IJARCCE A. ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023

DOI: 10.17148/IJARCCE.2023.12825

A Novel Intrusion Detection System for Wireless Enterprise Networks using Ensembled Machine Learning Models

Gururaja H S[1], M Seetha[2]

Assistant Professor, Dept. of ISE, B.M.S. College of Engineering, Bangalore, India[1]

Professor and Head, Dept. of CSE, G.Narayanamma Institute of Technology and Science, Hyderabad, India[2]

Abstract: For contemporary enterprises, the significance of wireless enterprise networks has expanded due to their heightened adaptability and mobility in terms of connectivity and access to information. However, these networks are also vulnerable to various forms of cyber-attacks, including intrusions from external parties, breaches of data, and outbreaks of malware. To counter these threats, it is imperative to possess efficient intrusion detection systems (IDSs). One potential strategy to enhance the performance of IDSs for wireless enterprise networks is the utilization of ensembled machine learning models. To construct an IDS model in a wireless environment employing the AWID dataset, this investigation integrates the prognostications of three distinct classification techniques: specifically, the hybrid model of CNN-SVM, the ensemble model of SVM-MLP, and the ensemble model of DT-KNN. The efficacy of the model is assessed based on the statistical information derived from the confusion matrix, such as accuracy, recall, precision, and F1-scores.

Keywords: Intrusion Detection System (IDS); Machine Learning; Ensemble models; AWID.

I. INTRODUCTION

In the contemporary era, a substantial amount of sensitive data is transmitted via various devices like smartphones and computers. As computer network technology progresses, concerns about security are becoming more frequent and cannot be disregarded. Adversaries continue to develop novel methods of attack, underscoring the importance of safeguarding our data. The role of Intrusion Detection Systems (IDS) is pivotal in protecting our information. Datasets consist of instances that possess multiple relevant features for the intrusion detection system. The widespread utilization of networking and internet systems has resulted in a growing prevalence of security issues, posing a significant obstacle for ordinary users, organizations, enterprises, and government agencies.

Attackers frequently devise new attack techniques that often escape the awareness of network administrators and security software. The widespread usage of networking and internet technologies has led to a heightened frequency of security concerns, imposing a substantial burden on the general populace, organizations, businesses, and governmental entities. On the internet, countless cyber security attacks have inflicted financial losses, damaged reputations, and disrupted business operations.

To ensure the privacy, reliability, and accessibility of numerous users and organizations, as well as to shield them from potential threats, the presence of a robust and dependable security system is indispensable. The monitoring and supervision of internet communication and data transfer have become integral components of contemporary services. Intrusion Detection Systems (IDSs) serve as crucial frontline defenders, providing a secondary line of defense against intruders. Given the escalating prevalence of attacks, the adoption of new detection techniques, particularly those employing machine learning (ML) for intrusion detection and prevention systems, is imperative.

II. BACKGROUND

With the widespread utilization of internet networks, ensuring information security has become a significant concern for both organizations and regular users. Safeguarding network communication devices against various threats and attacks is considered an urgent responsibility for administrators of networking systems. Various techniques are employed to establish secure communication and protect organizations' privacy, including cryptography, firewalls, and access control. However, attackers constantly adapt and devise innovative methods to breach system security. Consequently, Intrusion Detection Systems (IDS) play a vital role in safeguarding networks by upholding their confidentiality, integrity, and availability for authorized users. IDS can be implemented through hardware or software to automate intrusion detection

DOI: 10.17148/IJARCCE.2023.12825

processes. Depending on the settings and configurations, IDS continuously monitors system conditions and generates alerts to notify system administrators of potential attacks. The monitoring process covers incoming and outgoing data, ensuring efficient detection of suspicious activities and optimal security throughout the networking system. Numerous tools, such as firewalls, intrusion prevention systems, and IDS, have been developed to defend against internet-based threats. Machine Learning (ML), a subfield of artificial intelligence (AI), utilizes training data based on known facts to enhance its learning capabilities [1,2,3].

An IDS typically creates and sends alert signals for any illegal situations the networking system may be exposed to, such as unlawful emails, audio messages, and video messages. Its job is to look for unusual patterns in IP (Internet Protocol) packets as they travel over the network, gather information on attacks, and use countermeasures to stop them (if the availability of detection and prevention technology exists).

