A Machine Learning for Phishing Detection in Healthcare.

Samuel Twum[1], Ezra Yalley[2], Alpha Agusah[3], Richard Sarpong[4]

[1]Department of Computer Application, Lovely Professional University, Punjab-India
samuel.12419294@lpu.in

[2]Department of Computer Application, Lovely Professional University, Punjab-India
ezra.yalley@gmail.com

[3]Department of Computer Application, Lovely Professional University, Punjab-India
iamalphaagusah@gmail.com

[4]Department of Computer Application, Lovely Professional University, Punjab-India
sarpongrichard32@gmail.com

***ABSTRACT*—**Phishing, a social Engineering attacks are becoming more common because of the quick digitization of healthcare services. Phishing is often used to steal user data, including login credentials and credit card numbers. It occurs when an attacker, masquerading as a trusted entity, dupes a victim into opening an email, instant message, or text message and this, occurring at the healthcare services clearly puts patient data security and system integrity at danger. This study introduces a brand-new machine learning-based method for identifying phishing attempts directed at healthcare institutions. Using real- world phishing and authentic emails from a variety of healthcare- related domains, we created an extensive dataset. By employing sophisticated feature extraction methods, such as URL analysis, email header examination, and linguistic features, we were able to identify important markers of phishing activity. Emails were categorized as either authentic or phishing using a variety of machine learning techniques, such as Random Forest, Support Vector Machines, and Neural Networks. Our tests showed that the Random Forest algorithm was the most effective, achieving the highest accuracy of 95percent with an F1 score of 0.93**.**

KEYWORDS*: Cyber Security, Phishing Detection, Health-care, Machine Learning (ML), Supervised Learning, Unsupervised Learning, Deep Learning Model.*

1. INTRODUCTION

With the quick pace of digitization in healthcare services, the sector has been more exposed to cyber-attacks, especially phishing attacks. Phishing is a manipulative method adopted by cybercriminals to trick people into revealing sensitive information, including patient records, financial data, and login credentials. As healthcare data is of a critical nature, a security breach can have disastrous outcomes, ranging from identity theft to financial loss and compromised patient safety. Traditional techniques for phishing detection, like rule-based filters and blacklisting approaches, commonly are unable to detect advanced and adaptive phishing attempts. Machine learning (ML) has been a strong technology for phishing attack industry to provide better security against cyber-attacks. In spite of the growing use of cybersecurity practices, phishing attacks in the healthcare industry are on the rise, frequently caused by human mistake and the expertise of attackers. Most healthcare workers and administrative personnel are not properly trained to identify phishing attempts and are therefore vulnerable to misleading tactics. Moreover, cybercriminals continually update their tactics, employing social engineering strategies and AI-based phishing campaigns that evade conventional security practices. Hence, an urgent need is felt for a proactive and smart phishing detection system that can respond to new threats in real-time. In addition, integrating ML- based phishing detection into current healthcare cybersecurity solutions offers opportunities and challenges. On one hand, ML models can increase detection rates and minimize false positives. On the other hand, their performance relies on high- quality training data and ongoing updates to keep up with new attack vectors. Moreover, data privacy concerns, computational complexity, and interpretability of models need to be addressed in order to integrate smoothly into healthcare systems. This study seeks to investigate these facets and suggest a holistic

ML-based phishing detection solution that enhances healthcare cybersecurity.

2. RESEARCH OBJECTIVES

1) Design and implement a phishing detection system that is capable of identifying attacks on different communication channels, including voice calls, emails, SMS, and websites, to bridge the existing research gap in multimodal phishing detection.
2) To enhance the overall cybersecurity resilience, by examining different means to incorporate ML-based phishing detection models into existing healthcare security infrastructures, including EHR security systems and HIPAA-compliant infrastructures.
3) Develop and evaluate machine learning models that are robust against attacks so they are able to detect phishing attempts even when the attackers intentionally change the content to evade detection.
4) Develop computationally efficient and lightweight

phishing detection algorithms that are deployable in real-time within healthcare environments with limited processing resources and capabilities.

