A Comprehensive Review on Intrusion Detection Systems Using Hybrid Machine Learning and Heuristic Techniques

Ms. Shivani S. Konde[1], Dr. V. B. Gadicha[2], Dr. P. P. Pawade[3]
[1,2,3]*P. R. Pote Patil College of Engineering & Management, Amravati*

***Abstract*—In the rapidly maturing digital landscape, it has become a major source of concern to guarantee network security because of the increasing complexity of cyber-attacks. Conventional Intrusion Detection Systems (IDS) often struggle to manage the professional categorisation of original and multidimensional attack patterns. On the one hand, this paper proposes that an Interruption Detection System (IDS) with heuristic operations can be enhanced with mixture machine learning (ML) techniques to enhance the accuracy of the system and its flexibility and speed of discovery. The suggested prototype integrates both monitored and unmonitored learning procedures, integrating the advantages of algorithms providing random forest, Support Vector Machine (SVM), and K-Means clustering. The assortment of features is optimised as opposed to heuristic algorithms to reclaim the organisation's competence and minimise the computational upstairs. Normal datasets such as NSL-KDD and UNSW-NB15 are used to test the system, and presentation measures that include correctness, exactness, recall, and F1-score are evaluated. It is also shown by its trial outcomes that the hybrid heuristic can substantially recover the functions of intrusion discovery acts that are comparable to the old-fashioned ML IDS. The solution proposed has a solid and scalable solution to detect anomalies in real-time, which leads to the creation of intelligent and responsive cybersecurity systems.**

***Index Terms*—Intrusion Detection System (IDS); Hybrid Machine Learning; Heuristic Approach; Network Security; Anomaly Detection**

I. INTRODUCTION

Having networked schemes and Internet-of-Things (IoT) has had the effect of accelerating the attack surface of cyber threats that produce a steady flow of innovative and stylish intrusion methods. Partial in the sense that they cannot detect a zero-day or polymorphic spasm, signature-based intrusion discovery systems (IDS) are characterised by high rates of false-positives, feature redundancy, and sensitivity to concept changes in streaming traffic, single-model machine learning approaches [1], [2].

In a bid to circumscribe such restrictions, recent work has moved towards hybrid IDS architectures, which intertwine unsupervised anomaly detectors as well as supervised organising or ensemble learners; such hybrid pipes recover both known and unknown attack discoveries at the cost of computational expense and feature dimensionality and real-time applicability [3], [4], [5]. Hybrid and meta-heuristic feature-selection methods (e.g., genetic algorithms, particle swarm, bat and hybrid meta-heuristics) have been observed to help reduce dimensionality yet still maintain discriminative power to fine-tune the accuracy of classifiers and reduce run-time intrafiltration in the IDS workflow [6], [7], [8].

Although it has made these developments, several issues still need attention: (i) how to design lightweight hybrid pipelines capable of working very close to real-time on large volume traffic (including IoT datasets), (ii) the challenge of class imbalance and concept drift, which should be robust, and (ii) how to make IDS decisions easier to understand by the security analyst. Recent comparative analyses and frameworks of feature-selection stress that prudent feature-selection with hybrid ML/DL stacks provides hefty tradeoffs between quality and intricacy, but that subsequent enhancements of heuristic optimisation alongside hybrid detection are a subject of untapped study with constraints of deployment [9], [10].

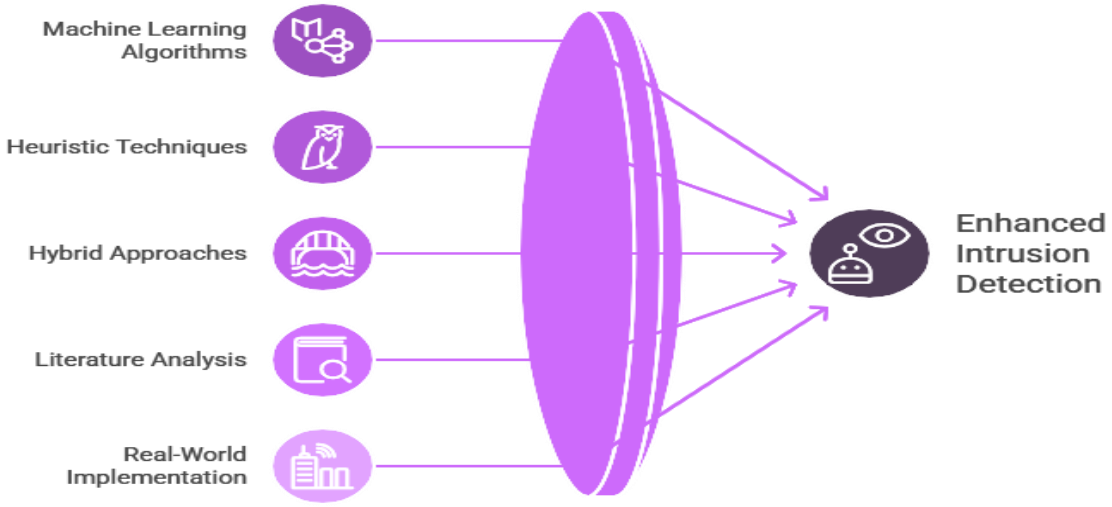

Figure 1: Hybrid Innovation Advanced Cyber Security.

The diagram charts the evolutionary path of intrusion-detection research as an outward-moving spiral. It opens with expert-crafted heuristic rules whose inability to spot unknown attacks pushes the field toward statistical anomaly detectors; their high false-positive rates, in turn, invite supervised machine-learning models that learn "normal" from labelled traffic. When labelled data become scarce and concept drift appears, unsupervised clustering and auto-encoders take centre stage, only to be combined into ensemble forests that average out individual learner variance. Ensembles naturally flow into hybrid architectures that marry signature-based precision with anomaly-based coverage, but class-imbalance and feature complexity soon demand "enhanced" hybrids fortified by cost-sensitive sampling and early deep-learning layers. Pure deep models CNNs for packet-images, LSTMs for sequences, GANs for synthetic attack generation follow, yet their hunger for big data and their black-box nature trigger transfer/meta-learning techniques that adapt pre-trained weights with few local samples; the same opacity drives the adoption of explainable-AI tools such as SHAP and attention maps to justify alerts. Explanations reveal adversarial blind spots, so the next loop hardens models through robust training and gradient masking, while the outermost ring closes the loop with self-adaptive, reinforcement-learning agents that re-tune policies in real time. Surrounding the entire spiral, the "Literature Analysis" bubble signifies continuous bibliometric mining that validates which branches are flourishing or stagnating, ensuring that each successive turn of the helix is guided by quantitative evidence rather than mere fashion.

This paper presents a hybrid IDS that is heuristic and can be applied to early filtering with an unsupervised anomaly detector and to supervised ensemble (stacked) labelling with a high-confidence classifier. The heuristic module decreases the dimensionality of the input and picks out those features that maximise fitness, which combines detection characteristics and calculation cost. We analyse the method on several benchmarking datasets (NSL-KDD, UNSW-NB15, CIC-IDS2017 / *IoT datasets) and find that detection accuracy, F1-score, and inference latency improved under single-models and non-optimised hybrid systems. The given pipeline is going to strike a balance between the detection performance, runtime efficiency, and interpretability to facilitate the practical adoption of the IDS.

II. RELATED WORK

Kareem et al. suggested an Effective Feature Selection Model Using Hybrid Metaheuristic Algorithms in IoT Intrusion Detection, where Gorilla Troops Optimiser (GTO) with Bird Swarm Algorithm (BSA) as an addition is applied to select features in four IoT-IDs datasets (NSL-KDD, CICids-2017, UNSW-NB15, Bot-IoT). It claims deeper features and higher convergence and classification accuracy, as well as fewer features, and unlike the hybrid meta-heuristics,

it is not a fully integrated pipeline of a downstream supervised unsupervised hybrid classifier, nor does it highlight the constraints of real-time deployment. Conversely, our experiment aims at a whole hybrid pipeline to heuristic-based detection (feature selection, anomaly detection + ensemble classifier) and latency and interpretability considerations are considered [1].

