scientific reports

OPEN

Prediction of DDoS attacks in agriculture 4.0 with the help of prairie dog optimization algorithm with IDSNet

Ramesh Vatambeti[1]✉, D. Venkatesh[2], Gowtham Mamidisetti[3], Vijay Kumar Damera[4], M. Manohar[5] & N. Sudhakar Yadav[6]

Integrating cutting-edge technology with conventional farming practices has been dubbed "smart agriculture" or "the agricultural internet of things." Agriculture 4.0, made possible by the merging of Industry 4.0 and Intelligent Agriculture, is the next generation after industrial farming. Agriculture 4.0 introduces several additional risks, but thousands of IoT devices are left vulnerable after deployment. Security investigators are working in this area to ensure the safety of the agricultural apparatus, which may launch several DDoS attacks to render a service inaccessible and then insert bogus data to convince us that the agricultural apparatus is secure when, in fact, it has been stolen. In this paper, we provide an IDS for DDoS attacks that is built on one-dimensional convolutional neural networks (IDSNet). We employed prairie dog optimization (PDO) to fine-tune the IDSNet training settings. The proposed model's efficiency is compared to those already in use using two newly published real-world traffic datasets, CIC-DDoS attacks.

Present-day farming is rapidly developing into a new period recognised as "agriculture 4.0". Agriculture 4.0 seeks to use new technology and approaches to address the problems plaguing current agriculture (such as climate change, illnesses, the overuse of chemicals and resources, etc.). This will hopefully improve efficiency and minimise risks. In order to achieve this goal, it makes use of a wide variety of cutting-edge forms of ICT[1]. In addition to these changes, the demand for food is on the rise; the UN's Food and Agriculture Organization estimates that demand will increase by 70% by 2050 compared to current production levels in order to meet the requirements of a global population of around 10 billion by that year[2,3]. Agriculture 4.0 is predicted to see massive market growth over the next several years as a result of continued technological advancements and the rising global need for food.

Solutions are widely used in Agriculture 4.0 because of the many advantages they offer to farmers (e.g., improved monitoring of environmental parameters related to crops, earlier detection of crop diseases, more accurate estimates of predicted yield, less time spent on manual labour)[4,5]. But the interconnectedness of diverse sensors and network devices allowed for numerous attacks[7]. This is because such devices frequently contain unpatched or outdated firmware or software[6]. Malware refers to any instance of a network attack[8,9].

Any form of disruption or distortion may offer significant obstacles and lead to severe repercussions in agriculture[10,11]. Monitoring and classifying network data has been a hot topic since the early days of the Internet because of its potential to thwart assaults[12]. Classifying network traffic to protect Internet of Things systems has been the topic of much research. Essential to Intrusion Detection Systems (IDS), it aids in the tracking down and elimination of potentially harmful network activities[13]. An IDS is a network monitoring device intended to identify suspicious or anomalous activity and allow preventative action against potential incursion threats. Consequently, there are two primary categories of intrusion detection systems: (1) NIDS and (2) HIDS. While HIDS systems may be used on any networked device with an Internet connection, NIDS are often implemented

[1]School of Computer Science and Engineering, VIT-AP University, Vijayawada, India. [2]Department of Computer Science and Engineering, GITAM School of Technology, GITAM University-Bengaluru Campus, Bengaluru, India. [3]Department of Computer Science and Engineering, Malla Reddy University, Hyderabad, India. [4]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad, India. [5]Department of Computer Science and Engineering, CHRIST (Deemed to be University), Bangalore, India. [6]Department of Information Technology, Chaitanya Bharathi Institute of Technology, Hyderabad 500075, India. ✉email: v2ramesh634@gmail.com

or situated at crucial network points to ensure they cover the sites where the traffic is most vulnerable to attack. IDS's most-used approaches for detecting intrusions[14] Signature-based intrusion detection systems (IDS), also known as misuse detection or knowledge-based detection, are as effective as real-time database updates because they focus on recognising the "signature," or unique pattern, of intrusion events. The anomaly-based IDS (also called behaviour-based detection) relies on frequent activity monitoring and machine learning methods to compare known, safe patterns of behaviour to any suspicious ones that may have emerged. Intrusion Prevention Systems (IPS) are used to thwart threats like Trojan horses, distributed denial of service attacks, and more once an administrator gets a warning from an IDS system[15].

In this work, we explore how to utilise deep learning to identify cyber risks (i.e., anomaly-based IDS). Recently suggested IDS arrangements use deep learning algorithms for IoT networks[10], large data environments[16], cyber-physical systems[17], SCADA systems[18], smart grids, internet-connected vehicles (IoVs), and cloud computing. Hail damage to crops, soil, etc. are all examples of areas where deep learning algorithms are employed in Agriculture 4.0. However, in the area of intrusion detection schemes for agriculture, there are eight major hurdles to overcome: One, collecting data on IIoT traffic cyberattacks; two, insufficient training data; three, training data that is not representative of the real world; four, poor data quality; five, irrelevant or unwanted features; six, seven, underfitting the training data; and eight, learning and deploying the model offline[18]. This difficulty is solved by the suggested model. Our article draws on widely-used, up-to-date datasets that are widely-utilized in the research community for the purpose of creating intrusion detection algorithms for IIoT networks. Limitations include IDSNet may have challenges in understanding the unique characteristics of Agriculture 4.0 environments, which involve diverse sensors, actuators, and communication protocols. Customizing IDSNet's detection rules and algorithms to consider the specific features and communication patterns of Agriculture 4.0 systems can help improve its accuracy.

The most important contributions to this study are:

- Provide a presentation, assessment, and proportional analysis of techniques for cyber security;
- propose a deep learning-based system for intrusion detection in agriculture 4.0; therefore, the proposal is called IDSNet-PDO.
- Each suggested deep learning model's presentation is evaluated across two classification types using data from two recently released real-world traffic datasets (the dataset and the TON IoT dataset). Important performance metrics were the focus of the research.

The remaining sections of this paper are organised as follows: In the second part, we'll examine some secondary sources. The use of IDS is described in "Proposed system" section. An in-depth look into IDS in Agriculture 4.0 is provided in "Results and discussion" section. In "Conclusions" section concludes with some last thoughts.

