

**An Intelligent Intrusion Detection Using Deep Learning on CICIDS2018
for Cloud Security**

Tilak Sharma

M.E Scholar, Department of Computer Science and Engineering
Maharana Pratap College of Technology, Gwalior, MP

Unmukh Datta

Associate Professor, Department of Computer Science and Engineering
Maharana Pratap College of Technology, Gwalior, MP

Abstract- By enabling on-demand application use, computation, and data storage across the internet, cloud computing has moved to the front of modern digital infrastructure. Its distributed and dynamic nature makes it susceptible to various security risks, including anomalies and malicious traffic. As cyberattacks develop increasingly sophisticated, conventional detection technologies are not always up to the task of delivering quick and reliable responses. Still, this work provides a solid framework for detecting aberrant and hazardous network traffic in cloud settings using deep learning. The suggested method relies on the CICIDS2018 dataset, which contains information on actual network traffic as well as several cyberattack scenarios. To handle class imbalance, the proposed method uses a full data preparation pipeline including data cleaning, normalisation, and SMOTE (Synthetic Minority Over-sampling Technique) application. There are also robust feature selection methods like Random Forest and Symmetric Uncertainty. The methods guarantee that the pertinent, high-quality input data utilised to train the model is of high quality. The data on network traffic is then automatically learnt by a Convolutional Neural Network (CNN) model. Techniques like routine normalisation and dropout help to enhance the model's performance and stop overfitting. The model demonstrates excellent performance with a training accuracy of 99.85% and a testing accuracy of 99.76%. This approach takes into account the necessity for scalability and agility in cloud infrastructures as they evolve, therefore providing excellent detection accuracy. Intelligent intrusion detection systems (IDS) tailored to contexts powered by Industry 4.0 and the Internet of Things (IoT) have advanced significantly with this development.

Keywords- Cloud Security, Anomaly Detection, Deep Learning, Convolutional Neural Network, Intrusion Detection System (IDS)

I. INTRODUCTION

The arrival of cloud computing has significantly changed how computer resources are handled, therefore opening more flexibility, scalability, or cost-effectiveness. Protecting cloud configurations has also become more and more important as more people depend on programs hosted, stored data, and processing by cloud-based services. [1]. Finding and prevention of hazardous or dubious network activity is a critical security issue for cloud computing platforms. Many different elements could compromise the integrity, privacy, and operation of systems stored on the cloud. Featuring a denial of (also known as assaults, unauthorised access, data hacking, and more.[2]. When faced with the complicated nature and volume of contemporary cloud communications, conventional security methods such firewalls & rule-driven detection

of attacks technologies can fail. These methods lack the ability to manage always changing dangers since they sometimes depend on set signatures and laws.[3]. The dynamic, multi-tenant nature of the system makes it difficult to spot aberrant behaviours since normal user behaviour could differ significantly among support or users in cloud settings. The strong pattern recognition, anomaly identification, and distinct threat analysis features of deep learning provide a potential answer to these constraints.[4].

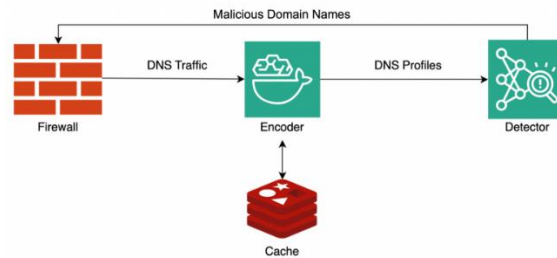


Fig.1 DNS-Based Malicious Domain Detection Architecture

Applying deep learning methods helps one to notably better understand network traffic data; models like CNNs, RNNs, and LSTMs show exceptional performance in identifying complex patterns and temporal correlations.[5]. These algorithms can automatically learn how to tell between the two and identify genuine and harmful activities in raw traffic data. Deep learning models' capacity to learn from large datasets directly reduces dependency on human feature scientists and continuous rule updates, enabling them to precisely identify new attack routes.[6].

Even if the attack was previously undiscovered, deep learning-based anomaly detection in clouds aims to identify traffic patterns that differ from the regular baseline. The ability to detect zero-day assaults and other advanced threats is a result of this.[7]. Conversely, malicious traffic detection seeks to categorise distinct attack types through labelled data, allowing cloud administrators to implement focused countermeasures. Improved incident management, automated responses, and real-time threat detection are all possible outcomes of incorporating such detection systems based on deep learning into cloud infrastructure, which in turn strengthens the overall security posture.[8].



