Dynamic Threat Prevention in Cloud Firewalls Using Hybrid Machine Learning for IP Reputation Intelligence

Nanayakkara Wawage Chanaka Lasantha [1,*], Madduma Wellalage Pasan Maduranga [2], and Ruvan Abeysekara [3,*]

[1] Faculty of Graduate Studies, IIC University of Technology, Phnom Penh, Cambodia
[2] Department of Electrical and Electronic Engineering, University of Sri Jayewardenepura, Nugegoda, Sri Lanka
[3] Faculty of Graduate Studies, BCAS Campus, Colombo, Sri Lanka
Email: chanaka.lasantha@gmail.com (N.W.C.L.); pasanm@sjp.ac.lk (M.W.P.M.); ruvan@bcas.lk (R.A.)
*Corresponding author

***Abstract*—The existing cloud security technologies have serious problems covering the management of the altered cyber threats, so the necessity to raise the Internet Protocol Reputation (IPR) validation and more efficient threat detection tools. There are traditional IPR validation techniques that prospectively consider the defining concept of fixed blocklists and rule-based detection techniques, which generate an astronomic number of False Positives (FP) and fail to differentiate between legitimate and malicious traffic. The proposed paper will introduce a hybrid Machine-Learning (ML) system on inference of the IPR in the Amazon Web Services (AWS) cloud infrastructure, which will contribute to the improvement of the automated defence and threat intelligence. The proposed framework implies using both of the following types of logs such as GuardDuty logs, as well as Security Operations Centre (SOC) and Web Application Firewall (WAF) security logs, to be able to identify threats in real-time. A Hybrid ML advanced model with Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), K-Nearest Neighbor (KNN) and eXtreme Gradient Boosting (XGBoost) are used to determine the intensity of IP threats and automatically created blocking policies in AWS WAF. Its architecture is capable of adaptive IP-list management and live threat learning (blocking). The experimental validation of weighted F1−Scores of benign traffic classification and attacker detection is 0.97 and 0.98, respectively, with the highest accuracy of 98.04 of RF Model. The Synthetic Minority Over-Sampling Technique (SMOTE) methodology would only be used on training data, with realistic imbalanced distributions in test data being retained to ensure realistic evaluation of performance.**

***Keywords*—hybrid machine learning, IP reputations validation, cloud web application firewall, sob-30 integration, dynamically driven threat detection, automated defence systems**

I. Introduction

The development of cloud computing is the ultimate transformation in the information technology infrastructure of enterprises in terms of cost effective and scalable service provision. However, the same transformation has also raised the count of attack preponements that may be applied by bad actors, and hence the deployment of advanced security features that can deal with the evolving threat areas is demanded. This part describes the motivation of the research, problems of the research and the major contributions of this work. The rapid development of Artificial Intelligence (AI) has led to the establishment of a new challenge to developers who need to uphold the high ambition demands of web application protection since new cybercriminals are incessantly developed. The existing Web Application Firewall (WAF) products suffer exceptionally large rates of Frames Per Second (FPs), and this prevents the efficient operation of dynamic Internet Protocol Reputation (IPR) validation and threat-management service [1]. The fundamental issue to which this paper has been directed is to distinguish between real user trafficking and the presence of bot-attack traffic in the cloud by using the current ensemble Hybrid ML models that can enhance the accuracy and efficiency of dynamic IPR validation.

The paper creates a continuously real-time IPR solution that will operate with AWS WAF and GuardDuty and will be stored with the assistance of Amazon Web Services (AWS) Simple Storage Service (S3). The hybrid ML model is adaptive since it is also trained and evolved to new threats founded on real-time SOC-categorized IP address logs, on top of GuardDuty and WAF logs. Ensemble model refers to a cluster of the classifier such that the RF standard and Support Vector Machines (SVM) algorithms and this identifies suspicious IPs and avoids irrelevant findings [2]. The system was able to accurately identify a malicious IP address and block criminals in real-time due to its adaptive element. It is a compounded approach of IPR metadata management and repository synchronization, which offers enhance dynamical IPR validation efficiency to the present cloud-based systems of application security [3]. Consequently, the produced defence operations eliminate false alarms and unwanted

Manuscript received December 29, 2025; revised January 24, 2026; accepted February 16, 2026; published May 13, 2026.

responses and increase the sustainability of cyber defense in web applications.

II. Background

This part forms a background on the key ideas that one should have to grasp the proposed framework. The main terminology, technical background, and technologies are introduced to set a context in the further literature and methodology discussion. IPR validation and cybersecurity measures are complementary and thus require preventative actions when it is processed to web application security breaches. A well-implemented framework can help organizations anticipate the risk of malware infection and WAFs regulate access to the backend services. IPR monitoring is an important aspect of security systems and a key element of the defending strategy used by them. The IP traffic permissions are calculated using reputation scoring using activities such as spam Distribution, Distributed Denial of Service (DDoS) attacks, cyberbullying, and botnet operations. Unlawful actions produce a poor reputation on the IP addresses, which in turn jeopardize the web applications and data security. Addressing such threats demands accurate execution of the measures that will incorporate the existing IP-tracking systems as well as the multi-layered system of cyber-protection and permanent development of defence systems. Such protective measures are implemented in organizations to safeguard critical information and network services against the upcoming cyber-attack strategies.

The initiative-taking approaches allow companies to protect against emerging threats and stay in line with the IPR policies. Organizational cybersecurity practices combined with systematic IPR monitoring allow firms to minimize risks, protect digital assets, and enhance the general security of cyber domains [4]. The conventional IPR model uses deny lists and behavioral records stored in databases by doing a static database search [5]. This method does not satisfy the conditions of multilevel pre-validation that is required to develop accurate datasets that can be used to train ML validation models. Simple data-gathering techniques such as honeypots, spam traps, and simple event logs are common in such databases but will not give much insight [6]. Conventional IPR systems rely on the ML analysis of IPR behavior based on history which does not update on the production level systems and does not properly validate reputations leading to scores becoming out-of-date at an extremely fast rate. The precision of IPR validation systems can be enhanced using real-time continuous validation, better metadata validation, and sequential validation processes to deal with the dynamism of cyber threats [7].

III. Literature Review

The section includes a complete overview of the literature in relationship to IPR validation, ML-based threat detection, and cloud security structures. The review has logically been divided thematically into subsections supported by tabulated comparative analysis outline of the research gap that the proposed methodology attempts to address.

A. Overview of IPR Validation Research

Studies on the validation of IPR have been made over the past ten years with numerous variants of the approaches such as the ones by static denylists and more advanced ML based detection systems. Such problems as a reduction of FPs, the realization of real-time processing needs, and the control over the contemporary trends of threats are some of the most critical issues in the field. Table I illustrates a systematic review of the literature available and categorizes them as methodology, technology, and key findings.

TABLE I. Systematic Literature Review

Work	Contribution	Technology	Remarks
IP Reputation Analysis of Public Databases and ML Techniques [8]	The development of an automated IPR analyzer tool that would do cross checking against various deny-list databases and assign security scores.	Blacklist databases, Neural Networks (NN), Decision Trees (DS), Linear Regression (LR)	Examine the limitations that exist in traditional deny-list techniques, including the high prevalence of FPs as well as the substantial maintenance requirements for their implementation.
Automatic IP Blacklisting Using ML from Security Logs [9]	A ML based framework for automated blocking of IP addresses in the context of security log analysis.	LR, RF, ML-based classification	Significant cut down of erroneous blocking and restricted activity windows of malicious IP addresses have been observed. ML models now come close to the ability of domain experts in diagnosing disease
Detect Malicious IP Addresses Using Cross-Protocol Analysis [10]	The development of a large-scale classification scheme using cross-protocol based telemetry is presented that uses IPR.	DR, RF, cross-protocol telemetry, real-world IPR	Among the salient challenges facing the field are the lack of appropriately labelled data, the substantial economic burden of the problem of false positive identifications, and the inherently dynamic
Detecting Malicious Websites content by Learning IPR Features [11]	The development of a detection structure based on machine learning for web-based malware, based on the features of the IP addresses, is presented herein.	LR, NN, Naive Bayes (NB), SVM	Highlights the flexibility of attackers and limitations that exist in current methods of detection.

Intrusion Detection Using Ranked Feature Bagging [12]	This study introduces an intrusion detection system with ensemble classifiers that use ranked feature bagging to select minimal feature sets in the NSL KDD data set.	Ranked Feature Bagging, Ensemble Learning, and Feature Selection.	A classification model was achieved with 99.71% accuracy and thereby proved the efficiency of the used feature selection methodology which enhanced the classification performance.
Optimized ML-Driven IDS for IoT Applications [13]	An optimized ML-driven intrusion detection model for IoT is assessed using a comparative analysis on real-time dataset.	RF, DT, NB, K-Nearest Neighbors (KNN)	RF was found to be the best modality for complex classification, which results in better performance in IoT security.
Sea Horse Optimization with ML on Cloud [14]	A framework for IDS using Sea Horse Optimization with Deep Echo State Networks for cloud computing environments.	SHO-DESNID, min-max normalization and hyper parameter optimization	Detection rates on benchmark IDS databases were increased by using metaheuristic optimization.
Ensemble Learning for Network IDS [15]	The ML-assisted method of network IDS uses an ensemble learning model, which is enhanced by the systematic optimization of hyperparameters.	The study looks at ensemble learning algorithms, systematic hyperparameter optimization, and NSL-KDD dataset evaluation.	The systematic hyper-parameter tuning obtained an accuracy of 89.32% and a FP rate of 1.95%.
OD-IDS2022 Dataset for ML Attack Classification [16]	The new IDS dataset was created to overcome the shortcomings of the existing datasets, including twenty-eight diverse types of attacks and including extensive metadata.	A large taxonomy of attacks, SVM and the creation of a dataset.	The SVM algorithm achieved the best prediction accuracy of the algorithms assessed on the new data.
Coati Optimization for VANET IDS [17]	IDS in VANETs using Coati Optimization Algorithm and Deep Belief Networks.	COA-DBN, automobile network security, and metaheuristic optimization.	Improved detection in vehicular ad-hoc network settings.
Cloud Architecture for Network IDS Using ANN [18]	A sophisticated cloud architecture plan with improved artificial neural networks to detect network intrusions.	The use of improved artificial neural networks, support of cloud-native deployment, and privacy-conscious design principles are all the foundations of the suggested system architecture.	Discusses privacy and interpretability issues when deploying a cloud-native IDS.