Related works

Three different types of deep learning-based IDS replicas have been developed by Ferrag et al.[19] They are based on deep networks. In this work, we compared and contrasted the efficacy of strategies for agribusiness 4.0 cyber security. The dataset and the TON IoT dataset, both of which include real-world traffic data, are used to analyse the presentation of each model across two categorization types (binary and multiclass). Key performance criteria favour deep learning techniques over conventional machine learning strategies. Furthermore, the CNN-based IDS model outperforms the IDS approaches as measured by their performance on the dataset with multiclass traffic finding, respectively.

An intrusion detection arrangement based on federated learning has been projected by Friha et al.[20] to protect agricultural IoT infrastructures. They call it FELIDS. In particular, the FELIDS system protects information by relying on local learning, which is when devices learn from each other by exchanging only model updates with an aggregate server. This makes the detection model more accurate. The FELIDS system uses deep learning classifiers to protect against attacks on agricultural IoTs. The proposed IDS is evaluated on the CSE-CIC-IDS2018 benchmark, the MQTTset, and the InSDN. It is clear from the findings that the FELIDS organisation is superior to traditional, non-federated types of machine learning in terms of both accuracy and effectiveness in safeguarding the privacy of data collected from IoT devices.

The review and analysis of intrusion have been completed by Ferrag et al.[21] In this paper, we detail the cyber security challenges facing Agriculture 4.0 and the criteria used to assess the effectiveness of intrusion detection systems. Next, we conduct an analysis of intrusion detection systems in light of current and forthcoming technological developments, such as the Internet of Things, autonomous tractors, drones, smart grids, and industrial agriculture. Based on the machine learning approach used, we present a detailed categorization of intrusion detection schemes in each developing knowledge area. We also showcase accessible tools used to assess the effectiveness of intrusion detection arrangements. Finally, we provide an overview of the difficulties and potential future research areas in intrusion detection for cyber safety in Agriculture 4.0.

IoT networks used in agriculture have been the target of invasions; however, a system for identifying and categorising these attacks has been established by Raghuvanshi et al.[22] All applications of the Internet of Things have the same fundamental problem: how to ensure the safety and privacy of their users. The NSL KDD data set is used as an example input in this framework. First, the NSL-KDD data set has all of its symbolic characteristics translated into numerical features as part of its pre-processing. Principal is used to extract features. To further categorise the gathered information, we apply machine learning methods, and precision and recall metrics are used to compare the effectiveness of various machine learning algorithms.

Through an examination of potential assaults and threats, Vangala et al.[23] want to learn more about the security scenarios that may be used in agriculture. Research on existing IoT testbeds in the agricultural sector has

been conducted. An architecture for smart farming is presented, as well as technologies that could be used in conjunction with the proposed architecture. The direction of advancement in each agricultural security sub-area is discussed, and the lack of current protocols is identified via a literature analysis of safety protocols for different sub-sectors of security in smart agriculture and verification protocols in smart requests. In addition, the state of the art of industry-based IoT-based tools and systems has been investigated.

Otoum et al.[24] present a novel (DLIDS) method for identifying potential security issues in IoT settings. While there is no shortage of IDSs described in the academic literature, many of them suffer from insufficient attack detection accuracy due to problems with feature learning and dataset management. In order to improve detection accuracy, we propose a module that uses a hybrid of the Spider Monkey Optimization algorithm (SMO) and the SDPN. The SMO algorithm is responsible for selecting the most informative features in the datasets, while the SDPN determines whether the data is typical or out of the ordinary. DoS U2R attacks, probing attacks, and remote-to-local (R2L) attacks are all recognised by DL-IDS. A battery of extensive experiments has shown that the suggested DL-IDS outperforms state-of-the-art tactics.

Sengan et al.[25] hope to provide a solution for healthcare data by using dynamic, secure, and aware routing through machine learning (DARML). In this work, we offer a DoS detection scheme that uses an ML algorithm. To see the permitted procedure, one must first get access to the user. Users may then register and utilise correlation factors between nodes to compare route information. The user then selects the gadget that will initiate the data key's automated activation and decryption. In the final module, the DAR-ML is linked to all healthcare records. Next, both users and the administrator will be able to provide feedback on the findings. Those are the benefits you get from simplifying everything using the internet. Based on the analysis of 21.19 percent of all data flow, the results show an attack finding accuracy of over 98.19 percent, along with an excellently low false alarm likelihood.

Lin et al.[26] suggest making adversarial harmful traffic records with a generative adversarial network architecture they call IDSGAN to fool and avoid being caught by intrusion detection systems. The adversarial attack examples carry out black-box assaults since the attackers do not know the fundamental structure and settings of the detection scheme. IDSGAN uses a generator to convert legitimate traffic records into adversarial data. A discriminator that also classes traffic instances learns the real-time black-box detection method. Moreover, the adversarial generation makes use of a controlled modification technique that was developed to protect the authentic attack capabilities of adversarial traffic records. Multiple algorithm-based detection models are subjected to various assault types to demonstrate the model's efficacy. By varying the sample size of the modifications, robustness can be tested. Through a controlled experiment utilising adversarial attack baselines, we are able to prove that our model is better.

Proposed system

Here, we take a look at the IDSNet model, which was developed to identify cyber-attacks in Agriculture 4.0 and makes use of a one-dimensional convolutional neural network and the PDO.

Network model. The agriculture 4.0 network model is provided, which is composed of the following three layers: (1) agricultural sensors; (2) fog computing; and (3) cloud computing. The agriculture industry uses data gathered by drones and other Internet of Things sensors. When certain thresholds are met in the data collected by the agricultural sensor layer, the actuators below are triggered. To ensure that Internet of Things (IoT) devices always have access to power, new energy technologies and smart grid design are implemented in the sensor layer. Every fog node has an embedded deep learning intrusion detection system. To perform analysis and machine learning algorithms, the IoT data is sent from the agricultural sensors layer to the fog computing layer, while cloud computing nodes offer storage and end-to-end services. Typically, intrusion detection systems that rely on deep learning to process alerts send their processing to fog nodes. We assume that there is a malicious party intent on disrupting the network's operations in order to compromise food security, the effectiveness of the agri-food supply chain, and output.

Pre-processing of the Cic-Ddos2019 dataset. There are a total of 50,063,112 records in the CIC-DDoS2019 dataset[29], consisting of 50,06,249 rows related to DDoS assaults and 56,863 rows related to normal traffic. with 86 characteristics in each row. Table 1 presents a summary of the dataset's attack statistics throughout both training and testing. SNMP and SSDP are used in the attacks.

