Evolutionary Algorithm-based Feature Selection for an Intrusion Detection System

Devendra Kumar Singh
Department of Computer Science & Engineering
Guru Ghasidas University
Bilaspur, India
devendra.singh1700@gmail.com

Manish Shrivastava
Department of Computer Science & Engineering
Guru Ghasidas University
Bilaspur, India
mannbsp@gmail.com

Abstract—Keeping computer reliability to confirm reliable, secure, and truthful correspondence of data between different enterprises is a major security issue. Ensuring information correspondence over the web or computer grids is always under threat of hackers or intruders. Many techniques have been utilized in intrusion detections, but all have flaws. In this paper, a new hybrid technique is proposed, which combines the Ensemble of Feature Selection (EFS) algorithm and Teaching Learning-Based Optimization (TLBO) techniques. In the proposed, EFS-TLBO method, the EFS strategy is applied to rank the features for choosing the ideal best subset of applicable information, and the TLBO is utilized to identify the most important features from the produced datasets. The TLBO algorithm uses the Extreme Learning Machine (ELM) to choose the most effective attributes and to enhance classification accuracy. The performance of the recommended technique is evaluated in a benchmark dataset. The experimental outcomes depict that the proposed model has high predictive accuracy, detection rate, false-positive rate, and requires less significant attributes than other techniques known from the literature.

Keywords-classification; feature selection; teaching learning-based optimization; intrusion detection

I. INTRODUCTION

Despite the increasing alertness in cyber-security problems, the present ongoing solutions are not suitable for shielding computer applications or undertaking security frameworks against the risk from consistently ever-propelling organized assaults [1]. Adaptable security methods were developed to solve this issue, but they turned out to have further issues. Typical cyber-security methods are insufficient to entirely defend web computer security, since they face problems from intruder attempts at the initiation of the security procedure, e.g. user authentication and firewalling [2, 3]. Thus, a different line of threat protection is acutely mentioned in the Intrusion Detection Systems (IDSs). An IDS is a program that observes the internet for venomous actions and policy contraventions [4]. At present, IDS along with safe-guard applications have resolved into an indispensable element of computer security of most companies. The union of the above-mentioned security machines provides increased opposition in network-attacks and improves network security.

II. RELATED WORK

Gradually, inexhaustible applications, e.g. choice and order models have been put in to intrusion datasets (i.e. KDD CUP 1999) for detecting network problems and attacks. Attribute selection with learning algorithms couldn't control or scale to very large volumes of datasets [5]. To beat this impediment, authors in [6, 7] proposed another hybrid feature selection technique that diminishes the non-applicable features and selects the best ideal component subsets. The recurring pattern study in [8] indicated that individual hunt calculation locates the most suitable subsets that amplify information over-fitting, while a probing interest is less prone to information over-fitting in the part assurance, developing a modest number of tests [9]. Authors in [10] proposed the use of ELM and alpha profiling to diminish the required time while superfluous highlights were disposed of utilizing a group of separated, relationship and consistency-based feature selection procedures.

A. Filter-based Methods

Optimum and appropriate feature subset selection is a task accomplished by choosing the qualities dependent on the high connections of concerning classes and uncorrelated features. From the Conditional Mutual Information Maximization (CMIM) method, Feature/attributes Subset Selection (FSS) is conducted depending on maximizing conditional mutual information [11] regarding the class. In addition, it is extremely close with class attributes and uncorrelated to attributes. It makes a compromise between the predictive power of the nominated competitor (significance for the class carrier) and its freedom from all recently chosen attributes. Mutual Information (MI) estimation between the class label y and attributes X is calculated in:

$$I(y;X) = H(y) - H(\frac{y}{X}) \quad (1)$$

where $H(y)$ and $H(\frac{y}{X})$ show the entropy and conditional entropy of the class change respectively. Some writers have mentioned issues using the Mutual Information-based Feature Selection (MIFS) technique [7, 12]. Therefore, we used this strategy to decrease the readability between class y and data attributes as shown in (2). The primary objective of CMIM is to choose the final feature subset that conveys as much information as possible from the sample S:

Corresponding author: Devendra Kumar Singh

$$M_{CMIM}(X) = \underset{x_j \in S}{min} \; I\left(y; {}^{x_k}/_{x_j}\right) \quad (2)$$

where M_{CMIM} estimates the mutual information between the full features set x_k and certain features x_j regarding class label *y*, whereas *S* shows the subsets of the selected features. $I\left(y; {}^{x_k}/_{x_j}\right)$ measures the quantity of the classification information that x_k affords when x_j has been selected [13]. Selected feature subset *S* cannot provide this information. As comparison to $I\,(y; x_k)$, $I\left(y; {}^{x_k}/_{x_j}\right)$ does not contain the superfluous data of pair wise attributes for categorization.

