International Journal of Scientific Research & Engineering Trends
Volume 8, Issue 5, Sep-Oct-2022, ISSN (Online): 2395-566X

System for Detecting Intruders Using Convolutional Neural Networks

M. Tech. Scholar Sanjay Soni, Assistant Prof. Aditi Khemariya
Department of Computer Science and Engineering,
Malwa Institute of Science & Technology,
Indore, India
Sanjaysuperfast786@gmail.com, adititiwari@mistindore.com

Abstract- On the internet, malicious activities may take place, which can infect a single computer or bring down a whole network. It's possible that they are chosen at random. The number of people connecting to the internet at an ever-increasing pace makes it more difficult to keep up. The internet, just like real life, may provide a number of potential safety hazards. The IDS software watches a network for activity that might be hostile or suspicious. IDS stands for intrusion detection system, and it is a kind of technology that helps detect attacks on computer systems and determine who carried them out. Machine learning (ML) strategies were used in the past to increase IDS accuracy and the results of intrusion detection. In this article, we will go through the process of constructing IDS by using CNN approach. This is a method that may be used when developing effective IDS. On the KDD, the technique that was proposed will be implemented (Knowledge Discovery Dataset). The accuracy of the technique that has been provided is clearly superior to that of SVM, Naive Bayes, and Decision Tree. It is possible to have faith in this because: The following outcomes were a consequence of using our method: 3.24 minutes is the amount of time required for performance, with an accuracy of 96.78% and an error rate of 0.21 percent.

Keywords – IDS, Knowledge Discovery Dataset, PCA, Random Forest.

I. INTRODUCTION

Because of the rapid acceleration of technological progress, the internet's presence in people's daily lives is rapidly expanding and becoming more pervasive. The internet has become so ubiquitous that almost everyone's life is reliant on it in some way, shape, or form. These days, having an internet connection and knowing how to use it are both necessities for everyone. As a result, protecting the system from potentially hazardous behaviour is becoming more and more necessary as the number of individuals utilising the internet for their own personal activities continues to rise.

There are many different kinds of attacks that may be observed on the system or the network. The purpose of these assaults is to either steal information from a system or modify the data that is already there on any system [1]. Attackers will employ a wide number of methods to get access to the system's data in order to misuse it. Some of these methods include denial of service (DoS), probe, sniff, r2l, and more methods that are quite similar.

As a direct result of this, an intrusion detection system was built and installed to guard the system from attacks of this kind. System intrusion detection systems, often known as IDS, are responsible for monitoring any attacks made against the system and working to protect it from any dangers.

1. Intrusion Detection System, Version:

Intrusion is a term that refers to the act of entering a system without authorisation and inflicting harm to the information that is held inside of the system [1]. Any system that has this infiltration risk having its hardware compromised as a result of the intrusion. The phrase "keeping the system from being compromised" has come to place a great deal of importance on the word "intrusion," which has developed into a very important noun. Depending on the circumstances, the intrusion detection system (IDS) may either be used to control any intrusions that take place inside a system or keep track of any intrusions that take place within the system.

In spite of the fact that several forms of intrusion detection systems have been used in the past, the accuracy of each method has been called into doubt over the course of the previous few years. In order to establish the system's level of accuracy, it is necessary to investigate two separate metrics, namely the detection rate and the false alarm rate [2]. It is essential that the system be developed in such a manner that the rate of false alarms is reduced to an absolute minimum while simultaneously increasing the detection rate. As a direct consequence of this, the IDS make use of the PCA in combination with the random forest.

The IDS may take one of two forms in nature; both of these forms are relevant to its usefulness, and they are as follows:

In this system, the network traffic is analysed, and any intrusions that occur as a consequence of the traffic are detected and examined. These actions take place in response to the monitoring of the network traffic.

Host-based intrusion detection systems, often known as HIDS, are systems that are used in the process of detecting network intrusions. These systems monitor system files that are accessible over the network.

An additional point to consider is that there is a subset of IDS types. Those variants that depend on signature identification and anomaly detection are the ones that are seen the most often.

Signature-based: In this instance, the system identified a number of specific patterns that malicious software employs in order to conceal its true nature. The patterns that have been uncovered are referred to as signatures. When it comes to detecting fresh assaults, this does not perform as well as it does when recognising already committed crimes; the signature detection procedure is where it falls short.

