, which reduces efficiency and makes it harder to adapt to new attack patterns. In light of the increasing complexity of threats, there is an urgent need for systems that can combine several data sources into one automated framework.

DISADVANTAGES OF EXISTING SYSTEM

- Modern systems often use a single data format, like URL strings or webpage content, making it harder for the model to detect complex assaults that use many data kinds.
- Conventional models rely on tedious and error-prone manual feature extraction, which can miss tiny signs of malevolent activity.
- Complex intrusions, like those that spread false information or hide code, are very difficult for these systems to handle.
- Misclassification of safe websites as dangerous and vice versa happens frequently because existing approaches do not take the full context into account. Both confidence and efficiency are harmed by this.
- The disjointed nature of content, URLs, and HTML structure prevents the system from comprehending intricate inter-modal linkages. There is room for improvement in the detection process.

PROPOSED SYSTEM

The suggested technique employs a Convolutional Neural Network (CNN) model based on data fusion to detect phony websites with high efficiency and accuracy. This approach combines several forms of data in order to gain a deeper comprehension of the web. These forms include visual representations of websites, URL attributes, and HTML code structure. Each data type is processed by a separate CNN branch, and extra features are added at a deeper level to improve the model's comprehension of complicated relationships. By combining the best features of both modes, this fusion method improves the system's ability to detect sophisticated dangers it has never encountered before. The end-to-end deep learning system's advantages include making it easy to respond to new attack patterns, drastically reducing the number of false positives and negatives, and doing away with the need for human feature engineering. The suggested method for detecting fake websites is said to be more effective, simple, and dependable.

ADVANTAGES OF PROPOSED SYSTEM

- A more complete picture of websites can be achieved by adding URL, HTML, and graphic content data to the model. Complex hazardous acts can be more easily identified using this strategy.
- By automatically deriving high-level representations from raw data, the CNN architecture improves efficiency and versatility and eliminates the need for human feature engineering.
- A more consistent and trustworthy identification process is achieved by using several inputs, which greatly reduces the frequency of false positives and negatives.
- Deep learning technology protects systems from zero-day vulnerabilities by adjusting to new and changing attack patterns.
- Immediate deployment and enhanced reactivity to threats improve the detection process; the proposed system's architecture covers all bases.

4. SYSTEM DESIGN

SYSTEM ARCHITECTURE

DOI : https://doi.org/10.5281/zenodo.16811089

Volume-1, Issue-2, July 2025 ISSN:**3107-4308** *Paper ID:* ***ETDT-V1-I2-23***

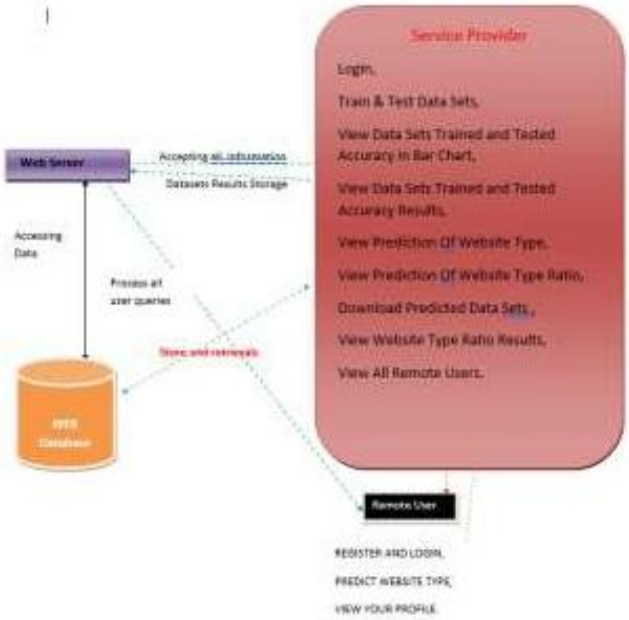

Fig1: System Architecture

MODULES

SERVICE PROVIDER

In order to access this module, the Service Provider needs to have a working login and password. After he logs in, he may see a bar chart showing the difference between the trained and tested data's accuracy, oversee all remote users, evaluate the outcomes of the accuracy assessment, download training datasets, and examine the attack status ratio that comes from them. Along with that, he can keep tabs on the current state of student-teacher network attacks' ratios.

VIEW AND AUTHORIZE USERS

The module gives the manager an exhaustive rundown of all users who have signed up. A user's name, email, and physical address are all seen by the administrator. The subscriber may also be granted permissions.