Fig.2 Symbolic Representation of Digital Security in a Cloud Environment

Finally, a major step forward in cybersecurity is the use of deep learning to identify suspicious activity and harmful communications in cloud computing settings. [9]. There is a rising demand for security solutions that are intelligent, adjustable, and adaptive due to the increasing adoption of cloud computing. Improving detection rates, reducing false positives, and taking a

preventive approach to securing the cloud are all benefits of deep learning's robust framework, which can handle the intricacies of modern cloud traffic.[10]. Lighter models for real-time identification, privacy-preserving approaches, and complete threat intelligence through multi-source data integration will likely be the emphasis of future research in this field.

II. LITERATURE REVIEW

Abhishek Sharma (2022) et al Several potential dangers were recognised in this setting. Used to choose the best features with the use of linear regression techniques after picking features and risk factor identification. Next, dangers in the CC environment are examined using AI-ML methods such as decision tree (DTC), Randomizable Filter Algorithm, and k-star with RMSE. The results of the experiments showed that, out of all the partitioning options, dividing the dataset to 95%-5 % produced the best results.[11]

Rojalina Priyadarshini (2022) et al In this research, we present a new method for defending against distributed denial of service assaults (DDoS) that relies on a single point of failure (SBP) and works equally well in cloud and fog environments. The DDoS defender module is deployed at the SDN controller using Software Defined Network (SDN), which detects abnormal behaviour of DDoS attacks at the network and transport levels.[12].

Lumbardha Hasimi (2023) et al In order to better understand how to include ANN models into cloud security measures, this study presents a feed-forward propagation model and delves into the essential procedures for doing so. Using a dataset from Kaggle.com for validation, the study highlights procedures involved in the development and evaluation of the ANN model. It comes to light that the performance of the model depends on criteria such as training data quantity, structure of the network, and weight adjustment algorithms.[13].

Amol D. Vibhute (2023)et al Datasets from CSE-CIC-IDS2018 were used for the experiments. Testing accuracy for the suggested CNN model was 97.07%, and the error rate was 2.93%, according to the data. Accuracy (98.11), recall (96.93), and f1-score (97.52%) were additional metrics used to evaluate the suggested model's performance. These findings show great promise for real-time Industry 4.0 systems, as they are more accurate, precise, and capable of detecting network irregularities with the utmost accuracy.[14].

Dr.P. Bharath Kumar Chowdary (2024)et al We have tested the proposed ML-EIDS system and compared it to other approaches. Experiments are carried out using the NSL-KDD dataset. Findings from performance evaluations and comparisons show that the proposed system outperforms existing approaches in reliably detecting anomalies with high precision and low false alarm occurrence rates using the ML-EIDS hybrid method.[15].

TABLE.1 LITERATURE SUMMARY

Author /Year	Methodology	Result	Limitation	References
Andrea Sharon (2022)	The system leverages NSL-KDD, uses Sparse and Stacked Contractive	The model strong performance, achieving 99% precision, 98% recall, and over	The system may struggle to generalize to unknown attacks due to reliance on	[16]

	Autoencoders for feature extraction,	98% accuracy, effectively.	the NSL-KDD dataset.	
Priya Parameswarappa (2023)	A machine learning firewall uses a “most frequent decision” strategy, combining past node decisions with current predictions, and is trained on the UNSW-NB-15 dataset for cloud intrusion detection.	The method achieves 97.68% anomaly detection accuracy and increases classification accuracy from 95% to 97%.	The system's accuracy may decrease with limited attack samples and struggle with real-time or unknown attacks without deep learning improvements..	[17]
TAIWO JOSEPH AKINBOLAJI (2023)	This study combines AI, deep learning, and reinforcement learning in real-time cloud threat detection, using ensemble learning to boost accuracy, cut false positives, and optimize performance.	The AI/ML model improves anomaly detection by 30%, achieving >95% accuracy with faster detection and higher precision, recall, and F1-score, while adapting to evolving threats.	The model requires large labeled datasets, ongoing updates for new threats, significant computational resources, and its AI integration adds complexity and scalability challenges.	[18]
P. Sherubha(2023)	The model uses autoencoder-based feature selection and a Naïve Bayes classifier on the NSL-KDD dataset to enhance NIDS by removing	The model achieves 93% accuracy, J48, Random Forest, and SVM, a 0.3 FAR and 0.99 TPR, indicating strong anomaly detection.	Limitations include reliance on a static dataset and lack of zero-day attack detection.	[19]

