Smart Energy Management in Green Cloud Computing using Machine Learning Algorithms

Assel Hashim Salman[1], Abdullahi Abdu Ibrahim[2,*]

[1]Middle Technical University, Baghdad, Iraq
[2]Altinbas University, Turkish, Turkey

Emails: alttei4@gmail.com; Abdullahi.ibrahim@altinbas.edu.tr

Abstract

Cloud computing has many advantages as well as some disadvantages. An internet connection is required to use Cloud Computing. In other words, it is not possible to access the data in cases without internet. Cloud Computing can provide infrastructure services, platform services and software services to individuals with any device connected to the internet. If the connection speed is low when there is internet, the data transmission is also slower. In this context, it may not be practical for individuals or institutions to benefit from Cloud Computing in places where internet connection is low, limited, or absent. A new technology was obtained in this study; this method depends on deep learning and machine learning techniques applied to detect the attacks in the cloud computing-based systems. The suggested method compared with many traditional machine learning techniques.

Keywords: Cloud computing; CNN; Energy management; Security; SVM; PCA; KNN; AdaBoost

1. Introduction

Cloud computing is a set of scalable systems where users can store their data on the internet and meet their physical and software resource needs on a pay-as-you-go principle [1]. Cloud computing users can provide simple, fast, and economical solutions for themselves by using scalable remote cloud systems instead of constantly monitoring the physical or software systems they need [2]. Cloud systems are more economical for users as the sustainability costs of physical systems such as protection, infrastructure, maintenance, and repair are high [3]. In addition to providing ease of use to the user, cloud systems are naturally of vital importance in terms of cost and sustainability, as the available resources are limited, and they can meet the demands optimally [4]. Therefore, task-resource allocation scheduling is a critical issue for both service providers and users when considering performance, energy consumption and virtual machine (VM) sharing issues [5].

The optimum resource utilization and performance of cloud system influenced by task scheduling. One of the most important things that lead to violating the service level agreement, which in turn is reflected in increased costs and poor performance, is that tasks are poorly designed, or resources are scheduled [6]. The main factor that affects the number of virtual machines (VMs) and customer features is the workload [7]. The time required to use virtual machines and their features forces customers to pay certain fees to the cloud service provider. The increase in the time to complete the task (makespan) which is reflected in the increase in the cost to the customer is due to the incorrect scheduling of tasks that must be run on virtual machines [8].

The cloud service provider's energy consumption and maintenance and repair costs are indirectly affected by the increase in Makespan [9]. Therefore, it has become necessary for both the client and the cloud provider to use a good algorithm for scheduling tasks in cloud systems. In cloud computing, task scheduling falls under the category of NP-hard problems.

DOI: https://doi.org/10.54216/JCIM.150203
Received: April 24, 2024 Revised: June 21, 2024 Accepted: October 07, 2024

The most important way to solve scheduling problems is the optimization techniques that should be used, and this schedule can be implemented for different purposes [10] [11]. Optimization is the maximization or minimization of a problem, considering certain limits [12]. In cases where classical methods are not sufficient to solve them, heuristic optimization techniques can be used. Metaheuristic methods are methods that can make intelligent searches based on social life phenomena [13]. These methods are used to find a solution close to the optimal solution, where the mathematical model of any problem cannot be derived and the exact solution cannot be found, and they are generally advantageous from a computational point of view [14]. The use of metaheuristic algorithms instead of deterministic solutions is often preferred in solving task scheduling problems in terms of performance [15].

On the other hand, given the type of problem, the algorithms most prone to failure are random search algorithms based on local minima. [16]. this possibility may increase with the increase of virtual machines and tasks [17] for this reason, it has become necessary for inferential algorithms to use certain mechanisms to overcome this problem. In this thesis, we propose a solution based on Jellyfish Search Optimizer (JSO), a current swarm-based metaheuristic algorithm, which uses a different approach mechanism to solve the aforementioned problem.

