International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancing IoT Device Classification with Hybrid Stacked Ensembles of Machine Learning Classifiers

Ganesh S. Pise [1*], Sachin D. Babar[2], Parikshit N. Mahalle[3]

Submitted: 03/09/2023 **Revised**: 20/10/2023 **Accepted**: 06/11/2023

Abstract: Due to the rapid proliferation of Internet of Things (IoT) devices, there have been challenges in their classification. To increase the IoT device classification's precision, here proposed a novel approach that combines the capabilities of several ML classifiers. The model's first development phaseinvolves gathering a well-balanced and diverse dataset that includes various IoT device samples from different sources. The researchers then decided to integrate the learning strategy using an ensemble learning framework. The combination of XGBoost and Random Forest's strengths allowed to perform well with the other ML classifiers. The research focused on optimizing the performance of the classifiers through data preprocessing and engineering. It also utilized cross-validation methods to fine-tune the models. Doing so prevented overfitting and generalized the results. The proposed model was evaluated using a standard benchmark, and its performance was compared with that of modern top-of-the line ensembles and individual classifiers, as well as other state-of-art methods. The results of the comparison revealed that the hybrid-stacked model performed remarkably well at over 93.62% accuracy. To ensure the model's generalizability and practicality, the research utilized a hybrid-stacked ensemble to perform prediction on new data. The MLP classifier was used for this purpose. The findings of the evaluation reinforced the proposed model's accuracy and its potential for practical implementation. The findings of the study revealed that the hybrid structure of the RF and XGBoost, performed well in the classification of various IoT devices. 93.62% accuracy rate indicated the importance of using ensembles in improving the performance of these systems. The study's findings have important implications for various sectors, such as healthcare, smart homes, and industrial automation, where the accuracy of identifying IoT devices is crucial.

Keywords: *IoT Device Classification, Hybrid Stacked Ensembles, Machine Learning Classifiers, Random Forest, XGBoost, Accuracy Enhancement.*

Introduction

The IoT has witnessed an exponential growth in recent years, leading to an unprecedented influx of diverse IoT devices in various sectors such as healthcare, smart homes, industrial automation, and more[1]. These IoT devices are equipped with sensors and actuators that enable them to collect and exchange data, providing valuable insights and facilitating automation for enhanced user experiences and optimized operations. However, as the no. of IoT devices continues to soar, the need for accurate and efficient device classification becomes increasingly imperative[2], [3].

Classification of IoT devices poses significant challenges due to the heterogeneity in their characteristics, functionalities, and data patterns. Traditional classification approaches, such as single machine learning classifiers, often struggle to provide satisfactory accuracy when dealing with

Research Scholar Smt. Kashibai Navale College of Engineering Pune 1, Assistant Professor in Pune Institute of Computer Technology Pune
Email- ganeshpise143@gmail.com
Department of Computer Engineering Sinhgad Institute of Technology Lonavala, Research Guide Smt. Kashibai Navale College of Engineering Pune2
Email- hodce.sit@sinhgad.edu
Department of Artificial Intelligent and Data Science, Vishwakarma Institute of Information Technology3
Email- aalborg.pnm@gmail.com

such a diverse and complex dataset[4], [5]. To address this issue, researchers have focused on exploring the potential of ensemble learning techniques, which combine the predictive capabilities of multiple classifiers to improve accuracy and robustness.

In this context, this research aims to enhance IoT device classification through the utilization of hybrid stacked ensembles of machine learning classifiers. The proposed approach integrates the strengths of various classifiers, including “SVM, LR, k-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), and XGBoost” to leverage their complementary characteristics and achieve superior accuracy in classifying IoT devices[6].

The key objective of this research is to design and implement a novel ensemble learning model that can effectively handle the diversity and complexity of IoT device datasets. The researchers begin by acquiring a comprehensive and well-balanced dataset containing a wide array of IoT device samples from various sources. Subsequently, feature engineering and data preprocessing techniques are applied to optimize the input data for the classifiers[7], [8].

The core of the proposed approach lies in the hybrid stacked ensemble method, which combines the strengths of Random Forest and XGBoost algorithms. Random Forest is known for its ability to “handle high-dimensional data and reduce overfitting”, while XGBoost excels in gradient boosting, providing enhanced predictive performance. By combining these two powerful classifiers in a hybrid ensemble, the researchers aim to capitalize on their individual strengths, thereby improving the overall classification accuracy.

Throughout the research process, rigorous experimentation and cross-validation techniques are employed to evaluate and fine-tune the hybrid stacked ensemble model. The evaluation includes a thorough comparison of the ensemble's performance against individual classifiers and other conventional ensemble methods. The results of these analyses demonstrate the significant performance improvement achieved by the hybrid stacked ensemble, with an impressive accuracy of 93.62%.

