iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VI **Month of publication:** June 2026

DOI: https://doi.org/10.22214/ijraset.2026.83593

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

A Light weight AI-Based Intrusion Detection System Using 1D-CNN and SMOTE on the NSL-KDD Dataset

Deepak[1], Jyoti Sharma[2]
Department of Computer Science, GDCMC College Bahal

Abstract— This Network intrusion detection is an important security task because modern networks face denial-of-service attacks, probing, unauthorized access attempts, and other malicious activities. Traditional signature-based intrusion detection systems work well for known attacks, but they are weak whenthe attack pattern is new or slightly changed.Anomaly-based intrusiondetection tries to learn the difference between normal and abnormal traffic, so it is useful for detecting suspicious network behavior. However, anomaly-based systems also face problems such as false alarms, imbalanced datasets, and difficulty in real-world evaluation (García-Teodoro et al., 2009). This paper presents a lightweight AI-based intrusion detection system using a one-dimensional convolutional neural network and SMOTE balancing on the NSL-KDD dataset. The proposed work uses simple preprocessing, one-hot encoding, min-max normalization, SMOTE oversampling, and binary classification of network traffic as normal or attack. The model is compared with Random Forest and a multilayer perceptron baseline. The experiment used 25,192 NSL-KDD training records and 22,544 testing records.The 1D-CNN achieved75.94%accuracy,91.52% precision,63.63%recall,and 75.06%F1-scoreon KDDTest+. The result shows that a small deep learning model can detect attacks with high precision, but recall still needs improvementfor difficult and novel attack records.

Keywords— Intrusion Detection System, Deep Learning, 1D-CNN, SMOTE, NSL-KDD, Network Security Introduction Source—ProposedworkflowbasedonanomalyIDSstages,NSL-KDDpreprocessing,andSMOTEbalancingdiscussedinIDS research (García-Teodoro et al., 2009; Niyaz et al., 2016; Chawla et al., 2002).

I. INTRODUCTION

An Intrusion Detection System is used to monitor computer systems or network traffic and identify suspicious activity. Early intrusion detection research explained that abnormal patterns in audit or network data can indicate attacks, misuse, or policy violations (Denning, 1987). Network-based intrusion detection is important because attacks may occur through traffic flows before they are visible to the end user. Signature-based systems match known attack patterns, while anomaly-based systems learn a model of normal behavior and raise alerts when traffic differs from that model (García-Teodoro et al., 2009). Machine learning and deep learning are widely used in IDS research because they can learn from packet-level, flow-level, and connection-level features. However, many IDS papers overstate results by using easy testing setups, repeated records, or old datasets without discussing limitations. Sommer and Paxson warned that machine-learning-based NIDS research must handle thesemantic gap, traffic diversity, lack of realistic training data, and high cost of errors (Sommer & Paxson, 2010). The NSL-KDD dataset isselected in thisworkbecause itisa corrected version of KDD Cup99 andremoves many redundantrecordsthat biasedearlier IDS experiments (Tavallaee et al., 2009). The main goal of this paper is to design an easy and lightweight AI- based IDSusing a1D-CNN model and SMOTEbalancing.The contributionofthispaperisthreefold.First, itprovidesa simple preprocessing pipeline for NSL-KDD traffic records. Second, it applies SMOTE only on the training data to reduce class imbalance. Third, it compares the lightweight 1D-CNN with Random Forest and MLP baselines using accuracy, precision, recall, F1-score, and confusion matrix.