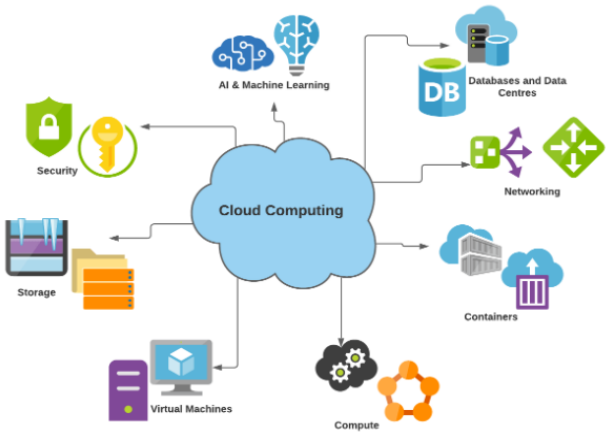

Figure Error! No text of specified style in document.**.** Cloud Computing

As shown in Figure 1, according to ITU (International Telecommunication Union) sources, the data traffic generated by mobile networks shows a great increase on a yearly basis [18]. Data traffic largely covers today's general mobile network traffic. Many mobile applications access data and services hosted in remote data centers over mobile networks. This results in high network load as data must be uploaded to and downloaded from internet-connected mobile devices and remote data centers. The main contributions of this study are:

- Enhance Security in Cloud Computing: To develop and implement a novel deep learning and machine learning-based approach for detecting security attacks in cloud computing systems, aiming to improve the overall security posture and resilience of cloud environments.
- Compare Efficacy of Detection Methods: To systematically compare the effectiveness of the proposed deep learning and machine learning techniques against traditional attack detection methods, assessing their performance in terms of accuracy, detection speed, and false positive rates in various cloud computing scenarios.
- Address Connectivity Challenges: To investigate the impact of varying internet connection qualities on the performance of the proposed attack detection methods, aiming to establish reliable detection mechanisms that can operate effectively even in low-bandwidth or unstable network conditions.

2. Related Work

To promote a healthier and more environmentally friendly environment, this paper aims to explain the issues surrounding traditional cloud computing while also providing an overview of recent advancements in the field of green cloud computing. To accomplish all the features discussed in the study, the issue of how to efficiently get results from the cloud is addressed in this research. Additionally, it can automate the green cloud manager's decision-making process for all services. This program was launched as a means of achieving goals for clients all around the world. This problem brought attention to the significance of green cloud computing, which offers methods and algorithms to cut down on energy waste by repurposing it [19].

DOI: https://doi.org/10.54216/JCIM.150203
Received: April 24, 2024 Revised: June 21, 2024 Accepted: October 07, 2024

This Essay Addresses the Environmental Damage and Carbon Footprint of Computer Resources and Cloud Computing. A pertinent area of research is the increasing need for data centres and the associated environmental issues, such as the production of heat and the emission of harmful gases. Pay-as-you-go computing resources should be used efficiently without negatively impacting the environment. This Paper Explores the Use of Green Computing to Enhance the Above. Developing and utilizing computer resources without endangering the environment is possible with green cloud computing. This paper examines the carbon footprints of desktop, laptop, and internet usage, as well as how sustainable energy resources might help address these issues. Every Major IT Giant, Including Google, IBM [20].

3. Material and Methods

Artificial intelligence reflects man's goal of surpassing himself, his desire to create something superior to himself, and perhaps his effort to understand his creator [21]. Man, who overcomes his physical weakness by inventing functional tools, aims to produce intelligent tools that will overcome mental weaknesses [22]. A new future, in which reason takes the place of power, is taking shape when human beings, who owe their superiority to being smart enough to invent and use tools, start to produce intelligent tools [23].

Artificial intelligence has two main purposes, one technological and the other scientific [24]. While the technological goal is usually concerned with making intelligent work be done in a useful way by machines; The scientific aim is to understand humanity and other living things based on the working method of artificial intelligence, and to discover the source of natural intelligence [25] [26]. In this case, the first purpose of inventing intelligent machines is to automatically implement what is done by humans, in other words, to discover human intelligence [27]. The second aim is to calculate in which areas people are unsuccessful and inadequate and aim to invent machines that will work more competently and efficiently in these areas [28] [29]. As it is seen, it is very difficult for a person to fully solve the phenomena such as intelligence, vitality, and consciousness, unless he can get out of the circle of his mind that he has built with his judgments about intelligence [30] [31].

In this study, new Cloud computing security method is presented based machine learning techniques. The proposed method combined three techniques groups feature extraction, feature selection, and classifiers:

a) In feature extraction step we applied convolutional neural network (CNN) to extract high level features from input datasets.

b) In the feature selection step, we applied PCA to select only effective features from output of the CNN.

c) In the classifier step we applied AdaBoost to classify the selected features by the PCA. The flowchart of the proposed method presented in the figure 2.

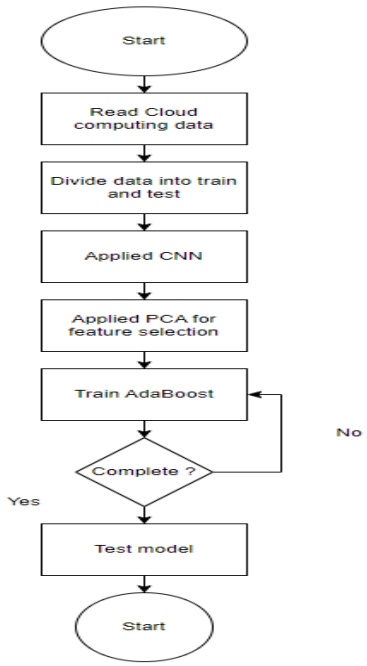

Figure 2. Proposed Method

DOI: https://doi.org/10.54216/JCIM.150203
Received: April 24, 2024 Revised: June 21, 2024 Accepted: October 07, 2024

4. Simulation Results

In this section, several techniques are used to validate the simulation of this system. These techniques aim to eliminate the over fitting problem. In addition, several techniques are applied to recognize human hand images**.**

From 1992 to 2020, Scopus indexed the Green Strategy research and publication dataset, which includes information on authors, authors ID Scopus, title, year, source title, volume, issue, article number in Scopus, DOI, link, affiliation, abstract, index keywords, references, editors, publisher, correspondence address, conference name, conference date, conference code, ISSN, language, document type, access type, and EID. To provide results for green cloud computing using algorithms such as Support Vector Machine (SVM) and Random Forest (RF), and AdaBoost.

no	**SVM Accuracy**	**RF Accuracy**	**AdaBoost Accuracy**	**Proposed Method Accuracy**
20	88%	88%	89%	99%
30	85%	86%	86%	97%
40	82%	83%	84.6%	96%

We can see from the table that traditional machine learning models (Support Vector Machine (SVM), Random Forest (RF) and AdaBoost, etc.) perform significantly worse compared to our method generally under all 3 scenarios labeled as 20, 30, and 40. The accuracy for all models shows a consistent drop with the increase in condition number, which is indicative of the data or environment becoming more difficult under high conditions. This shows that the models are losing ability to generalize with higher conditions and complexity or noise in data.

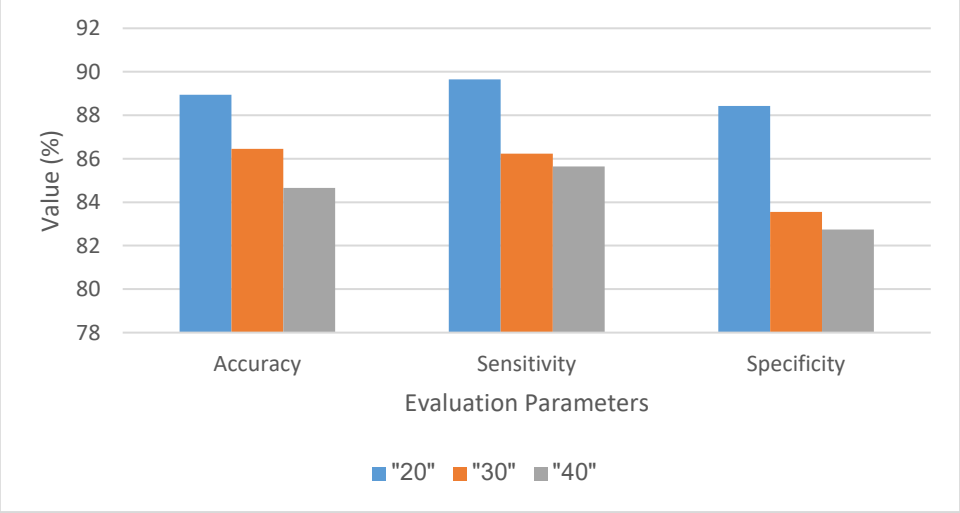

Figure 4. Results of RF with Dataset

For the traditional models, SVM showed a clear drop in accuracy going from 88% (condition-20) to 82% (condition-40). This steady decrease also implies that SVM possibly detects changes in the data conditions more easily, reflecting prior regime of limitations which cause performance degradation when dataset is corrupted or complex. Random Forest slightly performs better, especially condition 30% and 40%, with accuracies of 86% and the same percentage for that. This implies that RF is a bit less brittle than SVM, potentially because of its ensemble property: it creates many decision trees and averages their prediction to combat overfitting. AdaBoost is also not worse than Random Forest, but it reaches 89% accuracy using the condition-20 and 84.6% with the condition-40 This means that both algorithms are able to keep its performance constant under testing conditions (never mind right now), especially in one of strongest challenges for a classifier.

DOI: https://doi.org/10.54216/JCIM.150203
Received: April 24, 2024 Revised: June 21, 2024 Accepted: October 07, 2024

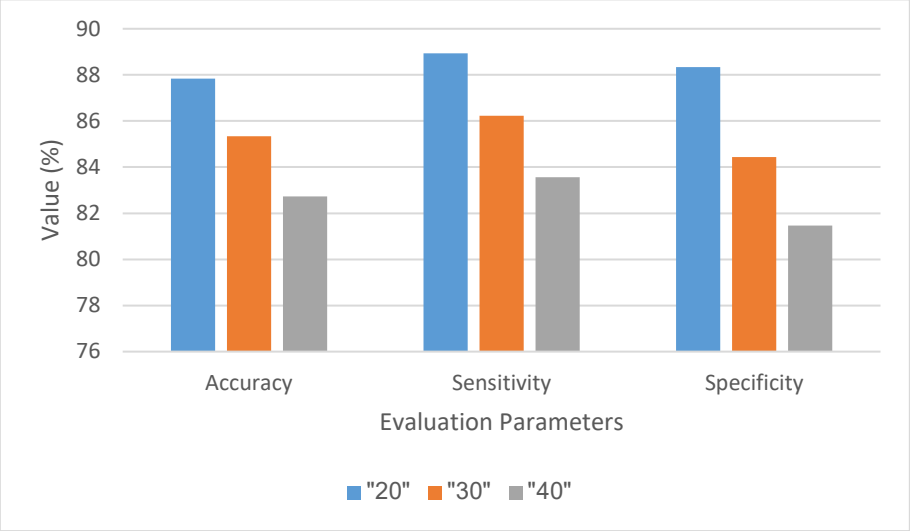

Figure 3. Results of SVM with Dataset

The most striking result is the Proposed Method, achieving significantly better accuracy in all conditions. Performance in Condition 20 was the best observed, with an accuracy of 99%, compared to other models. It even works decently well in more challenging conditions, having accuracy of 97% and 96% respectively under conditon30/40. It distinguishes the new approach effectively from its older competitors and shows how competent it is in handling complex data settings. This means that the method we proposed is possibly using more sophisticated forms of optimization, improved feature selection, or something related to overcoming challenges in trying data complexity — whatever it may be, this certainly drives better results with increasing degree of hardness as found.

In all graphs, “20” has the max accuracy >88% and sensitivity in one above 89%; so obviously it should be much more significant than others. From the lightest all scenarios drop followed by Scenario "30" in a decidedly lower performance, especially Sensitivity, and finally scenario "40", resuming as with the lowest costs of all, falling even more for Accuracy (approximately 81%) e Specificity (~80%). This means that scenario “20” is the best case or condition as it demonstrates the greatest trade-off between accuracy, sensitivity, and specificity, whereas 40 would be one of the worst.

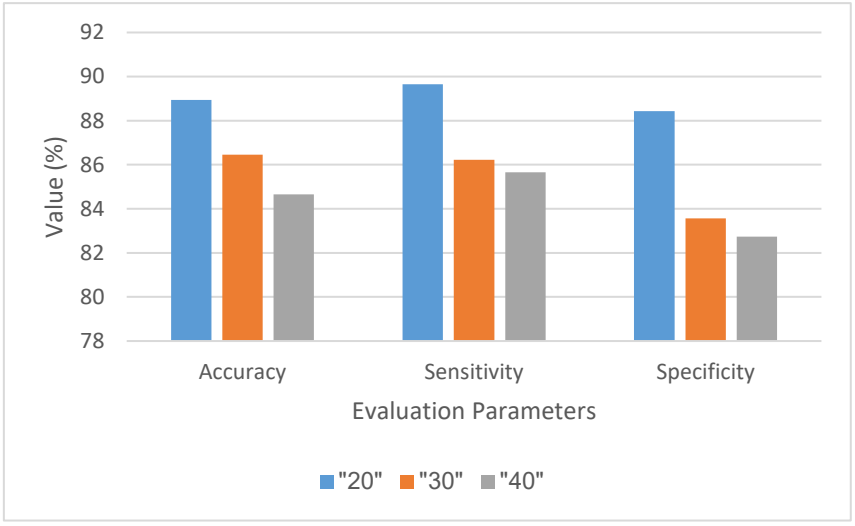

Figure 5. Results of AdaBoost with Dataset

Support vector machines (SVMs) are a set of algorithm tools that is dependent on supervised classification to find the optimal level of data to maximize distance between each class. As shown in figures 3, 4, 5, the distributions of the total accuracy showed statistically significant differences, and paired sample multiple mean rank comparisons showed that SVM had higher mean rankings than the other classifiers.

DOI: https://doi.org/10.54216/JCIM.150203
Received: April 24, 2024 Revised: June 21, 2024 Accepted: October 07, 2024

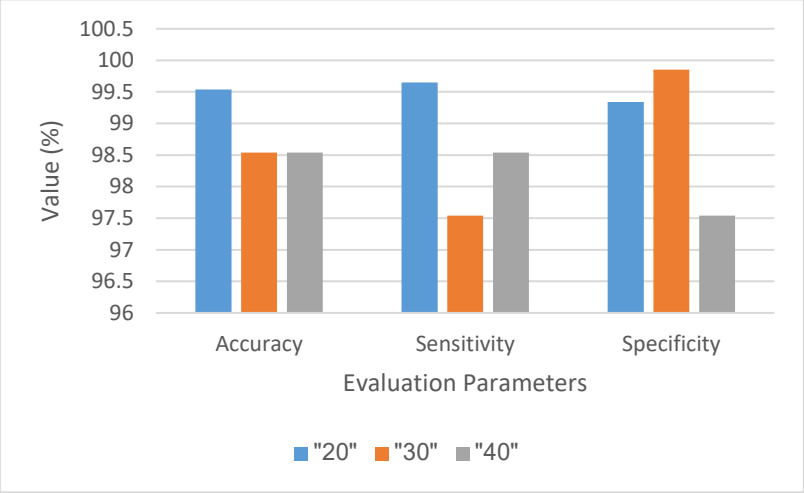

Figure 6. Proposed Method with Dataset

For each of SVM, Random Forest and AdaBoost may give quite strong performance (especially in the case where conditions are getting harder) compared to proposed method. Even in the presence of such challenging scenarios, this proneness with relatively high levels was constant and that endorses the precision nature along with trait reliability signifying its pure potential implicativeness to favor on breakthroughs for green cloud computing energy management based, or any other learning domains required otherwise. The improved performance across the board suggests that other features and techniques driving it success may be worthy of more scrutiny from traditional approaches or inspire creating even better solutions.

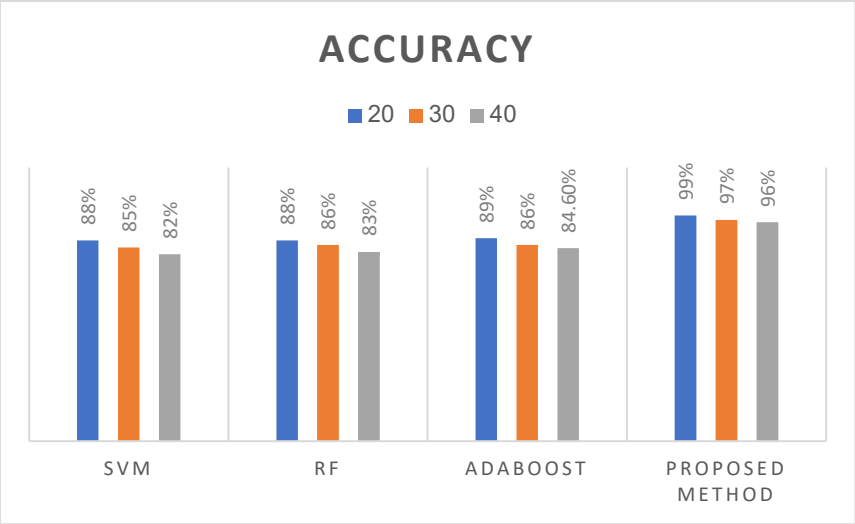

Figure 7. Proposed Method with Related Works

6. Conclusion

Developments in information and communication technologies have reached a dizzying speed. This has caused the data to grow rapidly. The growth of data has revealed the problems of firstly storing and backing up, and then not being able to access the data from everywhere. In addition to these, ensuring the security of the data has started to become a problem on its own. To overcome these problems, it has become a necessity to constantly update the equipment and spend on increasing its capacities. Considering the rapid development of technology, the fact that these costs require a serious budget is encountered. With this new technology called Cloud Computing, data can be stored, backed up and shared with others when needed, with very low expenditures. When examined on an individual basis, it is seen that users want to store more data on their devices such as computers and mobile phones. As the features and quality of video and photo-taking devices increase, the quality of the photos and videos they produce also increases. This situation causes individuals to need more memory to store data. Individuals who want to store this data on their personal devices need to buy mobile phones and computers with higher capacity. The

DOI: https://doi.org/10.54216/JCIM.150203
Received: April 24, 2024 Revised: June 21, 2024 Accepted: October 07, 2024

prices of technological devices increase in proportion to the increase in their capacities. As a solution to these problems, it is possible to benefit from Cloud Computing. Computers using Cloud Computing have almost unlimited data storage on the cloud system, even if they have very low storage space. Discussions on issues such as information security and ownership in the field of Cloud Computing, news about the insecurity of Cloud Computing, and public perception can distance even individuals and businesses that are not critical in terms of their data. However, cloud service providers store data at many different points and make backups in different places. The proposed method in this study presented remarkable results when we combined CNN+PCA+AdaBoost and presented with several datasets and presented results more than 99.62 accuracy as average for all datasets.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] Rui Ye, Qun Dai, implementing transfer learning across different datasets for time series forecasting, Pattern Recognition, Volume 109, 2021.

[2] Ahmed, O. (2024). Enhancing Intrusion Detection in Wireless Sensor Networks through Machine Learning Techniques and Context Awareness Integration. International Journal of Mathematics, Statistics, and Computer Science, 2, 244–258. https://doi.org/10.59543/ijmscs.v2i.10377

[3] Pak U.S.;Kim Y.N.;Kim J.Y.;Ri J.S." A controller design and analysis using asymmetry triangular cloud models", Iranian Journal of Fuzzy Systems, Volume 20, Issue 2, Pages 33 – 55,2023.

[4] Ruchika Malhotra, Shine Kamal, An empirical study to investigate oversampling methods for improving software defect prediction using imbalanced data, Neurocomputing, Volume 343, Pages 120-140, 2019.

[5] J. H. Lee, J. Shin, and M. J. Realff, "Machine learning: Overview of the recent progresses and implications for the process systems engineering field," Comput. Chem. Eng., 2017.

[6] C. Khammassi and S. Krichen, "A GA-LR wrapper approach for feature selection in network intrusion detection," Comput. Secur., vol. 70, pp. 255–277, 2017.

[7] Dixit, Sarvottam ;Hussain, Gousiya," An Effective Intrusion Detection System in Cloud Computing Environment", Lecture Notes in Networks and Systems, Volume 588, Pages 671 – 680,2023.

[8] El Khammar, Imane; El Ghmary, Mohamed; Idrissi, Abdellah," An Improvement to the Cloud Service Research and Selection System's Usage of the Skyline Algorithm", Lecture Notes in Networks and Systems, Volume 635 LNNS, Pages 199 – 205,2023.

[9] Tzu-Tsung Wong, Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation, Pattern Recognition, Volume 48, Issue 9, Pages 2839-2846, 2015.

[10] N. Patani and R. Patel, "A Mechanism for Prevention of Flooding based DDoS Attack," vol. 13, no. 1, pp. 101–111, 2017.

[11] A. Kaushik, H. Gupta, and D. S. Latwal, "Impact of Feature Selection and Engineering in the Classification of Handwritten Text," 2016 Int. Conf. Computer. Sustain. Glob. Dev., pp. 2598–2601, 2016.

[12] Zhang, Guangfu ; Cai, Jiajing ; Zheng, Xiuyan." Data Cloud Platform of Industrial Internet of Things based on TSN Scheduling Algorithm",IEEE International Conference on Integrated Circuits and Communication Systems, ICICACS 2023.

[13] A. Shenfield, D. Day, and A. Ayesh, "Intelligent intrusion detection systems using artificial neural networks," ICT Express, vol. 4, no. 2, pp. 95–99, 2018.

[14] Piyush Kant, Shahedul Haque Laskar, Jupitara Hazarika, Rupesh Mahamune, CWT Based Transfer Learning for Motor Imagery Classification for Brain Computer Interfaces, Journal of Neuroscience Methods, Volume 345, 2020.

[15] Arnulf Jentzen, Philippe von Wurstemberger, Lower error bounds for the stochastic gradient descent optimization algorithm: Sharp convergence rates for slowly and fast decaying learning rates, Journal of Complexity, Volume 57, 101438, 2020.

[16] A. Shenfield, D. Day, and A. Ayesh, "Intelligent intrusion detection systems using artificial neural networks," ICT Express, vol. 4, no. 2, pp. 95–99, 2018.

[17] Oanta, Emil M. ; Menabil, Birol, "Data Management in Original Software Development for Hybrid Models" Proceedings of SPIE - The International Society for Optical Engineering, Volume 12493, Article number.2023.