TABLE II. CASE STUDIES FOR IP REPUTATION VALIDATION

Title	Contribution	Technology	Remarks
Intelligent Dynamic Malware Detection using ML in IPR for Forensics Data Analytics [19]	Suggests the hybrid model of Dynamic Malware Analysis, Cyber Threat Intelligence, ML, and Data Forensics to predict IPR.	Big Data Forensics, Machine Learning, and Cyber Threat Intelligence.	It deals with the problem of high management costs and FP rates of traditional systems.
IPR Scoring with Geo-Contextual Feature Augmentation [20]	Presents an IPR scoring model which uses geo-contextual data to support a complete evaluation of threat.	Geo-Contextual Feature Augmentation of Anomaly Detection Modelling.	It is empirically indicated that network- and geo-contextual information integration improves assessment.
Blacklist-based Malicious IPR Traffic Detection [21]	The creation of a blacklist-based mechanism of identifying malicious IP traffic through signature matching.	A basic method of modern cybersecurity systems is represented by static deny lists, which are defined by signature-based detection systems.	The baseline method has an accuracy of eighty-five, thus showing the weakness of the static methodologies.

Table I illustrates large scope of the academic base regarding the issues of IPR validation and intrusion detection, and it can be established that the following tendencies are relevant. Always the ML-based methods prove to be better than the traditional rule-based methods and the accuracy rates of the ensemble methods witness the above-89 per cent rates. Recent developments in IoT security ecosystem hybrid deep learning systems have proven the usefulness of hybrid libraries of neural networks to anomaly detectors and therefore offers a theoretical rationale behind the utilization of hybrid methodologies to IPR validation in cloud platforms frequently used on the backends of IoT systems. Moreover, the study community has come to devote more interest to the decrease in FP rates, which is still a significant issue in all the discussed methodologies.

B. Real-World Case Studies for IP Reputation Validation

Besides their contribution to theoretical understanding, some studies have also been empirically valid in proving the application of IPR validation systems in real life. Table II is a list of case studies which offer empirical data

to demonstrate the profitability of ML-based methods in the operations context.

The empirical investigation of the use of ML-based IPR methods is supported by cases provided in Table II. It was demonstrated that hybrid methods where various sources of intelligence are combined have a significant impact on reducing FP rates than their monolithic and single-source techniques. The study IPR Scoring with Geo-Contextual Feature Augmentation demonstrated that the effect of geo-contextual features on ensuring the accuracy of the threat assessment is beneficial as it allows attributing IPR behavioral patterns to geographic dimensions. Based on these observations, the proposed framework is designed through synthesizing several data repositories and contextual attributes.

IV. Limitations of Existing IPR Validation Architectures

The section is a review of the weaknesses identified in the current methods of IPR validation as such that form the rationale of the proposed methodology.

A. Comparative Analysis of Existing Approaches

Table III conducts a comparative analysis when reviewing existing strategies in terms of critical dimensions needed to be taken when realizing enterprise cloud security deployments. This review gives the research gap that is being filled by the proposed methodology.

As Table III below indicates, the existing methodologies do not concomitantly support real-time processing, cloud-native integration, high accuracy, profile response ability, and multi-source data integration, which is a requirement in enterprise cloud security deployments [22, 23]. Though lone experiments have been far more successful on the test accuracy (see stress-free Bagging with a success of 99.71%), none of them presents the complete set of functions required in an operational cloud environment. Recently developed deep learning integrations like Long Short-Term Memory-Intrusion Detection System (LSTM-IDS), GRU based systems and Autoencoder-IDS can show good sequential pattern storage, but do not have cloud-native operability and active response subject. It is in these aggregate inadequacies, that the framework below addresses with a holistic framework that leads to the ultimate accuracy of 98.04% and simultaneously as well as an integration with AWS that is cloud-native, an automated update of WAF rules [24], and aggregation of threat intelligence among various sources.

TABLE III. Comparative Analysis of Existing Approaches

Reference	Methodology	Real-time	Cloud	Accuracy	Auto Resp	Multi-Src
[25]	LSTM-IDS	Yes	No	96%	No	No
[26]	GRU-Based IDS	Yes	No	95%	No	No
[27]	Autoencoder-IDS	Partial	No	93%	No	No
Proposed	Hybrid Ensemble	Yes	Yes	98%	Yes	Yes

V. Proposed Hybrid ML-Based Solution

The WAFs available traditionally do not respond fast enough as the high amount of FP detection levels are evident, as well as the outdated IP address database that do not enable the WAFs to address the current threat vectors. The proposed solution consists of several sophisticated ML models, which are facilitated by enhanced decision-making and improved accuracy. This is the method to join the data of AWS WAF and GuardDuty into Amazon S3 and analyze them which were complemented by alerts of the SOC. The RF, SVM, LR, KNN, and eXtreme Gradient Boosting (XGBoost) models' integration enables the system to validate traffic in real-time by having a sophisticated model-based infrastructure.

The hybrid model is useful in the broadest conditions since it distinguishes actual traffic and the fake bot traffic. The system incorporates advanced ML algorithms to assist legitimate users and, in the process, detecting and blocking network threats on a real-time basis. It has a dynamic IPR-tagging-system and blocking-mechanism to respond to anomalous traffic and mitigate malicious traffic without affecting the operational safety. It is structured in a manner that the detection mechanism has extremely low rates of FP and consequently the services will be precise and dependable. Both proper security measures and initiative-taking risk management are a more reasonable method towards cybersecurity, as they can make certain that new threats do not affect web applications, and that normal operations of users will not be interrupted.

Algorithm of Proposed Hybrid ML-Based IPR Validation. Fig. 1 depicted that the present algorithm presents an integrated processing of IP address validation and classification based on the utilization of both internal security logs and external threat intelligence and ML. The system takes four diverse kinds of inputs which include a list of IP addresses that need to be checked, WAF logs, Amazon GuardDuty logs as well as alerts issued by the SOC. It then yields two key products, i.e., classification labels to every IP address and revised WAF rules to avert the recognized threats.

This begins with the preparation of data. The logs received in WAF, GuardDuty, and SOC alerts are, firstly, aggregated into one database. This is a composite data, and the preprocessing phase involves data cleansing and normalization therefore ensuring consistency within the data internally. Depending on the filtered data, pertinent characteristics are derived to form inputs of the subsequent ML models. As security datasets are typically highly skewed by far benign samples outnumbering malicious ones, the algorithm only uses SMOTE on the training set to balance the classes to enable the models to be effectively learn on both classes without distorting the real-world distributions in test data [28]. This strategy of dealing with

extremely imbalanced data is in line with the latest developments in hybrid ML systems in detecting fraud and threats [29]. They are then subsequently trained on the balanced training dataset on five varying ML models, which are the RF classifier, SVM, LR, KNN, and XGBoost. This is achieved by the fact that the number of models is employed, therefore, the system can capitalize on the strengths of each of the algorithmic methods and therefore the potential of misclassification which would have occurred otherwise by use of a single model is avoided.

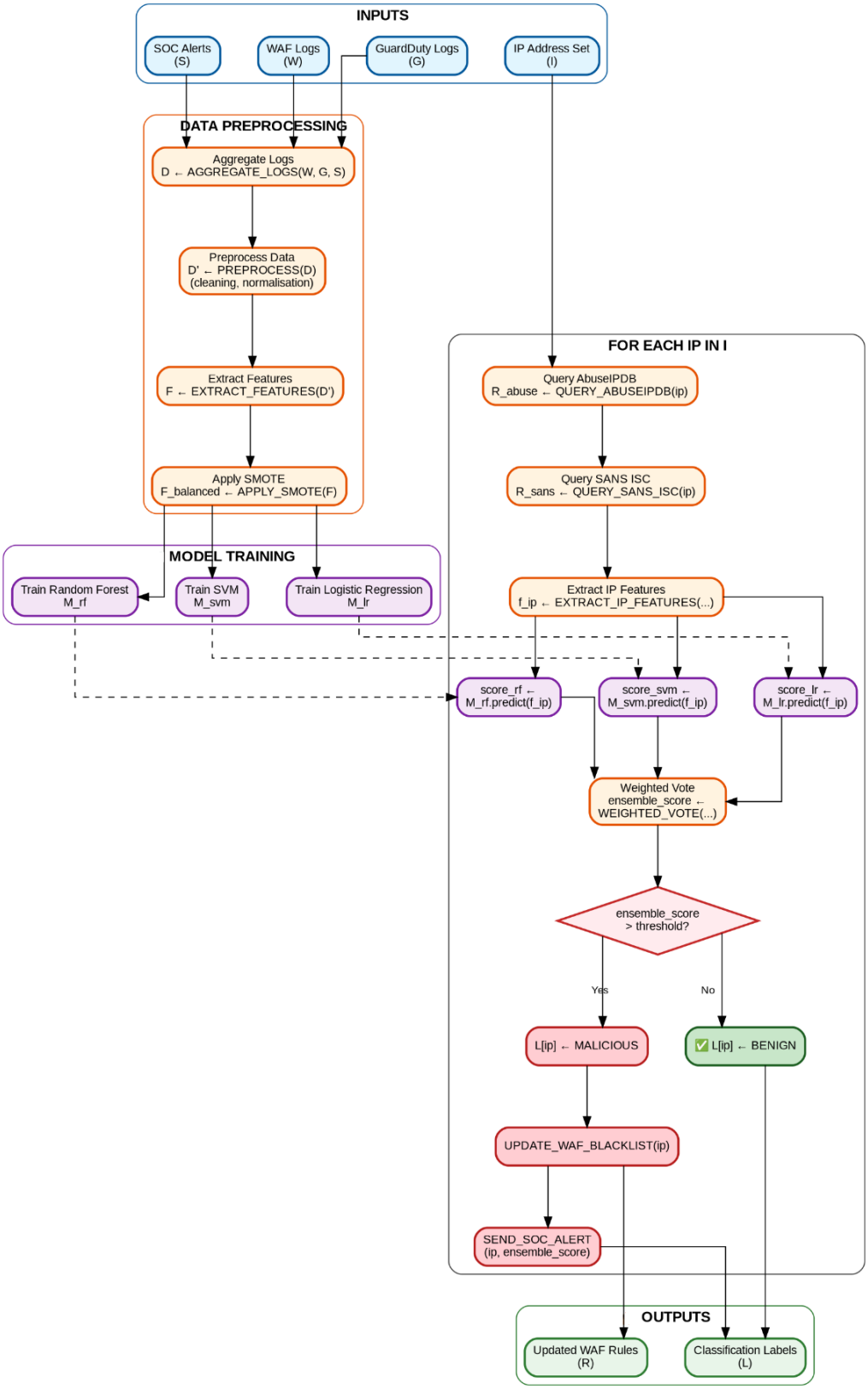

Fig. 1. Algorithm of proposed hybrid ML-Based IPR validation.

