Journal of Contemporary Issues in Business and Government Vol. 27, No. 06, 2021
https://cibgp.com/

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

Intrusion Detection System through ID3-PCA-BP

L K Suresh Kumar
University College of Engineering, Osmania University, India

Abstract
Currently, problems such as a slow processing speed, a low detection rate, a high false-positive rate, and a big feature dimension are all part of intrusion detection. To handle these issues, Iterative Dichotomiser 3 (ID3), Principal Component Analysis (PCA), and Back Propagation (BP) algorithms are available. Through a lower false-positive rate, and a higher detection rate, the research-based intrusion detection model ID3-PCA-BP increases the processing speed of intrusion detection systems. To decrease the overall data volume and accelerate processing, ID3 is used to initially differentiate the data. Differentiate ID3s and keep the temporary training sample set for intrusion data in order to retrain and optimize the ID3 and BP, treat the ID3 judges as standard data, and delete the added intermediate data. Then, we should reduce the dimension of the data using PCA and then introduce the data to BP for secondary discrimination. However, the ID3 algorithm operates a shallow structure in order to prevent an extreme quantity of intermediate numbers from being analyzed as intrusion data. As a result, further BP processing cannot effectively raise the accuracy. BP accelerates data processing by employing the ReLU activation function from the simplified neural network calculation approach and the faster convergence ADAM optimization algorithm.

Keywords
Intrusion Detection, Machine Learning, Iterative Dichotomiser 3, Principal Component Analysis, Back Propagation

1. Introduction

Cyber-Security is always a hot topic in the market because cybercriminals are constantly finding new ways to access sensitive information to sell on the dark web or for future attacks. Researchers are trying hard to safeguard private information from cybercriminals or unauthorized persons. The main aim of the researchers is to detect the intrusion. In recent years, Machine Learning is a very popular algorithm tool that deserves experts to try its application in intrusion detection [1]**.**

The intrusion detection system can be classified into two types Host-Based Intrusion Detection systems (HIDS) and Network-Based Intrusion Detection systems (NIDS) [2]**.** Host-Based Intrusion Detection systems detect intrusion behavior by detecting various log files,

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

system information, and disk resource information. Network-Based Intrusion Detection systems detect intrusion behavior by detecting the data packets in and out of the local network data flow.

Different types of cyber-attacks are Malware attacks, Phishing attacks, Password attacks, Man-in-the-Middle attacks, Denial-of-Service attacks, and many more. A survey reveals that the Yahoo data breach caused a loss of 350 million US dollars and the Bitcoin breach caused a loss of about 70 million US dollars [3]. Table 1 depicts the data breaches of different companies in recent years.

Table 1: Data breaches of different companies in recent years

Year	Company	Data Breach
October 2017	Yahoo	3 billion user accounts
March 2018	Aadhaar	1.1 billion Indian citizens
April 2019	Facebook	533 million users
March 2020	CAM4	10.88 billion records
June 2021	LinkedIn	700 million users

According to the intrusion detection models presented by different scholars, it is observed that the majority of research studies pay insufficient lookout to real-time intrusion detection, while a few intrusion detection models with more additional in-depth real-time performance research suffer from low detection accuracy. This research offers the ID3-PCA-BP model, to thoroughly investigate the real-time problem that is required for intrusion detection and to ensure intrusion detection accuracy. The trained ID3 algorithm is basically a series of if-else statements that manage huge collections of data quickly but with inadequate precision; the BP algorithm has a slow real-time performance but high precision when processing extensive volumes of high-dimensional data. The experimental results illustrate that the model significantly accelerates training and detection while maintaining an increased detection rate.

The rest of the paper is organized as follows: Section 2 presents a literature survey for research works in NIDS. Section 3 provides the basic theory about the decision trees, principal component analysis, and deep neural networks. Section 4 elaborates on the system design of the ID3-PCA-BP model. Section 5 represents the experimental simulation, and results of the proposed ID3-PCA-BP model. Finally, Section 6 concludes the paper.

2. Relevant Work

M. Crosbie et al. [4] proposed a prototype architecture of a defense mechanism for computer systems. A finer-grained approach is proposed, where small, independent agents monitor the system.

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

In 2003, G. Vigna et al. [5] demonstrated that decision trees (DTs) could detect intrusions faster than the Snort detection engine at the time. The proposed method of joint optimization of feature selection and SVM training model is demonstrated using the intrusion detection dataset. The results indicate that the joint optimization method outperforms the SVM in terms of performance and convergence speed.

Y. Li et al. [6] used support vector machines to identify intrusions and explored the real-time problem; however, the accuracy rate was low.

N. P. Moorthi et al. [7] suggested reducing dimensionality by principal component analysis (PCA) and then detecting intrusions with support vector machines (SVMs).

G. Liu et al. [8] demonstrated that the self-optimization technology increases the classifier's accuracy and decreases training and testing time.

In 2019, J. Cao et al. [9] developed an intrusion detection model based on an upgraded convolution neural network that has a high intrusion detection accuracy and true positive rate, while exhibiting a low false-positive rate.

In the same year, Fernandez et al. [10] proposed training an intrusion detection system using a feedforward fully connected deep neural network (DNN) (IDS). Due to the fact that DNN demonstrated robustness in the scenario of dynamic IP address allocation, the model they developed has a broader variety of real-world applications.

Still in 2019, Y. Gu et al. [11] introduced the ICA-DNN intrusion detection model, which is based on ICA (Independent Component Analysis) and DNN and has a higher feature learning capability than some shallow machine learning models. This has increased classification accuracy, but the algorithm's prediction time is not particularly tested and the model performs poorly in real-time.

M. Soni et al. [12] presents a lightweight user verification and key-management protocol for medical users to achieve privacy.

M. Soni et al. [13] proposed a secure and lightweight health authentication and key agreement protocol using low-cost operations. Based on the computational analysis, the proposed protocol comparatively takes less execution cost, computation time, and power consumption.

M. Soni et al 14] proposed a blockchain-based technique to hide medical information and keep it safe. This will ensure the data is accessible and reliable to study reviews and patients.

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

M. Soni et al. [15] proposed an energy-efficient and secure communication scheme for mobile node applications, achieving user identity privacy.

Resende et al. [16] present a survey of intrusion detection systems using Random Forest-based methods for classification, feature selection, etc.

Benkhelifa et al. [17] present a survey on intrusion detection techniques presented in the literature for Internet of Things (IoT). They also present an architecture for intrusion detection systems which are suitable for IoT.

Liu et al. [18] present a survey on different security threats against algorithms such as Naive Bayes, DT, SVM, etc.