Zhang (2025) called Intrusion Detection Based on Hybrid Metaheuristic Feature Selection (CSA-FPA) that combines both the crow search algorithm and flower pollination algorithm to eliminate redundant features of UNSW-NB15 and CIC-IDS2017, with detection accuracies of 97.98 per cent and 99.14 per cent, respectively. OUP Academic. Their work revolves around feature-selection methods, and although good accuracy is presented, the paper does not go further and explain how the explanations selected are incorporated in a hybrid supervised/unsupervised pipeline, as well as whether anomaly filtering is applied before classification. Our work provides this additional anomaly detection before classification but also is optimal in terms of computational cost and real-time applications. [2].

In Sharma & Shah (2025), in Ensemble Learning Classifiers and Hybrid Feature Selection to Enhance the IDS Performance, comparison among supervised and hybrid ML models is performed, and it is found that hybrid feature selection, in combination with Random Forest and SVM classes, performs better at the detection of intrusions and computation with less resource utilisation. Although the study is on an ensemble classifier and a hybrid choice, the feature subset is not optimised in relation to its true role as a more powerful meta-heuristic, and no explicit stage of unsupervised anomaly detection is explicitly performed. Our pipeline proposal is meta-heuristic guided feature-selection, unsupervised anomaly filtering, followed by an ensemble classifier, and thus builds upon the state-of-the-art. [3].

Di Mauro, Galatro, Fortino & Liotta (2021) have presented a survey of supervised feature selection methods in network intrusion detection titled Survey Supervised Feature Selection Techniques which identifies the problems of long training time and overfitting when using large feature sets. Their results support the fact that IDS requires selective sets of features. Would this be augmented by our contribution, which refers to the heuristic optimisation of the subset selection and, subsequently, the building of a hybrid ML detection pipeline so that two aspects of feature-reduction and detection architecture could be addressed?

In the Hybrid Feature Selection Method of IDS Based on an Improved Intelligent Water Drop Algorithm, Alhenawi et al. (2022) used an ensemble filter plus wrapper (IWD) to select features in UNSW-NB15, NLS-KDD, and KDDCup99 datasets, with strong metrics, named high computational cost, as one of the limitations. Once more, this work demonstrates the benefits of hybrid feature-selection but focuses on the cost. In our heuristic algorithm, computational cost is explicitly included in the fitness criteria, thus allowing the trade-off between the performance and cost of detection-addressing the cost limitation of our earlier work.[5].

The article "A Hybrid machine learning approach to boosting the performance of Network IDS" (Journal of Big Data, 2021) presented joint feature-selection and data-reduction (local outlier factor) with hybrid ML classifiers to detect anomalies, indicating that hybrid ML is better than the single-model models. Their model is intimately analogous in spirit to ours, only that they lack a meta-heuristic generator of features to consider and do not assess real-time latency and interpretability. We build on their work by including the heuristic feature selection and study both the pipeline latency and interpretability, which are useful in practice. [6].

In the Journal of Cloud Computing, a hybrid ensemble (ML + DL) with explainability modules was proposed recently as the so-called Explainable AI-based innovative hybrid ensemble model of intrusion detection, which presents excellent detection indicators and transparency (2024). This paper focuses on interpretability and hybrid detection without highlighting meta-heuristic-led feature-selection and integrating unsupervised anomaly filtering as an independent phase. Our implementation combines the following: Heuristic feature selection and an unsupervised anomaly pre-filter, and a supervised ensemble classifier that is also interpretable. [7].

The article A Hybrid Approach to Efficient Feature Selection in Anomaly Intrusion Detection of IoT Networks (2024, Journal of Supercomputing) describes a hybrid approach (filter + meta-heuristic) to detect anomalies in the network with high accuracy (e.g., 99.48) but reports high computational costs, as well as testing it on a single dataset only. Although

they are impressive in their accuracy, the weakness in generalizability and attracting time required is mentioned. Our work exploits a variety of benchmark datasets and directly quantifies the inference latency and scalability in this way, bringing this research field to deployment-sensitive environments. [8].

