ISSN: 2393-8048
International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double Blind, Open Access, Peer-Reviewed, Refereed-International Journal.
SJIF Impact Factor =8.152, January-June 2025, Submitted in May 2025

Machine Learning-Driven Security Frameworks for Protecting User Privacy in IOT Ecosystems

Srikanta Kolay, Research Scholar (Computer Science & Engineering) Sardar Patel University, Balaghat, Madhya Pradesh
Dr. Swati Jaiswal (Supervisor) Sardar Patel University, Balaghat, Madhya Pradesh

Abstract

The proliferation of Internet of Things (IoT) devices has achieved transformations in different areas by empowering consistent organization and information trade. In any case, there have additionally been critical security and privacy challenges achieved by this expanded interconnectedness. In IoT ecosystems, machine learning-driven security frameworks have arisen as a powerful solution to safeguard client privacy. These organizations are exposed against a great many online dangers and privacy concerns. Disastrous actions that can actually hurt an organization are called intrusions. IoT networks are especially exposed against dangers to their security. Utilizing the ideal features chose during the Optimal Feature Vector Selection (OFVS) step, the Bi-Layer Intrusion Detection Model (BIDM) identifies intrusions. This technique hinders assaults while likewise going about as a cradle against real dangers. A significant Intrusion Detection System (IDS) benchmark, the KDD CUP 99 dataset, was utilized to assess the recommended approach. The proposed conspire's presentation was likewise broke down with the NSL-KDD and CICIDS-2018 datasets. Research on the OFVS show was likewise conducted utilizing the generally new IoT Organization Intrusion dataset. NumPy, pandas, Matplotlib, and Scikit-learn were a portion of the bundles utilized in the implementation. This system can altogether further develop IoT organization security and moderate the dangers related with DDoS assaults.

Keywords: Machine Learning-Driven Security, Frameworks, Protecting, Privacy, Internet of Things (IoT), Ecosystems

1. INTRODUCTION

The Internet of Things, or IoT, has completely changed the way we interact with the outside world. By integrating a vast array of devices into our daily lives, we are able to create networked systems that increase productivity, comfort, and effectiveness. IoT devices generate and exchange enormous amounts of data, from smart homes to linked medical services systems, which raises serious concerns about customer privacy and security. Strong security frameworks that can protect sensitive data are essential because these devices frequently operate in environments with lax security rules and are defenseless against other threats. Promising solutions to these challenges can be found in machine learning-based security frameworks, which use sophisticated analytics and data analysis to enhance the guarantee of user privacy in IoT environments.

ML techniques are especially appropriate for the perplexing and dynamic situations found in the Internet of Things. These cycles can continuously deconstruct huge measures of information, distinguish inconsistencies, and anticipate conceivable security breaks before they happen. ML-driven security frameworks can give a proactive guard instrument by continuously learning from new information and adjusting to arising dangers. Because of the sheer size and variety of the devices in question, traditional security gauges regularly miss the mark in IoT environments. Hence, this feature is fundamental.

The limit of machine learning to deal with different and unstructured information produced by different sensors and devices is one of the vital benefits of this innovation for IoT security. Machine learning calculations can recognize examples and relationships inside this information, working with the identification of irregular way of behaving that might demonstrate a security risk. For instance, machine learning models can dissect occasions of organization traffic to distinguish unforeseen information streams that might demonstrate an approaching assault. Fundamentally, machine learning (ML) can be utilized to monitor device

conduct, distinguishing take offs from regular use designs that could show a device has been captured.

The capacity of machine learning-driven security systems to provide ongoing threat detection and response is another fundamental feature. Conventional security systems frequently rely on established guidelines and indicators to identify threats, which may not be sufficient to fend off fresh and sophisticated attacks. It's interesting to note that ML models may learn from validated data and adapt to new threats as they appear, providing a more adaptive and unique security setup. In the IoT ecosystem, where the pace and scale of the information age need fast and automated security reactions, this ongoing capability is essential.