3. Basic Theory

3.1 Decision Tree

A Decision Tree is a tree-like graph with nodes representing the features or attributes where we pick an attribute and ask a question, and edges represent the answers to the question. The decision tree starts from the root node and continuously splits according to the characteristics of the data until all the data reach the leaf nodes. In the decision tree, internal nodes represent a feature or attribute [12], and the leaf nodes represent the actual output or class label.

Consider a sales office wants to judge whether the surveyed person has a car purchase demand based on the person's identity information, age, and experience. The survey results are shown in Table 2 and the corresponding decision tree is shown in Figure 1.

Table 2: Respondent data and willingness.

Role	Age	Experience	Whether to buy a car
Employee	28	7	None
Manager	35	3	None
Manager	30	5	Have
Employee	25	2	None
Manager	40	8	None

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

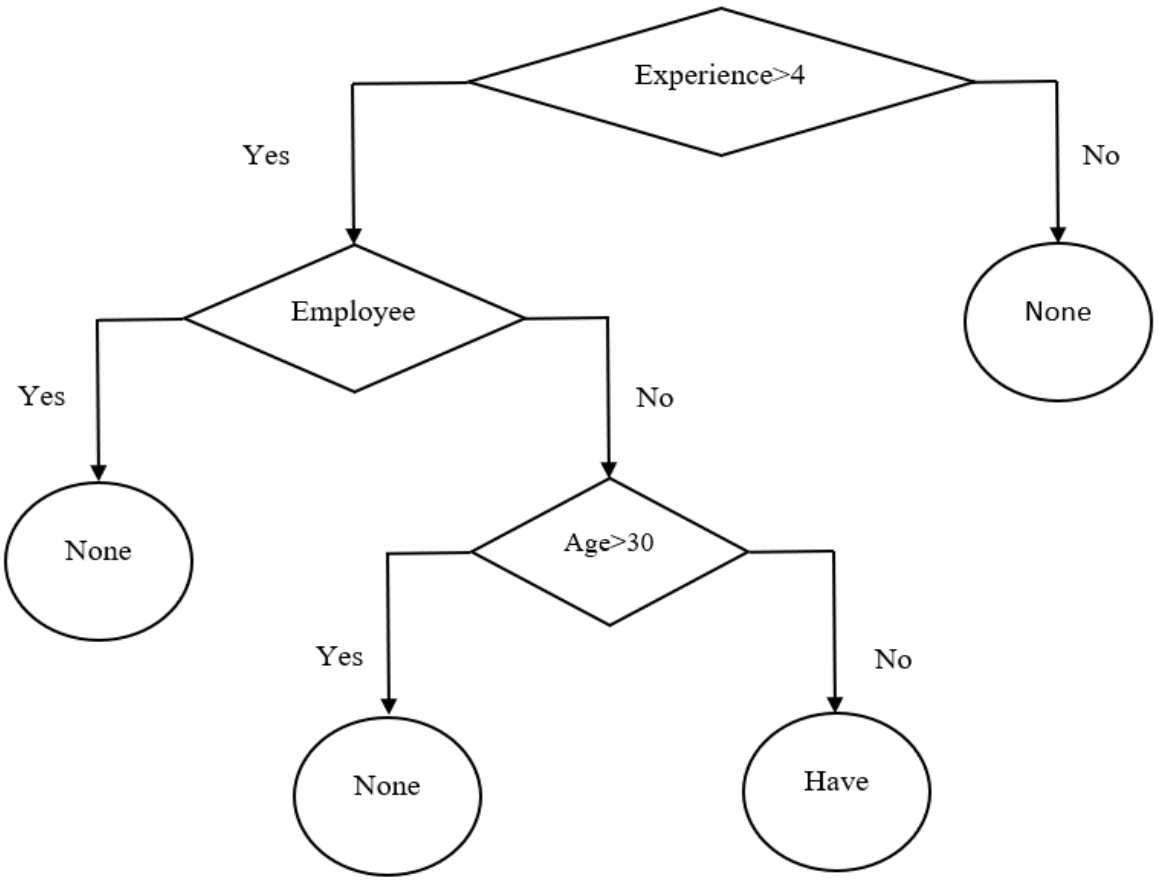

Figure 1: Decision tree of the respondent

Assume the given data as a person is a company manager, is 29 years old, and has experience of 6 years. From the decision tree, it can be inferred that the person has a car demand. In the above example, the feature selection order of the survey objects is different, and different decision trees can be generated.

There are three judgment bases for the selection of different splitting features, namely information gain, gain rate, and Gini index. The Iterative Dichotomiser 3 (ID3) algorithm uses information gain, the C4.5 algorithm uses gain ratio, and the Classification And the Regression Tree (CART) algorithm uses the Gini index to select the features. Due to space reasons, the decision tree pruning and specific feature selection will not be repeated [19].

3.2 Principal Component Analysis

Principal Component Analysis (PCA) is a linear dimensionality-reduction method. It is often used to reduce the dimensionality of large datasets, by transforming a large set of variables into a smaller one that still contains most of the information in the large dataset. Generally, dependent variables and independent variables are identified and the dataset is reduced in such a way that the dataset comprises independent variables.

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

3.3 Deep Neural Networks

A neural network is a large parallel interconnected network comprised of uncomplicated adaptive units that can be used to simulate the interactive response of the biological nervous system to real-world objects [20], where machine learning is used to interact with neural networks in a broader sense. In a biological neural network, each neuron is connected to other neurons, and when it is stimulated, it releases neurotransmitters to the connected neurons, altering their potential. When a neuron's potential strikes a threshold, the neuron becomes triggered and begins delivering neurotransmitters to its associated neurons [21].

