# Literature Survey: Anomaly Detection in Web Traffic Using Machine Learning Ensembles

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Study** | **Objective** | **Dataset** | **Methods** | **Results** | **Limitations** | **Relevance** |
| **Nguyen et al.** **(2012)** | Develop an adaptive intrusion detection system (A-IDS). | CSIC-2010 | Ensemble of Naive Bayes, Bayes Network, Decision Stump, RBF Network | Achieved 90.52% accuracy, outperforming majority voting and boosting methods. | Limited scalability for large-scale data. | Shows the advantage of adaptive ensemble techniques in improving anomaly detection. |
| **Kozik et al. (2014)** | Propose an algorithm for cyberattack detection in web applications. | Simulated HTTP requests dataset | Decision Tree (C4.5), Naive Bayes, Adaboost, PART | C4.5 outperformed other classifiers in attack detection. | Only evaluated on simulated HTTP dataset. | Reinforces the use of decision trees for effective cyberattack detection. |
| **Parhizkar and Abadi (2015)** | Propose one-class SVM for web traffic anomaly detection. | CSIC-2010v2, CSIC2012 | One-class SVM | Reasonable performance in terms of TPR, FPR, and F1-score. | Limited application to HTTP-specific features. | Emphasizes the effectiveness of one-class SVM in web traffic anomaly detection. |
| **Zhang et al. (2018)** | Apply deep neural networks for network intrusion detection. | CICIDS-2017 | Deep Neural Networks (DNN) | Achieved significant improvements in accuracy and detection rates. | High computational resources required. | Proves the utility of deep learning for real-time anomaly detection in network traffic. |
|  |  |  |  |  |  |  |
| **Ali Moradi Vartouni et al. (2018)** | Propose a web application firewall using deep neural network and Isolation Forest. | CSIC 2010 | Stacked Auto-Encoder (SAE), Isolation Forest | Deep learning with feature extraction improves detection accuracy. | Struggled with high-dimensional feature sets. | Demonstrates deep learning's utility in web attack detection. |
| **Pubudu et al. (2022)** | Detect malicious traffic in IoT and local networks using ensemble learning. | UNSW-NB15, IoTID20 | Stacked ensemble classifier combining tree-based models (EBF). | High accuracy of 98.5% for binary and 98.4% for multiclass classification. | High computational cost. | Shows ensemble learning's high effectiveness in detecting malicious traffic. |
| **João B.D. Cabrera et al. (2007)** | Study distributed intrusion detection in Mobile Ad-Hoc Networks (MANETs). | Simulated MANET environment | Ensemble methods, clustering, and machine learning for local anomaly detection. | Improved detection rates through averaging in a node-cluster-manager hierarchy. | Vulnerable to communication losses in MANETs. | Highlights ensemble methods for distributed anomaly detection in wireless networks. |
| **Rajagopal et al. (2020)** | Develop a stacking ensemble model for network intrusion detection. | UNSW NB-15, UGR’16 | Stacking ensemble of classifiers (Logistic Regression, KNN, Random Forest, SVM). | 97% accuracy on real-time datasets, outperforming individual classifiers. | Less effective on emulated datasets. | Shows the benefit of stacking ensemble techniques for network intrusion detection. |
| **Imran et al. (2021)** | Design an ensemble model combining autoML and Kalman filters to improve network intrusion detection accuracy. | UNSW-NB15, CICIDS2017 | AutoML, Deep Neural Networks (DNN), Kalman filter, ensemble voting. | Achieved 98.801% accuracy on UNSW-NB15 and 97.02% accuracy on CICIDS2017 | High dimensional data increases complexity; requires large computational resources. | Highlights the effectiveness of combining autoML and Kalman filters for anomaly detection. |