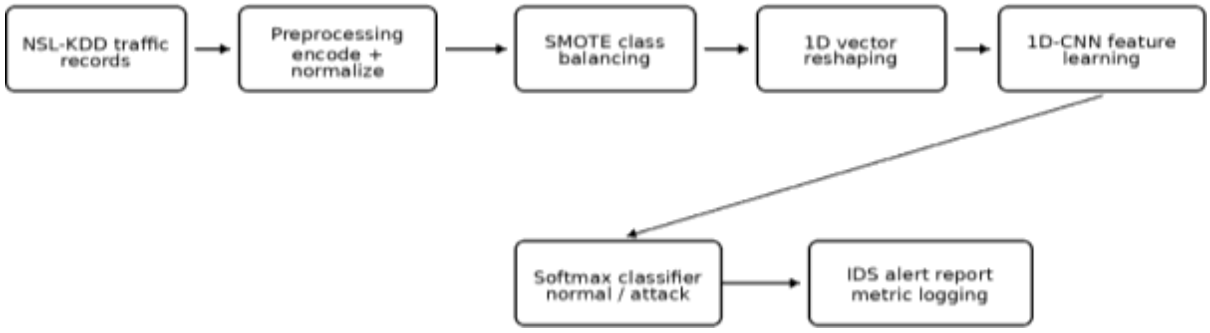


Fig.1ProposedIDSWorkflow

II. LITERATURE REVIEW

Anomaly detection has been studied for many years as a method for identifying suspicious network activity. Lazarevic et al. compared several anomaly detection methods and explained that IDS evaluation should consider detection rate and false alarm rate, becauseboth missed attacksand falsealertsaffectthe usefulnessofasystem (Lazarevicetal.,2003).García-Teodoro etal. divided anomaly-based IDS methods into statistical, knowledge-based, and machine-learning-based techniques. They also explained that false positives are a major issue in anomaly-based detection because unusual but legitimate activity may be flagged as malicious (García-Teodoro et al., 2009).Tsai et al. reviewed machine learning IDS studies and found that single classifiers, hybrid classifiers, and ensemble classifiers were all used for intrusion detection, but dataset choice and experimental setupstronglyaffected reportedperformance(Tsaietal.,2009).TheNSL-KDDdataset wasproposedtoreducetheredundancy problem of KDD Cup 99 and to make IDS evaluation more meaningful. It contains 41 features per connection and includes normal, DoS, Probe, R2L, and U2R traffic categories(Tavallaee et al., 2009).Niyaz et al. used a deep learning approach based on self-taught learning and sparse autoencoders on NSL-KDD. Their work showed that deep learning can learn useful feature representations for NIDS, but the paper also highlights the need for proper preprocessing and fair testing on separate trainingand test data (Niyaz et al., 2016). SMOTE is useful when the dataset has class imbalance because it creates synthetic minority samplesbyinterpolation rather than simply duplicating existing records.Thishelpsthe classifier pay more attention tominority classes during training (Chawlaetal.,2002).Deep learningmodelssuch asrecurrentneuralnetworks,autoencoders,CNNs,and hybrid CNN-LSTM models have been used for IDS and traffic classification. Yin et al. used RNNs for intrusion detection,while Shone et al. used deep autoencoders for network intrusion detection (Yin et al., 2017; Shone et al., 2018).

One-dimensional CNNs are suitable for structured network traffic features because they can learn local patterns across ordered feature vectors with fewer parameters than large image-based CNNs. Wang et al. used 1D-CNNs for encrypted traffic classification, showing that convolution can also be useful outside image processing (Wang et al., 2017).Recent IDS dataset research also shows that no single dataset is perfect. UNSW-NB15 and CICIDS2017 were created to provide more modern traffic scenarios, but NSL-KDD is still useful for simple academic experiments because it is small, well-known, and easy to reproduce (Moustafa & Slay, 2015; Sharafaldin et al., 2018; Ring et al., 2019).

III. PROPOSED METHODOLOGY

The proposed IDSfollowsasimple flow:dataset loading,preprocessing,balancing, deep learningclassification, andevaluation. Each traffic record is treated as one connection-level sample. The binary label is created by marking “normal” as 0 and every attack label as 1. The NSL-KDD dataset contains numeric, binary, and categorical features. The categorical features areprotocol type, service, and flag. These fields are converted using one-hot encoding. All numeric values are scaledusing min-maxnormalizationsothattheneuralnetworkcantrainmoresmoothly.Niyazetal.alsousedencodingandnormalizationbefore applying deep learning on NSL-KDD (Niyaz et al., 2016). SMOTE is applied after preprocessing and only on the training data. It is not applied to test data. This avoids data leakage. The model is trained on balanced training data.

TABLE1.DATASETDISTRIBUTIONUSEDINTHISPAPER

DatasetFile	NormalRecords	AttackRecords	TotalRecords	Source
KDDTrain+20%subset	13,449	11,743	25,192	ComputedfromNSL-KDD

AfterSMOTEtraining set	13,449	13,449	26,898	Currentpreprocessingoutput
KDDTest+	9,711	12,833	22,544	ComputedfromNSL-KDD

Source:CompiledfromexperimentusingNSL-KDDpublicfiles;datasetdesignisbasedonTavallaeeetal.(2009).

