International Journal for Modern Trends in Science and Technology
Volume 11, Issue 09, pages 49-59.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol11issue09.html
DOI: https://doi.org/10.5281/zenodo.17148917

Intrusion Detection System using Machine Learning

Ch. Deepika[1], S Usha Baby[1], G.Asmitha[2] | M.V.S.Mounika[2] | G.Neha[2] | I.Hima Lakshmi[2] | K.Navya[2]

[1]*Assistant Professor, Department of CSE, Vijaya Institute of Technology for Women, Enikepadu, AP, INDIA.*
[2]*Department of CSE, Vijaya Institute of Technology for Women, Enikepadu, AP, INDIA.*

To Cite this Article
Ch. Deepika, S Usha Baby, G.Asmitha, M.V.S.Mounika, G.Neha, I.Hima Lakshmi & K.Navya (2025). Intrusion Detection System using Machine Learning. International Journal for Modern Trends in Science and Technology, 11(09), 49-59. https://doi.org/10.5281/zenodo.17148917

Article Info
Received: 07 August 2025; Accepted: 31 August 2025.; Published: 05 September 2025.

KEYWORDS

signature-based system, anomaly-based, malicious activity, prompt, seensors, analysis engine, alerting systems,false positives.

ABSTRACT

The purpose of an intrusion detection system (IDS) is to protect the confidentiality, integrity, and availability of a system. Intrusion detection systems (IDS) are designed to detect specific issues, and are categorized as signature-based (SIDS) or anomaly-based (AIDS). IDS can be software or hardware. Intrusion Detection System (IDS) is a network security system that monitors and analyzes network traffic for signs of unauthorized access or malicious activity. IDS is designed to detect and alert on potential security threats, allowing administrators to take prompt action to prevent or mitigate attacks,typically involves several key components, including ensors, analysis engines, and alerting systems. Sensors collect network traffic data, which is then analyzed by the analysis engine to identify potential threats. The alerting system notifies administrators of detected threats, allowing them to take prompt action. IDS can also be integrated with other security systems, such as firewalls and intrusion prevention systems (IPS), to provide comprehensive security solutions. Advanced IDS solutions may incorporate machine learning and artificial intelligence (AI) to improve detection accuracy and reduce false positives.

1. INTRODUCTION

An Intrusion Detection System (IDS) using machine learning is a cybersecurity solution that leverages artificial intelligence and machine learning algorithms to detect and respond to potential security threats in real-time. Traditional IDS rely on signature-based detection methods, which can be limited in their ability to detect unknown threats. Machine learning-based IDS, on the other hand, can learn from network traffic patterns and identify anomalies, allowing for more effective detection of complex attacks, including zero-day threats. By analyzing network traffic, system logs, and other data, machine learning algorithms can identify patterns and anomalies that may indicate

malicious activity, enabling organizations to respond quickly and effectively to potential security threats. This approach enhances the overall security posture of an organization, providing a robust defense against evolving cyber threats.

1.1. Objectives:

1. Detect Unauthorized Access: Identify and alert on potential security threats.
2. Monitor Network Traffic: Continuously monitor network traffic for suspicious activity.
3. Identify Malicious Activity: Detect and classify malicious activity.
4. Alert and Respond: Alert security teams and enable response to potential threats.
5. Prevent Security Breaches: Prevent security breaches by detecting and responding to threats in real-time.

Existing System

Existing Intrusion Detection Systems (IDS) using machine learning are designed to identify and classify network attacks. These systems leverage various machine learning techniques, including supervised, unsupervised, and deep learning methods.

Types of Machine Learning IDS

1. Supervised Learning: Supervised learning in Intrusion Detection Systems (IDS) using Machine Learning (ML) involves training ML models on labeled datasets to recognize patterns and anomalies in network traffic
2. Unsupervised Learning: Unsupervised learning in Intrusion Detection Systems (IDS) using Machine Learning (ML) involves training models on unlabeled data to identify patterns and anomalies
3. Deep Learning: Deep learning in Intrusion Detection Systems (IDS) using Machine

Learning (ML) leverages complex neural networks to detect patterns and anomalies in network traffic.

Proposed System:

1. Advanced Threat Detection: The proposed system would utilize machine learning and deep learning algorithms to detect complex threats, including zero-day attacks and advanced persistent threats (APTs)
2. Real-Time Monitoring: The system would provide real-time monitoring of network traffic, allowing for swift detection and response to potential threats
3. Anomaly Detection: The system would use anomaly detection techniques to identify unusual patterns in network traffic that may indicate a potential threat
4. Context-Aware Analysis: The system would provide context-aware analysis, taking into account the specific network environment, user behavior, and other relevant factors to improve detection accuracy.
5. Integration with Other Security Systems: The system would be designed to integrate with other security systems, such as firewalls and intrusion prevention systems (IPS), to provide a comprehensive security solution.

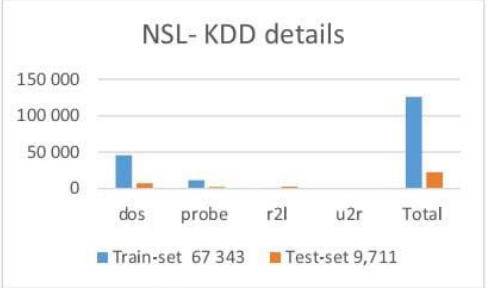
An intrusion detection system (IDS) is a network security tool that monitors for malicious activity. It can be a software application or hardware device.

It identifies potential threats by analyzing data for known attack patterns, unusual activity, or deviations from established baselines, and then alerts security personnel or a centralized security tool.

2.1 NSL-KDD

KDD'99 is outdated and contains redundant records, resulting in network intrusion detection inaccuracy. The problem is solved in NSL-KDD, which is a developed version of KDD'99. The training set of NSL-KDD has 125973 data points, whereas the testing set contains 22544 data points. It features 41 variables with numeric, binary, and nominal data types, as well as a label. Dos, probe, r2l, u2r, and regular class are the four major groups of attack types in the dataset. The distribution of each assault in training and testing sets is shown in Table

Dataset	Class	Train-set	Test-set
NSL-KDD	normal	67 343	9,711
	dos	45 927	7 458
	probe	11 656	2 421
	r2l	995	2 754
	u2r	52	200
	Total	125 973	22 544



Existing Intrusion Detection Systems (IDS) using machine learning are designed to identify and classify network attacks. These systems leverage various machine learning techniques, including supervised, unsupervised, and deep learning methods.

The training set of NSL-KDD has 125973 data points, whereas the testing set contains 22544 data points. It features 41 variables with numeric, binary, and nominal data types, as well as a label. Dos, probe, r2l, u2r, and regular class are the four major groups of attack types in the dataset. The distribution of each assault in training and testing sets is shown.

An intrusion detection system (IDS) is a software application or device that monitors system or network activity for policy violations or malicious behavior, and generates reports for the management system. The need for an intrusion detection system is undeniable; thus, an accurate model must be developed. In this field, machine learning has proven to be an effective investigation device that can detect any irregular event taking place in any system's traffic. An intrusion detection system (IDS) is a software application or device that monitors system or network activity for policy violations or malicious behavior, and generates reports for the management system. The need for an intrusion detection system is undeniable; thus, an accurate model must be developed. In this field, machine learning has proven to be an effective investigation device that can detect any irregular event taking place in any system's traffic. To build a good IDS it well be able to detect malicious traffics with a high efficacy; the accuracy of algorithms of classification well decide that efficacy.

3. Related works

Abhishek Divekar et al (A. Divekar, 2018) used classification algorithms such as Naıve. Bayes, K means, neural network, RF, SVM, and DT and compared performances for alternatives KDD'99. They found that UNSW-NB15 is a better and modern alternative for the KDD'99. The result of the study showed that classifiers trained in terms of f1-score were much better than those trained with KDD'99 and NSL KDD. The authors of (Srivastava, 2018) have attempted to assess the performance and effectiveness of NIDS.

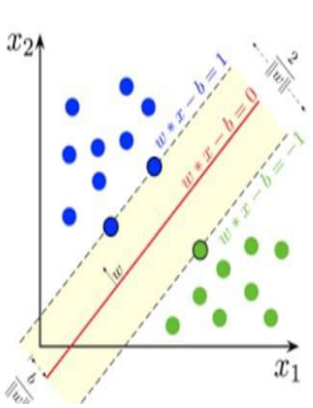
They have used two characteristic reduction methods, LDA and CCA. Seven classifiers were applied with different measurement parameters and metrics such as FPR, training time, accuracy, the ROC zone. The algorithms used are the random tree, the naive bayes, the rep tree, the RF, random committee, randomizable bagging, and filtered. The result with LDA and random tree on UNSW-NB15 was declared best.k2 In (M. Belouch, Performance evalution of intrusion detection based on machine learning using apache spark, 2018), the authors conducted experimental studies and evaluated the performance of some most commonly used classification ML algorithms such as NB, SVM, DT and RF on apache spark big data environment.

They measured time of detection, time of building and the time of prediction for network intrusion detection systems. They used UNSW-NB15 data-set for the purpose of performance evaluation and claimed that RF technique was outperformer with respect to specificity, accuracy, sensitivity, and also execution time among all the four other tested algorithms In (Slay N.M., 2015), the authors proposed a hybrid feature reduction approach based on attribute values' CP followed by an ARM.

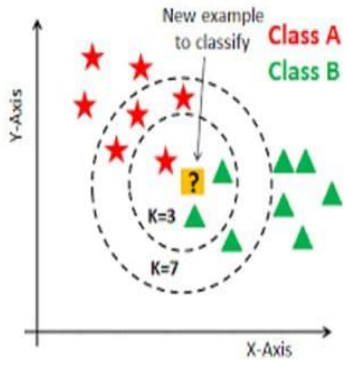
First, the dataset was splitted into partitions equally, the reason behind was to reduce the processing- time, then the output of CP technique was given as input to ARM to reduce number of feature. In the decision-engine expectation maximization clustering, logistic regression and naive bayes algorithms were employed for network intrusion detection to compare and evaluate the results. They claim

4. Classification Algorithms

Market analysis, science exploration, production control, and other applications can all benefit from the retrieved data. One of the key principles in the machine learning method is classification algorithms. They're used to sort unlabeled data into different categories. The following are the algorithms that were employed in the work: Support Vector Machine (SVM): When compared to other algorithms, SVM is one of the most reliable classification algorithms in machine learning since it offers a rapid and easy prediction process. It creates a hyperplane that separates the class labels into their associated classes by classifying data points based on support vectors in a data source.



