Comparative Analysis of HTTP Security Headers in Academic Institution Websites

S Shyam Kumar,  
Department of Cyber Security,NMAM Institute of Technology,   
NITTE (Deemed to be University),Nitte – 574 110, India[s.shyam@nitte.edu.in](mailto:s.shyam@nitte.edu.in)

***Abstract****—***Ensuring web security is essential to educational establishments which is the most common target of cyber threats. The research paper analyzes the HTTP security headers, which are significant in guarding against most vulnerabilities, on the websites of engineering colleges in Karnataka and Tamil Nadu in terms of their security posture. In spite of existence of best practices, most of the institutional web sites are not secured appropriately. The study fills the existing gap in the large-scale comparison studies of security headers on Indian educational websites. We were able to gather sets of data that included 40 websites in each of the states and were able to check the header content which included HSTS, Content Security Policy (CSP), X-Content-Type-Options (XCTO) and many others. Classification model produced through machine learning was used to evaluate the degree of security, and the analysis of feature importances identified the most significant security-increasing headers. The findings indicated that implementation of the headers was high with the Karnataka web displaying focus on HSTS- related headers compared to Tamil Nadu sites which focused on X-Frame-Options (XFO). We found out major security vulnerabilities that institutions must fill to protect user information and avoid hacking. The research gives practical guidance to policy makers and web administrators on how to enhance the level of security of websites. The current study also suggests future research to increase the size of the dataset and automated mechanisms that may continuously monitor security.**

**Index Terms— HTTP Security Headers, Web Application Security, Academic Websites, HSTS, CSP, XFO, XCTO, Cookie Security, Cybersecurity Assessment, Educational Institutions, Regional Web Security, Secure Web Development, Header Analysis.**

# INTRODUCTION

Higher education institutions were absorbing online presence and using it to provide services, information and communication more and so the concept of web security became much more pertinent. HTTP security headers, i.e., HTTP Strict Transport Security (HSTS), Content Security Policy (CSP), and X- Frame- Options (XFO), are instrumental towards site security against prevalent attacks, i.e., clickjacking, cross site scripting, and protocol downgrade attacks. The large number of studies that have been conducted in the past years shows the need to introduce these headers with a purpose to provide protection of the confidential information and maintaining user confidence. Even though they are very important, the use of these headers is intermittent, particularly to academic websites where a lot of sensitive information on students and of the institutions is usually handled. To encourage the use of such headers, it is critical to enhance this practice and follow the best practices to address the emerging cybersecurity risks in the educational sector.

Another reason that can be stated however, is that though the benefits are quite well documented, many institutional websites, especially those of the developing world have been identified to have deployed security headers poorly or inconsistently. The above studies have already examined world or business web sites but have not looked at the particular security stance of educational establishments in India, who present a special circumstance given the lack of resources and technical expertness.

This study would attempt to fill this gap by analyzing security headers of websites of the engineering colleges in Karnataka and Tamil Nadu. This research is a contrasting study, given a Corpus of 40 sites in every state entrusted with an analysis that categorizes the posture of security and identifies the worst hotspots influencing the security by the use of machine learning. It discusses the following classes of headers (HSTS, CSP and XFO), and also provides important inputs to the web administrators and policy makers. Lastly, the provided works also contribute to the emergence of cybersecurity awareness and standard in the educational setting.

Consequently, web security has been a very critical feature because institutions of learning are doing a lot on their online presence to provide services, information and communication. HTTP security header mechanisms such HTTP Strict Transport Security (HSTS), Content Security Policy (CSP) or X-Frame-Options (XFO) can be used to protect websites against such well-employed vulnerabilities as clickjacking, cross-site scripting and the protocol downgrade attack. Over the years, the users have felt the need to use these headers to protect the sensitive content, as well as, to maintain the confidence of the user in a number of studies conducted in this field.