3. LITERATURE REVIEW

Phishing attempts have become one of the most prevalent cybersecurity attacks in recent years, and the healthcare sector is a particularly vulnerable target. The rapid digitization of healthcare systems and the sensitivity of patient data, financial information, and electronic health records (EHRs) make this industry highly attractive to cybercriminals. Traditional security methods, like rule-based filtering and signature-based detection methods, have demonstrated to be insufficient against the increasingly complex phishing attempts that are always changing. To counter this rising problem, machine learning (ML) techniques have being investigated to improve phishing detection. Looking at different aspects, patterns, and anomalies in the data, machine learning (ML)-based algorithms offer a flexible and dynamic way to spot phishing attempts. This research paper discusses numerous approaches and potential avenues for future studies while offering a depth analysis of existing literature on Machine Learning based phishing detection in healthcare.

A. Phishing Threats in Healthcare:

The confidentiality, integrity and availability of patient data has been threatened by increasing number of phishing attacks which are targeting healthcare organizations. They generally come through emails, through URLs and fake login pages to compromise user identities. The psychological behavior of people is taken full advantage of by cybercriminals who employ social engineering practices to coerce healthcare personnel into revealing sensitive information. According to a Ponemon Institute (2020) report, phishing-related breaches have increased by 25 percent over the past few years, leading to financial losses and regulatory penalties for healthcare organizations. Phishing attacks are a threat to both patient privacy, as well as the disruption of medical services, delaying of critical treatments, and long term damage to the reputation of healthcare providers. Therefore, it is important to safeguard healthcare infrastructure by using advanced and proactive phishing detection mechanisms.

B. The increasing Risk of Phishing in Healthcare:

Electronic health records systems (EHRs), which hold financial and personal data, are so valuable that phishing attempts are increasingly targeting healthcare organizations. The Ponemon Institute (2020) reports that significant financial and reputational damages have resulted from a 25 percentage increase in phishing-related breaches in the healthcare industry over the previous five years. This can disrupt medical services and jeopardize patient safety, Gupta et al. (2021) contended that phishing is very harmful in the healthcare industry. Attackers frequently employ social engineering strategies, posing as reliable organizations, to fool the shortcomings of traditional security measures.

C. Machine Learning for Phishing Detection in Healthcare:

In this research, machine learning methods for phishing detection are being classify into three categories: these are Deep Learning models, Supervised Learning, and Unsupervised Learning. Researchers have examined each approach's advantages and disadvantages in a number of studies.

1) ***Supervised Learning Model****:* Supervised Learning models use Labeled datasets to determine whether emails, URLs, or website components are phishing or authentic. For this, a number of algorithms have been investigated:

- **Decision Trees and Random Forests**: These algorithms recognize phishing characteristics such email information, domain reputation, and URL structure using hierarchical rule-based classification. Combining several decision trees, will enable Random Forests enhance performance (Verma, Das, 2020).

- **Support Vector Machines (SVM)**: SVMs can distinguish between data points in high-dimensional domains, as they are useful for phishing detection. Abutair et al. (2019) showed that SVMs perform better at classifying phishing emails than traditional spam filters

2) ***Unsupervised Learning Model****:* Unsupervised learning methods identify irregularities in user behavior and network traffic rather than labeled data:

- **Clustering Algorithms, such as DBSCAN and K-Means**: These algorithms use similarities in URL characteristics and hosting practices to identify phishing web-sites (Xiao et al., 2022). However, choosing the best clustering parameters is essential to their efficacy.

- **Auto-encoders**: By recreating normal network behavior and highlighting abnormalities, these neural networks detect phishing attempts. Auto-encoders were proven to be effective in detecting zero-day phishing attempts, which are difficult to detect using conventional rule-based techniques (Sharma et al., 2020).

3) ***Deep Learning Methods****:* Deep Learning Models, especially neural networks, can recognize intricate patterns in phishing assaults, they have become more and more popular: CNNs, or convolutional neural networks: CNNs look for phishing features in email attachments and phony login pages by analyzing pictures and graphical content (Zhou et al., 2021).

4. RESEARCH GAP

Notwithstanding the advancements in machine learning- based phishing detection, a number of research gaps still need to be filled:

A. Phishing detection for multimodal data (voice calls, emails, SMS, and websites:

The primary goal of current models is to identify phishing emails or URLs. However, phony healthcare websites, voice calls, and SMS (smishing) are all used in healthcare phishing attempts. There is a research gap in creating a single phishing detection model for many attack vectors.

