AN ENHANCEMENT OF CLASSIFICATION TECHNIQUE BASED ON ROUGH SET THEORY FOR INTRUSION DETECTION SYSTEM APPLICATION

NOOR SUHANA SULAIMAN

DOCTOR OF PHILOSOPHY

(COMPUTER SCIENCE)

UNIVERSITI MALAYSIA PAHANG

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Doctor of Philosophy (Computer Science).

_____

(Supervisor's Signature)

Full Name : DR ROHANI BT ABU BAKAR

Position : ASSOCIATE PROFESSOR

Date :

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

_____

(Student's Signature)

Full Name : NOOR SUHANA BT SULAIMAN

ID Number : PCC 11001

Date :

AN ENHANCEMENT OF CLASSIFICATION TECHNIQUE BASED ON ROUGH SET THEORY FOR INTRUSION DETECTION SYSTEM APPLICATION

NOOR SUHANA SULAIMAN

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Doctor of Philosophy (Computer Science)

Faculty of Computer Systems & Software Engineering

UNIVERSITI MALAYSIA PAHANG

APRIL 2019

ACKNOWLEDGEMENTS

In the name of ALLAH, Most Merciful, Most Compassionate. Alhamdulillah, all praises to Allah for the strengths and blessings in completing this thesis. I would like to express the deepest sincere gratitude to my supervisor, Associate Professor Dr. Rohani Bt Abu Bakar, for her supervision and constant support. Her invaluable help of constructive comments and suggestions throughout the experimental and thesis works have contributed to the success of this research. My sincere appreciation also goes out to all my examiners; Dr. Mohamad Fadli Bin Zolkipli, Professor Dr. Kamal Zuhairi Bin Zamli, Professor Dr. Suhaidi Bin Hassan and Associate Professor Dr. Nur Izura Bt Udzir.

I am also obliged to express my greatest appreciation goes to my beloved parents; Mr. Sulaiman Bin Ramly and Mrs. Norihan Bt Abu Bakar, also goes out too Hj. Syed Noh Bin Syed Ali and Hjh C.A. Halimah Bt K.C. Ahammu. My deepest gratitude goes to my husband, Ansar Sadat bin M. Mohamed Kunju, childrens, Azeem, Ameen, Ameer, Sumayyah, Solehah and Syifaa, and also to my sisters, Noor Suhani Bt Sulaiman and Noor Suhaida Bt Sulaiman for their endless everlasting loves, sacrifices, prayers and encouragements throughout this journey.

I am very grateful towards all the staffs and lecturers of Faculty of Computer Systems & Software Engineering and Institute of Postgraduate Studies, Universiti Malaysia Pahang who have been directly or indirectly influential and supportive to this research. Sincere thanks to all my friends especially Sue, Kak Liza and others for their kindness and moral support during my study. Thanks for the friendship and memories. The sweet memory among us will never fade away. To those who indirectly contributed in this research, your kindness means a lot to me. Thank you very much.