- In a reflection-based DDoS assault known as an "NTP-based attack," an adversary hijacks a server running the Network Time Protocol (NTP) protocol to send an overwhelming amount of traffic across the User Datagram Protocol (UDP) to a single target. The target and its supporting network infrastructure may become inaccessible to legitimate traffic as a result of this attack.
- An attack that leverages the Domain Name System (DNS) to flood a target IP address with resolution requests is called a reflection-based DDoS assault.
- By sending queries to a publicly accessible vulnerable LDAP server, an attacker can generate massive (amplified) responses, which are then reflected to a target server, resulting in a DDoS attack.
- Reflection-based (DDoS) attacks, or "MSSQL-based attacks," include the attacker forging an IP address to make programmed requests seem to originate from the targeted server while really exploiting.
- NetBIOS-based attacks are a kind of reflection-based denial-of-service attack in which the attacker delivers forged "Name Release" or "Name Conflict" signals to the target system, causing it to reject any and all incoming NetBIOS packets.

Attack kind	Flow count
Benign	56,864
DDoS_SYNs	1,582,279
DDoS_TFTPs	20,082,581
DDoS_NetBIOSs	4,093,270
DDoS_NTPs	1,202,643
DDoS_SSDPs	2,610,612
DDoS_LDAPs	2,179,931
DDoS_MSSQLs	4,522,495
DDoS_UDPs	3,134,645
DDoS_UDP-Lags	366,471
DDoS_WebDDoSs	438

Table 1. Kinds of attacks in the CICDDoS dataset.

- To jam the target's network pipes, an SNMP-based assault will produce attack volumes in the hundreds of gigabits per second using the Simple Network Management Protocol (SNMP).
- The reflection-based SSDP attack is a DDoS attack in which the attacker uses UPnP protocols to deliver a flood of traffic to the intended victim.
- This kind of attack uses IP packets carrying UDP datagrams to deliberately saturate the network connection of the victim host and cause it to crash.
- To compromise a Web server or application, a WebDDoS-based attack will use seemingly innocuous HTTP GET or POST requests as a backdoor.
- Syn-based attacks use the standard TCP three-way handshake and respond with an ACK to exhaust the victim server's network resources and render it unusable.
- As its name suggests, an attack based on the TFTP protocol uses online TFTP servers to get access to sensitive information. An attacker makes a default request for a file, and the victim TFTP server delivers the information to the attacker's target host.
- An example of this is the PortScan-based attack, which is similar to a network security audit in that it scans the open ports of a target computer or the whole network. Scanning is performed by sending queries to a distant site in an effort to learn what services are available there.

We generate three datasets, respectively titled "Dataset 13 class," to examine the efficacy of learning approaches in binary and multi-class classification. Tables 2 and 3 describe the statistics for each dataset regarding attacks during training and testing, respectively. Table 4 describes the attack categories in Dataset 7 class.

Pre-processing of the Ton_IoT dataset. A novel testbed for an IIoT network, the TON IoT dataset[30] includes information on the network, the operating system, and telemetry. Seven files containing telemetry data from Internet of Things and industrial Internet of Things sensors are given in Table 5. Here's what you may expect to find within these files:

- File 1: "Train Test IoT Weather" includes the following conditions: Normal (35,000 rows), DDoS (5000 rows), injection (50,000 rows), Password (50,000 rows), and backdoor IoT data from a networked weather sensor, including temperature, pressure, and humidity values, are shown in the file.
- There are Normal (35,000 rows), DDoS, and Injection (2902 rows) in File 2 "Train Test IoT Fridge" (2942 rows). The file contains information on the sensor's temperature readings and environmental circumstances as they pertain to the Internet of Things.
- Train Test IoT Garage Door.txt has the following categories: normal (10,000 rows), ransomware (5804 rows). If you have a networked door sensor, this file will show you whether or not the door is open or closed.
- File 4 "Train Test IoT GPS Tracker" has the following categories and numbers of rows: Normal (35,000), DDoS (5,000), Injection (5,000), Password (5,000), Backdoor (5,000), Ransomware (2,833 rows), XSS (577 rows), and Scanning (550 rows). Data from a networked GPS tracker sensor is shown in the file, including its latitude and longitude readings, as an example of Internet of Things (IoT) data.

Class	Test	Training
Benign	17,147	56,102
Attack	314,717	997,055

Table 2. Attack categories in Dataset_2_class.

Category	Flow count	Category of attack	Training/test
BENIGN	56,101	BENIGN	Split the data test (x_train, x_test, stratify = y)
Reflection-based attacks	99,943	DrDoS_LDAP	
	98,576	DrDoS_SSDP	
	96,567	DrDoS_DNS	
	95,700	DrDoS_MSSQL	
	93,560	DrDoS_NetBIOS	
	91,578	DrDoS_SNMP	
	76,457	DrDoS_NTP	
	72,116	TFTP	
	439	WebDDoS	
Exploitation-based attacks	97,932	DrDoS_UDP	
	99,983	Syn	
	74,203	UDP-lag	

Table 3. Attack categories in Dataset_13_class.

Category	Category of attack	Test	Training
Reflection-based attacks	DrDoS_NetBIOS	136,729	619,700
	DrDoS_MSSQL	157,076	619,446
	DrDoS_LDAP	150,701	619,251
Exploitation-based attacks	DrDoS_UDP	150,706	618,696
	UDP-lag	1873	183,662
	Syn	150,416	790,662
Exploitation/reflection -based attacks	Others DoS attacks	28,12	938,733
Benign	Benign	17,146	56,101

Table 4. Attack categories in Dataset_7_class.

- You'll find the following data types in File 5: "Train Test IoT Modbus: Normal (35,000 rows), Injection (5,000 rows), Password (5,000 rows), Backdoor. IoT data file containing Modbus function code for reading an input register.
- There are 70,000 rows of normal data, 10,000 rows of DDoS data, 10,000 rows of injection data, 10,000 rows of password data, 10,000 rows of backdoor data, 4528 rows of ransomware data, 898 rows of XSS data, and 70,000 rows of scanning data in File 6 "Train Test IoT Motion Light" (3550 rows). In the file, we can see the Internet of Things data for a switch that may either be on or off.
- Included in File 7 "Train Test IoT Thermostat" are the following categories of data: Normal (35,000 rows), Injection (5,000 rows), Password (5,000 rows), Backdoor The file contains data from the Internet of Things that represents the temperature as it is right now according to a networked thermostat sensor.