Another area where ML-driven frameworks might provide significant benefits is privacy protection. Techniques like differential privacy and unified learning allow ML models to be developed on decentralized data sources without jeopardizing the privacy of specific client information. By enabling devices to collaboratively become proficient with a shared model, united learning restricts the amount of information that is available to other devices. Differential privacy makes the public aware of the data while ensuring that the privacy of specific customers is maintained even as the model benefits from the dataset as a whole. These approaches are particularly crucial in Internet of Things environments, where data handling involves delicate concepts and information privacy is of utmost importance.

In IoT ecosystems, machine learning-powered security frameworks have a great deal of promise to safeguard user privacy. These frameworks are able to tackle the unique security difficulties posed by IoT devices by employing machine learning (ML) to analyze large-scale data, identify anomalies, and provide continuous threat responses. Furthermore, privacy-preserving machine learning processes ensure that customer data remains protected throughout the security lifecycle, providing a comprehensive solution for protecting sensitive data in the networked world of the Internet of Things. As the IoT landscape continues to evolve, integrating state-of-the-art machine learning-driven security assessments will be essential to maintaining customer confidence and ensuring the secure operation of IoT devices.

2. LITERATURE REVIEW

Farooq et al. (2022) Examine the fundamental convergence of Internet of Things (IoT) security with machine learning (ML), including the two arrangements and ongoing challenges. They provide a thorough explanation of how anomaly detection, predictive assistance, and intrusion detection are three ways that ML techniques might enhance IoT security. The importance of machine learning in addressing security flaws in IoT devices and organizations is emphasized in the article. However, it also acknowledges several unresolved issues, including as scalability, privacy issues, and the need for robust machine learning models resistant to malicious attacks.

Javeed et al. (2021) provide a system that is partially driven by deep learning and enabled by Software-Defined Networking (SDN) for secure communication in Internet of Things environments. To improve the security of IoT communications, they combine deep learning techniques with SDN, focusing on continuous threat detection and mitigation. The report uses trial evaluations to demonstrate the feasibility of their methods, highlighting improved communication security and reduced susceptibility to various online threats. It contributes by providing a workable framework that strengthens IoT security by utilizing SDN and profound learning.

Mondal & Guha Roy (2022) Examine IoT security concerns and suggest a hybrid approach that combines blockchain technology and machine learning. They analyze the risks associated with IoT information security, such as data breaches and unauthorized access, and they suggest countermeasures that use blockchain technology for safe data storage and access management and machine learning (ML) for anomaly detection. The article addresses the emerging landscape of IoT security threats and provides insights on leveraging blockchain and machine

learning in tandem to enhance information integrity and confidentiality in IoT networks.

Oseni et al. (2022) provide a rational deep learning system aimed at robust intrusion detection in transportation companies with IoT capabilities. The focus of the study is on security challenges related to Internet of Things transportation, such as persistent inconsistency detection and mitigation of digital threats. Their architecture synchronizes interpretable deep learning models, enabling partners to understand and trust the decisions made by the intrusion detection system. The review emphasizes how important strength and dependability are to safeguarding the fundamental IoT infrastructure in transportation companies.

Petrov & Znati (2018) suggest a carefully considered, deeply learning-driven framework aimed at reducing security risks while using Bring Your Own Device (BYOD) settings. The study offers a novel approach that modifies deep learning protocols to analyze rational data and enhance security features specifically tailored to the novel concept of BYOD setups. Their approach emphasizes proactive risk assessment and threat detection, leveraging pertinent cues to improve the accuracy and effectiveness of security measures in diverse organizational settings.

Rajendran et al. (2022) Discuss the privacy and security issues surrounding edge knowledge in healthcare settings and make suggestions for machine learning-based solutions. The study addresses the challenges of safeguarding sensitive healthcare data handled at the perimeter of enterprises, emphasizing the combination of machine learning algorithms for anomaly identification and encryption techniques for data security. Their research enhances the development of safe edge intelligence arrangements that protect patient privacy and enable efficient medical care delivery through Internet of Things and artificial intelligence advancements.