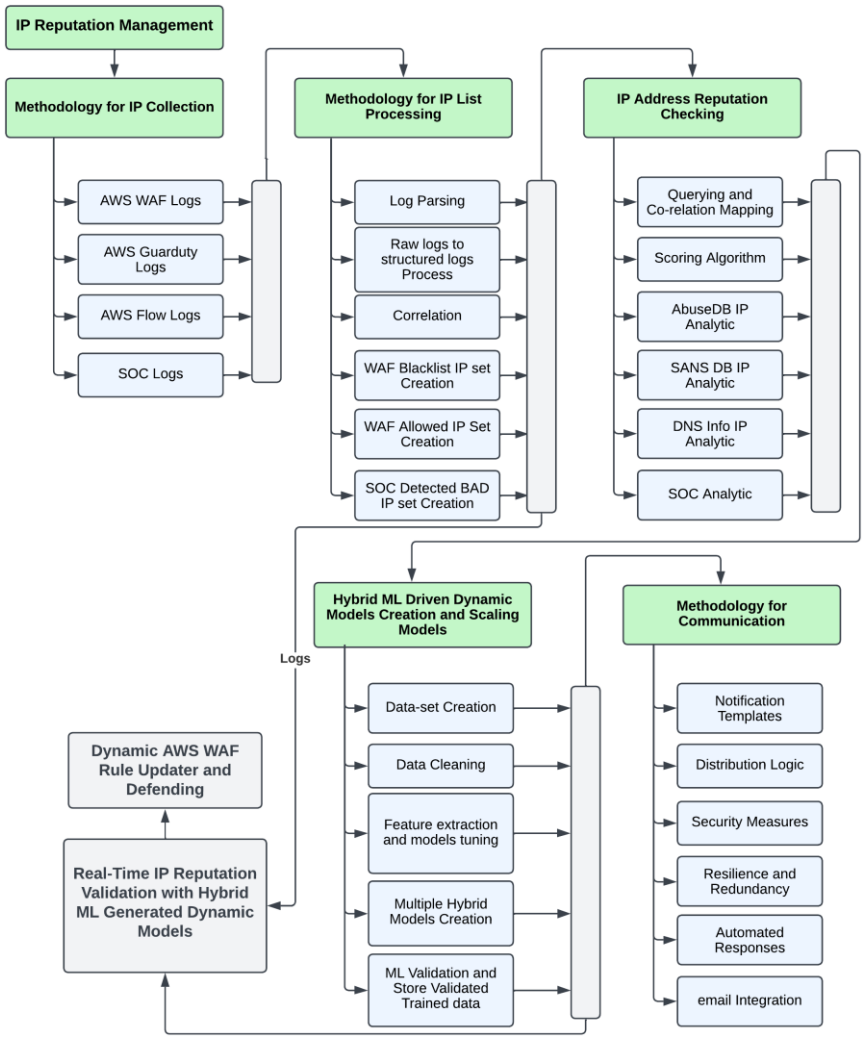

Fig. 2. Hybrid IPR validation methodology.

The algorithm queries the external threat intelligence databases, the AbuseIPDB, the SANS, on two occasions, on every IP address of the input group. According to the IP address, the features are derived according to the response of these external searches. Each of the three trained models then makes a prediction score on the IP address. A combination of these individual scores obtained by a weighted vote system is what forms an ensemble score that gives a more solid and dependable classification than could have been achieved by any of the single models. Each model of the weighted voting system has weights calculated by the cross-validation performance on the training dataset. In particular, the weight of each model (w i) takes the form of the division of F1 *i* by all the F1−Scores of the models (w_i = F1_*i* / ΣF1_*j*) and F1 *i* corresponds to the F1−Score of the model i and F1 j is the sum of model-*i* F1−Score. In this method more weight is put on models that exhibit greater discrimination ability when validating. According to our findings on experiments, the weights obtained are RF (0.24), XGBoost (0.23), SVM (0.20), LR (0.18), and KNN (0.15). The ensemble prediction is calculated as: $P\_ensemble = \Sigma\ (w\_i \times P\_i)$, *P i* being the probability of model *i*. When the threshold of P ensemble is more than 0.5, IP address is categorized as malicious.

Therefore, the ensemble score is taken against a fixed threshold. As defence tactic to avoid that, where the score will be above dynamically adjust the limit, the IP address will be termed as malicious, it will be registered on WAF deny list and an alert will be raised to SOC team to investigate further. Conversely, a score that is lower than the threshold is benign. The algorithm concludes with the delivery of the full set of the classification labels and modified WAF rules.

B. Hybrid ML-Driven IPR Validation

The overarching approach to IPR management presented in Fig. 2 combines a SOC model with a hybrid ML model. Strengths of the cybersecurity architecture are conditioned by the embedded components of detection of threats in real-time and a further possibility of auto-recovery measures based on analytical outcomes. This approach is a joint effort of SOC intelligence and real-time analytics and ML that provides a complete IPR management system. The system concentrates on the rapid data input, on the sophisticated reputation analysis and the

real-time confirmation options, and the standard communication criteria that are already developed.

C. IP Collection Approach

The IPR management starts by organizing the pertinent information that is available in diverse resources hence producing a comprehensive repository of IP addresses. To be accurate and dependable, the IP address reputation scoring relies on the exhaustive data collection. The AWS WAF logs are employed to monitor and filter the traffic on the sites that do not comply with the security rules set by the administrators and facilitate the tracking of the network. AWS GuardDuty logs provide precise information on the detection of threat through the identification of activity patterns, apparent anomaly identification, and minimum correlation with attack signatures, which is the improvement of network security. AWS Flow Logs capture much network traffic information that can be used to determine the occurrence of abnormal activity and easy detection of illegal data transfer and illicit network interactions.

The SOC logs offer security data using a centralized platform that synthesizes data of AWS WAF systems and SOC tools and therefore allows an incident response team to respond promptly to threats that are identified. The security analytics system ensures efficiency in developing the sound IPR management system through consolidating the entire log collection.

D. IP List Processing Approach

Raw logs are gathered and pass through a series of refinement processes to render them suitable in the threat detection process. In the log parsing, interesting components, i.e., IP addresses, timestamps, patterns of requests, and indicators of severity, are recognized and repackaged into form, capable of further analysis. The unstructured log data can be made more amenable to analysis by an improved chance of normalization into a structured corpus as normalization allows the use of more powerful detection mechanisms. The correspondence formed between different data points in the results of the correlation analysis will help to identify and detect threats continuously and identify organized accumulated sets of attacks that use multiple vectors. The dynamic nature of the WAF blocking rules and the associated rule set demands informing a repository of past based attack-pattern of recognized malicious IP addresses, and the continued incorporation of threat understanding. Having a backup of known IP ranges in the WAF can achieve a diminution of FP security alert in the system, therefore system integrity. IP addresses which are associated with malicious users are collectivized as per the SOC staff along with external threat feeds in order that the lists of denylists that are generated to comprise SOC-defined malicious IP addresses are dynamically changing.

E. IP Address Reputation Checking Approach

One of the significant techniques of the exclusion of legitimate and potentially malicious addresses is the assessment of the IPRs. Reputational metrics of individual IPs can form a baseline to establish a base on which analysts can use to conduct additional threat analysis. The information retrieval is done in the dual-layer query system, consisting of the internal and external data sets. Correlation- mapping analyses are capable of synthesis of various data sources therefore resulting in a total profile of behavior. The feedback loops enhance the behavioral signatures employed in the identification of bad actors at common occurrence. The scoring system relies on a series of variables, including behavioral analytics, cases of malicious activity, patterns of geolocation and history of prior incidents. Those factors have been calculated to produce risk scores to prioritize hierarchical threat prioritization systems. The AbuseIPDB analytic service is a query service, which cross-checks the IP addresses against the maintained database of known fraudsters. Such players are categorized into malware distributors, phishing attackers, or spammers and the service exposes them to all these types of threats. SANS DB IP analytics conducts cross-referencing protocols based on expert-curated data which is obtained out of the cybersecurity database of the SANS Institute. This helps in identifying the new or the latent risk vectors. Fast-flux networks and domain generation algorithms are some of the types of threats identified by DNS Info, based on exhaustive scans of DNS records, inverse scans, and scans on domain associations and could be utilized to deliver a second layer of situational awareness.

The AbuseIPDB analytic service is a query service, which cross-checks the IP addresses against the maintained database of known fraudsters. Such players are categorized into malware distributors, phishing attackers, or spammers and the service exposes them to all these types of threats. SANS DB IP analytics conducts cross-referencing protocols based on expert-curated data which is obtained out of the cybersecurity database of the SANS Institute. This helps in identifying the new or the latent risk vectors. Fast-flux networks and domain generation algorithms are some of the types of threats identified by DNS Info, based on exhaustive scans of DNS records, inverse scans, and scans on domain associations and could be utilized to deliver a second layer of situational awareness.

F. Hybrid ML-Driven Dynamic Model Creation and Scaling

The Hybrid ML provides an improvement of detection, due to effective mapping of massive data, detection of patterns, and adaptation to new attacks. The datasets are generated through the systematic collection and organization of vast repositories by the past attack history, the current network flow, and the third-party threat information, and therefore, contributes to the efficiency of the operations in training ML models. The process of data-cleaning eliminates any unnecessary, outdated, and irrelevant data and this leads to training data that is relevant and accurate. When the feature-extraction strategies and the model-tuning strategies are combined, the operational staff members recognize the major indicators of compromises, and this is why the predictive performance gets improved. Ensemble model that uses more than one hybrid ML is also a kind of supervised ML

algorithm that boosts the precision of the identification that is not supported by traditional security rule systems. When under the ML validation strict tests are put in place and validated training data is kept in archives to be utilized during production models.

