JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 138 - 145
https://publishoa.com
ISSN: 1309-3452

Home Net Shield

[1]Chempavathy B, [2]B Meenakshi Sundaram, [3]Ankit Datta, [4]Abhay Thoppal Shiva, [5]Gurpreet Singh

[1], [2], [3], [4], [5]New Horizon College of Engineering
[1]chempa.balaji@gmail.com,[2]bmsundaram@gmail.com, [3]ankitdatta46@gmail.com, [4]abhayts14@gmail.com, [5]gurpreet.1nh18cs721.cse@gmail.com

Ms. Chempavathy B
M.E in Computer Science and Engineering.
Professor, Department of Computer Science and Engineering,
New Horizon College of Engineering

Dr. B Meenakshi Sundaram
Professor, Department of Computer Science and Engineering,
New Horizon College of Engineering

Mr. Ankit Datta
Student, Department of Computer Science and Engineering,
New Horizon College of Engineering

Mr. Abhay Thoppal Shiva
Student, Department of Computer Science and Engineering,
New Horizon College of Engineering

Mr. Gurpreet Singh
Student, Department of Computer Science and Engineering,
New Horizon College of Engineering

ABSTRACT

The aim of our project is to detect intrusions in an IoT network by using deep learning algorithms to analyze various aspects of network traffic, such as packet flow rate, flow duration, packet volume, and so on. The CICIDS2018 dataset will be used since it contains attributes that are relevant to an IoT environment. The goal is to employ a hybrid network to analyze and classify a tuple as benign or malicious, such as a Botnet or DDoS attack. All the rows with NaN data must be removed from the data set. The types of attacks are classified using a single hot encoding. The Cu-DNNGRU and Cu-DNNBLSTM neural networks combine to form the hybrid network. Both networks' Cuda variants are employed because they have been demonstrated to be nearly 5x quicker than non-cuda variants. Both networks assist in classifying data points in a dataset as benign or malicious, and subsequently assigning a Botnet or DDoS class label. The goal is to catch such invasions in the act.

Index Terms—Botnet, CICIDS2018, DDoS, IoT.

I. INTRODUCTION

The term IoT(Internet of Things) was independently created by Kevin Ashton of Procter & Gamble(later MIT's Auto-ID Center) in 1999, however he prefers the phrase "Internet for things." He saw radio-frequency identification (RFID) as critical to the Internet of Things at the time, as it would allow computers to manage all individual things. The Internet of Things' fundamental idea is to implant short-range mobile transceivers in a variety of gadgets and everyday requirements in order to enable new types of communication between people and things, as well as between things themselves. Cisco Systems estimated that the Internet of Things was "born" between 2008 and 2009, defining it as "essentially the point in time when more 'things or items' were linked to the Internet than people.".

The Internet of Things (IoT) is a new technological realm that connects billions of things. Despite its many advantages, the diverse nature of the devices and their broad connectivity render them vulnerable to various cyberattacks that result in data breaches and financial loss. It is critical to protect the IoT environment from such threats. The goal of this project is to use Deep Learning Algorithms and Methodologies to secure IoT networks that are vulnerable to malicious assaults by identifying such attacks and quarantining malicious devices in the network.

II. STEPS

A. Data collection

The CICIDS2018 dataset employs the concept of profiles to create datasets in a systematic manner, which also include thorough descriptions of any intrusion as well as an abstract distribution model for apps, protocols, and even lower-level network elements. Brute force, Heartbleed, Botnet, DoS, DDoS, Web assaults, and Infiltration are among the seven attack scenarios included. The assault infrastructure consists of 50 machines, while the target organization consists of 5 departments and 420 machines with 30 servers.

Profiles are classified into

1) B-profiles

They contain the entity behaviors of users who utilize K-means, random forest, J48,SVM and other ML and statistical analysis approaches.

2) M-profiles

These attempt to describe an attack scenario in an unambiguous manner. Attacks are classified as

- DoS attack,
- DDoS + Port scan,
- Brute force attack,
- DDoS attack,
- Infiltration attack,
- Web attack,
- Botnet attack

Table 1. Dataset description

Classes	Attack	Instances
Benign	-	69,654
Bot	-	2977
Brute force	FTP	3066
DDoS	Loic-UDP	3015
	Hoic	3037
Infiltration	-	3043
Total		84,702

B. Data preparation

After collecting the data, it is prepared by putting together all the data collected and randomized to ensure that the data is evenly distributed, and the learning process is not affected by a specific order.