	redundant and noisy samples.			
Jibu K Samuel(2023)	IBMD uses deep learning and neural networks to analyze cloud resource behavior for malware detection, with clients submitting file samples that are behaviorally analyzed and classified in the cloud.	The system analyzes system calls, resources, and file types to detect malware, enhancing cloud security and integrating with existing tools to prevent spread.	The system may struggle in data-scarce or inconsistent cloud environments and in real-time detection of sophisticated or zero-day malware.	[20]

III. RESEARCH METHODOLOGY

In this project, we use a cleaned network traffic datasets (CSE-CIC-IDS2018) to build a deep learning system that can effectively detect anomalies and malicious traffic in cloud-based contexts. To guarantee data integrity, the first step of preprocessing is to remove columns that aren't important and deal with missing or infinite values. A process called feature normalisation is employed to ensure that all variables are scaled consistently. Applying SMOTE (Synthetic Minority Over-sampling Technique) helps alleviate class imbalance and improve model generalisation. This leads to a more equal representation of attack categories. In order to optimise features, we reduce dimensionality and improve model efficiency by selecting the top 30 most important features using a mixture of Random Forest or Mutual Information approaches. After that, we compare the results from the whole set of characteristics to those from the features that were chosen. In order to train a 1D Convolutional Neural Network (CNN), the dataset must first be cleaned and divided into a training set and a testing set. Conv1D, MaxPooling, BatchNormalization, and Dense layers are all part of the CNN model that helps it identify network data effectively. The model's sparse categorical crossentropy loss function training and Adam optimiser compilation yield good detection accuracy, proving the model's viability for immediate security in dynamic cloud infrastructures.



Fig.3 Flow chart of Methodology

A. Data collection:

The CSE-CIC-IDS2018 dataset, created by the Communications Security Establishment (CSE) and the Canadian Institute for Cybersecurity (CIC), mimics actual network traffic that includes both benign and malicious actions, including brute-force, botnet, and denial-of-service (DDoS) attacks. Over a span of ten days in a simulated corporate environment with thirty servers and one thousand and twenty machines, realistic threats were recorded. Every record relating to the managed network flows produced 80 characteristics using CICFlowMeter-V3. Comprising over 16 million tagged samples, the dataset is often used in studies on intrusion detection systems. Anyone may access it at the URL below: <https://www.unb.ca/cic/datasets/ids-2018.html>.

B. Preprocessing technique

The data cleaning and preparation pipeline comprised many vital steps that helped to prepare the dataset for deep learning. Unneeded columns like 'Unnamed: 0' were first deleted in an attempt to shorten the dataset. Usually from CSV imports, these areas were superfluous indexes. We then looked over the dataset for any erroneous entries including infinite values—both positive and negative. To guarantee the data was intact, we removed these and any rows with missing values as well, replacing them with NaN. The dataset was split into two sections: features (X) and labels (y) after cleaning. Labels were from the target classes and features came from all dataset columns save 'Label.' They were altered from a range of [1-5] to [0-4] so they could be utilised with a multi-class classification model depending on softmax. Data normalisation was done using the StandardScaler, a key step in improving training stability and convergence in neural networks. This meant standardising every characteristic to have unit variance and zero mean. Preprocessing also covered handling class imbalance by employing the Synthetic Minority Over-sampling Technique (SMOTE) to generate new minority class instances artificially, hence ensuring balanced representation without straightforward duplication. To ensure it could be repeated, we employed a set random seed (random_state=42). Features have to be selected next to reduce the dimensionality and avoid overfitting. A Random Forest Classifier was trained with one hundred estimators to start, therefore enabling feature importance assessment. The top 30 most informative features were preserved using the SelectKBest method. This guaranteed that only very relevant traits were