G. Real-Time IPR Validation with Hybrid ML Dynamically Generated Models

Validation in real-time also becomes extremely important in preventing cyber threats to allow responding to new risks quickly. AWS WAF uses dynamic rule updater and defensive framework to dynamically provide security updates provisions, thus conducting automated rules and firewall maintenance, and preventing suspicious traffic. Real-time anomaly detection will be available to the system as that of monitoring continuously on this will be used to analyze the network traffic to identify security anomalies and take automatically protective actions in case of a threat being registered. Security defence solutions lower the staffing requirements when using the human operators and permit AI-powered countermeasures to act promptly and precisely. This is followed by the fortification of the security stance by real time verification because any threats emanating is identified at early levels before they may escalate to uncontrollable levels.

VI. Experimental Design

The section is truly clear about the system diagram, details of implementation, data constructions as well as the setup of the experiment.

In Fig. 3, the AI-security defence platform is presented, providing protection to cloud infrastructure with the help of hybrid ML models, AWS security tools, real-time threat detection, automated denylists, and dynamic defenses against the emergence of new threats. The cybersecurity mechanism can identify threats in real-time coupled with automatic blocking of IP addresses and continuous adaptability to the new patterns of attacks. The system will be able to generate dynamically evolving security-rules to actively suppress cyberattacks using advanced ML models which are stimulated by threat intelligence using AI. This is an automated decision-making and real-time monitoring and AI-driven security algorithms that will provide the protection of infrastructure with minimum human involvement. The framework has multiple tiers of cybersecurity, which are aligned with different functional points. These layers consist of firewall security monitoring, threat analysis, automatic security rule implementation, SOC procedure optimization, and continual ML improvement, therefore, offering the complete defensive coverage.

A. Real-Time IPR Detection and Dynamic IPR based Blocking

The system executes its primary defence in an AWS VPC as illustrated in Fig. 3 whereby WAF instances offer the initial defence against cyber-attacks by malicious IP addresses and bots. The defence approach evaluates the received traffic through combining both intelligence feeds of threats and ML models, thereby enabling the dynamic management of the security policies to guard against possible cyber threats. The network traffic filtering has two broad functions to distinguish legitimate users and the potential attackers to grant seamless and authorized traffic; and to repress the unauthorized requests in the process. The requests go through a progressive security checking procedure that depends on AWS WAF layers that permit a predefined security policy to be set in action along with actual-time examination produced by ML. The system swiftly scans the suspicious activities that comprise botnet attacks and brutal force activities and undertakes automatic response activities. It is a system that sets request flagging mechanisms that lead to intensive review of the behavior observed when the behavior falls out of normal standards. The Web Access Control List (ACL) enforces access policies, it operates on comprehensive security policies that are used to block, allow, or filter inbound traffic. Real time blocking is a blocking mechanism of access requests by malicious IP addresses, which is done automatically. An effective IP exclusion system will keep authorized user access, and authentic users will never be blocked even in a wave of traffic suspected of attack.

B. Continuous Threat Intelligence and Data Processing

The system as illustrated in Fig. 3 involves real-time external threat-intelligence sources to know the threats, verify detections and provide information to decision-making to forecast modern cyber threats. The system also collects information, computes hostile IP addresses, and processes the findings of security, reducing the human factor. AbuseIPDB API, SANS-ORG API, IP Info API, FQDN Info API, and AWS GuardDuty Fetcher are used in the system as AbuseIPDB API, cyber-threat feeds, geolocation tracking and Autonomous System Number (ASN), domain analysis, and AWS security monitoring, respectively. The security algorithm also offers continuous threat feeds to the security system by means of the enormous number of APIs by which the feeds are considered after undergoing a stringent, multi-stage evaluation process to identify dangerous IP addresses and malicious behavior. It will be followed by the three steps of cleaning the data that will then be incorporated into the hybrid ML system which will thereby improve the performance of the cyber-threat detection.

All the processed information is served to the ML models which detect the trends, review the threat which might occur and amend automatically to the AWS WAF rules. This automatic process grants the security system the ability to continuously update itself on newer threats and conduct real time and dynamic security control in advance therefore limiting the escalation of risk. Also, concerning processing time, there are two industries that cause delays in the real-time classification pipeline, specifically, external API queries and internal ML inference. Experimental data of our production setup indicate the following latency distribution: AbuseIPDB API queries take an average of 180-250 ms per IP address, SANS ISC queries take an average of 150-220 ms and the total ML inference across all five models runs at 45-60 ms.

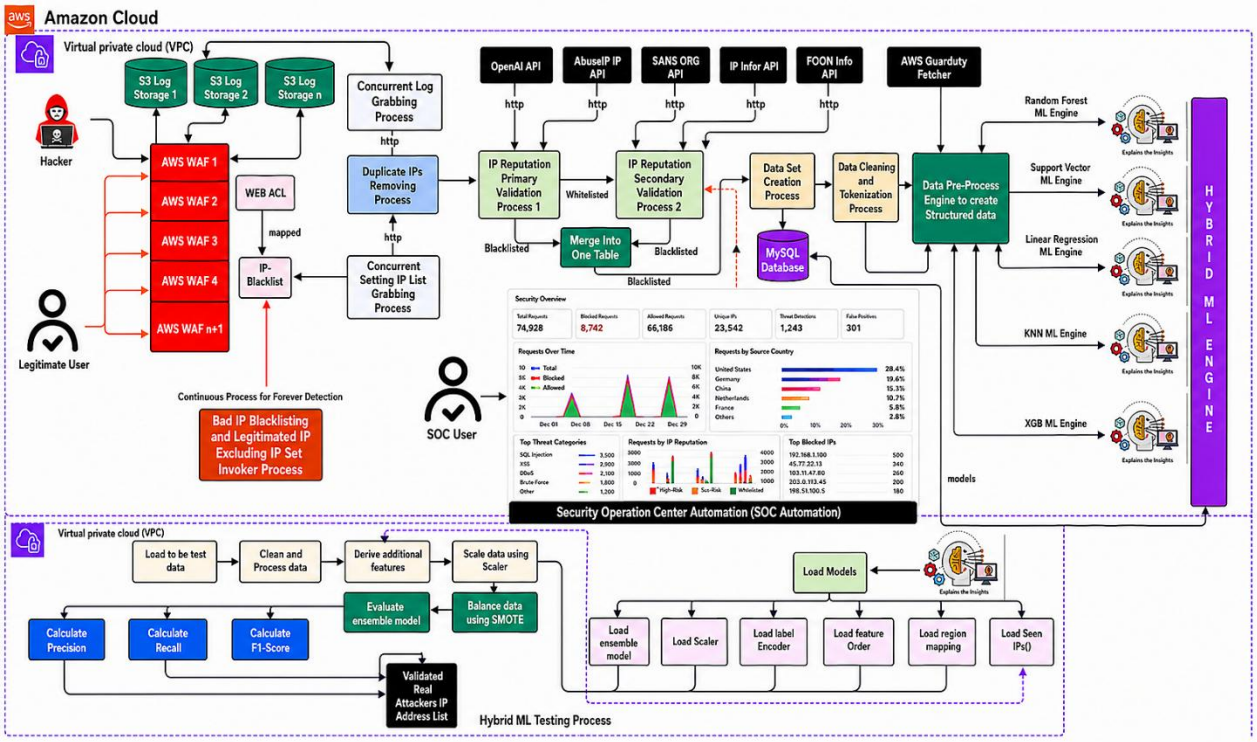

Fig. 3. Next Generation IPR validation architecture.

Also, the research considered to reduce the effects of external API latency on real-time processing, the system uses a multi-level caching approach based on Redis, in which recently accessed IP addresses are stored in 15-min long caches, which leads to the elimination of redundant API calls by about 65% at high traffic times. Also, asynchronous API querying is used where external database lookups are done parallel instead of sequentially and thus the cumulative external query time is cut down to 250 ms as compared to 400 ms. When the system needs to make blocking decisions using time-critical blocking information, it can run in fast-path mode with only the reputation information and ML inferences of cached information and classify IPs with sub-100 ms classification latency and then mark them as requiring full verification. In this architecture, the real-time processing claim is operationally valid, and end-to-end latency is 95% under normal operating conditions and less than 350 ms.

C. Hybrid ML Engine for Automated Security Enforcement

The hybrid ML engine as depicted in Fig.3 operates in real time to detect cyber threats and simultaneously performs classification and automatic security measures therewith, therefore maintaining an accurate defence against a threat. A number of ML algorithms are employed to support various threat-analysis activities with the help of RF which is a decision-tree model which recognizes suspicious network patterns, SVM, which is a traffic anomaly-classification model, LR, which is an attack-pattern-identification model in the long term and KNN, which is a model that detects threats based on past data. XGBoost has high accuracy in detecting activities thus providing rapid detection of malicious activities. The outputs of the security logs enforced through ML models are the detection of patterns that result in displaying the extent of danger before the IP addresses are classified as malicious or legitimate. The AWS WAF system can automatically append it denylist which contains risky IP addresses to be blocked out against attackers and its exclusion list which blocks false notifications. Integrated with the SOC Automation Dashboard is the ML-based decision making that allows visualization of the attacks in real-time and offers monitoring tools. Cybersecurity resilience is enhanced by the capability of the engine in learning new attack patterns, which is a more effective protection with a higher threat-detection rate.

D. SOC Automation Enhancement

The architecture has an optional C4 feature of Large Language Model (LLM) support in SOC operations. This additional tool, when activated, performs processing of security logs through natural language understanding to help SOC analysts in interpretation of threat and recommendation of response actions. Nevertheless, it should be mentioned that the fundamental IPR validation and threat classification are based solely on the hybrid ML ensemble (RF, SVM, LR, KNN, XGBoost) that is explained in the above sections. The integration of the LLM is not a detection mechanism but a human operator aide. The quantitative assessment of this element is not covered by this research and is named as a future activity. The system takes advantage of Sentence Transformers and Vector Databases to detect patterns and abnormalities concerning security in operational logs, and with the help of an inquiry field of question, a Security Q&A feature is presented, which can be accessed on demand by SOC operators.

E. Hybrid ML Model Training and Optimisation

As illustrated in Fig. 3, To achieve effects of high percentage of detection accuracy and reliability, the system is constantly validated, evaluated, and refined in its hybrid ML models. The performance of each model is strictly determined in terms of primary validation metrics, i.e., Precision, Recall and F1−Score, which are mentioned in Refs. [30−32]. The metrics are applied to optimize ML algorithms, to make them helpful in differentiating legitimate and malicious traffic and minimizing FPs. The system applies the most sophisticated data-preprocessing mechanisms, which eliminate redundant information to help increase the detection rates. Moreover, the biases within the training data were also addressed; SMOTE is applied to training data only to have balanced datasets to train the model but leave test datasets in their original imbalanced forms as this ensures that the performance is evaluated fairly. This methodology leads to more robust and generalized ML models having stable performance measurements.