A kind of detection called as anomaly-based detection is one that has been developed expressly for the purpose of identifying unknown forms of attack. This system, which incorporates ML into its operation, is used to create the model.

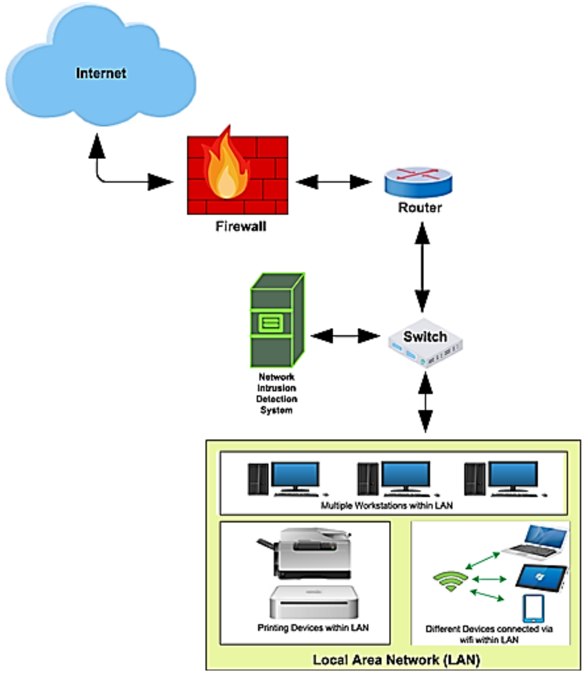
Internet
Firewall
Router
Switch
Network Intrusion Detection System
Multiple Workstations within LAN
Printing Devices within LAN
Different Devices connected via wifi within LAN
Local Area Network (LAN)

Fig 1. Intrusion Detection System.[2]

2. Random Forest:

When it comes to solving classification issues, RF is among the most effective approaches that can be found in the field of machine learning. Random forests are classified as supervised classification algorithms [2], which is the category that the random forest belongs to. This procedure is carried out in two distinct stages: the first stage is concerned with the formation of the forest using the dataset that has been provided, and the second stage is concerned with the prediction using the classifier that was generated in the very first stage.

The following is a pseudocode representation of how to generate a random forest:

- Pick a few characteristics, k, from a total of m to represent k/m.
- By applying the split point to k different characteristics, get node d
- In order to get the daughter nodes, use the optimal split.
- Continue with steps 3, 2, and 1 until we reach the first node.
- Establish a forest by repeating the procedures one through four in order to establish a forest.

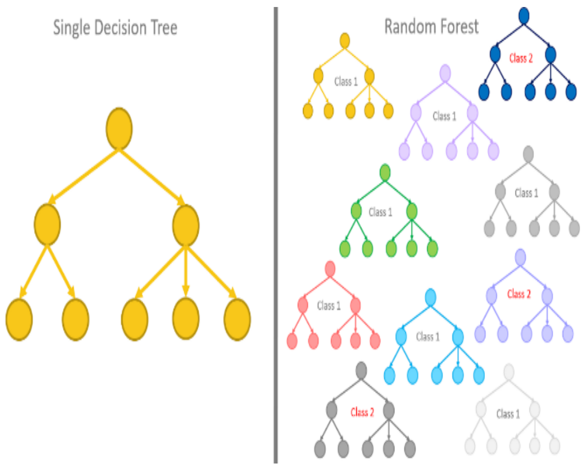

Fig 2.Random Forest Model.

3. PCA:

One of the methods that is utilised, in particular for the goal of lowering the size of the dimensions of a given dataset, is known as the principal component analysis technique. In addition to being one of the most effective and accurate methods currently available for reducing the dimensionality of data, it is also capable of producing the outcomes that are needed [3]. With the use of this method, the properties of a particular dataset may be whittled down to an acceptable amount of attributes, which are referred to as the primary components.

This method considers the whole input as a dataset, which since it has a big number of attributes also contains a large dimension. Since the dataset contains a large number of characteristics, this method has a large dimension. This strategy helps to reduce the amount of the dataset by positioning all of the data points along the same axis. Following the repositioning of the data points along one of the axes, the principal components analysis is performed.