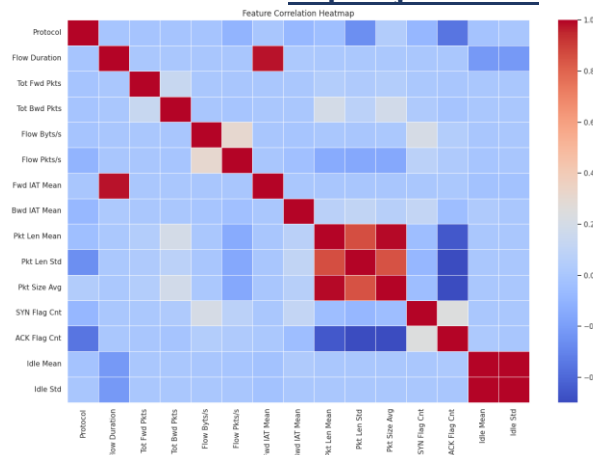


Fig.5 Heat Map Correlation Matrix

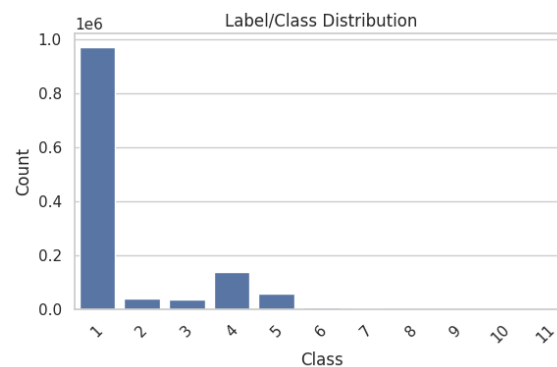


Fig.6 Class Disribution

D. Proposed Deep learning model

With an input shape of (30, 1), this 1D Convolutional Neural Network (CNN) architecture is built for multi-class classification of sequential or time-series data. The model starts with a 3x3 convolutional layer that uses 'same' padding and ReLU activation to retain input dimensions and capture local patterns. The layer consists of 64 filters. Following this, a 2D MaxPooling layer is used to reduce computational burden and dimensionality by picking the maximum values from each window. Then, a BatchNormalization layer is applied to standardise activations, speed up convergence, and make training more stable. A second MaxPooling1D layer reinforces the first convolutional layer by highlighting the most important features and reducing the likelihood of overfitting; this layer uses 128 filters and ReLU activation to extract more complicated and refined features. To prepare the 2D output for fully connected layers, a Flatten layer transforms it into a 1D vector after feature extraction and refinement. In order to avoid overfitting, a Dropout layer is used during training to randomly deactivate 30% of neurones. High-level abstract qualities are captured by the Dense layer, which has 64 units and runs ReLU activation. The last output layer is made up of neurones equal to the total amount of target classes, calculated using `len(np.unique(y_train))`, softmax activation producing class probabilities. Famous for its efficient and configurable learning rate, the Adam optimiser is utilised to construct the model. Used as an evaluation tool, the sparse categorised crossentropy loss function is suitable for multi-class classification tasks using integer labels. Due to the 100-epoch training duration, a 32-batch-size-trained network can acquire deep feature

representations and extend well to novel, unseen data. Useful in sectors handling sequential data, especially cybersecurity and sensor analysis, this approach finds a solid balance between rapidity, flexibility, and generalisability.

TABLE.2 HYPERPARAMETER TABLE

Parameter	Value
Input Shape	(X_train.shape[1], 1)
Conv1D Layer 1	64 filters, Kernel=3, ReLU
MaxPooling1D 1	Pool Size=2
Batch Norm	Yes
Dropout 1	0.3
Conv1D Layer 2	128 filters, Kernel=3, ReLU
MaxPooling1D 2	Pool Size=2
Flatten	Yes
Dense Layer 1	64 units, ReLU
Dropout 2	0.3
Dense Layer 2	len(np.unique(y_train)) units, Softmax
Optimizer	Adam
Loss	Sparse Categorical Crossentropy
Metrics	Accuracy
Epochs	100
Batch Size	32

IV. RESULT AND DISCUSSION

The proposed 1D neural networks based on con performance in detecting suspicious and hazardous traffic in cloud computing environments was evaluated using the CSE-CIC-IDS2018 dataset. Reflecting real network conditions and attack circumstances, this dataset comprises both harmful and benign network traffic. Key elements like timestamps, domain names, ports, and protocols are included, which makes it perfect for testing and training intrusion detection systems. Covering several attack kinds—e.g., DoS, DDoS, exploits—the dataset lets the model learn distinct attack patterns and properly distinguish between normal and harmful traffic in cloud settings.

