

## Combining Convolutional Neural Networks and XGBoost for Phishing URL Classification

Muhammad Narendra Bagus Nurseto<sup>1</sup>, Fauzi Adi Rafrastara<sup>2</sup>, Wildanil Ghozi<sup>3</sup>,  
Ramadhan Rakhmat Sani<sup>4</sup>, Ajib Susanto<sup>5</sup>, Emad Hamadeh<sup>6</sup>

<sup>1,2,3,4,5</sup>Faculty of Computer Science, Universitas Dian Nuswantoro  
Semarang, Indonesia

<sup>6</sup>Department of Computer Engineering, Al-Quds University,  
Palestine

[111202314983@mhs.dinus.ac.id](mailto:111202314983@mhs.dinus.ac.id)<sup>1</sup>, [fauziadi@dsn.dinus.ac.id](mailto:fauziadi@dsn.dinus.ac.id)<sup>2</sup>, [wildanil.ghozi@dsn.dinus.ac.id](mailto:wildanil.ghozi@dsn.dinus.ac.id)<sup>3</sup>,  
[ramadhan\\_rs@dsn.dinus.ac.id](mailto:ramadhan_rs@dsn.dinus.ac.id)<sup>4</sup>, [ajib.susanto@dsn.dinus.ac.id](mailto:ajib.susanto@dsn.dinus.ac.id)<sup>5</sup>, [ehamadeh@staff.alquds.edu](mailto:ehamadeh@staff.alquds.edu)<sup>6</sup>

### Article Info

#### Article history:

Received 2026-09-04

Revised 2026-09-22

Accepted 2026-10-02

#### Keyword:

Convolutional Neural Network,  
LightGBM,  
Phishing Detection,  
Spatial RGB,  
Tree-Based Ensemble.

### ABSTRACT

Cyber threat detection failures stem from the complexity of psychological exploitation tactics leveraging target anxiety and the limitations of traditional models in analyzing non-linear correlations within the structural features of malicious web URLs. To improve classification performance and computational efficiency in feature engineering, a sequential hybrid framework is developed by combining the spatial feature extraction capabilities of RGB Color Mapping image tensors via a Convolutional Neural Network (CNN) with tree-based boosting classifiers (XGBoost and LightGBM). Experiments are executed using a public dataset containing 11,055 observations, applying an 80:10:10 data splitting schema and isolated MinMaxScaler normalization to eliminate data leakage risks. The evaluation results demonstrate that the hybrid CNN + XGBoost model achieves the highest performance, with a validation accuracy of 97.83% and an F1-score of 98.06%, outperforming the replicated State-of-the-Art Random Forest baseline accuracy (97.38%). The metric stability confirms that the proposed architecture possesses robust generalization capabilities for implementation as a near real-time cyber defense system.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

### I. INTRODUCTION

Exploitation in cybersecurity is designed to manipulate targets into handing over personal data, which is technically classified as a phishing attack [1], [2]. This action is usually carried out through the distribution of manipulative URL links or emails posing as legal entities [1],[3], where scammers impersonate financial institutions or reputable organizations to harvest credentials without the victim's consent [2]. These attackers proactively engineering deceptive websites and register malicious domains by manipulating SSL certificates or hosting history to visually convince unwary users [4]. The impact is that it can trigger macro-scale financial destruction with a projected global loss surpassing the \$10.5 trillion mark in 2025 [5], [6], inducing severe economic damage and posing a direct threat to

corporate and personal data integrity [2]. The threat is validated by the Anti-Phishing Working Group report, which has recorded nearly 1 million incidents in the fourth quarter of 2024 [7].

The increase in digital attacks is tangibly recorded through the high volume of manipulative intrusions on a global scale. The Anti-Phishing Working Group (APWG) report noted a cumulative total of 3.8 million phishing attacks throughout 2025 [8]. This exploitation figure is validated to have experienced an increase of 13.8 percent in the first quarter of 2026. The surge in the quantity of this threat was engineered by attackers with a total of 971,181 attacks detected in just a three-month period [9].

The level of exploitation in cybersecurity has been specifically documented through incident reports of fake websites monitored from mid-2025 to early 2026. The peak

of intrusions in the third quarter of 2025 was recorded in July with 351,590 attacks, later shifting to February 2026 with a total of 329,625 manipulative websites. The scope of attacks indicates a target transition, where the initial dominance of targets on webmail service providers (21.2 percent) was replaced by the telecommunications sector with a proportion of 33 percent. These manipulative acts are continuously engineered to exploit legal entities by forging between 311 and 475 unique institutional brands every month [8], [10].