All the unwanted data, missing data, repeating data, data type conversion, etc. is solved by cleaning the data.

To know how our data is structured and the relationship between variables and classes, we visualize the data.

Then we split our dataset into 2,

a. Training set – to train the model.

b. Testing set – to check the accuracy of the model after it is trained.

C. Choose a model

D. Train the model

In this phase, we pass the prepared data to the model in order to find patterns

E. Evaluate the model

F. Parameter tuning

G. Make predictions

Finally, we use our model on unseen data to accurately make predictions about the results.

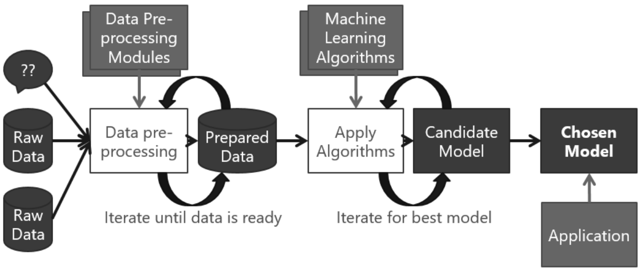

Fig. 1 Machine learning process

III. Goals and Objectives

A. To detect intrusions in IoT networks using deep learning methodologies.

B. To implement Software-Defined-Networks and incorporate Deep-Learning models in the control plane of the networks.

C. To build a robust deep neural network which is capable of classifying threats from the live data

D. Evaluate the model built to ensure higher accuracy and computational complexity.

E. Social outcomes

Almost everything in today’s world is “smart” and is connected to a network in some or the other manner, with advancement in technology comes vulnerability to attacks, hackers can get unauthorized access to confidential and sensitive data and/or incur monetary loss. According to a report, almost about 1.5 billion IoT breaches were recorded in the first 3 months of 2021. This project is aimed to detect such vulnerabilities in IoT network and ensure user safety to an extent.

IV. Requirements

A. Hardware requirements

Table 2. Hardware requirements

CPU	Intel core i5, >8th generation with ~2GHz processor
RAM	8GB
GPU	>=Nvidia GTX 970
Hard disk space	100GB

B. Software requirements

Table 3. Software requirements

OS	Windows 10 64 bit
Language	Python 3.8
Libraries	Pandas, TensorFlow, Numpy, Scikitlearn, Keras

V. Methodology

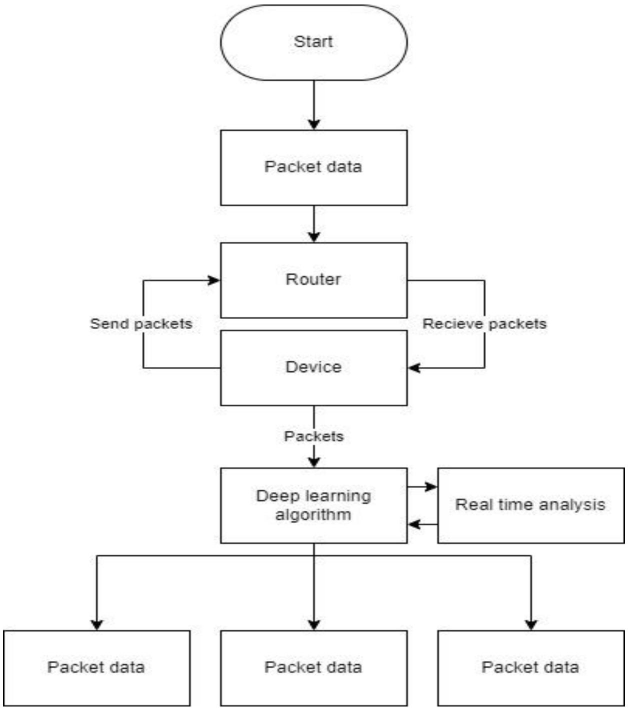

Fig. 2 Methodology

VI. IMPLEMENTATION

A. Deep Learning

Computer algorithms represent the core of machine learning. In the context of deep learning, machine learning can be considered as a subset of that field. Deep learning employs artificial neural networks, which are designed to imitate the human beings and how they learn. Machine learning employs simpler concepts. Neural networks were previously limited in complexity due to computational power constraints. Big Data analytics advances have enabled larger, more powerful neural networks, allowing computers to monitor, understand, and react to complex events quicker than humans. It has aided image categorization, language translation, and speech recognition. It can solve any pattern recognition problem without requiring human intervention.