III. RELATED WORK

Extensive investigation has been carried out regarding the implementation of machine learning techniques in the identification of unauthorized access within wireless enterprise networks. Ensembled machine learning models have garnered considerable attention among these techniques due to their capability to augment the accuracy and efficiency of intrusion detection systems (IDS).

One strategy entails the utilization of decision tree ensembles, which amalgamate the forecasts from numerous decision trees to enhance the overall performance of IDS. Decision trees are a prevalent machine learning approach that classifies data based on a collection of decision rules. They offer the benefits of interpretability and swift training, but they may encounter challenges associated with overfitting when confronted with intricate data. By employing ensembled decision trees, the risk of overfitting can be mitigated as the forecasts of multiple trees are merged, thus yielding a more resilient IDS.

Another widely applied method is the application of random forest ensembles, which serve as a type of decision tree ensemble. Random forests employ a random subset of features to train each tree, thereby diminishing the likelihood of overfitting and enhancing the generalization performance of the IDS. Random forest ensembles have evidenced their efficacy in diverse IDS applications, encompassing network intrusion detection, malware detection, and spam filtering.

Boosting ensembles represent another form of machine learning model that has been employed for IDS in wireless corporate networks. In these models, a robust learner that can precisely categorize the data is constructed by stacking a series of weak learners, each of which is trained on a distinct subset of the data. IDS applications have exhibited the effectiveness of boosting algorithms such as Adaboost and Gradient Boosting, particularly when coupled with weak learners like decision trees.

In the study mentioned in reference [4], the author introduced a framework that is grounded in agent technology for the purpose of identifying harmful behavior. Within this framework, the actions of intruders are detected by means of utilizing an Artificial Neural Network (ANN). Experiments were conducted on the AWID-CLS-R subset in order to classify each incident as either normal or indicative of an attack. The results of these experiments revealed that the proposed framework achieved a precision rate of 99.3 percent when applied to the AWID-CLS-R subset. Furthermore, in reference [5], the same subset was utilized for a multi-class classification experiment, in which the author employed a deep learning approach with the objective of enhancing the overall accuracy to 98.67 percent.

With regard to reference [6], the author's focus was on the classification of attacks on a subset of AWID-CLS-R through the use of eight conventional supervised machine learning classifiers. In order to train these classifiers, the author consolidated a total of 20 features into a single feature and manually selected the pertinent features. Various algorithms, including AdaBoost, OneR, J48, Naive Bayes, Random Forest, ZeroR, and Random Tree, were employed to assess the performance of attack classification. The study was successful in raising the overall precision from 89.43 percent to 96.2 percent.

In the study mentioned in reference [7], a proposed framework utilized the Stack Auto Encoder (SAE) for active assault detection. The SAE, being an unsupervised learning strategy, was employed for the purpose of feature selection. By incorporating a regression layer, supervised learning technique, and SoftMax activation function, the framework was able to achieve a maximum accuracy of 97.7%. Additionally, the research aimed to identify the most effective feature among three different machine learning methods. Moreover, in references [8,9], the emphasis was placed on enhancing impersonation attack detection through the utilization of the AWID-CLS-R subset. This involved eliminating two out of

DOI: 10.17148/IJARCCE.2023.12825

the four classes within the subset, resulting in the presence of the impersonation attack class and the normal traffic class. To classify the attacks, the study employed the Artificial Neural Network (ANN) and leveraged the Decision Tree and Support Vector Machine (SVM) techniques. As a result, an accuracy of 99.86% was achieved in detecting impersonation attacks.

Furthermore, in references [10,11], the author conducted an analysis on the reduced classification version of the CLS and ATK class subsets using five different supervised machine learning classifiers. The study employed algorithms such as AdaBoost, Random Forest, Random Tree, OneR, and J48. Prior to applying these classifiers, the features were thoroughly examined and ranked using Information Gain and Chi-Square. The classifiers were then applied to the respective subgroups, yielding optimal accuracies with 41 features. The findings revealed that the Random Tree classifier achieved an accuracy of 95.12% on the AWID-CLS-R subset, while the Random Forest classifier achieved an accuracy of 94.97% on the AWID-ATK-R subset. Furthermore, it was observed that reducing the number of features to a specific limit led to an improvement in the accuracy of the models.

In reference [12], the proposal of TermID, a distributed network intrusion detection system, was made. This particular approach was developed with the intention of increasing productivity while also ensuring the confidentiality of private data. The development process involved the utilization of Classification Rule Induction (CRI) and Swarm Intelligence Optimization (SIP) to create a useful model. This model consists of two functional components, namely the central node and the monitoring node. The AWID-ATK-R subset was carefully considered and physically divided for each respective node. It is worth noting, however, that the accuracy of the system was not made public by the author.