ABSTRAK

Sistem Pengesan Pencerobohan mampu mengesan pencerobohan yang tidak dibenarkan ke dalam sistem dan rangkaian komputer dengan mencari punca serangan yang diketahui atau penyimpangan aktiviti normal. Walau bagaimanapun, prestasi ketepatan adalah salah satu isu dalam aplikasi Sistem Pengesan Pencerobohan. Sementara itu, pengkelasan adalah salah satu teknik dalam perlombongan data yang digunakan untuk meningkatkan prestasi Sistem Pengesan Pencerobohan. Untuk meningkatkan masalah prestasi klasifikasi, algoritma pemilihan ciri dan pembekasan adalah penting dalam memilih sifat yang berkaitan yang dapat meningkatkan prestasi klasifikasi. Algoritma pembekasan telah dicadangkan baru-baru ini akan tetapi, algoritma pembekasan tersebut hanya mampu mengendalikan atribut kategori dan tidak dapat menangani atribut berangka. Di dalam algoritma pembekasan, adalah sukar untuk menentukan bilangan selang dan lebar yang diperlukan. Oleh itu, untuk menangani dataset yang besar, teknik perlombongan data boleh diperbaiki dengan memperkenalkan algoritma yang berupaya untuk meningkatkan prestasi klasifikasi. Generasi peraturan dianggap sebagai proses penting dalam perlombongan data, malahan peraturan yang dihasilkan adalah dalam jumlah besar. Oleh itu, adalah mustahak untuk menentukan peraturan yang penting dan relevan untuk proses seterusnya. Oleh itu, tujuan kajian ini adalah untuk meningkatkan prestasi klasifikasi dari segi ketepatan, kadar pengesanan dan pengurangan kadar penggera positif palsu untuk aplikasi Sistem Pengesan Pencerobohan. Di dalam penyelidikan ini mencadangkan peningkatan algoritma pembekasan berdasarkan Pembekasan Tong dalam Teori Set Kasar untuk meningkatkan prestasi klasifikasi dan juga untuk meningkatkan strategi peraturan generasi dalam Teori Set Kasar dalam meningkatkan prestasi klasifikasi. Kedua-dua penambahbaikan ini dinilai dari segi ketepatan, penggera positif palsu dan kadar pengesanan terhadap data KDD Cup 99 dalam aplikasi Sistem Pengesan Pencerobohan. Beberapa algoritma pembekasan seperti Kesamaan Frekuensi Tong, Entropy / MDL, Naïve dan pembekasan yang dicadangkan telah dianalisis dan dibandingkan dalam kajian. Hasil eksperimen menunjukkan teknik yang dicadangkan mampu meningkatkan peratusan klasifikasi ketepatan sehingga 99.95%; dan bilangan tong yang minimum menentukan algoritma pembekasan yang baik. Impak dari kajian penyelidikanyang dicadangkan, peratusan kadar pengesanan serangan adalah meningkat dan kadar penggera positif palsu diminimumkan. Algoritma yang dicadangkan menghasilkan kompromi yang memuaskan antara bilangan tong dan juga ketepatan prestasi teknik klasifikasi.

ABSTRACT

An Intrusion Detection System (IDS) is capable to detect unauthorized intrusions into computer systems and networks by looking for signatures of known attacks or deviations of normal activity. However, accuracy performance is one of the issues in IDS application. Meanwhile, classification is one of techniques in data mining employed to increase IDS performance. In order to improve classification performance problem, feature selection and discretization algorithm are crucial in selecting relevant attributes that could improve classification performance. Discretization algorithms have been recently proposed; however, those algorithms of discretizer are only capable to handle categorical attributes and cannot deal with numerical attributes. In fact, it is difficult to determine the needed number of intervals and their width. Thus, to deal with huge dataset, data mining technique can be improved by introducing discretization algorithm to increase classification performance. The generation of rule is considered a crucial process in data mining and the generated rules are in a huge number. Therefore, it is dreadful to determine important and relevant rules for the next process. As a result, the aim of the study is to improve classification performance in terms of accuracy, detection rate and false positive alarm rate decreased for IDS application. Henceforth, to achieve the aim, current research work proposed an enhancement of discretization algorithm based on Binning Discretization in RST to improve classification performance and to enhance the strategy of generation rules in RST to improve classification performance. Both enhancements were evaluated in terms of accuracy, false positive alarm and detection rate against state-of-the-practice dataset (KDD Cup 99 dataset) in IDS application. Several discretization algorithms such Equal Frequency Binning, Entropy/MDL, Naïve and proposed discretization were analysed and compared in the study. Experimental results show the proposed technique increases accuracy classification percentage up to 99.95%; and the minimum number of bins determine good discretization algorithm. Consequently, attack detection rate increases and false positive alarm rate minimizes. In particular, the proposed algorithm obtains satisfactory compromise between the number of cuts and classification accuracy.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