The vulnerability of digital infrastructure to manipulation is evident through various hacking incidents on national-scale institutions. It is documented in a financial sector fraud case that replicated the official Bank Central Asia (BCA) portal to hack customer data. In camouflage patterns, a similar case was found in the government sector through the forgery of the Directorate General of State Assets (DJKN) website to visually deceive targets [11]. These incidents have confirmed that these attackers are more focused on intruding into domains that have a higher level of public trust [11]. Furthermore, academic environments and universities have emerged as highly vulnerable targets for phishing vectors, primarily driven by remote learning transitions and a general lack of institutional cybersecurity awareness among students [12].

Black-box architecture becomes a traditional threat that is unable to rationalize the reasons behind the determination of an anomaly. An ensemble learning architecture based on Random Forest and XGBoost is used as a technical resolution, which is proposed alongside Explainable AI (SHAP) technology [13], [14]. The algorithm design is used to detect malicious links and simultaneously serves as a visual weight mapping for each triggering feature [14]. The detection output is automatically extracted into a structured threat intelligence archive to meet security standards and governance [15].

## II. RESEARCH METHOD

This research proposes a hybrid-based phishing website detection framework with a combination of the spatial feature extraction power from images by a Convolutional Neural Network (CNN) and the speed and classification sharpness of Tree-based Ensemble algorithms (XGBoost and LightGBM) [14]. This approach is designed to overcome the limitations of traditional models in recognizing non-linear correlations between the security features of site drafts [14], [16]. All stages in the experiment are integrated into a systematic pipeline system to ensure testing fairness, prevent data leakage. The following is the comprehensive methodological flow in this research illustrated through architectural visualization in Figure 1.

### A. Dataset Preparation

A public dataset containing 11,055 website observations is used as the basic material, encompassing data of 30 technical features and 1 binary label [1], [17]. The comprehensive statistical layout, class attributes, and initial data dimensions

of this repository are explicitly presented in Table 1. Data splitting is locked at an 80:10:10 ratio; in this scheme, it fully separates the training, validation, and testing data.

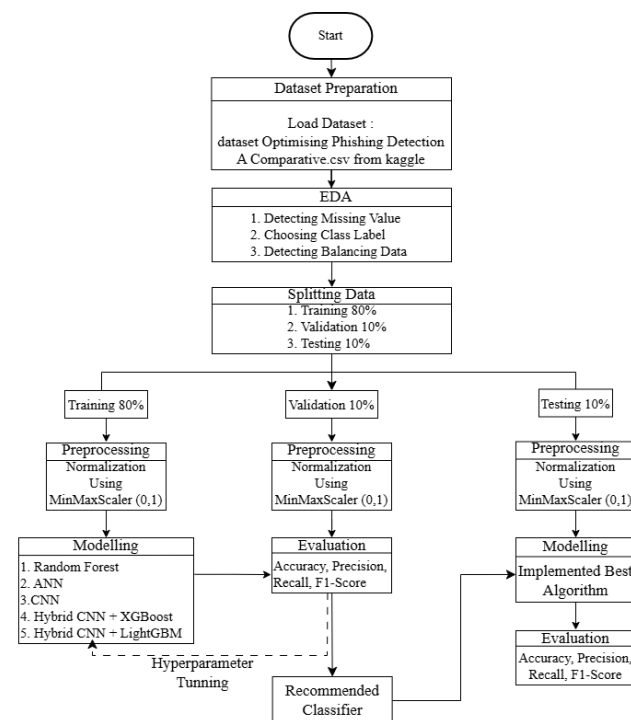


Figure 1. Pipeline Research Method

TABLE I  
DETAILS OF PHISHING WEBSITE DATASET

| Dataset Name       | Phishing website dataset |
|--------------------|--------------------------|
| Number of Rows     | 11.055                   |
| Number of Features | 30                       |
| Number of Classes  | 2 (Legitimate, Phishing) |
| Missing Values     | -                        |

### B. Exploratory Data Analysis (EDA)

Missing value detection is carried out for data quality inspection in the initial stage on each table column because real-world datasets are highly prone to incomplete records that significantly degrade model classification capability [18]. This exhaustive verification confirms that structured observation matrices must remain uncorrupted before feature scaling objects map the data vectors [18]. The results show that all rows are fully populated without any missing or defective data. The target class label is identified through the 'Result' column which contains binary numeric symbols to distinguish cyber objects. Evaluation of the data proportion shows a balanced distribution between phishing samples amounting to 4,898 rows and legitimate samples amounting to 6,145 rows [1],[17].