Reference [13] discusses the implementation of an ensemble learning algorithm technique that makes use of the AWID-CLS-R dataset for the purpose of multi-class classification. The author achieved a highly commendable accuracy rate of 95.88% in this particular study. Additionally, an interesting observation was made when the assault classes were merged into a single class, resulting in an accuracy of approximately 99.11%. However, it is important to highlight that distinguishing between impersonation and injection attacks proved to be quite challenging and led to a decrease in accuracy. To tackle this issue, a novel machine learning model was employed to effectively differentiate between these specific types of assaults.

Reference [14] presents a classification framework that aims to distinguish between difficult and easier samples. The Wireless Network Intrusion Detection System (WIDS) implemented within this framework demonstrated an impressive accuracy rate of 98.54% for multi-class classification and 99.54% for binary class classification through the application of deep learning techniques.

Machine learning techniques played a significant role in the majority of the aforementioned studies, resulting in distinct outcomes during their respective result cycles. Consequently, a number of relevant works centered around the AWID dataset were presented. In conclusion, ensembled machine learning models possess considerable potential to greatly enhance the performance of Intrusion Detection Systems (IDS) in wireless enterprise networks.

IV. METHODOLOGY

The primary aim of this research is to create a robust intrusion detection system (IDS) specifically designed for wireless environments. This will be accomplished by utilizing ensembled machine learning (ML) models to enhance the detection of attacks compared to using a single model. The research methodology is depicted in Figure 1, consisting of five distinct steps: preprocessing, feature selection, identification of attack types, deployment in a client-server environment, model evaluation, and comparison with state-of-the-art techniques.

The research objectives associated with the main aim include the following: analyzing and addressing the limitations of the AWID dataset, employing feature selection techniques to mitigate overfitting and enhance the performance of the models, implementing and analyzing various ML techniques for IDS, constructing a reliable IDS specifically tailored for wireless environments using ensembled ML models to improve attack identification, and finally, comparing and contrasting the outcomes of the new enhancements with existing approaches.