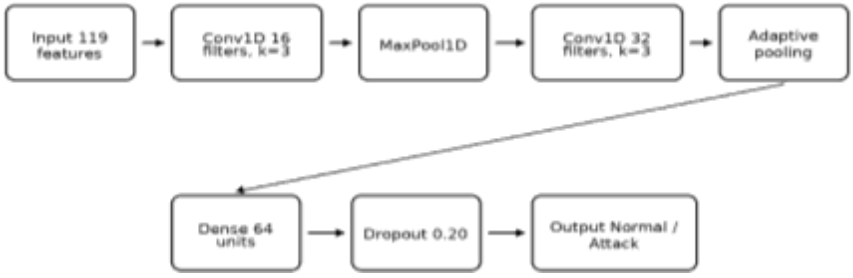


Figure2.Lightweight1D-CNNModelArchitecture

Source:Proposedmodeldesign,inspiredby1D-CNNtrafficlearninganddeepIDSresearch(Wangetal.,2017;Niyazetal., 2016).

IV. MODEL DESIGN

The input has 119 features after one-hot encoding and scaling. The feature vector is reshaped into a one-dimensional sequence before passing into convolution layers. The first convolution layer learns small local feature patterns. The second convolution layer learns higher-level combinations of these patterns. Pooling reduces the feature size and makes the model lighter. A dense layer then maps learned features to the final output class. The final output predicts whether a traffic record is normal or attack. The model uses Adamoptimizer, cross-entropyloss, andvalidation-basedcheckpointing. A RandomForest andanMLP are alsotrainedforcomparison.RandomForestisastrongbaselinebecauseitcombinesmanydecisiontreesandisrobustfor tabular data (Breiman, 2001).