TABLE. 3 EVALUATION METRICS

Model Name	Training Accuracy	Training Loss	Precision	F1 Score	Recall
CNN	0.9985	0.0083	0.8628	0.85781	0.85372

Often known as a Convolutional Neural Network, the performance measures of the model are summarised in this table. With a training accuracy of 0.9985 and a training loss of 0.0083, it is obvious that the model fit the training data well. Its accuracy (0.8628), recall (0.85372), and F1 score (0.85701) underline the model's fair ability in rightly classifying positive data. samples, reducing false positives, and finding pertinent positive cases. Together, these

measures show a highly-performing model with great accuracy and little loss, as well as strong precision and recall, which qualifies it for its intended categorisation duties.

TABLE. 4 MODEL EVALUATION METRICS

Model Name	Testing Accuracy	Validation Accuracy	Validation Loss
CNN	0.9976	0.9977	0.0179

The CNN model, which also obtains a high training precision (0.9985) with a low loss (0.0083), shows regular play in testing (exactness: 0.9976) or validation (accuracy: 0.9977, loss: 0.0179). Among the balanced classification metrics shown are recall (0.85372), the F1 rating (0.85781), and precision (0.8628).

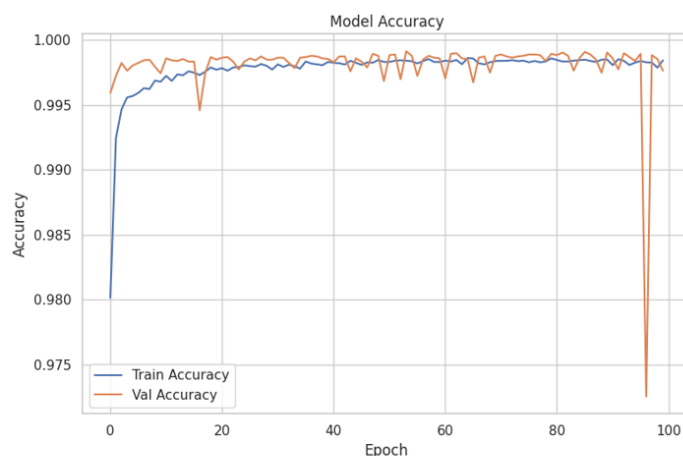


Fig.7 Accuracy Curve

This curve reveals the training or testing precision after 100 iterations. While the validation accuracy trails closely behind with little changes, the training accuracy increases and is very stable, indicating that the generalising is consistent.

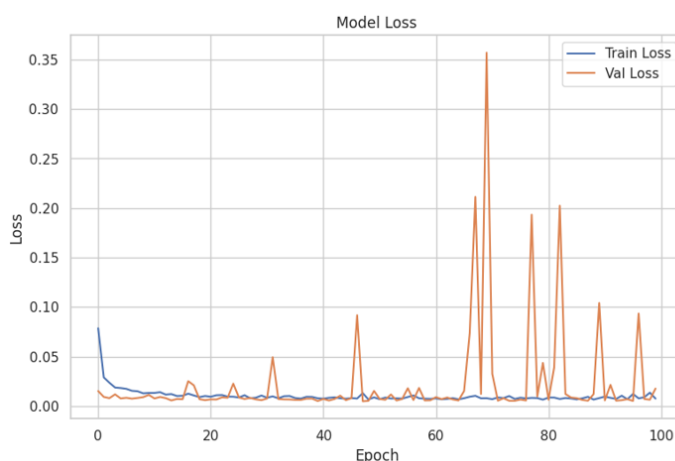


Fig.8 Loss curve

This graph depicts the loss across 100 iterations—including training and validation. Though the validation loss fluctuates significantly, the training loss decreases and remains modest, suggesting possible overfitting.