Deep learning is powered by artificial neural networks with several layers. Deep Neural Networks (DNNs) are networks that make sense of images, music, and text by performing complicated operations such as representation and abstraction at each layer. Deep learning is growing at a rapid pace in the field of machine learning. It is a an established technology that is being adopted by huge number of businesses for developing new business models.

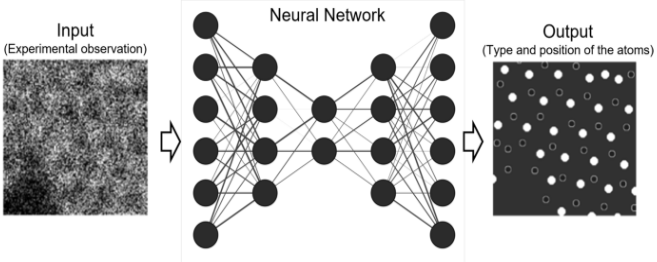

Fig. 3 Deep learning architecure

B. Naïve Bayes

Naïve Bayes is a wonderful example that justifies that anything simple is perhaps stronger. Even though there has been a progress in Machine Learning, it has still shown that it is uncomplicated, swift, precise, and reliable. This is widely used in many different domains, but it is excellent at solving natural language processing complications.

It is The Bayes Theorem that is considered as the framework for the Naïve Bayes algorithm, which has variety of applications, which is utilized in a wide range of classification problems. In this essay, we will learn about the Nave Bayes algorithm and all the key ideas so that there are no questions. Several implementations of the practical nature estimate parameters through the approach of maximum likelihood in these models i.e., the model can be utilized without implementing Bayesian Probability/Methods.

For estimation of conditional probabilities, the Bayes' Theorem has a very basic and easy formula. Conditional probability refers to the possibility of an event occurring if another event has already occurred (by an assumption, inference, statement, or evidence).

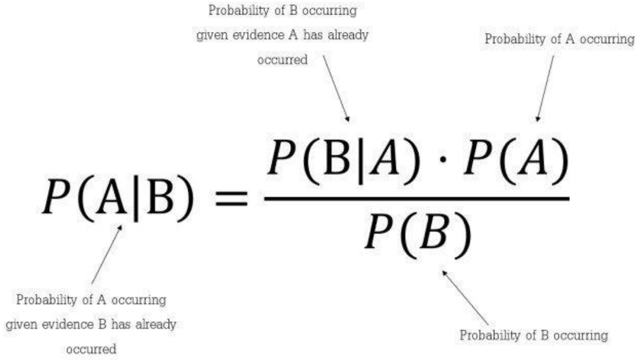
Probability of B occurring given evidence A has already occurred
Probability of A occurring
$P(\mathrm{A}|\mathrm{B}) = \frac{P(\mathrm{B}|A) \cdot P(A)}{P(B)}$
Probability of A occurring given evidence B has already occurred
Probability of B occurring

Fig. 4 Naïve bayes formula

C. Honeypots

A honeypot is a system that is attached to a network that attracts cyber attackers and detects, diverts, and investigates attempts made at hacking to obtain access that is not legal. It's major idea is to display itself on the internet/intranet as a target for hackers/ attackers, gather data, and inform users to any illegal attempts.

A honeypot operation, in general, comprises of a computer, programs, and data that simulate the behavior of a real person. such as a financial system, internet of things (IoT) devices, a public utility, or another system that could be alluring to attackers

Honeypots are widely used in the demilitarized zone of a network (DMZ). This method keeps it linked while separating it from the main production network. While attackers access a honeypot in the DMZ, it may be monitored from afar, lowering the risk of the main network being infiltrated. They can be put outside the external firewall, facing the internet, to detect efforts to enter the internal network. The exact position of the honeypot is determined by its complexity, the type of traffic it seeks, and its proximity to vital business resources. Regardless of where it is placed, it will always be isolated from the production environment.