IoT	Internet of Things
UCL	University College London
ARPANET	Advanced Research Projects Agency Network
TCP	Transport Control Protocol
SSL	Secure Socket Layer
TLS	Transport Layer Security
HTTPS	Hyper Text Transfer Protocol Secure
DNS	Domain Name System
DDoS	Distributed Denial of Service
IT	Information Technology
WWW	World Wide Web
KDD	Knowledge Discovery and Data Mining
RST	Rough Set Theory
NIDS	Network intrusion detection systems
HIDS	Host Intrusion Detection System
DoS	Denial of Service
R2L	Remote to Local
U2R	Remote to User
GA	Genetic Algorithm
DT	Decision Tree
NNIV-RS	Neural Network with Indicator Variable Using Rough Set Theory for Attribute Reduction
ROC	Receiver Operating Characteristic
AUC	Area Under Curve
TP	True Positive
FN	False Negative
TN	True Negative
FP	False Positive
FPR	False Positive Rate
FAR	False Caution Rate
FNR	False Negative Rate
TPR	True Positive Rate

TNR	True Negative Rate
MIT	Massachusetts Institute of Technology
DARPA	Defense Advanced Research Projects Agency
RIM	Rule Important Measure
MADAM ID	Mining Audit Data for Automated Model for Intrusion Detection

LIST OF APPENDICES

REFERENCES

Aburomman, A. A., & Reaz, M. B. I. (2013). Evolution of Intrusion Detection System Based on Machine Learning Methods. *Australian Journal of Basic and Applied Sciences, 7*(7), 799-813.

Aggarwal, P., & Sharma, S. K. (2015). An Empirical Comparison of Classifiers to Analyze Intrusion Detection. *Procceding of the 5th International Conference on Advanced Computing & Communication Technologies 2015, IEEE,* India, 446-450.

Agrawal, R., & Srikant, R. (1994). Fast Algorithms For Mining Association Rules. *20th Procceding of the International Conference on Very Large Databases,* Chile, 487–499.

Ahmad, P., Qamar, S., & Rizvi, S. Q. A. (2015). Techniques of Data Mining in Healthcare: A Review. *International Journal of Computer Applications, 120* (15), 38-50.

Akbar, S., Rao, D. K. N., & Chandula, J. A. (2010). Intrusion Detection System Methodologies Based on Data Analysis. *International Journal of Computer Application, 5*(2), 10-20.

Allen, J. (1990). Boolean Analysis. *Boolean Reasoning* (pp. 87-122). Retrieved from http://www2.fiit.stuba.sk/~kvasnicka/Free%20books/Brown_Boolean%20Reasoning.pdf (last accessed January, 2013).

Altwaijry, H., & Algarny, S. (2012). Bayesian based Intrusion Detection System. *Journal of King Saud University - Computer and Information Sciences, 24*(1), 1-6.

Anderson, J. (1980). *Computer Security Threat Monitoring and Surveillance.* (Report No 79F296400). Fort Washington: James P. Anderson Co.

Atilla, O., & Hamit, E. (2016). A Review of KDD99 Dataset Usage in Intrusion Detection and Machine Learning between 2010 and 2015. *Journal of PeerJ PrePrints, 4*, 1–21.

Azevedo, P. J., & Jorge, M. (2007). Comparing Rule Measures for Predictive Association Rules. *Procceding of the 18th Conference of European on Machine Learning,* Poland, 510-517.

Bahrainian, S., & Dangel, A. (2013). Sentiment Analysis uses Sentiment Features. Lecture Notes in Computer Science, Vol: 4701. *Procceding of the International Joint Conference of Web Intelligence and Intelligent Agent Technologies* (pp. 26–29).

Biswanath, M., Todd , L. H., & Karl N. L. (1994). Network Intrusion Detection. *IEEE Network, 8*(3), 26–41.

Borgohain, R. (2012). FuGeIDS: Fuzzy Genetic Paradigms in Intrusion Detection Systems. *International Journal of Advanced Networking and Applications, 3*(6), 1409-1415.

Bose, I. (2006). Deciding the Financial Health of Dot-Coms using Rough Sets. *Journal of Information and Management*, *43*(7), 835-846.

Bruha, I. (1997). Quality of Decision Rules: Definitions and Classification Schemes for Multiple Rules. *Machine Learning and Statistics: The Interface* (pp. 107–131). New York : John Wiley.