3. PROPOSED METHODOLOGY

The suggested framework consists of the following:

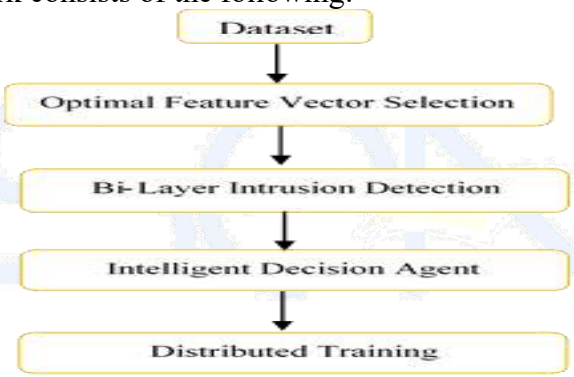

Figure 1: Proposed Framework

3.1.Feature Vector Selection

One of the layered diminishing procedures in machine learning is feature selection, which plans to choose the best or most splendid subset of features from the underlying feature course of action. In a few fields, a talented technique is commonly used to diminish the component of the information. The dataset may contain dreary, loud, and unimportant features. It is fundamental for both diminishing dimensionality and patching up the presentation.

The Feature selection calculation separates the promotion of the features connected with the undertaking. When contrasted with the classifier made utilizing the chose subset of significant qualities, the exactness of the classifier made utilizing the whole feature course of action is low. Feature selection offers a few benefits, for example, exact expectations, decreased dealing with time, etc. The exactness of a learning system might be influenced by the presence of inconsequential features. In the unlikely event if at least two features are present in a same set of data, it is typically regarded as having too many features. Pointless and boring features should be removed in order to optimize the learning process. Qualities The methods related to

the feature selection process are subset age, trait subset assessment, end measures, and approval. The working feature selection model is shown in Figure 2.