II. LITERATURE REVIEW

The authors claim that a solution to the IDS was discovered by using the SVM and Nave Bayes algorithms, with the SVM algorithm showing to be better to both of the other approaches that were tried. Their experiment was carried out by utilising information from the KDD dataset, and they provide their results in terms of the detection rate as well as the false alarm rate. [4]

The writers of this paper carried out three independent experiments, each of which is described in extensive detail in the authors' respective paragraphs. Both in the analysis and in the design, they made use of the feature selection technique. In addition, the naive Bayes algorithm, the adaptive boost algorithm, and the partial decision tree were presented. They investigated each and every option for detecting infiltration that was accessible. [5]

The writers of this article have arrived at the conclusion, with the assistance of this article, that the approach of artificial neural networks with feature selection would produce higher results when compared to the Support vector machine method. The NSL-KDD dataset was used throughout the course of the project. It turned out that the approach that was offered was effective. [6]

The purpose of this study is to provide an overview of intrusion detection systems that employ a machine-learning algorithm, as described by the authors of this paper. On the basis of the findings that they acquired, the authors offered a performance comparison of several machine learning methods. They examined the survey's detection rates as well as its rates of false alarms so that they could evaluate it. [7]

The authors have created a method for detecting intrusions that takes use of logistic regression and belief propagation methods. This approach was developed by the writers. As was said earlier, the proposed technique has shown that it provides a quicker average detection time in comparison to the approaches that have been used in the past. [8]

The authors created a technique of in-depth learning that was used to extract features from the dataset. This approach was successful in accomplishing this task. They made an effort to extract features from a dataset in order to make the dataset more useful for utilisation, and as a result of their efforts, they arrived at the realisation that they might provide improved input to the intrusion detection system. [9]

They used a process known as machine learning to carry out the research for this segment, which focused on intrusion detection systems. In their research, the authors examined the entire machine learning algorithms that have been used up to this point. Based on their findings, they determined that the machine learning algorithms that were supplied by Md. Nasimuzzaman Chowdhury and the ANN algorithms that were provided by Alex Shenfield, Aladdin Ayesh, and David Day were the most successful. [10]

The authors of this study evaluated a variety of different machine learning strategies that may be used into an intrusion detection system. They examined a variety of approaches, some of which were the SVM, the Extreme learning machine, and the random forest, amongst others. The authors come to the conclusion that the Extreme machine learning methodology is superior to every other method by a significant margin in their conclusion. [11]

The authors of this study made an effort to improve the quality of the dataset before making it accessible for examination by the intrusion detection system. In order to improve upon the dataset, which they have previously discussed, they have implemented a fuzzy rule-based feature selection approach. They used the KDD dataset, and the findings of the IDS revealed a dynamic rise in the total number of outcomes. [12]

III. PROBLEM DOMAIN

The computer networks that are connected to the internet are vulnerable to a broad variety of malicious activities. The intrusion into the system with the intention of breaking the information has been deemed to be the most significant problem that has been seen in this sector. The construction of an intrusion detection system allows for the identification of this incursion; nevertheless, in order for this system to be useful, it must be accurate and efficient in its detection of intruders. For the purpose of intrusion detection, machine learning algorithms were used. These approaches included SVM, Naive Bayes, and other variations of the methodology.

In spite of this, the data suggest that there is perhaps some space for development in a variety of areas, including but not limited to accuracy, detection rates, and the occurrence of false alarms. It is possible that approaches that have been utilised in the past may need to be replaced with new methods. Some of these new methods are SVM and Nave Bayes. In addition, the study suggests that the dataset may be improved in some way if certain procedures were applied to it. It is required to in order to improve the overall quality of the data that is fed into the suggested system.

IV. PROPOSED SOLUTION

The purpose of the intrusion detection system is to improve the overall performance of the system, which is negatively influenced by the presence of intruders. This apparatus is able to recognise suspicious activity on the grounds and alert the appropriate personnel. The proposed method makes an effort to address the problems that have surfaced as a consequence of the work that has been done before.

One of the proposed components of the system is known as principal component analysis, while the other is known as random forest. It is recommended that both of these components be included in the system. The dimension of a dataset can be reduced in size through the application of the technique of principal component analysis; the quality of the dataset will be improved as a result of this approach due to the fact that the dataset will include the appropriate characteristics as a result of this method.