K-Nearest Neighbor (KNN): is another reliable classification algorithm used for classifying data classes. One of its promising features is that it can be used for both classification regression purposes.



Naïve Bayes (NB): They are capable to forecast the probability that whether the given model fits to a particular class. It is based on Bayes' theorem. It constructed on the hypothesis that for instance, for aNgiven class, the attribute value is independent to the values of the attributes. This theory is called Class Conditional Independence.

$$P(H/X) = P(X/H).P(H)/P(X) \quad (1)$$

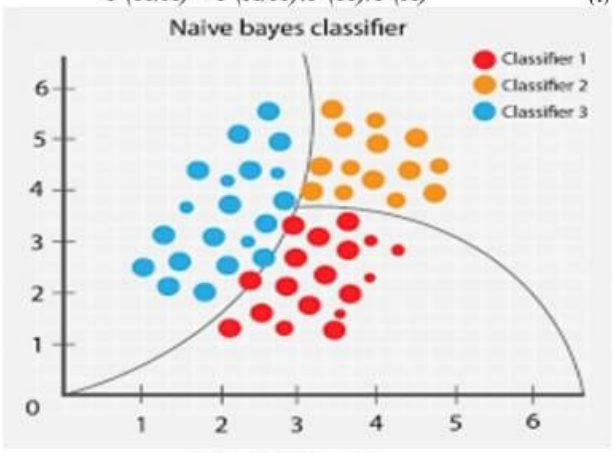


Fig. 5. Naïve Bayes

5. Methodology

Comparative analysis done between SVM KNN and Naïve Bayes for classification of dataset, to analyze their accuracy. At first raw dataset is taken and the class attribute contains 19 different types of attack which get labeled under 5 categories. They are normal, Dos, Probe, r2l and u2r. Figures

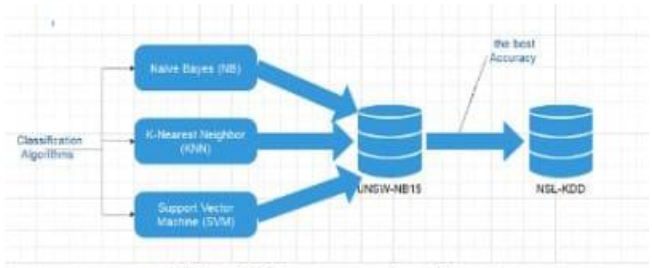


Fig. 6. Processes of testing

	Accuracy
KNN (k=3)	93.3333%
NB	95.55555%
SVM	97.77777%

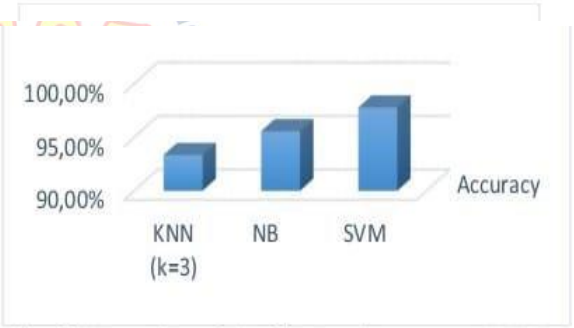


Fig. 7. Comparison of classifiers' performance on UNSW-NB15

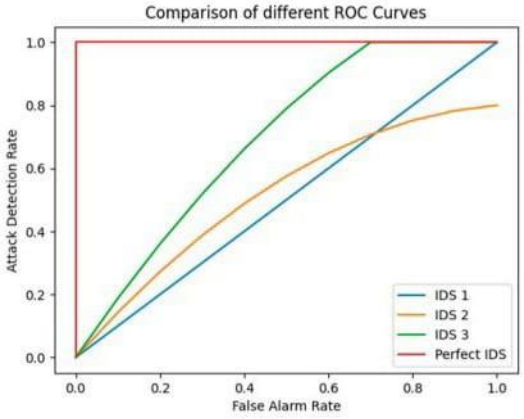
5.1. Precision Corresponds

To the ratio of correctly predicted attack samples to all the predicted attack samples. 3.2. Recall Precision = TP TP+FP Corresponds to the ratio of correctly predicted attack samples to all the samples that correspond to an attack. This metric is also known as the Detection Rate. TP Recall = TP+FN Ap

Since the attack detection rate and false alarm rate are often opposed to each other, evaluation of IDSs is also performed using Receiver Operating Characteristics (ROC) analysis. A ROC curve, as shown in Figure 1, represents the trade-off between attack detection rate and false alarm rate. The closer the ROC curve is to the top left corner the more effective the IDS is [4]. As shown

in Figure 1, ROC Curves can also be used to compare different IDS using the same dataset

IDSs can also be evaluated according to their time performance. The time performance corresponds to the total time that the IDS needs to detect an intrusion. This time is composed of the processing time and the propagation time. The closer the ROC curve is to the top left corner the more effective the IDS is [4]. As shown in Figure 1, ROC Curves can also be used to compare different IDS using the same dataset



The propagation time is the time needed to propagate the information to the security analyst or the Security Operation Centre (SOC). Both times, processing time and propagation time need to be as short as possible to allow security analysts to have enough time to react to an attack in real-time [4]. time needed by the IDS to process the information to detect an attack. The processing speed of the IDS needs to be as fast as possible, if not, real-time processing of intrusion is not feasible. The propagation time is the time needed to propagate the information to the security analyst or the Security Operation Centre (SOC). Both times, processing time and propagation time need to be as short as possible to allow security analysts to have enough time to react to an attack in real-time

6. Machine Learning

Machine learning is closely linked to Artificial Intelligence (AI) technology. It trains an algorithm to find regular patterns in a dataset. This training results in a model that can Machine learning is closely linked to Artificial Intelligence (AI) technology. It trains an algorithm to find regular patterns in a dataset. This training results in a model that can be used to predict or automate things. For IDSs, machine learning can be used to detect either known attacks or unknown attacks if the model has been sufficiently trained. be used to predict or automate things. For IDSs, machine learning can be used to detect either known attacks or unknown attacks if the model has been sufficiently trained. As shown in Figure 2, there are three main types of machine learning methods: supervised, unsupervised and semi-supervised machine learning [5]. These methods As shown in Figure 2, there are three main types of machine learning methods: su pervised, unsupervised and semi-supervised machine learning [5]. These methods are discussed further in this section. Supervised machine learning uses labelled data to generate a function that maps an input to an output. The function is constructed from labelled training data. Supervised machine learning uses labelled data to generate a function that maps an input to an output.

The function is constructed from labelled training data. Supervised learning can be categorized into two types of models, classification and regression: Supervised learning can be categorized into two types of models, classification and regression: Classification models are used to put data into specific categories. From a dataset it recognizes the category it belongs to based on its features. The model will be trained onlabelled input and output data to understand what the features of the input data are Classification models are used to put data into specific categories. From a dataset it recognizes the category it belongs to based on its features. The model will be trained on labelled input and output data to understand what the features of the input data so as to correctly classify it. Classification models are very useful to detect attacks.

For instance, for traffic coming into our network, a well-trained classification model can classify the traffic as normal traffic or as abnormal traffic. If the model performs well, the abnormal traffic can then be classified into subcategories of well-known attacks such as DoS, phishing, worms, port scan, etc. Common classification algorithms are decision tree, k-nearest neighbour, support vector machine, random forest, and neural network [6]. Regression models are used to predict continuous outcomes. The model is trained to understand the relationship between independent variables and a dependent variable.

Regression is used to find patterns and relationships in datasets that can then be applied to a new dataset.

Regression models are mainly used for forecasting the evolution of market prices or predicting trends [7]. Common regression algorithms are linear regression, logistic regression, decision tree, random forest, and support vector machine. One of the Supervised machine learning uses labelled data to generate a function that maps an input to an output.

The function is constructed from labelled training data. Supervised machine learning uses labelled data to generate a function that maps an input to an output. The function is constructed from labelled training data. Supervised learning can be categorized into two types of models, classification and regression: Supervised learning can be categorized into two types of models, classification and regression:

Classification models are used to put data into specific categories. From a dataset it recognizes the category it belongs to based on its features. The model will be trained onlabelled input and output data to understand what the features of the input data are

Classification models are used to put data into specific categories. From a dataset it recognizes the category it belongs to based on its features. The model will be trained on labelled input and output data to understand what the features of the input data so as to correctly classify it.

Classification models are very useful to detect attacks. For instance, for traffic coming into our network, a well-trained classification model can classify the traffic as normal traffic or as abnormal traffic. If the model performs well, the abnormal traffic can then be classified into subcategories of well-known attacks such as DoS, phishing, worms, port scan, etc

Common classification algorithms are decision tree, k-nearest neighbour, support vector machine, random forest, and neural network [6]. Regression models are used to predict continuous outcomes. The model is trained to understand the relationship between independent variables and a dependent variable

Regression is used to find patterns and relationships in datasets that can then be applied to a new dataset. Regression models are mainly used for forecasting the evolution of market prices or predicting trends [7]. Common regression algorithms are linear regression, logistic regression, decision tree, random forest, and support vector machine. One of the main advent controlled by a simple process control system assisted by a level switch connected to the floating gas holder and a hermetic compressor. The pressure swing that took place during mixing process enabled a self discharge system for digested slurry. A simple gravity filter was used for dewatering digested slurry and filtrate is safe to discharge into drainage.

Fig 2 High rate digestersusing oilcake

Fig 3 High rate digester using canteen waste

The present compact digesters available in India are just sealed tanks holding cow dung slurry with a floating gas holder .No scientific principles are applied in its engineering The present compact digesters available in India are just sealed tanks holding cow dung slurry with a floating gas holder. No scientific principles are applied in its engineering design in order to improve efficiency. Pre-fabricated compact digesters by M/s Biotech, Sintex, ARTI or many other similar agencies use different material of construction but basically design is same. Manual operations involved in its use make it inconvenient and it is also associated with serious operational problems. The basic design of it was first developed by Khadi and Village Industries Commission (KVIC) around 50 years ago and it is called KVIC digesters and popularly known as gober gas plant. Feeding these digesters is a difficult proposition because the waste needs to be fed through a 4 inches dia. pipe after external homogenisation using a grinder. Choking

is another major problem and surroundings become dirty due to overflow. Discharge of secondary effluents from the digester with high amount of solids is another difficulty. Loading rate in a conventional 1000 lit. digester is less than 2 kg/day and hence it is suitable only for a single household. Whereas potential customers of a compact digester are restaurants, apartments, canteens, housing colonies, small scale food processing units/centres and small agro farms. The prefabricated compact digester proposed here as a new concept is based on high rate biomethanation technology and it is designed to overcome the drawback of conventional digesters. The new equipment has almost 15 times more loading rate with increased treatment efficiency and short retention time compared to conventional digester. No operational difficulties are associated with this new variant of high rate compact digester.

This study utilized and aim to develop an high rate, two-stage reactor and phase-separation was employed to limit the reactions of anaerobic digestion to occur simultaneously in order to optimize the process. The same pilot-scale reactor is designed and fabricated, is installed at canteen to analyze the real time problem at user site.

6.2 BMP test procedure

The lab-scale experimental procedure is briefly illustrated in figure 3.1. Moreover, figure 3.2 clearly indicates the four substrates for BMP test with common blank reactors. Nevertheless, figure 3.3 illustrates the BMP test gas sampling.

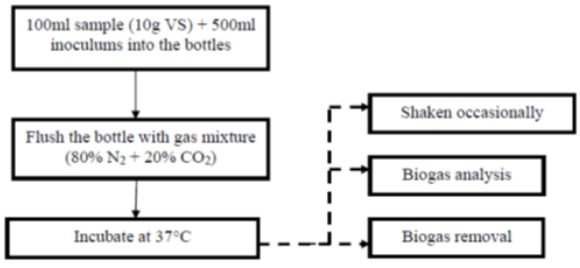


Fig 4.: Schematic representation of lab-scale (BMP test) procedure

6.3Mechanism of combined process:

The solid waste was digested in high-solids single-reactor in batch mode and undergoes a combined anaerobic digestion process followed by reaeration stage. It is said to be a combined process because all the reactions involved in anaerobic digestion takes place in single digester. Figure 3.4 illustrates the anaerobic digestion system used in this experiment.

As described, anaerobic digestion is the biomethanization of organic waste which can be roughly separated into two main steps, the pre-stage that involves liquefaction and acidification and the main stage that consist of methane generation. In order to enhance the process, phase separation and leachate recirculation is imperative. The following statement explains the concept of combined process:

Initially, flushing the fresh waste bed with water can wash-out the dissolve organic compounds and biodegradable materials present in the digester. Eventually, this could remove and reduce the possibility of VFA accumulation. The water produced after flushing the waste is known as leachate. Since this leachate will carry some fractions of VFA, it is said to be acidified.

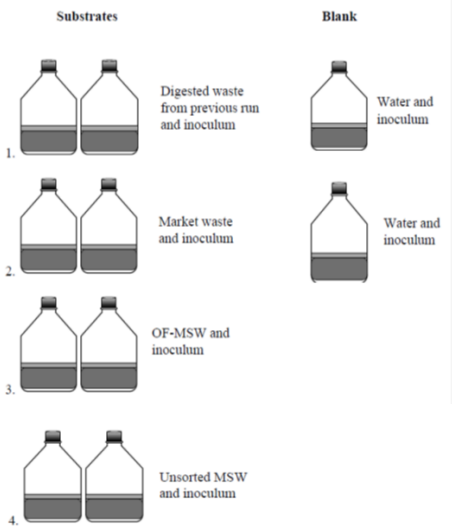


Fig 5 Different substrates used in BMP analysis.

Fig 6. General illustration of the reactor for BMP test sampling

6.4.Feed stock and inoculum preparation:

As the moisture content in organic fraction of oil cake was very high, the feed stock was comprise of solid and liquid phase. The amount of solid to be added was determined by the formulae

$$HRT = V/Q$$

HRT = Volume of the digester/feeding rate, where HRT and volume is fixed. The percentage of inoculum for the fermentation of solid waste is approximately 30%

(Carriero et al 2006). The cowdung was added to the reactor initially to enhance the start up process.

Fig 7. feed stock preparation

a) Feed with 10 % total solids:

2kg's Castor de-oiled cake was soaked in 18 liters of water over night to accomplish the first stage of digestion that is hydrolysis,so as to make the digestion a multistage process.

b) Feed with 12% total solids:2.4kg's castor de-oiled cake was soaked in 17.6 liters of water over night to accomplish the first stage of digestion that is hydrolysis,so as to make the digestion a multistage process.

c) Feed with 15% total solids:

3 kg castor de-oiled cake was soaked in 17 liters of water over night to accomplish the first stage of digestion that is hydrolysis, so as to make the digestion a multistage

7.RESULTS & DISCUSSION

7.1Biogas generation and quality:

One of the main ojective of this research was to determine the performance of the AD process when operated at different loading rates. Due to this, it is highly important to evaluvate the biogas production and methane content to various organic loading rate. The experimental results showed that during the loading rate of 2, the biogas production was very high as compared to 1 and the decline of biogas production started during the loading rate of 3 . The gas production was high when it started because of entrapped air inside the reactor, but the methane percentage was found to be very low. When the start –up reached methanogenesis, it was found to be high methane content. The biogas production increased with increase in organic loading rate to an extent and started declining after the optimum organic loading rate. The highest volume of biogas produced (1200 l/d) during the period of 15- 20 days and highest methane production (67%) was found to be these days. The biogas production rate fell after 20th day indicating exhausting of readily accessible substrate for biogas production. The reactor allowed to run until the gas production peaked and then dropped below 300 L per day. The biogas production of conventional digesters were also analyzed. The biogas production reached to the maximum of 500L in the case of conventional digesters. The fig.4.1 shows variation of biogas production at different loading rate.

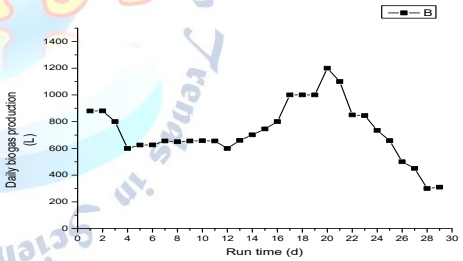


Fig 8 Daily biogas production of high rate digester

The biogas production of conventional digesters were also analyzed. The biogas production reached to the maximum of 500L in the case of conventional digesters. The fig. shows variation of biogas production of conventional digesters at different loading rate.

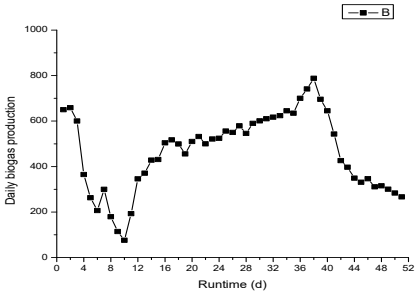


Fig9. Biogas production from conventional digester.

7.2METHANE PRODUCTION:

The concentration of methane in biogas was observed maximum of 67% during the period of 15th-20th day of digestion in the case of high rate digester. Intially the methane content was less due to the formation of acids but it gradually increased with in 3-5 days. Because of the less efficiency it was found that the percentage of methane present produced by conventional digesters are very less compared to the production of high rate digesters. The fig. 4.3 shows the analysis of variation in the production of methane.

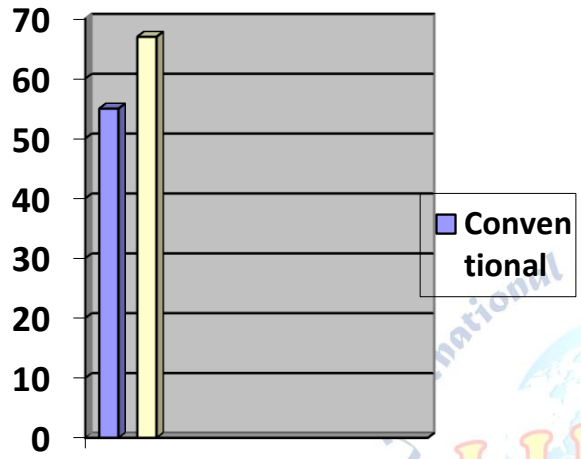


Fig: 10shows methane concentration of two reactors.

7.3 BMP test (lab-scale)

BMP test was conducted on four different substrates giving emphasize to the results of castrol oil cake and canteen waste. Other substrates such as stabilized (digested) oil cake and karanja oil cake were also included to investigate the methane potential of the waste.

It is important to emphasize that in this study, incubation at mesophilic temperature (37°C) for 100 days was found sufficient to ensure full degradation of degradable organic matter contained in the waste. The BMP test is not dependent on temperature, incubating at lower temperature may influence the incubation time but the BMP of the waste would just be the same.

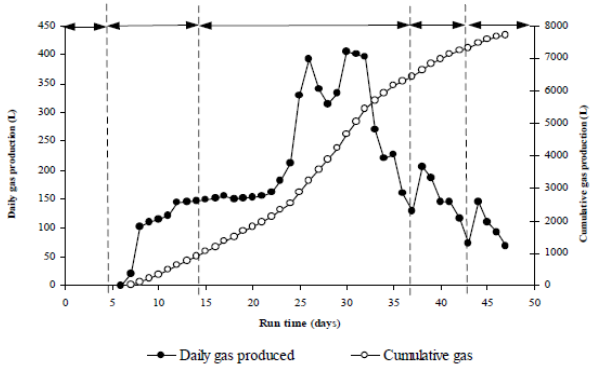
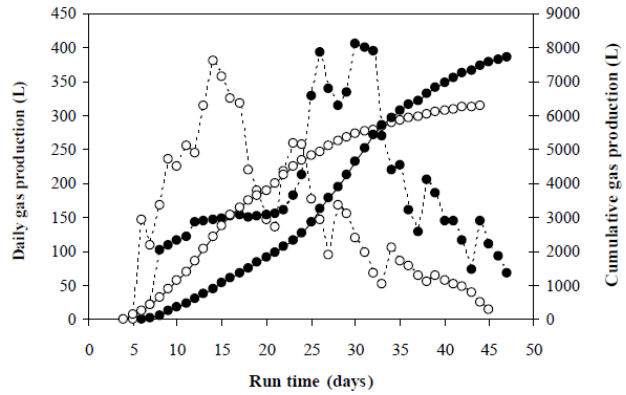
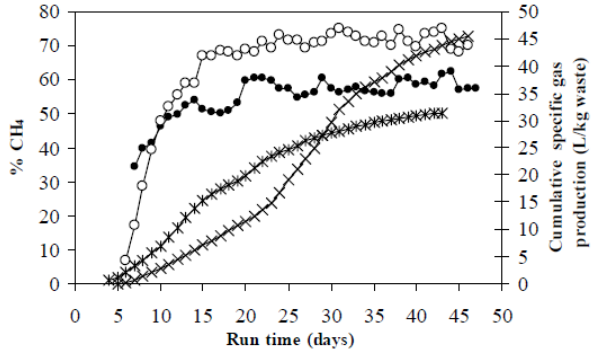


Fig 11 Daily and cumulative biogas produced



Castrol de-oiled cake –daily gas
canteen waste - daily gas
Castrol de-oiled cake –cumulative gas
canteen waste – cumulative gas

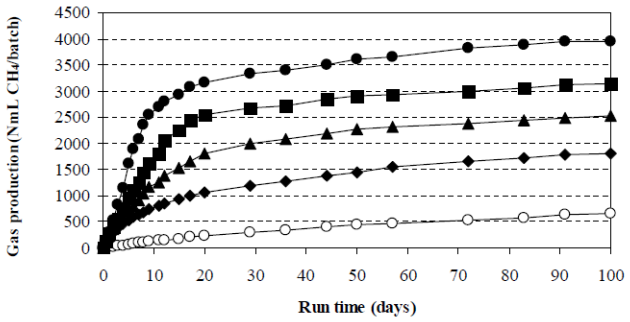
Fig 12 Biogas produced between Castrol de-oiled cake and canteen waste



% CH4 (canteen waste)
% CH4 (castrol de-oiled cake)
L/kg (canteen waste)
L/kg (castrol de-oiled cake)

Fig 13. variation between cantten waste and castrol de-oiled cake in terms of CH_4 content and total biogas produced per kilogram of waste

Fig13 Represent the methane potential of different waste together with the blank reactor. The cumulative volume of methane at STP is presented. The blank sample which includes water and inoculum represent the gas production produced by the inoculum itself. The methane production from the inoculum was subtracted from the methane production of the waste samples. The results thus represent only the methane production from the waste and not from the inoculum. Figure 4.38 Illustrates the corrected methane production generated by the waste sample in terms of milliliters of CH4/g VS at STP

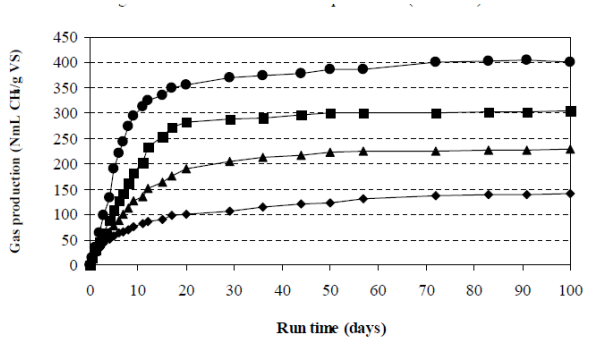


digested castrol de-oiled cake
castrol de-oiled cake
blank
canteen waste
karanja de-oiled cake

Fig 14 cumulative methane production(lab scale)

Digested de-oiled cake
canteen waste
castrol de-oiled cake
karanja de-oiled cake

Fig 15 corrected cumulative methane production(lab- scale run)



The final results after 100 days of mesophilic incubation showed that for each kg of VS of Castrol de-oiled cake, around 400 L of methane could be produced. Nguyen (2004) who conducted BMP test on fresh market waste obtained a value of around 300 L CH4/kg V Safter 40 days of incubation. The result was higher than the previous study and the difference may be attributed with the length of incubation time. It can be said that in order to ensure full degradation of waste, 100 days of incubation period is suitable and enough to determine the biochemical methane potential of the waste under mesophilic condition. Among the substrate used in this test, the castrol de-oiled cake showed the highest yield of methane. Also, the digested castrol de-oiled cake was included in this lab-scale run and implied that that around 35% of methane can be retained in the waste after digestion. However, it should be noted that the digested waste used here was obtained from the previous experiment . It is important to mention that more than 60% of the organicmatter can be converted into biogas, in which if the process conversion can be optimize, higher conversion efficiency can be achieved likewise.

Table1:Overall assessment of pre-stage(flushing and acidification)

Run no./Reactor no.	Hydrolysis yield (g C/kg TS)	% C removal into leachate	Acidification yield (g VFA/g TS)
Run 1: 5 days (600 L)			
R1: Ambient (30 mm)	143.2	33.3	251.6
R2: Mesophilic (30 mm)	169.4	39.4	313.4
R3: Thermophilic (30 mm)	180.5	42.0	355.4
Previous run: Ambient (60 mm)	129.2	30.0	193.2
Run 2: 3 days (360 L)			
R1 and R2: Thermophilic (ave.)	187.8	43.7	337.3
Run 3: 3 days (360 L)			
R3: OF-MSW	63.6	16.3	131.3

Table2 :Overall assessment of main stage

Run no./Reactor no.	Duration of operation (days)	Cumulative gas production (L)	% VS reduction	Methane yield (L CH_4/ kg VS)	CH_4 Conversion rate (L CH_4/kg VS. day)	% Mass reduction	% Volume reduction	Efficiency of process (%)
Run 1								
R1: Mesophilic	47	4719	65.2	285.7	6.1	66.0	44.3	71.4
R2: Mesophilic	47	3648	64.4	222.8	4.7	66.0	46.6	55.7
R3: Thermophilic	47	5367	71.2	319.7	6.8	74.0	58.2	80.0
60 mm (Mesophilic)	60	4706	61.3	266.8	4.5			66.7
Run 2								
R1: Thermophilic	45	6276	86.7	321.6	7.1	84.2	74.4	80.4
R2: SEBAC concept	40	5429	81.7	270.8	6.8	81.0	76.7	67.7
Run 3								
R1: SEBAC concept	28	6199	85.5	334.0	11.9	85.5	79.1	83.5
R3:	47	7728	71.1	229.3	4.9	53.7	49.0	76.4

Results and discussions from anaerobic wastewater treatment studies typically focus on several key aspects to evaluate the effectiveness and efficiency of the treatment process. Here's an outline of what would typically be included in the results and discussion sections:

8.RESULTS

High Accuracy:

1)Many studies demonstrate that machine learning algorithms can achieve very high accuracy in detecting intrusions, often exceeding 99%.

Ensemble Methods:

Ensemble methods, which combine multiple machine learning models, have shown strong performance in intrusion detection, achieving high accuracy across various datasets.

Deep Learning:

Deep learning approaches, like those using neural networks, have also been shown to be effective in intrusion detection, particularly in detecting complex attacks and anomalies.

Challenges:

Unknown Attacks: A significant challenge is the ability of IDS to detect previously unseen attacks, which require continuous adaptation and learning.

False Positives:

While accuracy is high, false positives (incorrectly flagging normal traffic as malicious) can still be an issue, requiring careful tuning of the models.

Data Imbalance: Real-world network traffic data often has a large imbalance between normal and malicious activities, which can affect the performance of some machine learning algorithms.

Specific Algorithms:

- Decision Trees and SVM: These are commonly used and can achieve high accuracy, with decision trees often favored for their interpretability.
- Random Forest: This ensemble method has also shown strong performance in intrusion detection, especially when combined with other algorithms.
- Genetic Algorithms: These algorithms can be used to optimize the rules for intrusion detection, potentially leading to higher accuracy.

Datasets:

- KDD Cup 99 and NSL-KDD: These are widely used benchmark datasets for evaluating intrusion detection systems.
- IoTID20 and UNSW-NB15: These datasets are specifically designed for evaluating intrusion detection systems in the Internet of Things (IoT) environment.

9.Future Work

Implementing an Intrusion Detection System (IDS) using Machine Learning (ML) requires a strategic approach. Start by collecting and preprocessing high-quality network traffic data. Select suitable ML algorithms and train models on labeled or unlabeled data, depending on the approach. Integrate the trained model into the IDS architecture, ensuring seamless data flow and real-time detection capabilities. Continuously monitor and update the model to adapt to evolving threats and network changes. Consider factors like computational resources, data storage, and alerting mechanisms to ensure effective deployment and maintenance of the ML-based IDS.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] M. Belouch, S. El Hadaj, and M. Idhammad. A two-stage classifier approach using reptree algorithm for network intrusion detection. International Journal of Advanced Computer Science and Applications, 8(6), pp.389-394 (2017)

[2] M. Belouch, S. El Hadaj, & M. Idhammad. Performance evaluation of intrusion detection based on machine learning using Apache Spark. Procedia Computer Science, 127, 1-6,(2018).

[3] N. Moustafa, N. (2017). Designing an online and reliable statistical anomaly detection framework for dealing with large high-speed network traffic(Doctoral dissertation, University of New South Wales, Canberra, Australia). (2017)

[4] W. Richert, L. P. Coelho, "Building Machine Learning Systems with Python", Packt Publishing Ltd., ISBN 978-1-78216-140-0

[5] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine learning techniques in cognitive radios," IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1136–1159, 2012.

[6] A. Iftikhar, M. Basheri, M. Javed Iqbal, A. Raheem; "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection",IEEE ACCESS, Survivability Strategies for Emerging Wireless Networks, 6 ,pp.33789-33795, (2018).