IDSNet: design and configuration. The current concept took some cues from CNN's practical uses. However, this model just needs a single raw input, and its reduced number of layers helps save time during training.

The current concept takes some cues from CNN's practical uses. However, this model only needs a single raw input, and its reduced number of layers helps save time during training. Figure 1 depicts the design process as it was carried out. The first step was to fine-tune the training and optimization methods as well as the layer count, filter size, and filter amount. It was also necessary to tweak the network's hyper settings. These included the training lot size, learning rate, number of training cycles (epochs), and number of training signals (batch size). Table 6 provides the suggested values. And second, a CNN structure was built, and it's laid out in Table 6. The number of layers in the model network determines the number and size of filters available in each convolutional layer. In this situation, the network layout shown by the bold fonts in the table below performed the best after being optimised by altering a few stated choices in the literature. Figure 1 depicts the filter setup and internal structure of the kernel.

The network employs algorithms to discover and prioritise the most relevant aspects of raw data for mining purposes. To achieve this goal, we apply a convolution process (convolutive layer) to the input data, resulting in a longer vector from which we use a maximum clustering criterion (max-pooling layer) to extract the most representative features. Table 6 shows that the same steps are performed four times with a different number of kernels added to each Convolutive plus Max-Pooling set. This adjustment is made so that feature maps may be

TON_IoT dataset	Attack category	Flow count
Train_Test_IoTs_Weathers	Normal s	35,000
	DDoSs	5000
	Injection s	5000
	Password s	5000
	Backdoor s	5000
	Ransomware s	2865
	XSS s	866
	Scanning s	529
Train_Test_IoT_Fridge	Normal s	35,000
	DDoSs	5000
	Injections	5000
	Password s	5000
	Backdoor s	5000
	Ransomware s	5000
	XSS s	2942
Train_Test_IoT_Garage_Door	Normal s	70,000
	DDoSs	10,000
	Injection s	10,000
	Password s	10,000
	Backdoor s	100,000
	Ransomware	5804
	XSS s	2312
	Scanning s	1058
Train_Test_IoT_ Trackers	Normal s	35,000
	DDoSs	5000
	Injection s	5000
	Password s	5000
	Backdoor s	5000
	Ransomware	2833
	XSS s	577
	Scanning s	550
Train_Test_IoT_Modbus	Normal s	35,000
	Injection s	5000
	Password s	5000
	Backdoor s	5000
	XSS s	577
	Scanning s	529
Train_Test_IoT_Motion_Light	Normal s	70,000
	DDoSs	10,000
	Injection	10,000
	Password s	10,000
	Backdoor s	10,000
	Ransomware s	4528
	XSS s	898
	Scanning s	3550
Train_Test_IoT_Thermostat	Normal s	35,000
	Injection s	5000
	Passwords	5000
	Backdoor s	5000
	Ransomware s	2264
	XSS s	449
	Scanning s	61

Table 5. Attack categories in TON_IoT dataset.

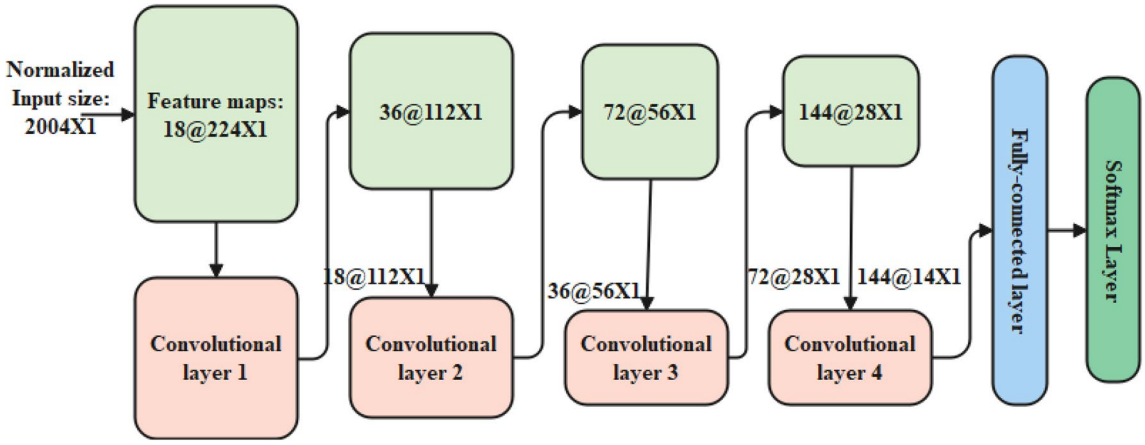

Figure 1. Internal structure of IDSNet.

Type	Stride	Filters size	Filters size	Padding
Convolutive	1	2	2	'Same'
Max-pooling	2	2	2	'Same'
Convolutive 1	1	64	64	'Same'
Max-pooling	2	2	2	'Same'
Convolutive	1	2	2	'Same'
Max-pooling	2	2	2	'Same'
Convolutive	1	2	2	'Same'
Max-pooling	2	2	2	'Same'

Table 6. Structure of IDSNetwork.

generated that accurately depict the signals' non-linearity. Using a filter with a duration of three samples and a sliding pass of one sample, the first three values of a feature map are generated in sequence. The procedure is performed on each convolutional layer. It is possible to fine-tune this procedure by adjusting the number and size of filters (u), as well as the window's sliding factor (stride). Since the output vector of the final convolutional layer is the input vector of the fully connected layer, only its map length needs to be calculated during network design. The PDO method is used to fine-tune the IDSNet's hyper-parameters like momentum, learning rate, and epochs, as shown below.

Prairie dog optimization. The following were assumed to facilitate the development of models for the proposed PDO:

Each prairie dog belongs to one of the m coteries in the colony, and there are n prairie dogs in each coterie. (i) Prairie dogs are all the same and can be classified into m subgroups, (ii) Each coterie has its own ward inside the colony, which represents the search area for the corresponding issue.

Nesting activities generate an increase from ten burrow openings per ward to as many as one hundred. Both an antipredator call and a new food supply (burrow construction) call are used. It's only individuals of the same coterie that engage in foraging and burrow construction activities (exploration), communication, and anti-predator (exploitation) actions. Exploration and exploitation are repeated m (the number of coteries) times since other coteries in the colony undertake the same tasks at the same time and the whole colony or problem space has been partitioned into wards (coteries).

