

28/11/2025

Assignment 1: Innovation & Research

Changemakers: Creativity and
Value Creation

Module: 5033



Michael Kofi Awua Asiedu (2411698)
LECTURER: DR FAHAD AHMED

Table of Contents

Introduction.....	2
Innovation and its Impact	2
Mindmap	2
I. Introduction	3
II. Applications of Machine Learning in Cybersecurity	4
III. Ideation and Project Definition	5
IV. Project Description	5
V. Research Sources	6
VI. Project Scope	6
Proposed Solution	7
VII. Development and Design	9
Chosen Methodology	10
Design: High-Level System Architecture Diagram	10
Data flow diagram (DFD) for URL Processing.....	12
Conclusion	13
References	13

Introduction

The aim of this assignment is to identify opportunities, shape and improve ideas, carry out research, and apply critical analysis to generate meaningful value for society through innovation. As discussed in Unit 1, innovation goes beyond simple invention; it is the effective implementation of creative concepts that deliver substantial benefits to both individuals and communities. This process highlights the connection between design thinking and research, demonstrating that when combined, they can uncover insights that drive impactful solutions and create value on both personal and societal levels.

Innovation and its Impact

Innovation and the process of value creation is deeply linked to societal development; technological innovation can address global risks, foster digital inclusion, and solve "wicked problems"—complex, evolving issues with no definitive solution including cybersecurity[1]. Cybersecurity has become one of the most pressing challenges of the digital age. Among the wide range of threats, phishing remains particularly devious because it exploits human psychology rather than purely technical vulnerabilities. Human Factor: According to Verizon DBIR 2024, 74% of breaches involve human actions such as social engineering, mistakes, or misuse.

Phishing & Credentials: Phishing contributes to more than one-third of breaches, while stolen credentials remain the leading cause, implicated in 50% of incidents[2. 2. 3].

Mindmap

Mindmap organises ideas visually around a central concept, mind mapping enhances learning and creativity. This technique, which uses keywords and radiating branches, aligns with the brain's natural workflow to improve memory and understanding. Its versatility makes it effective for diverse tasks, from helping students deconstruct complex subjects to aiding business teams in project planning and brainstorming[4]

The mindmap below places **Creativity and Innovation** at the centre, branching into key areas like idea generation, design thinking, research, implementation, and impact[5]. Under *Impact*, you'll see a dedicated **Cybersecurity** child node with sub-branches for threat detection, fraud prevention, vulnerability management, and adaptive defence.



The diagram vividly bridges theory and practice, illustrating how abstract concepts become actionable outcomes. It highlights the role of imaginative problem-solving, showing that creativity alone is not enough; when guided through structured, disciplined processes, it evolves into practical solutions. These solutions manifest as resilient cybersecurity tools designed to withstand complex threats. Moreover, the diagram emphasizes proactive strategies, demonstrating that disciplined innovation not only addresses existing vulnerabilities but also anticipates emerging risks. By making this connection clear, the diagram underscores how theory, creativity, and rigor converge to produce effective, forward-thinking approaches to cybersecurity and threat mitigation in real-world contexts

A Hybrid Machine Learning Architecture for Adaptive Anti-Phishing Defence (AAD) in a Browser Extension

I. Introduction

Phishing is a major cybersecurity challenge, exploiting human vulnerabilities and bypassing traditional defences. Conventional security methods, while useful, often fail to counter evolving attacks. Phishing remains particularly insidious because it exploits human psychology rather than purely technical vulnerabilities. Attackers craft deceptive emails, websites, or messages that trick users into revealing sensitive information such as login credentials, financial details, or personal data. Traditional defences, such as

blacklists or rule-based filters, often fail to keep pace with the creativity and adaptability of attackers. This reality has created a demand for intelligent, adaptive solutions that can evolve alongside the threat landscape[6].