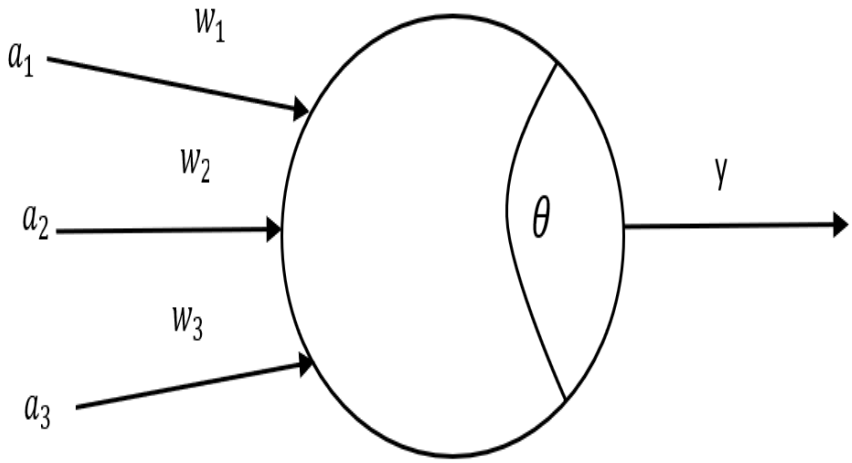

Figure 2: M-P neuron model

Foreign scholars outlined the above situation in 1943 into a simple model depicted in Figure 2, namely the "M-P neuron model" [22]. M-P neuron model stands for the "McCulloch-Pitts Neuron model". In this model, a_i represents the input from the i^{th} neuron, w_i represents the i^{th} neuron's connection weight and θ represents the threshold. The neuron accepts input signals from 'n' other neurons. The signals are dispatched via weighted connections, and neurons receive them. The neuron's output y is acquired by processing the activation function f(a) as specified in

$$o = f(\sum_{i=1}^{n} a_i w_i - \theta). \quad (1)$$

Frequently used activation functions are the sigmoid function, the tanh function, and Rectified Linear Unit (ReLU) function [23]. A neural network is constructed by connecting many neurons [24-26]. The term "Deep Neural Network" directs to a neural network with more than two layers and more than two hidden layers. A neural network with input and output layers is only competent in solving linear problems. The hidden layer is introduced to manage the nonlinear separable problem.

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

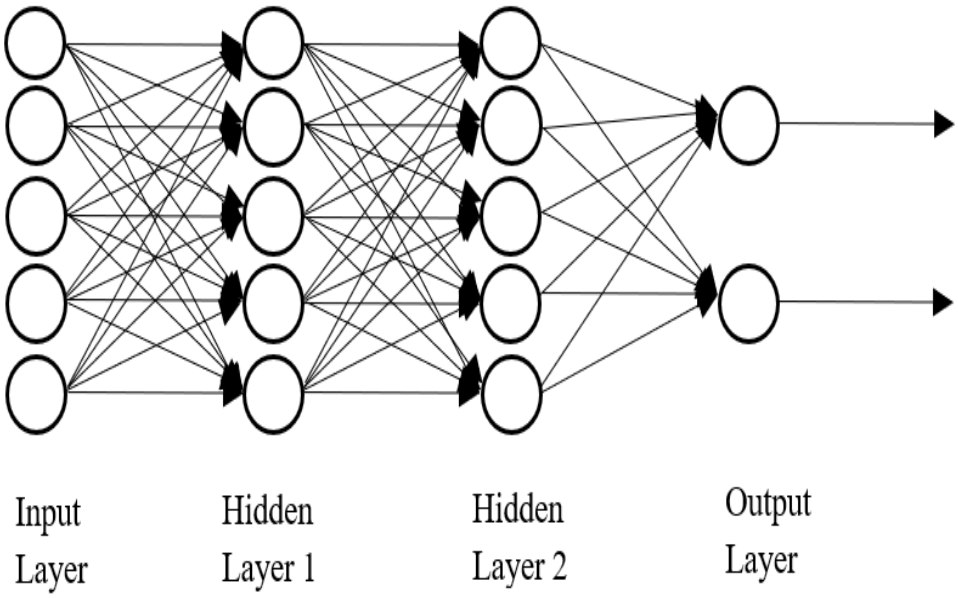

Figure 3: Double hidden layers of fully connected feedforward DNN

Figure 3 illustrates a fully connected feedforward neural network which means that any neuron in the preceding layer must be connected to any neuron in the successive layer. It is also called a feed-forward network as the output of one layer acts as input to the next layer. The neurons in the input layer accept only information and do not process any functions. The neural network's learning process is fundamentally one of constantly adjusting the connection weights and thresholds of neurons to approach the output results of the training samples. Among them, the most notable method is the error Back Propagation (BP) algorithm, which is used for the majority of neural network training today.

4. System Design

Figure 4 depicts the architecture of the model.

Journal of Contemporary Issues in Business and Government Vol. 27, No. 06, 2021
https://cibgp.com/

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

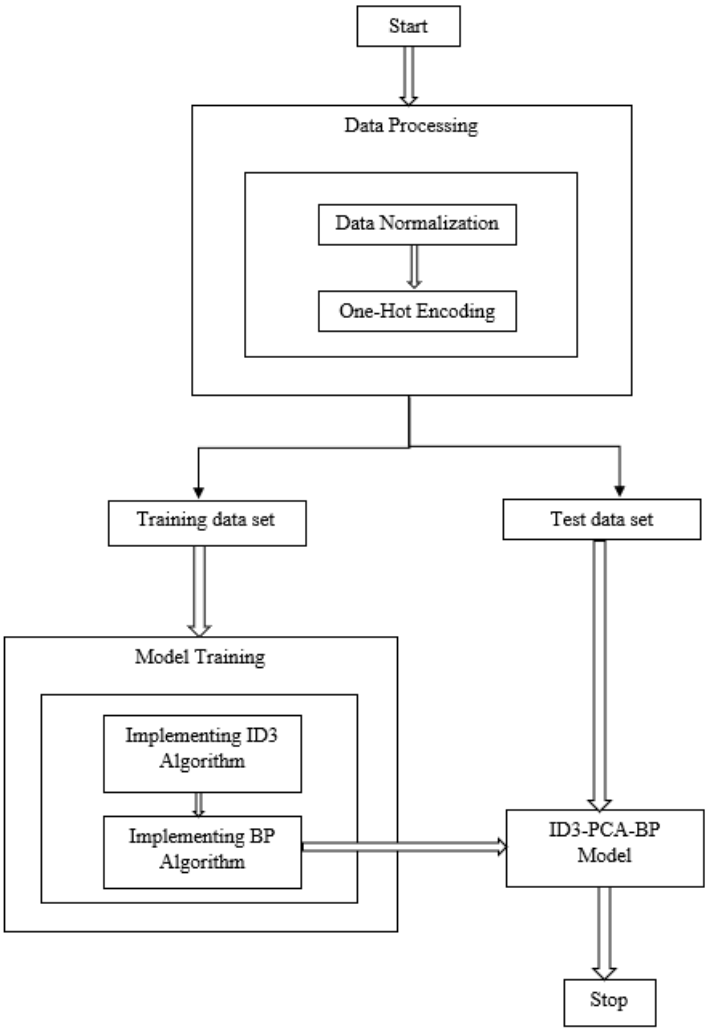

Figure 4: Architecture of the model

4.1. Data Processing

The data processing should be divided into two parts. First, Data Normalization - the continuous data are normalized, and then One-Hot Encoding - the discrete-valued data are encoded.