Like other population-based algorithms, the prairie dog optimization (PDO) relies on a random initialization of the placement of the prairie dogs. The search agents are the prairie dog populations themselves, and each prairie dog's position is represented by a vector in d-dimensional space.

Initialization. Each prairie dog (PD) is a member of one of m coteries, where n is the total number of PDs. Because prairie dogs live and work together in groups called "coteries," each prairie dog's position within a given coterie may be uniquely determined by a vector. Positions of all coteries (CT) in a colony are shown by the matrix in Eq. (1):

$$CT = \begin{bmatrix} CT_{1,1} & CT_{1,2} & \cdots & CT_{1,d-1} & CT_{1,d} \\ CT_{2,1} & CT_{2,2} & \cdots & CT_{2,d-1} & CT_{2,d} \\ \vdots & \vdots & CT_{i,j} & \vdots & \vdots \\ CT_{m,1} & CT_{m,2} & \cdots & CT_{m,d-1} & CT_{m,d} \end{bmatrix} \tag{1}$$

When talking about a colony, the jth dimension of the ith coterie is denoted as CT (i,j). All of the prairie dogs in a coterie may be found at the coordinates given by Eq. (2):

$$PD = \begin{bmatrix} PD_{1,1} & PD_{1,2} & \cdots & PD_{1,d-1} & PD_{1,d} \\ PD_{2,1} & PD_{2,2} & \cdots & PD_{2,d-1} & PD_{2,d} \\ \vdots & \vdots & PD_{i,j} & \vdots & \vdots \\ PD_{n,1} & PD_{n,2} & \cdots & PD_{n,d-1} & PD_{n,d} \end{bmatrix} \tag{2}$$

where $PD(i,j)$ stands for the jth dimension of the ith prairie dog in a pack and nm is the total number of dogs in the pack. Equations 3 and 4 depict the uniform distribution used to assign each prairie dog to its coterie.

$$CT_{i,j} = U(0,1) \times \left(UB_j - LB_j\right) + LB_j \tag{3}$$

$$PD_{i,j} = U(0,1) \times \left(ub_j - lb_j\right) + lb_j \tag{4}$$

where UB_j and LB_j of the optimization problem, $ub_j = \frac{UB_j}{m}$ and $lb_j = \frac{LB_j}{m}$, and U(0,1) is a random sum with a uniform distribution among 0 and 1.

Fitness function evaluation. By plugging the solution vector into the predefined fitness function, we can get the fitness value for each prairie dog site. To keep track of the results, we may use the array defined by Eq. (5).

$$PD = \begin{bmatrix} f_1([PD_{1,1} & PD_{1,2} & \cdots & PD_{1,d-1} & PD_{1,d}]) \\ f_2([PD_{2,1} & PD_{2,2} & \cdots & PD_{2,d-1} & PD_{2,d}]) \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ f_1([PD_{n,1} & PD_{n,2} & \cdots & PD_{n,d-1} & PD_{n,d}]) \end{bmatrix} \tag{5}$$

An individual prairie dog's fitness function value is a measure of the quality of food available at a given location, the likelihood of successfully excavating and populating new burrows, and the efficacy of its anti-predator alarm system. The fitness function values array is sorted, and the element with the lowest value is designated the optimal solution to the minimization issue. In addition to the following three, the greatest value is taken into account while designing burrows that help animals hide from predators.

Exploration. The PDO has four parameters it uses to determine when to switch between exploration and exploitation. The total number of possible cycles is cut in half, with the first half going toward exploration and the second half toward exploitation. There is a causal relationship between the two investigation tactics. on $iter < \frac{max_{iter}}{4}$ and $iter \leq \frac{max_{iter}}{4} < iter < \frac{max_{iter}}{2}$, while the two strategies for exploitation are conditioned on $\frac{max_{iter}}{2} \leq iter < 2\frac{max_{iter}}{4} \leq iter \leq max_{iter}$.

Equation (6) describes how our algorithm updates its location throughout the foraging phase of its exploration phase. The second plan of action is to analyse the digging strength and the quality of the found food sources thus far. The digging power used to create new burrows is calibrated to decrease with time. This limitation aids in controlling the burrowing population. Position updates during tunnel construction are described by Eq. (7).

$$PD_{i+1,j+1} = GBest_{i,j} - eCBest_{i,j} \times \rho - CPD_{i,j} \times Levy(n) \forall iter < \frac{max_{iter}}{4} \tag{6}$$

$$PD_{i+1,j+1} = GBest_{i,j} \times rPD \times DS \times Levy(n) \forall iter < \frac{max_{iter}}{4} \leq iter < \frac{max_{iter}}{2} \tag{7}$$

As demonstrated in Eq. (8), where $GBest_{i,j}$ is the best global solution so far achieved, $eCBest_{i,j}$ assesses the impact of the currently obtained best answer. In this experiment, q is the frequency of the specialised food source alert, which has been set at 0.1 kHz; rPD is the location of a random solution; and $CPD_{i,j}$ is defined as the random cumulative impact of all prairie dogs in the colony. The digging strength of the coterie, denoted by DS, varies with the quality of the food supply and is determined at random by Eq. (10). The Levy(n) distribution is recognised to promote more effective and thorough investigation of the search space of a topic.

$$eCBest_{i,j} = GBest_{i,j} \times \Delta + \frac{PD_{i,j} \times mean\left(PD_{n,m}\right)}{GBest_{i,j} \times \left(UB_j - LB_j\right) + \Delta} \tag{8}$$

$$CPD_{i,j} = \frac{GBest_{i,j} - rPD_{i,j}}{GBest_{i,j} + \Delta} \tag{9}$$

$$DS = 1.5 \times r \times \left(1 - \frac{iter}{max_{iter}}\right)^{\left(2\frac{iter}{max_{iter}}\right)} \tag{10}$$

where r adds the stochastic property to guarantee exploration by taking either − 1 or + 1 as its value depending on whether the current iteration is odd or even, Despite the fact that the prairie dogs are considered to be identical in the PDO implementation, the small number represented by helps explain for these variances.