Machine Learning (ML) has emerged as a transformative technology in this space. Machine learning (ML), a branch of artificial intelligence (AI), enables systems to perform tasks by recognizing patterns in data rather than relying on manually coded rules. Algorithms are trained on datasets to build models, which are then applied to new information. The learning process is typically categorized into approaches such as supervised and unsupervised learning[7]. In cybersecurity, ML is applied to **threat detection**, identifying anomalies in user behaviour or network traffic; **fraud prevention**, spotting unusual financial transactions; and **vulnerability management**, prioritizing risks likely to be exploited. Its adaptability makes ML essential for combating phishing and strengthening digital security[8].

II. Applications of Machine Learning in Cybersecurity

Threat Detection

ML models analyse network traffic in real time, identifying anomalies and behaviours indicative of malicious activity. For example, data exfiltration attempts often involve unusual traffic patterns, while botnet activity may manifest as repetitive, coordinated requests. Static signature-based systems can only detect known threats, but ML models generalize from patterns, enabling them to flag novel or zero-day attacks. This proactive capability represents a fundamental shift from reactive defence to anticipatory protection.

Fraud Prevention

In the financial sector, ML algorithms scrutinize millions of transactions per second. By learning from historical data, they detect fraudulent patterns such as unusual spending behaviours, geographic inconsistencies, or account takeovers. Consumers benefit from reduced exposure to fraud, while institutions save billions in potential losses. Importantly, ML systems adapt to evolving fraud tactics, ensuring resilience against attackers who constantly innovate.

Vulnerability Management

ML-powered systems scan codebases and networks to prioritize vulnerabilities based on their likelihood of exploitation. Instead of overwhelming security teams with thousands of alerts, ML models rank risks intelligently, allowing resources to be focused on the most pressing threats. This prioritization improves efficiency and reduces the window of exposure.

III. Ideation and Project Definition

Project Title

A Hybrid Machine Learning Architecture for Adaptive Anti-Phishing Defence (AAD) in a Browser Extension

Aim

To design, develop, and evaluate a hybrid client-server browser extension that utilizes machine learning to provide real-time, adaptive protection against phishing attacks for end-users.

Objectives

1. Conduct a comprehensive literature review on ML applications in phishing detection, hybrid system architectures, and adversarial evasion techniques.
2. Design a system architecture featuring a lightweight client-side model for instant classification and a robust server-side model for in-depth analysis.
3. Develop a functional prototype browser extension using TensorFlow.js for client-side inference and a Python-based backend with a BERT-style model for email analysis.
4. Evaluate the prototype's performance using standard metrics (Precision, Recall, F1-Score) and test its resilience against adversarial attacks.

IV. Project Description

This project envisions an integrated defence system embedded directly within a web browser. The investigation will cover several dimensions:

1. **History of the Technology:** Anti-phishing defences have evolved from simple blacklists to investigative filters, and now to ML and deep learning models. Early systems relied on static rules, which attackers easily bypassed. ML introduced adaptability, enabling detection of novel threats.
2. **Industry Embrace:** Finance, healthcare, and e-commerce industries have adopted ML for security. Banks use ML to monitor transactions, hospitals protect patient data, and online retailers safeguard customer accounts. These examples demonstrate the broad applicability of ML.
3. **Impact on Cybersecurity:** ML has shifted the industry from reactive to proactive defence. Zero-day attacks, once nearly impossible to detect, can now be flagged

by anomaly detection models. This capability reduces the damage window and enhances resilience.

4. **Contribution to the Modern World:** By embedding ML into a browser extension, this project democratizes advanced phishing protection. Individuals and small businesses, often unable to afford enterprise solutions, gain access to cutting-edge defences.
5. **Positive and Negative Impacts:**
 - *Positive:* Enhanced detection rates, adaptability, and automation.
 - *Negative:* Computational overhead, false positives, and susceptibility to adversarial attacks.
6. **Implementation Challenges:** Challenges include acquiring quality datasets, requiring specialized AI expertise, and addressing model obsolescence due to evolving phishing tactics.
7. **Stakeholder Benefits:** End-users gain seamless protection, organizations reduce breach risks, and security teams focus on complex threats rather than routine alerts[9].