4.1.1.Data Normalization

Min-Max normalization is used to normalize the data. It performs the linear transformation on the original data. The transformed data ranges from 0 to 1. The transformation function used is as

$$x^* = \frac{x - min}{\max - min} \tag{2}$$

Let us assume the dataset contains m entities and each entity has n-dimensional features, where x represents the j^{th} eigenvalue of the i^{th} entity, min represents the feature's

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

minimum value and max represents the feature's maximum value in the j^{th} dimension before normalization. After normalization, x^* is the feature's j^{th} dimension value in the i^{th} piece of data.

4.1.2.One-Hot Encoding

One-Hot Encoding is also called as One-Bit Efficient Encoding. For discrete data encoding, N states are encoded using an N-bit status register. Each state has its register bit and only one bit is valid at any time. Table 3 shows a sample set. In table 3, the sample feature dimension is 3.

Table 3: Feature distribution of the sample set.

Sample	Feature 1	Feature 2	Feature 3
A	1	0	2
B	1	3	0
C	0	1	1

Feature 1 has two values [0,1], so the encoding rule should be as follows:

(i) 0 → 10
(ii) 1 → 01

The corresponding feature 2 has four values[0,1,2,3], so the encoding rule should be as follows:

(i) 0 → 1000
(ii) 1 → 0100
(iii) 2 → 0010
(iv) 3 → 0001

The encoding rule of feature 3 is the same as given above and will not be repeated. The results of samples A, B, and C after one-hot encoding are shown in Table 4.

Table 4: One-hot encoded result of the sample set.

Sample	Feature 1	Feature 2	Feature 3
A	01	1000	001
B	01	0001	100
C	10	0100	010

4.2. Dataset

In machine learning, datasets are split into two subsets namely, the training dataset and the testing dataset.

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

4.2.1. Training Dataset

The training dataset is 80% of the original dataset that is fed to the machine learning model to discover and learn patterns.

4.2.2. Test Dataset

The test dataset is 20% of the original dataset that is used to evaluate the performance and progress of training and optimize it for improved results.

4.3. Model Training

Model Training is divided into 2 parts namely building a Decision Tree (DT) and training a Deep Neural Network (DNN).

4.3.1. Implementing ID3 Algorithm

The information gain used by the ID3 algorithm has a selection for attributes with a large number of possible values, and the model used uses DT before the experimental data are dimensionally reduced. The depth of DT should not be too deep because the primary aim to use the ID3 is to misjudge average data as intrusion data as little as possible but not to identify as much intrusion data as possible.

4.3.2. Implementing BP Algorithm

DNN uses the Back Propagation (BP) algorithm for training and Rectified Linear Unit (ReLU) as the activation function to simplify the calculation process of the neural network. BP algorithm requires a fairly large number of hidden layers to analyze high-dimensional data, the underfitting phenomena will be severe if the number of hidden layers is too low. As the number of hidden layers increases, the time spent training the neural network grows exponentially, which is inconsistent with the real-time needs of this work. When PCA pair data are introduced after the dimensionality reduction process, the relationship between data feature dimensions and data redundancy is reduced, BP algorithm training is faster, and BP algorithm ensures accuracy.

4.3.3. ID3-PCA-BP Model Optimization

As shown in Figure 5, firstly, the pre-processed test dataset is classified with the trained ID3. Next, the data whose classification result is intrusion is considered as intrusion and stored in the temporary training sample. The information whose classification result is usual is removed from this ID3 classification. Given the label, go for the second judgment type of data. The ID3 layer is identical to a filter screen, which filters out the intrusion data effortlessly to filter. Since the trained ID3 is a series of if-else statements, the processing speed of large collections of data is extremely high, which enormously reduces the workload of BP algorithm and enhances the running speed of the BP algorithm.

Journal of Contemporary Issues in Business and Government Vol. 27, No. 06, 2021
https://cibgp.com/