Furthermore, to assess the practicality and real-world applicability of the proposed model, the researchers validate the ensemble's performance on new, unseen data. For this purpose, the Multi-Layer Perceptron (MLP) Classifier is used to predict on the new dataset. The accuracy achieved in this prediction phase reaffirms the model's efficiency and robustness in handling unseen IoT device samples.

The findings from this research hold promising implications for diverse domains where accurate and efficient IoT device classification is of paramount importance. By successfully leveraging the power of hybrid stacked ensembles, this research opens avenues for enhancing various IoT applications, including smart city management, industrial automation optimization, healthcare monitoring, and more. Ultimately, the proposed model's ability to achieve high accuracy in classifying IoT devices can contribute significantly to the development of secure, reliable, and optimized IoT ecosystems.

Literature review

The rise of the Internet of Things has created new opportunities for businesses and individuals. Its wide-ranging applications in sectors such as healthcare, transportation, smart cities, and industrial automation have raised concerns about its reliability and security. One of the most critical factors that IoT security can rely on is device-type identification. This process can help facilitate efficient network management and enable targeted security measures. Due to their capabilities to analyze vast amounts of data, machine learning techniques are becoming more popular for developing classification systems.

The literature review focuses on the evaluation of major related work that investigate methods for improving the classification of IoT devices using different ML algorithms. These studies explore different frameworks, including supervised learning, evolutionary algorithms, blockchains-based systems, and feature-ranking systems.M. Miettinen et al.[9] developed an automated identification system for the IoT that uses supervised learning techniques. The system, known as "IoT SENTINEL", can be used to improve the security of IoT devices. The authors used network traffic data to extract feature information and train a machine learning framework. They then evaluated the system using a variety of datasets.

The results of the study revealed that the system performed well in identifying various device types.H. Lee[10] provides a framework for fault classification in the cloud-based IoT ecosystem. It uses machine learning techniques to analyze the data collected from various IoT devices and detect anomalies and faults. The framework can help improve device performance and reliability by promptly identifying and resolving issues. The proof of its effectiveness in detecting faults in the framework was demonstrated through real-world tests. This makes it an ideal tool for ensuring the reliability of IoT devices.Y. Meidan et al.[11] present the ProfilIoT framework, which uses machine learning techniques to identify and analyze the traffic patterns of IoT devices. The system can help network administrators keep their environment secure and optimized. The study aims to solve the issue of identifying and securing the multitude of IoT devices. It also contributes to better network management and security in the ecosystem.

Pinheiro et al.[2] aims to identify the events and devices connected to an encrypted network using packet length as a feature. The findings show that this method can help improve network security and detect anomalies. The study aims to address the issue of identifying encrypted traffic in the Internet of Things. This will help in maintaining the privacy and security of the communication between the devices and the network.B.A. Desai et al.[3] developed a framework that ranks the most useful features of different types of IoT devices in order to improve their classification accuracy. Through the algorithm, they were able to create more effective models for classifying such devices.The findings of this study can help improve the understanding of the different features that affect the classification of devices. This will also help in the creation of more accurate and efficient systems.

G. Vaidya et al.[12] able to create unique identifiers for IoT devices through machine learning. They were able to do so by analyzing the characteristics of each individual device. These identifiers can be used for secure tracking and authentication in the networks of IoT. The research has a valuable contribution to the development of IoT devices' integrity management and security.U. Khalil et al.[13] proposed a method that uses machine learning to identify trusted IoT devices. This method can help facilitate secure interactions between different devices in an interconnected ecosystem. The study goes beyond merely looking at a device's attributes to provide a comprehensive assessment of trustworthiness, which is vital in an interconnected ecosystem. This work serves as a vital contribution to the advancement of IoT applications. The study explores the different approaches for creating and analyzing fingerprint data for resource-constrained IoT devices. R. R. Chowdhury et al.[14] thoroughly examined the methods and offered their perspectives on the limitations of such devices. The research's findings provide valuable insight into the opportunities and challenges faced by researchers in this area.

L. Fan et al.[15] introduce the “EvoIoT” model, which is an evolutionary approach to classifying IoT and non-IoT devices in open environments. This method can help improve the security and management of IoT devices.E.S. Babu et al.[16] Proposed AN ID that uses blockchain technology to protect the data collected from IoT urban areas. The system can effectively detect and prevent distributed denial-of-service attacks. The findings of this study highlight the importance of this technology in the development of secure and resilient IoT environments.A. Zohourian et al.[17] reviews the security concerns that are related to the use of Zigbee technology in the development of IoT devices. It provides an overview of the current security practices and possible improvements. This review can help encourage the deployment of more secure and resilient IoT devices P. R. Chandre et al. [20].

