PySpark Orchestrated Machine Learning Paradigms for Advanced Network Intrusion Detection

T Harinadh[1], Pothuguntla Anusha[2], Roja D[3], Sai Srinivas Vellela[4], P.Muthukumar[5]

[1]Dept. of CSE-Data Science, Chalapathi Institute of Technology, Guntur-522016, A.P, India.
[1]Dept. of CSE, Chalapathi Institute of Technology, Guntur-522016, A.P, India.
[3,4]Dept. of CSE-Data Science, Chalapathi Institute of Technology, Guntur-522016, A.P, India.
[5]Dept. of EEE, Saveetha school of Engineering, Saveetha University, Chennai, India-602105
Email id: harinadh2497@gmail.com[1], anushapothuguntla08@gmail.com[2], rojad510@gmail.com[3], sais1916@gmail.com[4], muthukumarvlsi@gmail.com[5]

Article Received: 2 Aug 2025 *Article Accepted: 22 Sep 2025* *Article Published: 02 Oct 2025*

Citation
T Harinadh, Pothuguntla Anusha, Roja D, Sai Srinivas Vellela, P.Muthukumar (2025), "PySpark-Orchestrated Machine Learning Paradigms for Advanced Network Intrusion Detection", Journal of Next Generation Technology,5(6), pp. 11-22, October 2025.

Abstract

Cyber-attacks are on the rise, more and more, in this world full of Artificial Intelligence and Internet. It is very useful to have a ML model which can identify the network attack based on the network parameters. According to the proposed model in this paper, it is more accurate to detect intrusion in computer networks with a brand new and comprehensive method. The machine learning rules placed into the network data are established with the aid of shared computing power in PySpark. It goes a long way in helping to make a first move of putting things in order. For improving and refining the feature space two processes are used in an orderly manner; first is the Chi-Square Selector for choose important features and second is the Principal Component Analysis for minimizing the features. These methods are applied in sequences to enhance how well the machine learning models operate. This makes them more reliable and versatile in determining unauthorized attacks on the computer systems. It provides valuable contribution to the field of cyber security and Big Data processing. From the experiments. The model which gave the best results is the Multi- layer Perception model with an accuracy of 0.993.

Keywords: Text Bus Classifier algorithms, Feature, Feature Selectors, Feature Reducers, Principal Component Analysis, Chi Square Selector, PySpark, Machine learning classifications, Synthetic Minority Over-sampled Technique

I. Introduction

Security of the networks [1, 2, 3, 4] is now emerging as one of the concerns when the world is becoming more connected today. There is no day that computer network safety is threat- ened by bad activities from various sources. NID or Network Intrusion Detection is a very important defense system which should be in each organization. It seeks to identify and prevent intruders from accessing networks, identify abnormal behavior in the networks. As has been observed in Big Data Analytics (BDA) world, network data is more extensive and complex today. This is why there is need to develop new method of making it easy to know when an intruder is invading our

system in live time environment using robust approaches. This project focuses on the application of Machine Learning techniques on PySpark to tackle the issue of analyzing big labelled network data for identifying the appropriate location for situations to be classified as normal or strange.

This task also aims at exploiting the distributed nature of computer skills in PySpark in order to handle large datasets with labels appropriately. The analysis of network traffic details in this work employs some Machine Learning techniques such as Random Forest, Logistic Regression, and Gradient Boosting techniques. It wants to make strong models that detect the pop off events that look nothing like the usual behavior of the networks. Moreover, feature engineering methods for PySpark used in the work are also aimed to improve models for the given objects. This ensures that the intrusions in network are detected perfectly. This research employs PySpark a big data machine learning method for processing labeled data. Its goal is a more efficient and faster Network Intrusion Detection system within the context of Big Data Analytics. When the beginning of the project, we applied numerous sorts of apparatus from machine learning algorithms to the given data set. This enabled us to determine the effectiveness of basic class sorts with no additional assistance or enhancements. Net- work traffic data is analyzed through Random Forest, Logistic Regression and Gradient Boosting.