V. Research Sources

Key resources include:

- Google Scholar.
- UWTSD Library Assistant.
- IEEE Xplore Digital Library for peer-reviewed papers.
- Coursera's Supervised Machine Learning: Regression and Classification.
- Technical documentation for TensorFlow.js and Python frameworks.
- Datasets from OpenPhish, PhishTank, and PhiUSIIL.

VI. Project Scope

The chosen technology is **Machine Learning in Cybersecurity**, with a focus on phishing defence. Departments such as HR and Finance are particularly vulnerable. HR staff often handle sensitive credentials, while Finance teams manage wire transfers, making them prime targets.

Benefits

Implementing ML reduces successful phishing attacks, automates threat intelligence, and decreases reliance on human judgment. Unlike employees who may overlook subtle cues, ML systems consistently analyse data without fatigue.

Drawbacks

Costs include development, retraining to counter concept drift, and computational resources. Deep learning inference can be resource-intensive, especially for real-time applications.

Future Directions

The integration of **Explainable AI (XAI)** is crucial. Users must understand why a threat was blocked. For example, an extension might explain: *“This email failed DMARC authentication and contains a homoglyph link.”* Domain-based Message Authentication, Reporting and Conformance (DMARC) is an email security protocol that enables domain owners to safeguard their domains against unauthorized use, particularly email spoofing. Such transparency educates users and strengthens human defences.

Proposed Solution

Guided by a **pragmatist research philosophy**, the project combines quantitative metrics (accuracy, recall) with qualitative feedback (usability). The motivation stems from rising phishing losses and the lack of accessible tools for non-enterprise users.

Expected Outcomes

The prototype aims for >95% recall and <2% false positives. Delivering an open-source solution ensures accessibility and fosters community collaboration.

Research Tools

- Peer-reviewed journals and government reports.
- Citation analysis tools like “Publish or Perish.”
- Reference management with RefWorks.
- Cross-verification of claims for credibility.

Implementation Technologies

- **Frontend:** JavaScript, TensorFlow.js, Browser APIs.
- **Backend:** Python, Flask/Django, PyTorch/TensorFlow, PostgreSQL.
- **DevOps:** Docker, Git, Celery for task queues.

Realistic Benchmarking: Create evaluation frameworks that reflect actual attacker behaviour, such as obfuscated URLs, temporary phishing sites, and black-box attack simulations.