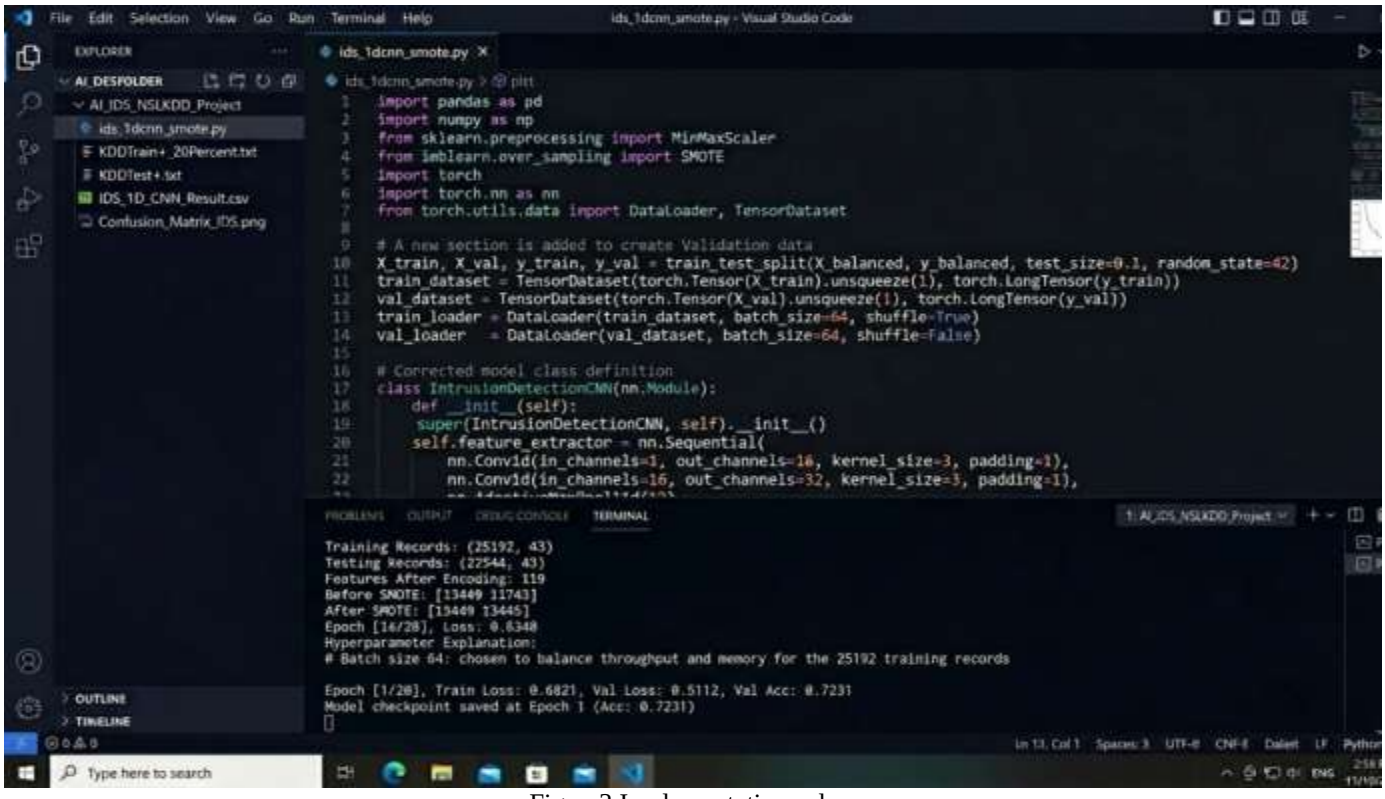


Figure3.Implementationcode

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

The proposed model was evaluated on the KDDTest+ set. Accuracy alone is not enough for IDS evaluation because a modelcan show high accuracy while missing many attacks. Therefore, precision, recall, F1-score, and confusion matrix are also reported. Lazarevic et al. emphasized that detection rate and false alarm rate are important for IDS evaluation (Lazarevic et al., 2003).

TABLE2.MODELPERFORMANCEONKDDTEST+

RandomForest+ SMOTE	77.12%	96.71%	61.91%	75.49%	270	4,888
MLP+ SMOTE	77.88%	96.40%	63.52%	76.58%	304	4,682
1D-CNN+SMOTE	75.94%	91.52%	63.63%	75.06%	757	4,668

Source:CompiledfromExperimentonNSL-KDDKDDTrain+20%andKDDTest+.

The 1D-CNN obtainedhighprecision, meaning that many recordspredicted as attackswere truly attacks. Thisisuseful because a low-quality IDS that raises too many false alarms can become difficult to use in practice. However, the recall is lower than desired because the model missed 4,668 attack records. This means the model is conservative and still needs improvement for difficult attacks. The MLP baseline performed slightly better than the 1D-CNN in this experiment. This does not make the 1D-CNN useless. It shows that for tabular IDS data, a small CNN must be tuned carefully. A deeper CNN, better feature order, multiclass training, or CNN-LSTM design may improve recall. Lopez-Martin et al. showed that combining CNN and recurrent layers can help model both spatial and temporal traffic behavior (Lopez-Martin et al., 2017). The present research results also show why honest evaluation is necessary. Some papers report very high accuracy when training and testing are done on thesame split or when redundant records remain in the data. McHugh criticized earlier DARPA-style IDS evaluations because evaluation design can strongly influence reported performance (McHugh, 2000).