IJARCCE B. ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023

DOI: 10.17148/IJARCCE.2023.12825

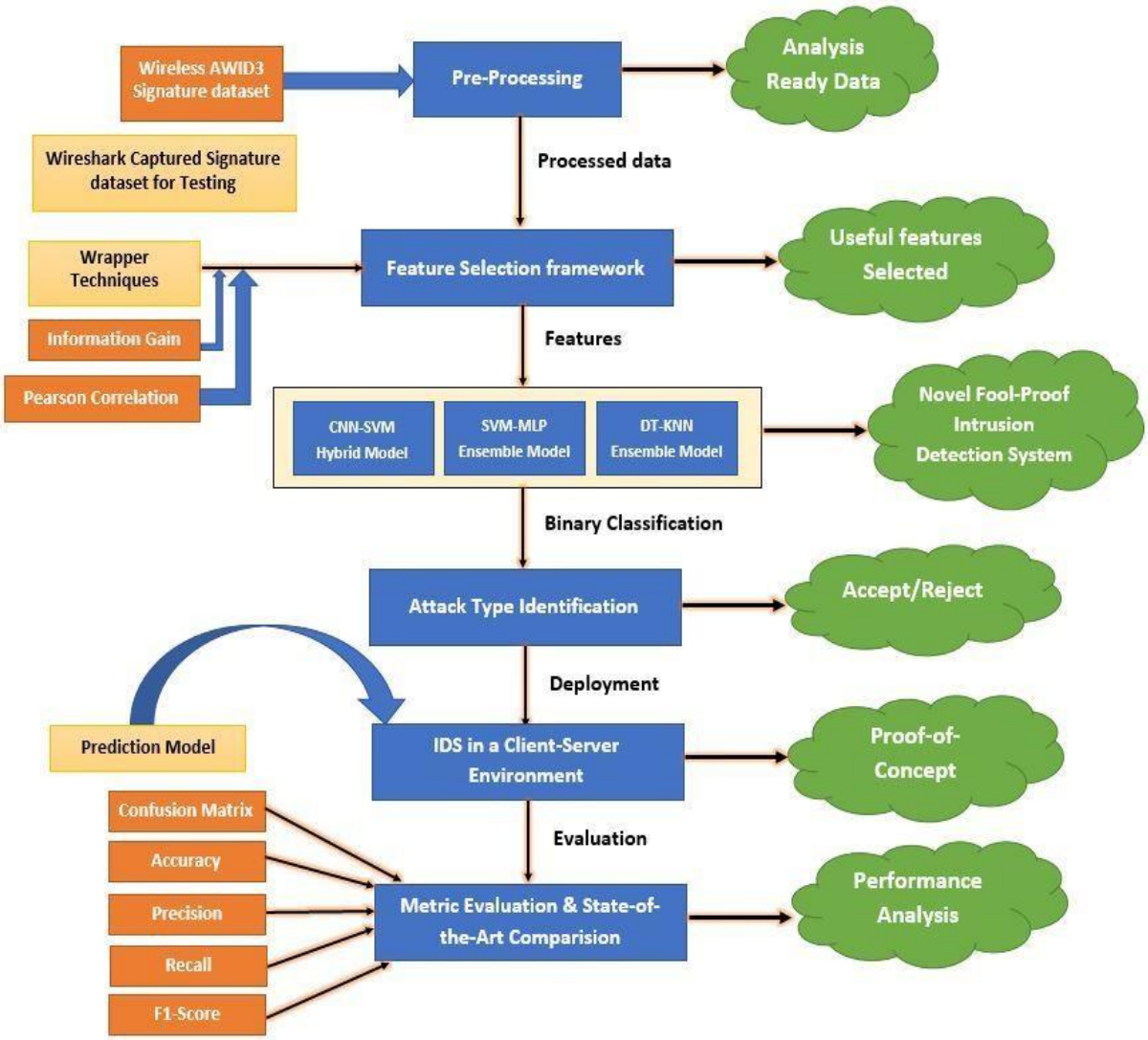

Fig 1. The methodology to build the proposed IDS

A. Pre-Processing

The AWID dataset [6,15] is a corpus that includes both legitimate and malicious traffic in an effort to aid in the creation of effective detection and security mechanisms. The AWID dataset consists of 155 columns and 17,95,575 rows.

You can calculate the number of records for each class in the AWID dataset using the value_counts() method on the target variable (the variable that identifies the class of each data point). The dataset is likely to have a class imbalance if there is a large disparity in the number of records for each class.

Algorithm: Check for Class Imbalance in AWID Dataset

```
# Import the necessary libraries
import pandas as pd
# Load the dataset
df = pd.read_csv("AWID.csv")
# Calculate the number of records for each class
class_counts = df["class"].value_counts()
# Print the class counts
print(class_counts)
```

Fig 2. Check for Class Imbalance distribution in AWID dataset

IJARCCE B. ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023

DOI: 10.17148/IJARCCE.2023.12825

The methodology for identifying the presence of class imbalance distribution in the AWID dataset is illustrated in Figure 2. The results of this analysis are presented in Table 1, which confirms the existence of class imbalance distribution in the AWID dataset. To address the shortcomings of the dataset, the following measures were implemented: Basic Preprocessing and Synthetic Minority Oversampling Technique (SMOTE).

Several preprocessing steps were performed on the AWID dataset, including:

1. Appending column names to the dataset.
2. Replacing "?" with NaN values. Additionally, columns with over 60% NaN values were dropped, resulting in the removal of 7 columns.
3. Eliminating columns containing at least one NaN value, which affected a total of 1973 rows.
4. After the completion of the preprocessing phase, the AWID dataset consisted of 1,793,602 records and 148 features.

Table 1. Count of each class in AWID dataset

Number	Class	Class Count
0	Amok	31180
1	Arp	64609
2	Authentication_req	3500
3	Beacon	1799
4	Café_latte	45889
5	Deauthentication	10447
6	Evil_twin	2633
7	Fragmentation	770
8	Normal	1633190
9	Probe_response	1558

Table 2. AWID dataset after applying SMOTE

Number	Class	Before SMOTE	After SMOTE
0	Amok	31180	30240
1	Arp	64609	64610
2	Authentication_req	3500	3500
3	Beacon	1799	1700
4	Café_latte	45889	46330
5	Deauthentication	10447	10770
6	Evil_twin	2633	2670
7	Fragmentation	770	790
8	Normal	1633190	1631490
9	Probe_response	1558	1510

SMOTE is utilized to oversample the minority class without modifying the number of records for each class. To control the oversampling ratio, the "ratio" parameter in the SMOTE function is employed. This parameter determines the ratio of synthetic data points to be generated for the minority class.