### C. Splitting Data

The controlled random data splitting using the stratified sampling method divides the data into a ratio of 80% for training, 10% validation, and 10% testing to maintain class distribution consistency, intentionally differentiating from the first SOTA research which only utilized an 80:20 scheme [1]. The largest allocation of 80% focuses on the training phase to ensure the model's internal weights learn attack patterns optimally in complete isolation. Training data isolation is absolutely necessary to prevent the data leakage phenomenon and allow abstract characteristics of cyber indicators to be separated in depth.

The remaining partitions allocate 10% for validation and 10% for testing as separate, independent partitions. The validation partition serves as an objective guide to monitor overfitting and structurally guides weight adjustments to maintain high stability across different network layers without being influenced by outside data. Meanwhile, the 10% testing partition acts as a real simulation to test the model's performance against unseen objects, ensuring the scientific validity of the model's final accuracy and determining the overall success of the proposed method innovation [19].

1) *Preprocessing* : The application of scale normalization using the MinMaxScaler method in the training phase data transformation. All numerical values in the features are repositioned into a new value range between the numbers 0 to 1 [20]. In the mathematical formula, the Scaler is calculated purely based on the minimum and maximum values that have been found in this training data [20],[21]. This stage equalizes the feature unit values to prevent the dominance of certain variables during the calculation of the model gradient [20].

The validation data is adjusted mathematically using parameters that have been obtained from the previous training phase. The transformation process is prohibited to recalculate new minimum and maximum values from the internal validation data itself. The Scaler scheme utilizes the MinMaxScaler object trained on the 80% main data partition. This approach establishes a consistent feature space mapping that synchronizes multiple sub-datasets without inducing dimensional distribution drifts [22]. This approach can maintain the consistency of the feature space so that it remains in the same mathematical range [21].

The equalization of testing data scale by following standard operating procedures that use the training transformation object. Data values will be converted to a binary range that uses the function of the scaler formed initially. No parameters are calculated from the internal characteristics of the testing data during the ongoing process. This processing stage can guarantee that the model evaluates new data without any outside intervention [21].

2) *Modelling* : The experiment that evaluates the performance of 5 (five) model architectures, namely Random Forest, ANN, CNN, and 2 (two) Hybrid variants. In the hybrid

model, it is built with a combination of visual feature extractors based on CNN with advanced decision tree classification [23]. This sequential framework optimizes classification capability by transforming structural parameters into highly descriptive visual-based feature mappings [24]. The working mechanism of the Hybrid model runs sequentially, not simultaneously [23]. Tabular data is rearranged into a two-dimensional spatial matrix, then passed to the CNN convolution layer to extract abstract features [23]. In the feature vector resulting from the cut of the deepest CNN layer, it will be fed as a single input for the XGBoost and LightGBM algorithms [14].

The architectural mechanism of the sequential hybrid framework utilizes a Convolutional Neural Network (CNN) acting purely as a deep feature extractor to filter complex non-linear parameters from transformed image tensors. Upon stripping the final classification layer of the network, the extracted dense abstract vector blocks are systematically transferred as a single cohesive input array. Finally, the Extreme Gradient Boosting (XGBoost) algorithm executes its gradient tree-boosting decision boundaries on these refined abstract features to produce the ultimate biner threat classification [25].

The sequential hybrid framework operates in two distinct phases. In the first phase, the Convolutional Neural Network (CNN) acts as a deep spatial feature extractor. The 2D image tensors are passed through multiple convolutional layers equipped with ReLU activation and max-pooling layers to capture local spatial hierarchies and non-linear feature correlations. Prior to the final classification, the multi-dimensional feature maps are flattened into a one-dimensional dense vector. In the second phase, this high-level abstract vector is systematically exported and formatted as a DMatrix input for the XGBoost algorithm. XGBoost then constructs an ensemble of decision trees using gradient boosting mechanics on these extracted deep features, effectively separating the complex spatial representations into the final binary classification (legitimate or phishing). This decoupling allows the CNN to handle representation learning while XGBoost handles the robust, non-linear decision boundary mapping.

The second hybrid variant implemented in this study establishes a sequential framework leveraging a Convolutional Neural Network (CNN) paired with a LightGBM final classifier. Under this operational pipeline, the layered CNN filters function strictly as a spatial representation extractor to condense high-dimensional structural embeddings into low-latency dense vectors. These refined deep parameters are subsequently routed into the LightGBM algorithm, which utilizes efficient leaf-wise tree splitting mechanics to compute the final binary cybersecurity decisions, expanding the collaborative potential of combined deep learning and statistical ensemble architectures [26].