These algorithms assist in categorizing or sorting out various incidences according to the particular classification AIMD submits to the algorithms. We realized that the network data is several-tiered. Therefore, we included methods to lessen features so that our model could function optimally and at a faster pace. The reduction of features was applied as a strong PCA way in the PySpark place [5,6,7,8]. This was assisting us in filtering out all but necessity components of data and also helped in saving time for creating workload. Beginning with the first results from machine learning we embarked upon a step by step process. This included deflating features and rechecking our models one more time. The data set, not one but two times had features being dropped from it. This was done employing a procedure known as Proportionality of Covariance Analysis on each of those occasions.

This approach also assisted us in the determination of the right balance of how accurate our model is and how fast the processes take. We all know the role of features election in designing a model and in making a model to work much better. Therefore, to do just that, we adopted something called Chi-Squares elector in our plan. This maths way allowed us to select and preserve the best elements to stop bad attacks and make Machine Learning models more targeted. To handle scale and complexity of network datasets, we utilized the PySpark that is designed for big computers. PySpark unions benefited from faster and easier collaboration and machine learning. It ensured that large analysis can be made quickly from lots of information, labeled together in large analysis setup. This work is a comprehensive research on how the field of the ML domain identifies security issues. It integrates various approach, employs methods to strip down the characteristics it doesn't need and select the most suitable all under PySpark program frame. The following parts detail how, what and discuss that happens with each of them. They enable one to know the way our method is possible or good. The key contributions of the paper are: Distributed computing of big data and to perform complex computation for big data analysis the PySpark can be used. In the course of the project, Random Forest, Decision Tree, SVM, Naive Bayes, MLP Classifier were used in different capacities with aims to perform ensemble learning, construction of hyperplane, probabilistic classification and deep learning for classifying the data.

II. Literature Review

The current section presents the results from a survey limited to network intrusion

detection models which include both the ML and DL The main idea to enhance the strength of network security by enhancing the selection of features, the establishment of classifiers, and enhancing parameters to possibly get the best results. This is the case with simple maximization tactics, for example, searching for useful sets of features or filter options such as detailed designs of filters or wrappers. The research aims to address issues on problem of identifying network attacks by improving models and selecting significant measures. Liu and team have looked at how best to find and categories online attacks using Support Vector Machines (SVM) method and another one known as Principal Component Analysis [10]. This is under the activities of network traffic behavior. There is one ultimate aim and that is to sort out bad items right and have minimal false positives.

They created a model in a simpler method for Network Intrusion Detection Systems using only two divisions or categories which proved how effective their concept was by testing the same records set. All the main classification features are studied and tested in detail with the focus on utilization of the ML tools for security breaches detection. It was explained the above [11] concerning a smart system built on knowledge looks as, principally, the rule resolving part in such a system should be focalized at detection of unauthorized access. The application of intelligent approaches, such as those presented by expert systems, is illustrated. Each of the six points has a focus; one of the focuses is on the coordination between the Rule reading part and data. The paper also turn to deep learning model such as Convolution Neural Networks (CNN)to enhance accuracy and flexibility for auditing. The primary purpose is to improve security by SMART exploiting of the definition of hacking occurrences or the application of enhanced learning models. It also provides a suggestion for improving the LSTM neural network classification model [12].

This is done using the features that are chosen by QPSO to reduce the number of details required and create a thick feature list. The primary objective of the research work is to improve on it with effectiveness in detecting network attacks. This way is preferable to the other ways to find things, tested by higher F1-scores and lower false good rate as compared to benchmark on the real Internet traffic data. QPSO is used to select this relevant features and LSTM is used to do the classification. This pair makes identification of different threats that are characterized by distinct timing more effective and swift. The paper is concerning employment of new and zero-day cyber threats[13] Learning models. They propose a hybrid method which combines Simple Bayes, Support Vector Machine and Random Forest classifiers in an attempt to improve the discovery of attacks. It also helps prevent inside threats from happening. Thus, the main goal of this work is to enhance checking for intrusions based on the analysis of the user's forensic data and security, with a particular focus on the use of CRF and the spider monkey optimization methods. The decision as to which features are important is made using SMO, while the separations between the attacks and normal activity is done with the help of CNN.

This article discusses [14] one that is involved with working on making an intrusion detection system excellent. Thus, the main aim of this research is identify and prevent the network attacks in the shortest possible time. To this end, the study employs Random Forest as its primary technique of identifying unwanted access. Network data, and characteristics of intrusion episodes used in the model helps to identify potential threats to Networks. Finally, the study emphasizes the call for identifying and preventing wrong people to join network. As a very efficient approach, the Random Forest method is applied to this big security issue. Galatro and team have also tried to understand different techniques to come out with features to be selected for intrusion detection [15]. Its primary intended purpose is to provide a more careful and explicit look at a range of FS methods. The paper provides a synopsis on network traffic analysis and its security measures. It discusses the following FS algorithms' performance: speed;

connections between features; how long it takes to come up with the results that assist with managing networks/security.

A specific paper titled [16] examines how to enhance NIDS works with the help of different learning techniques from machines with SDN. The researcher painstakingly examines whether standard and efficient machine learning algorithms –Random Forest (RF), Support Vector Machine (SVM), Decision Tree, and Naïve Bayes algorithms, apply with the security in detection dataset. It has things like 'data count', 'data sweep', 'size of feature' and 'selection of features for learning to proceed well'. The aim is to achieve superior performance with lower dimensions and data samples. This study assists in selecting excellent algorithms suitable for various types of data, both large and small, and network response. The part taken out of the paper discusses various methods [17] to identify intrusions into the network. This includes analysis of attack on traffic rules plans and moving information patterns. It also provides the rationale for using weighted average in errors for classification of things. This reveals how punishing wrong alarms as well as detecting rates are distinct if all weights are equal. In addition, it refers to another research on web team learning and also an algebraic Ada boost method for network intrusion detection (Moshfeghi et al., 2014).

The paper also mentions about employing techniques such as CNN and FSL or the learning without much data for the systems that presuppose detecting of bad things occurred in computer networks. This also displays the output of tests that would substantiate these methods as very accurate in tracing hackers' attacks with a lot of definition. However, the paper also demonstrates how to enhance protection of the networks as well as lessen security concerns in current life. However, there is one important note: with deep learning- powered approaches in network management, malicious actors are particularly vulnerable to attacking systems. The new models which are being proposed as better are in fact much better than the older models for sensing an attack. They also describe what could be done in the future to accelerate and improve this method with a help of transfer learning.

Ahamad and Zeeshan have reviewed papers on systems that look for attacks in computer networks through machine learning and deep learning. Different studies are shown. The first one is recurrent neural networks, the second one is group methods improved with BAT algorithm optimization, the third one is non-symmetric deep auto encoder paired with random forest learning style and the last one is fast network study followed by particle swarm idea. Still, more types among those studied concern data mining algorithms called 'deep neural nets'. It also examines them merits and demerits of these ways in relation to the fact that they either get caught quicker or are trained faster but are not effective with types of attacks [18].

This article is about how to build a system to monitor computer security issues with machine learning, but emphasis on how to select good features fast. It describes how data should be prepared, how the Chi squared Test method should be used to select features, and how Intrusion Detection Systems can be developed using Advanced Security Network Metrics' corpus along side with both the Logistic Regression and Neural Networks algorithms. The unique feature of this paper is illus- trading how one can employ approaches for selecting relevant features. This makes the modeless complex and shortens the training time, to attempt to address a gap in writing from previous eras relative to machine learning approaches for datasets where additional information has been created and such purposes of feature selection are also employed as well. The survey discusses the ML and DL based network intrusion detection techniques, including SVM, PCA, CNN and random forest. There focus on feature selection methods, optimization algorithms such as QPSO and BAT and application of other models of learning with an aim of improving precision as well as speed in identifying cyber-attacks. The implications profiled in

this research stress the need to enhance network security through smart feature selection, model optimization, and the use of deep learning methodologies despite barriers such as vulnerabilities to attack of systems that capitalize on DL. The next section describes the proposed methodology where experiments ML models, feature selection by Chi square and feature reduction by PCA in applying classifiers.

III. Methodology

The proposed methodology is explained in detail in the following subsections and is presented in Fig.1.

A. Dataset Description

The dataset is sourced from Kaggle which contains 105000 rows and 42 columns. It provided an environment for obtaining raw TCP/IP dump data for a network by emulating a typical US Air Force LAN. The LAN was focused as a real setting and bombarded with many attacks. A connection is a sequence of TCP packets that begin and finish at a specific time interval, during which data travel to and from a source IP address to a target IP address using a well-defined protocol. Furthermore, each link is classified as either normal or an attack, with just one specific attack type. Each connection record consists of around 100 bytes.

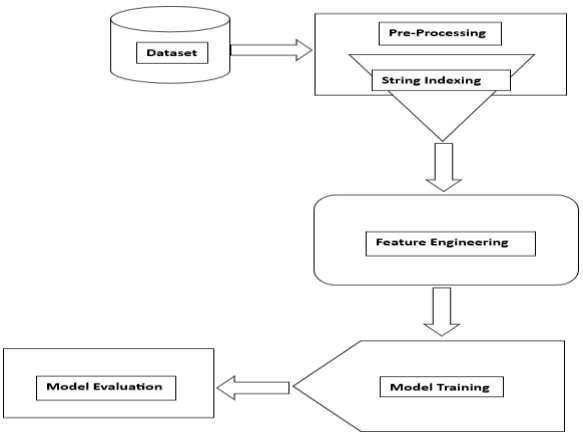

Figure 1: The Proposed Methodology of Network Intrusion Detection

B. Pre-Processing

The last step of the preprocessing stage was string indexing, which converts categorical features of the data set such as protocol_type, service, flag and class into numerical indices. This step was important especially for algorithms that needed numerical input data because it converted categorical data to annotated format used in machine learning models.

C. Feature Engineering

1) **Feature Selection using Chi-square Selector**: The selector is useful in eliminating non-diagnostic or noisy characteristics; consequently enhancing both the reliability and effectiveness of the-Disposition model since it selects characteristics with desirable relations with the target.

2) **Principal Component Analysis: PCA** is used in an effort to apply dimensionality reduction to the standardized features. Made a pipeline by including the avenues of feature assembling, scaling, and PCA. Fitted the model in the data and converted the data set into a data set containing PCA features.

D. Model Training

Divided the dataset into training and testing sets in the ratio 70:30. The classifiers experimented

are trained on the training data.

E. Classification

Naive Bayes: Naive Bayes is a predicting method of classifying items by using normal rules of mathematics with the principle that every feature is in dependent of the other. It is an odds based program that is frequently applied to text categorization, anti spam and all other tasks where it is possible to depict data with odds.

Support Vector Machines: SVM are a kind of learning technique which belongs to supervised learning technique for sorting and calculation of numbers. SVMs are really nice when the working space is high dimensional and widely used for machine learning and pattern recognition. SVM lies on the fundamental premise of identifying a line that efficiently classifies different datasets. Decision Tree: Decision Tree is a supervised learning methodology categorized into two data mining functions; classification and numeric prediction. By partitioning the data down to its subsets, over and over again, it achieves this. At each stage it focuses on which feature is the most important first then looks to see if that is predictive for the next step. The idea is to construct a tree which makes its decisions moving down the branches of the tree.

Random Forest: Random Forest is the group learning way which forms many decision trees throughout the training process. After passing the data through each tree it chooses the most frequently occurring resultant classification or mean output in the case of a regression task to provide its solution. Sometimes it can be a good concept, if it has to be that way because it is robust, and one of the most popular topics in machine learning.

Multi-Layer Perception: MLP is actually an imitation of the brain like structure, utilized in learning systems to make computers smarter. Feed forward neural networks are also known as MLPs. This means that the information flow is unidirectional only from the input portion through other hidden portions and then to the output portion.

F. Model Evaluation

In order to compare the performance of all algorithms, we compute classification measures including accuracy, precision, recall, and F1-score with respect to the testing dataset. Out of all the metrics, these gave an indication of how properly the model was able to classify instances.

IV. Results And Discussion

The experiments carried out included the evaluation of the contribution of each phase carried out. The first set of classifiers is then trained on the preprocessed data and the performance of the classifiers is presented in Table 1 below. From Table 1, the Random Forest Classifier appeared to be the best with the F1 score at 98.47%. Accuracy of Decision Tree model has is almost in par with F1 score of 98.26%. Consequently, the accuracies and F-measures obtained by the experimented models are graphically depicted in Graph 2 and Graph 3 respectively below.

Algorithm	Accuracy	Precision	Recall	F1
NaiveBayes	83.56	84.52	83.56	83.56
SVM	95.35	95.36	95.35	95.34
DecisionTree	98.26	98.26	98.26	98.26
RandomForest	98.47	98.51	98.47	98.47
MLP	96.34	96.37	96.34	96.34

Table 1: Evaluation of Classifiers in Terms of the performance when applied on the Pre-processed data

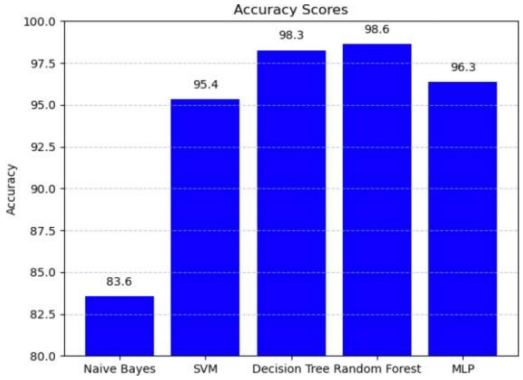

Figure 2: Display of Accuracy Scores of the Models That Are Employed on Data That Has Undergone Pre-Processing

For the next set of experiments, the classifiers are trained on applying feature selection (Chi-Square) by choosing finest 25featuresaspresentedintheTable.2.Besides, MLP and Decision Tree algorithms produced highest F1-score of 99.24%. By applying the feature selection procedure performance increased compared to the data provided in Table.1 Figure 4.

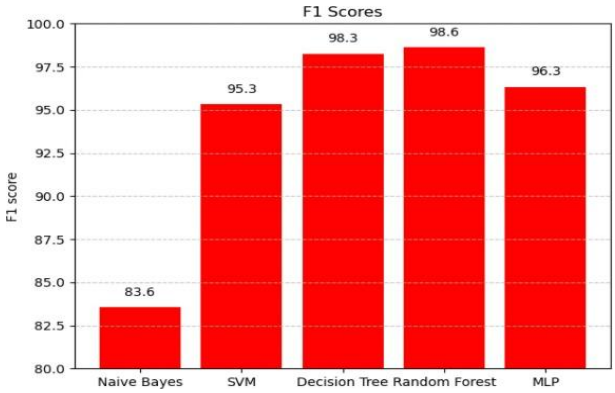

Figure 3: The F1 scores of the pre-processed data are presented in the following figure below

Table 2: The following looks at the performance metrics of classifiers in use after applying feature selection, specifically the Chi Square Selector.

Algorithm	Accuracy	Precision	Recall	F1
NaiveBayes	65.25	72.00	65.25	60.58
SVM	94.43	94.43	94.43	94.43
DecisionTree	97.70	97.75	97.70	97.71
RandomForest	98.43	98.45	98.43	98.44
MLP	96.08	96.14	96.08	96.08

Below represents the accuracies achieved by the experimented modelswhilefigure5representstheF-measuresofthevarious models.

The subsequent experiments involved Feature Reduction (PCA) for dimensionality, on preprocessed data, depicted in table three below. By comparing the metrics of each of the models implemented, Random Forest demonstrated the maximumscoreof98.44%ofF1-score. Please note that Naive Bayes cannot be used for the data after performing PCA as it gives Negative values. Naive bayes is based on probabilistic Functions. The accuracies and F-measures obtained by the experimented models are presented graphically in Fig.6 and Fig.7.

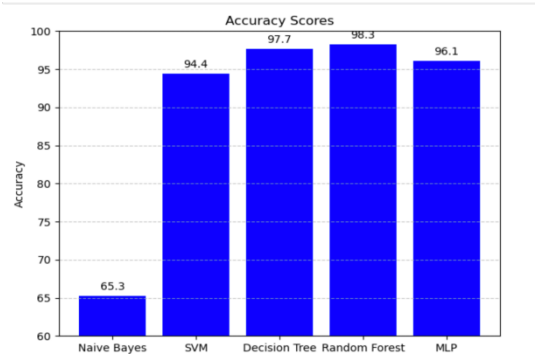

Figure 4: Doing the feature selection using chi square selector the following is the graph representing the accuracy of the models.)

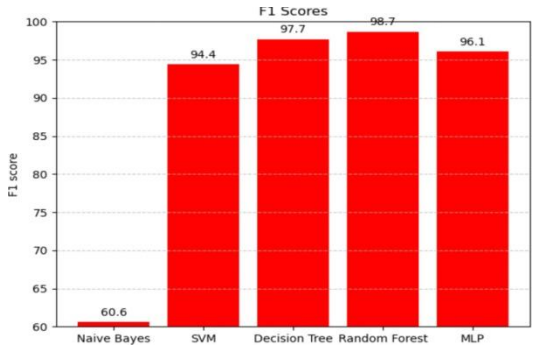

Figure 5: Bar Chart of F1 Scores of the models after applying Feature Selection (Chi Square Selector)

Table 3: The Performance Observations after applying feature Reduction on Classifiers

Algorithm	Accuracy	Precision	Recall	F1Score
SVM	94.35	94.35	94.35	94.35
DecisionTree	95.73	95.73	95.73	95.73
RandomForest	96.28	96.47	96.28	96.27
MLP	99.30	99.30	99.30	99.30

Preprocessed data, depicted in table three below. By comparing the metrics of each of the models implemented, Random Forest demonstrated the maximum score of 98.44% of F1-score. Please note that Naive Bayes cannot be used for the data after performing PCA as it gives Negative values. Naive bayes is based on probabilistic Functions Which contains so many pros but one main con is it doesn't support Negative values. The accuracies and F-measures obtained by the experimented models are presented graphically in Fig.6 and Fig.7.

Table 4: Performance Metrics of Machine Learning Algorithms

Algorithm	Accuracy	Precision	Recall	F1Score
SVM	94.21	94.21	94.21	94.21
Decision Tree	99.24	99.24	99.24	99.24
Random Forest	96.58	96.68	96.58	96.57
MLP	99.24	99.24	99.24	99.24

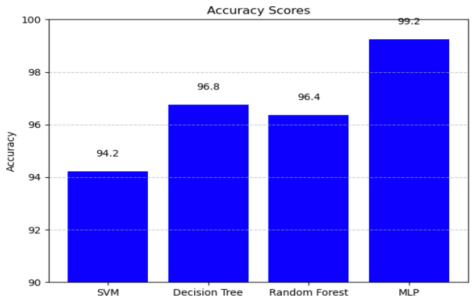

Figure 6: Accuracy scores of the models after performing Feature Reduction Activity pictorial representation

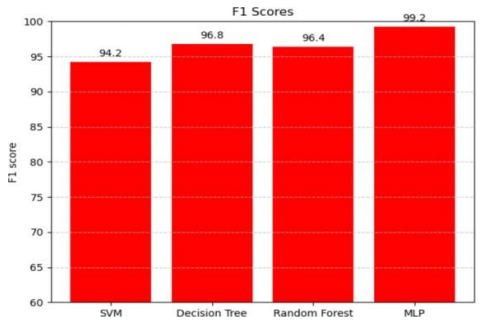

Figure 7: Comparison of F1 Scores of the models after Feature Reduction (PCA)

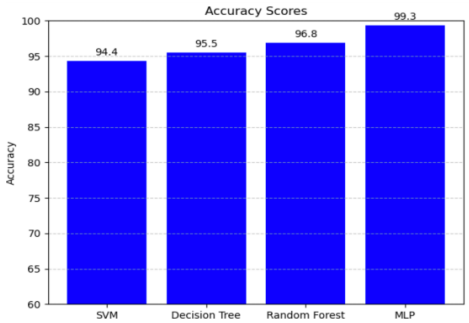

Figure 8: Accuracy of selected models after the feature selection followed by feature reduction has been as follows

Figure 9: The next part of the study involves graphical representation of the F1scores of the models after the feature selection followed by feature reduction.

V. CONCLUSION

There are several ML models that were tested in the proposed model identification process and the feature selection and feature reduction methods were also evaluated which made the identification of the best suitable model, Multi-Layer Perception and it not only gave F-measure

and accuracy of 99.3% but also incorporated Chi-Square based feature selection and PCA based feature reduction. The results have evaluated and confirmed the placement of feature selection, and feature reduction methods in the ML pipeline the model that has been proposed here can be extended with Deep Learning models as well as the ensemble model. This means that the proposed model can be easily implemented using real world network environments and in addition, it can perform comprehensive evaluation and validation of the IDS.

References

[1] Mandava, R., Vellela, S. S., Malathi, N., Haritha, K., Gorintla, S., & Dalavai, L. (2025, May). Exploring the Role of XAI in Enhancing Predictive Model Transparency in Healthcare Risk Assessment. In 2025 International Conference on Computational Robotics, Testing and Engineering Evaluation (ICCRTEE) (pp. 1-5). IEEE.

[2] Mandava, R., Vellela, S. S., Gorintla, S., Dalavai, L., Malathi, N., & Haritha, K. (2025, May). Evaluating the Impact of Explainable AI on User Trust in Financial Decision-Support Systems. In 2025 International Conference on Computational Robotics, Testing and Engineering Evaluation (ICCRTEE) (pp. 1-6). IEEE.

[3] Vellela, S. S., & Balamanigandan, R. (2024). Optimized clustering routing framework to maintain the optimal energy status in the wsn mobile cloud environment. Multimedia Tools and Applications, 83(3), 7919-7938.

[4] Polasi, P. K., Vellela, S. S., Narayana, J. L., Simon, J., Kapileswar, N., Prabu, R. T., & Rashed, A. N. Z. (2024). Data rates transmission, operation performance speed and figure of merit signature for various quadurature light sources under spectral and thermal effects. Journal of Optics, 1-11.

[5] Biyyapu, N., Veerapaneni, E. J., Surapaneni, P. P., Vellela, S. S., & Vatambeti, R. (2024). Designing a modified feature aggregation model with hybrid sampling techniques for network intrusion detection. Cluster Computing, 27(5), 5913-5931.

[6] Vuyyuru, L. R., Purimetla, N. R., Reddy, K. Y., Vellela, S. S., Basha, S. K., & Vatambeti, R. (2025). Advancing automated street crime detection: a drone-based system integrating CNN models and enhanced feature selection techniques. International Journal of Machine Learning and Cybernetics, 16(2), 959-981.

[7] Vullam, N. R., Geetha, G., Rao, N., Vellela, S. S., Rao, T. S., Thommandru, R., & Rao, K. N. S. (2025, February). Optimized Multitask Scheduling in Cloud Computing Using Advanced Machine Learning Techniques. In 2025 International Conference on Intelligent Control, Computing and Communications (IC3) (pp. 410-415). IEEE.

[8] Reddy, N. V. R. S., Chitteti, C., Yesupadam, S., Desanamukula, V. S., Vellela, S. S., & Bommagani, N. J. (2023). Enhanced speckle noise reduction in breast cancer ultrasound imagery using a hybrid deep learning model. Ingénierie des Systèmesd'Information, 28(4), 1063-1071.

[9] Vellela, S. S., & Balamanigandan, R. (2023). An intelligent sleep-awake energy management system for wireless sensor network. Peer-to-Peer Networking and Applications, 16(6), 2714-2731.

[10] Vellela, S. S., & Balamanigandan, R. (2024). An efficient attack detection and prevention approach for secure WSN mobile cloud environment. Soft Computing, 28(19), 11279-11293.

[11] Vellela, S. S., Roja, D., Purimetla, N. R., Thalakola, S., Vuyyuru, L. R., & Vatambeti, R. (2025). Cyber threat detection in industry 4.0: Leveraging GloVe and self-attention mechanisms in BiLSTM for enhanced intrusion detection. Computers and Electrical Engineering, 124, 110368.

[12] Vellela, S. S., Malathi, N., Gorintla, S., Priya, K. K., Rao, T. S., Thommandru, R., & Rao, K. N. S. (2025, March). A Novel Secure and Scalable Framework for a Cloud-Based Electronic Health Record Management System. In 2025 3rd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT) (pp. 131-135). IEEE.

[13] Praveen, S. P., Nakka, R., Chokka, A., Thatha, V. N., Vellela, S. S., & Sirisha, U. (2023). A novel classification approach for grape leaf disease detection based on different attention deep learning techniques. International Journal of Advanced Computer Science and Applications (IJACSA), 14(6), 2023.

[14] Vellela, S. S., Rao, M. V., Mantena, S. V., Reddy, M. J., Vatambeti, R., & Rahman, S. Z. (2024). Evaluation of Tennis Teaching Effect Using Optimized DL Model with Cloud Computing System. International Journal of Modern Education and Computer Science (IJMECS), 16(2), 16-28.

[15] Vellela, S. S., Vullam, N. R., Gorintla, S., Rao, T. S., & Harinadh, T. (2025, July). Exploring the Anti-Inflammatory Potential of Green-Synthesized Pyrazolines. In 2025 6th International Conference on Data Intelligence and Cognitive Informatics (ICDICI) (pp. 814-819). IEEE.

[16] Vellela, S. S., Manne, V. K., Trividha, G., Chaithanya, L., & Shaik, A. (2025). Intelligent Transportation Systems AI and IoT for Sustainable Urban Traffic Management. Available at SSRN 5250812.

[17] Vellela, S. S., Singu, K., Kakarla, L. S., Tadikonda, P., & Sattenapalli, S. N. R. (2025). NLP-Driven Summarization: Efficient Extraction of Key Information from Legal and Financial Documents. Available at SSRN 5250908.

[18] Vuyyuru, L. R., Purimetla, N. R., Reddy, K. Y., Vellela, S. S., Basha, S. K., & Vatambeti, R. (2025). Advancing automated street crime detection: a drone-based system integrating CNN models and enhanced feature selection techniques. International Journal of Machine Learning and Cybernetics, 16(2), 959-981.

[19] Dalavai, L., Purimetla, N. M., Vellela, S. S., SyamsundaraRao, T., Vuyyuru, L. R., & Kumar, K. K. (2024, December). Improving Deep Learning-Based Image Classification Through Noise Reduction and Feature Enhancement. In 2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA) (pp. 1-7). IEEE.

[20] Haritha, K., Vellela, S. S., Vuyyuru, L. R., Malathi, N., & Dalavai, L. (2024, December). Distributed Blockchain-SDN Models for Robust Data Security in Cloud-Integrated IoT Networks. In 2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS) (pp. 623-629). IEEE.

[21] Madhuri, A., Jyothi, V. E., Praveen, S. P., Sindhura, S., Srinivas, V. S., & Kumar, D. L. S. (2024). A new multi-level semi-supervised learning approach for network intrusion detection system based on the ‘goa’. Journal of Interconnection Networks, 24(supp01), 2143047.

[22] Biyyapu, N., Veerapaneni, E. J., Surapaneni, P. P., Vellela, S. S., & Vatambeti, R. (2024). Designing a modified feature aggregation model with hybrid sampling techniques for network intrusion detection. Cluster Computing, 27(5), 5913-5931.

[23] Praveen, S. P., Sarala, P., Kumar, T. K. M., Manuri, S. G., Srinivas, V. S., & Swapna, D. (2022, November). An adaptive load balancing technique for multi SDN controllers. In 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS) (pp. 1403-1409). IEEE.

[24] Vellela, S. S., & Balamanigandan, R. (2022, December). Design of Hybrid Authentication Protocol for High Secure Applications in Cloud Environments. In 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS) (pp. 408-414). IEEE.

[25] Vellela, S. S., Balamanigandan, R., & Praveen, S. P. (2022). Strategic survey on security and privacy methods of cloud computing environment. Journal of Next Generation Technology, 2(1).

[26] SrinivasVellela, S., Praveen, S. P., Roja, D., Krishna, A. R., Purimetla, N., Rao, T., & Kumar, K. K. (2024, April). Fusion-Infused Hypnocare: Unveiling Real-Time Instantaneous Heart Rates for Remote Diagnosis of Sleep Apnea. In 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS) (Vol. 1, pp. 1-5). IEEE.

[27] Vellela, S. S., & Krishna, A. M. (2020). On Board Artificial Intelligence With Service Aggregation for Edge Computing in Industrial Applications. Journal of Critical Reviews, 7(07).

[28] Praveen, S. P., Vellela, S. S., & Balamanigandan, R. (2024). SmartIris ML: harnessing machine learning for enhanced multi-biometric authentication. Journal of Next Generation Technology (ISSN: 2583-021X), 4(1).

[29] Vellela, S. S., Chandra, S. S., Thommandru, R., Mastan Basha, S., & Sri Ram, D. (2023). Novel Approach to Mitigate Starvation in Wireless Mesh Networks. Available at SSRN 5262254.

[30] Vellela, S. S., Singu, K., Kakarla, L. S., Tadikonda, P., & Sattenapalli, S. N. R. (2025). NLP-Driven Summarization: Efficient Extraction of Key Information from Legal and Financial Documents. Available at SSRN 5250908.