Carvalho, D. R., Freitas, A. A., & Ebecken, N. (2005). Evaluating the Correlation Between Objective Rule Interestingness Measures and Real Human Interest. Lecture Notes in Computer Science, Vol: 3721. *Procceding of the Conference of European on Principles of Data Mining and Knowledge Discovery* (pp. 453-461).

Catania, C. A., Bromberg, F., & Garino, C. G. (2012). An Autonomous Labeling Approach to Support Vector Machines Algorithms for Network Traffic Anomaly Detection. *Expert Systems with Applications, 39*(2), 1822-1829.

Catlett, J. (1991). On Changing Continuous Attributes into Ordered Discrete Attributes. Lecture Notes in Computer Science, Vol: 482. *Procceding of the European Working Session on Learning* (pp. 164-168).

Chen, X., Vorvoreanu, M., & Madhavan, K. (2014). Mining Social Media Data to Understand Student's Learning Experiences. *IEEE Transaction on Learning Technologies*, *7*(3), 246–259.

Chi, C., Wee, P. T., & Guang, B. H. (2012). Extreme Learning Machines for Intrusion Detection. *Procceding of the 2012 International Joint Conference on Neural Networks,* USA, 1-8.

Dai, J. H., & Li, Y. X. (2002). Study on Discretization based on Rough Set Theory. *1st Procceding of the International Conference on Machine Learning and Cybernetics*, China, 3, 1371-1373.

Debar, H. (2009). An Introduction to Intrusion Detection Systems. *IBM Research, Zurich Research Laboratory*, 1-18.

Desale, K. S., & Ade, R. (2015). Genetic Algorithm based Feature Selection Approach for Effective Intrusion Detection System. *Procceding of the 3rd International Conference of Computer Communication and Informatics,* India, 1-6.

Devendra, K., & Jain, R. C. (2012). Improve Intrusion Detection using Decision Tree with Sampling. *International Journal of Computer Technology & Applications*, *3*(3), 1209–1216.

Dhakar, M., & Tiwari, A. (2014). A Novel Data Mining based Hybrid Intrusion Detection Framework. *Journal of Information and Computing Science*, *9*(1), 037–048.

Dougherty, J., Kohavi, R., & Sahami, M. (1995). Supervised and Unsupervised Discretization of Continuous Features. *Procceding of the 12th International Conference on Machine Learning*, USA, 194–202.

Dougherty, J., Kohavi, R., & Sahami, M. (1995). Supervised and Unsupervised Discretization of continuous Features. *Procceding of the 12th International Conference on Machine Learning*, USA, 194–202.

Elfeshawy, N. A., & Faragallah, O. S. (2012). Divided Two Part Adaptive Intrusion Detection System. *Springer Science+Business Media*, 301-321.

Garcia, S., Luengo, J., Saez,J., Lopez,V., & Herrera, F. (2013). Survey of Discretization Techniques: Taxonomy and Empirical Analysis in Supervised Learning. *IEEE Transactions on Knowledge and Data Engineering*, *25*(4), 734-750.

Gautam, G., & Yadav, D. (2014). Sentiment Analysis of Twitter Data using Machine Learning Approaches and Semantic Analysis. *Procceding of the 7th International Conference on Contemporary Computing*, India, 437–442.

Gera, M., & Goel, S. (2015). Data Mining - Techniques, Methods and Algorithms: A Review on Tools and their Validity. *International Journal of Computer Applications*, *113*(18), 22-29.

Ghorbani, A. A., Lu, W., & Tavallaee, M. (2010). Network Intrusion Detection and Prevention Concepts and Techniques. *Advances in Information Security, Springer*.

Gokulakrishnan, B., Plavnathan, P.,Thiruchittampalam, R., Prasat, N., & Ashehan, P. (2012). Opinion Mining and Sentiment Analysis on a Twitter Data Stream. *Procceding of the International Conference on Advances in ICT for Engineering Regions*, Sri Lanka, *7*(3), 182–188.

Govindarajan, M., & Chandrasekaran, R. M. (2011). Intrusion Detection Using Neural Based Hybrid Classification Methods. *International Journal of Computer and Telecommunications Networking*, *55*(8), 1662-1671.