In the algorithm represented in Figure 3, the input data is denoted as X, and the target variable is denoted as y. By using the fit_resample method, SMOTE is applied to the dataset, and the resampled dataset is returned as X_resampled and y_resampled. The resulting dataset will have nearly an equal number of records for each class as the original dataset. The outcomes obtained after applying the SMOTE algorithm to the AWID dataset are presented in Table 2.

Algorithm: SMOTE technique for AWID Dataset

```
# Import the necessary libraries
from imblearn.over_sampling import SMOTE
# Define the SMOTE function
smote = SMOTE(ratio=1.0)
# Apply SMOTE to the dataset
X_resampled, y_resampled = smote.fit_resample(X, y)
```

Fig 3. Algorithm for SMOTE

IJARCCE B. ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023

DOI: 10.17148/IJARCCE.2023.12825

B. Feature Selection

In order to optimize the model's performance and reduce the number of input variables, it is beneficial to employ a Feature Selection framework. This framework involves selecting a subset of important features from the original set of variables based on predefined criteria. By doing so, the dimensionality of the dataset is minimized, and the performance of machine learning algorithms is enhanced. Feature selection aims to decrease the number of features, eliminate noisy and irrelevant features, and retain only the most relevant ones.

In this research paper, the Feature Selection framework utilizes the Random Forest classifier as a wrapper method. Figure 4 illustrates the algorithm for selecting the most significant features in a dataset using a random forest classifier in Python. The code segment initially divides the dataset into input features (X) and target labels (y). It then trains a random forest classifier on the input features and evaluates the importance of each feature based on the trained classifier. Finally, it selects the top N features according to their importance and returns the selected features. This approach proves valuable for reducing the dataset's dimensionality and improving the classifier's performance.

The results obtained for the Random Forest (RF) classifier are as follows. Since the CNN model for the wireless environment was reshaped to 2x5, only the top 10 features were considered for the AWID dataset. Table 3 presents the list of these top 10 features selected for the AWID dataset using the RF classifier.

Algorithm: Feature Selection using Random Forest Classifier for AWID Dataset

Input: AWID training dataset X
Output: Selected features

```
# Split the data into X and y
X = dataset.copy ( )
y = X.pop ('Label')
split_data(X, y)
# Fit random forest classifier on the dataset
rf_classifier = random_forest_classifier.fit (X, y)
# For each feature in the dataset, calculate its importance
for each feature in dataset:
importance = calculate_feature_importance (rf_classifier, feature)
# Select the top N features according to their importance
selected_features = select_top_features (feature, importance)
Return selected_features
```

Fig 4. Feature selection using Random Forest Classifier for AWID Dataset

Table 3. Selected features for AWID Wireless dataset using Random Forest Classifier

Feature	Importance
1. wlan.duration	0.41784713
2. frame.time_delta_displayed	0.36132384
3. wlan.fc.type	0.21510416
4. wlan.fc.frag	0.00464401
5. wlan.fc.moredata	0.00108086
6. wlan.fcs_good	0.00049427
7. wlan.fc.version	0.00035655
8. frame.offset_shift	0.00028519
9. radiotap.rxflags.badplcp	0.00019761
10. radiotap.flags.shortgi	0.00005629

C. Fool-Proof IDS for Wireless Environments

The goal of the research study is to combine a number of ensembled machine learning (ML) models to create a unique and incredibly reliable intrusion detection system (IDS). The IDS includes three distinct models, namely:

IJARCCE B. ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023

DOI: 10.17148/IJARCCE.2023.12825

1. CNN-SVM Hybrid Model: In this model, Convolutional Neural Network (CNN) and Support Vector Machine (SVM) approaches are combined into a single model or strategy.

2. SVM-MLP Ensemble Model: In this model, the ensemble output of the individual SVM and Multilayer Perceptron (MLP) models is taken into account for prediction.

3. DT-KNN Ensemble Model: The ensemble output from the individual Decision Tree (DT) and K-Nearest Neighbours (KNN) models is utilized to make predictions.

The architecture of the proposed fool-proof IDS for the wireless environment is illustrated in Figure 5. In this architecture, if any one of the three models predicts an attack, it is classified as an attack, ensuring a robust and reliable detection system.