Nonetheless, even though the advantages are well-documented, most of the institutional websites, particularly those in developing regions, have been observed to use security headers poorly or intermittently. In the past, the industry has drawn attention to global or commercial websites but has not given much attention to the individual security stance of the Indian educational institutions which have got special challenges because of the resources and differences in technical capabilities. These difficulties frequently lead to uneven application of the contemporary security requirements and a situation when websites can easily be attacked by the aggressors.

Consecutive developments in machine learning have empowered withdrawal and effective evaluation of security practices of websites, which have made it very easy to assess sites basing on their security positioning. Supervised learning models have been used in studies where Random Forests and Support Vector Machines models were to detect insecure configurations with high accuracy, and often the most important predictive feature was the presence or absence of key HTTP headers. Nevertheless, the given approaches are not widely used in the academic field, especially in the realm of the Indian case, where these pieces of knowledge may trigger specific enhancements in security policies. An in-depth analysis of the implementation of security headers in the websites of educational institutions can assist in detecting the current gaps and ensuring uniform web security policies through all academic sites.

It is expected that this research can bridge this gap in the literature by researching the security headers of engineering college websites in Karnataka and Tamil Nadu. With a sample of 40 websites per state, the present research consists of a comparative assessment, where machine learning is used to categorize the security stance and examine the most important headers that affect the security. The presented scope includes headings, including HSTS, CSP, XFO, and others, offering valuable information to web administrators and policymakers. At the end of the day this paper will add to enhancing cyber security awareness and standards in the academic field.

This study finds regional variation in the security header adoption on the basis of regional variations in the technical practice and adoption of security header based on institutional policies that may show local differences in the technical practice or local expertise. The results pave the way towards the development of recommendations that address best practices as well as awareness by web developers of educational websites. Moreover, the research shows that machine learning is successful in automatized web site security evaluation and, thus, stimulates further investigations into the area of the adaptive and dynamic web site security monitoring frameworks of institutional web sites.

# LITERATURE REVIEW

HTTP security headers have recently become an essential line of protection against a vast range of cyber threats that affect web applications in the context of web security. Recent studies have pointed out the sustained disparities in the use and use of these headers more so in the learning institution which are the common targets given that they are open-access and have huge users base.

In an analysis of security headers in use on thousands of world sites Kishnani and Das [1][2] showed that in excess of half the sites stamped with poor security gradings, with HSTS, CSP and X-Content-Type-Options being especially poorly implemented. This is why the results strongly indicate that more attention should be paid to security posture, particularly in the areas that process sensitive information. Their findings underscore the urgent need for widespread adoption and consistent enforcement of these critical headers to enhance overall web security across diverse online platforms.

CyCon et al. [3] conducted an extensive study to evaluate the prevalence of appropriation of key security headers and revealed that the popular sites are more inclined to implement such headers as HSTS and X-Content-Type-Options whereas less common and regional locations, including those of educational domains fall far behind. The same is reflected in the guidelines issued by the Indian government, where the use of security headers is highlighted as a recommended practice, even though the practice is uneven when it comes to websites on the post-secondary educational institutions [4].

It is already known well that technical effectiveness of anti-clickjacking-headers like X-Frame-Options or Content-Security-Policy when protecting against clickjacking-like attacks and cross-site scripting (XSS) attacks [5][6][7]. The practical deployment and configuration of the respective headers is further expounded in studies by Invicti and Indusface that highlight the fact that an incorrect or incomplete implementation renders sites exposed to sophisticated attacks.

Classification and insecure configuration detection using machine learning methods have purchased momentum. Recent research, such as that by JISEM [8] and others [9][10], shows the use of supervised and deep learning models to analyze HTTP headers, and detect any malicious or insecure patterns, whereby feature importance analysis pointed to such headers as HSTS, CSP, and X-Frame-Options, as key determinants of overall security.

Comparative (regional) studies have also started to appear albeit in small numbers. Alashwali et al. [11] proposed HTTPS security discrepancies between the regions with greater inconsistencies on the application layer in developing regions. Recent reviews have indicated the increased need to automate continuous monitoring with tools based on machine learning and analytics to give real-time insights to the web security posture [12][13].