The SVM approach will not be employed for intruder detection because the random forest algorithm, which has a greater detection rate and a lower false alarm rate than SVM, will be used instead.

1. Algorithm for the proposed solution:
The coordination degree of the original attribute for the split node standard has been replaced with compatibility as the replacement attribute.

1. Attribute compatibility:
Let the modulus for the primary decision set be "Pr," and the modulus for the secondary set be "Se." The definition of attribute compatibility is as follows:

$$\mathbf{CO\ (X \rightarrow D) = Pr\ -SeX\ (1)}$$

X is the subset for non-empty C in this context. When the impact of the secondary set can be demonstrated to have over the primary set, we term this situation "strict compatibility." The primary text and the additional set both include errors or inconsistencies. The expression puts the finishing touches to the secondary set.

$$\mathbf{CO\ (X \rightarrow D) = Pr\ X\ (2)}$$

Here X is the subset for non-empty C. In this, the wide compatibility of the second set is seen.

Algorithm for the Base Classifier Improvement:
In the first step, the active attribute of the data set is given its starting value by marking every condition attribute.
The second step is to compute the modulus for each condition attribute that is included in either the main or secondary set.
Phase 3: The computation of compatibility between all conditional attributes is carried out in this step by applying equation (1). If other characteristics with a similar compatibility are observed, equation (2) should be used.
Step 4: To divide the sample, pick the node with the most comprehensive compatibility for splitting to use as the split node, and then remove the tag that is now active.
Step 5: continue picking the active attribute for splitting up to the point when we receive the active attribute when we reach the leaf node.
The sixth and last step is the generation of the basic classifier.

2. Flowchart for the Proposed Algorithm:

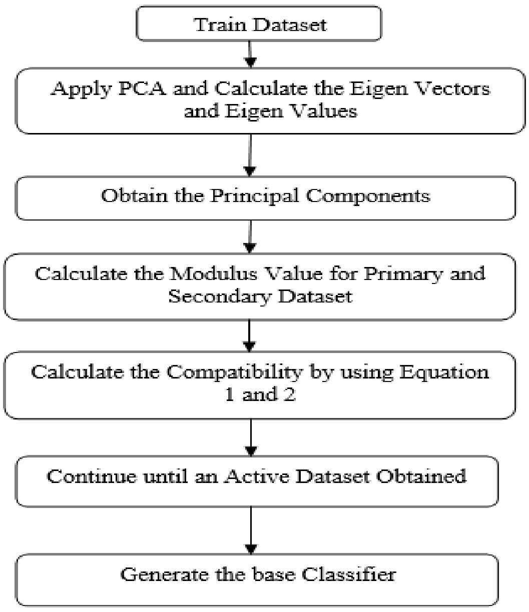
Train Dataset
Apply PCA and Calculate the Eigen Vectors and Eigen Values
Obtain the Principal Components
Calculate the Modulus Value for Primary and Secondary Dataset
Calculate the Compatibility by using Equation 1 and 2
Continue until an Active Dataset Obtained
Generate the base Classifier

Fig 3. Flowchart for The Proposed Approach.

V. RESULTS

It was feasible to acquire good results with the experiment that was carried out to test the recommended approach when the KDD dataset was used as the foundation for the experiment. The use of principal component analysis (PCA) in combination with convolutional neural networks (CNN) fared very well in comparison to other prevalent approaches such as SVM, Naive Bayes, decision trees, and CNN. In the next part, a table will be shown that compares the performance time (in minutes), accuracy rate (in percentage), and error rate (in percentage) of a number of different approaches:

Table 1. Result Comparison with other Classifiers.

Method	Performance time (min)	Accuracy rate (%)	Error rate (%)
SVM	4.57	84.34	2.67
Naïve Bayes	9.12	80.85	3.49
Decision Tree	12.36	89.91	0.78
PCA with CNN	3.42	96.78	0.21

When compared to other algorithms such as SVM, Naive Bayes, and Decision Tree, it is clear that the technique that has been provided here is superior in terms of performance. PCA combined with Random Forest performed much better based on three different factors. Its representation may be seen in table 1.