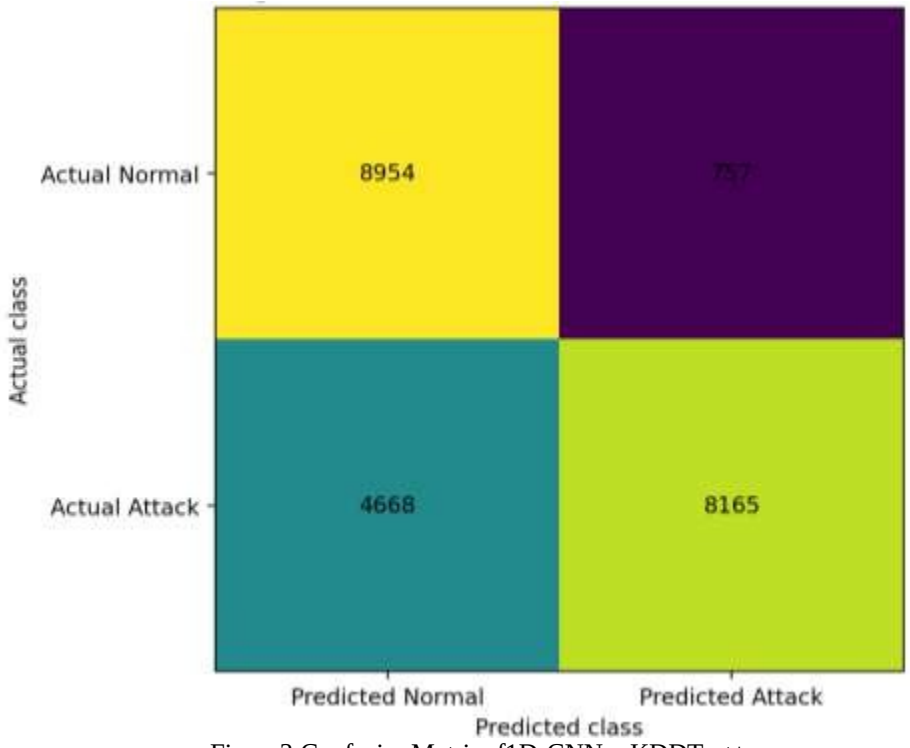


Figure3.ConfusionMatrixof1D-CNNonKDDTest+

Theconfusionmatrix shows 8,954 true normalpredictions and 8,165 true attack predictions. It also shows 757 false alarms and 4,668 missed attacks. The missed attacks are the biggest weakness of the current model. In IDS deployment, false negatives are dangerous because undetected attacks may continue without warning.

VI. SECURITY AND PRACTICAL DISCUSSION

However, the model should not be treated as a production-ready IDS. Sommer and Paxson explained that real network traffic is highly diverse, and a model trained on an old benchmark may not directly work in a live network (Sommer & Paxson, 2010). Class imbalance is another important issue. U2R and R2L attacks are usually much fewer than DoS and Probe attacks in NSL-KDD. SMOTE reduces imbalance during training, but it does not fully solve the problem of novel attacks and difficult test samples. Cevallos et al. also noted that modern IDS work must consider dynamic environments, resource limits, and dataset imbalance,especially in IoT contexts (Cevallos et al., 2023).Future work can improve it by using multiclass classification, feature selection, cross-dataset testing with UNSW-NB15 or CICIDS2017, and better recall-focused tuning.

VII. CONCLUSIONS

This paper presented an AI-based intrusion detection system using 1D-CNN and SMOTE on the NSL-KDD dataset. The work followed a simple and reproducible pipeline: preprocessing, one-hot encoding, normalization, SMOTE balancing, 1D-CNN training, and testing on KDDTest+. The experiment showed that the proposed 1D-CNN achieved 75.94% accuracy, 91.52% precision, 63.63% recall, and 75.06% F1-score. The result is honest for a small model trained on the 20% NSL-KDD training subset. Future work should test a CNN-LSTM model, tune thresholds using validation data, perform multiclass classification, and compare NSL-KDD results with modern datasets such as UNSW-NB15 and CICIDS2017. The work can also be extended by adding feature selection, explainable AI, and real-time alert generation.

