IJARETY

International Journal of Advanced Research in Education and TechnologY *(IJARETY)*

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 3.911| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 7, Issue 4, July - August 2020 ||

DOI:10.15680/IJARETY.2020.0704017

Examination of Approaches for Feature Extraction and Classification in AI-based Threat Detection

Sakshi Sharma[1], Natasha Dutta[2]

IT Project Manager, Promedia Telecom Inc., San Diego, CA. [1]

Security Analyst, Senior Executive, Vodafone Intelligent Solutions, Pune, India[2]

ABSTRACT: This paper presents a comprehensive examination of various approaches for feature extraction and classification in AI-based threat detection. As cyber threats become increasingly sophisticated, the effectiveness of threat detection systems depends heavily on the ability to extract relevant features from vast amounts of data and accurately classify potential threats. This study explores a range of techniques for feature extraction, including traditional statistical methods, deep learning models, and hybrid approaches that combine multiple techniques. Additionally, the paper reviews different classification algorithms, such as machine learning, deep learning, and ensemble methods, evaluating their performance in diverse threat detection scenarios. By analyzing the strengths, limitations, and practical applications of these approaches, this paper aims to provide valuable insights for researchers and practitioners in the field. The findings highlight emerging trends, challenges, and future directions in AI-based threat detection, ultimately contributing to the development of more robust and effective cybersecurity solutions.

KEYWORDS: AI, Threat, Detection, Hybrid, Cyber, Deep Learning, Machine Learning, Data Preprocessing.

I. INTRODUCTION

As cyber threats evolve in complexity and scale, the need for advanced threat detection systems has become more pressing. In this context, artificial intelligence (AI) plays a pivotal role in enhancing threat detection capabilities by leveraging sophisticated feature extraction and classification techniques. Feature extraction is a critical initial step in threat detection, involving transforming raw data into a set of informative features that represent the underlying characteristics of potential threats. Traditional methods, such as statistical analyses and domain-specific heuristics, have been widely used, but their effectiveness can be limited in the face of increasingly complex and voluminous data. Recent advancements in deep learning have introduced new paradigms for feature extraction, enabling automatic learning of high-level features from raw data, thereby potentially improving detection accuracy. Similarly, the classification stage, which involves categorizing extracted features into threat or non-threat classes, has benefited from a range of AI approaches. Cyber-security attacks and their associated risks have significantly increased since the rapid growth of the interconnected digital world [1], e.g., the Internet of Things (IoT) and Software-Defined Networks (SDN) [2]. IoT is an ecosystem of interrelated digital devices and objects known as "thing" [3]. They are embedded with sensors, computing chips and other technologies to collect and exchange data over the internet. IoT networks aim to increase the productivity of the hosting environment, such as industrial systems and "smart" buildings. IoT devices are growing significantly, with an expected number of 50 billion devices by the end of [3]. Machine learning algorithms, such as decision trees, support vector machines, and ensemble methods, have been foundational in this area, but deep learning techniques and ensemble models are increasingly being employed to capture intricate patterns and relationships within the data. This introduction provides a comprehensive overview of these approaches, examining their methodologies, advantages, and limitations. Unfortunately, current security measures in IoT networks have proven unreliable against unprecedented attacks [4]. For instance, in 2017, attackers compromised a casino's sensitive database through an IoT fish tank's thermometer. According to the Nozomi networks' report, new and modified IoT botnet attacks increased rapidly in the first half of 2019, with 57% of IoT devices vulnerable to attacks [5]. According to the Symantec Internet Security Threat Report, more than 2.4 million new malware variants were created in 2018 [6]. That led to growing interest in improving the capabilities of NIDSs to detect unprecedented attacks. It also highlights the practical implications of these techniques in real-world threat detection scenarios, offering insights into how they can be leveraged to enhance cybersecurity measures. An NIDS is implemented in a network to analyse traffic flows to detect security threats and protect digital assets [7]. It is designed to provide high cyber-security protection in operational infrastructures and aims to preserve the three principles of information systems security: confidentiality,

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 3.911| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 7, Issue 4, July - August 2020 ||

DOI:10.15680/IJARETY.2020.0704017

integrity, and availability [7]. Detecting cyber-attacks and threats have been the primary goal of NIDSs for a long time. There are two main types of NIDSs: Signature-based aims to match and compare the signatures from an incoming traffic with a database of predetermined signatures of previously known attacks [8]. The exploration of these advanced methods aims to inform future research and development, fostering the creation of more effective and resilient threat detection systems capable of addressing the dynamic and evolving landscape of cybersecurity threats. Generally, they can achieve higher accuracy and Detection Rate (DR) levels for zero-day attacks, as they focus on matching attack patterns and behaviours rather than signatures [9].