This real-time learning aspect of self-learning enables the cybersecurity defence architecture to continue improving with more accuracy of the detective, enhance real-time security automation, and provide scalable and initiative-taking defence of the advanced assaults. In addition, this study has been covered in terms of concept drift, where immoral behavior of IPs changes over time and patterns of attack are changed, the architecture enforces a systematic retraining protocol. Concept drift is observed with the constant monitoring of model performance measures on a 7-day rolling production data window.

The retraining triggers are as follows some such as Performance-based trigger automatic retraining occurs automatically when the rolling F1−Score or the FP rate reaches or exceeds 3%; Scheduled retraining automatic retraining occurs automatically regardless of the performance measure when the SOC analysts detect significant new attack vectors or external threat intelligence sources (AbuseIPDB, SANS ORG) include substantial updates to their databases; Event-driven retraining automatic retraining occurs automatically regardless of the performance measure when the SOC analysts detect significant new attack vectors or when external threat intelligence providers (AbuseIPDB, SANS ORG) The retraining is based on the use of incremental learning whenever possible and this enables models to fit into new patterns whilst preserving the knowledge about threats that have been previously identified. During deployment, model versioning and testing are used, whereby new models are evaluated on a holdout dataset and then substitute production models without the retraining having the unintentional side effect of reducing detection capabilities.

F. Dataset

The data sample used in this research is a proprietary set collected on the production AWS locations across the three geographical locations (us-east-1, ap-southeast-1, and ap-southeast-2) and was taken over six months (between January and June 2024). The data set comprises of the raw security logs and processed logs of WAF, GuardDuty, AbuseIPDB and SANS API fetchers that will allow a holistic security detection solution. The whole dataset is 45,231 unique IP address records that have twenty-four engineered features obtained out of the union of the log's sources. The extraction of the features was conducted through the automated pipelines that featured the metadata of the connection (frequency of requests, duration of the session, geographic location), behavioral trait information (blocked request ratio, rule trigger patterns), and third-party reputation information (AbuseIPDB confidence score, SANS threat level).

TABLE IV. DATASET COMPOSITION AND SOURCES

Data Source	Records	Features	Reference
AWS WAF Logs	125,000	24	AWS Documentation
GuardDuty Findings	45,000	18	AWS Documentation
AbuseIPDB API	80,000	12	AbuseIPDB
SANS ISC Data	35,000	8	SANS ISC
SOC Alert Logs	15,000	16	Internal Collection

Labelling was done in two phases, the first one was automated classification based on threshold criteria (AbuseIPDB score more than 80, GuardDuty severity more than 7 or WAF block rate more than 50 percent), then manual validation by 3 SOC analysts who selected 15% of edge cases to validate the labelling quality. Cohen attained agreement between the raters at 0.87 kappa. The last distribution of classes is 38,446 benign samples (85%) and 6,785 attacker samples (15%), which is a realistic imbalance in operations. Table IV contains the list of sources and the dataset.

This part of data set was subjected to a set of preprocessing algorithms that were intended to enhance the characteristics of this data and enhance its predictive capabilities. The initial phase was eradication of data bias, and it minimized the unbalanced distribution of classes that had caused malfunctions of classification algorithms in the past. The archive operations eliminated unnecessary records of data and preserved the major security indicators thereby maintaining a dataset biased towards the major threat analysis. The introduction of geolocation data gave a new perspective of correlation between risky activities and the geographic areas.

The process of model optimization made use of adversarial training and generated perturbed training causing the system to face attack variations previously undeclared. This was one of the adaptation methods which made the models less vulnerable as they could observe the threats that the training data was not able to find. All these advances led to better models with the capability to differentiate between credible and non-credible traffic streams which enhanced the protection of security as well as reduced FPs in real-time.

G. Hybrid ML Training Architecture

Fig. 4 reveals that the hybrid ML process delivers a set of typical workflows to create, educate, evaluate, and launch the ML models which are specifically designed to apply and given to the cybersecurity applications. The

process may be divided into two large sections, i.e., Data Preprocessing and Feature Engineering and Model Training, Evaluation, and Deployment. The prepared labelled data is then subjected to a range of transformations in the data preprocessing to satisfy the quality and consistency requirements. This begins with importation of labelled data, cleaning, normalization, removal of outliers, distribution analysis, and leakage scrutiny to prevent the generation of bias.

Imputation is also implemented in implementation which tends to fill in the missing values and creation of new attributes, which are useful in prediction. The training data and the test data are separated, and additional scaling of the data is performed by Minmax scaling or standardization. This data is then iterated through the revising of sampling with SMOTE applied only to the training data through the consideration of the balance and a variety of different ML algorithms are trained, which are XGBoost, RF, SVM, LR, and KNN.

The application of SMOTE to train data only and hyperparameter optimization procedures is best practice by the definition of hybrid deep learning processes of sequential behavior analysis and non-uniformity of classes. Model training will be followed by verification datasets, which evaluate the performance of the model based on evaluation measures such as accuracy, F1−Score, recall, and precision statistics. The system conducts feature-importance analysis to identify whether there are important predictive factors in the dataset and model diagnosis to evaluate the risk of bias, variance, and the risks of over-fitting.

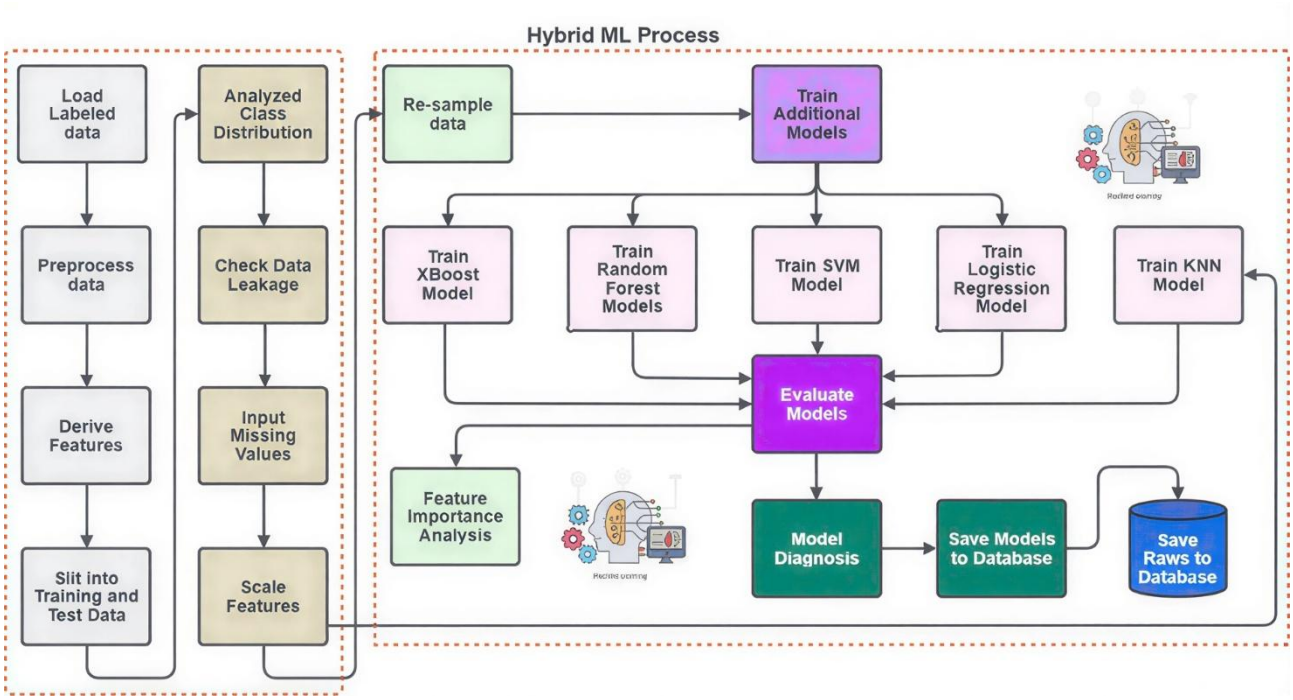

Fig. 4. Hybrid ML training and prediction H. hybrid ML prediction architecture.

Fig. 5 illustrates that IPR validation performs the malicious IP authentication process of scanning IP addresses using ML and achieves correct verification. A ML ensemble model is defined as an integer which integrates two or more classifiers to reduce FPs and FNs as well as enhance security. The model loading and preparation process relates to activities that entail database consistency and scalability by seeking to ensure that model database retrieval is conducted.

The loading dedicated part incorporates the scalers, label encoders, though also incorporates the IPs and feature mappings that have been found to enhance the reliability of the system. Through the mapping of the area, security analysts can monitor the distribution of potential security threats throughout the world. The system carries an intensive process of data processing ensuring methodological fidelity.

SMOTE is implemented only on the training data to eliminate the imbalanced distribution and the test data maintains the disproportional distribution to compromise with the real-life settings. Such a methodology predetermines the fact that the performance metrics reflect the real ability of the model to cope with real threat situations. In prediction, the incoming data are processed by cleaning operation to preserve the quality of the data and the features intact, then scaled by the parameters which are learned in the training set. The ensemble model then analyses the data which has been conducted to identify threats.