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

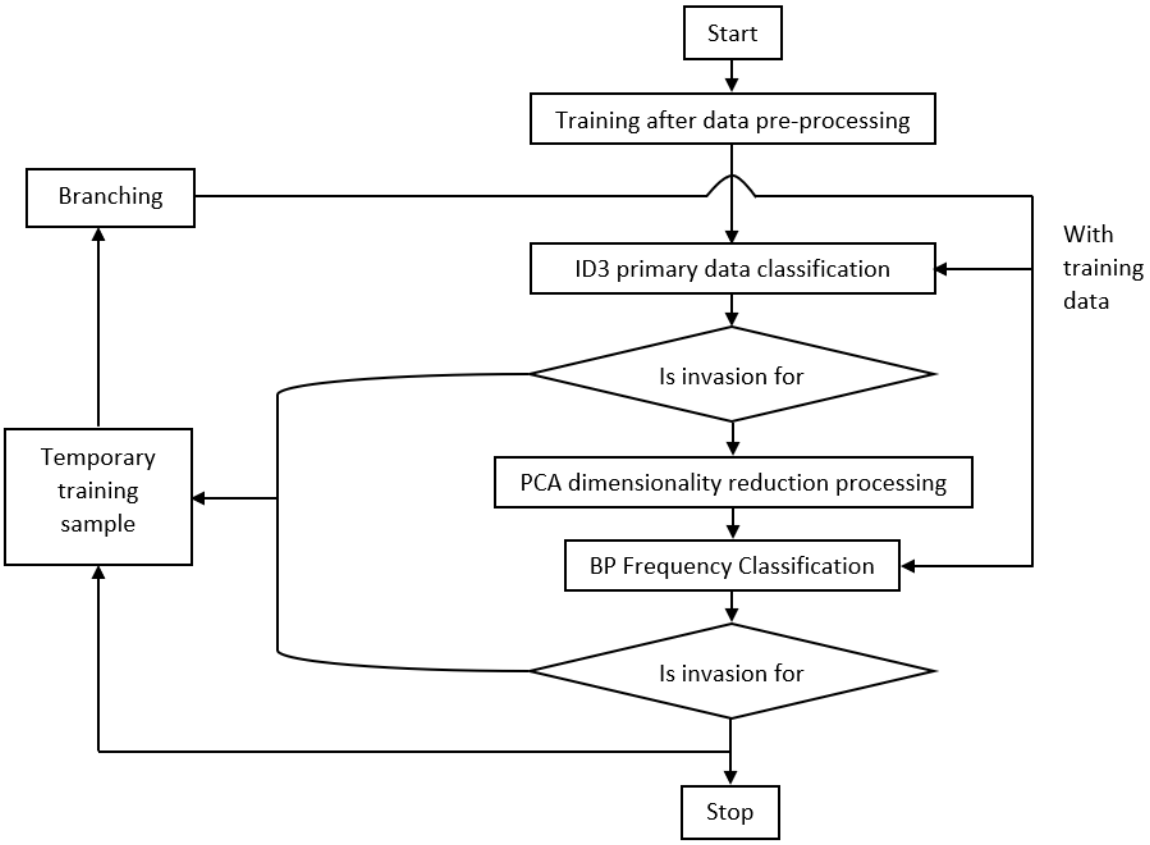

Figure 5: ID3-PCA-BP Model Optimization

The second step is to perform PCA dimensionality reduction on the data judged by ID3 to be predicted, but the labels have been removed. The trained BP algorithm classifies the low-dimensional data result after PCA processing for the second time. If the classification result is an intrusion, the intrusion label is added, and the temporary training sample is kept. Since ID3 and BP algorithms belong to supervised learning, the labels assigned to the data need to be used when using the brief training sample set for retraining.

5. Experimental Simulation

5.1. Dataset

The Knowledge Discovery in Databases (KDD) 99 dataset was employed in this experiment. This dataset was published in 1999 and is considered one of the baseline datasets to evaluate different intrusion detection systems including machine learning-based intrusion detection systems[27]. Each record in the KDD99 dataset has 41 features and records are either labeled as normal or an attack type. There are basically four types of attacks in the dataset[28]. They are Denial of Service Attack (DoS), User to Root Attack (U2R), and Remote to Local Attack (R2L), and Probing Attack. The dataset has 78% and 75% of duplicate records in training and testing datasets respectively. The U2R (in training dataset) and R2L (in training dataset) attacks are inadequate to train a machine learning model and that makes the KDD99 dataset

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

imbalanced. Table 5 contains information about the data distribution for each class of the dataset KDD99[29].

Table 5: Data distribution of KDD99 dataset

Class	**Amount of training dataset**	**Amount of test dataset**
Normal	972780	60595
DoS	3883389	231450
U2R	52	245
R2L	1106	14574
Probe	41104	4169

5.2. Metrics

The confusion matrix is listed as shown in Table 6. True Positive (TP) represents the number of data pieces whose real data type is normal and the model prediction result is still normal. True Negative (TN) represents the number of data pieces whose real data type is an intrusion and the model prediction result is also intrusion. False Positive (FP) indicates the number of data pieces whose real data type is an intrusion and the model prediction result is normal. False Negative (FN) represents the number of data pieces whose real data type is normal and the model prediction result is an intrusion. Of course, the size of different data models is not enough as a standard for evaluating the experimental results. Therefore, a relatively reasonable evaluation standard is designated based on the above parameters: the Accuracy rate (AC), Detection rate (DR), Precision rate (PR), and False Alarm Rate (FAR), and the definitions are as follows:

$$\text{Accuracy Rate}(AC) = \frac{TP+TN}{TP+FP+FN+TN}$$

$$\text{Detection Rate } (DR) = \frac{TP}{TP+FN}$$

$$\text{False Alarm Rate } (FAR) = \frac{FP}{FP+TN}$$

$$\text{Precision Rate } (PR) = \frac{TP}{TP+FP} \quad (3)$$

Table 6: Confusion Matrix

Confusion Matrix		**Actual Value**	
		Normal Data	**Intrusion Data**
Predictive Value	**Normal Data**	TP	FP
	Intrusion Data	FN	TN

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

5.3. Parameter Setting

After preprocessing the data, including normalization and one-hot encoding, all data values are located in the interval [0, 1]. Next, we use the ID3 algorithm to perform the first test on all training data. For the secondary screening, the primary parameters of the ID3, PCA, and BP used are shown in Tables 7–9, and then the PCA dimensionality reduction is carried out. Because the developed system is linear, all parameters can be fetched one by one by fixing other parameters to obtain the optimal parameters.

In ID3, for Criterion (attribute segmentation criterion), the value is of string type. There are two values to choose from: "Gini" and "Entropy." For Splitter (segmentation point), the value is of string type. There are two standards to choose from, "best" and "random," where "best" means that in all features encountering the optimal segmentation point and "random" means encountering the optimal segmentation point in the randomly selected part of the features. For max_depth (the constructed decision tree), the value can be an integer or none. For random_state (multiple states used to generate random numbers), the value can be an integer, an instance of a random state, or none. Experiments have revealed that the best effect is achieved when the value is 392.

Table 7: ID3 primary parameters

Primary Parameters	Value
Criterion	Entropy
Splitter	Best
max_depth	5
random_state	392

n_components (feature dimension after dimension reduction) can be the number of dimensions decreased or the percentage of data enclosed. whiten (whether whitening) decreases the correlation between features and all features have the same variance. svd_solver (singular value decomposer) is a string when its value is "auto," and certain conditions are met, the complete singular value decomposition function is called.

Table 8: PCA primary parameters

Primary Parameters	Value
n_components	11
Whiten	True
svd_solver	Auto

hidden_layer_sizes (hidden layer sizes) is of tuple type. The number of hidden layers and the number of neurons in the hidden layer are determined by changing this value. Two hidden layers are presented here, with 140 neurons in the first layer and 70 neurons in the second layer. activation is the activation function which is ReLU. solver (weight optimization

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

function) is selected by selecting different strings of the related weight optimization function which is Adam.

Table 9: BP primary parameters

Primary Parameters	Value
hidden_layer_sizes	[140,70]
Activation	ReLU
Solver	Adam

5.4. Experimental Results

This experiment uses a Windows10 system and 64-bit operating system. The processor version is Intel® CoreTMi7-9750H CPU@2.60 GHz. Total physical memory is 16.0 GB. The development language used is Python3.5, and the software package used is sklearn.

5.4.1. Experiment 1

This experiment mainly compared the two-class prediction accuracy and training time of FC [13], ID3, PCA-BP, EDF [23], CNN [30], and ID3-PCA-BP models. To reflect the characteristics of ID3-PCA-BP, the test data used are all the data in the KDD test dataset. For the convenience of observance, Figure 6 is obtained from Table 10. In the figure, because the FC training time is too long, the impact on the preferred vertical axis interval is too significant, so it is not listed.

Table 10: Results of experiment 1.