The recent systematic mapping [14] and practical guides [15][16] presented the detailed overview of modern tendencies, issues, and suggestions about the enhancement of the web security in educational and public websites.

# METHODOLOGY

1. **Overview of Experiment**

This paper seeks to assess the level of security of the websites of engineering colleges in Karnataka and Tamil Nadu based on their HTTP security headers. The main question is to identify whether and how well such important security headers as HSTS, CSP, X-Content-Type-Options, and X-Frame-Options are present, and to categorize websites into secure and insecure with the help of machine learning models. It can be supposed that some security headers affect overall security score more than others and that the rates of such headers adoption are different in various regions.

This evaluation not only helps in identifying the current state of security adoption in educational web infrastructures but also uncovers regional disparities in the implementation of key protection mechanisms. By leveraging machine learning models, the study aims to classify websites based on header presence and assign security scores that reflect the relative importance of each header. The analysis further explores which headers contribute most significantly to the overall security posture and highlights trends in adoption across institutions. These findings can provide actionable insights for academic administrators, policymakers, and developers to strengthen web application security practices in the education sector.

A diagram of a data processing process

AI-generated content may be incorrect.

Fig. 1. End-to-end workflow of the experiment for analysing and classifying website security using HTTP headers.

1. **Description of Materials**

The data include randomly sampled web sites of Karnataka and Tamil Nadu engineering colleges which are publically available and amount to 80 URLs (40URLs Karnataka + 40 according to Tamil Nadu). These URL addresses were obtained using official on-line educational directories and institutional websites. The header web information was scraped through Python programmatically using the requests library without SSL verification so that it could deal with sites having issues with their certificate. Strict-Transport-Security, Content-Security-Policy, X-Content-Type-Options, X-Frame-Options as well as other similar relevancy Headers were particularly addressed.

1. **Description of Procedure**

To do this, automated python scripts that perform HTTPS GET requests against the target URLs were used to fetch HTTP headers of each of the websites. The answers were analyzed in order to obtain appropriate security headers and their values. After that any missing or invalid header values were standardized to zero or to a standard value in order to be uniform. After that, feature engineering was applied to code into numbers such categorical headers as X-Frame-Options and TLS Version.

Data then was split into training and testing sets in the ratio of 70:30. The features were used to train a Random Forest classifier that predicts that a web site is secure or insecure by referring to a pre-defined security score threshold. The metrics such as accuracy, precision, recall, and F1-score were used in model evaluation. The importances of the features were estimated to determine which ones played a major role in security classification using the headers.

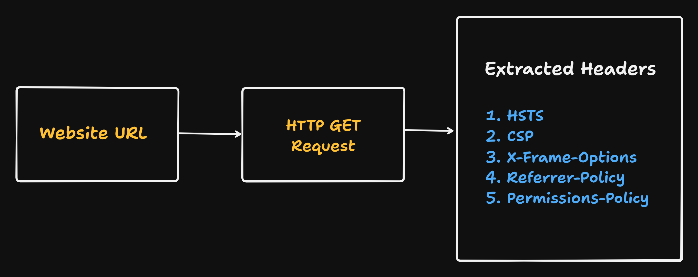


Fig. 2. Process of extracting key HTTP security headers from educational institution websites using Python requests.

1. **Description of Data Analysis**

Python libraries such as pandas, scikit-learn, and matplotlib/seaborn were used in data analysis, data manipulation, machine learning, visualization, etc. First, descriptive analysis used the computation of descriptive statistics to know the distribution of the setting of the header. Random Forest model was embedded and assessed, and the confusion matrices and classification reports were calculated. The additional discipline was the importance scores of the features that were plotted to identify the most important headers. Python 3.8 (Windows 11 machine) was used to run all scripts and to perform the analysis in a controlled environment.