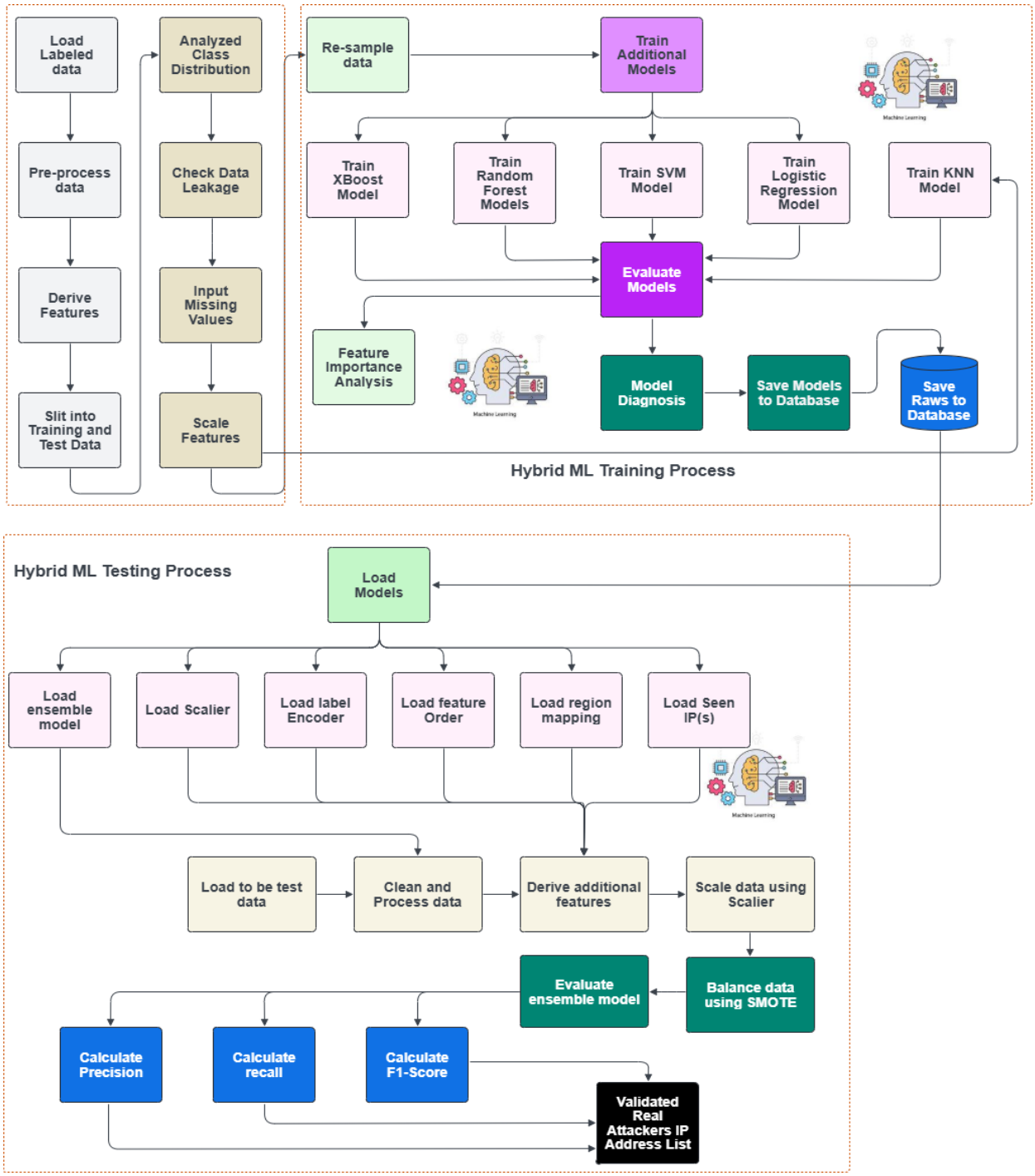

Fig. 5. Hybrid ML training and prediction architecture.

VII. Results

This section includes the findings of the experiment which evidences the efficiency of the proposed hybrid ML framework in a vast scope of assessment metrics.

A. Overall Hybrid ML Predictions on Multiple AWS Regions

Fig. 6 depicts the measures of precision, recall and F1 account of disparate AWS accounts that are spread across different geographies. These metrics have a monotonic increase in all the reviewed accounts, therefore, indicating the capability of the model to predict TP. A more detailed examination reveals that the model has a prominent level of true positives across the board, hence a steady precision and recall rate. Strongness of the model lies in that the model holds prominent levels of precision, recall and F1−Score despite minor variations in the scores of an individual when predicting TP across different regions of AWS. Another fact that helps to prove the correctness of the model performance is the consistency of the F1−Score of the model which became possible due to a balanced trade-off between the precision and the recall of the model in all the areas of the accounts. All these measurements

have been witnessed to a constant performance profile, and which is a testimony that the model can provide a high achieving level of accuracy irrespective of the AWS account or region in question.

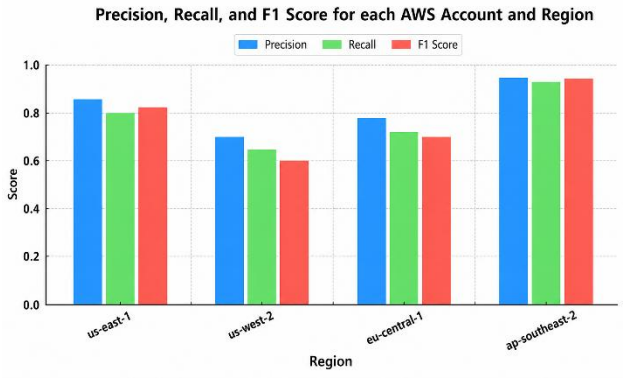

Fig. 6. Hybrid ML prediction over multiple regions.

B. Hybrid Supervised ML Models Comparison

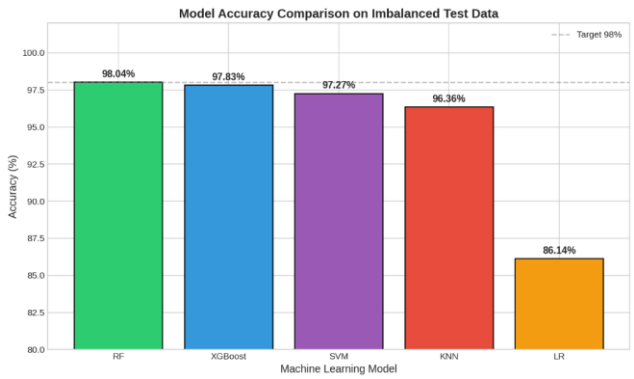

Fig. 7. The Hybrid supervised ML models comparison.

Fig. 7 proposes a comparative analysis of five different ML models, i.e., LR, KNN, RF, SVC, and XGB. The models possess some unique operational characteristics, being optimized to accuracy whilst being ineffective in training and the possibility of overfitting. The LR happens to be consistent with the data sets of any type and does not demonstrate an evident overfitting, therefore the most appropriate to consider the use in case of balanced collections. Although KNN model has the best training accuracy, the fact that it is slightly inclined to overfit may impact on predictive ability on new cases. RF performs optimally in general, since it is more precise with training data, and as powerful with validation data, making it a candidate in complicated datasets. As the amount of data being fed to SVC is increased, the performance of the SVC improves but the sensitivity to the hyperparameters used also implies that it is also sensitive to hyperparameters. Finally, extreme gradient boosting is the fastest of the pack both in terms of large input volumes and the excellent out-of-sample performance, and it is, therefore, the most suitable option when it concerns advanced designs of ML.

C. Learning Curve Analysis

The data of Fig. 8 demonstrates that the level of generalization with LR model is high, and the reason is due to its ability to maintain the level of bias and variance with the increasing number of samples of training, the score of training beginning at approximately 0.92 with increasing growth and addition of more samples indicating a strong learning ability. The cross-validation score has a performance trend of the same nature and is equaled with the training score with increase in data size. The difference between the two score points demonstrates proper learning behavior without either major over or under-fitting during the assessment period. The areas colored in show the range of data performance, which is less when there is more data injected into the model, showing more stability. The training and validation scores converge on about 0.96 indicating that they have low bias and variance and realistic levels of performance. The results obtained show that the model optimization is successful, and it means that the LR model has generalized.

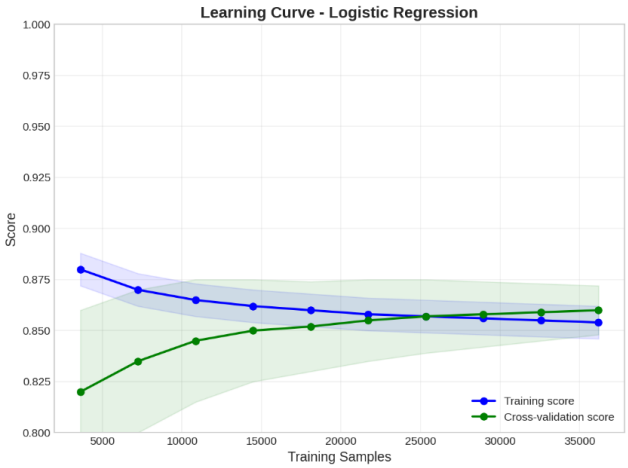

Fig. 8. LR learning curve behavior.

As can be presented in Fig. 9, the KNN model demonstrates near-perfect results with the training and cross-validation accuracy of 0.98 using a limited amount of training cases. The training accuracy is more than the cross-validation accuracy; this implies that there is a some-time mismatch. The larger the number of training samples, the larger the model will become acquainted with the distribution of data, and at the end, the values of concordant accuracy will be 0.98. The high variability of the colored regions during the first stages of the training process is an indicator of unstable data but as more data points are being added, variability decreases hence more model stability is achieved. The KNN algorithm however performs very well in this data and converts to the error rate of zero in case enough examples are available. Nonetheless, high performance values may also be a sign that they are overfitting the data, and at the same time the problem of separation is also trivially solvable due to clearly defined data groups.

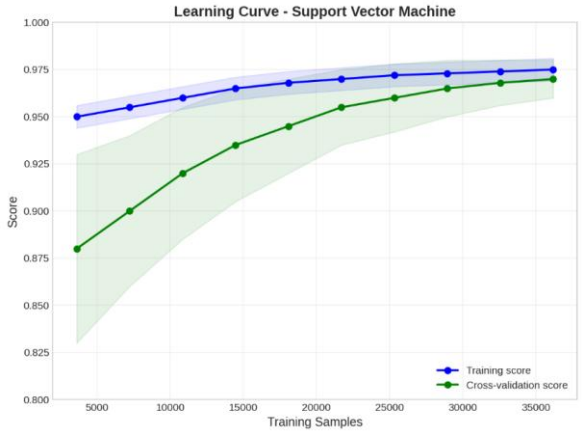

Fig. 9. KNN learning curve behavior.

As illustrated in Fig. 10, RF model possesses great learning capacity, and the training and cross-validation scores are reached 98 as the number of training samples increases. The evaluation metrics are extremely high and keep on increasing with supplementary data during the first training stage. The cross-validation score of the smaller is not as variable in the smaller samples as the training progresses. The larger the size of the training set, the less uncertain and less predictive is the cross-validation prediction, and the more predictive precision one has. This model can be generalized well, as the training regime can use sufficient examples, and there can be an over-fitting problem. Consequently, the model is biased, and its variance is low leading to high confidence in the set of data.