The importance of the input attributes defined by the JMI is shown in (3):

$$M_{JMI}(X) = \sum_{x_j \in S} I\left(x_k; x_j; y\right) \propto \sum_{x_j \in S} I\left(y; {}^{x_k}/_{x_j}\right) \quad (3)$$

where $I\left(x_k; x_j; y\right)$ signifies the mutual information between the novel attribute subset x_k and the selected attributes x_j regarding class *y*. In linguistics of mutual information, the determination of attribute choice is to reduce attribute subsets *S* with *N* attributes with a maximum holding on the target class *c*. This structure, called Max-Dependency, has the form of:

$$max\, w(X, y) = I(y; x_1, x_2, \ldots, x_N) = H(y)H\left(\frac{y}{x_1, x_2, \ldots, x_N}\right) \quad (4)$$

In (3), the holding among attribute *X* is resultant and can have a high value [14]. The correspondence between readability between attributes is expressed in (5) and (6):

$$\min \mathrm{Z}\,(\mathrm{X}, \mathrm{c}\,) = 1/|\, s^2\,| \sum_{x_j \in s} I\left(x_j; x_k\right) \quad (5)$$

$$\mathrm{Max}\, \emptyset(w, Z) = \mathrm{w} - \mathrm{Z} \quad (6)$$

The incorporation (i.e. integration) of (5) and (6) is known as the Minimal-Redundancy-Maximal-Relevance (mRMR) [15]:

$$j_{mRMR}(\emptyset) = I(c; X) - 1/|\, s^2\,| \sum_{x_j \in s} I\left(x_j; x_k\right) \quad (7)$$

where x_j is a selected subset of attributes *S* and x_k is a native feature set.

B. The Proposed Ensemble Feature Selection

The pre-owned Feature Selection strategies are mRMR, JMI, and CMIM which can relegate the position of the IDS datasets and the output is aggregated utilizing a combination strategy [7].

C. Frequency Vote

Frequency Vote (FV) is a cooperative decision making framework that has been proposed as more useful than other increasingly complex plans [16]. Thus, we can follow the most voted prediction as to the last prediction or expectation as per (8):

$$\sum_{n=1}^{£} d_{n,j} = argmax_{i \in \{1,2,\cdots,L\}} \sum_{n=1}^{£} d_{n,i} \quad (8)$$

where *£* shows the number of attribute choice methods, and *L* is a selection of some attributes. For attribute *j*, the sum $\sum_{n=1}^{£} d_{n,j}$ tabulates the number of votes for *j*.

D. Using Teaching Learning-Based Optimization (TLBO)

TLBO [17-19] is the best and most powerful metaheuristic method to apply high convegence rate with less adjusting parameters. It is an easy and simple computation of tuning the control parameters with less memory requirments. The working methodology of the TLBO algorithm can create better evaluation outcome [20]. The position of the *i*th learner is :

$$X_{i,k} = \{X_{i,1}, X_{i,2}, \ldots, X_{i,D}\} \quad (9)$$

where L_b shows the lower limit and U_b shows the upper limit of the *D* dimension in the search area $X_{i,D} \in [\,L_b\,, U_b]$ [21]. The learner *X* is unplanned, initialized in the search area [22]. The development (i.e. evolution) of $X_{i,k}$ is generated by:

$$X_{i,k} = L_{b,k} + r_1 * \left(U_{b,k} - L_{b,k}\,\right) \quad (10)$$

where *i*=1, 2, 3, ..., *nPop*, $k = 1,2,3, \ldots, D$, r_1 signifies the unplanned variable, $L_{b,k}$ shows the lower limit and $U_{b,k}$ shows the upper limit value, and *nPop* denotes the population count [23]. The simulation of an old-style initiation procedure is arranged into two critical stages of the TLBO calculation: the teacher stage and the learner stage.