The selected model architecture in the Recommended Classifier stage is loaded to run the testing data. The hybrid model can process the spatial image tensor from the remaining 10% website samples. There is no retraining

process or parameter weight changes in this computational phase. The recommended classification works by making predictions based on knowledge patterns that have been previously locked.

**Hyperparameter Tuning :** Structural optimization is designed through a parameter adjustment mechanism that uses the GridSearchCV algorithm. In this scheme, it is run automatically if the accuracy on the training data is much higher than the validation data. Parameter combinations such as learning rate (Learning Rate) and tree depth (Max Depth) are tested repeatedly [14]. Through parameter simulation, the prediction error rate of the hybrid classification can be suppressed to the lowest point.

**Recommended Classifier :** The determination of the best model based on the acquisition of the highest F1-Score value during the validation data testing. The model that occupies the top rank will be selected as the main candidate positioned for the recommended classification. This selection decision ignores the researcher's subjectivity because it is based on quantitative evaluation metrics [20]. This optimization strategy targets maximum precision balance to deliver customizable and resilient detection layers [24].

The output of the validation data testing produces one best architecture, which is the one having the most stable performance. The Hybrid algorithm based on LightGBM or XGBoost that is scanned automatically so that it occupies the recommended classification position. Evaluation at this stage locks all internal weights as well as the hyperparameter configuration of the selected model. The recommended classification will be declared ready to be sent to the final testing phase. This selected model will later be considered the most feasible to represent the cyber threat detection system.

**Evaluation :** Testing the model's performance on the training data is measured directly using the validation data. The evaluation parameters used are such as accuracy score level, precision, recall value, and F1-Score weight. Prediction results are recorded systematically to find out the most flexible baseline model. This stage can provide an objective picture regarding the initial generalization capability of the designed convolutional and Hybrid architectures [20].

The evaluation method on the validation data which becomes the main tool in the SOTA comparison assessment. The baseline Random Forest and ANN models trained using PCA dimensionality reduction and RFE feature elimination in this space. These stages are taken to create a fair comparison on the superiority claims of previous research algorithms. All performances of the conventional architecture and the proposed architecture are tested under the same data conditions [1].

The final phase in this methodology is by conducting calculations on the testing data prediction performance. The accuracy, precision, recall, and F1-Score values are recalculated which are used as the final result of the research. The evaluation metrics obtained at this stage become the validity of the designed hybrid model [20]. Those quantitative

numbers will be juxtaposed directly against the performance record of the SOTA method [1].

### III. RESULTS AND DISCUSSION

The initial cleaning stage by removing the index feature produces a clean data structure with dimensions of  $11,055 \times 31$  ready to be processed to the modeling stage. The locking of the 80:10:10 data splitting ratio successfully mitigated the risk of information contamination between script subsets from the beginning of the experiment. The truncation of this non-predictive column is proven to simplify the computational matrix without reducing the important variance value of cyber data.

The data quality investigation confirms a missing value rate of 0% across all database columns. The distribution evaluation shows that the dataset has adequate balance with a distribution of 4,898 phishing sample rows and 6,157 legitimate sample rows. This proportion balance frees the testing from majority class bias so that the calculation of the final accuracy value is objective.

The application of stratified sampling successfully maintained an identical percentage of the target class in each training, validation, and testing partition. The locking of this three-independent-part split provides a stricter evaluation foundation when compared to the first SOTA script which ignored a separate validation phase. The existence of this 10% independent partition is proven to be a sensitive validation instrument in detecting overfitting symptoms.

The model testing process on the 80% training partition shows a learning curve that varies depending on the type of algorithm architecture. Cyber threat typing patterns can be absorbed deeply by the internal model due to the large volume of training data allocation provided. The weighting visualization results show that this isolated training script is able to transfer solid basic knowledge to all testing models.

The testing phase using the 10% validation data acts as a filtering instrument that shows the comparative strength map of the script. The performance record in this isolated space confirms the stability of the developed architecture against external data. Testing on this subset successfully detected the upper limit of the model's generalization capability before the final script was executed.

Testing on the 10% blind test data (unseen data) provides final proof regarding the robustness of the proposed hybrid model. The model's ability to face new attack pattern variations is radically tested in this operational phase. Strict isolation of the testing data guarantees that the generated metrics are free from internal optimization manipulation.

The application of the MinMaxScaler function to the training data successfully unified the entire feature domain into a binary mathematical range of  $[0, 1]$ . This scale equalization is proven to accelerate the gradient descent convergence in models based on artificial neural networks. This isolated normalization scheme also successfully eliminated the absolute number dominance of the URL length parameter.