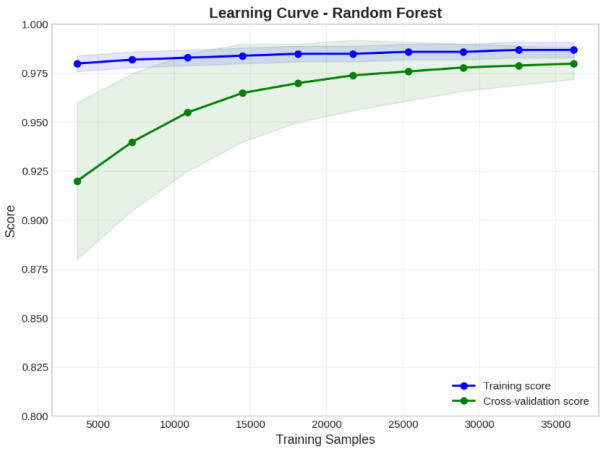

Fig. 10. RF learning curve behavior.

Fig. 11 indicates that training and cross validation score of the SVC algorithm steadily increase with an increment in the training data. The training score always has a high base, and the cross-validation score also has a lower base which is the natural variance. The two curves demonstrate that there is more generalization associated with the learning curves due to the appearances of coinciding curves with the addition of more training examples. These areas of shaded variance with the starting data and reduce with areas of shaded variance with the introduction of more samples in the model to strengthen the performance. The fact that the two measures of evaluation are slowly rising shows that the SVC model is a viable learner and generalizer and does not over-fit seriously. This model is thus even-handed in terms of bias and variance on prediction data and does not affect its reliability.

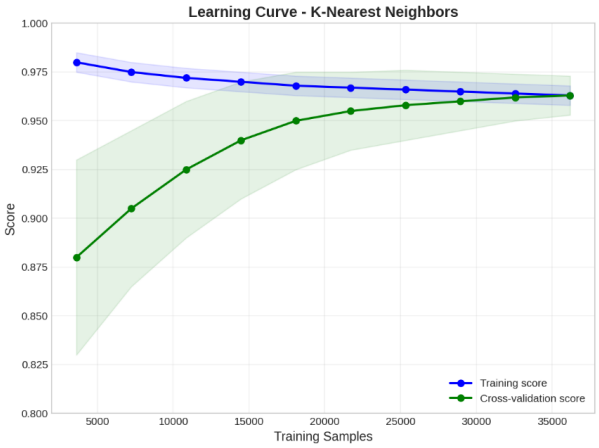

Fig. 11. SVC learning curve behavior.

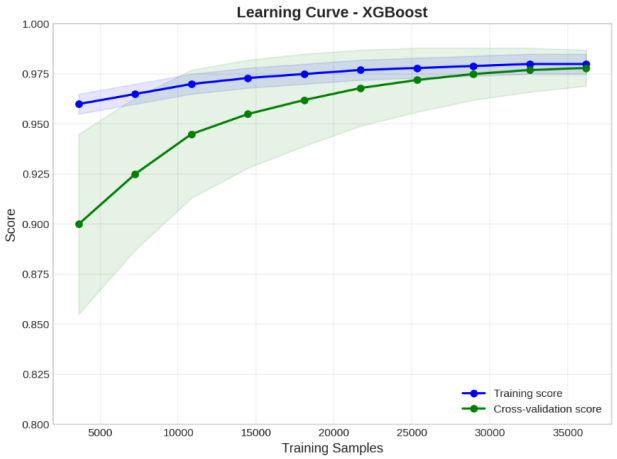

Fig. 12. XGB learning curve behavior.

As is presented in Fig. 12, model achieves the accuracy of 0.98 with a short training. The training also has an accuracy of 0.98, which indicates that it perfectly conforms to the training set. The cross-validation accuracy reflects by first reducing and then stabilizing to the figures of training accuracy. It means that the accuracy of the validation decreases as the number of the training samples grows, which testifies to the greater stable condition. This tendency indicates that the XGBoost model can represent recurrent patterns in the dataset successfully and, therefore, reduce bias and variance. The absence of errors in the training results implies that there has not been any false or unjustifiable overfitting due to the occurrence of this one cluster of data that can easily be separated to belong in certain classes.

D. Hybrid ML-Driven IPR Against Bad Actors

Fig. 13 demonstrates the distribution of F1−Scores of two classes, Not Attacker and Attacker. According to the statistics, the most stable category of identification is the Not Attacker, with the F1−Scores being centered around 0.97. The classification of normal activities is highly specific to the model and is easily remembered in the

classification. Attackers, however, attain F1−Scores of approximately 0.98. Consequently, the detecting model has extremely high success rates in the detecting process of attackers without reducing computational efficiency. RF has the best accuracy of 98.04, the second highest is XGBoost (97.83), and the third highest is SVM (97.27). The concentration distribution of F-1-scores in the two categories confirms the fact that the model is efficient in differentiating FP and FN cases. The prominent level of differentiation between the two labels points to the level of robustness in the model used in cybersecurity application with robust threat-detection tools and low levels of a FP occurrence.

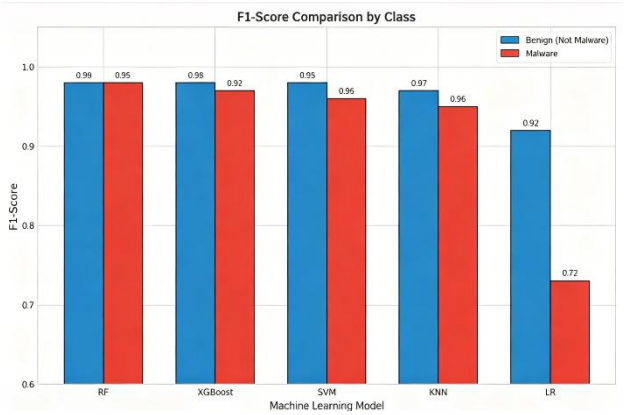

Fig. 13. The hybrid ML-driven IPR against bad actors.

E. Definitions of Hybrid ML Classification Metrics

The performance measures used in the model evaluation are in line with the established conventions. The definition of precision is as stated in Eq. (1):

$$Precision = TP / (TP + FP) \quad (1)$$

The percentage of TPs that is correctly recognized is referred to as recall and is defined as Eq. (2):

$$Recall = TP / (TP + FN) \quad (2)$$

The $F1$−Score is a harmonic average of precision and recall, which is defined in Eq. (3):

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (3)$$

FP Rate (FPR) measures the rate of instances that are falsely identified as positive as illustrated in Eq. (4):

$$FPR = FP / (FP + TN) \quad (4)$$

VIII. Discussion

The research shows the current research gap; it is demonstrated that the implementation of several ML models combined to form an IPR validation system is achievable and can help detect more malicious users without providing excessive security warnings. AWS WAF including GuardDuty and SOC log data are applied in the experimental set up to surmount the conventional security system constraints and achieve complete encoding of the threat intelligence and lower ratio of FP. A combination of real-time validation of suspicious IP addresses and high-speed response and accurate threat-detection methods allows to improve the detection of WAF and Web Applications security threats. Compared to the traditional approaches, the model is dynamically receptive of the emerging form of attacks, and therefore, the modern cyber-attack can be controlled on an ongoing basis. The synergy between the algorithms offers enhanced detection of the threats and simultaneously allows being able to track and instate control against the evolving attacks.

The resulting accuracy (98.04) of the RF is better than the resulting accuracy (85%), (90%), and (94%), of Static Blacklist, Automatic IP Blocking Using ML through Security Logs, and Dynamic ML, respectively. Thus, along with the last one, the proposed system will offer real-time processing simultaneously, cloud-native integration, and automatic response. The primary advantage of this strategy is that it will deal with the integration of information on multiple sources of threat-intelligence that will lessen the dependence on one reputation database. Given the scalability and efficiency in its operation, the system is best applicable in the circumstances of the enterprise level and cloud infrastructure.

Limitations and Potential Biases have been illustrated that the results have several limitations that need to be considered. The dataset has been sampled in selected AWS zones (us-east-1, ap-southeast-1, ap-southeast-2) and might not be accurate to the entire world's patterns of threats. Attack distributions may have geographic and temporal biases which have the potential of influencing model generalization to the other cloud environments. Also, its labelling methodology was based on criteria connected to thresholds and validated by SOC analysts, which might introduce the confirmation bias to the established version of the attack, but overlook unknown threats. Third, even with the balancing using SMOTE, the distribution of classes in the training data might not be the same in the working conditions, when attack traffic is usually a smaller percentage.

The system relying on the use of SOC-categorized alerts as one of its main input sources, though, introduces the risk of further distribution of human error into the system flow. The IP addresses may sometimes be misclassified when the SOC analysts lack enough information, are too overwhelmed at the high alert times, and when there are discrepancies in the classification criteria usage. The study has addressed these weaknesses by including a number of mitigation measures in the overall architecture, such as consensus-based labelling, where different analysts have to concur on borderline cases, confidence weighting, where the SOC-labelled IP patterns are weighted lower by the system whenever they are widely refuted by external databases, periodic audit sampling, whereby a random 10 percent of the SOC-labelled IPs are audited by senior analysts to identify systematic classification errors, and feedback loops, whereby the system raises red flags when Hybrid ML predictions are repeatedly contradicted by The procedures minimized (but not eliminate) the spread of human error through the use of automated defence mechanisms. Also, the overfitting mitigation behaviors were measured to prevent the possibility of overfeeding

such as hyperparameter tuning which used stratified 5-fold cross-validation, separate holdout test set with the original imbalanced distribution, regularization (L2 in LR and SVM) in tree-based models, and early stopping criteria in XGBoost. The learning curves show that there are proper convergence and not close scores of perfections that would show data leakage.

Future Work Future study opportunities include phases of comparison of DL systems (LSTM, Transformer models) to sequential attack patterns detection, quantitative analysis of the integration component of a LLM (to assist SOC), deployment analysis of the measured FP rates and impact of its operation on actual activity, federated learning to share threat intelligence between organizations without data exposure.

IX. Ethical Considerations and Data Privacy

The formulation and implementation of the proposed IPR validation system are based on ethical principles in the research of cybersecurity and data privacy laws such as Data Privacy Compliance such that Appropriate organizational authorization was applied to all IP address data which was addressed as production AWS environments while no personal identifying information was identified or stored. Most privacy models would consider IP addresses in the data set to be network infrastructure identifiers as opposed to personal information. This study is in conformity to AWS Acceptable Use Policy and security research guidelines.

Also, the General Data Protection Regulation (GDPR) considers that by aligning the extent that EU-origin traffic is deployed, the IP addresses can serve as personal data, according to GDPR. The use of such a framework by organizations must include Data Protection Impact Assessments (DPIAs) and the need to have proper justification of processing based on legal basis. The automated blocking system incorporates the legitimacy of interest balancing and is not a system of automated personal account.

Furthermore, there they did not identify or disclose any zero-day vulnerabilities or exploits during this study. The threat intelligence sources adopted (AbuseIPDB, SANS ISC, SOC Platform) are open security sources that are used as defensive architecture. Although this architecture is intended to work on the defensive side, autoprotective mechanisms have a risk of false positives on the legitimate usage. The system has allow-list management and manual reviewing provisions. The affected parties should have proper monitoring and appeal mechanisms that are used in the organization.

X. Conclusions

The pros of the strong ensemble approach whereby the IPR validation is evaluated by means of ML rather than conventional models, such as accuracy, scalability and interpretability, surpass the flaws of the latter. The proposed framework includes the data of the AWS WAF, GuardDuty, and SOC logs and, therefore, enables several ML models (RF, SVM, LR, KNN, and XGBoost) to evaluate the classification results and enhance the security of the AWS web applications. The revised experimental setup, which trains SMOTE on training only data, delivers the valid results of RF, which is 98.04 accurate with an F1−Score of 0.98, and then XGBoost (97.83%) and SVM (97.27%). These outcomes are realistic as far as real-world performance on an imbalanced threat data is concerned. The model offers scalability in performance and real-time data analytics enhanced by the computational simulation capabilities of the system operating on an extensively diverse palette of cloud-based service offerings.

The system has threat detection capabilities as well as feature imports that manage the privacy and interpretability issue thus is a versatile cybersecurity system. By default, the hybrid ML architecture and the continuous enhancement of web protection by real- time capabilities cast the modern vulnerabilities in the threat environment as detected by the system. In the future, scalable deep learning components will be incorporated, and the federated learning method will be researched to develop more effective detection of threats in modern clouds.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Nanayakkara Wawage Chanaka Lasantha is the brainchild of the research idea, the system architecture designer, the author of the experiments and the first draft of the manuscript; Madduma Wellalage Pasan Maduranga helped in data preparation, model implementation, validation of the results through experiments and also helped in the analysis of the results; Ruvan Abeysekara also guided the research, methodological advice, and reviewed the technical design and also helped in revising and editing the manuscript; all the reviewed manuscripts were approved and the authors were in agreement with the published manuscript.

References

[1] X. Li, Y. Xue, and B. Malin, "Detecting anomalous user behaviors in workflow-driven web applications," in *Proc. IEEE 31st Symp. Reliab. Distrib. Syst.*, Irvine, CA, USA, 2012, pp. 1–10.

[2] T. Blauth, O. Gstrein, and A. Zwitter, "Artificial intelligence crime: An overview of malicious use and abuse of AI," *IEEE Access*, vol. 10, pp. 77110–77122, 2022.

[3] C. Berghoff, M. Neu, and A. V. Twickel, "Vulnerabilities of connectionist AI applications: Evaluation and defence," *Front. Big Data*, vol. 3, Art. no. 23, 2020.

[4] T. H. Lacey, R. Mills, B. Mullins, R. Raines, M. Oxley, and S. Rogers, "RIPsec: Using reputation-based multilayer security to protect MANETs," *Comput. Secur*., vol. 31, no. 1, pp. 122–136, Feb. 2012.

[5] N. W. C. Lasantha, R. Abeysekara, and M. W. P. Maduranga, "A novel framework for real-time IP reputation validation using artificial intelligence," *Int. J. Wireless Microw. Technol*., vol. 14, no. 2, pp. 1–16, 2024.

[6] N. C. Lasantha, V. Tilwari, M. Maduranga, R. Abeysekara, N. Chakraborty, and D. Sharma, "Validating IP reputation in cloud firewall systems using machine learning driven signature generation and detection techniques," in *Proc. IEEE Ind. Electron. Appl. Conf. (IEACon)*, Kuala Lumpur, Malaysia, 2024, pp. 230–235.

[7] C. Huang, J. Han, X. Zhang, and J. Liu, "Automatic identification of honeypot server using machine learning techniques," *Secur. Commun. Netw*., 2627608, 2019.
[8] J. L. Lewis, G. F. Tambaliuc, H. S. Narman, and W. S Yoo, "IP reputation analysis of public databases and machine learning techniques," in *Proc. Int. Conf. Comput. Netw. Commun. (ICNC)*, Big Island, HI, USA, 2020, pp. 181–186.
[9] D. Jeon and B. Tak, "BlackEye: Automatic IP blacklisting using machine learning from security logs," *Wireless Netw*., vol. 28, no. 3, pp. 937–948, Apr. 2022.
[10] Y. Huang, J. Negrete, A. Wosotowsky, J. Wagener, E. Peterson, A. Rodriguez, and C. Fralick, "Detect malicious IP addresses using cross-protocol analysis," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI*), Xiamen, China, 2019, pp. 664–672.
[11] D. Chiba, K. Tobe, T. Mori, and S. Goto, "Detecting malicious websites by learning IP address features," in *Proc. IEEE/IPSJ Int. Symp. Appl. Internet*, Munich, Germany, 2012, pp. 29–39.
[12] M. Azhagiri, A. Rajesh, and S. Karthik, "An intrusion detection system using ranked feature bagging," *Int. J. Inf. Technol*., vol. 16, no. 2, pp. 1213–1219, Feb. 2024.
[13] K. Alemerien, S. Al-suhemat, and M. Almahadin, "Towards optimized machine-learning-driven intrusion detection for IoT applications," *Int. J. Inf. Technol*., vol. 16, no. 8, pp. 4981–4994, Dec. 2024.
[14] C. J. S. Mary and K. Mahalakshmi, "Modelling of intrusion detection using sea horse optimization with ML on cloud," *Int. J. Inf. Technol*., vol. 16, no. 3, pp. 1981–1988, Apr. 2024.
[15] N. Ruhela, R. Agrawal, A. Sharma, and A. Sharma, "A supervised ML-based solution for network intrusion detection using ensemble learning," *Int. J. Inf. Technol*., vol. 14, no. 7, pp. 3651–3658, Dec. 2022.
[16] N. D. Patel, B. M. Mehtre, and R. Wankar, "OD-IDS2022: Generating a new intrusion detection system dataset for machine learning-based attack classification," *Int. J. Inf. Technol*., vol. 15, no. 8, pp. 4349–4363, Dec. 2023.
[17] K. Sarathkumar, P. Sudhakar, and A. C. Kanmani, "Enhancing intrusion detection using coati optimization with deep learning on VANETs," *Int. J. Inf. Technol*., vol. 16, no. 5, pp. 3009–3018, Jun. 2024.
[18] J. K. Periasamy and K. Periasamy, "A novel cloud architecture to detect network intrusions using enhanced ANN," *Int. J. Inf. Technol*., vol. 16, no. 7, pp. 4267–4276, Oct. 2024.
[19] N. Usman, S. Usman, F. Khan, M. Jan, A. Sajid, M. Alazab, and P. Watters, "Intelligent dynamic malware detection using machine learning in IP reputation for forensics data analytics," *Future Gener. Comput. Syst*., vol. 118, pp. 124–141, May 2021.
[20] H. Sainani, J. Namayanja, G. Sharma, V. Misal, and V. Janeja, "IP reputation scoring with geo-contextual feature augmentation," *ACM Trans. Manage. Inf. Syst*., vol. 11, no. 4, pp. 1–29, Dec. 2020.
[21] I. Ghafir and V. Přenosil, "Blacklist-based malicious IP traffic detection," in *Proc. IEEE Glob. Conf. Commun. Technol. (GCCT)*, Thuckalay, India, 2015, pp. 229–233.
[22] S. Shaw and P. Choudhury, "A new local area network attack through IP and MAC address spoofing," in *Proc. Int. Conf. Comput. Commun. Syst*., Springer, 2018, pp. 347–350.
[23] F. Barbhuiya, S. Biswas, and S. Nandi, "An active host-based intrusion detection system for ARP-related attacks and its verification," *Int. J. Netw. Secur. Appl*., vol. 3, no. 3, pp. 95–107, May 2011.
[24] J. K. Lee, Y. Chang, H.-Y. Kwon, and B. Kim, "Reconciliation of privacy with preventive cybersecurity: The bright internet approach," *Inf. Syst. Front*., vol. 22, no. 1, pp. 45–57, Feb. 2020.
[25] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," *IEEE Access*, vol. 7, pp. 41525–41550, 2019.
[26] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Commun. Surv. Tutor*., vol. 21, no. 3, pp. 2224–2287, 2019.
[27] M. Zavrak and M. Iskefiyeli, "Anomaly-based intrusion detection from network flow features using variational autoencoder," *IEEE Access*, vol. 8, pp. 108346–108358, 2020.
[28] A. Yazdinejad, R. M. Parizi, A. Dehghantanha, and K.-K. R. Choo, "AI-powered security for IoT ecosystems: A hybrid deep learning approach to anomaly detection," I*EEE Trans. Ind. Inf*., vol. 19, no. 11, pp. 11234–11243, Nov. 2023.
[29] S. Alharbi, M. Alshammari, and A. Almutairi, "Fraudulent account detection in social media using hybrid deep transformer model and hyperparameter optimization," *IEEE Access*, vol. 12, pp. 45892–45907, 2024.
[30] A. M. Carrington *et al*., "Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit, and explanation," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 45, no. 1, pp. 329–341, Jan. 2023.
[31] Z. Wen, R. Zhang, and K. Ramamohanarao, "Enabling precision/recall preferences for semi-supervised SVM training," in *Proc. 23rd ACM Int. Conf. Inf. Knowl. Manage*, Shanghai, China, 2014, pp. 1399–1408.
[32] C. K. I. Williams, "The effect of class imbalance on precision-recall curves," *Neural Comput*., vol. 33, no. 3, pp. 853–857, Mar. 2021.