Griner, P. F., Mayewski, R. J., Mushlin, A. I. & Greenland, P. (1981). Selection And Interpretation of Diagnostic Tests and Procedures. *Annals of Internal Medicine*, *94*(4), 557–-592.

Guoyong, W. (2001). Rough Sets Theory and Knowledge Acquisition. *Xi'an Jiaotong University Press.*

Hacibeyoglu, M., Arslan, A., & Kahramanli, S. (2011). Improving Classification Accuracy with Discretization on Datasets Including Continuous Valued Features. *International Journal of Computer and Information Engineering*, *5*(6), 623–626.

Hamid, Y., Sugumaran, M., & Journaux, L. (2016). Machine Learning Techniques for Intrusion Detection: A Comparative Analysis. *International Journal of Computer Application*, 1-6.

Han, J., Kamber, M. (2006). Datamining: Concepts and Techniques. *Morgan Kaufmann*. 3rd Edition.

Hassan, M. M. M. (2013). Network Intrusion Detection System using Genetic Algorithm and Fuzzy Logic. *International Journal of Innovative Research in Computer and Communication Engineering*, *1*(7), 1435-1445.

Hopgood, B. (2001). History of the Web, *Oxford Brookes University*.

Horng, S. J., Su, M. Y., Yuan, H. C., Tzong, W. K., Rong, J. C., Jui, L. L., & Citra, D. P. (2011). A Novel Intrusion Detection System Based on Hierarchical Clustering and Support Vector Machines. *Expert Systems with Applications*, *38*(1), 306-313.

KDD Cup 1999 Data. (1999). *The UCI KDD Archive* [Data file). Retrieved from http://kdd.ics.uci.edu/databases/KDD Cup99/task.html (last accessed January, 2012).

Hui, J. Y. (2016). Research of Network Intrusion Detection System based on Machine Learning and Rough Set Theory. *Advanced Science and Technology Letters*, 120–124.

Internet Users. (2016). Internet World Stats. Retrieved from https://www.internetworldstats.com/stats.htm (last accessed January, 2017).

Jaisankar, N., Ganapathy, S., & Kannan, A. (2012). Intelligent Intrusion Detection System Using Fuzzy Rough Set Based C4 . 5 Algorithm. *Procceding of the International Conference on Advances in Computing, Communications and Informatics*, India, 596–601.

Janmejay, P., Kamlesh, P., Himanshu, P. (2015). Rough Set Approach for Feature Selection in IDS [Special Issue]. *International Journal of Innovations & Advancement in Computer Science*, 4, 8-11.

Jerzy, W. G. B. (2005). Introduction to Rough Set Theory and Applications. *Real World Applications of Computational Intelligence*, 76-147.

Julisch, K., & Dacier, M. (2002). Mining Intrusion Detection Alarms for Actionable Knowledge. *Procceding of the 8th International Conference on Knowledge Discovery and Data Mining, ACM*, 366–375.

Kabir, E., Hu, J., Wang, H., & Zhuo, G. (2018). A Novel Statistical Technique for Intrusion Detection Systems. *Future Generation Computer Systems*, *79*(1).

Kale, S. V., Kale, S. R., Naphade, R. A., & Dande, P. A. A. (2016). A Review of Various Intrusion Detection Approaches. *International Journal of Advanced Research Engineering, Computer Science and Software*, *6*(3), 261–262.

Kaur, H., Singh, G., & Minhas, J. (2013). A Review of Machine Learning Based Anomaly Detection Technique. *International Journal of Computer Applications Technology and Research*, *2*(2), 185-187.

Kaur, K., & Kaur, N. (2015). A Hybrid Approach of Fuzzy C-Mean Clustering and Genetic Algorithm (GA) to Improve Intrusion Detection Rate. *International Journal of Science and Research*, *5*(5), 955-959.

Kausar, N., Samir, B. B., Sulaiman, S. B., Ahmad, I., & Hussain, M. (2012). An Approach Towards Intrusion Detection using PCA Feature Subsets and SVM. *Procceding of the International Conference on Computer & Information Science 2012*, Malaysia, 569-574.