VI. CONCLUSION

In recent years, there has been an increase in the number of systems that are connected to the internet, which has been followed by an increase in the number of security concerns. The solution that was recommended deals with the detection of intruders successfully via the use of the internet, and it does so in a cost-effective manner. The approach that was recommended performed much better than other algorithms that have been utilised in the past, such as SVM, Naive Bayes, and Decision Tree.

The approach that was presented has the potential to dramatically rise, in a number of different ways, not only the detection rates but also the rates of false errors. The knowledge discovery dataset is the one that has been used for the purpose of demonstrating this example. Using the strategy that we suggested, we were able to acquire the following results: The performance time (in minutes) is 3.24, the accuracy rate (in percentage) is 96.78, and the error rate (in percentage) is 0.21. The performance time (in minutes) is 3.24, the accuracy rate (expressed as a percentage) is 96.78, and the error rate (expressed as a percentage) is 0.21.

REFERENCE

[1] Jafar Abo Nada; Mohammad Rasmi Al-Mosa, 2018 International Arab Conference on Information Technology (ACIT), A Proposed Wireless Intrusion Detection Prevention and Attack System

[2] Kinam Park; Youngrok Song; Yun-Gyung Cheong, 2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (BigDataService), Classification of Attack Types for Intrusion Detection Systems Using a Machine Learning Algorithm

[3] Le, T.-T.-H., Kang, H., & Kim, H. (2019). The Impact of PCA-Scale Improving GRU Performance for Intrusion Detection. 2019 International Conference on Platform Technology and Service (PlatCon). Doi:10.1109/platcon.2019.8668960

[4] Anish Halimaa A, Dr K.Sundarakantham: Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) 978-1-5386-9439-8/19/$31.00 ©2019 IEEE "Machine Learning Based Intrusion Detection System."

[5] Rohit Kumar Singh Gautam, Er. Amit Doegar; 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence) "An Ensemble Approach for Intrusion Detection System Using Machine Learning Algorithms."

[6] Kazi Abu Taher, Billal Mohammed Yasin Jisan, Md. Mahbubur Rahma, 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST) "Network Intrusion Detection using Supervised Machine Learning Technique with Feature Selection."

[7] L. Haripriya, M.A. Jabbar, 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)" Role of Machine Learning in Intrusion Detection System: Review"

[8] Nimmy Krishnan, A. Salim, 2018 International CET Conference on Control, Communication, and Computing (IC4) "Machine Learning-Based Intrusion Detection for Virtualized Infrastructures"

[9] Mohammed Ishaque, Ladislav Hudec, 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS) "Feature extraction using Deep Learning for Intrusion Detection System."

[10] Aditya Phadke, Mohit Kulkarni, Pranav Bhawalkar, Rashmi Bhattad, 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) "A Review of Machine Learning Methodologies for Network Intrusion Detection."

[11] Iftikhar Ahmad , Mohammad Basheri, Muhammad Javed Iqbal, Aneel Rahim, IEEE Access (Volume: 6) Page(s): 33789 – 33795 "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection."

[12] B. Riyaz, S. Ganapathy, 2018 International Conference on Recent Trends in Advanced Computing (ICRTAC)" An Intelligent Fuzzy Rule-based Feature Selection for Effective Intrusion Detection.

[13] N. Arya, U. Singh, and S. Singh, "Detecting and avoiding of worm hole attack and collaborative blackhole attack on MANET using trusted AODV routing algorithm," in IEEE International Conference on Computer Communication and Control, IC4 2015, 2016, doi: 10.1109/IC4.2015.7375649.

[14] U. Singh, M. Samvatsar, A. Sharma, and A. K. Jain, "Detection and avoidance of unified attacks on MANET using trusted secure AODV routing protocol," in 2016 Symposium on Colossal Data Analysis and Networking, CDAN 2016, 2016, doi: 10.1109/CDAN.2016.7570908.

[15] U. Singh, M. Shukla, A. K. Jain, M. Patsariya, R. Itare, and S. Yadav, Trust Based Model for Mobile Ad-Hoc Network in Internet of Things, vol. 98. 2020.

[16] M. Muwel, P. Mishra, M. Samvatsar, U. Singh, and R. Sharma, "Efficient ECGDH algorithm through protected multicast routing protocol in MANETs," in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212743.