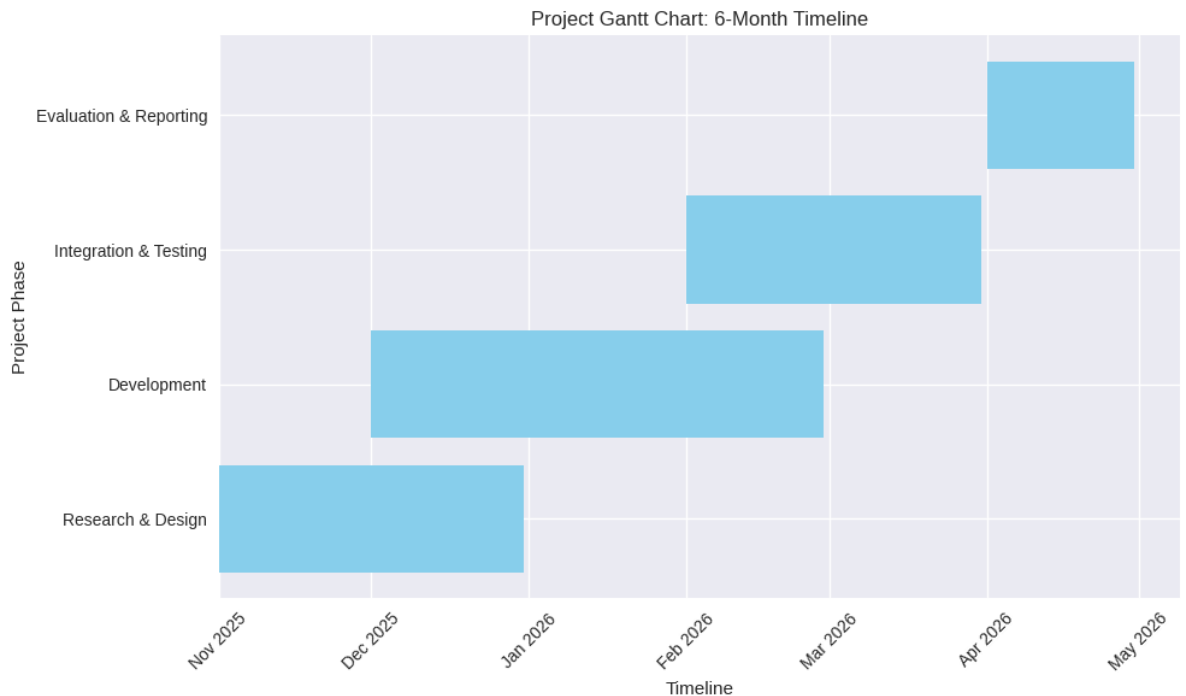
Transparent Hybrid Detection: Develop detection systems that blend AI with explainable rules, ensuring analysts understand why threats are flagged and reducing unnecessary alerts.

Resilient Deployment: Implement multi-stage pipelines that merge fast rule-based checks with lightweight device-level ML and scalable cloud analysis, incorporating automated updates to maintain accuracy against changing threats

Task Schedule/Critical Path Method using Gantt Chart

A Gantt chart is a graphical project scheduling tool that represents project activities as horizontal bars against a time axis. This visualization illustrates task breakdowns, timelines, and dependencies, enabling managers to identify the critical path and monitor advancement by comparing planned versus actual completion. Below are the task schedule and Gantt Chart showing research & design, development, integration & testing and evaluation & reporting[10].

- Months 1–2: Research & Design.
- Months 2–4: Development.
- Months 4–5: Integration & Testing.
- Month 6: Evaluation & Reporting.



Risk Assessment

- **Underperforming ML Model:** Mitigated by pre-trained baselines and hyperparameter tuning.
- **Scope Creep:** Controlled via Minimum Viable Product (MVP) definition and Must Have, Should Have, Could Have and Would Have (MoSCoW) prioritization.
- **Data Privacy:** Ensured by local processing of sensitive data, with only metadata sent to servers.

VII. Development and Design

Software Development Life Cycle (SDLC) models provide a disciplined framework for software projects, aiming to deliver quality results efficiently. However, while traditional methodologies like Waterfall, V-Model, and Spiral have been widely used, they often come with significant drawbacks. These include rigidity that leads to high risk and cost, as well as difficulties in ensuring customer satisfaction due to their inherent time complexity and limited flexibility[11].

SDLC Models Considered

- **Waterfall:** Linear but inflexible.
- **V-Model:** Emphasizes validation but lacks adaptability.
- **Agile:** Iterative, responsive to feedback.
- **Scrum:** Uses sprints to deliver increments.

Chosen Methodology

A **Hybrid Agile-Waterfall model** is selected. Initial research and design follow a structured approach, while development and testing adopt Scrum sprints. This hybrid ensures upfront planning while retaining flexibility for ML experimentation.

Development Phases

1. Research & Design

Literature review establishes the foundation. Architectural design specifies client-server interactions, data flow, and security considerations.

2. Client-Side Development

A lightweight TensorFlow.js model runs in the browser, classifying URLs and emails instantly. This ensures real-time protection without server latency.

3. Server-Side Development

A Python backend hosts a BERT-style model for deeper analysis. Complex cases are escalated to the server, which provides robust classification.

4. Integration

Client and server components are integrated via secure APIs. Docker containers ensure portability and reproducibility.