In TLBO algorithm, the teacher is a quantification of obtaining an ideal output gained from optimization problems. Therefore, the teacher can grow the mean result of a classroom to a specific result which relies on the ability of the complete classroom. Let $M_{i,k}$= (1/n*P*op) ($\Sigma X_{i,k}$) be the mean value of the particular subject where *k*=1, 2,…,*D*. Equation (11) shows the updating equation process:

$$X_{i,k}^{new} = X_{i,k}^{old} + r_2 * \left(X_{teacher} - T_f * M_{i,k}\right)$$

$$\&\; T_f = \mathrm{round}[1+\mathrm{rand}(0,1)] \quad (11)$$

where $X_{teacher}$ is the greatest begineer of the embrance group (i.e. population) at the current duplication of the algorithm, r_2 represents random numbers, T_f behaves as a teaching element that chooses the merit of the mean to be changed. In each iteration, $X_{i,k}^{new}$ is updated from the old merit $X_{i,k}^{old}$. $X_{i,k}^{new}$ and $X_{i,k}^{old}$ denote the *k*-th beginer choice after or before it is modernized by the teacher.

E. The Fitness Function

The fitness function must maximize the categorization Accuracy of the calculations accomplished by the best attributes during the progresive (i.e. evolutionary) process, which is defined as:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

TP, *TN*, *FP*, and *FN* stand for True Positive, True Negative, False Positive, and False Negative respectively.

III. Experimental results and discussion

During the experiments, every record had features such as feature name, records, and feature portrayal. Most IDS numerical studies have been performed on NSL-KDD [24]. This data set have varying data importance and feature integrity. Authors in [25] analyzed the deliberate intrusion dataset called KDD Cup 1999 [36]. Every record is tagged as normal or as an attack type in the dataset. The flowchart of the proposed method can be seen in Figure 1.

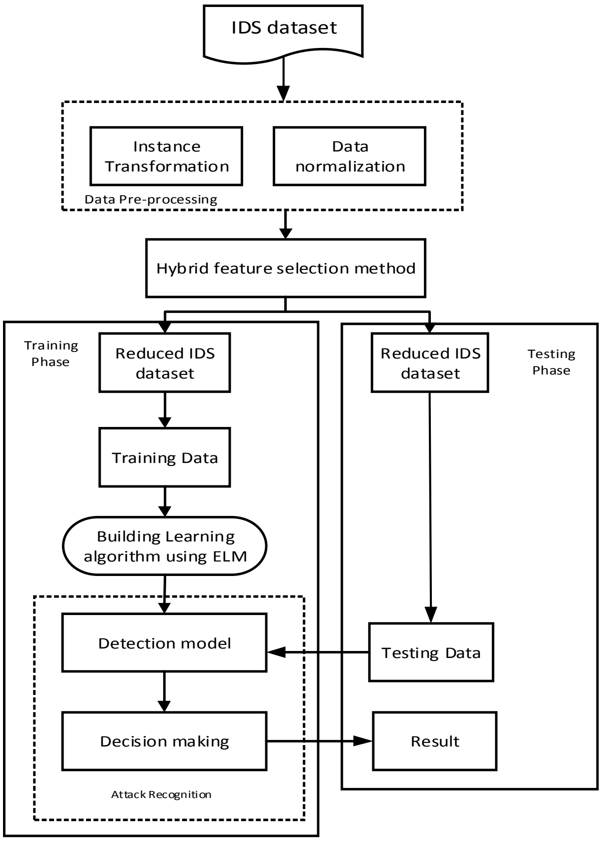

Fig. 1. General structure of the proposed model for intrusion detection.

A. *Results and Discussion*

Different exploratory tools and techniques were used on the NSL-KDD dataset (see Table I) [14, 25, 26]. The classification performance is estimated with the assistance of the support vector machine categorization with four execution variables. These exhibition measures, along with Accuracy, are [27-29]:

$$\text{Detection Rate: DR} = \frac{TP}{TP+FN}$$

$$\text{Precision: Pr} = \frac{TP}{TP+FP}$$

$$\text{F-measure} = \frac{2*Pr*Re}{Pr+Re}$$

$$\text{False Alarm Rate: FAR} = \frac{TP}{TN+FP}$$

B. *Result Comparison*

Tenfold cross-validation was applied to ELM [7] and other classifiers, namely SVM [30, 31] and NB in the IDS dataset. Table I shows the comparison of the performance of the proposed algorithm with existing known algorithms. The result shows that the proposed algorithm performs better on the basis of parameters like feature, DR, FPR and Accuracy in the same data set. Only five of the attributes have been selected by the proposed method which can identify intrusion attacks in the network with maximum Accuracy.