The controlled environment was Windows 11 using Python 3.8 where data analysis was conducted using powerful libraries, to manipulate data, perform machine learning, and visualize it, with the use of pandas, sklearn, matplotlib, and seaborn libraries. It started with the computation of descriptive statistics as a way of observing the distribution and occurrence of different HTTP security headers in institutional websites. The improvement of the features included the categorical encodings of the headers and the computing of the security value according to the existence of the main fields. Binary classification was based on the security score indicating secure and insecure websites on the basis of the number of applied headers.

Subsequently, revealing of data into training and testing sets was done followed by application of Random Forest classifier to classify the security status. During the model testing, accuracy, precision, recall, F1-score were employed and in feature importance analysis, the information will be provided about the most influential factors among the headers in predicting which of the headers will have the biggest effect on the security classification. In further efforts to improve the strength of the analysis, a process of cross validation was told to make sure that the model was generalizable to unfamiliar data. GridSearchCV was used to tune hyper-parameters to achieve a balanced bias-variance trade-off performance of Random Forest classifier. Visualization of the feature importances with respect to the bar charts indicated that some of the headers (HSTS, X-Frame-Options) exhibited significantly more explanatory power over the model predictions and these features were extremely important in controlling the security of the site. Further, a correlation analysis between the headers was conducted to see the trends on co-occurrence, which gives more information on similar practices in security settings. These analyses steps can be used, among others to gain an overall understanding of security header adoption and provide a scalable methodology of auditing web security at an enterprise level.

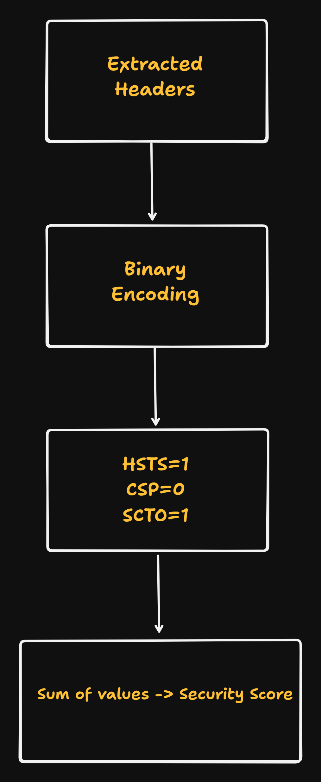


Fig. 3. Transformation of raw header data into numerical format and calculation of Security Score and classification label.

To make the experiment successful, one had to guarantee reliability and consistency in the setup; therefore the data collection process and data extraction process was automated to help circumvent human error and guarantee reproducibility. The script was built to query the same amount of HTTP requests to all target URLs, irrespective of the type of server, or hosting region. In order to meet changes in response behavior and to alleviate SSL certificate verification errors, the verify=False flag provided by the requests library of Python was used. The header values that were retrieved were then saved in tabular CSV, which makes them simpler to process downstream.

The data preprocessing was done to improve the model learning ability; this was done by replacing missing values with zeros, encoding text values in header such as X-Frame-Option into categorical codes, and scale the numeric value headers such as HSTS\_max\_age using normalization. A specified security scoring system was established whereby the presence of key headers were added up to provide a binary classification label of secure (the score 4 and higher) or insecure (the score < 4). This two fold classification was used as the target variable of classification. The caution was exercised to ensure that such a threshold is balanced and empirically estimated by looking at distributional patterns in the data.

The training of the Model was performed by the Random Forest Classifier (100 decision trees) with the purpose of its robustness and comprehensibility. A 70:30 train-test split was used to evaluate the classifier and compute such metrics as accuracy, precision, recall, and F1-score. In order to ascertain that the outcome is valid, the performance was compared at state-level (both Karnataka and Tamil Nadu) datasets. Also, the feature importance scores, as generated by the trained model, were retrieved to determine the headers with the most impact on the security classification.