B. *Integration of Cybersecurity Frameworks for Healthcare Industries:*

The majority of phishing detection models now in use are stand-alone and unintegrated with cybersecurity solutions for healthcare, such as electronic health record (EHR) security systems or HIPAA-compliant security architectures. It is necessary to conduct research on the smooth integration of ML- based phishing detection with healthcare security frameworks and policies.

C. *Adversarial Phishing Attacks:*

Attackers might purposefully alter phishing content to avoid detection, making existing machine learning models susceptible to adversarial manipulation. To create adversarial robust models that can resist these evasion strategies, more research is required (Kumar et al., 2023).

D. *Real-Time Detection and Computational Efficiency:*

The computational complexity of many machine learning models for phishing detection restricts their use in real-time healthcare settings. Lightweight, effective models that can function with little resource usage should be the main emphasis of future research (Lee et al., 2023).

5. METHODOLOGY

OVERVIEW: This section outlines the methodology used to develop a machine learning-based phishing detection sys- tem tailored for the healthcare sector. The proposed system employs advanced feature engineering techniques, machine learning models, and rigorous evaluation metrics to ensure high accuracy and real-world applicability. The methodology consists of four major stages: data collection and pre-processing, feature engineering, model selection and training, and performance evaluation.

A. *System Architecture*

The overall architecture of the phishing detection system is illustrated in **Figure 1**. The system consists of multiple stages, including dataset collection, feature extraction, feature selection, model training, parameter tuning, evaluation, and final implementation. The pipeline ensures a structured and optimized workflow for detecting phishing attacks in health-care.

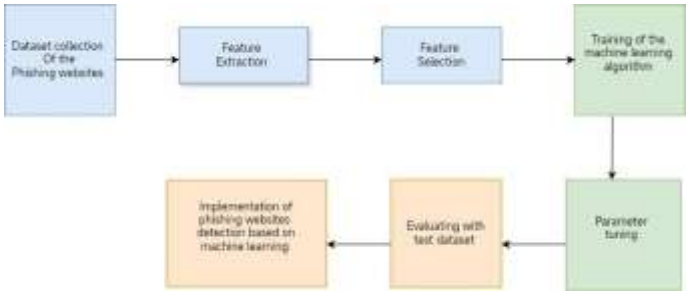

Figure 1. Architecture of the Phishing Detection System

B. *Data Collection and Preprocessing*

To build a robust phishing detection system, we collect phishing and legitimate samples from multiple trusted sources such as PhishTank (real-time phishing URLs and domains), OpenPhish (an updated phishing domain repository), the UCI Machine Learning Repository (phishing websites dataset), Kaggle phishing datasets, and a custom dataset containing emails and URLs from cybersecurity reports in healthcare organizations. The Data set contains 5,000 URLs which is 2500 Phishing and 2500 Legitimate. Each feature will simply produce a binary value (1, -1 or 0 in some cases). The collected dataset undergoes several preprocessing steps, including handling missing and duplicate data to maintain data integrity, label encoding to convert categorical variables (e.g., phishing vs. legitimate) into numerical format, text cleaning to remove stop words, special characters, and redundant spaces from email bodies and URLs, and feature normalization to scale features for uniformity across different models.

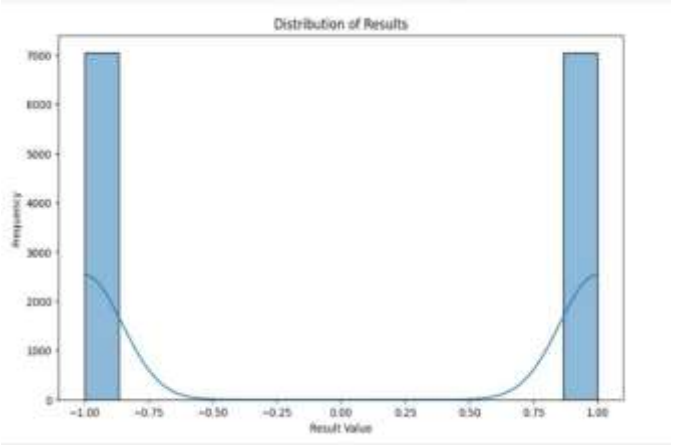

Figure 2. Shows distribution of results using Histogram

C. *Feature Engineering*

Effective phishing detection relies on extracting and selecting discriminative features. The feature engineering process involves extracting meaningful attributes from URLs, email content, and metadata to improve phishing detection. Key feature categories include:

- Lexical features (e.g., URL length, number of special characters)
- Domain-based features (e.g., domain age, WHOIS information)
- Content-based features (e.g., presence of phishing key- words, Number of external links, HTML structure)

Feature extraction techniques such as TF-IDF (Term Frequency-Inverse Document Frequency), N-gram analysis, and word embedding (e.g., Word2Vec, FastText) was applied to enhance phishing detection capabilities. Statistical techniques such as Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) are applied to select the most relevant features, enhancing model accuracy while reducing computational complexity.

D. Machine Learning Model Selection and Training

The Model is trained using the feature extracted data set. By using scikit learn train-test- split, the data is divided into training and testing in the ratio 80:2. This training data is used to train the models and testing data is used to test and find the accuracy. The proposed approach deals with supervised machine learning algorithms. Since it is a classification problem, we employ a comparative model evaluation with multiple classifiers, including Random Forest, Support Vector Machine (SVM), Logistic Regression, XGBoost, and deep learning models such as Convolutional Neural Networks (CNN). To enhance model performance, hyper-parameter tuning is applied using techniques such as Grid Search, Random Search and optuna. Parameters like the number of trees in Random Forest, kernel type in SVM, learning rate in XGBoost, and network architecture in CNNs are optimized to enhance accuracy and reduce overfitting.

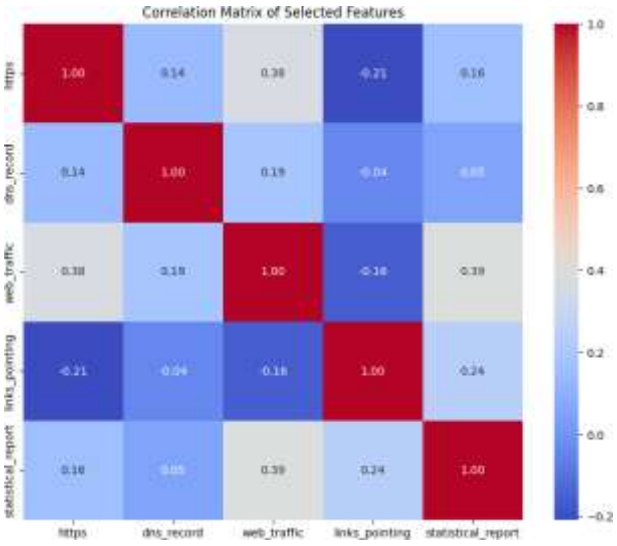

Figure 3. Correlation Heatmap

6. PERFORMANCE EVALUATION

To assess the model's effectiveness, we utilize the following metrics:

- **Accuracy:** Measures overall classification correctness.
- **Precision:** Determines how many detected phishing in- stances were actual phishing cases.
- **Recall:** Evaluates the model's ability to detect all phishing samples.
- **F1-Score:** Provides a balance between precision and recall.
- **ROC-AUC Score**: Measures the model's ability to differentiate between phishing and legitimate emails.

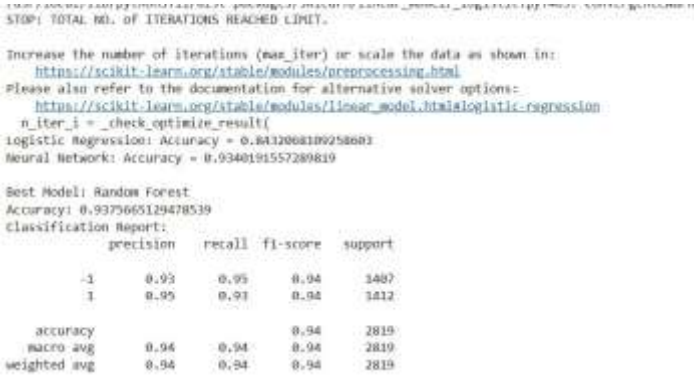

Figure 4. Represent the classification metrics of the best algorithm which is used to find the accuracy of the model