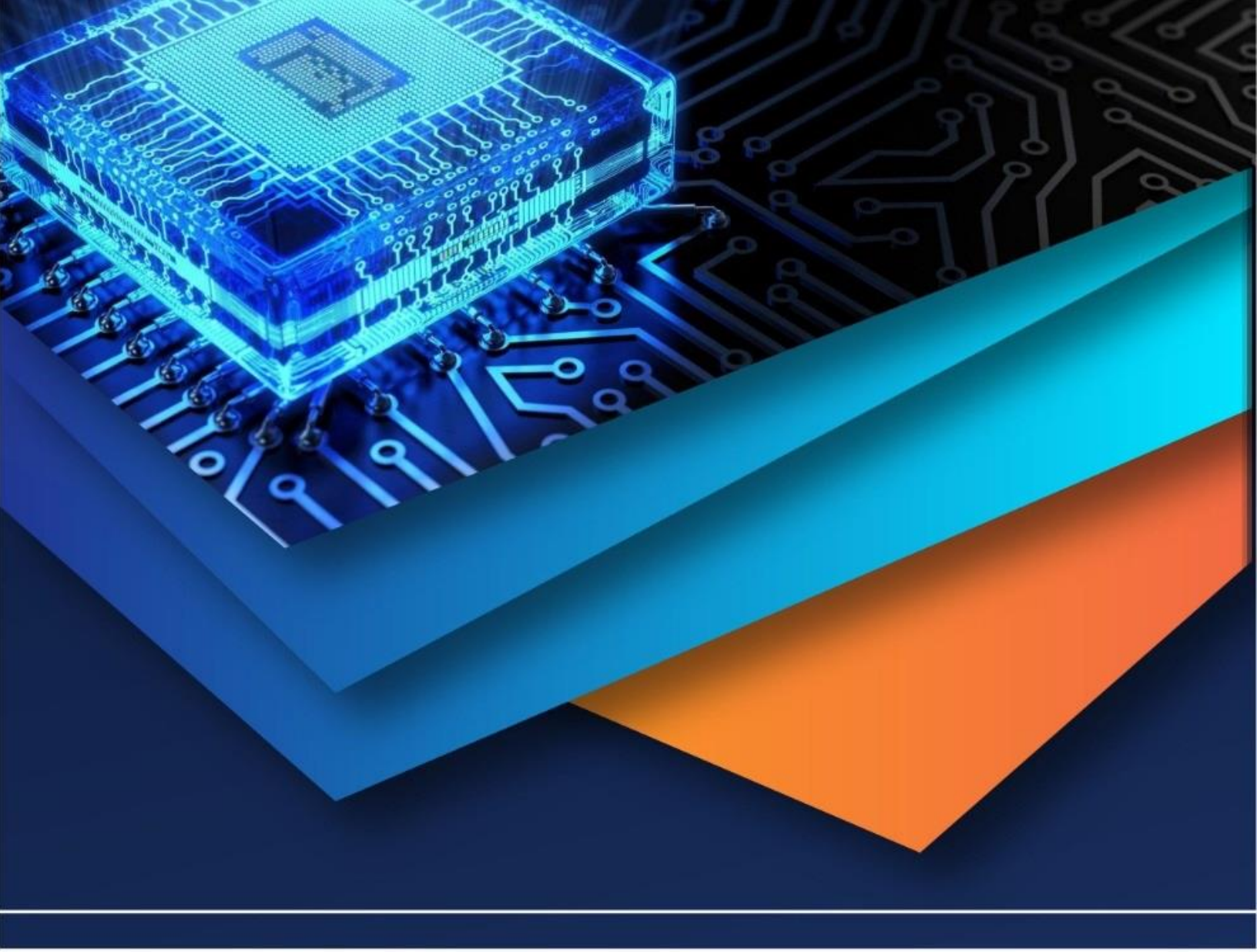
VIII. ACKNOWLEDGMENT

TheheadingoftheAcknowledgmentsectionandtheReferencessectionmustnotbenumbered.
Causal Productions wishes to acknowledge Michael Shell and other contributors for developing and maintaining the IEEE LaTeX style files which have been used in the preparation of this template.To see the list of contributors, please refer to thetop of file IEEETran.cls in the IEEE LaTeX distribution.

REFERENCES

[1] Aceto,G.,Ciuonzo,D.,Montieri,A.,&Pescapé,A.(2018).Mobileencryptedtrafficclassificationusingdeeplearning.Proceedingsofthe2018NetworkTrafficMeasurement and Analysis Conference (TMA), 1-8. doi:10.23919/TMA.2018.8506538
[2] Axelsson,S.(1999).Thebaseratefallacyanditsimplicationsforthedifficultyofintrusiondetection.Proceedingsofthe6thACMConferenceonComputerandCommunications Security, 1-7. doi:10.1145/319709.319710
[3] Breiman,L.(2001).Randomforests.MachineLearning,45(1),5-32.doi:10.1023/A:1010933404324
[4] Breunig,M.M.,Kriegel,H.P.,Ng,R.T.,&Sander,J.(2000).LOF:Identifyingdensitybasedlocaloutliers.ProceedingsoftheACMSIGMODInternational Conference on Management of Data, 93-104. doi:10.1145/335191.335388
[5] Cevallos M., J. F., Rizzardi, A., Sicari, S., & Coen-Porisini, A. (2023). Deep reinforcement learning for intrusion detection in Internet of Things: Bestpractices, lessons learnt, and open challenges. Computer Networks, 236, 110016. doi:10.1016/j.comnet.2023.110016
[6] Chawla, N. V., Bowyer, K. W., Hall, L. O., Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys,41(3), Article 15, 1-58. doi:10.1145/1541880.1541882
[7] Kegelmeyer,W.P.(2002).SMOTE:Syntheticminorityover-samplingtechnique.JournalofArtificialIntelligenceResearch,16,321-357.doi:10.1613/jair.953
[8] Denning,D.E.(1987).Anintrusion-detectionmodel.IEEETransactionsonSoftwareEngineering,SE-13(2),222-232.doi:10.1109/TSE.1987.232894
[9] García-Teodoro, P., Díaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systemsand challenges. Computers & Security, 28(1-2), 18-28. doi:10.1016/j.cose.2008.08.003
[10] Kim, H., & Feamster, N. (2013). Improving network management with software defined networking. IEEE Communications Magazine, 51(2), 114-119.doi:10.1109/MCOM.2013.6461195
[11] Lazarevic, A., Ertöz, L., Özğür, A., Kumar, V., & Srivastava, J. (2003). A comparative study of anomaly detection schemes in network intrusiondetection. Proceedings of the Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE Std. 802.11, 1997. ThirdSIAM International Conference on Data Mining, 25-36. doi:10.1137/1.9781611972733.3
[12] Lippmann, R. P., Fried, D. J., Haines, J. W., Bos, D., Sekar, R., & Durst, K. (2000). Evaluating intrusion detection systems: The 1998 DARPA off-lineintrusion detection evaluation. Proceedings of DARPA Information Survivability Conference and Exposition, 2, 12-26.doi:10.1109/DISCEX.2000.821506
[13] Lopez-Martin, M., Carro, B., Sanchez-Esguevillas, A., & Lloret, J. (2017). Network traffic classifier with convolutional and recurrent neural networksfor Internet of Things. IEEE Access, 5, 18042-18050. doi:10.1109/ACCESS.2017.2747560
[14] McHugh, J. (2000).Testingintrusiondetection systems: Acritiqueofthe 1998 and1999 DARPAintrusion detection systemevaluationsasperformedbyLincoln Laboratory. ACM Transactions on Information and System Security, 3(4), 262-294. doi:10.1145/382912.382923