[17] U. Singh, V. Vankhede, S. Maheshwari, D. Kumar, and N. Solanki, Review of Software Defined Networking: Applications, Challenges and Advantages, vol. 98. 2020.

[18] U. Singh, M. Samvatsar, A. Sharma, and A. K. Jain, “Detection and avoidance of unified attacks on MANET using trusted secure AODV routing protocol,” in 2016 Symposium on Colossal Data Analysis and Networking, CDAN 2016, 2016, doi: 10.1109/CDAN.2016.7570908.

[19] A. S. Chouhan, V. Sharma, U. Singh, and R. Sharma, “A modified AODV protocol to detect and prevent the wormhole usingh using hybrid technique,” in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212740.

[20] L. Baghel, P. Mishra, M. Samvatsar, and U. Singh, “Detection of black hole attack in mobile ad hoc network using adaptive approach,” in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212741.

[21] N. Arya, U. Singh, and S. Singh, “Detecting and avoiding of worm hole attack and collaborative blackhole attack on MANET using trusted AODV routing algorithm,” in IEEE International Conference on Computer Communication and Control, IC4 2015, 2016, doi: 10.1109/IC4.2015.7375649.

[22] A. Sharma, D. Bhuriya, and U. Singh, “Secure data transmission on MANET by hybrid cryptography technique,” in IEEE International Conference on Computer Communication and Control, IC4 2015, 2016, doi: 10.1109/IC4.2015.7375688.

[23] S. Singh, A. Mishra, and U. Singh, “Detecting and avoiding of collaborative black hole attack on MANET using trusted AODV routing algorithm,” in 2016 Symposium on Colossal Data Analysis and Networking, CDAN 2016, 2016, doi: 10.1109/ CDAN.2016.7570906.

[24] R. Verma, R. Sharma, and U. Singh, “New approach through detection and prevention of wormhole attack in MANET,” in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212719.

[25] D. Wagh, N. Pareek, and U. Singh, “Elimination of internal attacksfor PUMA in MANET,” in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212710.

[26] R. Parihar, A. Jain, and U. Singh, “Support vector machine through detecting packet dropping misbehaving nodes in MANET,” in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212711.

[27] Gupta, P., Shukla, M., Arya, N., Singh, U., Mishra, K. (2022). Let the Blind See: An AIIoT-Based Device for Real-Time Object Recognition with the Voice Conversion. In: Al-Turjman, F., Nayyar, A. (eds) Machine Learning for Critical Internet of Medical Things. Springer, Cham. https://doi.org/10.1007/978-3-030-80928-7_8

[28] B. Bamne, N. Shrivastava, L. Parashar and U. Singh, "Transfer learning-based Object Detection by using Convolutional Neural Networks," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 328-332, doi: 10.1109/ICESC48915.2020.9156060.

[29] A. Saxena, A. Vyas, L. Parashar and U. Singh, "A Glaucoma Detection using Convolutional Neural Network," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 815-820, doi: 10.1109/ICESC48915.2020.9155930.

[30] H. Soni, A. Vyas, and U. Singh, “Identify Rare Disease Patients from Electronic Health Records through Machine Learning Approach,” in Proceedings of the International Conference on Inventive Research in Computing Applications, ICIRCA 2018, 2018, doi: 10.1109/ICIRCA.2018.8597203.

[31] A. Roshan, A. Vyas, and U. Singh, “Credit Card Fraud Detection Using Choice Tree Technology,” in Proceedings of the 2nd International Conference on Electronics, Communication and Aerospace Technology, ICECA 2018, 2018, doi: 10.1109/ICECA.2018.8474734.

[32] S. Sable, A. Porwal, and U. Singh, “Stock price prediction using genetic algorithms and evolution strategies,” in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, doi: 10.1109/ICECA.2017.8212724.

[33] V. Prakaulya, R. Sharma, U. Singh, and R. Itare, “Railway passenger forecasting using time series decomposition model,” in Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017, vol. 2017-Janua, 10.1109/ICECA.2017.8212725.

[34] Gupta, P., Shukla, M., Arya, N., Singh, U. and Mishra, K., 2022. Let the Blind See: An AIIoT-Based Device for Real-Time Object Recognition with the Voice Conversion. In Machine Learning for Critical Internet of Medical Things (pp. 177-198). Springer, Cham.