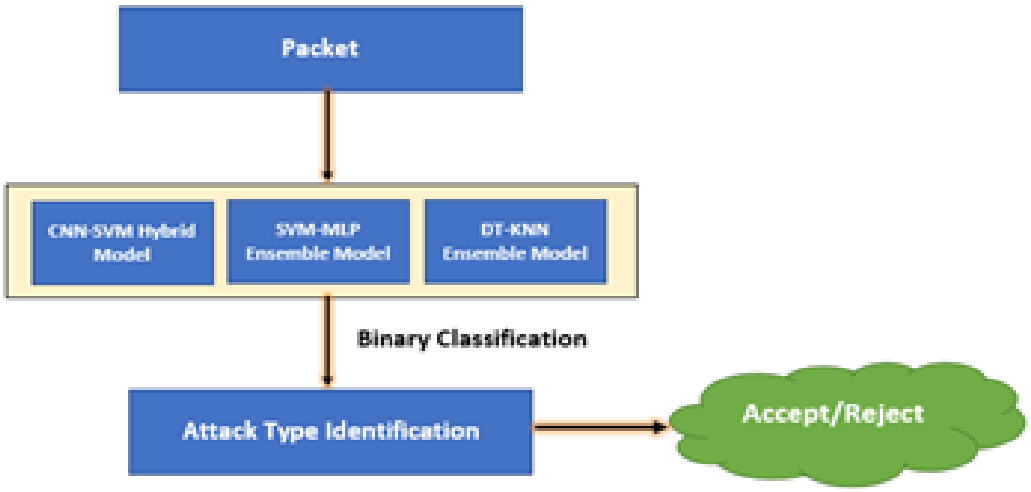

Fig 5. Fool-Proof IDS Architecture for Wireless Environment

Algorithm: The CNN-SVM Hybrid Model for Wireless Environment

```
1 SET model TO Sequential()
2 model.add(tf.keras.Input(shape=(10)))
3 model.add(layers.Reshape((2, 5)))
4 model.add(layers.Conv1D(512, 3, 1, padding='same',activation= 'elu'))
5 model.add(layers.BatchNormalization())
6 model.add(layers.Conv1D(256, 3, 1, padding='same', activation= 'elu' ))
7 model.add(layers.BatchNormalization())
8 model.add(layers.Conv1D(128, 2, 1, padding='same', activation= 'elu' ))
9 model.add(layers.BatchNormalization())
10 model.add(layers.Dropout(0.4))
11 model.add(layers.Flatten())
12 model.add(layers.Dense(64, activation= 'elu'))
13 model.add(layers.Dense(32, kernel_regularizer=tf.keras.regularizers.l2(0.01), activation='linear',
name='SVM'))
14 model.add(layers.Dropout(0.2))
15 model.add(layers.Dense(10))
16 SET opt TO keras.optimizers.Adam(learning  rate=1e-5)
17 SET loss TO tf.keras.losses.SparseCategoricalCrossentropy()
18 model.compile(loss=loss, optimizer=opt, metrics=[ 'accuracy' ])
```

Fig 6. Algorithm for CNN-SVM Hybrid Model

DOI: 10.17148/IJARCCE.2023.12825

We have generalized the CNN-SVM model to include anomaly detection. CNN can identify even the slightest trend that corresponds to the earlier attack. The model is tuned using Early Stopping and ReduceLRonPlateau. The 90:10, 80:20, and 70:30 splits were used to train and test the model. The CNN-SVM hybrid model's methodology and model summary in a wireless environment are shown in Figure 6.

Algorithm: The SVM-MLP Ensemble Model for Wireless Environment

```
1 #Create a svm Classifier
2 SET svc TO svm.SVC(probability=True, kernel='poly', max_iter=100,
decision  function  shape='ovo', verbose=True)
3 #Create a mlp classifier
4 SET mlp TO MLPClassifier(hidden_layer_sizes=(50,), learning_rate='adaptive',
alpha=0.001, learning  rate  init=1e-05, n  jobs=-1)
5 # Create Ensemble Model of SVM and MLP
6 SET combined  model TO [('SVM', svc), ('MLP', mlp)]
7 SET EnsembleClassifier TO VotingClassifier(estimators=combined_model, voting='hard',
n  jobs=-1)
```