Algorithm	AC%	DR%	Training time per second	Prediction time/ms
FC	79.44	N/A	1856.37	221.15
ID3	75.63	65.70	0.62	58.87
EDF	79.37	64.38	17.74	N/A
CNN	78.68	68.73	92.69	N/A
PCA-BP	79.66	66.64	15.62	47.82
ID3-PCA-BP	88.59	84.51	14.13	57.83

Observing Figure 6, we can see that the accuracy AC of PCA-BP and FC is identical, but the training time of FC is much higher than that of PCA-BP, the prediction time is slightly longer, and the real-time detection is insufficient. On the other hand, the training time of the EDF algorithm is marginally more elongated than that of PCA-BP, and the accuracy of AC is the same as that of PCA-BP. The training time of the CNN algorithm is as long as 90 s, and the accuracy rate is barely lower than that of EDF and PCA-BP, which is inferior to both. On the other hand, although the training speed of ID3 is breakneck, the accuracy rate is four percentage points less than that of EDF and PCA-BP.

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

Compared with PCA-BP without ID3, ID3-PCA-BP takes 1.32 s longer to train and takes about 10 ms more to predict, but the accuracy AC has improved by nearly ten percentage points remarkably. The introduction of ID3 has the tiniest impact on training time and prediction time but greatly improves prediction accuracy.

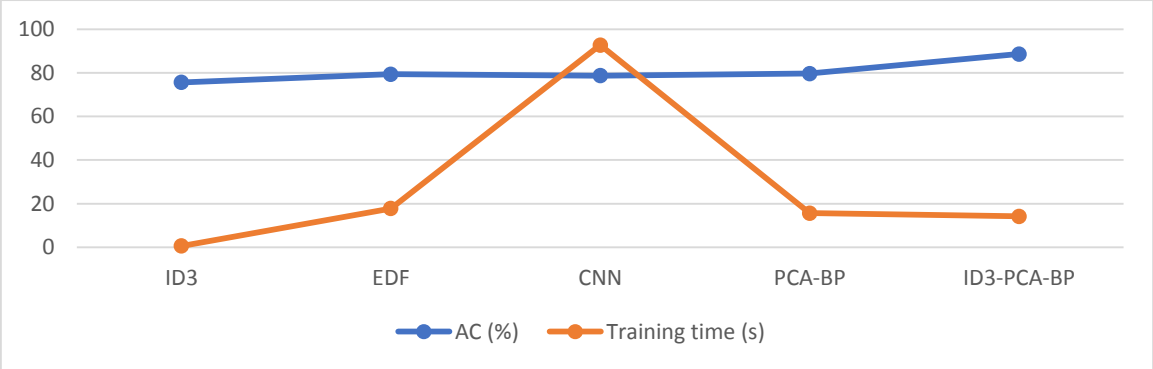

Figure 6: Results of experiment 1.

5.4.2. Experiment 2

This experiment mainly studies the five-classification detection time of the ID3-PCA-BP model, the real-time detection problem under five classifications. Mark normal samples as 0, DoS samples as 1, probe samples as 2, U2R samples as 3, and R2L samples as 4. The analysis in Table 11 shows that the speed advantage of ID3-PCA-BP in the five classifications is absolute. Still, the prevalent accuracy is slightly inferior to EDF and CNN, and we compare ID3 and PCA-BP. However, the training time is somewhat short. The overall accuracy is low, and the performance is lacking. Comparing PCA-BP and ID3-PCA-BP in the five-classification experiment, the training time is 3 s longer. The accuracy rate is improved by six percentage points, confirming that the introduction of DT does not cause much time while ensuring the accuracy rate loss.

Table 11: Total results of experiment 2.

Algorithm	Five-classification training time/s	Total accuracy/%
ID3	0.62	78.86
EDF	41.21	87.23
CNN	93.34	87.11
PCA-BP	14.43	77.01
ID3-PCA-BP	17.41	83.19

Journal of Contemporary Issues in Business and Government Vol. 27, No. 06, 2021
https://cibgp.com/

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

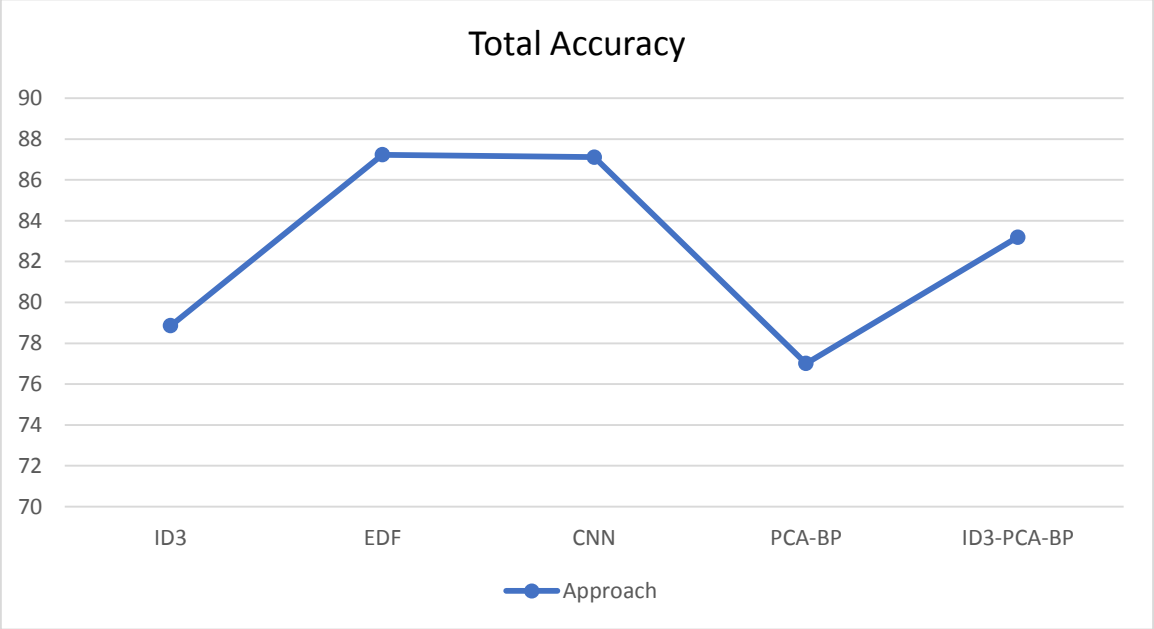

Figure 7: Comparison of accuracy of the model

Table 12: Five-classification results of experiment 2.

Algorithm	Evaluation criteria	Normal	DoS	Probe	R2L	U2R
ID3	PR	67.11	91.80	25	71.45	53
	DR	95.36	84.20	0.43	61.93	5
	FAR	34.33	3.12	0.13	2.66	0.06
EDF	AC	91.68	98.13	88.18	59.51	15.33
CNN	AC	92.51	98.44	92.52	35.47	26.51
PCA-BP	PR	64.72	95.66	96.69	82.43	0
	DR	98.12	75.41	16.18	67.48	0
	FAR	39.29	1.57	0.04	1.68	0
ID3-PCA-BP	PR	68	92.29	91.05	78.95	0
	DR	98.79	83.22	36.93	61.61	0
	FAR	26.14	4.4	0.31	1.46	0

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

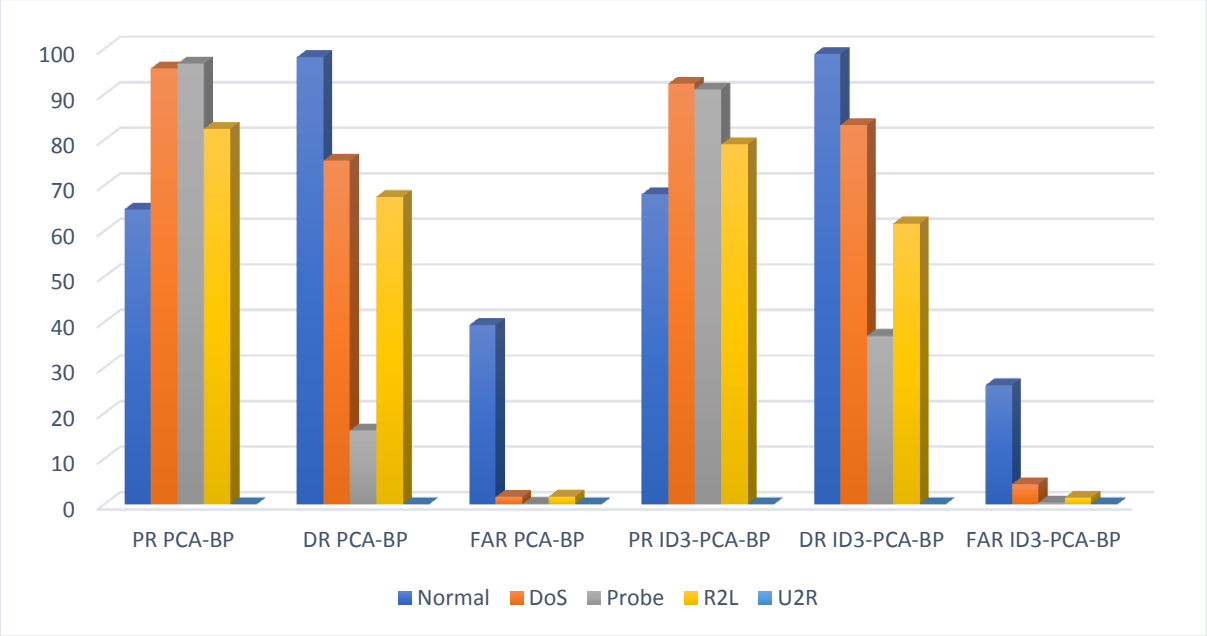

Figure 8: Classification performance over different attack scenarios for the PCA-BP and ID3-PCA-BP Models

The analysis of Table 12 and Figures 7–8 shows that ID3-PCA-BP may have processed part of the data during ID3 pre-screening, resulting in no U2R outcomes (small U2R sample size). The advantage of ID3-PCA-BP is mainly reflected in the capability to recognize R2L, because of its small proportion in the dataset.

As a result, the overall accuracy rate is lower than that of EDF and CNN. And also the model has a relatively high false alarm rate for normal data when the detection rate is relatively high. This is an issue that needs to be pointed out.

6. Conclusion

The ID3-PCA-BP intrusion detection model described in this study greatly enhances the training and detection speed while preserving accuracy. The model employs the ID3 algorithm to do a primary screening of the pre-processed data to be detected before employing PCA as an input to perform a secondary judgment through the BP algorithm. The addition of ID3 causes a small increase in training time but a large gain in accuracy. Simultaneously, ID3 pre-screening reduces the forthcoming BP burden, which has a considerable effect on the overall training speed. The following study concentrates mostly on overcoming the problem of the ID3-PCA-BP model having a high false alarm rate for normal data in the five-classification experiment while also increasing the ID3-PCA-BP model's five-classification capability [31].

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

References

[1] K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed, and M. Xu, "A survey on machine learning techniques for cybersecurity in the last decade," IEEE Access, vol. 8, pp. 222310– 222354, 2020.

[2] A. Phadke, M. Kulkarni, P. Bhawalkar, and R. Bhattad, "A review of machine learning methodologies for network intrusion detection," in Proceedings of the 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), pp. 272–275, Erode, India, March 2019.

[3] I. Sembiring, "Implementation of honeypot to detect and prevent distributed denial of service attack," in Proceedings of the 3rd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), pp. 345– 350, Semarang, Indonesia, October 2016.

[4] M. Crosbie and G. Spafford, "Defending a Computer System Using Autonomous Agents," Technical Report No. 95-022, Department of Computer Sciences, Purdue University, West Lafayette, Indiana, March 1996.

[5] G. Vigna, E. Jonsson, and C. Kruegel, Eds., RAID 2003, LNCS 2820, pp. 173–191, c Springer-Verlag, Berlin Heidelberg, 2003.

[6] Y. Li, J. Xia, S. Zhang, J. Yan, X. Ai, and K. Dai, "An efficient intrusion detection system based on support vector machines and gradually feature removal method," Expert Systems with Applications, vol. 39, no. 1, pp. 424–430, 2012.

[7] N. P. Moorthi and V. Mathivanan, "Hybrid optimization for feature selection in opinion mining," International Journal of Engineering & Technology, vol. 7, no. 1.3, p. 112, 2017.

[8] G. Liu, Z. Yi, and S. Yang, "A hierarchical intrusion detection model based on the PCA neural networks," Neurocomputing, vol. 70, no. 7-9, pp. 1561–1568, 2007.

[9] J. Cao, C. Wu, L. Chen, H. Cui, and G. Feng, Hindawi Computational Intelligence and Neuroscience, vol. 2019, Article ID 2060796, 12 pages, 2019.

[10] L. Ashiku and C. Dagli, "Network intrusion detection system using deep learning," Procedia Computer Science, vol. 185, pp. 239–247, 2021.

[11] Y. Gu, B. Zhou, and J. Zhao, "PCA-ICA ensembled intrusion detection system by pareto-optimal optimization," Information Technology Journal, vol. 7, no. 3, pp. 510–515, 2008.