Exploitation. The point of PDO's exploitation mechanisms is to conduct extensive searches in the promising regions discovered during the exploration phase. Equations (11) and (12) model the two approaches used during this stage. Earlier, we discussed how the PDO toggles between these two tactics. to $\frac{max_{iter}}{2} \leq iter < 2\frac{max_{iter}}{4}$ and $3\frac{max_{iter}}{4} \leq iter \leq max_{iter}$, respectively.

$$PD_{i+1,j+1} = GBest_{i,j} - eCBest_{i,j} \times \varepsilon - CPD_{i,j} \times rand \forall \frac{max_{iter}}{2} < iter < 3\frac{max_{iter}}{4} \tag{11}$$

$$PD_{i+1,j+1} = GBest_{i,j} \times PE \times rand \forall 3\frac{max_{iter}}{4} < iter < max_{iter} \tag{12}$$

As demonstrated in Eq. (8), where GBest (i,j) is the best global solution so far achieved, eCBest (i,j) assesses the impact of the currently obtained best answer. Equation (8) defines CPD (i,j) as the aggregate influence of all prairie dogs in the colony, where is a tiny integer representing the quality of the food supply. In Eq. (13), PE stands for the predator effect, and rand is a random integer between zero and one..

$$PE = 1.5 \times \left(1 - \frac{iter}{Max_{iter}}\right)^{\left(2\frac{iter}{max_{iter}}\right)} \tag{13}$$

where *iter* is the current iteration and Max_{iter} is the supreme sum of iterations.

Results and discussion

Performance evaluation. Agriculture 4.0 entails incorporating cutting-edge technology into standard farming practises to raise output and quality standards. Internet-of-Things gadgets, are all examples of such cutting-edge technology. We used and chose current data sets based on these technologies that include DDoS employed by Here, we focused on two recently released real-world traffic dataset[29] and the TON IoT dataset[30]. The TCP/IP communication stack compatibility, DDoS attack mitigation, and symbolic representation of Agriculture 4.0 all played roles in their selection. The TON IoT dataset was developed to mimic the functioning of actual operational IoT/IIoT networks via the use of interacting network parts and IoT/IIoT systems across the Edge, Fog, and Cloud. SDN and NFV technologies, such as those provided by the NSX-VMware platform, were used to better control the interplay between the three levels. The experiment is coded in Python 3 on a GPU using TensorFlow. The suggested model's hyper-parameters are summarised in Table 7.

Performance metrics. Important consideration should be given to the metrics used to assess the effectiveness of machine learning and deep learning approaches. Our analysis centres on the following key performance metrics: In Table 8, we see examples of four potential classifications, two of which are incorrect.

Hyper parameter	Value
Activation function s	Sigmoids
Classification function s	Soft-max
Batch size s	10.000
Hidden nodes (HN)	20–100
Sum of epoch s	100
Learning rate (LR)	0.01–0.5

Table 7. The hyper-parameters working in deep learning tactics.

		Predicted class	
		Negative-class	Positive-class
Class	Negative-class	(TN)	(FP)
	Positive-class	(FN)	(TP)

Table 8. Confusion matrix.

$$TNR_{BENIGN} = \frac{TN_BENIGN}{TN_BENIGN + FP_BENIGN} \tag{14}$$

$$FAR = \frac{FP_BENIGN}{TN_BENIGN + FP_BENIGN} \tag{15}$$

$$Precision = \frac{TP_Attack}{TP_Attack * FP_BENIGN} \tag{16}$$

$$Recall = \frac{TP_Attack}{TP_Attack * FN_Attack} \tag{17}$$

$$DR_{Attack} = \frac{TP_Attack}{TP_Attack + FN_Attack} \tag{18}$$

$$F - score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \tag{19}$$

$$Accuracy = \frac{TP_Attack + TN_BENIGN}{TP_Attack + FN_Attack + TN_BENIGN + FP_BENIGN} \tag{20}$$

$$DR_{Overall} = \frac{\sum TP_Each - Attack - Type}{\sum TP_Each - Attack - Type + \sum FN_Each - Attack - Type} \tag{21}$$

where innocuous data that is accurately identified as benign whereas False Positive (FP) suggests benign data that is wrongly identified as an attack. True Positive (TP) is information on an assault that has been appropriately identified as such. Attack data that is wrongly categorised as non-threatening is called a False Negative (FN).

The CICDDoS2019 dataset of seven classes is tested with different generic models and proposed models, which are shown in Table 9. The existing models are tested with other different datasets; therefore, generic models are considered for comparison. The results are averaged and provided in Table 9.

In Table 9, the various attack types are taken into account for comparative analysis of accuracy among RNN, LSTM, and the proposed model. TNR (BENIGN) gives detection accuracy of 95 in RNN and 98% in LSTM, and the proposed model achieves 99% accuracy. In the DrDos_LDAP attack, RNN achieves 96% accuracy, 95% accuracy in LSTM, and 97% accuracy in the proposed system. The accuracy of other attacks like DrDoS_MSSQL, DrDoS_NetBIOS, and DrDoS_UDP shows the results of RNN as 96%, 69%, and 60%, while LSTM achieves 94%, 95%, and 71%, and the proposed attack gives better accuracy of 95%, 96%, and 75%. Syns achieves 100% accuracy on RNN LSTM and the proposed IDSNet-PDO. The proposed model has a higher detection ratio. Multi-class analysis on the second dataset is presented in Table 10.

In Table 10, the various attack types are taken into account for a comparative analysis of accuracy among RNN and LSTM with the proposed model. Normal gives detection accuracy of 93% in RNN, 94% in LSTM, and the proposed model achieves 96% of accuracy, whereas in DDos attacks, RNN achieves 94%, 95% in LSTM, and 98% in the proposed system. The IDSNet-PDO model gives a ratio for all attack categories: backdoor, ransomware, and XSS. For the different 13 classes of the first dataset, the results are provided in Table 11.

In TNR (BENIGN), RNN, LSTM, and the proposed IDSNet-PDO achieve 100% detection accuracy. In DrDoS_DNS attacks, RNN achieves the least accuracy of 61%, LSTM has 56%, and proposed has a detection rate of 58%. In the DrDoS_LDAP attack, the existing technique as well as the proposed technique achieve a low value of 47%. DrDoS_SNMP also gives the same accuracy rate of 67% in RNN, LSTM, and the proposed model. DrDoS_SSDP gives 61% in RNN, 58% in LSTM, and the proposed achieves 52%. The attack DrDoS_UDP gives

Attack type	RNN	LSTM	IDSNet-PDO
TNRs	95	99	99
DrDoS_LDAPs	96	95	98
DrDoS_MSSQLs	97	94	95
DrDoS_NetBIOSs	68	96	96
DrDoS_UDPs	60	71	75
Syns	100	100	100
UDP-lags	0	0	0

Table 9. The performance experimental results comparative to benign and numerous kinds of attacks in Dataset_7_class.