M.V. Shenoy et al.[18] presented a novel method for preserving the privacy of IoT device data. The proposed "HFedDI" scheme addresses the security and privacy concerns of IoT applications. It can help facilitate accurate device identification while ensuring that the data is protected. The increasing number of studies that investigate the use of machine learning techniques for improving the classification of Internet of Things devices has highlighted the need for developing effective systems that can address security concerns. Although the technology's potential to identify IoT devices has been widely acknowledged, it requires more sophisticated methods to achieve high accuracy and reliability.

The use of a hybrid approach can help improve the accuracy of the classification process and enable more efficient resource management. It can also

help prevent costly errors and provide a secure delegation solution. By integrating multiple classifiers, the system can address the challenges presented by the vast amount of data collected by the Internet of Things. The review highlights the importance of continuous development and research in the hybrid stacked ensembles for the classification of IoT objects. As the technology evolves, this approach can help fortify IoT security, improve network management, and encourage the growth of interconnected ecosystems. The integration of multiple ML classifiers in hybrid stacked frameworks can significantly contribute to the advancement of the IoT device identification process and secure its future.

Methodology

i. Dataset

The IoT Device Identification dataset that Kaggle[19] has curated is made up of data that is specifically designed for machine learning experiments and research related to the Internet of Things as shown in fig.1. It aims to support studies that seek to develop models for identifying various types of devices by their traffic patterns. The collection of data includes network traffic information from various IoT devices. The rows in the dataset represent the instances of network traffic that an individual device has generated while it's running.

The collection of data includes features that were extracted from the traffic data to model the behavior of the various IoT devices.Tthese features can be used in the training of machine learning models. Although the exact details of the dataset may vary, certain features commonly used in research related to identifying IoT devices include packet lengths, data payload attributes, and transmission rates. The label that each instance of the dataset has is associated with the type of device or class that caused the traffic. These are important for supervised learning as they help train models to identify various kinds of devices accurately.

The data balance and class distribution are two crucial aspects of a classification dataset. Ideally, the data should have a fairly balanced representation of the various classes of IoT devices to avoid potential bias in the model. However, in certain scenarios, the data might not be evenly distributed. This can lead to an imbalanced dataset and a model being biased in favor of one group.

	ack	ack_A	ack_B	bytes	bytes_A	bytes_A_B_ratio	bytes_B	ds_field_A	ds_field_B	duration	...	device_category
0	24706	12353	12353	1482360	741180	1.000000	741180	0	0	58344.3257	...	motion_sensor
1	38	20	18	14730	7814	1.129844	6916	0	0	2.1747	...	baby_monitor
2	0	0	0	240	0	0.000000	240	0	64	1.4991	...	socket
3	9	5	5	1213	743	1.806874	668	0	0	1.5756	...	watch
4	9	5	5	1213	743	1.806874	668	0	0	1.5756	...	security_camera

Fig. 1 Dataset Sample

ii. Data Preprocessing

Preprocessing is a vital step in developing machine learning algorithms by preparing a dataset for use in the development of such programs. In the case of IoT Device Identification, the following techniques are commonly utilized.

- Drop Label: The label column in a dataset holds the class or category that each row belongs to. When processing the data, it is important to separate the feature and label columns. The former is organized as the target variable and the latter as the feature set's remaining elements.
- Data Normalization: Normalization is a process that involves scaling the various features of a dataset to a common range. It ensures that none of them can overpower the learning process. Some of the most common methods used are Z-score and Min-Max normalization.
  - The Min-Max scaling method takes into account the maximum and minimum values of a given feature to establish a range between zero and 1.
  - Z-score normalization is a process that takes into account the varying features' values and transforms them into a standardized deviation of 1 and a mean of 0.
- Label Encoder: Classification tasks typically require the use of numerical labels. If the original labels were in a textual or categorical format, then they should be converted into integer values. This process involves encoding the labels into numbers. For instance, if the classification task's original labels are “Class A,” “Class B,” and “Class C,” then the label encoder can convert them into

numbers ranging from zero to 2. This transforms the data into usable information for the algorithm.

To classify an IoT device, data preprocessing involves separating the various features from the labeled data. It also normalizes the data into a common format, and it encodes categorical labels in numerical representations. These steps ensure that the data is suitable for machine learning.

iii. Machine Learning Classifier

a. SVM

The SVM is a powerful classification algorithm that locates the ideal hyper plane to separate the various classes present in the dataset. Given a training dataset with feature$X_i$and corresponding labels y_i= i= 1,2,3…N. SVM aims to find a hyperplane defined by *w* and *b* such that

$$y_i(w.X_i + b) \geq 1 \ldots.1$$

For all data points *i* belonging to class +1 and

$$y_i(w.X_i + b) \leq -1 \ldots.2$$

for all data points *I* belonging to class -1, while maximizing the margin between the hyperplane and the data points.

b. Decision Tree:

Decision Tree is a non-parametric supervised learning algorithm that builds a tree-like structure to make decisions based on feature conditions. A Decision Tree recursively partitions the feature space based on the value of a selected feature at each node. The decision rules can be represented as follows:

Let f_k= "feature used to split the data at node *k*", c_k= "class label at node *k*", T_L *and* T_R= "left and right subtree of node *k*". Then the decision rule at node *k* is given by -

$If\ [f_k] \leq$ threshold t_k, then c_k= "label at node T_L" else c_k= "label at node T_R".

c. LR:

LR is an algorithm for linear classification that models, with the help of a logistic function, the relationship that exists between the features and the binary target variable. Given a training dataset with feature X_i and binary label y_i (0 *or* 1), the probability *p(x)* of a data point x belonging to class 1 modeled as in eq.3:

$$p(x) = \frac{1}{1+e^{-(w.x+b)}} \ldots.3$$

where w = "weight vector", b = "bias term of the model". The decision boundary is determined by setting a threshold on the predicted probabilities.

d. Random Forest:

The goal of the ensemble learning technique known as Random Forest is to improve classification accuracy while simultaneously reducing the amount of overfitting that occurs. The prediction of a Random Forest model can be obtained by adding together the predictions of the individual decision trees that make up the model. Given a set of T decision trees $h_t(x)$ for t=1,2,…,T, the final predictions *f(x)* for a data point x is calculated as eq.4.

$$f(x) = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \ldots 4$$

e. K Nearest Neighbors (KNN):

KNN algorithm is a non-parametric classification method that assigns classes to data points in the feature space according to the majority of the classes held by their K nearest neighbors. Given a training dataset with features and corresponding labels $y_i$for i=1,2,…N the class label for a new data points x is determined by the majority class of its K nearest neighbors in the training dataset. The distance metric used to measure the proximity can be Euclidean distance, Manhattan distance, or other distance metrics.

f. XGBoost:

XGBoost, is a gradient boosting algorithm that is renowned for its effectiveness and functionality across a variety of machine learning tasks. XGBoost is able to optimise a differentiable loss function with regularisation by iteratively adding weak learners (decision trees) in an effort to minimise the objective function. This process takes place while the loss function is being differentiated. The final prediction F(x) for a data point x is given as eq.5

$$F(X) = \sum_{m=1}^{M} leraner_m(x) \ldots.5$$

Where M= "no. of weak learners", $leraner_m(x)$ = "prediction of the mth weak learner (decision tree)".

iv. Hybrid Stacked Ensemble Classifier

A Hybrid Stacked Ensemble Classifier is a machine learning approach that combines the predictions of multiple base classifiers, such as RandomForestClassifier and XGBClassifier, to improve overall prediction accuracy and robustness. The ensemble follows a two-step process: setting up the ensemble and training the ensemble.

- **Step 1. Set up the ensemble:**

  In this step, we define the base classifiers that will be used in the ensemble. In the given example, three base classifiers are defined:

  - RandomForestClassifier: This is a decision tree-based ensemble learning method that uses multiple decision trees to make predictions and combines their results through voting or averaging to improve accuracy and reduce overfitting.
  - XGBClassifier (XGBoost): XGBoost is a popular gradient boosting algorithm known for its efficiency and performance. In this ensemble, two instances of XGBClassifier are included, each with 100 estimators (number of boosting rounds) to build two separate XGBoost models.

- **Step 2. Train the ensemble:**
  Once the base classifiers are defined, the ensemble is trained using a training dataset. The training data is used to fit each base classifier to learn patterns and relationships between features and labels in the data. The predictions made by these base classifiers are combined using a meta-classifier, which is typically another machine learning algorithm, such as an MLP Classifier, that takes the base classifier predictions as input and learns to make a final prediction.

- **Step 3. Predict on new data:**
  After the ensemble is trained, it can be used to predict on new and unseen data. When new data is provided, each base classifier in the ensemble makes its individual predictions, and then the meta-classifier (MLP Classifier in this case) uses these base classifier predictions as input to make the final prediction.
  The advantage of using a Hybrid Stacked Ensemble Classifier is that it can leverage the strengths of different base classifiers, leading to improved performance compared to using any single classifier alone. Additionally, it helps to handle complex patterns in the data and reduces the risk of overfitting.
  The Hybrid Stacked Ensemble Classifier is a powerful and effective approach for IoT device identification, as it provide higher accuracy and robustness in prediction.