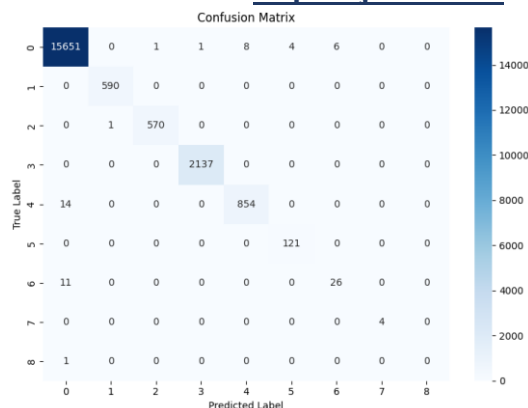


Fig.9 Confusion Matrix

This confusion matrix represents the classification performance of a model across 10 classes. The diagonal cells show the correct predictions, with higher values indicating stronger accuracy for specific labels. For example, the model excels in correctly predicting label 0, as evidenced by the highest count (13651) in the matrix. Off-diagonal cells indicate misclassifications, which appear to be minimal, suggesting the model performs well overall with high precision and recall for most classes. Visually highlighting the spread and strength of forecasts is the colour gradient from light darker blue. Its superior precision and low loss metrics show that the CNN model excels in classifying several classes job. The model performs well in terms of precision, scoring F1, or recall validating its durability across many assessment criteria.

V. CONCLUSION

This study trained a 1D a CNN using the CSE-CIC-IDS2018 dataset to help find questionable or harmful traffic in cloud computing environments. The dataset was beneficial for model training and evaluation as it precisely represented network activity and contained both benign and harmful traffic. The model excelled in training, showing its data-learning ability with a training accuracy of 99.85%. The model obviously showed a remarkable ability to generalise to new data throughout evaluation with a validation success rate of 99.77% and a testing accuracy of 99.76%. The model's Equal performance in recall, precision, and F1 score makes it a great candidate for spotting benign and harmful network traffic.

Important preprocessing techniques like data cleaning, normalisation, and SMOTE class imbalance correction helped to further improve the model's great accuracy and robustness. Feature selection techniques deleted superfluous data to further hone the dataset, therefore improving the model's performance and lowering computing complexity. The proposed CNN model, which offers an extendable foundation for real-time detection systems for intrusions, helps to find anomalies and harmful traffic in cloud computing environments. Its low loss and high accuracy allow it to be applied in dynamic cloud environments to strengthen cloud-based systems as they evolve.

REFERENCES

- [1] S. Sreenivasa Chakravarthi, R. Jagadeesh Kannan, V. Anantha Natarajan, and X. Z. Gao, "Deep Learning Based Intrusion Detection in Cloud Services for Resilience Management," *Comput. Mater. Contin.*, vol. 71, no. 2, pp. 5117–5133, 2022, doi: 10.32604/cmc.2022.022351.