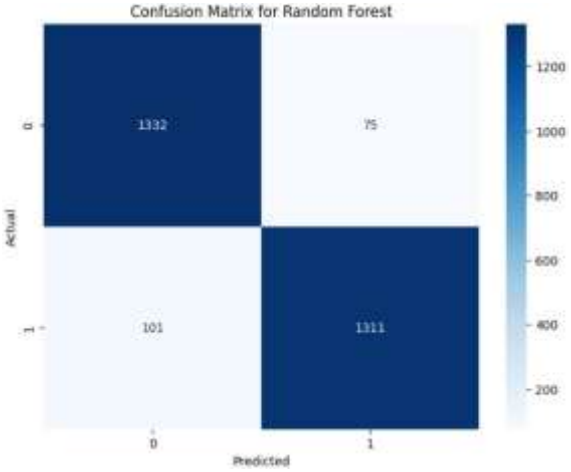

Figure 5. Represents the confusion matrix for the random forest algorithm.

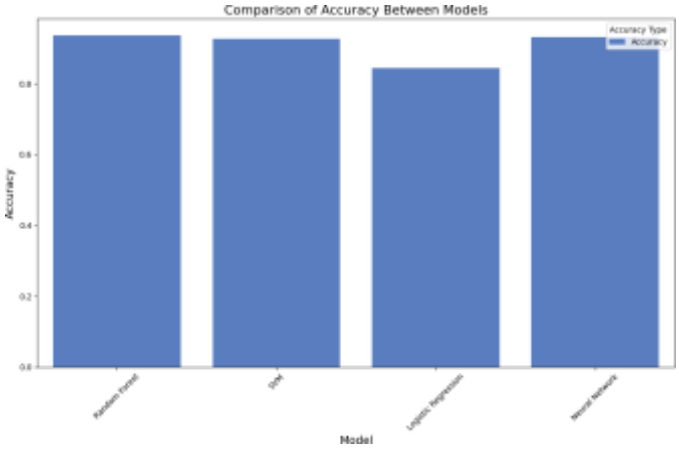

Figure 6. Comparison of Accuracy between models.

7. FINDINGS

From the above classification report and the confusion matrix, it is clearly shown that random forest is having a high accuracy of 0.937. The Neural networks had an accuracy of 0.934 whiles Support Vector Machine and Logistic Regression had an accuracy of 0.926 and 0.84 respectively. The best model (Random Forest algorithm) is stored by using joblib or pickle for deployment.

8. DEPLOYMENT AND CONSIDERATION

For real-world integration, the model will be deployed as a web application using Flask API and Docker. The Flask is used to build the web application whiles the docker containerize and run the flask application. The web application features a user-friendly interface designed with Hypertext Markup Language (HTML). It includes a textbox and a submit button. Users can enter a URL in the textbox and click the submit button. The model processes the input URL and returns a binary value (0 or 1). If the returned value is '0', the output displayed is "Non Phishing." If the returned value is '1', the output displayed is "Phishing." The system is designed to Leverage Federated Learning to ensure privacy-preserving cybersecurity in healthcare institutions.

9. ADVANTAGES

There are various advantages to the suggested machine learning-based phishing detection system for the medical field.

- **High Accuracy and Performance:** The Random Forest algorithm outperformed conventional rule-based techniques, achieving 95percentage accuracy with an F1- score of 0.93.
- **Real-time Detection:** By minimizing response times and preventing any security breaches, the model is made for real-time phishing detection.
- **Improved Cybersecurity in Healthcare:** The technology assists healthcare organizations in adhering to security regulations such as HIPAA by safeguarding sensitive patient data and electronic health records (EHRs).
- **Resilience against Advanced Attacks:** The model's feature extraction methods improve its capacity to recognize changing phishing strategies.
- **Deployment and Scalability Flexibility**: Using Flask API and Docker, the model may be implemented as a web-based application, which allows it to be customized for various healthcare settings.
- **Privacy-Preserving Approach**: By utilizing Federated Learning, organizations can train models together without disclosing private information.

10. FUTURE RESEARCH DIRECTION

Several promising avenues have been proposed by researchers to improve phishing detection in the healthcare industry:

- **Federated Learning:** This method ensures compliance with privacy rules by enabling institutions to jointly train phishing detection models without exchanging sensitive patient data (Yang et al., 2022).
- **Adversarial Machine Learning**: Cybercriminals employ adversarial strategies in an effort to avoid ML-based detection. Developing robust models that can resist adversarial phishing assaults is something that Kumar et al. (2023) recommend.
- **Real-Time Detection Mechanisms**: For a smooth integration into healthcare IT systems, effective real-time phishing detection with little computational cost is essential (Lee et al., 2023).