The application of the training scaler parameters to the validation data is proven to maintain the consistency of the feature value distribution without causing data leakage symptoms. This transformation converts the discrete logic values of -1, 0, and 1 into normalization scales of 0.0, 0.5, and 1.0 precisely. This scheme successfully maintained the purity of the validation test data from outside internal parameter interventions.

The scale equalization of the testing data purely utilizes the training scaler object reference without involving new parameter calculations. This locking step successfully maintained testing objectivity in a real-world simulation environment. Feature conversion values are mapped consistently towards a two-dimensional image matrix representation.

The execution of modeling on the training data produces performance variations that separate the superiority of the hybrid model from conventional models. The hybrid algorithm based on spatial-RGB convolution successfully reduced data dimensions into 384 new abstract features efficiently. This relay operation provides very clean data input for the final decision-making algorithm.

The implementation of the best CNN + XGBoost architecture on the testing subset successfully executed predictions responsively. The model processes spatial image tensors from the remaining 10% website samples without experiencing computational latency constraints. The locking of parameter weights since the training phase ensures no structural modifications in this operational stage.

Parameter optimization simulation through the GridSearchCV mechanism over a period of 2.3 hours yielded an accuracy value of 97.20%. Although these results are considered high, the accuracy obtained is still below the hybrid model with default standard parameters. This finding proves the existence of an overfitting phenomenon in cross-validation due to overly aggressive parameter tuning.

The calculation of initial performance metrics on the training subset shows that the hybrid model recorded superiority across all test parameters. Empirical data shows high performance in the proposed model group when compared to traditional architectures. The recording of values is carried out periodically to map the comparison graph between training and validation performance.

The final calculation on the testing data confirms the achievement of a consistent accuracy value at the 97.47% level and a precision value of 97.57%. The capture performance of dangerous objects is strengthened by the acquisition of a recall rate of 97.89% and an f1-score weight of 97.73%. The stability of these metric numbers that remain high proves the superiority of spatial-RGB analysis in breaking the performance dominance of conventional SOTA models.

TABLE II  
RESEARCH PERFORMANCE

| No. | Method | Acc | Precision | Recall | F1-Score |
|-----|--------|-----|-----------|--------|----------|
|-----|--------|-----|-----------|--------|----------|

|    |                 |               |               |               |               |
|----|-----------------|---------------|---------------|---------------|---------------|
| 1. | Proposed Method | <b>0.9747</b> | <b>0.9757</b> | 0.9789        | <b>0.9773</b> |
| 2. | RF Base         | 0.9738        | 0.9650        | <b>0.9886</b> | 0.9767        |

Given the marginal accuracy gap between the proposed model and the baseline, evaluating performance solely on accuracy is insufficient. Instead, the assessment prioritizes the F1-Score and Precision metrics, which are more critical for real-world cybersecurity applications. The proposed CNN + XGBoost model achieves a superior F1-Score of 97.73% compared to the baseline's 97.67%. This indicates a more optimal balance between detecting threats and minimizing false alarms.

Furthermore, the proposed architecture demonstrates higher precision (97.57%) than the Random Forest baseline (96.50%). In a live network environment, higher precision directly translates to fewer legitimate URLs being incorrectly blocked. This reduction in false positives ensures smoother user browsing experiences without compromising security. Therefore, the architectural enhancement provides measurable operational advantages beyond marginal accuracy differences.

The performance comparison in Table III establishes the CNN 3 (CNN + XGBoost) hybrid model as the best classifier, achieving a peak validation accuracy of 97.83% and an F1-score of 98.06%. This absolute superiority is influenced by the efficient gradient boosting architecture applied by the XGBoost algorithm, which can execute abstract dense vector parameters from the CNN quickly without experiencing overfitting symptoms. Subsequently, after being rigorously evaluated on the unseen testing data (Table II), the proposed method successfully produced a final accuracy score of 97.47%, a precision of 97.57%, a recall of 97.89%, and an F1-score of 97.73%. This proposed method is proven to be superior to the State-of-the-Art (SOTA) baseline Random Forest model, which culminated at an accuracy level of 97.38% on the same testing partition.