- [2] L. Lahesoo, U. Do, R. M. Carnier, and K. Fukuda, *SIURU: A Framework for Machine Learning Based Anomaly Detection in IoT Network Traffic*, vol. 1, no. 1. Association for Computing Machinery, 2023. doi: 10.1145/3630590.3630601.
- [3] G. Padmavathi, D. Shanmugapriya, and S. Asha, "A Framework to Detect the Malicious Insider Threat in Cloud Environment using Supervised Learning Methods," *Proc. 2022 9th Int. Conf. Comput. Sustain. Glob. Dev. INDIACom 2022*, no. March, pp. 354–358, 2022, doi: 10.23919/INDIACom54597.2022.9763205.
- [4] W. H. Aljuaid and S. S. Alshamrani, "A Deep Learning Approach for Intrusion Detection Systems in Cloud Computing Environments," *Appl. Sci.*, vol. 14, no. 13, 2024, doi: 10.3390/app14135381.
- [5] T. Vaiyapuri and A. Binbusayyis, "Deep self-taught learning framework for intrusion detection in cloud computing environment," *IAES Int. J. Artif. Intell.*, vol. 13, no. 1, pp. 747–755, 2024, doi: 10.11591/ijai.v13.i1.pp747-755.
- [6] K. Mitropoulou, P. Kokkinos, P. Soumplis, and E. Varvarigos, "Anomaly Detection in Cloud Computing using Knowledge Graph Embedding and Machine Learning Mechanisms," *J. Grid Comput.*, vol. 22, no. 1, 2024, doi: 10.1007/s10723-023-09727-1.
- [7] H. Attou *et al.*, "Towards an Intelligent Intrusion Detection System to Detect Malicious Activities in Cloud Computing," *Appl. Sci.*, vol. 13, no. 17, 2023, doi: 10.3390/app13179588.
- [8] M. Waqas *et al.*, "Botnet attack detection in Internet of Things devices over cloud environment via machine learning," *Concurr. Comput. Pract. Exp.*, vol. 34, no. 4, 2022, doi: 10.1002/cpe.6662.
- [9] S. M. T. Nizamudeen, "Intelligent intrusion detection framework for multi-clouds – IoT environment using swarm-based deep learning classifier," *J. Cloud Comput.*, vol. 12, no. 1, 2023, doi: 10.1186/s13677-023-00509-4.
- [10] A. V. Songa and G. R. Karri, "An integrated SDN framework for early detection of DDoS attacks in cloud computing," *J. Cloud Comput.*, vol. 13, no. 1, 2024, doi: 10.1186/s13677-024-00625-9.
- [11] A. Sharma and U. K. Singh, "Modelling of smart risk assessment approach for cloud computing environment using AI & supervised machine learning algorithms," *Glob. Transitions Proc.*, vol. 3, no. 1, pp. 243–250, 2022, doi: 10.1016/j.gltp.2022.03.030.
- [12] R. Priyadarshini and R. K. Barik, "A deep learning based intelligent framework to mitigate DDoS attack in fog environment," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 3, pp. 825–831, 2022, doi: 10.1016/j.jksuci.2019.04.010.
- [13] L. Hasimi, D. Zavantis, E. Shakshuki, and A. Yasar, "Cloud Computing Security and Deep Learning: An ANN approach," *Procedia Comput. Sci.*, vol. 231, no. 2023, pp. 40–47, 2024, doi: 10.1016/j.procs.2023.12.155.
- [14] A. D. Vibhute and V. Nakum, "Deep learning-based network anomaly detection and classification in an imbalanced cloud environment," *Procedia Comput. Sci.*, vol. 232, no. 2023, pp. 1636–1645, 2024, doi: 10.1016/j.procs.2024.01.161.
- [15] P. B. K. Chowdary, R. Udayakumar, C. Jadhav, B. Mohanraj, and V. R. Vimal, "An

- Efficient Intrusion Detection Solution for Cloud Computing Environments Using Integrated Machine Learning Methodologies,” *J. Wirel. Mob. Networks, Ubiquitous Comput. Dependable Appl.*, vol. 15, no. 2, pp. 14–26, 2024, doi: 10.58346/jowua.2024.i2.002.
- [16] A. Sharon, P. Mohanraj, T. E. Abraham, B. Sundan, and A. Thangasamy, “An Intelligent Intrusion Detection System Using Hybrid Deep Learning Approaches in Cloud Environment,” *IFIP Adv. Inf. Commun. Technol.*, vol. 651 IFIP, pp. 281–298, 2022, doi: 10.1007/978-3-031-11633-9_20.
- [17] P. Parameswarappa, T. Shah, and G. R. Lanke, “A Machine Learning-Based Approach for Anomaly Detection for Secure Cloud Computing Environments,” *IDCIoT 2023 - Int. Conf. Intell. Data Commun. Technol. Internet Things, Proc.*, no. March, pp. 931–940, 2023, doi: 10.1109/IDCIoT56793.2023.10053518.
- [18] T. J. Akinbolaji, “Advanced Integration of Artificial Intelligence and Machine Learning for Real-Time Threat Detection in Cloud Computing Environments,” vol. 6, no. 10, pp. 980–991, 2023.
- [19] P. Sherubha *et al.*, “An Efficient Unsupervised Learning Approach for Detecting Anomaly in Cloud,” *Comput. Syst. Sci. Eng.*, vol. 45, no. 1, pp. 149–166, 2023, doi: 10.32604/csse.2023.024424.
- [20] J. K. Samuel, M. T. Jacob, M. Roy, P. M. Sayoojya, and A. R. Joy, “Intelligent Malware Detection System Based on Behavior Analysis in Cloud Computing Environment,” *Proc. Int. Conf. Circuit Power Comput. Technol. ICCPCT 2023*, no. August 2023, pp. 109–113, 2023, doi: 10.1109/ICCPCT58313.2023.10245065.