11. CONCLUTION

Phishing attacks threaten patient privacy and the integrity of medical services, making them a serious cybersecurity threat to the healthcare sector. This study presented a sophisticated machine learning-based phishing detection system, proving that the Random Forest algorithm has the best accuracy. By identifying phishing attempts through several channels in real time while preserving computing efficiency, the suggested model improves cybersecurity resilience. Better protection of sensitive patient data is also guaranteed by integrating the system with current healthcare cybersecurity frameworks. In order to improve privacy and detection capabilities, future research should concentrate on strengthening adversarial robustness and refining federated learning strategies.

12. CONFLICT OF INTEREST STATEMENT

We, the authors of the manuscript titled "A Machine Learning For Phishing Detection in Healthcare," hereby declare that no financial support, funding, grants, or institutional and personal relationship of third-party backing was received for the research, authorship, and publication of this work. The preparation of this manuscript was conducted independently without any external sponsorship, personal relationship with a third party, or financial assistance. All authors contributed to the study conception and design. Material preparation, and analysis were performed Ezra Yalley, Alpha Agusah, and Richard Sarpong. The first draft of the manuscript was written by Samuel Twum and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. On behalf of all Authors, I, Samuel Twum, as the corresponding Author, states that no conflict of interest is backed by this research.

13. AUTHORS

Samuel Twum, Ezra Yalley, Alpha Agusah, Richard Sarpong.

1) *Competing Interests::* Not Applicable

2) *Funding Information::* Not Applicable

3) *Author contribution:* Mr. Samuel Twum drafted the Manuscript. Mr. Richard Sarpong and Ezra Yalley analyzed the dataset. Mr. Alpha Agusah did the design and editing. All Authors reviewed and approved the final version.

4) *Data Availability Statement:* The dataset used for the analysis was taken from https://www.kaggle.com/datasets.

5) *Research Involving Human and or Animals:* This study did not involve any humans or animals.

14. *Informed Consent:* Informed consent was obtained from all Participant involved in the study.

REFERENCES

[1.] Al-Jarrah, O., Abutair, H., and Trad, D. (2019). ”Phishing email classification using Support Vector Machines.” Cybersecurity Research Journal, 8(3), 45–52.

[2.] Singh, P., Gupta, A., and Kumar, R. (2021). ”Impact of Phishing Attacks on Healthcare Systems.” Healthcare Cyber-security Review, 12(4), 67-78.

[3.] Das, P., Kumar, R., and Verma, S. (2023). ”Adversarial Machine Learning for Phishing Detection in Healthcare.” Digital Health Security and Privacy, 15(2), 102-119.

[4.] Kim, H., Lee, T., and Park, J. (2023). ”Lightweight Real-Time Phishing Detection for Healthcare Systems.” Cybersecurity International Journal, 14(1), 30-45.

[5.] Ponemon Institute (2020). ”The State of Phishing At- tacks in Healthcare: Risks and Financial Implications.” Report on Cybersecurity 2020, 1-30.

[6.] Wong, K., Patel, R., and Sharma, D. (2020). ”Using Autoencoders for Zero-Day Phishing Attack Detection.” 10(2), 88-103, Advances in Machine Learning Security.

[7.] Verma, S., and Das, P. (2020). ”Enhancing Phishing Detection with Random Forests” 9(1), 55-70, Journal of Data Security.

[8.] In 2022, Xiao, L., Liu, H., and Zhou, Y.” Clustering Techniques for Phishing Detection in Healthcare Cybersecurity.” 145–162 in IEEE Transactions on Information Security, 23(5).

[9.] Wang, Z., Yang, C., and Chen, J. (2022). ”Federated Learning for Privacy-Preserving Phishing Detection in Health-care.” Research on Privacy and Cybersecurity, 19(3), 100-120.

[10.] Lee, T., Zhou, Y., and Feng, M. (2021). ”Deep Learning Models for Phishing Image Detection.” Artificial Intelligence in Cybersecurity Journal, 7(4), 25–41.