TABLE I. PERFORMANCE COMPARISON OF THE PROPOSED AND EXISTING ALGORITHMS IN THE SAME DATABASE

Algorithm	Feature	DR	FPR	Accuracy
LSSVM-IDS + FMIFS [32]	18	98.93	0.28	99.94
TUIDS [33]	All	98.88	1.12	96.55
HTTP based IDS [34]	13	99.03	1.0	99.38
Hybrid IDS [35]	All	99.10	1.2	*
Proposed	5	99.31	0.19	99.95

IV. CONCLUSION

In this study, a novel hybrid model called EFS-TLBO with ELM is proposed, to easily identify threats by using the attribute choice algorithm [7] which increases the perceptive power for better class distinction. For exhibiting the superiority of the proposed technique, the NSL-KDD database of intruders was employed. The results show that the proposed technique provides an important depletion to the required features and outperforms the advanced attribute selection techniques from the literature. The practical results show that the suggested technique achieved an accuracy of 99.95% in the NSL-KDD data set of intruders [36, 37], surpassing the other techniques.

Future work is going to be focused on multi-objective algorithms that combine ensemble filter and classification methods for pattern analysis and intrusion attack detection. Also, some different optimization algorithms for ELM parameter optimization are going to be researched.

REFERENCES

[1] A. Praseed and P. S. Thilagam, "DDoS Attacks at the Application Layer: Challenges and Research Perspectives for Safeguarding Web Applications," *IEEE Communications Surveys Tutorials*, vol. 21, no. 1, pp. 661–685, Firstquarter 2019, https://doi.org/10.1109/COMST.2018.2870658.

[2] S. Pontarelli, G. Bianchi, and S. Teofili, "Traffic-Aware Design of a High-Speed FPGA Network Intrusion Detection System," *IEEE Transactions on Computers*, vol. 62, no. 11, pp. 2322–2334, Nov. 2013, https://doi.org/10.1109/TC.2012.105.

[3] N. Fallahi, A. Sami, and M. Tajbakhsh, "Automated flow-based rule generation for network intrusion detection systems," in *24th Iranian Conference on Electrical Engineering*, Shiraz, Iran, May 2016, pp. 1948–1953, https://doi.org/10.1109/IranianCEE.2016.7585840.

[4] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R. P. Liu, "A System for Denial-of-Service Attack Detection Based on Multivariate Correlation Analysis," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 2, pp. 447–456, Feb. 2014, https://doi.org/10.1109/TPDS.2013.146.

[5] Y. Wang, W. Meng, W. Li, J. Li, W.-X. Liu, and Y. Xiang, "A fog-based privacy-preserving approach for distributed signature-based intrusion detection," *Journal of Parallel and Distributed Computing*, vol. 122, pp. 26–35, Dec. 2018, https://doi.org/10.1016/j.jpdc.2018.07.013.

[6] S. A. Medjahed, T. A. Saadi, A. Benyettou, and M. Ouali, "Kernel-based learning and feature selection analysis for cancer diagnosis," *Applied Soft Computing*, vol. 51, pp. 39–48, Feb. 2017, https://doi.org/10.1016/j.asoc.2016.12.010.

[7] A. K. Shukla and P. Singh, "Building an Effective Approach toward Intrusion Detection Using Ensemble Feature Selection," *International*

Journal of Information Security and Privacy, vol. 13, no. 3, pp. 31–47, 2019.

[8] T. F. Ghanem, W. S. Elkilani, and H. M. Abdul-kader, "A hybrid approach for efficient anomaly detection using metaheuristic methods," *Journal of Advanced Research*, vol. 6, no. 4, pp. 609–619, Jul. 2015, https://doi.org/10.1016/j.jare.2014.02.009.

[9] S. Dwivedi, M. Vardhan, and S. Tripathi, "Incorporating evolutionary computation for securing wireless network against cyberthreats," *The Journal of Supercomputing*, vol. 76, no. 11, pp. 8691–8728, Nov. 2020, https://doi.org/10.1007/s11227-020-03161-w.

[10] R. Singh, H. Kumar, and R. K. Singla, "An intrusion detection system using network traffic profiling and online sequential extreme learning machine," *Expert Systems with Applications*, vol. 42, no. 22, pp. 8609–8624, Dec. 2015, https://doi.org/10.1016/j.eswa.2015.07.015.

[11] A. K. Shukla, P. Singh, and M. Vardhan, "A two-stage gene selection method for biomarker discovery from microarray data for cancer classification," *Chemometrics and Intelligent Laboratory Systems*, vol. 183, pp. 47–58, Dec. 2018, https://doi.org/10.1016/j.chemolab.2018.10.009.

[12] C. Liu, W. Wang, Q. Zhao, X. Shen, and M. Konan, "A new feature selection method based on a validity index of feature subset," *Pattern Recognition Letters*, vol. 92, pp. 1–8, Jun. 2017, https://doi.org/10.1016/j.patrec.2017.03.018.

[13] P. E. Meyer, C. Schretter, and G. Bontempi, "Information-Theoretic Feature Selection in Microarray Data Using Variable Complementarity," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 3, pp. 261–274, Jun. 2008, https://doi.org/10.1109/JSTSP.2008.923858.

[14] A. K. Shukla, P. Singh, and M. Vardhan, "A hybrid gene selection method for microarray recognition," *Biocybernetics and Biomedical Engineering*, vol. 38, no. 4, pp. 975–991, Jan. 2018, https://doi.org/10.1016/j.bbe.2018.08.004.

[15] A. K. Shukla, P. Singh, and M. Vardhan, "Gene selection for cancer types classification using novel hybrid metaheuristics approach," *Swarm and Evolutionary Computation*, vol. 54, May 2020, Art. no. 100661, https://doi.org/10.1016/j.swevo.2020.100661.

[16] L. Barolli and O. Terzo, Eds., *Complex, Intelligent, and Software Intensive Systems*, 1st edition. New York, NY, USA: Springer, 2017.

[17] R. V. Rao, V. J. Savsani, and D. P. Vakharia, "Teaching–learning-based optimization: A novel method for constrained mechanical design optimization problems," *Computer-Aided Design*, vol. 43, no. 3, pp. 303–315, Mar. 2011, https://doi.org/10.1016/j.cad.2010.12.015.

[18] A. Rajasekhar, R. Rani, K. Ramya, and A. Abraham, "Elitist Teaching Learning Opposition based algorithm for global optimization," in *IEEE International Conference on Systems, Man, and Cybernetics*, Seoul, Korea (South), Oct. 2012, pp. 1124–1129, https://doi.org/10.1109/ICSMC.2012.6377882.

[19] Y. Oubbati and S. Arif, "Transient stability constrained optimal power flow using teaching learning based optimization," in *8th International Conference on Modelling, Identification and Control*, Algiers, Algeria, Nov. 2016, pp. 284–289, https://doi.org/10.1109/ICMIC.2016.7804124.

[20] A. K. Shukla, S. K. Pippal, and S. S. Chauhan, "An empirical evaluation of teaching–learning-based optimization, genetic algorithm and particle swarm optimization," *International Journal of Computers and Applications*, pp. 1–15, Nov. 2019, https://doi.org/10.1080/1206212X.2019.1686562.

[21] M. M. Polycarpou, A. de Carvalho, J.-S. Pan, M. Wozniak, H. Quintian, and E. Corchado, Eds., *Hybrid Artificial Intelligence Systems.* New York, NY, USA: Springer International Publishing, 2014.

[22] A. K. Shukla, P. Singh, and M. Vardhan, "An adaptive inertia weight teaching-learning-based optimization algorithm and its applications," *Applied Mathematical Modelling*, vol. 77, pp. 309–326, Jan. 2020, https://doi.org/10.1016/j.apm.2019.07.046.

[23] P. K. Nayak, S. Mishra, P. K. Dash, and R. Bisoi, "Comparison of modified teaching–learning-based optimization and extreme learning machine for classification of multiple power signal disturbances," *Neural Computing and Applications*, vol. 27, no. 7, pp. 2107–2122, Oct. 2016, https://doi.org/10.1007/s00521-015-2010-0.

[24] A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," *IEEE Communications Surveys Tutorials*, vol. 18, no. 2, pp. 1153–1176, Secondquarter 2016, https://doi.org/10.1109/COMST.2015.2494502.

[25] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *IEEE Symposium on Computational Intelligence for Security and Defense Applications*, Ottawa, Canada, Jul. 2009, pp. 1–6, https://doi.org/10.1109/CISDA.2009.5356528.

[26] G. V. Nadiammai and M. Hemalatha, "Effective approach toward Intrusion Detection System using data mining techniques," *Egyptian Informatics Journal*, vol. 15, no. 1, pp. 37–50, Mar. 2014, https://doi.org/10.1016/j.eij.2013.10.003.

[27] K. Ray, S. N. Sharan, S. Rawat, S. K. Jain, S. Srivastava, and A. Bandyopadhyay, Eds., *Engineering Vibration, Communication and Information Processing*, 1st edition. New York, NY, USA: Springer, 2018.

[28] N. Shakhovska, Ed., *Advances in Intelligent Systems and Computing: Selected Papers from the International Conference on Computer Science and Information Technologies, ... in Intelligent Systems and Computing, 512)*, 1st ed. New York, NY: Springer, 2016.

[29] S. R. Basha and J. K. Rani, "A Comparative Approach of Dimensionality Reduction Techniques in Text Classification," *Engineering, Technology & Applied Science Research*, vol. 9, no. 6, pp. 4974–4979, Dec. 2019, https://doi.org/10.48084/etasr.3146.

[30] F. Kuang, S. Zhang, Z. Jin, and W. Xu, "A novel SVM by combining kernel principal component analysis and improved chaotic particle swarm optimization for intrusion detection," *Soft Computing*, vol. 19, no. 5, pp. 1187–1199, May 2015, https://doi.org/10.1007/s00500-014-1332-7.

[31] J. Gu, L. Wang, H. Wang, and S. Wang, "A novel approach to intrusion detection using SVM ensemble with feature augmentation," *Computers & Security*, vol. 86, pp. 53–62, Sep. 2019, https://doi.org/10.1016/j.cose.2019.05.022.

[32] S. Senthamarai Kannan and N. Ramaraj, "A novel hybrid feature selection via Symmetrical Uncertainty ranking based local memetic search algorithm," *Knowledge-Based Systems*, vol. 23, no. 6, pp. 580–585, Aug. 2010, https://doi.org/10.1016/j.knosys.2010.03.016.

[33] P. Gogoi, M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Packet and Flow Based Network Intrusion Dataset," in *International Conference on Contemporary Computing*, Noida, India, Aug. 2012, pp. 322–334, https://doi.org/10.1007/978-3-642-32129-0_34.

[34] M. M. Abd-Eldayem, "A proposed HTTP service based IDS," *Egyptian Informatics Journal*, vol. 15, no. 1, pp. 13–24, Mar. 2014, https://doi.org/10.1016/j.eij.2014.01.001.

[35] G. Kim, S. Lee, and S. Kim, "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection," *Expert Systems with Applications*, vol. 41, no. 4, Part 2, pp. 1690–1700, Mar. 2014, https://doi.org/10.1016/j.eswa.2013.08.066.

[36] "NSL-KDD Datasets," *Canadian Institute for Cybersecurity | UNB*. https://www.unb.ca/cic/datasets/nsl.html (accessed Apr. 23, 2021).

[37] A. S. A. Aziz, S. E.-O. Hanafi, and A. E. Hassanien, "Comparison of classification techniques applied for network intrusion detection and classification," *Journal of Applied Logic*, vol. 24, pp. 109–118, Nov. 2017, https://doi.org/10.1016/j.jal.2016.11.018.

AUTHORS PROFILE

Devendra Kumar Singh got his B.E. from the B.U. University of Bhopal (MP) in 2000 and his M.E from AAI-DU Allahabad (UP) in 2006. He is currently pursuing his Ph.D. and works as an Assistant Professor in the Department of Computer Science & Engineering, SOS, E&T, GGV Bilaspur (CG) since 2005. He is a member of AIENG. He has published more than 10 research papers in reputed international journals. His main research work focuses on Intrusion Detection Systems. He has 15 years of teaching experience and 6 years of research experience.

Manish Shrivastava, is an Assistant Professor at the Department of Computer Science & Engineering, Institute of Technology, Guru Ghasidas University, Bilaspur. he obtained his M.Tech. Degree from DAVV, Indore, and Ph.D. from Guru Ghasidas University, Bilaspur. He has about 15 years of teaching and research experience, and a number of papers in various national and international journals to his credit. His field of interest is network security. He has served as a chairman of the board of studies, computer science & engineering, Guru Ghasidas, Bilaspur. He is a life member of the Indian Society for Technical Education and a senior member of IACSIT .