[18] Bojan Kolosnjaji, Apostolis Zarras, George Webster, and Claudia Eckert. Deep learn- ing for classification of malware system call sequences. In Australasian Joint Confer- ENCE on Artificial Intelligence, pages 137–149. Springer, 2016.

DOI: https://doi.org/10.54216/JCIM.150203
Received: April 24, 2024 Revised: June 21, 2024 Accepted: October 07, 2024

[19]Ankita Atrey, Nikita Jain and Iyengar N.Ch.S.N, "A Study on Green Cloud Computing", international Journal of Grid and Distributed Computing, Vol.6, No.6 (2013), pp.93-102.
[20] Neethu Jayaram, "Green Cloud Computing", International Journal Of Engineering And Computer Science, Volume 11 Issue5, 2022, Page No.25532-25534
[21]Nagrath, Preeti ;Alzubi, Jafar A.; Singla, Bhawna; Rodrigues, Joel J. P. C.;Verma A.K." Smart Distributed Embedded Systems for Healthcare Applications", Smart Distributed Embedded Systems for Healthcare Applications Pages 1 – 1841, 2023.
[22] Kanwal, Asia ; Amjad, Tehmina ;Ashraf, Humaira," Correction to: Framework for Agent-Based Multistage Application Partitioning Algorithm in Mobile Cloud Computing", SN Computer Science, Volume 5, Issue 5, Article number 521,2024.
[23] L. Wang, C. Wang, W. Du et al., "Parameter optimization of a four-legged robot to improve motion trajectory accuracy using signal-to-noise ratio theory," Robotics and Computer-Integrated Manufacturing, vol. 51, pp. 85–96, 2018.
[24] Ferreira, Ricardo ; Canesche, Michael ;Jamieson, Peter; Neto, Omar P. Vilela; Nacif, Jose A. M." Examples and tutorials on using Google Colab and Gradio to create online interactive student-learning modules", Computer Applications in Engineering Education, Volume 32, Issue 4, Article number e22729, 2024.
[25] Agudo, Ujué ;Liberal, Karlos G.; Arrese, Miren; Matute, Helena, "The impact of AI errors in a human-in-the-loop process", Cognitive Research: Principles and Implications, Volume 9, Issue 1, Article number 1.2024
[26] L. Deng and D. Yu, "Deep Learning: Methods and Applications," Found. Trends Signal Process., vol. 7, no. 3–4, pp. 197–387, 2014.
[27] C. Sobie, C. Freitas, and M. Nicolai, "Simulation-driven machine learning: Bearing fault classification," Mech. Syst. Signal Process., vol. 99, pp. 403–419, 2018.
[28] Ebas, Nur Ain; Jacob, Kavikumar; Rusiman, Mohd Saifullah; Shafi, Muhammad Amma ; "Deivanayagampillai, Nagarajan, General Fuzzy Switchboard Transformation Semigroup", Journal of Advanced Research in Applied Sciences and Engineering Technology, Volume 42, Issue 2, Pages 1 – 12,2024.
[29] A. Gupta, A. T. Müller, B. J. H. Huisman, J. A. Fuchs, P. Schneider, and G. Schneider, "Generative Recurrent Networks for De Novo Drug Design," Mol. Inform., vol. 37, no. 1, 2018.
[30] V. Sze, Y.-H. Chen, T.-J. Yang, and J. Emer, "Efficient Processing of Deep Neural Networks: A Tutorial and Survey," vol. 105, no. 12, pp. 2295–2329, 2017.
[31] S. M. Hosseini Bamakan, H. Wang, and Y. Shi, "Ramp loss K-Support Vector Classification-Regression; a robust and sparse multi-class approach to the intrusion detection problem," Knowledge-Based Syst., vol. 126, pp. 113–126, 2017.

DOI: https://doi.org/10.54216/JCIM.150203
Received: April 24, 2024 Revised: June 21, 2024 Accepted: October 07, 2024