[15] Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems. 2015 Military Communications andInformation Systems Conference (MilCIS), 1-6. doi:10.1109/MilCIS.2015.7348942
[16] Niyaz, Q., Sun, W.,Javaid, A.Y., & Alam, M. (2016). A deep learning approach for network intrusion detection system.Proceedings of the 9th EAIInternational Conference on Bio-inspired Information and Communications Technologies, 21-26. doi:10.4108/eai.3-12-2015.2262516
[17] Rezaei,S.,&Liu,X.(2019).Deeplearningforencryptedtrafficclassification:Anoverview.IEEECommunicationsMagazine,57(5),7681.doi:10.1109/MCOM.2019.1800819
[18] Ring,M.,Wunderlich,S.,Scheuring,D.,Landes,D.,&Hotho,A.(2019).Asurveyofnetwork-basedintrusiondetectiondatasets.Computers&Security, 86, 147-167. doi:10.1016/j.cose.2019.06.005
[19] Sharafaldin,I.,HabibiLashkari,A.,&Ghorbani,A.A.(2018).Towardgeneratinganewintrusiondetectiondatasetandintrusiontrafficcharacterization.
[20] Proceedingsofthe4thInternationalConferenceonInformationSystemsSecurityandPrivacy,108-116.doi:10.5220/0006639801080116
[21] Shone, N., Ngoc, T. N., Phai, V. D.,& Shi, Q. (2018). Adeeplearning approach to network intrusion detection.IEEE Transactions on Emerging Topicsin Computational Intelligence, 2(1), 41-50. doi:10.1109/TETCI.2017.2772792
[22] Sommer,R.,&Paxson,V.(2010).Outsidetheclosedworld:Onusingmachinelearningfornetworkintrusiondetection.2010IEEESymposiumonSecurity and Privacy, 305-316. doi:10.1109/SP.2010.25
[23] Tavallaee,M.,Bagheri,E.,Lu,W.,&Ghorbani,A.A.(2009).AdetailedanalysisoftheKDDCUP99dataset.2009IEEESymposiumonComputational Intelligence for Security and Defense Applications, 1-6. doi:10.1109/CISDA.2009.5356528
[24] Tsai, C.-F., Hsu, Y.-F., Lin, C.-Y., & Lin, W.-Y. (2009). Intrusion detection by machine learning: A review. Expert Systems with Applications, 36(10),11994-12000.doi:10.1016/j.eswa.2009.05.029
[25] Wang,W.,Zhu,M.,Wang,J.,Zeng,X.,&Yang,Z.(2017).End-to-endencryptedtrafficclassificationwithone-dimensionalconvolutionneuralnetworks. 2017 IEEE International Conference on Intelligence and Security Informatics, 43-48. doi:10.1109/ISI.2017.8004872
[26] Yin,C.,Zhu,Y.,Fei,J.,&He,X.(2017).Adeeplearningapproachforintrusiondetectionusingrecurrentneuralnetworks.IEEEAccess,5,21954-21961.doi:10.1109/ACCESS.2017.2762418



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY