Infotek : Jurnal Informatika dan Teknologi
Vol. 3 No. 2, Juli 2020, hal. 123-130

Penggunaan Metode Backpropagation Pada Jaringan Syaraf Tiruan Untuk Intrusion Detection System

Ramli Ahmad[1]*, Bq Andriska Candra P[2], Amri Muliawan Nur[3]
[1]Program Studi Teknik Komputer, Universitas Hamzanwadi
[2]Program Studi Teknik Informatika, Universitas Hamzanwadi
[3]Program Studi Teknik Informatika, Universitas Hamzanwadi
*iosram81@gmail.com

Abstrak

Intrusion detection system (IDS) adalah sebuah sistem untuk mendeteksi gangguan pada jaringan yang biasa kita sebut dengan hacker. Kenyamanan merupakan tujuan dari pemanfaatan teknologi pada saat ini, akan tetapi seiring dengan berjalannya waktu dan kemajuan teknologi itu sendiri keamanan dan privasi pengguna lambat laun semakin terganggu dan menjadi suatu faktor yang memprihatinkan. Keberadaan IDS (Intrusion Detection System) diyakini dapat membantu dalam mencapai keamanan dalam penggunaan jaringan dimanan sistem deteksi ini bekerja dengan cara mengamati prilaku jaringan yang abnormal. IDS menekankan perlunya penggunaan jaringan syaraf tiruan untuk melakukan pendeteksian terhadap serangan tersebut. Dari penelitian yang telah dilakukan, penggunaan jaringan syaraf tiruan dengan tingkat pembelajaran 0,1 telah diterapkan dan dataset KDDCup-'99 telah digunakan untuk melatih dan membuat tolak ukur jaringan. Untuk melakukan perbandingan telah dilakukan pula pelatihan pada dataset yang sama dengan menggunakan beberapa algoritma pembelajaran yang lain . Jumlah layer yang digunkan pada jaringan syaraf tiruan ini mulai dari 1 hingga 5, dan hasilnya telah dibandingkan sehingga diperoleh hasil kesimpulan bahwa jaringan syaraf tiruan dengna 3 layer memiliki kinerja yang lebih unggul di banding algoritma pembelajaran mesin learning yang lain.

Kata kunci: jaringan syaraf tiruan, pembelajaran mesin , pembelajaran struktural mendalam, IDS.

Abstract

Convenience is the main goal of existence of technology this days, but over time these technological advancements have made user privacy increasingly difficult and become a cause of concern. The existence of IDS (Intrusion Detection System) is believed to help in achieving security in network usage where this detection system work by observing abnormal network behavior. IDS emphasizes the need to use artificial neural network to detect these attack. From the research tht has been done, the use of artificial neural network with a learning rate of 0.1 has been applied and the KDDCup-99 dataset has been used to train and make network benchmarks. To conduct a comparison, training has also been carried out on the same dataset using several other learning algorithms. The number of layer used in this artificial neural network start from 1 to 5, and the result have been compared so that the conclusion said that the artificial neural network with 3 layers has a superior performance compare to other machine learning algorithms.

Keywords: backpropagation network, IDS, deep learning, machine learning.

1. Pendahuluan

Kenyamanan merupakan suatu tujuan dari penggunakan teknologi saat ini, akan tetapi tingkat keamanan dan pivasi pengguna menjadi terganggu seiring dengna perkembangna teknologi tersebut, dan hal ini menjadi hal yang sangat menghawatirkan. Penggunaan Internet Of Things (IoT) banyak membantu manusia dalam banak hal, akan tetapi juga menyebabkan timbulnya banyak kelemahan terutama hal yang terkait dengan jaringan, infrastruktur , hal – hal yang terkait dengan komunikasi dan lain – lain. Oleh karena banyaknya IoT yang jumlahnya bisa mencapai jutaan, menyebabkan sulitnya penerapan keamanan pada setiap perangkat. Untuk itu guna memantau data melalui jaringan maka dibutuhkan suatu keamaan berbasis jaringan.

Solusi keamanan jaringan berbasis IoT dapat diimplementasikan dengan melakukan beberapa perubahan kecil pada IoT itu sendiri. Untuk memungkinkan suatu akses melalui jaringan, perangkat IoT harus didaftarkan pada pusat keamanan untuk memastikan bahwa jaringan bebas dari penyusup. Hal tersebut diperlukan untuk memantau semua lalu lintas masuk dan keluar pada setiap objek IoT serta menjaga agar lalu lintas jaringan tetap normal. Apabila suatu prilaku dinyatakan gagal masuk pada kategori berprilaku normal, maka akan di identifikasi sebagai serangan pada jaringan selanjutnya akan ada alarm yang menjadi sinyal pemberitahuan pada pemilik perangkat. IDS (Intrusion Detection System) dapat membantu menjaga keamanan jaringan untuk melakukan pengamatan terhadap prilaku yang abnormal pada suatu jaringan [1].

Yang menjadi pertanyaan adalah dimana menempatkan IDS dalam suatu sistem. Karena apabila IDS di letakan pada node ataupun di distribusikan secara acak, maka hal tersebut dikenal sebagai IDS berbasis jaringan, sedangkan apabila IDS ditempatkan pada workstation, hal tersebut dikenal sebagai IDS berbasis host. Network Intrusion Detection System (NIDS) lebih cenderung menggunakan IDS. Untuk itu dapat dilakukan penggabungan IDS dengna beberapa teknologi pembelajaran mesin atau yang sering di sebut teknologi machine learning sehingga diperoleh hasil yang lebih akurat. Penggunaan machine learning merupakan bagian yang penting dalam AI yang berguna untuk menganaisis dan membangun sistem berdasarkan pengetahuan yang di peroleh dari sekumpulan data. Beberapa algoritma pembelajaran mesin yang umumnya digunakan diantaranya backpropagation neural network, Boost, K-Nears Neighbor, Regresi Linier, Naif Bayes , Random forest dan lain lain.

Metode backpropagation meurpakan metode yang dianggap sangat baik dalam menangani jaringan syaraf tiruan terkait masalah mengenali pola, yaitu keseimbangan antara kemampuan jaringan untuk mengenali pola yang digunakan selama pelatihan dan kemampuan jaringan untuk memberikan respon yang benar terhadap pola serupa dengan input seperti pola yang digunakan selama pelatihan.

Prediksi dapat dapat dihitung dengan menggunakan berbagai metode salah satu metode prediksi yang sering digunakan dan sudah berkembang saat ini adalah backpropagation. Meskipun metode ini memiliki beberapa kelemahan, seperti hasil pelatihan yang bisa dikatakan tidak konsisten dan tidak diketahui secara detail bagaimana hasil prediksi diperoleh karena metode ini tidak dapat memberikan informasi tentang bobot yang paling berpengaruh diantara input yang diberikan. Akan tetapi metode ini juga memiliki kelebihan yaitu kemampuan dalam merumuskan pengalaman dan fleksibel dalam perubahan aturan [2].

Oleh sebab itu, jelas bahwa propagasi balik (backpropagation) dan IDS dapat bekerja bersama sama ketika digabungkan dan dapat mempermudah manusia. Dikarenakan IDS tidak sesuai dengna infrastruktur dan serangan umum seperti Denial of Service (DoS) yang utamanya bertujuan untuk membatasi band jaringan untuk mendapatkan akses dari host, kinerja tersebut tidak terhambat sehingga laposan keamanan ini tidak dapat merusak mitigasi.

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Penelitian yang telah dilakukan oleh penulis [3] telah mengusulkan model dengna menggunakan dua fase untuk mendeteksi dan melakukan pengelompokan anomaly. Yang pertema dilakukan Random Forest berdasarkan akurasi tertinggi dari sebelas algoritma yang umum di gunakan kemudian diuji dengan menggunakan dataset yang sama. Yang kedua menggunakan Neural Network dengan menggunakan data yang memiliki fitur adanya serangan atau tidak untuk membedakan pada kategori seranga. Dalam hal ini dipergunakan cukup banyak dataset, salah satu diantaranya adalah KDDCUP99.

Penelitian selanjutnya [4] menyatakan bahwa untuk melakukan pendeteksian anomaly secara dinamis diperlukan melakukan pengelompokan online secara real time dengan mempertimbangkan masalah umum dan skalabilitas . Kemampuan dalam membedakan anomali yang bersifat dinamis dan dianggap kuat seperti yang ada pada sistem jaringan saat ini. Keadaan geografis jaringan komputer yang terhubung melalui internet memotivasi straming data dilakukan dengan cepat dimana pembelajaran online diperlukan.

2.2. Landasan Teori

1. Back propagation Neural Network (BPN)

Back Propagation Network adalah suatu sistem informasi pemrosesan yang memiliki karakteristik yang mirip dengan jaringan syaraf biologis. Inti dari pelatihan jaringan syaraf ini adalah suatu metode pembobotan fine-tuning berdasarkan tingkat kesalahan yagn di peroleh pada masa sebelumnya , contohnya iterasi. Penyesuaian bobot yang tepat memungkinkan terjadinya pengurangan tingkat kesalahan dan membuat model menjadi dapat diandalkan dengna meningkatkan generalisasinya. Backpropagation merupakan bentuk sederhana untuk menerjemahkan propagasi kesalahan yang

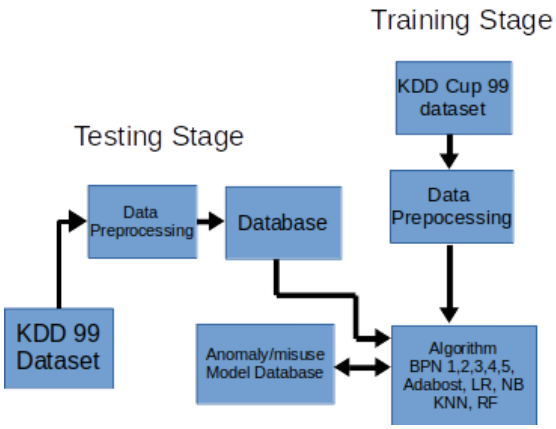
dilakukan secara mundur. Bentuk ini merupakan metode standard pelatihan jaringan syaraf tiruan untuk menghitung gradient fungsi yang hilang dengan memperhatikan semua bobot dalam jaringan.

2. Application of rectified linear unit (ReLU)

Faktanya, masalah ReLU lebih efisien dan memiliki kapasitas yang cukup untuk digunakan dalam mempercepat seluruh proses pelatihan secara keseluruhan [10]. Biasanya, jaringan syaraf tiruan in menggunakan fungsi aktivasi sigmoidal atau fungsi aktivasi tanh (tangen hiperbolik), akan tetapi fungsi – fungsi in biasanya menghilankan atau mengesampingkan masalah gradient. Gradient menghilan gketika lapisan bawah dari backpropagation neural network (BPN) memiliki gradient yang hampir nol, karena unit lapisan yang lebih tinggi hampir jenuh. ReLU merupakan alternative untuk sigmoid non linier dimana yang mengatasi masalah yang disebutkan sejauh ini [11].

3. Data Preprocessing

Klasifikasi untuk data yang dianggap berbahaya dan data normal tidak akan efektif jika data yang terdapat pada dataset KDDCUP 99 diproses pada format asalnya. Oleh sebab itu diperlukan untuk pra proses data sebelum sistem algoritma klasifikasi dibangun. Serangan dipetakan kedalam lima kelas, 0 untuk normal, 1 untuk DoS (Denial of Service), 2 untuk U2R (user to root : akses tidak sah ke root), 3 R2L (remote to local ; akses tidak sah ke lokal dari mesin remote), 4 Probe (probing : informasi mengumpulkan serangan).



Gambar 1 Pemrosesan data

2.3. Tahapan Penelitian

Adapun tahapan dari penelitian ini meliputi beberapa tahapan penelitian diantaranya :

1. Tahap pengumpulan data

   Merupakan tahap pengumpulan data-data dan informasi tentang IDS
2. Tahap Pengelompokan data

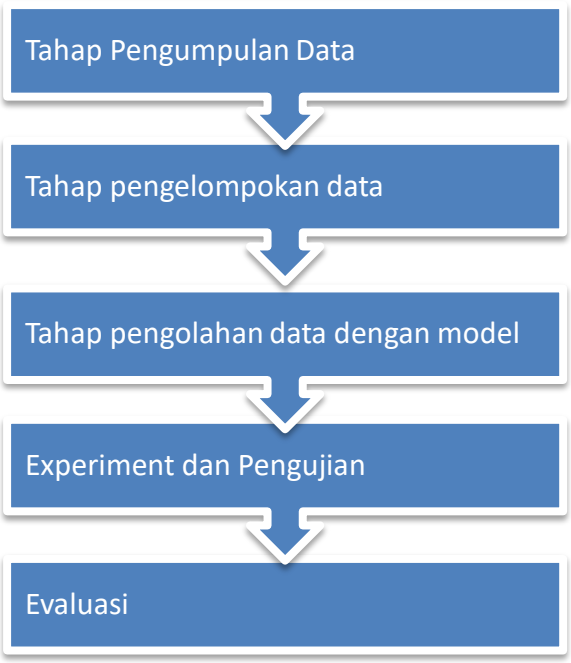
   Merupakan tahap untuk mengecek jika ada data data yang redundan atau tidak sesuai
3. Tahap pengolahan data

   Tahap melakukan pengolahan data dengan model yang telah direncanakan
4. Experimen dan Pengujian

   Hasil yang telah dilakukan, di uji coba untuk melihat apakah hipotesis yang dihasilkan dapat sesuai dengan yang diinginkan peneliti.
5. Evaluasi

   Setelah dilakukan pengujian dan experimen terhadap hasil yang dilakuan, maka bisa dilihat perbedaan dari hasil yang dilakukan

sebelumnya, apakah apakah terjadi peningkatan atau tidak.



Gambar 2. Tahapan Penelitian

3. Metode Penelitian

3.1. Metode Penelitian yang digunakan

Untuk mendapatkan data yang benar – benar bisa digunakan, akurat, dan relevan terhadap hasil nyata, maka peneliti menerapkan beberapa cara dalam pengumpulan data diantaranya adalah :

1. Observasi

   Observasi merupakan suatu teknik pengumpulan data melalui pengamatan dan pencatatan terhadap suatu peristiwa yang berhubungan dengan object penelitiannya. Pengamatan dilakukan berkaitan dengan kehidupan sehari hari yang cukup bergantung dengan kebutuhan teknologi namun memiliki kekurangan yaitu hilangnya privasi.

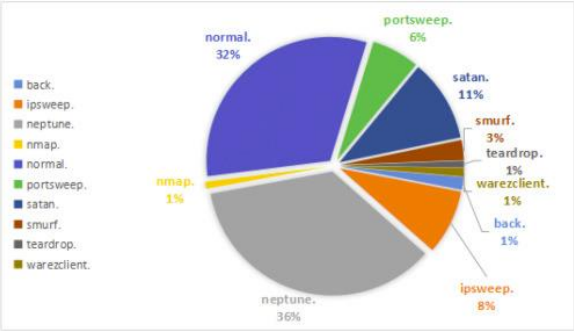
2. Studi Pustaka

   Suatu metode yang dilakukan degan membaca dan mempelajari literatur, dokuman – dokumen atau buku - buku yang bisa digunakan sebagai petunjuk dalam melancarkan penelitian.

4. Hasil dan Pembahasan

Penelitian ini menggunakan Keras [12] sebagai pembungkus diatas tensor flow sebagai kerangka kerja perangkat lunak guna meningkatkan tingkat kecepatan dalam pemrosesan data eksponsial dalam arsitektur deep learning. Kumpulan data digunakan untuk melakukan evaluasi model, baik yang digunakna untuk mendeteksi serangan secara akurat maupun tidak. Kualiitas daya yang ditetapkan akan mempengaruhi hasil dari setiap NIDS. Fitur dalam dataset KDD'99 dikategorikan menjadi empat kelompok, yaitu fitur dasar (1 sampai 9), fitur konten (10 sampai 22), waktu berdasarkan fitur lalu lintas (23 sampai 31) dan catatan host lebih besar berdasarkan lalu lintas (32 sampai 41). KDD'99 [14] terdiri dari 4.898.430 catatan yang lebih besar dari kumpulan data lainnya. Banyak teknik penambangan data yang telah di terapkan pada kumpulan data KDD'99 untuk mendeteksi gangguan pada trafik jaringan. Sebagian besar dataset KDDCup'99 digunakan untuk membangun IDS. Dataset KDD memiliki dua masalah kritis yang disimpulkan oleh analisis statistik , yang sangat mempengaruhi kinerja sistem. Jumlah besar replikasi data dapat menyebabka algoritma perbelajaran parsial dan bukan banyaknya data. Oleh karena itu,

algoritma ini akan berhenti mempelajari data yang tidak umum, karena mungkin akan berbahaya bagi jaringan seperti U2R, R2L dll.



Gambar 3. Distribusi dataset KKDCup 99

Simulasi serangan secara umum dikategorikan sebagai berikut :

1. Denial of Service (DoS) : serangan yang menyebabkan penyerang membuat agar sumberdaya sibuk melakukan pencegahan agar pengguna yang sah menggunakan sumber daya.
2. Remote to Local (R2L) : serangan dimana hacker berupaya dengan menggunakan account pengguna normal untuk mendapatkan hak super user.
3. Scanning atau Probing : serangna dimana hacker melakukan scanning terhadap mesin atau jaringan untuk mendapat informasi tentang mesin yang menjadi targetnya.

Evaluasi matrik untuk menikatkan kinerja model adalah dengan akurasi, recall, dan tingkat presisi juga harus di hitung. Disini telah di pilih penggunaan akurasi, recall, presisi dan ukuran F1 untuk melakukan evaluasi. Akurasi 1 merupakan persentase deteksi benar atas sejumlah kasus. Recall merupakan seberapa sering suatu hal di prediksi dengan benar. Recall 2 dikenal pula sebagai True Positive Rate (TPR) atau tingkat sensitivitas. Presisi 3 menyatakan bahwa ketika suatu hal diprediksi dengan benar, seberapa sering prediksi tersebut sebenarnya benar. F1 4 merupakan rata rata dari recall dan presisi. Representasinya terlihat pada gambar confusion matrix . Confusion matrix merupakan tabel yang berkaitan dengan kinerja model klasifikasi dalam serangkaian pengujian dengan mengidentifikasi nilai sebenarnya.

Total instances	**Predicted NO**	**Predicted YES**	
Actual NO	TN (True Negative)	FP (False Positive}	
Actual YES	FN(False Negative)	TP (True Positive)	**Recall**
		Precision	**Accuracy**

Gambar 4 Confusion Matrix

$$\text{Accuracy} = \frac{TP+TN}{Total\ Instances} \quad (1)$$

$$\text{Recall} = \frac{TP}{Total\ Actually\ YES} \quad (2)$$

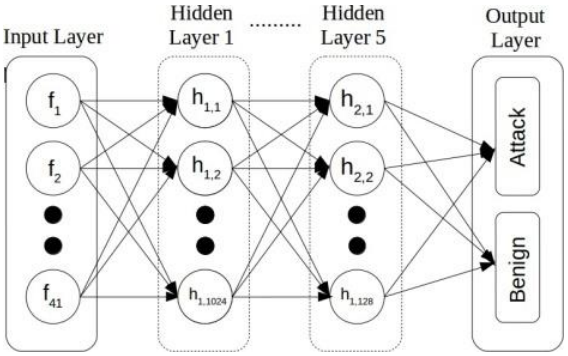
$$\text{Precision} = \frac{TP}{Total\ Predicted\ YES} \quad (3)$$

$$\text{F1Measure} = 2 * \text{Precision} * \text{Recall Precision} + \text{Recall} \quad (4)$$

Data diatas merupakan representasi matematika langkah kerja dimana input dan hiden layer terdiri dari 41 neuron yang dinyatakan sebagai input. Selanjutnya input tersebut dimasukan dalam hiden layer menggunakan ReLU sebagai fungsi aktivasi non linier. Selanjutnya bobot ditambahkan untuk memberi jalan menuju hiden layer selanjutnya. Perhitungan neuron untuk setiap hiden layer terus berkurang mulai dari input awal yang menuju output pertama kali

untuk membuat hasil lebih akurat, dan disaat yang sama dapat mengurangi biaya komputasi.

Regulasi dilakukan untuk menghemat waku dan membuat proses menjadi lebih efisien dengan nilai dropout (0.01). Fungsi dropout ini adalah untuk membuat model menjadi lebih kuat, secara acak membawa neuron neuron sampai pada proses training data. Oleh karena 1024 neuron yang terdapat pada layer sebelumnya harus di buat menjadi 2 neuron saja, maka digunakan fungsi aktivasi sigmoid.



Gambar 5 Bentuk arsitektur yang di capai

Tabel 1. Hasil Perbandingan Beberapa Algoritma

Algorithm	Accuracy	Precision	Recall	F1score
BPN 1	0.929	0.998	0.915	0.954
BPN 2	0.929	0.998	0.914	0.954
BPN 3	0.93	0.997	0.915	0.955
BPN 4	0.929	0.999	0.913	0.954
BPN 5	0.927	0.998	0.911	0.953
Ada Boost	0.925	0.995	0.911	0.951
K-Nearest Neighbour	0.929	0.998	0.913	0.954
Linear Regression	0.846	0.988	0.819	0.896
Naive Bayes	0.929	0.988	0.923	0.954
Random Forest	0.926	0.999	0.91	0.952

Dari hasil ujicoba yang telah dilakukan, semua model dibandingkan untuk skor F1, akurasi, memori dan ketepatan dari tes daaset. Hasil perbandingan secara rinci telah dituliskan pada tabel 1. Dari percobaan didapatkan 3 lapisan BPN mengungguli semua algoritma mesin learning klasik lainnya.

5. Kesimpulan

Dari hasil analisis yang telah dilakukan diperoleh kesimpulan terkait manfaat penggunaan backpropagation network secara komprehensif. Sebagai referensi, telah dilakukan pula penghitungan dengan menggunakan algoritma machine learning klasik yang lain yang telah dibandingkan hasilnya dengan backpropagation network. Dataset yang digunakan pada penelitian ini adalah dataset KDDCup-'99 dimana keuntungan dari backpropagation network dibandingkan dengan algoritma lain yang dianggap sebanding. Hasil yang diperoleh adalah bahwa dengan penggunaan backpropagation network dengan jumlah layer tersembunyi yang berbeda dan menyimpulkan bahwa dengan menggunakan hidden layer 3 mendapatkan hasil lebih efektif dan akurat.

Berdasarkan hasil eksperimen didapatkan bahwa metode backpropagatin merupakan metode yang menjanjikan untuk tugas cybersecurity, walau hasil kinerja pada dataset luar biasa masih diperlukan aplikasi yang sama pada lalu lintas jaringan real time dengan struktur yang lebih kompleks.

6. Daftar Pustaka

[1] H. R. A. Talompo, A. S. Ahmad, Y. S. Gondokaryono, and S. Sutikno, "NAIDS design using ChiMIC-KGS," in *International Symposium on Electronics and Smart Devices,* 2017.

[2] Y. Yang, K. Zheng, C. Wu, and Y. Yang, "Improving the classification effectiveness of intrusion detection by using improved conditional variational autoencoder and deep neural network," *Sensors (Switzerland)*, 2019.

[3] D. A. Mohammad Kazim Hooshmand, "Machine Learning Based Network Anomaly Detection," *Int. J. Recent Technol. Eng.*, 2019.

[4] L. Rettig, M. Khayati, P. Cudre-Mauroux, and M. Piorkowski, "Online anomaly detection over Big Data streams," *Proc. - 2015 IEEE Int. Conf. Big Data*, 2015.

[5] V. Timčenko and S. Gajin, "Machine Learning based Network Anomaly Detection for IoT environments," *Icist*, 2018.

[6] N. Moustafa and J. Slay, "The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set," *Inf. Secur. J.*, 2016.

[7] D. G. Mogal, S. R. Ghungrad, and B. B. Bhusare, "NIDS using Machine Learning Classifiers on UNSW-NB15 and KDDCUP99 Datasets," *Ijarcce*, 2017.

[8] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep Learning Approach for Intelligent Intrusion Detection System," *IEEE Access*, 2019.

[9] S. Alhamouz and A. Abu-Shareha, "Hybrid Classification Approach Using Self-Organizing Map and Back Propagation Artificial Neural Networks for Intrusion Detection," *Proc. - Int. Conf. Dev. eSystems Eng. DeSE*, 2018.

[10] Z. Wang, "Deep Learning-Based Intrusion Detection with Adversaries," *IEEE Access*, 2018.

[11] S. M. Kasongo and Y. Sun, "A deep learning method with wrapper based feature extraction for wireless intrusion detection system," *Comput. Secur.*, 2020.

[12] O. Faker and E. Dogdu, "Intrusion detection using big data and deep learning techniques - UNSW-NB15 - CICIDS2017," *ACMSE 2019 - Proc. 2019 ACM Southeast Conf.*, 2019.

[13] J. Yan, D. Jin, C. W. Lee, and P. Liu, "A Comparative Study of Off-Line Deep Learning Based Network Intrusion Detection," in *International Conference on Ubiquitous and Future Networks, ICUFN*, 2018.

[14] S. Choudhary and N. Kesswani, "Analysis of KDD-Cup'99, NSL-KDD and UNSW-NB15 Datasets using Deep Learning in IoT," *Procedia Comput. Sci.*, 2020.

[15] T. Janarthanan and S. Zargari, "Feature selection in UNSW-NB15 and KDDCUP'99 datasets," in *IEEE International Symposium on Industrial Electronics*, 2017.

[16] N. Moustafa and J. Slay, "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.*, 2017.

[17] M. V, "A Survey on Performance Analysis through Dimensional Reduction and Classification Algorithm using KDD Cup and UNSW-NB15 Dataset," *Int. J. Res. Appl. Sci. Eng. Technol.*, 2019.

[18] C. Xu, J. Shen, X. Du, and F. Zhang, "An Intrusion Detection System Using a Deep Neural Network with Gated Recurrent Units," *IEEE Access*, 2018

.