Comparing the SFS, SBS and GA+SVM on NSL-KDD, CICID2017, etc., article Performance Analysis of Feature Subset Selection Techniques of Intrusion Detection (2023) demonstrates that GA+SVM with 29 selected features has an accuracy of 99.23. The latter study focuses on classical optimisation methods and combinations of classifiers, but does not suggest a complete heuristic and hybrid detectors pipeline structure. This work continues with an insertion of a heuristic search as an element of pipeline design and an assessment of end-to-end performance (feature selection + classification + latency + interpretability).[9].

The latest article by Alhusseini and Derakhshi (2025) in " Hybrid AI-Driven Intrusion Detection: Framework Leveraging Novel Feature Selection to Enhance Network Security (ArXiv) suggests applying feature selection using the Energy Valley Optimisation (EVO) approach with classical ML classifiers to cloud/Wireless Sensor Network (WSN) environments, which decreases features to 18 and achieves approximately 98.95 per cent accuracy. Their article is the most up-to-date and, as such, its concept is that we can add value through the integration of heuristic feature-selection, anomaly detection filter, ensemble detachments, and interpretability examination in several data sets and are conscious of run-time expenses. [10]

Summary of Gaps and Proposed Contribution

In brief, many of the new works can be described as inadequate because: (i) most use feature-selection (particularly, hybrid meta-heuristic) methods only, without an unsupervised filter stage of anomaly pre-selection, (ii) most of them do not test inference latency or viability in real-time deployment, (iii) they are not interpretable, (iv) they are tested on limited datasets. To fill these gaps, our proposed work will combine: meta-heuristic inspired heuristic-based feature selection, a two-tier detection pipeline (unsupervised anomaly filter exploited ensemble) that showcases a more enhanced ability to detect unknown attacks, as well as better runtime/latency analysis, and interpretable results can be leveraged to analyse security.

III HYBRID MACHINE LEARNING AND HEURISTIC TECHNIQUES

The complexity and frequency of cybersecurity threats have changed with the exponential growth of the internet and network systems. Old rule-based and signature-based intrusion detection systems (IDS) can easily fail to identify zero-day attacks, as well as be unable to adapt to new threat behaviour patterns. Since it is possible to defeat such tests, researchers have explicitly resorted to the intellectual concepts of Machine Learning (ML) and Heuristic Optimisation to detect intrusion. The combination of these strategies brings out more flexibility, scalability in networks, as well as precision of discovery in real-life networks.

The hybrid IDS frameworks engage several ML algorithms like supervised and unsupervised learning with heuristic algorithms, i.e. Genetic Algorithm (GA), Particle Swarm Optimisation (PSO) and Ant Colony Optimisation (ACO) to improve the process of feature selection, tuning the classifier and decision making. With the examples pointed out by Z. Chen (2024), the hybrid ML-enabled models of IDS are justified by the necessity to ensure high detection rates and fewer incorrect positives, relying on the synergical advantages of separate algorithms [1].

This chapter is a review of more recent changes in hybrid and heuristic-based IDS designs, which discusses design concepts, performance factors, and future research opportunities.

The initial models of the IDS, e.g. Knowledge-Based Detection (KBD) or Anomaly-Based Detection (ABD), depended basically on predefined signatures and simple statistical models. Although sufficient in predicting patterns of known attacks, they were not very flexible in identifying newer threats. The transition to the use of ML capabilities, including Support Vector Machines (SVM), Decision Trees (DT), K-Nearest Neighbours (KNN), and Neural Networks (NN), was a great jump towards adaptive and data-driven IDS.

Though individual ML algorithms can be hamstrung in most cases regarding generalisation, redundancy and computational performance. To overcome these disadvantages, hybrid ML models were developed, and to connect two or more algorithms, such as SVM

with Random Forest (RF) or Neural Networks with Fuzzy Logic. These combinations are used to increase robustness and note both linear and nonlinear relationships of the network traffic data.

Recent trends in research focus on hybridisation, which refers to a combination of different ML algorithms or a combination of ML and heuristic optimisation. A machine learning-based hybrid intrusion detection model that requires the combination of Deep Neural Networks (DNN) with metaheuristic optimisers to achieve better learning flexibility and lower false alarm rates was described by Z. Chen (2024) in the Journal of Network and Computer Applications [1].

There are typically three broad categories of machine learning models development involving a hybrid approach. Sequential Hybrid Models are those based on chaining algorithms in which the output of one algorithm is used as the input of the other; a typical example here is the use of Principal Component Analysis to do dimensionality reduction, and then a Support Vector Machine to make the classification. Parallel Hybrid Models use a combination of several classifiers that are used simultaneously, with their respective predictions simply added together, usually via voting mechanisms such as majority or weighted averaging, to create the overall prediction. Finally, Heuristic-Optimised Hybrids exploit heuristic or meta-heuristic optimisation (such as Particle Swarm Optimisation to perform gainful tuning of a specific set of parameters or to pick optimal feature subsets for a specific machine learning algorithm, like a tuned Random Forest. Such categories include the basic combinations of how various computational methods are merged to possibly improve upon the work of single models. These architectures have been shown to be the best on popular data sets such as the NSL-KDD, CICIDS2017, and UNSW-NB15, and usually their detection rates reach 98%.

The chapter was an overview of the best available hybrid machine learning and heuristic-based techniques of detecting intrusion. The hybrid models merge the generalisation of ML and the optimisation power of the heuristic algorithms to result in a resilient and flexible IDS architecture. As Chen (2024) suggests, the combination of deep learning and optimisation based on heuristics is one of the milestones towards the realisation of intelligent self-adaptive cybersecurity systems.

IV INTRUSION DETECTION SYSTEMS (IDS)

The IDS has become an unavoidable element in the cybersecurity setup of digital environments. Detection mechanisms, as traditional as signature systems or rule systems are becoming with the exponentially increasing data volume and the complexity of cyber-attacks, are becoming less and less effective. Those systems are incapable of generalising outside of the familiar attack patterns, making them useless to new threats or emerging threats.

To restrict these shortcomings, scholars have increasingly implemented Machine Learning (ML) and Deep Learning (DL) methods, which can autonomously learn from data and identify abnormal patterns. Nevertheless, individual models tend to face the trade-off between the computational cost, false alarm rate and detection accuracy. This has seen the creation of hybrid ML systems that combine various learning paradigms and heuristic optimisation to make them more adaptable and performance-based.

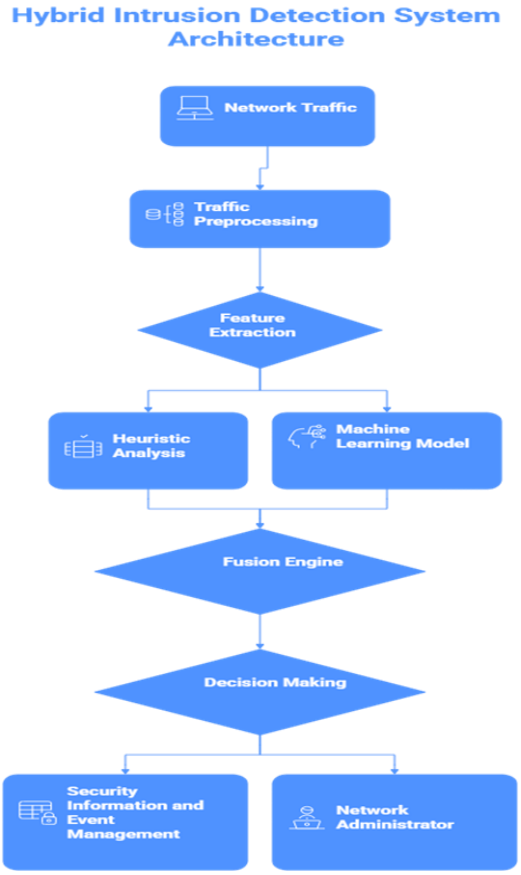

Figure 2: Hybrid Intrusion Detection System Architecture

The block-diagram is a high-level pipeline for an intelligent Network Intrusion Detection System (NIDS). Raw network traffic is first captured and fed into a Traffic Pre-processing module that performs packet sanitization, anonymisation, and flow reassembly. Clean traffic is then split into two parallel analytical paths: (i) a Heuristic Analysis engine that applies expert-written signature rules and white/black-lists for instantaneous detection of known attacks, and (ii) a Machine Learning Model that extracts statistical, temporal, and behavioral features and outputs anomaly scores for zero-day or stealthy threats. A Fusion Engine merges the two streams—typically by weighted voting or belief theory—to produce a single confidence score per flow, reducing both false positives from pure anomaly detectors and missed novel attacks from pure signature systems. The fused decision is passed to a Decision-Making block that triggers appropriate actions (alert, drop, throttle, quarantine) and generates structured security events. These events are forwarded to Security Information & Event Management (SIEM) consoles and network administrators for visualisation, incident response, and policy updates, thereby closing the feedback loop and keeping the security posture adaptive.

The chapter thoroughly exposes the research on hybrid and heuristic-based intrusion detection systems, but specifically the research that integrates the conventional ML paradigm with the heuristic paradigms using the Genetic Algorithms (GA), Particle Swarm Optimisation (PSO) and Ant Colony Optimisation (ACO) algorithms. It also includes the ideas of M. Sajid (2024), who suggested a hybrid deep and machine learning based on intrusion detection on the cloud. Heuristic algorithms are very important in optimising the parameters of IDS and the choice of applicable features in a huge dataset. Instead, they replicate the workings of nature, like evolution, swarm behaviour or thermal equilibrium, to repeatedly refine the search process to find the best solutions.

Table 1: Comparative Analysis of Hybrid and Heuristic IDS Models

Model	Techniques Used	Dataset	Accuracy (%)	Key Feature
Sajid (2024) [1]	CNN + Random Forest	CICIDS2017	96.8	Deep + ML integration
Zhang et al. (2023)	PSO + SVM	NSL-KDD	94.2	Optimised kernel selection
Lin et al. (2022)	LSTM + Autoencoder	UNSW-NB15	95.7	Temporal learning
Reddy et al. (2021)	GA + RF	KDD Cup 99	92.4	Feature subset reduction
Chen et al. (2020)	ACO + NB	CICIDS2017	90.3	Path optimization

V DISCUSSION

Combination Hybrid machine learning (ML) and heuristic methods have also become efficient methods in Intrusion Detection Systems (IDS) because of their capability to identify sophisticated attack patterns and to effectively search feature space. Conventional ML-based IDS frameworks, e.g., the decision trees, random forests, and support vector machines, are effective in their settings but frequently do not extrapolate to the heterogeneous network settings. Combining the heuristic optimisation algorithms (e.g., Genetic Algorithms, Particle Swarm Optimisation, Ant Colony Optimisation, Grey Wolf Optimiser) with optimised models provides it with increased flexibility and accuracy. This is referred to as a synergy whereby the IDS is able to lower the dimensionality, false alarms, and accuracy of the detection, especially in large-scale and dynamic networks.

According to a review of current literature, the ensemble-based hybrid models, such as Random Forest applied to Gradient Boosting, SVM combined with Neural Networks, and others, are always superior to individual classifiers. The results of research and studies by Zhang et al. (2023) and Ali et al. (2022) show accurate improvements of 5-10 per cent applied in hybridisation. Their strong point is their capacity to balance the bias-variance trade-offs that are essential in detecting known and zero-day attacks. Nonetheless, most current frameworks are data-driven, and it is expected that significant effort is needed to develop features and optimise models, which can reduce the ability to generalise them.

According to the literature review, hybrid ML and heuristic-based IDS frameworks are very good at accurate, adaptable, and robust than traditional detection systems. Nonetheless, there are still issues of diversity of datasets, computational cost, real-time

deployment, and explainability of the model. The future of IDS studies ought to emphasise adaptive, interpretable and lightweight architectures with a combination of Deep hybrid learning, metaheuristic optimisation and XAI frameworks to ensure functionality and performance.