TABLE III  
EVALUATION OF CNN, SPATIAL, AND HYBRID MODELS

| No. | Method                        | Acc           | Precision     | Recall        | F1-Score      |
|-----|-------------------------------|---------------|---------------|---------------|---------------|
| 1.  | CNN 1 (RGB Color Mapping)     | 0.9719        | 0.9664        | 0.9837        | 0.9750        |
| 2.  | CNN 2 (Spatial Rearrangement) | 0.9638        | 0.9443        | <b>0.9934</b> | 0.9683        |
| 3.  | CNN 3 (CNN + XGBoost)         | <b>0.9783</b> | <b>0.9728</b> | 0.9886        | <b>0.9806</b> |
| 4.  | CNN 4 (CNN + LightGBM)        | 0.9729        | 0.9650        | 0.9867        | 0.9759        |

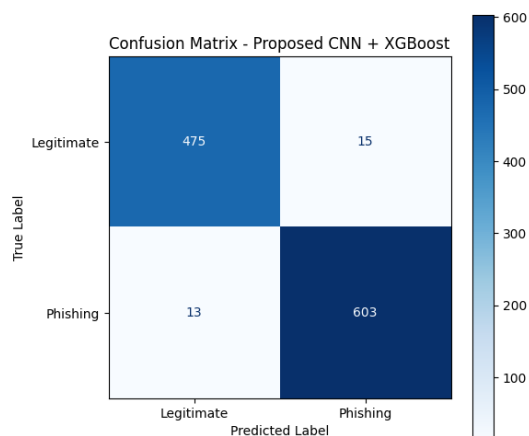


Figure 2. Confusion Matrix Performance of the Proposed CNN + XGBoost Model on Unseen Data.

#### A. Image Variant (RGB Color Mapping)

Image transformation with RGB Color Mapping successfully recorded an accuracy value of 0.9719 (97.19%). This proves that color contrast can help convolution filters extract features to be sharper.

#### B. Spatial Rearrangement

Theoretically, rearrangement on correlation-based features is expected to improve performance, with accuracy results reaching 0.9638 (96.38%). This proves that the basic CNN architecture still requires a more robust classification component to handle the extracted data.

#### C. Hybrid Architecture

The increase in performance that occurs in the use of hybrid models shows a stable efficiency. The CNN + XGBoost model successfully achieved the highest accuracy with 0.9783 (97.83%), followed by CNN + LightGBM with 0.9729 (97.29%). Both officially surpassed the SOTA RF Base accuracy record belonging to the previous reference research

To justify the RGB Color Mapping technique, an ablation study compares it against the Spatial Rearrangement baseline. Theoretically, mapping tabular features into RGB channels allows convolutional filters to exploit inter-channel correlations. Color gradients effectively represent the semantic proximity between lexical and host-based features. The RGB variant achieved 97.19% accuracy, outperforming the Spatial Rearrangement at 96.38%. This empirical evidence confirms that preserving feature correlation through color space provides a richer spatial hierarchy.

Strict deduplication was applied prior to splitting to prevent direct data leakage from identical URL strings. Furthermore, the MinMaxScaler was fitted exclusively on the training set to avoid distribution leakage across partitions. However, a random stratified split may still allow structurally similar subdomains to cross partition boundaries. While the 80:10:10 hold-out validates immediate generalization, it remains a

limitation for robust real-world application. Future work will incorporate domain-aware splitting and external temporal validation on newly collected phishing URLs.

The diagnostic transparency of the proposed framework is further validated through the extraction of confusion matrix parameters during the unseen testing phase, as illustrated in Figure 2. Based on the evaluation of 1,106 blind testing instances, the model successfully identifies 475 true negative (TN) samples and 603 true positive (TP) classifications. This high density along the diagonal axis demonstrates that the sequential mapping layers are highly resilient in preserving structural pattern alignment across transformed spatial image tensors.

Conversely, the model exhibits minimal classification discrepancies, yielding only 15 false positive (FP) anomalies and 13 false negative (FN) errors. The critical suppression of the false negative rate down to 2.11% is highly strategic for active network perimeter defenses. This structural optimization ensures that malicious web links are trapped responsively, thereby effectively preventing credential harvesting activities and mitigating macro-scale financial data breaches.

## IV. CONCLUSION

The cyber threat detection experiment on malicious website links using the spatial-RGB transformation approach and hybrid classifier architecture has been comprehensively completed. Empirical testing results prove that the integration of the CNN + XGBoost model successfully recorded the highest performance, achieving a final testing accuracy of 97.47% and a recall of 97.89%. The acquisition of these metrics is quantitatively proven valid, surpassing the highest record of the replicated baseline State-of-the-Art Random Forest model, which culminated at an accuracy level of 97.38%. The main limitation in the implementation of this research lies in high computational processing load as well as the extensive time duration while the visualization of the conversion of tabular features into two-dimensional image tensors is running. Advanced assessment in the future is suggested to integrate a non-linear feature selection algorithm before the transformation phase in order to optimally prune the image pixel dimensions [16]. Future research will incorporate external and temporal validation to ensure robust generalization against evolving threats. Testing the model against newly emerging phishing campaigns will verify its adaptability to dynamic attacker tactics. Additionally, domain-aware splitting will be implemented to completely isolate URL families across future data partitions. These enhancements will further validate the framework's reliability and scalability in diverse real-world environments.