5. Testing

Performance is measured using precision, recall, and F1-score. Adversarial testing evaluates resilience against crafted attacks designed to fool ML models.

6. Evaluation

Results are analysed, and user feedback is collected. The final report documents findings, limitations, and future directions.

Design: High-Level System Architecture Diagram

Unified Modelling Language (UML) diagrams serve as the industry standard for illustrating object-oriented designs. By presenting systems visually, they make it easier to clarify requirements and determine the scope of applications[12].

The UML diagram below illustrates a hybrid security architecture designed for efficient and intelligent threat detection. Browser extensions act as the first line of defence, intercepting URLs and emails to perform immediate client-side lexical analysis. Obvious threats are blocked locally in real time. For ambiguous cases, only extracted features—never raw data—are transmitted to the server. The

backend leverages deep learning models and external threat intelligence sources to conduct comprehensive analysis, then sends the final verdict back to the client. This architecture balances the speed of local processing with the depth of cloud-based AI, enabling fast mitigation of clear threats while preserving robust detection for complex scenario

Hybrid Flow of URL for Analysis

CLIENT-SIDE LIGHTWEIGHT MODEL

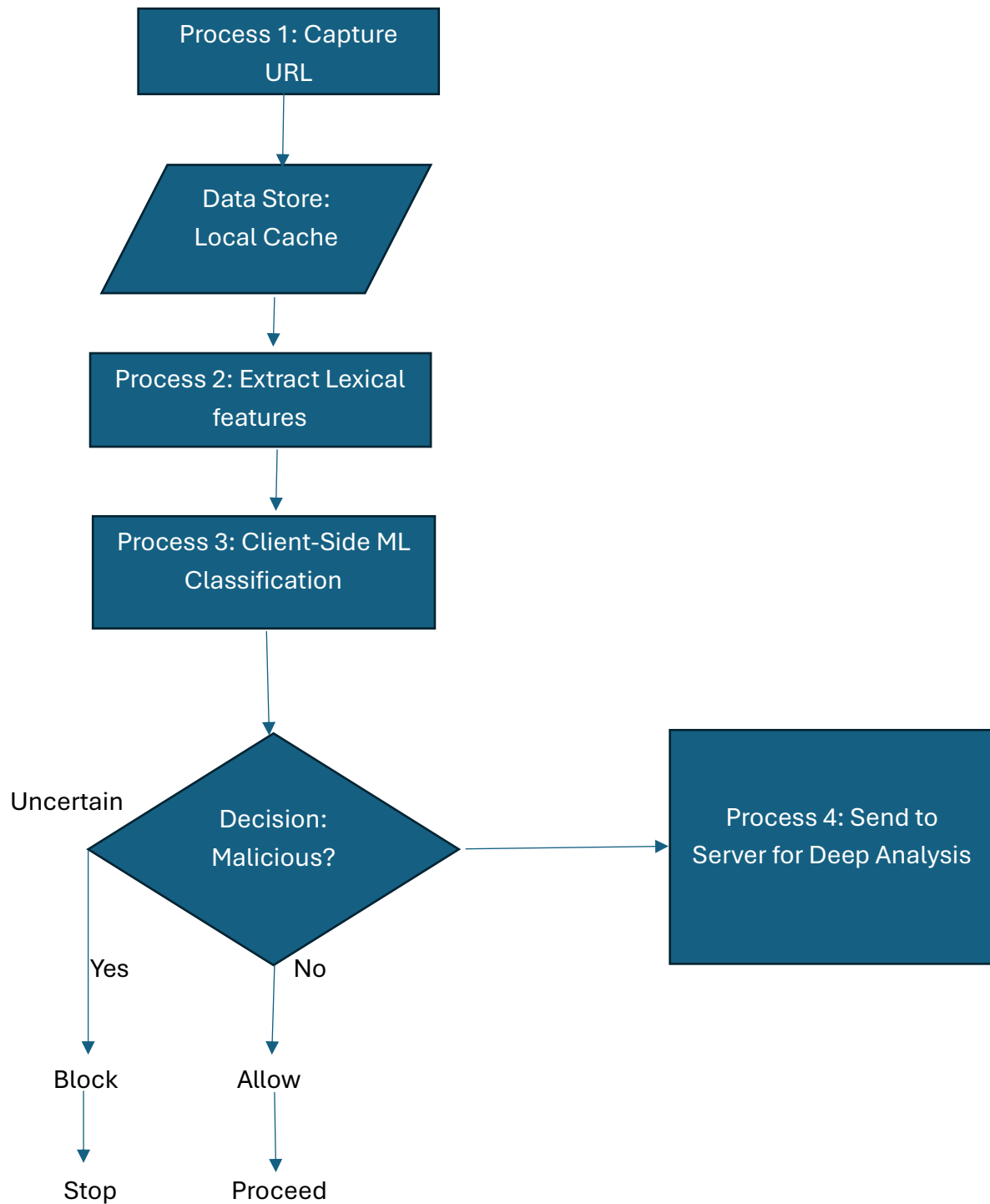


SERVER-SIDE HEAVYWEIGHT MODEL



SERVER-SIDE HEAVYWEIGHT MODEL

Data flow diagram (DFD) for URL Processing