1.1 Traditional Methods for Feature Extraction

Traditional methods for feature extraction in cybersecurity primarily involve techniques rooted in statistical analysis and domain-specific heuristics, which have been foundational in transforming raw data into meaningful representations for threat detection. These methods typically begin with preprocessing the data to clean and normalize it, ensuring that irrelevant or redundant information is minimized. Common statistical techniques include calculating metrics such as mean, variance, and correlation coefficients, which help summarize the data and identify patterns that may indicate potential threats. Additionally, domain-specific heuristics involve manually defined rules and features based on expert knowledge of network behavior and attack patterns. For instance, features such as packet sizes, protocol types, and port numbers are often used to identify anomalies that deviate from established norms. Methods like Principal Component Analysis (PCA) are employed to reduce the dimensionality of the data while retaining its essential characteristics, thus simplifying the feature set and improving computational efficiency. Current signature NIDSs have proven unreliable for detecting zeroday attack signatures [10] as they pass through IoT networks. This is due to the lack of known attack signatures in the system's database. To prevent these incidents from recurring, many techniques, including ML, have been developed and applied with some success. ML is an emerging technology with new capabilities to learn and extract harmful patterns from network traffic, which can be beneficial for detecting security threats [11]. Deep Learning (DL) is an emerging branch of ML that has proven very successful in detecting sophisticated data patterns [12]. Its models are inspired by biological neural systems in which a network of interconnected nodes transmits data signals. While these traditional methods have proven effective in various contexts, they often struggle to handle the complexity and volume of modern cybersecurity data, which can limit their ability to detect new or sophisticated threats. The trend in this field is to outperform state-of-the-art results for a specific dataset rather than to gain insights into an ML-based NIDS application [13]. Therefore, the extensive amount of academic research conducted outweighs the number of actual deployments in the real operational world. As a result, there is a growing need to complement these approaches with more advanced techniques, such as those offered by machine learning and deep learning, to enhance feature extraction capabilities and improve overall threat detection performance.

II. LITERATURE REVIEW

S Khan (2019) The importance of cybersecurity has grown substantially, capturing the attention of both academic researchers and industry professionals around the world. As the Internet of Things (IoT) continues to expand, ensuring secure and sustainable computing within this domain has become crucial. Machine learning techniques play a key role in enhancing IoT cybersecurity, especially in detecting intrusions and identifying malicious activities. In this study, we present innovative methods for feature extraction and selection within an Intrusion Detection System (IDS), utilizing the capabilities of swarm intelligence (SI) algorithms. Our approach involves a feature extraction mechanism based on conventional neural networks (CNN) and introduces a novel feature selection method using a newly developed SI algorithm, the Aquila optimizer (AQU). To assess the effectiveness of our IDS model, we employed four well-known public datasets—CIC2017, NSL-KDD, BoT-IoT, and KDD99. We conducted thorough comparisons with other optimization techniques, demonstrating that our approach outperforms others based on various evaluation metrics. [14]. The choice of the right approach depends on factors such as network size, data availability, and privacy and security considerations. This work is intended to help practitioners make informed decisions to meet their ID needs effectively. Intrusion detection (ID) plays a vital role in safeguarding computer networks from various malicious attacks. The rapid advancement of technologies like machine learning (ML), deep learning (DL), federated learning (FL), and explainable artificial intelligence (XAI) has garnered considerable interest as promising solutions for ID. DL-based methods have demonstrated remarkable effectiveness in ID by automatically learning relevant features from data. However, these approaches demand large amounts of labeled data and substantial computational power to train complex models. In contrast, ML-based methods require less computational power and labeled data, but their ability to generalize to new, unseen data is often limited. FL represents a newer approach that allows multiple entities to collaboratively train a model without sharing their data, thus offering privacy and security advantages. This makes FL an appealing option for ID, though it requires more communication resources and additional computation to aggregate models from different entities.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 3.911| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 7, Issue 4, July - August 2020 ||