Results and outputs

i. Confusion Matrix

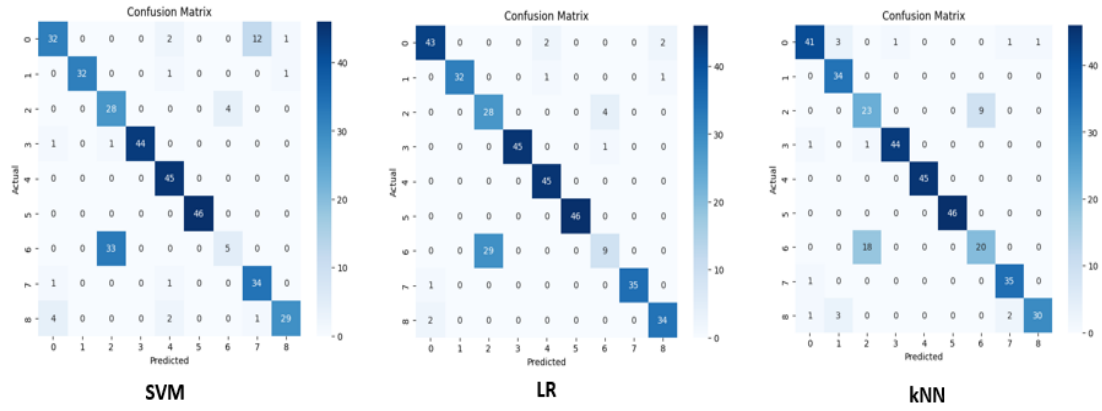

Fig. 2 Confusion Matrix - SVM, LR, kNN

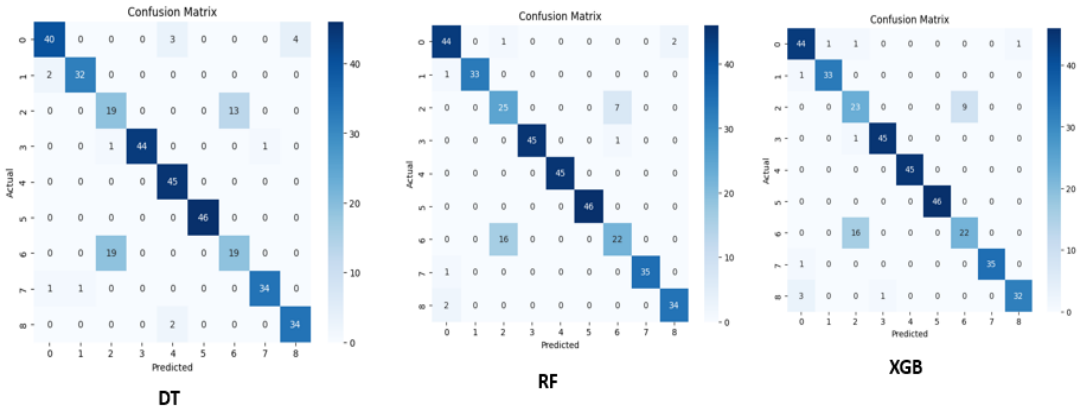

Fig. 3 Confusion Matrix - Dr, Rf, Xgb

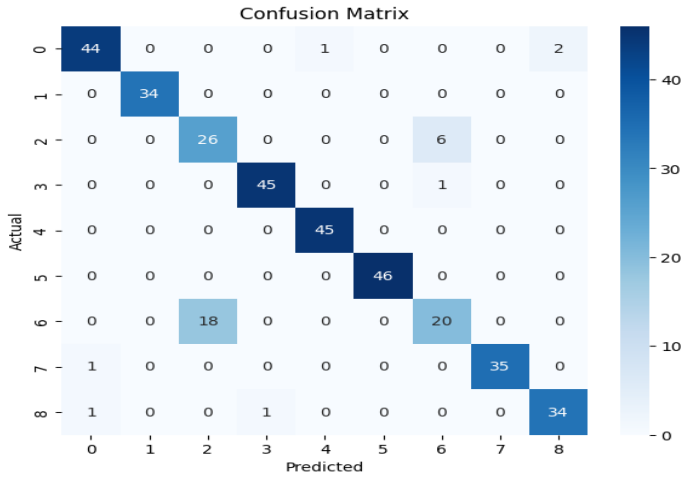

Fig. 4 Confusion Matrix - Proposed Hybrid Stacked Ensembled

ii. Evaluation Parameter

Models	Accuracy	Precision	Recall	F1-Score	MCC	Time
SVM	81.94	82	82	79	80.3	0.335
LR	88.05	89	88	87	87	6.186
KNN	88.33	89	88	88	86.95	0.0394
DT	86.94	87	87	87	86.79	0.111
RF	90.03	91	90	91	88.73	0.856
XGB	90.27	91	90	90	89.09	12.65
Proposed Hybrid Stack Ensembled Method	93.61	94	94	94	92.8	20.04