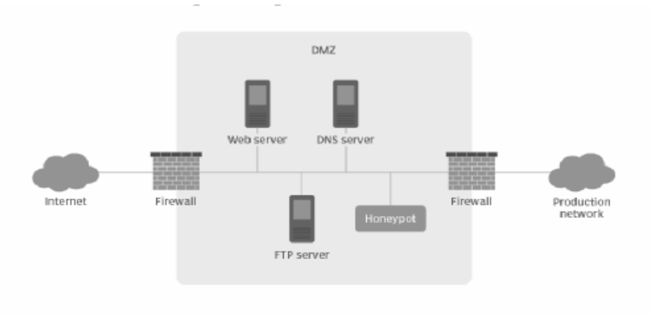

Fig. 5 Honeypots in a network

D. Software Defined Networks

SDNs (Software Defined Network) are a networking technique which communicates with the hardware infrastructure underlying and re-directs the traffic on a network using a software based controller or APIs (Application Programming Interface)

On the other hand, traditional networks require hardware devices that are specialized. These devices include but not limited to routers, switches, in order to govern the underlying network traffic. SDNs can make use of software to achieve access and control over a virtual network.

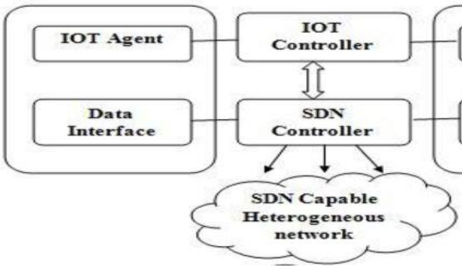

Fig. 6 SDN architecture

E. LSTM

Long Short Trem memory Network, or LSTM, is a variation of Deep Neural Networks. It comes under the category of recurrent neural networks (RNNs) that have the capability to learn long period sequences, which is useful for solving sequential context issues. Apart from singular data points such as pictures, LSTM has feedback connections, that means it can run through the complete context of sequence. This is beneficial in speech recognition, machine language synthesis, and other areas. The LSTM is a subtype of RNN that performs extremely well on a wide range of problems. An LSTM model's primary function is carried by a memory unit called a 'cell state,' which maintains its context across time. The horizontal line that runs through the top of the diagram below represents this unit.

LSTM can be related to a conveoyr belt ontop of which we just pass data, without any modifications. In an LSTM, information can be put in to or taken out from the cell state, which is controlled by gates. These gates control data to move into and out of the cell if desired. The method is helped by a pointwise multiplication method and a sigmoid activated neural net layer. The sigmoid layer outputs integers between 0 and 1, where 0 indicates "nothing be let through" and 1 indicates "everything be let through."

F. F Score

The F score or popularly known as F1 score, used to measure the accuracy of a model on some dataset. It's used to evaluate binary classification algorithms that categorize examples into 2:
a. positive groups and
b. negative groups.

The F-score is a way for integrating the precision and recall of a model. It is defined as the harmonic mean of the model's precision and recall.

The F-score is a popular metric for assessing information retrieval systems like search engines, as well as a variety of machine learning models, particularly in natural language processing. It's possible to tweak the F score so that precision takes precedence over recall, or recall over precision. The F0.5 and F2 scores and the traditional F1 score are commonly adjusted to F scores.

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \qquad (1)$$

$$precision = \frac{TP}{TP + FP} \qquad (2)$$

$$recall = \frac{TP}{TP + FN} \qquad (3)$$

Fig. 6 F-score calculation

G. Standard Evaluation Metrics

Accuracy, F-score, recall, precision, etc. are used as the evaluation metrics for this project

$$Accuracy = \frac{Tpos + Tneg}{Tpos + Tneg + Fpos + Fneg} \qquad (1)$$

$$Recall = \frac{Tpos}{Tpos + Fneg} \qquad (2)$$

$$Precision = \frac{Tpos}{Tpos + Fpos} \qquad (3)$$

$$F1 - score = \frac{2 * Tpos}{2 * Tpos + Fpos + Fneg} \qquad (4)$$

Fig. 7 evaluation metrics

VII. FUTURE WORK

The proposed model can be improved in terms of efficiency and security by utilizing hybrid deep learning algorithms with SDNs and Blockchain for detection of threats and intrusions in an IoT network.

