**Introduction**

The rapid evolution of technology and the increasing dependence on internet-based services have led to the proliferation of cyber-attacks, particularly Distributed Denial of Service (DDoS) attacks. These attacks pose significant threats to the availability and reliability of online services by overwhelming targeted systems with a flood of malicious traffic, rendering them inaccessible to legitimate users. In recent years, the detection and mitigation of DDoS attacks have become critical areas of research within the cybersecurity domain.

The integration of Internet of Things (IoT) devices into various sectors such as healthcare, smart homes, and industrial control systems has further exacerbated the challenge of DDoS attacks. IoT devices, often characterized by limited computational power and inadequate security measures, are particularly vulnerable to such attacks​(P1\_\_ 1-s2.0-S0045790622…)​. As IoT networks continue to expand, ensuring their security has become paramount. Traditional Intrusion Detection Systems (IDS), whether signature-based or anomaly-based, often fall short in effectively identifying and mitigating sophisticated DDoS attacks due to their reliance on predefined attack patterns and inability to adapt to new attack vectors​(P1\_\_ 1-s2.0-S0045790622…)​​(P23\_\_ S0146411619050043)​.

Recent advancements in machine learning and deep learning have shown promise in enhancing the detection capabilities of cybersecurity systems. Machine learning algorithms, such as Support Vector Machine (SVM) and Random Forest, have demonstrated high accuracy in identifying DDoS attacks by learning from historical attack data and recognizing patterns indicative of malicious activity​(P5\_\_ Analysis\_and\_Detec…)​​(P6\_\_ Classification\_Bas…)​. However, these approaches often struggle with interpretability, making it difficult for cybersecurity professionals to understand and trust the decision-making process of these models.

To address these challenges, this research proposes a novel hybrid deep learning framework for the detection of DDoS attacks, integrating multiple machine learning models with explainable AI techniques. The primary objective is to enhance both the accuracy and interpretability of DDoS attack detection systems. By leveraging the strengths of Random Forest, Logistic Regression, and Neural Networks in conjunction with SHapley Additive exPlanations (SHAP), this approach aims to identify the most influential features contributing to DDoS attack detection and provide clear, actionable insights to cybersecurity analysts.

**Background**

The literature on DDoS detection using machine learning highlights the efficacy of various algorithms. SVM and Random Forest classifiers, for example, have shown notable success in different studies. SVM achieved an accuracy of 99.7% in detecting DDoS attacks in cloud environments​(P5\_\_ Analysis\_and\_Detec…)​, while Random Forest classifiers demonstrated an accuracy of 99.68%​(P6\_\_ Classification\_Bas…)​. Hybrid models, combining multiple algorithms, have been particularly effective. A hybrid approach integrating SVM and Self-Organized Map (SOM) improved detection rates and reduced false alarms​(P12\_\_ Detection\_of\_DDoS…)​. Additionally, deep learning models such as Deep Feed Forward (DFF) networks have shown superior performance, achieving accuracies as high as 99.63%​(P17\_\_ Performance\_Compa…)​.

**Motivation and Objectives**

The motivation behind this research is to overcome the limitations of existing DDoS detection methods by developing a comprehensive hybrid framework that combines machine learning models with explainable AI. The integration of SHAP helps in understanding the feature importance, thereby making the detection process transparent and trustworthy. This research aims to meticulously preprocess the dataset, including the removal of missing values, encoding of categorical variables, and normalization of numerical features. By splitting the data into training and testing subsets, and employing rigorous training protocols, the study seeks to achieve high detection accuracy.

Our Random Forest classifier demonstrated exceptional performance, achieving an accuracy of 99.96%, an F1 score of 99.96%, a precision of 100%, and a recall of 99.92%. Logistic Regression and Neural Network models also performed substantially well, with accuracies of 92.95% and 96.91%, respectively​(Abstract)​. To further enhance detection capabilities, a hybrid model approach combining Stacking, Boosting, and Voting classifiers was implemented. The Stacking model, integrating Logistic Regression, Random Forest, and Gradient Boosting as base learners with Logistic Regression as the meta-learner, achieved a superior accuracy of 99.96%​(Abstract)​. SHAP provided valuable insights into feature importance, identifying key features such as 'Flow Duration', 'Total Fwd Packets', and 'Fwd Packet Length Mean' as critical determinants in DDoS attack detection​(Abstract)​.

**Contributions**

This research significantly contributes to the field of cybersecurity by:

1. **Developing a Hybrid Deep Learning Framework**: Integrating multiple machine learning models to enhance detection accuracy.
2. **Implementing Explainable AI Techniques**: Using SHAP to provide transparency and interpretability in model decisions.
3. **Achieving High Detection Performance**: Demonstrating superior accuracy and low false positive rates through rigorous testing and validation.
4. **Identifying Key Features**: Highlighting the most influential features contributing to DDoS attack detection, thus informing future developments in threat detection systems.

The findings of this research underscore the efficacy of hybrid models in achieving high detection accuracy while maintaining interpretability, thereby enhancing the security infrastructure against DDoS attacks. This comprehensive framework not only improves detection capabilities but also provides a robust and interpretable solution that can be readily adopted in practical cybersecurity applications.

**References**

* Mihoub, A., et al. (2022). Denial of Service Attack Detection and Mitigation for Internet of Things using Looking-Back-Enabled Machine Learning Techniques. Computers and Electrical Engineering, 98, 107716.
* Bindra, N., & Sood, M. (2019). Detecting DDoS Attacks Using Machine Learning Techniques and Contemporary Intrusion Detection Dataset. Automatic Control and Computer Sciences, 53(5), 419–428.
* Tuan, T. A., et al. (2020). Performance Evaluation of Botnet DDoS Attack Detection Using Machine Learning. Evolutionary Intelligence, 13, 283–294.