Fig 7. Algorithm for SVM-MLP Ensemble model

Algorithm: The DT-KNN Ensemble Model for Wireless Environment

```
1 #Create a Decision Tree Classifier
2 SET dt TO DecisionTreeClassifier(max_depth TO 20)
3 #Create a KNN Classifier
4 SET knn TO KNeighborsClassifier(n  neighbors TO 15, n  jobs=-1)
5 # Create Ensemble Model of DT and KNN
6 SET combined_model TO [('Decision Tree Classifier', dt, ('KNN', knn)]
7 SET EnsembleClassifier TO VotingClassifier(estimators=combined_model, voting='hard',
n  jobs=-1)
```

Fig 8. Algorithm for DT-KNN Ensemble model

We generalized the SVM-MLP model to include anomaly detection. This makes use of hard voting, which combines forecasts using a majority vote. The algorithm for the SVM-MLP Ensemble model in a wireless environment is shown in Figure 7.

The DT-KNN model has been overfitted to the available dataset in order to eliminate assaults with even a small number of records. This makes use of hard voting, which combines forecasts using a majority vote. The algorithm for the DT-KNN Ensemble model in a wireless environment is shown in Figure 8.

V. RESULTS AND DISCUSSIONS

A binary classification procedure using several classifier models based on an ensemble of ML algorithms is used for intrusion detection. Three splits—90:10, 80:20, and 70:30—are assessed for each model. The accuracy, precision, recall, F1-Score, and confusion matrix performance of each model is calculated using a variety of evaluation metrics.

IJARCCE B. ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023

DOI: 10.17148/IJARCCE.2023.12825

Table 4. Summary of classification report results for Wireless Enterprise Networks

Name of the Model	Wireless		
CNN-SVM	**Precision**	**Recall**	**Accuracy**
90:10	96%	97%	96%
80:20	96%	97%	96%
70:30	96%	97%	96%
SVM-MLP	**Precision**	**Recall**	**Accuracy**
90:10	97%	53%	67%
80:20	97%	54%	69%
70:30	95%	94%	94%
DT-KNN	**Precision**	**Recall**	**Accuracy**
90:10	97%	97%	97%
80:20	97%	97%	97%
70:30	97%	97%	97%

Table 4 provides a summary of the classification reports for CNN-SVM, SVM-MLP, and DT-KNN models in the context of Wireless Enterprise Networks. From the results of CNN-SVM, it can be concluded that the 80:20 data split is suitable as the model performs well on both the training and validation data. On the other hand, the 90:10 and 70:30 splits show inconsistent performance on the validation data. For SVM-MLP, the 70:30 data split is deemed appropriate as the model demonstrates better performance on both the training and validation data. In contrast, the 90:10 and 80:20 splits do not align with the experimental results. As for DT-KNN, the results indicate that the 90:10, 80:20, and 70:30 splits yield similar outcomes since the DT-KNN model is prone to overfitting.

Table 5 compares the proposed method with existing approaches in terms of prediction accuracy. The scores clearly indicate that the newly proposed method surpasses the state-of-the-art techniques in terms of accuracy for Wireless IDS environments.

[[

Table 5. State-of-the-Art Comparison for Wireless IDS Environment

Reference	Method	Algorithm	Dataset	Accuracy
Udaya Sampath K. et al. (2016)	ML	OneR	AWID-ATK-R	92.07%
Udaya Sampath K. et al. (2016)	ML	J48	AWID-ATK-R	94.37%
Udaya Sampath K. et al. (2016)	ML	Random Forest	AWID-CLS-R	93.39%
Udaya Sampath K. et al. (2016)	ML	Ada Boost	AWID-CLS-R	91.85%
Proposed Method	ML/DL	CNN-SVM SVM-MLP DT-KNN	AWID	96% 94% 97%

VI. CONCLUSION

The research paper introduces and evaluates an ensemble approach for intrusion detection in wireless enterprise networks. The findings of the study indicate that the ensemble technique is highly effective for intrusion detection. Among the models evaluated, the CNN-SVM model achieved the highest accuracy of 96% in wireless environments, followed by the SVM-MLP model with an accuracy of 94%, and the DT-KNN model with an accuracy of 97%.

DOI: 10.17148/IJARCCE.2023.12825

In summary, the study concludes that employing ensembled machine learning models significantly enhances the accuracy of intrusion detection systems, resulting in an average improvement of 3% for wireless environments compared to individual models.

The ensembled model also demonstrates the ability to detect a wider range of intrusions and exhibits a lower false positive rate compared to individual models. These findings highlight the potential of ensembled machine learning models to greatly enhance the accuracy and effectiveness of intrusion detection systems.