CONCLUSION

IoT with its internal complexity requires a flexible, reliable and a secure infrastructure. Deep learning has lately gained the attention of the world with regular advancements. In this project, we propose an SDN-enabled hybrid deep leaning driven architecture to protect the IoT environment against attacks like DoS, DDoS, botnets, infiltration, etc. The result of this proposed architecture is par with the Cuda-GRULSTM and Cuda-DNNLSTM, the existing hybrid algorithms.

The performance of the model is tested against the evaluation metrics – accuracy, recall, precision, F-score, and speed efficiency.

REFERENCES

[1] C. B, V. M. Deshmukh, A. Datta, A. T. Shiva and G. Singh, "An Exploration Into Secure IoT Networks Using Deep Learning Methodologies," 2022 International Conference for Advancement in Technology (ICONAT), 2022, pp. 1-4, doi: 10.1109/ICONAT53423.2022.9725988.

[2] Vairale V.S., Shukla S. (2019) Recommendation Framework for Diet and Exercise Based on Clinical Data: A Systematic Review. In: Mishra D., Yang XS., Unal A. (eds) Data Science and Big Data Analytics.

[3] Nithya B., Ilango V., Mohan Kumar S. (2019). Cryptographic system models and algorithms for network security Journal of Advanced Research in Dynamical and Control Systems.

[4] Veerasamy, V., Wahab, N. I. A., Othman, M. L., Padmanaban, S., Sekar, K., Ramachandran, R., ... & Islam, M. Z. (2021). LSTM recurrent neural network classifier for high impedance fault detection in solar PV integrated power system.

[5] Raiesh, G., Saroia, B., Dhivya, M., & Gurulakshmi, A. B. (2020, October). DB-Scan Algorithm based Colon Cancer Detection and Stratification Analysis. In 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud).

[6] Nguyen, T.D.; Marchal, S.; Miettinen, M.; Fereidooni, H.; Asokan, N.; Sadeghi, A.-R. DÏoT: A Federated Self-learning AnomalyDetection System for IoT. In Proceedings of the 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), Dallas, TX, USA, 7–10 July 2019; pp. 756–767.

[7] Godwin, J.J., Krishna, B.V.S., Rajeshwari, R., Sushmitha, P., Yamini, M. (2021). IoT Based Intelligent Ambulance Monitoring and Traffic Control System. In: Balas, V.E., Solanki, V.K., Kumar, R. (eds) Further Advances in Internet of Things in Biomedical and Cyber Physical Systems. Intelligent Systems Reference Library, vol 193. Springer, Cham. https://doi.org/10.1007/978-3-030-57835-0_20

[8] M. S. Shoba, S. D, K. L. Suchala, R. H. Shravya and B. S. Soundhaaryha, "Survey on IoT based E-Farming Technology Enabled Farming," 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), 2022, pp. 989-995, doi: 10.1109/ICSCDS53736.2022.9760870.

[9] S. S. Priya, R. P, M. B. M, S. Aramoti and S. Fathima, "Home Automation by Speech Detection System using Deep Learning," 2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), 2022, pp. 1-5, doi: 10.1109/ICSTSN53084.2022.9761303.

[10] V. Adaickalam, S. P. Manikandan, N. Kanagavalli and P. S. Dinesh, "A Vibrant Multiple Object Detection Using Machine Learning Techniques," 2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), 2022, pp. 1-5, doi: 10.1109/ICSTSN53084.2022.9761357.

[11] V. Cp, S. Kalaivanan, R. Karthik and A. Sanjana, "Blockchain-based IoT Device Security," 2022 2nd International Conference on Artificial Intelligence and Signal Processing (AISP), 2022, pp. 1-6, doi: 10.1109/AISP53593.2022.9760674.

[12] J. Karthiyayini, A. C. Vantipalli, D. S. Tanti, K. Malvika Ravi and K. Kannan, "IOT based AquaSwach," 2022 2nd International Conference on Artificial Intelligence and Signal Processing (AISP), 2022, pp. 1-12, doi: 10.1109/AISP53593.2022.9760657.