This data flow diagram depicts the system's internal operations, starting with the capture of a URL and progressing through feature extraction and swift client-side classification. At the core lies a decision node that presents three possible outcomes: immediate blocking of malicious content, direct approval of benign activity, or escalation to advanced server-side analysis for a definitive verdict. By emphasizing rapid, local decision-making, this architecture aligns with the "21-second rule," ensuring timely protection while maintaining the capacity for deeper evaluation when uncertainty arises

Conclusion

This proposal outlines a structured approach to tackling the persistent challenge of phishing by leveraging the capabilities of machine learning. At its core, the Adaptive Anti-Phishing Defence (AAD) system employs a hybrid client-server architecture designed to balance the demand for immediate response with the need for comprehensive and precise threat analysis. The research strategy guiding this project is pragmatic, ensuring that theoretical insights are paired with practical application. Development will follow a hybrid Agile-Waterfall methodology, combining the rigor of upfront planning with the flexibility of iterative refinement. The anticipated outcome is a significant contribution to applied cybersecurity: a scalable, efficient, and accessible defence mechanism capable of protecting a diverse range of users from evolving phishing threats.

References

References

- [1] J. Lönngren and K. van Poeck, "Wicked problems: a mapping review of the literature," *International Journal of Sustainable Development & World Ecology*, vol. 28, no. 6, pp.481–502, 2021. doi: 10.1080/13504509.2020.1859415. Available: <https://doi.org/10.1080/13504509.2020.1859415>