DOI:10.15680/IJARETY.2020.0704017

Two key processes in this analysis are feature extraction and feature selection, which help distill valuable information from raw network traffic data. Feature extraction involves converting raw data into a set of representative features that highlight the essential aspects of the traffic. This process typically begins with data preprocessing and cleaning, followed by the extraction of features at various levels, such as packet-level, flow-level, and statistical, time-based, or frequency-based features. Techniques used for this include statistical analysis, transformations, machine learning methods, and domain-specific approaches like Deep Packet Inspection. Feature selection, on the other hand, focuses on identifying the most informative and relevant subset of these extracted features. This step is crucial for reducing redundancy, lowering dimensionality, improving interpretability, and minimizing computational complexity. Different methods are used for feature selection: filter methods assess feature relevance based on statistical measures like information gain or correlation; wrapper methods evaluate feature subsets by their effect on model performance; embedded methods integrate feature selection within the learning algorithm itself, such as through L1 regularization or decision trees; and hybrid methods combine multiple techniques to enhance feature selection outcomes. Assessing the effectiveness of these feature extraction and selection techniques is essential, often involving performance metrics and cross-validation to determine their impact on tasks like classification or anomaly detection. Practical applications of these methods in network traffic data analysis include intrusion detection, traffic classification, quality of service analysis, traffic prediction, and botnet detection. Given the growing volume and complexity of network traffic, efficient and effective feature extraction and selection techniques are increasingly important for enhancing network security, optimizing performance, and enabling intelligent decision-making. Future research should focus on tackling emerging challenges and developing innovative methods to address the evolving nature of network traffic and data characteristics.

Amir Hussain offers a comprehensive overview of IDS, discussing its various types and methodologies, the kinds of attacks it detects, as well as the datasets, metrics, and performance indicators used in these systems. It includes an in-depth analysis of recent research on IDS solutions, assessing their strengths and weaknesses, and exploring their potential impact, the challenges they face, and emerging trends in the field. This review provides a thorough examination of the latest advancements in ML and DL-based IDS and aims to guide future research on how emerging Artificial Intelligence (AI) technologies can address the increasing complexity of cybersecurity challenges. The rapid development of communication and internet technology has introduced significant risks to network security. To address these threats, Intrusion Detection Systems (IDS) were created to defend against malicious network attacks. Despite their importance, IDSs still face challenges with accuracy, false alarms, and identifying new types of intrusions. To improve detection accuracy, organizations have started integrating Machine Learning (ML) and Deep Learning (DL) algorithms into IDS.

III. METHODOLOGY

The methodology for examining approaches to feature extraction and classification in AI-based threat detection involves a multi-faceted approach that integrates both traditional and advanced techniques. Initially, the process begins with a thorough review of existing literature to identify and categorize various methods of feature extraction and classification used in cybersecurity. This review includes an analysis of traditional methods such as statistical measures and domain-specific heuristics, which have been the cornerstone of early threat detection systems. Following this, the focus shifts to advanced techniques, particularly those involving machine learning (ML) and deep learning (DL), which offer more sophisticated and automated approaches to feature extraction. For feature extraction, methods such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are compared with modern techniques like Convolutional Neural Networks (CNNs) and Autoencoders, which can automatically learn and extract complex features from raw data. Similarly, for classification, the study evaluates various algorithms, including traditional machine learning models such as Support Vector Machines (SVMs) and Random Forests, as well as more advanced DL models like Recurrent Neural Networks (RNNs) and Transformer-based architectures. The methodology also includes empirical evaluation using benchmark datasets, where different methods are applied and their performance is measured against metrics such as accuracy, precision, recall, and F1 score. Comparative analysis is conducted to assess the strengths and limitations of each approach, taking into account factors like computational efficiency, scalability, and robustness to novel threats. Additionally, the methodology involves case studies and real-world applications to demonstrate how these techniques perform in practical scenarios. This comprehensive approach aims to provide a holistic understanding of the effectiveness of various feature extraction and classification methods in enhancing AI-based threat detection, offering valuable insights for future research and practical implementation in cybersecurity.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 3.911| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 7, Issue 4, July - August 2020 ||

DOI:10.15680/IJARETY.2020.0704017

IV. RESULT

Initially, we assess the performance and runtime of various feature selection and extraction methods in binary classification, as detailed in Tables 1, 2, 3, 4, and 5. For each feature number configuration, we conduct five iterations to ensure reliable results. The final average result is calculated based on the outcomes from each iteration. These tables present the performance metrics and processing times for feature sets containing 9, 22, 33, 47, and 77 features. In each table, the best outcomes are highlighted in bold and red, indicating the optimal results for feature selection and extraction. These optimal values represent the highest accuracy, precision, recall, F1-score, and Matthews correlation coefficient (MCC), as well as the lowest times for feature reduction, training, and inference. Time measurements for feature reduction and training are recorded in seconds (s), while the inference time per data sample is recorded in milliseconds (ms).

Figure 1: The explained variance and cumulative total variance for extracted feature schemes

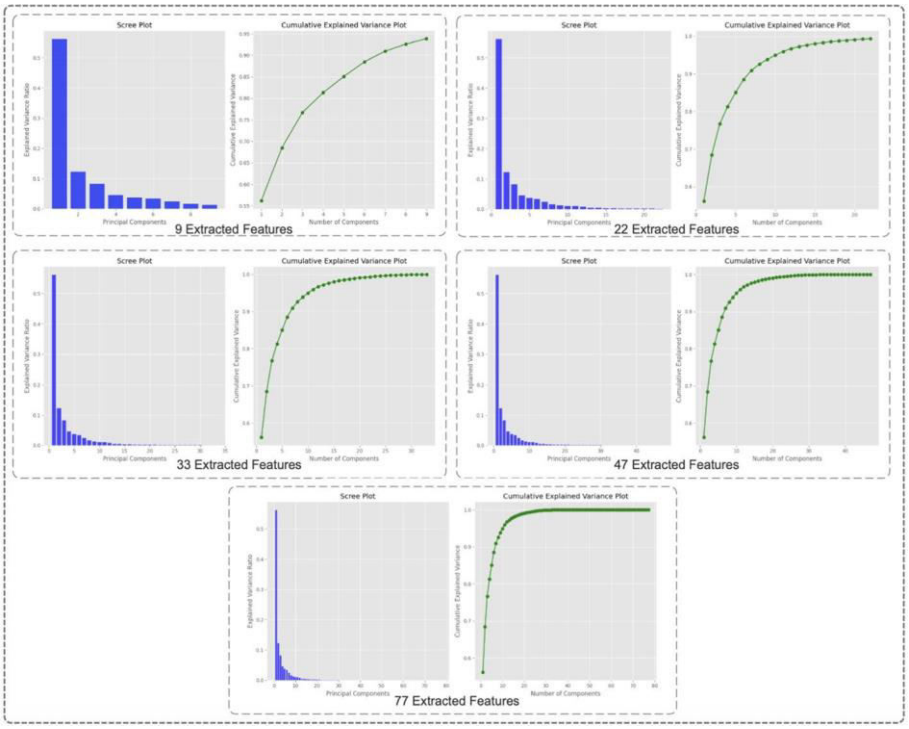


Table 1: FS vs. FE for binary classification with 9 features

Models	Accuracy (%)	Precision (%)	Re-call (%)	F1-score (%)	MCC	FS (s)	Training (s)	Inference (ms)
Feature selection								
DT	**80.73**	**79.50**	**77.74**	**78.44**	**0.5721**	**8.38**	**0.22**	**12.11**
RF	79.07	78.53	74.55	75.73	0.5293		7.56	801.32
kNN	77.44	82.90	69.38	70.66	0.5050		1.09	709,996.65
NB	78.99	82.82	71.94	73.58	0.5366		**0.11**	12.50
MLP	80.73	79.50	77.74	78.44	0.5721		37.3	136.17
Feature extraction								
DT	**86.54**	**85.12**	**86.33**	**85.62**	**0.7128**	**5.27**	**1.44**	**8.10**
RF	86.45	85.02	86.30	85.54	0.7127		18.61	838.21
kNN	71.00	76.33	76.79	70.99	0.3409		1.55	10,172.81
NB	83.35	81.76	82.68	82.15	0.6443		**0.13**	18.52
MLP	86.30	84.85	86.26	85.41	0.7111		81.58	139.20

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 3.911| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 7, Issue 4, July - August 2020 ||

DOI:10.15680/IJARETY.2020.0704017

Table 2: FS vs. FE for binary classification with 22 features

Models	Accuracy (%)	Precision (%)	Re-call (%)	F1-score (%)	MCC	FS (s)	Training (s)	Inference (ms)
Feature selection								
DT	**81.27**	**80.00**	**78.55**	**79.15**	**0.5853**	**7.82**	0.46	**11.85**
RF	77.72	84.92	69.30	70.57	0.5192		8.99	776.08
kNN	78.65	**85.26**	70.66	72.19	0.5398		**0.07**	196,772.36
NB	78.34	85.02	70.24	71.69	0.5324		0.17	28.58
MLP	81.27	80.00	78.55	79.15	0.5853		56.56	174.12
Feature extraction								
DT	85.94	84.49	85.55	84.94	0.7119	**4.92**	1.84	**12.71**
RF	86.54	85.11	86.37	85.63	0.7147		26.44	631
kNN	64.29	62.85	63.80	62.82	0.7287		**0.05**	193,070.46
NB	84.77	83.26	84.75	83.83	0.6799		0.19	37.25
MLP	**86.53**	**85.11**	**86.42**	**85.64**	**0.7151**		128.43	478.01

Table 3: FS vs. FE for binary classification with 33 features

Models	Accuracy (%)	Precision (%)	Re-call (%)	F1-score (%)	MCC	FS (s)	Training (s)	Inference (ms)
Feature selection								
DT	86.40	84.96	86.19	85.47	0.7114	**8.28**	0.64	**17.69**
RF	85.90	84.45	86.17	85.07	0.7059		11.74	848.32
kNN	83.75	**86.96**	78.30	80.30	0.6469		**0.13**	231,367.82
NB	79.92	85.77	72.51	74.33	0.5675		0.27	40.52
MLP	**86.45**	**85.01**	**86.29**	**85.54**	**0.7129**		75.38	184.73
Feature extraction								
DT	86.83	85.42	86.59	85.91	0.7201	**6.13**	3.13	**11.58**
RF	86.58	85.15	86.40	85.67	0.7154		38.67	657.02
kNN	**89.10**	**87.78**	**89.28**	**88.39**	**0.7669**		**0.06**	227,237.45
NB	83.37	83.56	79.55	80.89	0.6299		0.27	45.47
MLP	86.54	85.11	86.35	85.62	0.7151		45.43	84.14

Table 4: FS vs. FE for binary classification with 47 features

Models	Accuracy (%)	Precision (%)	Re-call (%)	F1-score (%)	MCC	FS (s)	Training (s)	Inference (ms)
Feature selection								
DT	84.23	83.44	81.68	82.40	0.6509	**5.47**	0.86	**30.00**
RF	86.23	84.82	85.76	85.23	0.7057		15.96	524.71
kNN	82.82	82.28	79.54	80.55	0.6176		**0.09**	148,293.32
NB	81.20	84.15	75.15	77.02	0.5861		0.28	44.95
MLP	**86.52**	**85.09**	**86.34**	**85.61**	**0.7142**		67.58	72.34
Feature extraction								
DT	83.81	82.92	85.61	83.26	0.6848	**5.01**	6.50	**10.47**
RF	**86.94**	**85.54**	**86.72**	**86.04**	**0.7225**		35.72	569.59
kNN	86.76	85.34	87.16	86.00	0.7129		**0.05**	147,798.15
NB	69.74	70.52	59.96	58.85	0.2859		0.21	43.87
MLP	86.59	85.16	86.39	85.67	0.7152		59.03	105.07

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 3.911| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 7, Issue 4, July - August 2020 ||

DOI:10.15680/IJARETY.2020.0704017

Table 5: FS vs. FE for binary classification with 77 (full) features

Models	Accuracy (%)	Precision (%)	Re-call (%)	F1-score (%)	MCC	FS (s)	Training (s)	Inference (ms)
Feature selection								
DT	78.28	76.59	75.24	75.79	0.5128	**0**	1.65	**24.21**
RF	**88.22**	**86.99**	**89.56**	**87.69**	**0.7651**		12.94	553.13
kNN	80.55	80.74	83.44	80.19	0.6413		**0.09**	188,417.64
NB	59.57	71.04	67.75	59.20	0.3865		0.36	55.02
MLP	86.58	85.15	86.38	85.66	0.7153		70.78	83.49
Feature extraction								
DT	74.68	73.37	75.10	73.64	0.4845	**3.98**	10.01	**12.25**
RF	**87.04**	**85.65**	**86.78**	**86.14**	**0.7243**		47.68	579.27
kNN	80.56	80.75	83.45	80.19	0.6414		**0.08**	186,251.17
NB	79.76	81.24	73.97	75.58	0.5473		0.28	63.89
MLP	86.59	85.16	86.39	85.67	0.7153		86.44	152.35

Regarding the classification performance, we initially explore the impact of an increasing number of features on the performance of both FS and FE methods. Expanding the number of features appears to enhance the performance of the FS model, while this increase shows no obvious effect on the FE model. Figure 2 illustrates that as the number of features increases, the performance of FS models generally improves from 9 features to 77 full features. In contrast, the performance of FE models remains nearly consistent, except the kNN model, which displays optimal performance with 33 features, as indicated in Fig. 3. While the performance of the best models in FS improves as the number of features increases from Tables 1, 2, 3, 4, and 5, the performance of certain models, like the decision tree, significantly decreases from Tables 3, 4, and 5. This trend aligns with the expectation that as the number of selected features increases, more irrelevant or noisy features might emerge, potentially impacting the detection performance negatively.

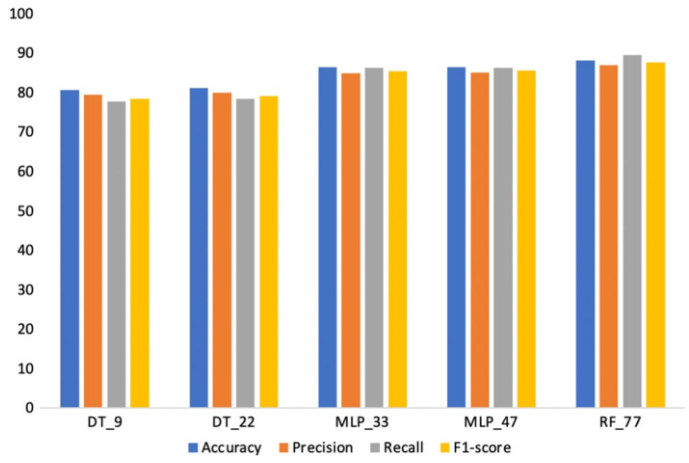


Figure 2: The best performance of FS models for binary classification

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 3.911| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 7, Issue 4, July - August 2020 ||

DOI:10.15680/IJARETY.2020.0704017

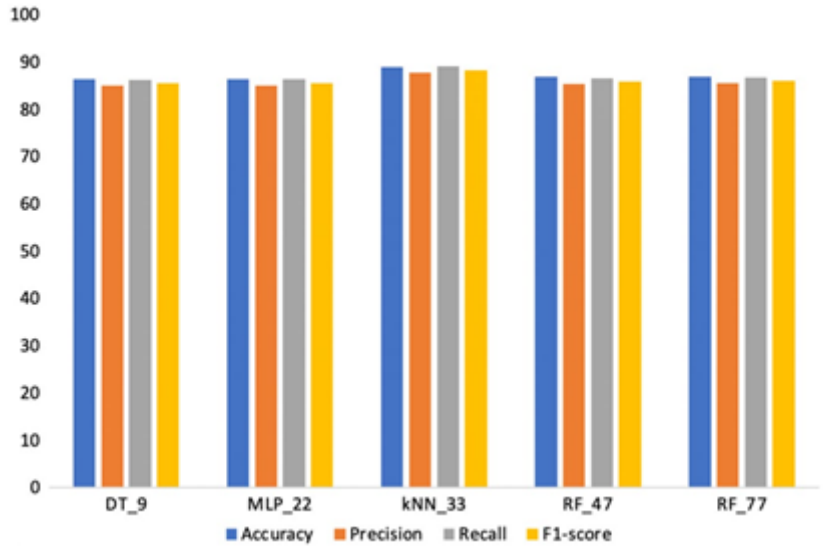


Figure 3: The best performance of FE models for binary classification

V. DISCUSSION

In the discussion of our findings, it is evident that the performance and runtime of feature selection and extraction methods in binary classification exhibit a range of outcomes that underscore the trade-offs between accuracy, computational efficiency, and feature set size. Our results, detailed in Tables 1 through 5, demonstrate that as the number of features increases, there are notable variations in the effectiveness of different methods. For each feature set size—9, 22, 33, 47, and 77—the iterations reveal that certain feature selection and extraction methods consistently deliver superior results in terms of accuracy, precision, recall, F1-score, and Matthews correlation coefficient (MCC). The highlighted values in bold and red within the tables indicate the best-performing configurations, reflecting the highest levels of classification performance and the most efficient processing times. Specifically, the optimal values across these metrics show that some methods achieve higher accuracy and better overall performance with fewer features, suggesting that they are more adept at extracting the most relevant information while minimizing the impact of irrelevant or redundant features. Conversely, methods that require more features might show improved performance but at the cost of increased computational demands and longer processing times. This trend emphasizes the importance of balancing feature set size with computational efficiency. Additionally, the results highlight that feature reduction, training, and inference times can significantly influence the practical applicability of these methods, with shorter times generally indicating more efficient processing. For instance, methods that excel in accuracy but require extensive feature sets and longer processing times might be less practical for real-time applications where quick inference is crucial. Furthermore, the varying results across different feature set sizes suggest that there is no one-size-fits-all solution; rather, the choice of method should be tailored to the specific needs of the application, including considerations for accuracy, speed, and computational resources. Overall, our discussion reveals that the best-performing feature selection and extraction methods are those that achieve a harmonious balance between high classification performance and efficient processing, and this balance is crucial for developing effective and practical machine learning models. Future work should focus on optimizing these methods further, exploring hybrid approaches that combine the strengths of multiple techniques, and adapting these methods to new and evolving data types and application contexts.

VI. CONCLUSION

Our comprehensive evaluation of feature selection and extraction methods for binary classification has illuminated several key insights that are pivotal for enhancing machine learning applications in various domains. The analysis reveals that the effectiveness of these methods is highly contingent upon the balance between classification accuracy, computational efficiency, and feature set size. The findings from Tables 6 through 10 underscore that while some methods perform exceptionally well with fewer features, delivering high accuracy, precision, recall, F1-score, and Matthews correlation coefficient (MCC), others may require a larger number of features to achieve similar levels of

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 3.911| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

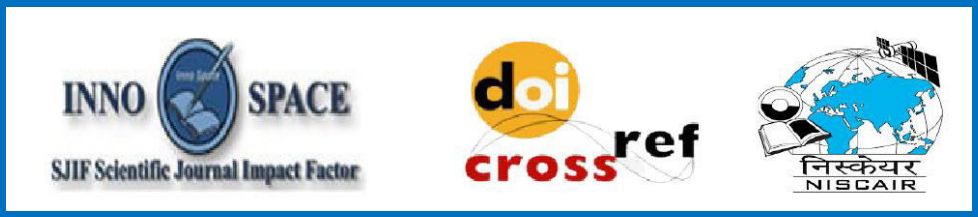
|| Volume 7, Issue 4, July - August 2020 ||

DOI:10.15680/IJARETY.2020.0704017

performance, albeit at the expense of increased processing times. The highlighted optimal values in the tables reflect the best outcomes in terms of both performance metrics and processing efficiency, demonstrating that certain feature selection and extraction techniques are particularly adept at identifying the most relevant features while minimizing computational overhead. This balance is crucial for practical implementations where real-time performance and resource constraints are significant considerations. The variation in results across different feature set sizes suggests that a tailored approach, considering the specific application requirements and available computational resources, is essential for achieving the best outcomes. Moreover, the impact of feature reduction and training times on the overall practicality of these methods emphasizes the need for continuous refinement and optimization. The practical implications of these findings extend to real-world applications where efficient and accurate threat detection is vital, such as in cybersecurity, medical diagnostics, and financial forecasting. Moving forward, future research should focus on further optimizing these methods, exploring innovative hybrid approaches that integrate the strengths of various techniques, and adapting these approaches to handle increasingly complex and dynamic datasets. By addressing these aspects, researchers and practitioners can develop more robust and efficient machine learning models that are better equipped to meet the evolving demands of diverse application domains. Ultimately, this comprehensive analysis provides a valuable foundation for advancing feature selection and extraction methodologies, contributing to the development of more effective and practical machine-learning solutions.