Attack type	RNN	LSTM	Proposed
Normals	93	94	96
DDoSs	94	95	98
Injections	92	91	94
Passwords	91	92	93
Backdoors	93	95	96
Ransomwares	94	96	97
XSS	94	96	97
Scannings	94	95	97

Table 10. The presentation of deep learning tactics relative to normal and many categories of attacks in TON_IoT dataset.

Attack type	LSTM	RNN	Proposed
TFTP	98	100	94
DrDoS_NTP	91	91	92
WebDDoS	24	23	20
TNR	100	100	100
DrDoS_DNS	56	61	59
DrDoS_UDP	48	47	46
DrDoS_NetBIOS	97	93	73
DrDoS_MSSQL	56	55	56
Syns	64	64	64
DrDoS_LDAPs	47	47	47
DrDoS_SNMPs	67	67	68
DrDoS_SSDPs	58	61	52
UDP-lags	98	99	97

Table 11. Experimental findings on the efficacy of deep learning methods against both benign and malicious assaults on Dataset 13 class.

an accuracy of 47% in RNN, 48% in LSTM, and 46% in the proposed model. DrDoS_NetBIOS gives a moderate accuracy of 93% in RNN and 97% in LSTM, where the proposed method gives a lesser accuracy of 73%. Therefore, we attack various attacks, such as DrDoS_MSSQL and TFTP, and this is the rate of this attack, which must be improved in future work. Syn gives 64% in RNN and LSTM, where the proposed model gives 65%. TFTP gives maximum accuracy of 100% in RNN, 98% in LSTM, and the proposed achieves 94%. The other attacks, like DrDoS_NTP and UDP-lag, give 91% and 99% of detection accuracy in RNN and 91% and 98% in LSTM, where propose gives 92% and 97% of detection rate. In the WebDDoS attack, the experiment results give 23% accuracy in RNN, 24% in LSTM, and the proposed model attains a lesser of 20% accuracy.

Table 12 illustrates that TNR (BENIGN) gives accuracy in RNN and LSTM of 96.99, whereas the proposed accuracy of TNR (BENIGN) attacks is 99%, and the attack gives 100% accuracy in RNN, LSTM, and proposed.

Comparative analysis of proposed with existing techniques. Most of the existing techniques mentioned in "Related works" section use machine learning techniques for DDoS attacks, but they have used various datasets. Therefore, these generic techniques are implemented with our system, and the results are averaged in Table 13.

The average result provides a comparative analysis of various techniques in terms of different metrics. In the analysis of accuracy, the proposed model achieved 95.62%, whereas the existing practices achieved 80% to 94% accuracy. The auto-encoder achieved 91.68% of F-measure, 92.44% of precision, and 92.15% of recall; the

Attack type	RNN	LSTM	IDSNet-PDO
TNR	95	99	99
Attack	100	100	100

Table 12. The performance experimental results of deep learning approaches relative in Dataset_2_class (Binary classification).

Algorithm	Accuracy	Recall	F-score	Precision
SVM	92.46	92.45	91.81	93.48
RF	89.52	89.54	89.03	90.21
DT	80.10	80.15	80.43	87.21
LSTM	85.71	85.93	83.45	84.32
Auto-encoder	92.11	92.15	91.68	92.44
RNN	94.53	92.52	92.25	96.61
1D-CNN	94.16	92.32	92.10	96.17
IDSNet-PDO	95.62	94.62	94.53	98.32

Table 13. Overall average results of the proposed techniques with existing techniques.

LSTM model achieved 83.45% of F-measure, 84.32% of precision, and 85.93% of recall. Among other techniques, DT achieved 80% of recall, an F-measure of 80.43%, and 87.21% of precision, while the other model, called RF, achieved 89.54% of recall, 90.21% of precision, and 89.03% of F-measure. But the projected model achieved 94.62% recall, 98.32% precision, and 94.53% F-measure, where the reason for better performance is the usage of PDO for the selection of optimal features (the learning rate of IDSNet).

Conclusions

In the context of Agriculture 4.0, the investigated methods may be employed for traffic categorization via networks. This article contains a related works section with a collection of papers discussing the monitoring and categorization of network traffic. In this work, we create an IDSNet model that uses PDO to foresee potential attacks. In this work, we compared and contrasted the efficacy of strategies for agribusiness 4.0 cyber security. The CICDDoS2019 dataset and the TON IoT dataset, both of which include real-world traffic data, are used to compare and contrast the models' performances across binary and multiclass classifications. The findings reveal that deep learning approaches outperform key performance measures. Also, with an accuracy of 95% and a precision of 98.32% on the whole dataset, the IDS model based on CNN beats the best deep learning IDS approaches that were tested using the dataset. The study's findings on the use of ensemble techniques in network traffic categorization seem highly encouraging. This research will then be integrated into an application that requires historical and near-real-time studies for network assault categorization, allowing threats and anomalous traffic to be detected, isolated, and/or alerted to. We also recommend testing these kinds of models on data from different sources and in other application areas. Moreover, similar approaches may be used in fields other than agriculture to learn more about the opportunities and limitations of various datasets.