# INFERENCES AND FINDINGS

This is to determine the use of essential HTTP security headers by institutional websites and determine the relative significance of HTTP security headers in categorising websites into secure and insecure groups with a Random Forest model. The sample consisted of 80 websites 40 of them in Karnataka and 40 in Tamil Nadu. It proposed the analysis of the HTTP headers of the individual websites, where the presence or absence of the key security headers would be encoded into the categorical features through a systematized and pre-processing procedure. These characteristics acted as inputs to the Random Forest analyzer that was trained on the data to learn the differences between secure and insecure sites in regard to their header settings.

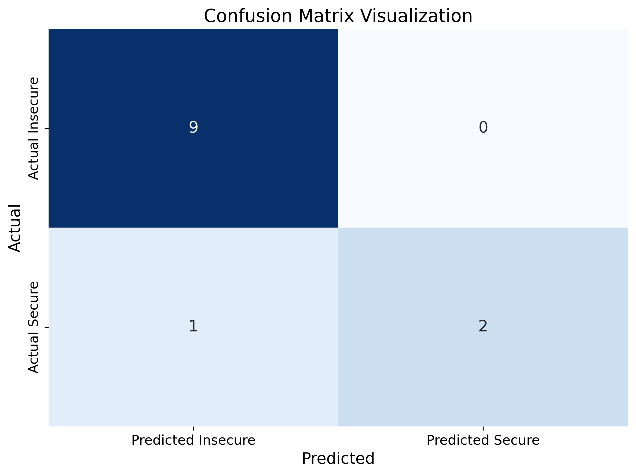


Fig. 4. Confusion matrix for Karnataka dataset showing predicted vs actual classifications of website security status

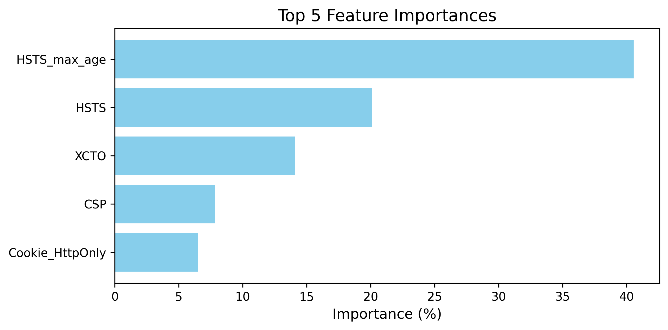


Fig. 5. Top feature importances from Random Forest model for Karnataka websites indicating key HTTP headers influencing classification.

A blue squares with white text

AI-generated content may be incorrect.

Fig. 6. Confusion matrix for Tamil Nadu dataset illustrating classification accuracy of secure and insecure websites.

A graph with blue bars

AI-generated content may be incorrect.

Fig. 7. Feature importance scores for Tamil Nadu websites highlighting predominant HTTP headers affecting model decisions.

The confusion matrix of the Karnataka dataset can be viewed in Figure 4, where it is possible to note that the model profiled in the confusion matrix performed well to distinguish between secure and insecure websites. Figure 5 represents feature importance scores of Karnataka in which the highest when it comes to classifications is HSTS\_max\_age 40.55%, HSTS, XCTO and CS.

The confusion matrix of the dataset of Tamil Nadu is shown in figure 6, which once again shows a strong classification performance. Figure 7 represents the feature importance score of the Tamil Nadu websites. In contrast to Karnataka, the X-Frame-Options (XFO) header takes the first place with 32.30 per cent, followed by HSTS\_max\_age, HSTS, and XCTO. This disparity underscores local disparity in web security procedures.