Future research can expand on these findings by incorporating larger datasets and conducting tests on real-world intrusion scenarios. Additionally, there is potential to offer Intrusion Detection Systems as a service based on these advancements.

REFERENCES

[1] Tarter A. (2017): Importance of cyber security. Community Policing-A European Perspective: Strategies, Best Practices and Guidelines. New York, NY: Springer; 213-230.
[2] Lunt TF. (1993): A survey of intrusion detection techniques. Comput Sec. 12(4); 405-418.
[3] Lew J, Shah DA, Pati S, et al. (2019): Analyzing ML workloads using a detailed GPU simulator. Proceedings of the IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS). Madison, WI, USA: IEEE; 151-152.
[4] D. Kaleem and K. Ferens (2017): A cognitive multi-agent model to detect malicious threats, Proc. Appl. cognitive Comput. Conf., pp. 58–66.
[5] V. L. L. Thing (2017): Attack Classification: A Deep Learning Approach, 2017 IEEE Wirel. Commun. Netw. Conf., pp. 1–6.
[6] C. Kolias, G. Kambourakis, A. Stavrou, and S. Gritzalis. (2016): Intrusion detection in 802.11 networks: Empirical evaluation of threats and a public dataset. IEEE Commun. Surveys Tuts., vol. 18, no. 1, pp. 184-208.
[7] M. E. Aminanto and K. Kim (2016): Detecting Active Attacks in WiFi Network by Semi-supervised Deep Learning, pp. 1–4.
[8] N. Moustafa and J. Slay (2015): The significant features of the UNSW-NB15 and the KDD99 data sets for Network Intrusion Detection Systems, Proc. - 2015 4th Int. Work. Build. Anal. Datasets Gather. Exp. Returns Secur. BADGERS 2015, pp. 25–31.
[9] M. E. Aminanto, P. D. Yoo, H. C. Tanuwidjaja, and K. Kim (2017): Weighted Feature Selection Techniques for Detecting Impersonation Attack in Wi-Fi Networks, Symp. Cryptogr. Inf. Secur., pp. 1–8.
[10] U. S. K. P. M. Thanthrige, J. Samarabandu, and X. Wang (2016): Machine learning techniques for intrusion detection on public dataset, Can. Conf. Electr. Comput. Eng., vol. 2016-Octob, pp. 7–10.
[11] U. Sampath, K. Perera, and M. Thanthrige (2016): Scholarship @ Western.
[12] C. Kolias, V. Kolias, and G. Kambourakis (2017): TermID: a distributed swarm intelligence-based approach for wireless intrusion detection," Int. J. Inf. Secur., vol. 16, no. 4, pp. 401–416.
[13] F. D. Vaca and Q. Niyaz (2018): An ensemble learning based Wi-Fi network intrusion detection system (WNIDS), NCA 2018 - 2018 IEEE 17th Int. Symp. Netw. Comput. Appl.
[14] J. Ran, Y. Ji, and B. Tang (2019): A semi-supervised learning approach to IEEE 802.11 network anomaly detection, IEEE Veh. Technol. Conf., vol. 2019.
[15] AWID. (2022, Nov.). Aegean wireless intrusion dataset, Samos, Greece. [Online]. Available: www.icsd.aegean.gr/awid/downloads.

BIOGRAPHY

Gururaja H S has completed his B.E. in Computer Science and Engineering from Visvesvaraya Technological University and M.Tech in Computer Network Engineering also from Visvesvaraya Technological University, Belgaum. He is currently pursuing his Ph.D. from JNTU, Hyderabad in the field of Network Security under the supervision of Dr. M. Seetha. He has around 15+ years of teaching experience. His research interests include Network Security, Informatics and Computing.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 8, August 2023
DOI: 10.17148/IJARCCE.2023.12825

Dr. M. Seetha received her B.Tech from Nagarjuna University in 1992, M. S. from BITS, Pilani in 1999 and Ph.D in Computer Science and Engineering in the area of image processing in December 2007 from Jawaharlal Nehru Technological University, Hyderabad. She is currently working as a Professor and Head, Department of CSE in GNITS, Hyderabad. She has a vast teaching experience of 26 years and worked in reputed colleges including CBIT, Hyderabad and Bapatla Engineering College, Bapatla. Her areas of research include Image processing, Soft Computing, Artificial Intelligence, Data Mining and Optimization Algorithms.