Ethics approval. The submitted work is original and has not been published elsewhere in any form or language.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 24 December 2022; Accepted: 13 September 2023
Published online: 16 September 2023

References

1. Huh, J. H. Implementation of lightweight intrusion detection model for security of smart green house and vertical farm. *Int. J. Distrib. Sens. Netw.* **14**(4), 1550147718767630 (2018).
2. Eskandari, M., Janjua, Z. H., Vecchio, M. & Antonelli, F. Passban IDS: An intelligent anomaly-based intrusion detection system for IoT edge devices. *IEEE Int. Things J.* **7**(8), 6882–6897 (2020).
3. Ciklabakkal, E., Donmez, A., Erdemir, M., Suren, E., Yilmaz, M. K. & Angin, P. ARTEMIS: An intrusion detection system for MQTT attacks in Internet of Things. In *2019 38th Symposium on Reliable Distributed Systems (SRDS)* 369–3692. IEEE (2019).
4. Ioulianou, P., Vasilakis, V., Moscholios, I. & Logothetis, M. A signature-based intrusion detection system for the internet of things. *Information and Communication Technology Form* (2018).
5. de Araujo Zanella, A. R., da Silva, E. & Albini, L. C. P. Security challenges to smart agriculture: Current state, key issues, and future directions. *Array* **8**, 100048 (2020).
6. Veena, S., Mahesh, K., Rajesh, M. & Salmon, S. The survey on smart agriculture using IOT. *Int. J. Innov. Res. Eng. Manag. (IJRIREM)* **5**(2), 63–66 (2018).
7. Pirozmand, P., Ghafary, M. A., Siadat, S. & Ren, J. Intrusion detection into cloud-fog-based iot networks using game theory. *Wirel. Commun. Mobile Comput.* (2020).
8. Sivabalan, S. & Radcliffe, P. J. Feasibility of eliminating IDPS devices from a web server farm. *Int. J. Netw. Secur.* **20**(3), 433–438 (2018).
9. Kfoury, E., Saab, J., Younes, P. & Achkar, R. A self organizing map intrusion detection system for RPL protocol attacks. *Int. J. Interdiscip. Telecommun. Netw. (IJITN)* **11**(1), 30–43 (2019).
10. Cristiani, A. L. *et al.* A fuzzy intrusion detection system for identifying cyber-attacks on iot networks. In *2020 IEEE Latin-American Conference on Communications (LATINCOM)*. IEEE (2020).
11. Santhoshi, K. & Bhavana, S. Intruder recognition in a farm through wireless sensor network. *Int. J. Adv. Res. Ideas Innov. Technol.* **4**(3), 667–669 (2018).

12. Mohapatra, H., Rath, S., Panda, S. & Kumar, R. Handling of man-in-the-middle attack in wsn through intrusion detection system. *Int. J.* **8**(5), 1503–1510 (2020).
13. Astillo, P. V., Kim, J., Sharma, V. & You, I. SGF-MD: behavior rule specification-based distributed misbehavior detection of embedded IoT devices in a closed-loop smart greenhouse farming system. *IEEE Access* **8**, 196235–196252 (2020).
14. Rahman, S. A., Tout, H., Talhi, C. & Mourad, A. Internet of things intrusion detection: Centralized, on-device, or federated learning?. *IEEE Netw.* **34**(6), 310–317 (2020).
15. Choudhary, G., Sharma, V., You, I., Yim, K., Chen, R. & Cho, J. H. Intrusion detection systems for networked unmanned aerial vehicles: A survey. In *2018 14th International Wireless Communications and Mobile Computing Conference (IWCMC)*, 560–565. IEEE (2018).
16. Chockwanich, N. & Visoottiviseth, V. Intrusion detection by deep learning with tensorflow. In *2019 21st International Conference on Advanced Communication Technology (ICACT)* 654–659. IEEE (2019).
17. Othman, S. M., Alsohybe, N. T., Ba-Alwi, F. M. & Zahary, A. T. Survey on intrusion detection system types. *Int. J. Cyber-Security Digit. Forensics* **7**(4), 444–463 (2018).
18. Saharkhizan, M. *et al.* A hybrid deep generative local metric learning method for intrusion detection. In *Handbook of Big Data Privacy* 343–357 (Springer, Cham, 2020).
19. Ferrag, M. A., Shu, L., Djallel, H. & Choo, K. K. R. Deep learning-based intrusion detection for distributed denial of service attack in Agriculture 4.0. *Electronics* **10**(11), 1257 (2021).
20. Friha, O. *et al.* FELIDS: Federated learning-based intrusion detection system for agricultural Internet of Things. *J. Parallel Distrib. Comput.* **165**, 17–31 (2022).
21. Ferrag, M. A., Shu, L., Friha, O. & Yang, X. Cyber security intrusion detection for agriculture 4.0: Machine learning-based solutions, datasets, and future directions. *IEEE/CAA J. Autom. Sin.* **9**(3), 407–436 (2021).
22. Raghuvanshi, A., Singh, U. K., Sajja, G. S., Pallathadka, H., Asenso, E., Kamal, M., Singh, A. & Phasinam, K. Intrusion detection using machine learning for risk mitigation in IoT-enabled smart irrigation in smart farming. *J. Food Quality* (2022).
23. Vangala, A., Das, A. K., Chamola, V., Korotaev, V. & Rodrigues, J. J. Security in IoT-enabled smart agriculture: architecture, security solutions and challenges. *Clust. Comput.* **26**(2), 879–902 (2022).
24. Otoum, Y., Liu, D. & Nayak, A. DL-IDS: A deep learning–based intrusion detection framework for securing IoT. *Trans. Emerg. Telecommun. Technol.* **33**(3), e3803 (2022).
25. Sengan, S., Khalaf, O. I., Sharma, D. K. & Hamad, A. A. Secured and privacy-based IDS for healthcare systems on E-medical data using machine learning approach. *Int. J. Reliab. Quality E-Healthc. (IJRQEH)* **11**(3), 1–11 (2022).
26. Lin, Z., Shi, Y. & Xue, Z. Idsgan: Generative adversarial networks for attack generation against intrusion detection. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* 79–91 (Springer, Cham, 2022).
27. Maranhão, J. P. A., da Costa, J. P. C., de Freitas, E. P., Javidi, E. & de Sousa, R. T. Noise-robust multilayer perceptron architecture for distributed denial of service attack detection. *IEEE Commun. Lett.* **25**(2), 402–406 (2020).
28. Alsaedi, A., Moustafa, N., Tari, Z., Mahmood, A. & Anwar, A. TON_IoT telemetry dataset: A new generation dataset of IoT and IIoT for data-driven intrusion detection systems. *IEEE Access* **8**, 165130–165150 (2020).
29. Sharafaldin, I., Lashkari, A. H., Hakak, S., Ghorbani, A. A. Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy. In *Proceedings of the 2019 International Carnahan Conference on Security Technology (ICCST)*, 1–8, Chennai, India, 1–3 October (2019).
30. TON_IOT DATASETS. Available online: https://ieee-dataport.org/documents/toniot-datasets (accessed on 4 April 2021).

Author contributions

R.V. and D.V.: Literature Review and Proposed Algorithm; G.M. and V.K.D.: Implementation; M.M. and N.S.Y.: Results and Discussion.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to R.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023