## REFERENCES

- [1] M. A. Daniel, S.-C. Chong, L.-Y. Chong, and K.-K. Wee, "Optimising Phishing Detection: A Comparative Analysis of Machine Learning Methods with Feature Selection," *J. Inform. Web*

- Eng., vol. 4, no. 1, pp. 200–212, Feb. 2025, doi: 10.33093/jiwe.2025.4.1.15.
- [2] S. Prasad Panda, “The Evolution and Defense Against Social Engineering and Phishing Attacks,” *Int. J. Sci. Res. IJSR*, vol. 14, no. 5, pp. 397–408, May 2025, doi: 10.21275/SR25504223645.
- [3] I. Abid, M. Abid, and A. Aqsa, “AI-Based Phishing Detection for Emails and URLs,” 2025, SSRN Electron. J. doi: 10.2139/ssrn.5336075.
- [4] D. Komalasari, T. B. Kurniawan, D. A. Dewi, M. Z. Zakaria, Z. Abdullah, and A. Alanda, “Phishing Domain Detection Using Machine Learning Algorithms,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 15, no. 1, pp. 318–327, Feb. 2025, doi: 10.18517/ijaseit.15.1.12553.
- [5] R. Salsabila, A. F. Haq, and D. Putriana, “Memetakan Konsep Dan Faktor-Faktor Psikologis Yang Mempengaruhi Kerentanan Individu Sebagai Korban Phishing: A Scoping Review,” *J. Psikol. Forensik Indones.*, vol. 5, no. 2, May 2026, doi: 10.71088/jpfi.v5i2.103.
- [6] “205 Cybersecurity Stats and Facts for 2026.” Accessed: Jun. 15, 2026. [Online]. Available: <https://www.vikingcloud.com/blog/cybersecurity-statistics#1>
- [7] M. A. S. Aritonang, M. J. Simanulang, T. P. Batubara, I. Zega, and M. H. Afrizal, “Natural Language Processing (NLP) and Support Vector Machine (SVM) Optimization in Detecting Phishing Website URLs,” *J. Tek. Inform. Jutif*, vol. 7, no. 1, pp. 552–570, Feb. 2026, doi: 10.52436/1.jutif.2026.7.1.5334.
- [8] APWG, “Phishing Activity Trends Report, 1st Quarter 2026,” May 2026. [Online]. Available: [https://docs.apwg.org/reports/apwg\\_trends\\_report\\_q1\\_2026.pdf](https://docs.apwg.org/reports/apwg_trends_report_q1_2026.pdf)
- [9] APWG, “Phishing Activity Trends Report, 3rd Quarter,” Dec. 2025. [Online]. Available: [https://docs.apwg.org/reports/apwg\\_trends\\_report\\_q3\\_2025.pdf](https://docs.apwg.org/reports/apwg_trends_report_q3_2025.pdf)
- [10] APWG, “Phishing Activity Trends Report, 4th Quarter,” Feb. 2026. [Online]. Available: [https://docs.apwg.org/reports/apwg\\_trends\\_report\\_q4\\_2025.pdf](https://docs.apwg.org/reports/apwg_trends_report_q4_2025.pdf)
- [11] Z. Fatima *et al.*, “Explainable AI for IOT Devices and Robotic Communication Phishing Detection: A Machine Learning Approach Using LIME and SHAP,” *Eng. Technol. Appl. Sci. Res.*, vol. 15, no. 5, pp. 26478–26486, Oct. 2025, doi: 10.48084/etasr.11595.
- [12] Z. Shahbazi, R. Jalali, and M. Molaevand, “AI-Based Phishing Detection and Student Cybersecurity Awareness in the Digital Age,” *Big Data Cogn. Comput.*, vol. 9, no. 8, p. 210, Aug. 2025, doi: 10.3390/bdcc9080210.
- [13] K. M. M. Uddin, N. Biswas, S. T. Rikta, Md. Nur -A-Alam, and R. Mostafiz, “Explainable Machine Learning for Phishing Site Detection: A High-Efficiency Approach Using Boosting Models and SHAP,” *J. Eng.*, vol. 2025, no. 1, p. e70110, Jan. 2025, doi: 10.1049/tje2.70110.
- [14] K. Dev, P. Pratim, B. Elly, and G. Azziz, “Unlocking Cybersecurity Value through Machine Learning Analytics: A Case Study of Anomaly Detection,” Jan. 2025.