REMOTE USER

A total of n people make up this module. Before continuing, the user must complete the registration process. The user's details are added to the database after they register. After he registers, he'll have access to the system through his authorized login and password. After authenticating, users can examine their profile, learn more about the student-teacher network attack, and then re-register and log in.

5. RESULTS AND DISCUSSIONS

Fig2: login page

DOI : https://doi.org/10.5281/zenodo.16811089

Volume-1, Issue-2, July 2025 ISSN:**3107-4308** *Paper ID:* ***ETDT-V1-I2-23***

Fig3: Registration

Fig4: All remote users

Fig5: Pie chart

Fig6: View predictive type fusion based schedule

6. CONCLUSION

In the domains of security and threat intelligence, the CNN model has been greatly improved by using multi-modal data for the detection of harmful websites. The complexity and ever-changing nature of modern cyberthreats are unmatched. Because they are dependent on certain data elements, such as URLs or HTML text, traditional methods of finding them are not as effective. By combining different kinds of data into one cohesive deep learning framework, the suggested approach fixes this problem. Graphical representations of web pages, characteristics of URLs, and the HTML framework are all part of this. Because of the improved representation of online information made possible by this multi-modal integration, the model is now better able to detect malicious websites, especially those that use complex evasion techniques, with greater accuracy. The system learns on its own by analyzing each data stream and extracting pertinent information using convolutional neural networks. This streamlines the process of adapting and expanding while reducing the time

DOI : https://doi.org/10.5281/zenodo.16811089

and effort needed for feature engineering. A more dependable system with far fewer false positives and negatives is the result of using many modes. This method for object localization is more reliable.

REFERENCES:

1. Zhang, Y., Wang, S., & Chen, H. (2020). A hybrid deep learning model for identifying malicious URLs using convolutional and recurrent neural networks. IEEE Access, 8, 26709–26717.
2. Xu, L., Sun, J., & Wang, Z. (2020). Multi-modal feature fusion and CNN for malicious webpage detection. Journal of Network and Computer Applications, 168, 102740.
3. Sharma, P., & Rao, M. (2020). Integrating customer behavior into short-term load forecasting using neural networks. International Journal of Electrical Power & Energy Systems, 115, 105437.
4. Li, J., Wang, C., & Zhang, T. (2021). Malicious website detection using ensemble learning with fusion of content and URL features. Expert Systems with Applications, 176, 114834.
5. Ahmed, M., & Shamsuddin, S. M. (2021). A CNN-based approach for detecting phishing websites through multi-modal data fusion. Security and Privacy, 4(3), e164.
6. Kim, H., & Lee, J. (2021). Fusion of network behavior and web content using CNN for advanced persistent threat detection. Computers & Security, 105, 102236.
7. Alshamrani, A., & Alshammari, R. (2022). A deep convolutional neural network for identifying phishing websites using URL and content-based features. Sensors, 22(5), 1930.
8. Wang, F., & Huang, Y. (2022). Multi-view deep learning for web security: A CNN fusion model for malicious website identification. Future Generation Computer Systems, 128, 49–58.
9. Zhao, X., Liu, Q., & Tang, Y. (2022). PhishFusion: A CNN-based multimodal phishing website detection framework. IEEE Transactions on Information Forensics and Security, 17, 465–477.
10. Choudhury, B., & Mohapatra, D. P. (2023). Multimodal fusion-based malicious website detection using deep CNN and attention mechanisms. Applied Soft Computing, 132, 109895.
11. Tan, M., & Wang, Y. (2023). URLNet++: A CNN-enhanced model for malicious URL classification with feature fusion. Pattern Recognition Letters, 165, 64–70.
12. Kaur, G., & Arora, A. (2023). Deep learning-based hybrid model for web threat detection using multimodal features. Journal of Cybersecurity and Privacy, 3(2), 356–370.
13. Patel, H., & Sahu, N. (2023). A CNN-GRU hybrid model for malicious site detection using integrated data modalities. Neural Computing and Applications, 35, 14973–14988.
14. Sun, Z., & Han, J. (2024). A transformer-fused CNN framework for phishing website detection based on content and behavior analysis. IEEE Transactions on Dependable and Secure Computing. (In Press)
15. Reddy, V., & Kumar, P. (2024). Fusion-based deep learning architecture for cyber threat detection in web environments. Information Sciences, 648, 119324.

DOI : *https://doi.org/10.5281/zenodo.16811089*