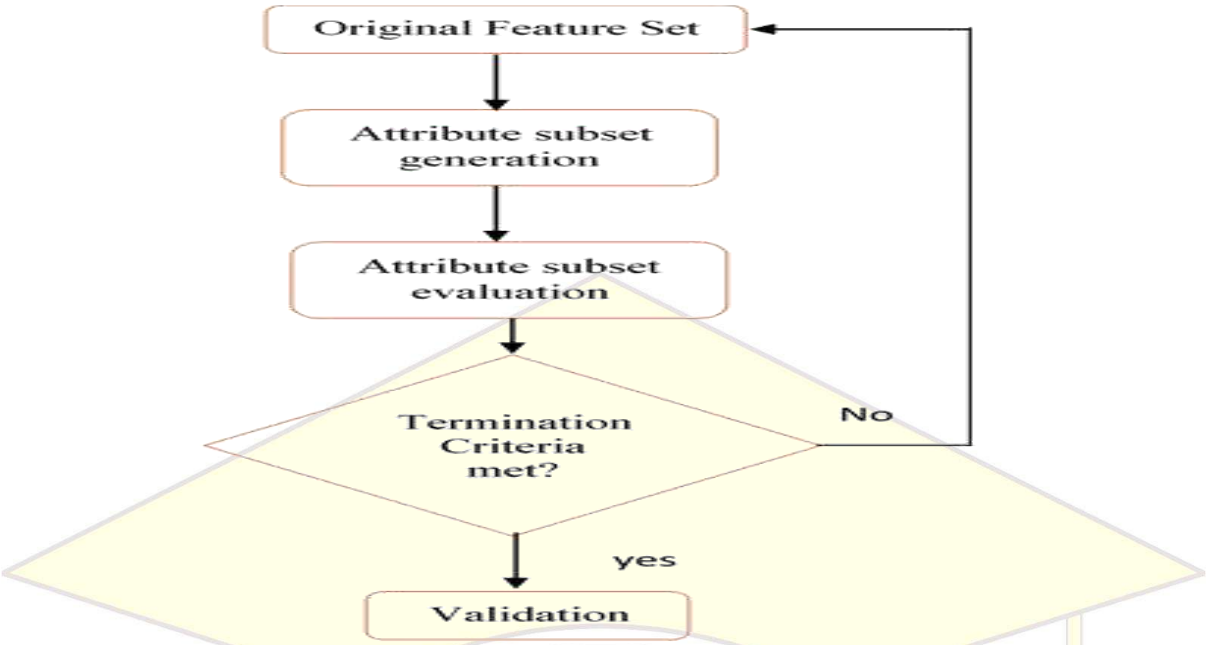

Figure 2: Steps in Feature Selection

3.2.Bi-Layer Intrusion Detection

An intrusion detection system (IDS) that utilizes two layers of detection to recognize malevolent movement is known as a bi-layer IDS. An assault mark informational index is utilized by an imprint-based intrusion detection system (IDS) as the essential layer of detection to distinguish malignant traffic. An intrusion detection system (IDS) that investigates organization traffic conduct to distinguish odd development that might demonstrate an assault fills in as the second layer of detection. There are a couple of benefits to utilizing a bi-layer strategy to intrusion detection over a solitary layer method. Most importantly, it can give a more complete inclusion of known assaults. Assaults that have previously been seen and contributed to the informational collection of imprints can be perceived by signature-based intrusion detection systems. If a singular continues to show odd way of behaving, conduct-based intrusion detection systems (IDSs) can recognize new endeavours to catch them. Second, the quantity of misleading up-sides created by the IDS can be decreased by utilizing the bi-layer method. IDSs in view of marks can produce a lot of misleading up-sides since they can contrast marks with genuine traffic that eventually contains cases that look like an assault signature. Conduct-based intrusion detection systems (IDSs) can decrease the quantity of bogus up-sides by only raising admonitions for traffic that acts peculiarly.

Third, the bi-layer method could give additional information with respect to the concept of an assault. IDSs in light of marks can distinguish the kind of assault, yet they can't give information with respect to the particular assault vector or the aggressor's goal. IDSs that depend on conduct can give additional information about the assault, for example, the particular orders that were completed or the records that were gotten to. This information can be utilized to examine assaults for criminological targets and to concentrate on responses to attacks. Albeit the bi-layer intrusion detection model is a generally new way to deal with intrusion detection, it has demonstrated guarantee as far as giving full inclusion, decreasing misdirecting potential gains, and giving additional information about assaults. Considering that IoT devices are much of the time less strong to go after than traditional IT systems, the bi-layer way to deal with intrusion detection is probably going to turn out to be progressively significant as the Internet of Things (IoT) continues to develop.

ISSN: 2393-8048
International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double Blind, Open Access, Peer-Reviewed, Refereed-International Journal.
SJIF Impact Factor =8.152, January-June 2025, Submitted in May 2025

3.3.Distributed Training on IoT Node

One kind of machine learning training that appropriates the training system among numerous IoT nodes is called distributed training on IoT nodes. It ought to be doable to deal with the training system's presentation, abbreviate the time it takes to create a model, or further develop the training system's security. On IoT nodes, training can be distributed in various ways. A well-known approach is to utilize a unified learning designing. Each IoT centre point in a combined learning designing makes a close by model utilizing its own information. In this manner, the nearby models are accumulated to make an international model. This technique can be applied to further develop the training system's showcase by diminishing how much information that should be sent between the Internet of Things nodes and the central server.

$$A_{i,j} = A_{i,j}\alpha\frac{\delta(F_{loss}(A,B|S))}{\delta A_{i,j}} \quad (1)$$

$$B_{i,j} = B_{i,j}\alpha\frac{\delta(F_{loss}(A,B|S))}{\delta B_{i,j}} \quad (2)$$

Utilizing a shared learning design is an additional technique for dealing with scattered training on IoT nodes. The Internet of Things nodes discuss transparently with one more in a distributed learning designing to share updates and information. Decreasing how much information that should be moved to a central server is one method for working on the security of the training system. On IoT nodes, distributed training can be a very complicated strategy. There are various variables to consider, including the model's size and intricacy, how much information that is accessible, the information move limit and latency between the IoT nodes, and the security prerequisites. Regardless, chipping away at the scalability, security, and show of machine learning applications for IoT can profit from distributed training.

4. EXPERIMENTAL RESULTS

We utilized four intrusion detection datasets: KDD CUP 99, NSL-KDD, CICIDS 2018, and IoT Organization Intrusion Dataset, to assess the suitability of a few inconsistency detection strategies. Different attacks are organized in these datasets into classifications like DoS, U2R, R2L, and Test. With a sum of 15 sub-attacks, the CICIDS 2018 dataset consists of six divisions: Savage power, DoS, web attack, infiltration, Botnet ARES, and Ports can. About a third of the KDD CUP 99, NSL-KDD, IoT Organization Intrusion, and CICIDS 2018 datasets were provided for testing in our examination.

Table 1: Measuring Performance of Various Attacks in KDD CUP 99 Dataset

Feature Extraction Method		**PCA**				**KPCA**				**LDA**				**SVM-CA**			
Attack	**Classifier**	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**Accuracy**	**Precision**	**Recall**	**F1 Score**
DoS Attack	NB	53	52	55	57	53	54	60	61	57	61	55	53	57	58	55	61
	RF	55	76	79	74	81	75	68	80	87	86	84	79	85	86	87	85
	ID3	75	80	77	74	77	74	76	78	80	77	80	76	84	84	86	86
	Ada Boost	82	77	76	75	76	77	81	77	78	79	68	72	85	85	91	86
	LR	77	76	79	74	81	75	68	80	87	86	84	79	85	86	87	85
	KNN	76	79	74	71	66	66	65	66	82	74	77	82	84	82	78	80
	NB	80	74	79	75	71	67	65	67	52	44	51	57	59	54	44	57
	RF	69	70	72	73	77	76	75	76	79	81	80	82	64	68	65	64

ISSN: 2393-8048
International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double Blind, Open Access, Peer-Reviewed, Refereed-International Journal.
SJIF Impact Factor =8.152, January-June 2025, Submitted in May 2025

U2R Attack	ID3	87	88	85	86	85	85	80	82	76	78	79	81	96	95	99	97
	Ada Boost	88	86	88	91	78	88	85	86	84	86	84	83	96	94	96	95
	LR	82	85	85	87	80	87	75	75	69	72	75	76	95	99	98	99
	KNN	82	77	76	79	75	75	77	79	81	77	77	77	88	86	87	87
R2L Attack	NB	79	75	76	81	80	76	70	72	78	79	80	81	62	58	56	59
	RF	80	70	72	78	80	77	79	82	82	88	91	89	67	72	66	70
	ID3	78	77	77	77	80	77	81	82	82	77	74	75	91	86	88	88
	Ada Boost	90	89	89	91	89	81	89	89	82	88	86	85	97	99	97	97
	LR	88	87	88	87	90	86	79	77	82	86	87	85	97	96	91	89
	KNN	89	87	86	82	78	79	77	77	76	79	80	77	89	90	90	96
Probe Attack	NB	77	78	77	80	75	79	78	76	69	61	66	67	69	69	68	69
	RF	72	77	86	84	80	76	75	74	71	72	71	67	77	73	76	78
	ID3	85	86	87	89	88	87	87	85	84	85	87	92	95	94	96	97
	Ada Boost	87	87	88	89	86	87	85	86	76	80	85	89	92	96	95	95
	LR	88	89	85	85	82	87	86	84	80	82	80	86	89	92	97	94
	KNN	80	78	78	77	91	90	89	90	78	79	78	81	98	92	92	96

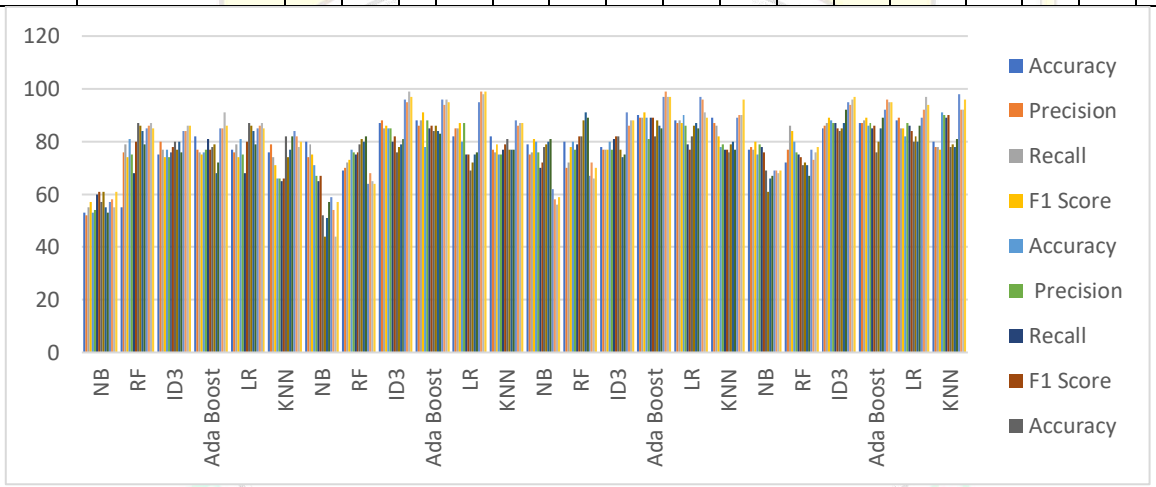

Figure 3: Measuring Performance of Various Attacks in KDD CUP 99 Dataset

It is vital to look at the model's showcase utilizing different measurements that catch particular parts of viability during the evaluation interaction. For this situation, the model's precision, precision, survey, and F1-Score were discovered and contrasted and different models. One commonly utilized measurement is precision, which shows the level of exactly requested occasions contrasted with the absolute number of cases. It gives an expansive outline of the prescient force of the model in doling out the suitable class name. Notwithstanding, depending entirely on exactness is probably not going to give a total comprehension of the model's result, especially when datasets are lopsided or the expense of mistaken categorization fluctuates between classes.

$$\textbf{Accuracy} = \frac{T\_P + T\_N}{T\_P + F\_P + F\_N + T\_N} \qquad (3)$$

Precision is defined as the proportion of genuine up-sides (very much portrayed occasions) to all anticipated up-sides (cases expected to fall into a particular class). It evaluates the model's ability to smother misleading up-sides, demonstrating the accuracy with which it distinguishes occasions that fall into a specific class. Assess, on the other hand, computes the level of genuine

advantages in relation to each genuine assurance (cases that really fall into a particular class). It assesses how well the model lessens misleading negatives, demonstrating how well it surveys or responds to examples of a specific class.

At the point when accuracy and survey are combined, we can acquire a more profound comprehension of the model's result, especially in situations where misclassifying occasions in one class might bring about various expenses or results when contrasted with another class.

$$\mathbf{P} = \frac{\mathbf{T\_P}}{\mathbf{T\_P + F\_P}} \qquad \mathbf{(4)}$$

A high F1-Score implies that the model accomplishes a good harmony between survey (which decreases misdirecting negatives) and precision (which limits deceptive up-sides). When the dataset is unequal or there is a critical monetary expense related with inaccurately characterizing cases into various gatherings, it is critical.

$$\mathbf{R} = \frac{\mathbf{T\_P}}{\mathbf{T\_P + F\_N}} \qquad \mathbf{(5)}$$

$$\mathbf{F_1\ Score} = \frac{\mathbf{2 * P * R}}{\mathbf{P + R}} \qquad \mathbf{(6)}$$

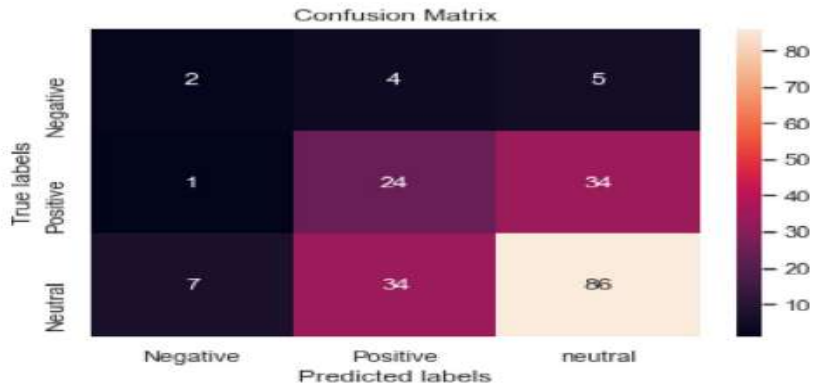

Figure 4: Confusion matrix

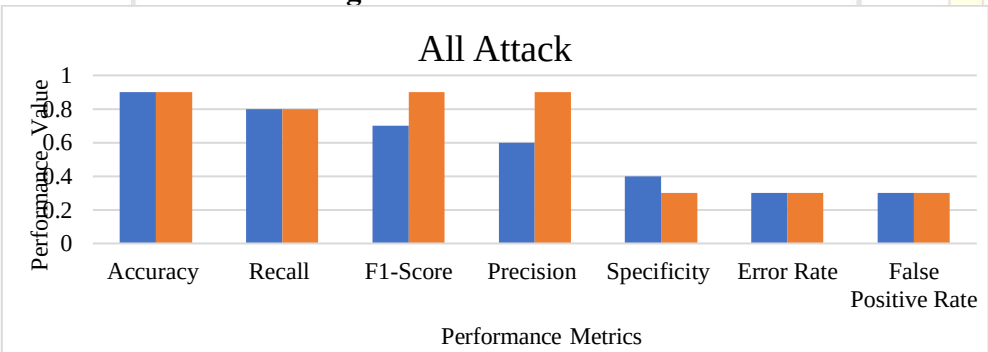

Figure 5: suggested model comparison

5. CONCLUSION

The research focuses on developing a machine learning framework to protect user privacy in Internet of Things environments by enabling persistent threat identification, adaptable reaction strategies, and aggressive anomaly detection. Three stages make up the system: distributed training in hazy nodes, feature extraction, and bi-layer intrusion detection. Feature extraction using a sophisticated Help Vector Machine with a Correlation Calculation is part of the primary stage. High level machine learning procedures are utilized in the second stage to distinguish and portray potential intrusions or oddities inside the IoT venture precisely. Redesigning ability and scalability through scattered training in cloudy nodes is the last step. Four datasets are utilized to assess the machine learning model's exhibition: the NSL-KDD dataset, the KDD CUP 99 dataset, the IoT network intrusion dataset, and the CICIDS 2018 dataset. The goal is to improve forecasts as well as the classifier's overall accuracy and suitability for identifying anomalies and breaches within IoT enterprises.

REFERENCES

1. *Alsharif, M., & Rawat, D. B. (2021). Study of machine learning for cloud assisted iot security as a service. Sensors, 21(4), 1034.*
2. *Bhayo, Jalal, Jafaq, Riaz, Ahmed, Awais, Hameed, Sufian, Shah, Syed Attique, 2022. A time-efficient approach toward DDoS attack detection in IoT network using SDN. IEEE Internet Things J. 9 (5), 3612–3630. http://dx.doi.org/10.1109/JIOT.2021.3098029.*
3. *Chen, Yi-Wen, Sheu, Jang-Ping, Kuo, Yung-Ching, Van Cuong, Nguyen, 2020. Designand implementation of IoT DDoS attacks detection system based on machine learning. In: 2020 European Conference on Networks and Communications (EuCNC). IEEE, pp. 122–127.*
4. *Farooq, U., Tariq, N., Asim, M., Baker, T., & Al-Shamma'a, A. (2022). Machine learning and the Internet of Things security: Solutions and open challenges. Journal of Parallel and Distributed Computing, 162, 89-104.*
5. *Javeed, D., Gao, T., Khan, M. T., & Ahmad, I. (2021). A hybrid deep learning-driven SDN enabled mechanism for secure communication in Internet of Things (IoT). Sensors, 21(14), 4884.*
6. *Mondal, K. K., & Guha Roy, D. (2022). Iot data security with machine learning blckchain: Risks and countermeasures. In Deep Learning for Security and Privacy Preservation in IoT (pp. 49-81). Singapore: Springer Singapore.*
7. *Oseni, A., Moustafa, N., Creech, G., Sohrabi, N., Strelzoff, A., Tari, Z., & Linkov, I. (2022). An explainable deep learning framework for resilient intrusion detection in IoT-enabled transportation networks. IEEE Transactions on Intelligent Transportation Systems, 24(1), 1000-1014.*
8. *Petrov, D., & Znati, T. (2018, October). Context-aware deep learning-driven framework for mitigation of security risks in BYOD-enabled environments. In 2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC) (pp. 166-175). IEEE.*
9. *Rajendran, S., Mathivanan, S. K., Jayagopal, P., Purushothaman Janaki, K., Manickam Bernard, B. A. M., Pandy, S., & Sorakaya Somanathan, M. (2022). Emphasizing privacy and security of edge intelligence with machine learning for healthcare. International Journal of Intelligent Computing and Cybernetics, 15(1), 92-109.*
10. *S. Suganyadevi, S. S. Priya, R. Menaha, S. Sathiya, P. Jha and S. B. S, "Smart Healthcare in IoT using Convolutional Based Cyber Physical System," 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), Mysuru, India, 2022, pp. 1-6, doi: 10.1109/MysuruCon55714.2022.9972679.*
11. *Taylor, R., Baron, D., Schmidt, D., 2015. The world in 2025 – predictions for the next ten years. In: 2015 10th International Microsystems, Packaging, Assembly and Circuits Technology Conference (IMPACT). pp. 192–195. http://dx.doi.org/10. 1109/IMPACT.2015.7365193.*
12. *Verma, Abhishek, Ranga, Virender, 2020. Machine learning based intrusion detection systems for IoT applications. Wirel. Pers. Commun. 111 (4), 2287–2310.*
13. *Wang, Jingjing, Jiang, Chunxiao, Zhang, Haijun, Ren, Yong, Chen, Kwang-Cheng, Hanzo, Lajos, 2020. Thirty years of machine learning: The road to Pareto-optimal wireless networks. IEEE Commun. Surv. Tutor. 22 (3), 1472–1514.*
14. *Xiao, Liang, Wan, Xiaoyue, Lu, Xiaozhen, Zhang, Yanyong, Wu, Di, 2018. IoT security techniques based on machine learning: How do IoT devices use AI to enhance security? IEEE Signal Process. Mag. 35 (5), 41–49.*
15. *Yaqoob, Ibrar, Hashem, Ibrahim Abaker Targio, Ahmed, Arif, Kazmi, SM Ahsan, Hong, Choong Seon, 2019. Internet of things forensics: Recent advances, taxonomy, requirements, and open challenges. Future Gener. Comput. Syst. 92, 265–275.*