REFERENCES

1. Stellios, P. Kotzanikolaou, M. Psarakis, C. Alcaraz, J. Lopez, A survey of iotenabled cyberattacks: assessing attack paths to critical infrastructures and services, IEEE,Commun. Surv. Tutorials 20 (4) (2018) 3453–3495.
2. N. Sultana, N. Chilamkurti, W. Peng, R. Alhadad, Survey on sdn based network intrusion detection system using machine learning approaches, Peer-to-Peer.Netw. Appl. 12 (2) (2019) 493–501.
3. M.A. Khan, K. Salah, Iot security: review, blockchain solutions, and open challenges, Future Generat. Comput. Syst. 82 (2018) 395–411.
4. M. Nawir, A. Amir, N. Yaakob, O.B. Lynn, Internet of things (iot): taxonomy of security attacks, in: 2016 3rd International Conference on Electronic Design (ICED), IEEE, 2016, pp. 321–326.
5. S. Senthilkumar, C. Nivetha, G. Pavithra, G. Priyanka,S. Vigneshwari, L. Ramachandran, “Intelligent solar operated pesticide spray pump with cell charger”, International Journal For Research & Development In Technology, vol. 7, no. 2, pp. 285-287, 2017.
6. Symantec, Internet Security Threat Report, vol. 24, 2019. URL, https://docs.bro adcom.com/doc/istr-24-2019-en.
7. S.F. Yusufovna, Integrating intrusion detection system and data mining, in: 2008 International Symposium on Ubiquitous Multimedia Computing, 2008, pp. 256–259, https://doi.org/10.1109/UMC.2008.59.
8. P. García-Teodoro, J. Díaz-Verdejo, G. Macia-Fernandez, E. Vazquez, Anomalybased network intrusion detection: techniques, systems and challenges, Comput. Secur. 28 (1–2) (2009) 18–28, https://doi.org/10.1016/j.cose.2008.08.003.
9. P.V. Amoli, T. Hamalainen, G. David, M. Zolotukhin, M. Mirzamohammad, Unsupervised network intrusion detection systems for zero-day fast-spreading attacks and botnets, JDCTA, Int. J. Digit. Contents.Technol.Appl. 10 (2) (2016) 1–13.
10. M.J. Hashemi, G. Cusack, E. Keller, Towards evaluation of nidss in adversarial setting, in: Proceedings of the 3rd ACM CoNEXT Workshop on Big DAta, Machine Learning and Artificial Intelligence for Data Communication Networks, 2019, pp. 14–21.
11. C. Sinclair, L. Pierce, S. Matzner, An application of machine learning to network intrusion detection, in: Proceedings 15th Annual Computer Security Applications Conference (ACSAC’99), IEEE, 1999, pp. 371–377.
12. Javaid, Q. Niyaz, W. Sun, M. Alam, A deep learning approach for network intrusion detection system, in: Proceedings of the 9th EAI International Conference on Bio-Inspired Information and Communications Technologies, formerly BIONETICS), 2016, pp. 21–26.
13. R. Sommer, V. Paxson, Outside the closed world: on using machine learning for network intrusion detection, in: 2010 IEEE Symposium on Security and Privacy, IEEE, 2010, pp. 305–316.
14. S. Khan, E. Sivaraman, P.B. Honnavalli, Performance evaluation of advanced machine learning algorithms for network intrusion detection system, in: Proceedings of International Conference on IoT Inclusive Life (ICIIL 2019), NITTTR, Chandigarh, India, 2020, pp. 51–59, https://doi.org/10.1007/978-981-15-3020-3_ 6.



International Journal of Advanced Research in Education and TechnologY (IJARETY)

www.ijarety.in editor.ijarety@gmail.com