[13] A. P. Nirmala, V. Asha, P. Chandra, H. Priya and S. Raj, "IoT based Secure Smart Home Automation System," 2022 IEEE Delhi Section Conference (DELCON), 2022, pp. 1-7, doi: 10.1109/DELCON54057.2022.9753086.

[14] S. K. B V, S. Sharma, K. S. Swathi, K. R. Yamini, C. P. Kiran and K. Chandrika, "Review on IoT based Healthcare systems," 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA), 2022, pp. 1-5, doi: 10.1109/ICACTA54488.2022.9753547.

[15] D. K. S. Nadiger, B. V. Santhosh Krishna, P. Piruthiviraj, K. Vinay Kumar, B. N. Hirebidari and Vignesh, "IoT Based Alive Human Detection in War Field and Calamity Area Using Microcontroller," 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), 2022, pp. 1233-1238, doi: 10.1109/ICAIS53314.2022.9742741.

[16] R. Thirukkumaran, B. Rajalakshmi, A. Priyedarshni, K. Abhigna and A. Kumar, "Soil and Crop Health Analysis Using IoT and ML," 2022 International Conference for Advancement in Technology (ICONAT), 2022, pp. 1-4, doi: 10.1109/ICONAT53423.2022.9726043.

[17] V. M. Deshmukh, R. B, G. B. Krishna and G. Rudrawar, "An Overview of Deep Learning Techniques for Autonomous Driving Vehicles," 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), 2022, pp. 979-983, doi: 10.1109/ICSSIT53264.2022.9716433.

[18] D. K. S. Nadiger, J. Dhanush, R. Vikas, S. K. B V, A. R. Naik and C. G. M, "E-Health Tracker: An IoT-Cloud Based Health Monitoring System," 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), 2022, pp. 35-39, doi: 10.1109/ICSSIT53264.2022.9716540.

[19] M. Swarnamugi and R. Chinnaiyan, "IoT Hybrid Computing Model for Intelligent Transportation System (ITS)," 2018 Second International Conference on Computing Methodologies and Communication (ICCMC), 2018, pp. 802-806, doi: 10.1109/ICCMC.2018.8487843.

[20] P. Manojkumar, M. Suresh, Alim Al Ayub Ahmed, Hitesh Panchal, Christopher Asir Rajan, A. Dheepanchakkravarthy, A. Geetha, B. Gunapriya, Suman Mann & Kishor Kumar Sadasivuni (2021) A novel home automation distributed server management system using Internet of Things, International Journal of Ambient Energy, DOI: 10.1080/01430750.2021.1953590

[21] C. Kavyashree, H. K. Sowmya and N. V. U. Reddy, "A Survey on the Cervical Cancer Detection using Deep Learning methods," 2021 International Conference on Forensics, Analytics, Big Data, Security (FABS), 2021, pp. 1-6, doi: 10.1109/FABS52071.2021.9702701.

[22] Kumar K, Vinoth. "Network Intrusion Detection System in Latest DFA Compression Methods for Deep Packet Scruting." Design, Applications, and Maintenance of Cyber-Physical Systems, edited by Pierluigi Rea, et al., IGI Global, 2021, pp. 219-243. https://doi.org/10.4018/978-1-7998-6721-0.ch010

[23] Karuppusamy, L, Ravi, J, Dabbu, M, Lakshmanan, S. Chronological salp swarm algorithm based deep belief network for intrusion detection in cloud using fuzzy entropy. Int J Numer Model. 2022; 35(1):e2948. doi:10.1002/jnm.2948

[24] D. Kalaivani, L. Srinivasan and K. Saravanan, "Using multipath TCP and opportunistic routing in IoT network," 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), 2022, pp. 94-104, doi: 10.1109/ICSSIT53264.2022.9716336.

[25] R. Bhargava and J. Dinesh, "Deep Learning based System Design for Diabetes Prediction," 2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), 2021, pp. 1-5, doi: 10.1109/SMARTGENCON51891.2021.9645906.

[26] Vinodha, K., Vaishali M. Deshmukh, and Subhashree Rath. "Secured Online Learning in COVID-19 Pandemic using Deep Learning Methods." 2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC). IEEE, 2021.