[12] M. Soni and D. K. Singh, "privacy preserving authentication and key management protocol for health information system," Data Protection and Privacy in Healthcare: Research and Innovations, CRC Publication, vol. 37, 2021.

[13] M. Soni and D. K. Singh, "LAKA: Lightweight Authentication and Key Agreement Protocol for Internet of Things Based Wireless Body Area Network," Wireless Personal Communication, 2021.

[14] M. Soni and D. K. Singh, "Blockchain Implementation for Privacy Preserving and Securing the Healthcare Data," in Proceedings of the 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT), pp. 729–734, Bhopal, India, June 2021.

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

[15] M. Soni, G. Dhiman, B. S. Rajput, R. Patel, and N. K. Tejra, "Energy-Effective and Secure Data Transfer Scheme for Mobile Nodes in Smart City Applications," Wireless Pers Commun, 2021.
[16] Resende P., Drummond A."A survey of random forest based methods for intrusion detection systems", ACM Comput. Surv., 51 (3) (2018), pp. 1-36.
[17] Benkhelifa E., Welsh T., Hamouda W., "A critical review of practices and challenges in intrusion detection systems for IoT: Toward universal and resilient systems", IEEE Commun. Surv. Tutor., 20 (4) (2018), pp. 3496-3509, 10.1109/COMST.2018.2844742.
[18] Liu Q., Li P., Zhao W., Cai W., Yu S., Leung V.C.M., "A survey on security threats and defensive techniques of machine learning: A datadriven view", IEEE Access, 6 (2018), pp. 12103-12117, 10.1109/ACCESS.2018.2805680.
[19] R. Nian, G. Ji, and M. Verleysen, "An unsupervised Gaussian mixture classification mechanism based on statistical learning analysis," in Proceedings of the 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery, pp. 14–18, Shandong, China, October 2008.
[20] R. Aggarwal and Y. Song, "Artificial neural networks in power systems. Part 1: general introduction to neural computing," Power Engineering Journal, vol. 11, no. 3, pp. 129–134, June 1997.
[21] Z.-H. Zhou, Machine Learning, Springer Nature Singapore Pte Ltd, Singapore, 2021.
[22] V. R. Sargsyan, "Formation of human higher nervous activity and new biological theories, HSOA journal of brain & neuroscience research, Sargsyan VR," J Brain Neursci, vol. 2, p. 004, 2018.
[23] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), in Proceedings of the 2017 International Conference on Engineering and Technology (ICET), pp. 1–6, Antalya, Turkey, August 2017.
[24] V. K. Gupta, S. K. Shukla, and R. S. Rawat, "Crime tracking system and people's safety in India using machine learning approaches," International Journal of Modern Research, vol. 2, no. 1, pp. 1–7, 2022.
[25] P. K. Vaishnav, S. Sharma, and P. Sharma, "Analytical review analysis for screening COVID-19 disease," International Journal of Modern Research, vol. 1, no. 1, pp. 22–29, 2021.
[26] I. Chatterjee, "Artificial intelligence and patentability: review and discussions," International Journal of Modern Research, vol. 1, no. 1, pp. 15–21, 2021.
[27] KDD cup 1999 data (2021) http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html, October 2007.
[28] Tavallaee M., Bagheri E., Lu W., Ghorbani A.A. A detailed analysis of the KDD CUP 99 data set Proceedings of 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, IEEE (2009), pp. 1-6.
[29] Özgür A., Erdem H. A review of KDD99 dataset usage in intrusion detection and machine learning between 2010 and 2015 PeerJ Preprints, 4 (2016), p. e1954v1.

P-ISSN: 2204-1990; E-ISSN: 1323-6903
DOI: 10.47750/cibg.2021.27.06.141

[30] X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, "An adaptive ensemble machine learning model for intrusion detection," IEEE Access, vol. 7, pp. 82512–82521, 2019.

[31] Q. Qin, K. Poularakis, K. K. Leung, and L. Tassiulas, "Line-speed and scalable intrusion detection at the network edge via federated learning," in Proceedings of the 2020 IFIP Networking Conference (Networking), pp. 352–360, Paris, France, June 2020.

[32]. P R Anisha, Dr. C Kishor Kumar Reddy, "Early Detection of Diabetes using Machine Learning Algorithms and Internet of Things: ADPA", Springer INDIA, 2019, India.

[33]. Dr B V Ramana Murthy , Kishor Kumar Reddy C, Anisha P R, Rajasekhar Shastry, "IOT Based Smart Stale Food Detector", Springer INDIA, 2019, India.

[33]. Kishor Kumar Reddy C, Anisha P R, Rajasekhar Shastry, Dr B V Ramana Murthy and Dr Vuppu Padmakar "Automated rainwater Harvesting System", IEEE ICCES, 2019.

[34]. Kishor Kumar Reddy C, Anisha P R, Rajasekhar Shastry, Dr B V Ramana "Comparative Study on Internet of Things: Enablers and Constraints", Springer-ICDECT, 2019.

[35]. Rajasekhar Shastry, Dr B V Ramana Murthy, Kishor Kumar Reddy C, Anisha P R " Automated Lighting Smart Parking Using Internet of Things", Springer-ICICCS, 2019

[36]. R Gangadhara Reddy, M Ramesh Babu, CH jaya Prakash Rao, Kishor Kumar Reddy C , PPSIC: Pulsated Power Supply Inverter Circuit", IEEE CICN, 2017

[37]. Kishor Kumar Reddy C and Vijaya Babu B, "ISPM-OC: Improved Snow Prediction Model Using Optimal k-Means Clustering and Decision Tree to Nowcast Snow/No-Snow", SCESM 2017

[38]. Kishor Kumar Reddy C and Vijaya Babu B, "ISLIQ-OC: Improved Supervised Learning in Quest using Optimal k-means Clustering Mechanism to Nowcast Snow/No-Snow", SCESM , 2017.

[39]. Kishor Kumar Reddy C, Anisha P R and G V S Raju, "A Novel Methodology to Detect Bone Cancer Stage Using Mean Intensity of MRI Imagery and Region Growing Algorithm", Springer International Congress on Information and Communication Technology, 2016

[40]. Kishor Kumar Reddy C, Anisha P R and G V S Raju, "Detection of Pancreatic Cancer using Clustering and Wavelet Transform Techniques", IEEE International Conference on Computational Intelligence and Communication Networks , December 2015.

[41]. Kishor Kumar Reddy C, "A Novel Approach for Detecting the Tumor Size and Bone Cancer Stage using Region Growing Algorithm", IEEE International Conference on Computational Intelligence and Communication Networks , December 2015.