- [2] L. Xu, "Trends in US Healthcare Data Breaches," *2025 IEEE International Conference on AI and Data Analytics (ICAD)*, 2025. doi: 10.1109/ICAD65464.2025.11114030.
- [3] Z. Angyalos-Porkoláb and R. Szilágyi, "Cybersecurity Risks in Critical Infrastructures: Insights from CISA and ENISA Data," *Journal of Agricultural Informatics (ISSN 2061-862X)*, vol. 16, no. 2, pp.1–11, 2025.
- [4] J. R. Mammen and C. R. Mammen, "Beyond concept analysis: Uses of mind mapping software for visual representation, management, and analysis of diverse digital data," *Res.Nurs.Health*, vol. 41, no. 6, pp.583–592, 2018. doi: 10.1002/nur.21920.
- [5] J. Wöhler and R. Reinhardt, "The users' perspective on how creativity techniques help in the idea generation process—A repertory grid study," *Creativity and innovation management*, vol. 30, no. 1, pp.144–163, 2021. doi: 10.1111/caim.12424.
- [6] Saleem Raja Abdul Samad, et al., "Analysis of the Performance Impact of Fine-Tuned Machine Learning Model for Phishing URL Detection," *Electronics*, vol. 12, no. 7, 2023. doi: 10.3390/electronics12071642. Available: <https://www.proquest.com/scholarly-journals/analysis-performance-impact-fine-tuned-machine/docview/2799623424/se-2?accountid=130472>
https://libkey.io/libraries/1422/openurl?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest&rft_val_fmt=journal&rft.genre=article&rft.jtitle=Electronics&rft.atitle=Analysis+of+the+Performance+Impact+of+Fine-Tuned+Machine+Learning+Model+for+Phishing+URL+Detection&rft.au=Saleem+Raja+Abdul+Samad%253BBalasubaramanian%252C+Sundarvativazhagan%253BAI%252C+Kaabi%252C+Amna+Salim%253BSharma%252C+Bhisham%253BChowdhury%252C+Subrata%253BMehbodniya%252C+Abolfazl%253BWebber%252C+Julian+L%253BBostani%252C+Ali.&rft.aulast=Saleem+Raja+Abdul+Samad&rft.aufirst=&rft.date=2023-01-01&rft.volume=12&rft.issue=7&rft.spage=1642&rft.isbn=&rft.btitle=&rft.title=Electronics&rft.issn=&rft_id=info:doi/10.3390%252Felectronics12071642
<https://doi.org/10.3390/electronics12071642>
- [7] Amir Masoud Rahmani, et al., "Machine Learning (ML) in Medicine: Review, Applications, and Challenges," *Mathematics*, vol. 9, no. 22, 2021. doi: 10.3390/math9222970. Available: <https://www.proquest.com/scholarly-journals/machine-learning-ml-medicine-review-applications/docview/2602102241/se-2?accountid=130472>
https://libkey.io/libraries/1422/openurl?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest&rft_val_fmt=journal&rft.genre=article&rft.jtitle=Mathe

[maths&rft.atitle=Machine+Learning+%2528ML%2529+in+Medicine%253A+Review%252C+Applications%252C+and+Challenges&rft.au=Amir+Masoud+Rahmani%253BYousefpoor%252C+Efat%253BMohammad+Sadegh+Yousefpoor%253BMehmood%252C+Zahid%253BHaider%252C+Amir%253BHosseinzadeh%252C+Mehdi%253BRizwan+Ali+Naqvi.&rft.aulast=Amir+Masoud+Rahmani&rft.aufirst=&rft.date=2021-01-01&rft.volume=9&rft.issue=22&rft.spage=2970&rft.isbn=&rft.btitle=&rft.title=Mathematics&rft.issn=&rft_id=info:doi/10.3390%252Fmath9222970](https://doi.org/10.3390/math9222970)
<https://doi.org/10.3390/math9222970>

- [8] Saeid Sanei and Jonathon A. Chambers, *Machine Learning for EEG Analysis*, 2021. doi: 10.1002/9781119386957.ch7
- [9] X. Lu, K. Wu and Y. Chen, "Enhancing Situational Awareness in Industrial Automation and Cybersecurity Through an Integrated Framework That Leverages Both Machine Learning and Deep Learning Technologies," *2024 IEEE 5th International Conference on Pattern Recognition and Machine Learning (PRML)*, 2024. doi: 10.1109/PRML62565.2024.10779917.
- [10] N. Bdair, W. Alzyadat and S. AlZubi, "Intelligent Model for Optimizing Gantt Chart in the Planning Stage," *2023 International Conference on Information Technology (ICIT)*, 2023. doi: 10.1109/ICIT58056.2023.10225762.
- [11] A. Garg, R. Kumar Kaliyar and A. Goswami, "PDRSD-A systematic review on plan-driven SDLC models for software development," *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2022. doi: 10.1109/ICACCS54159.2022.9785261.
- [12] H. Koc, et al., "UML diagrams in software engineering research: a systematic literature review," *Proceedings*, 2021.