- [15] K. Owa and O. Adewole, “Benchmarking Machine Learning Techniques for Phishing Detection and Secure URL Classification,” *Int. J. Comput. Sci. Mob. Comput.*, vol. 14, no. 1, pp. 20–37, Jan. 2025, doi: 10.47760/ijcsmc.2025.v14i01.003.
- [16] H. Jabbar and S. Al-Janabi, “AI-Driven Phishing Detection: Enhancing Cybersecurity with Reinforcement Learning,” *J. Cybersecurity Priv.*, vol. 5, no. 2, p. 26, May 2025, doi: 10.3390/jcp5020026.
- [17] A. Kumar, “dataset Optimising Phishing Detection A Comparative.” [Online]. Available: <https://www.kaggle.com/datasets/akashkr/phishing-website-dataset>. Accessed: Apr. 18, 2026.
- [18] B. Elaziz, Y. Ommame, M. Eddabbah, and Y. Laaziz, “Enhancing Diabetes Detection Through Data Preprocessing: A Comparative Analysis of Machine Learning Algorithms,” in *2024 International Conference on Computing, Internet of Things and Microwave Systems (ICCIMS)*, Gatineau, QC, Canada: IEEE, Jul. 2024, pp. 1–6. doi: 10.1109/ICCIMS61672.2024.10690679.
- [19] A. Nazarkar, H. Kuchulakanti, C. S. Paidimarry, and S. Kulkarni, “Impact of Various Data Splitting Ratios on the Performance of Machine Learning Models in the Classification of Lung Cancer,” in *Proceedings of the Second International Conference on Emerging Trends in Engineering (ICETE 2023)*, vol. 223, B. Raj, S. Gill, C. A. G. Calderon, O. Cihan, P. Tukkaraja, S. Venkatesh, V. M. S., M. Mudigonda, M. Gaddam, and R. K. Dasari, Eds., in *Advances in Engineering Research*, vol. 223, Dordrecht: Atlantis Press International BV, 2023, pp. 96–104. doi: 10.2991/978-94-6463-252-1\_12.
- [20] B. Deepa and K. Ramesh, “Epileptic seizure detection using deep learning through min max scaler normalization,” *Int. J. Health Sci.*, pp. 10981–10996, May 2022, doi: 10.53730/ijhs.v6nS1.7801.
- [21] Haniza Nahar, Zulkiflee Muslim, Mohammad Radzi Motsidi, Siti Rahayu Selamat, Warusia Yassin, and Fauzi Adi Rafrastara, “Psychological Effects of Phishing Email Exposure: A Review,” *SSRN Electron. J.*, vol. IX, no. XI, Dec. 2025, doi: 10.47772/IJRISS.
- [22] I. M. Pires, F. Hussain, N. M. M. Garcia, P. Lameski, and E. Zdravevski, “Homogeneous Data Normalization and Deep Learning: A Case Study in Human Activity Classification,” *Future Internet*, vol. 12, no. 11, p. 194, Nov. 2020, doi: 10.3390/fi12110194.
- [23] U. Şevik and F. S. Aydemir, “A computationally efficient hybrid framework combining deep feature extraction and gradient boosting for early diagnosis of Olive leaf diseases,” *Sci. Rep.*, vol. 16, no. 1, p. 2170, Dec. 2025, doi: 10.1038/s41598-025-31918-x.
- [24] W. S. Tyas, F. A. Rafrastara, and W. Khozi, “Optimizing URL-Based Phishing Detection Using XGBoost and Relief Feature Selection,” *sinkron*, vol. 10, no. 1, pp. 430–438, Jan. 2026, doi: 10.33395/sinkron.v10i1.15651.
- [25] S. Pawar, P. Dhane, D. Shelke, P. Dheple, and N. Doshi, “PCOS Detection using Hybrid CNN-XG Boost Model - A Multimodal Data Approach,” vol. 18, 2025, doi: <http://dx.doi.org/10.13005/bpj/3252>.
- [26] W. Bouchelligua, “PatternFusion: a hybrid model for pattern recognition in time-series data using ensemble learning,” *Sci. Rep.*, vol. 15, no. 1, p. 43371, Dec. 2025, doi: 10.1038/s41598-025-28649-4.