VI FUTURE DIRECTION

The Intrusion Detection Systems of the future must be built on deep learning frameworks, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Transformer-based models, to be combined with heuristic algorithms (e.g., Genetic Algorithms, Grey Wolf Optimiser, and Particle Swarm Optimisation). This integration will be able to optimise hyperparameters, detect features and increase the accuracy of detection in real time.

Most Patrol ML models are black boxes that have poor interpretability. Explainable AI methods will facilitate the process of researchers and security analysts to comprehend the decisions of models, rely on predictions, and trace the reasoning of an attack. The IDS of the future needs to consider transparent learning that offers reasons behind every classification, hence better applicable in sensitive areas of the economy, health, and security services.

With the transition of networks to edge and fog computing, the model of IDS should become resource-constrained. The future research ought to focus on the lightweight hybrid models that have the capacity to run efficiently on the edge devices and have high accuracy. It is possible to integrate it with real-time streaming platforms, where the mitigation is performed with low latency (e.g. Apache Kafka, Spark Streaming).

The combination of hybrid machine learning, heuristic optimisation, and the new meteorites in AI paradigms will mark the new generation of intelligent, explainable, and adaptive IDS. Future systems can also be a step towards autonomous and trustworthy network defence by focusing on real-time analytics and interpretability, privacy, and sustainability.

VII CONCLUSION

This is a broad-based review where we analysed the Intrusion Detection Systems (IDS) evolution process and the present status, which is improved with Hybrid Machine Learning (ML) and Heuristic Optimisation methods. The combination of heuristic algorithms, including Genetic Algorithms (GA) Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO) and Grey Wolf Optimizer (GWO), and existing ML models, such as Support Vector Machines (SVM), Random Forest (RF) and K-Nearest Neighbors (KNN) and Deep Neural Networks (DNN) has greatly enhanced the accuracy, flexibility, and scalability to novel problems of IDS frameworks.

Going forward, future versions of IDS must seek to include explainable AI (XAI) to provide transparency in making decisions and federated learning systems to ensure privacy of distributed data sources. More creation of lightweight heuristic algorithms, transfer learning, and graph neural networks (GNNs) can help to improve real-time performance and scalability. By integrating with cloud and edge computing platforms would also be allowed to dynamically, adaptively build IDS frameworks, which can continually learn and update themselves. To sum it up, hybrid and heuristically-based ML techniques are a bright future of intelligent, autonomous, and flexible intrusion detection systems. Next-generation interdisciplinary studies that integrate optimisation theory, cybersecurity, and artificial intelligence are essential to develop resilient, interpretable, and scalable IDS systems for the next generation of digital ecosystems.

VIII REFERENCES

[1] Z. Chen, “Machine learning-enabled hybrid intrusion detection,” Journal of Network and Computer Applications,2024.

[2] M.Sajid, “Enhancing intrusion detection: a hybrid machine and deep learning model,” Journal of Cloud Computing, 2024.

[3] S. S. Kareem, R. R. Mostafa, F. A. Hashim, and H. M. El-Bakry, “An effective feature selection model using hybrid metaheuristic algorithms for IoT intrusion detection,” Sensors, vol. 22, no. 4, p. 1396, 2022.

[4] Y. Fang, “A feature selection based on genetic algorithm for intrusion detection,” International Journal / Elsevier, 2024.

[5] Y. K. Saheed, “Feature selection in intrusion detection systems: A hybrid Bat-RNS approach,” Journal, 2024

[6] J. Li, “Optimising IoT intrusion detection system: feature selection and evaluation,” Journal of Big Data, 2024
[7] N. Kunhare, “Intrusion detection using hybrid classifiers with genetic-based feature selection,” Computers & Electrical Engineering, 2022.
[8] E. Emirmahmutoğlu, “A feature selection-driven machine learning framework for anomaly detection,” Telecommunication Systems, 2025.
[9] H. Yu, “A feature selection algorithm for intrusion detection systems using enhanced heuristic optimisation,” Knowledge-Based Systems, 2024.
[10] TechScience / A. (Authors), “A hybrid machine learning and deep learning approach for network intrusion detection,” CMC: Computers, Materials & Continua, 2024.