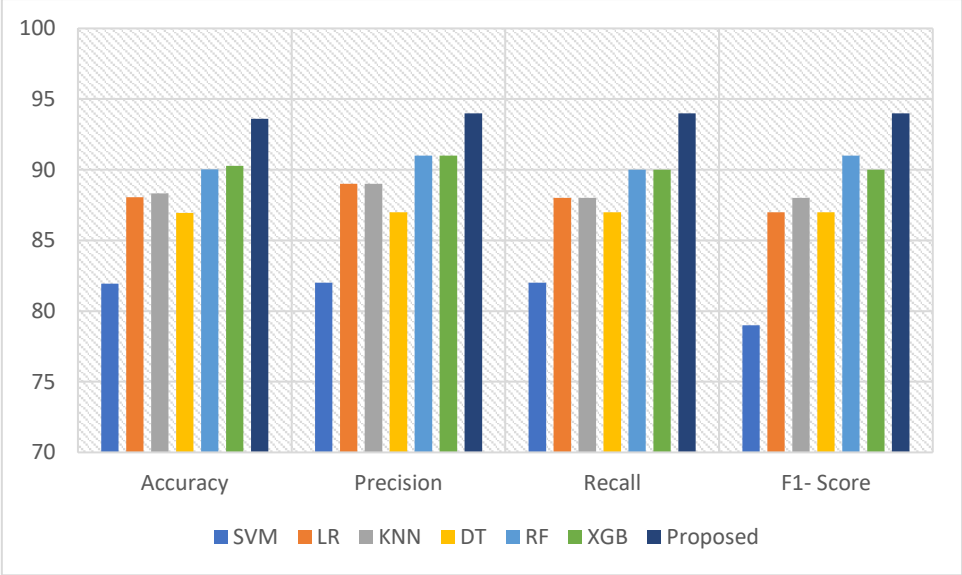

Fig. 5 Evaluation Parameters Comparison

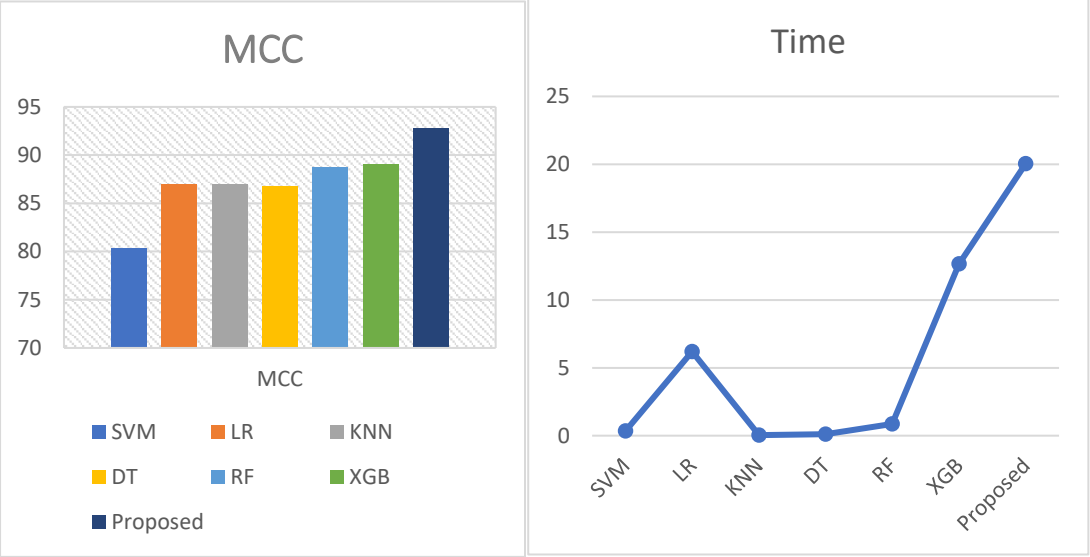

Fig. 6 Mcc And Time Comparison

The summary of the results provides an overview of the various classifiers that are used to identify the devices of the IoT. They are evaluated according to various metrics such as precision, recall, F1-Score, and MCC. The proposed hybrid stacked ensemble classifier has the highest accuracy among the various classifiers. It is able to achieve a 93.61% accuracy rate and outperforms the other options, such as the XGBoost and Random Forest classifiers. The proposed ensemble classifier was able to achieve an impressive 94% accuracy rate in the F1-Score, recall, and precision metrics. This shows that it can perform well in identifying various types of IoT devices.

The proposed ensemble also performed well in the MCC, which considers both false and true positives and negatives. This shows that the group approach is beneficial in handling the imbalanced data. Although the proposed group achieved impressive results, several individual classifiers, such as the K-Nearest Neighbors, LR, and Decision Trees, performed well. These are able to achieve an accuracy rate of around 88%.

The training time for the proposed ensemble classifier was longer compared to that of other classifiers due to the additional step involved in training the meta-classifier. Nevertheless, the

improved accuracy justifies the additional time spent training this model. The results indicate that the Hybrid-stacked ensemble classifier is an ideal tool for identifying the various types of devices that are part of the Internet of Things ecosystem. It can perform well in achieving high precision and recall rates while maintaining the security of the network.

Conclusion and future scope

The goal of this study was to investigate the use of various learning classifiers for identifying devices connected to the IoT. The results indicated that the proposed hybrid-stacked ensemble method performed the best. The proposed ensemble method was able to achieve an accuracy of 93.61%. It exceeded the performance of the individual classifiers, such as XGBoost and Random Forest, and also performed well with respect to recall and F1-Score. The approach's robustness was demonstrated by its superior precision and recall. Various individual classifiers, like LR, Decision Trees, and K-Nearest Neighbour, performed well and provided competitive results with an accuracy of around 88%. The SVM, on the other hand, performed poorly and had an accuracy of only 81.94%. The findings highlight the value of ensembles in improving IoT security and device identification. By employing the multiple classifiers' strengths, the proposed method was able to handle intricate patterns in the collected data and reduce the likelihood of overfitting, leading to superior predictive capabilities. The use of ensembles in IoT device identification is expected to be further improved by optimizing the base classifiers and hyper parameters. In addition, exploring various data augmentation techniques can help improve the model's capacity to identify complex network patterns. In order to ensure that the classifier can adapt to the changing environment and technologies of IoT, it can also be integrated with transfer learning methods and real-time implementations.

References

[1] S. Bhattacharya and M. Pandey, “Issues and Challenges in Incorporating the Internet of Things with the Healthcare Sector,” 2021, pp. 639–651.

[2] A. J. Pinheiro, J. de M. Bezerra, C. A. P. Burgardt, and D. R. Campelo, “Identifying IoT devices and events based on packet length from encrypted traffic,” *Comput. Commun.*, vol. 144, no. May, pp. 8–17, 2019, doi: 10.1016/j.comcom.2019.05.012.

[3] B. A. Desai, D. M. Divakaran, I. Nevat, G. W. Peter, and M. Gurusamy, “A feature-ranking framework for IoT device classification,” *2019 11th Int. Conf. Commun. Syst. Networks, COMSNETS 2019*, vol. 2061, pp. 64–71, 2019, doi: 10.1109/COMSNETS.2019.8711210.

[4] R. R. Chowdhury and P. E. Abas, “A survey on device fingerprinting approach for resource-constraint IoT devices: Comparative study and research challenges,” *Internet of Things (Netherlands)*, vol. 20, no. October, p. 100632, 2022, doi: 10.1016/j.iot.2022.100632.

[5] O. Salman, I. H. Elhajj, A. Chehab, and A. Kayssi, “A machine learning based framework for IoT device identification and abnormal traffic detection,” *Trans. Emerg. Telecommun. Technol.*, vol. 33, no. 3, pp. 1–15, 2022, doi: 10.1002/ett.3743.

[6] V. Khetani, Y. Gandhi, S. Bhattacharya, S. N. Ajani, and S. Limkar, “INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Cross-Domain Analysis of ML and DL : Evaluating their Impact in Diverse Domains,” vol. 11, pp. 253–262, 2023.

[7] I. Cvitić, D. Peraković, M. Periša, and B. Gupta, “Ensemble machine learning approach for classification of IoT devices in smart home,” *Int. J. Mach. Learn. Cybern.*, vol. 12, no. 11, pp. 3179–3202, 2021, doi: 10.1007/s13042-020-01241-0.

[8] W. Li, H. Chu, B. Huang, Y. Huan, L. Zheng, and Z. Zou, “Enabling on-device classification of ECG with compressed learning for health IoT,” *Microelectronics J.*, vol. 115, no. July, p. 105188, 2021, doi: 10.1016/j.mejo.2021.105188.

[9] M. Miettinen, S. Marchal, I. Hafeez, N. Asokan, A. R. Sadeghi, and S. Tarkoma, “IoT SENTINEL: Automated Device-Type Identification for Security Enforcement in IoT,” *Proc. - Int. Conf. Distrib. Comput. Syst.*, pp. 2177–2184, 2017, doi: 10.1109/ICDCS.2017.283.

[10] H. Lee, “Framework and development of fault detection classification using IoT device and cloud environment,” *J. Manuf. Syst.*, vol. 43, pp. 257–270, 2017, doi: 10.1016/j.jmsy.2017.02.007.

[11] Y. Meidan *et al.*, “ProfilIoT: A machine learning approach for IoT device identification based on network traffic analysis,” *Proc. ACM*

Symp. Appl. Comput., vol. Part F1280, pp. 506–509, 2017, doi: 10.1145/3019612.3019878.

[12] G. Vaidya, A. Nambi, T. V. Prabhakar, T. Vasanth Kumar, and S. Sudhakara, "Towards generating a reliable device-specific identifier for IoT devices," *Pervasive Mob. Comput.*, vol. 76, p. 101445, 2021, doi: 10.1016/j.pmcj.2021.101445.

[13] U. Khalil, A. Ahmad, A. H. Abdel-Aty, M. Elhoseny, M. W. A. El-Soud, and F. Zeshan, "Identification of trusted IoT devices for secure delegation," *Comput. Electr. Eng.*, vol. 90, no. January, p. 106988, 2021, doi: 10.1016/j.compeleceng.2021.106988.

[14] B. Chakraborty, D. M. Divakaran, I. Nevat, G. W. Peters, and M. Gurusamy, "Cost-Aware Feature Selection for IoT Device Classification," *IEEE Internet Things J.*, vol. 8, no. 14, pp. 11052–11064, 2021, doi: 10.1109/JIOT.2021.3051480.

[15] L. Fan *et al.*, "EvoIoT: An evolutionary IoT and non-IoT classification model in open environments," *Comput. Networks*, vol. 219, no. 30, p. 109450, 2022, doi: 10.1016/j.comnet.2022.109450.

[16] E. S. Babu *et al.*, "Blockchain-based Intrusion Detection System of IoT urban data with device authentication against DDoS attacks," *Comput. Electr. Eng.*, vol. 103, no. October, p. 108287, 2022, doi: 10.1016/j.compeleceng.2022.108287.

[17] A. Zohourian *et al.*, "IoT Zigbee device security: A comprehensive review," *Internet of Things (Netherlands)*, vol. 22, no. May, p. 100791, 2023, doi: 10.1016/j.iot.2023.100791.

[18] Sumitra and M. V. Shenoy, "HFedDI: A novel privacy preserving horizontal federated learning based scheme for IoT device identification," *J. Netw. Comput. Appl.*, vol. 214, no. March, p. 103616, 2023, doi: 10.1016/j.jnca.2023.103616.

[19] AMI, "IOT device identification | Kaggle." [Online]. Available: https://www.kaggle.com/datasets/fanbyprinciple/iot-device-identification.

[20] P. R. Chandre, P. N. Mahalle, and G. R. Shinde, "Machine Learning Based Novel Approach for Intrusion Detection and Prevention System: A Tool Based Verification," in 2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN), Nov. 2018, pp. 135–140, doi: 10.1109/GCWCN.2018.8668618.

Pise Ganesh Shivaji is a PhD scholar at the Savitribai Phule Pune university pune Department of Computer engineering research center Smt. Kashibai Navale College of engineering Pune. His current research interests are the IOT device classification and data accusation and knowledge management for machine to machine communication.

Email: ganeshpise143@gmail.com

Dr Sachin D. Babar is ISTE Life Member. He is also BOS Member for Computer Engineering, Savitribai Phule Pune University, Pune. He is awarded with the Phd Degree on 25th Feb 2015 from Aalborg University, Denmark in the area of Wireless communication. He is graduated in Computer Engineering from Pune University, Maharashtra, India in 2002 and received Master in Computer Engineering from Pune University, Maharashtra, India in 2006.He is currently working as Professor and Head in the department of Computer Engineering, SIT Lonavala, India. He has 21+ years of teaching experience. He has published more than 70 papers at national and international level. He has authored two books on subjects like Software Engineering and Analysis of Algorithm & Design. He has received the Cambridge International Certificate for Teachers and Trainers at Professional level under MISSION10X Program. He is IBM DB2 certified professional. His research interests are Data Structures, Algorithms, Theory of Computer Science, IoT and Security.

Email: hodce.sit@sinhgad.edu

Dr Parikshit Mahalle is a senior member IEEE and is Professor, Dean Research and Development and Head - Department of Artificial Intelligence and Data Science at Vishwakarma Institute of Information Technology, Pune, India. He completed his Ph. D from Aalborg University, Denmark and continued as Post Doc Researcher at CMI, Copenhagen, Denmark. He has 23 + years of teaching and research experience. He is an ex-member of the Board of Studies in Computer Engineering, Ex-Chairman Information Technology, Savitribai Phule Pune University and various Universities and autonomous colleges across India. He has 15 patents, 200+ research publications (Google Scholar citations-2950 plus, H index-25 and Scopus Citations are 1550 plus with H index -18, Web of Science citations are 438 with H index - 10) and authored/edited 54 books with Springer, CRC Press, Cambridge University Press, etc. He is editor in chief for IGI Global – International Journal of Rough Sets and Data Analysis, Inter-science International Journal of Grid and Utility Computing, member-Editorial Review Board for IGI Global – International Journal of Ambient Computing and Intelligence and reviewer for various journals and conferences of the repute. His research interests are Machine Learning, Data Science, Algorithms, Internet of Things, Identity Management and Security. He is guiding 8 PhD students in the area of IoT and machine learning and SIX students have successfully defended their PhD under his supervision from SPPU. He is also the recipient of “Best Faculty Award” by Sinhgad Institutes and Cognizant Technologies Solutions. He has delivered 200 plus lectures at national and international level.

Email: aalborg.pnm@gmail.com