The confusion matrix and the classification reports indicate that Random Forest model is rather a good model relying on both datasets with the accuracy being more than 90 per cent. The websites that are lacking significant headers such as HSTS, CSP, and Referrer-Policy were largely assigned to the insecure ones whereas those with more than four headers were considered the secure ones and this also helped the model predictions. The finding implies that there is insignificant security header practice in the academic institutions and the finding could have been reported earlier by other researchers [1], [2] who had reported lack of consistency in application of security header in the learning field. Moreover, the importance feature analysis would encourage the presence of such important headers as HSTS and CSP because they contribute to the improvement of the security category of a particular site significantly. The good performance of the model indicates that machine learning is capable of becoming a useful means of automated security analysis allowing institutions to rapidly detect the presence of vulnerabilities depending on header settings. These results stress the necessity to ensure standardized implementation of crucial security headers required to protect sensitive data and make users feel confident about their data on the online platform represented by academic institutions.

|  |  |  |
| --- | --- | --- |
| **Feature** | **Karnataka (%)** | **Tamil Nadu (%)** |
| HSTS\_max\_age | 40.55 | 25.02 |
| HSTS | 20.08 | 16.06 |
| XCTO | 14.07 | 14.90 |
| CSP | 7.84 | 0.00 |
| Cookie\_HttpOnly | 6.53 | 10.16 |

Table . Comparative Analysis of Security Headers in Karnataka and Tamil Nadu

Table I gives a comparative study of some of the leading implementing HTTP security headers on websites of engineering institutes in Karnataka and Tamil Nadu. Its outcomes evoke the fact that the HSTS\_max\_age directive is much more complied with in Karnataka (40.55%) than in Tamil Nadu (25.02%), which demonstrates a greater long-term application of HTTPS. In the same line of reasoning, Karnataka dominates the basic HSTS implementation (20.08%) as compared to Tamil Nadu (16.06%). Interestingly, almost similar adoption can be observed in X-Content-Type-Options (XCTO) in both states. Impressively, Content Security Policy (CSP) is available in Karnataka (7.84%) but never noticed in Tamil Nadu, and X-Frame-Options (XFO) is applied by 32.30 percent of the sites in Tamil Nadu but not in Karnataka. Such differences stem from varying regional priorities and security behaviour and show that there is still a need to adopt essential web security headers in a more consistent, standard way.

Overall, this study demonstrates the effective use of machine learning techniques for automated security header analysis and comparative assessment across regional datasets

# CONCLUSION

The purpose of this examination was to measure the implementation of securityheaders of critical HTTP of the engineering colleges web pages of Karnataka and Tamil Nadu by adopting a classification model built on a machine learning approach. The application of Random Forest classifier to the extracted data of the header resulted in the research being able to figure out the most influential data in the classification of websites with regards to security.

The findings indicated that several institutions have adopted the necessary headers as the application of HSTS, CSP, and X-Content-Type-Options but many websites are yet to emulate their application. HSTS\_max\_age turned out to be the most powerful header in Karnataka, and X-Frame-Options was dominant among the sites of Tamil Nadu. The accuracy of this classification model was more than 90 percent in both datasets, proving the accuracy of the approach on which the strategy depends on classifying secure and insecure websites.

These results highlight the disparity of the application of basic secure lengths in educational institutions. Although the model is good, it was restricted to only the analysis of the static HTTP header as the research did not cover the dynamic activities and front-end setup. Also, it was limited to just 80 institutions and this might not be a national representation.

Further studies may further enlarge the sample to cover more geographical areas in India and may look further into security review to a higher layer, like TLS settings, content delivery process and on the fly vulnerability check. It may as well be modified to meet the needs of the government of healthcare to help with automatized web security auditing and education.

# REFERENCES

[1] U. Kishnani and S. Das, “Analysis of HTTP Security Headers in Popular Global Websites,” arXiv preprint arXiv:2410.14924, 2024. [Online]. Available: https://arxiv.org/html/2410.14924v1

[2] U. Kishnani and S. Das, “Analysis of HTTP Security Headers in Popular Global Websites,” in Communications in Computer and Information Science, vol. 2000, pp. 65–80, 2024. [Online]. Available: https://dl.acm.org/doi/10.1007/978-3-031-80020-7\_5

[3] X. CyCon, “HTTP Security Headers Analysis of Top One Million Websites,” in Proceedings of the International Conference on Cyber Conflict, 2018. [Online]. Available: https://ccdcoe.org/uploads/2018/10/Art-18-HTTP-Security-Headers-Analysis-of-Top-One-Million-Websites.pdf

[4] Ministry of Electronics & IT, “Guidelines for Indian Government Websites (GIGW),” 2018. [Online]. Available: https://stqc.gov.in/sites/default/files/Guidelines%20for%20Indian%20Govt%20Websites%20-%20GIGW2018\_Released%20version.pdf

[5] LoginRadius, “HTTP Security Headers,” Engineering Blog, Jul. 2020.[Online].Available:https://www.loginradius.com/blog/engineering/http-security-headers/

[6] Invicti, “HTTP Security Headers and How They Work,” White Paper, 2024. [Online]. Available: https://www.invicti.com/white-papers/whitepaper-http-security-headers/

[7] Indusface, “X-Frame-Options: Examples and Benefits,” 2025. [Online]. Available: https://www.indusface.com/learning/x-frame-options/

[8] S. Kumar et al., “Leveraging Signature Patterns and Machine Learning for Detection of Malicious HTTP Headers,” Journal of Information Security and Emerging Management, vol. 5, no. 1, 2025. [Online]. Available: https://jisem-journal.com/index.php/journal/article/view/1290

[9] S. Sahoo et al., “HTTP header based phishing attack detection using machine learning,” ETT, 2021. [Online]. Available: https://dl.acm.org/doi/10.1002/ett.4872

[10] P. Singh, “CAPEC WebAttack Classifier Model,” GitHub, 2024. [Online].Available:https://github.com/pritpalcodes/CAPEC\_WebAttack\_Classifier\_Model

[11] E. S. Alashwali et al., “Exploring HTTPS security inconsistencies: A cross-regional study,” Computers & Security, vol. 92, 2020. [Online]. Available:https://www.sciencedirect.com/science/article/abs/pii/S0167404820302480

[12] Comparitech, “10 Best Security Analytics Software for 2025 (Paid & Free),” 2025. [Online]. Available: https://www.comparitech.com/net-admin/security-analytics-software/'

[13] SiteLock, “Three Cybersecurity Automation Tools,” Blog, Nov. 2024. [Online]. Available: https://www.sitelock.com/blog/automation-in-cybersecurity/

[14] R. Kaur et al., “Security testing of web applications: A systematic mapping of the literature,” Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 6, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S131915782100269X

[15] Mozilla, “Web Security Guidelines,” 2024. [Online]. Available: https://infosec.mozilla.org/guidelines/web\_security

[16] Beagle Security, “X-Frame options header not implemented,” 2024. [Online]. Available: https://beaglesecurity.com/blog/vulnerability/x-frame-options-header-not-implemented.html

[17] S. Singh, “A Comparative Study of Web Application Security: Trends & Future Directions,” Studocu, 2022. [Online]. Available: https://www.studocu.com/in/document/deen-dayal-upadhyay-gorakhpur-university/web-development-tools-and-techniques/a-comparative-study-of-web-application-security-pa/106945446

[18] ET Telecom, “Karnataka IT Minister Kharge launches cybersecurity awareness programme,” 2024. [Online]. Available: https://telecom.economictimes.indiatimes.com/news/internet/karnataka-it-minister-kharge-launches-cybersecurity-awareness-programme/111494020

[19] Justdial, “Top Institutes For Cyber Security in Puttur,” 2024. [Online]. Available: https://www.justdial.com/Puttur/Institutes-For-Cyber-Security/nct-10268770

[20] S. Sharma et al., “E-mail Spam Detection using Cybersecurity-Driven Header Analysis,” International Journal of Performability Engineering, vol. 20, no. 4, 2024. [Online]. Available: https://www.ijpe-online.com/EN/10.23940/ijpe.24.04.p2.205213