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine Learning and Data Mining Methods in Diabetes Research. *Journal of Computational and Structural Biotechnology*, 104–116.

Kerber, R., & Chimerge. (1992). Discretization of Numeric Value. *Procceding of the National Conference of Artificial Intelligent*, California, 123–128. Retrieved from https://sci2s.ugr.es/keel/pdf/algorithm/congreso/1992-Kerber-ChimErge-AAAI92.pdf (last accessed January, 2014).

Koller, D., & Sahami, M. (1996). Toward Optimal Feature Selection. *International Conference on Machine Learning*, Scotland, 284-292.

Kumar, A., Maurya, H. C., & Mishra, R. (2013). A Research Paper on Hybrid Intrusion Detection System. *International Journal and Advanced Technology*, *2*(4), 294-297.

Kumar, M., & Yadav, N. (2015). Fuzzy Rough Sets and Its Application in Data Mining Field. *Journal of Advances in Computer Science and Information Technology*, *2*(3), 237-240.

Kumar, Y., & Dhawan, S. (2012). A Review on Information Flow in Intrusion Detection System. *International Journal of Computational Engineering & Management*, 15(1), 91–96.

Kurgan, L.A., Cios, K. J. (2004). Caim Discretization Algorithm. *IEEE Transactions on Knowledge and Data Engineering*,*16*(2), 145–153.

Liu, L., Wan, P., Yingmei, W., Songtao, L. (2014). Clustering and Hybrid Genetic Algorithm Based Intrusion Detection Strategy. *Indonesian Journal of Electrical Engineering and Computer Science*, *12*(1), 762-770.

Leandros, A. M., Jianmin, J., Tiago, C. (2014). Integrated OCSVM Mechanism for Intrusion Detection In SCADA Systems. *IET Electronics Letters*, *50*(25), 1935-1936.

Li, J. (2007). Rough Set Based Rule Evaluations and Their Applicatio*n* (Doctoral dissertation, University of Waterloo) Retrieved from https://pdfs.semanticscholar.org/6c1b/9aa86e29577c6bd106c89a3279bc4bb81bd9.pdf (last accessed January, 2013).

Li, J., & Cercone, N. (2005). *Empirical Analysis on the Geriatric Care Dataset Using Rough Sets Theory*. (Report No CS-*2005*-05). Canada: University of Waterloo.

Li, Y., Xia, J., Zhang, S., Yan, J., Xi, X., & Dai, K. (2012). An Efficient Intrusion Detection System Based on Support Vector Machines and Gradually Feature Removal Method. *Expert Systems with Applications*, *39*(1), 424-430.

Lin, S. W., Ying, K. C., Lee, C. Y., & Lee, Z. J. (2012). An Intelligent Algorithm with Feature Selection and Decision Rules Applied to Anomaly Intrusion Detection. *Applied Soft Computing*, *12*(10), 3285-3290.

Liu, H., Hussain, F., Tan, C.L., & Dash, M. (2002). Discretization: An Enabling Technique. *Data Mining and Knowledge Discovery*,6(4), 393–423.

Liyana, N., Shuib, M., Bakar, A. A., & Othman, Z. A. (2011). Performance Study on Data Discretization Techniques Using Nutrition Dataset, *Proceeding of the International Symposium on Computing, Communication, and Control 2009,* Singapore, 304–308.

Majidi, F., Mirzaei, H., Irnapour, T., & Faroughi, F. (2008). A Diversity Creation Method for Ensemble Based Classification: Application in Intrusion Detection Systems, *Proceeding of the7th IEEE International Conference on Cybernetic Intelligent Systems 2008*, UK, 1-5.

Manocha, S., & Irolami, M. A. G. (2007). An Empirical Analysis of the Probabilist C K-Nearest Neighbor Classifier. *Pattern Recognition Letters, 28*(13), 1818-1824.

Metz, C. (1978). Basic Principles of ROC Analysis. *Seminars in Nuclear Medicine*, *8*(4), 283–298.

Ming, H., Wenying, N., & Xu, L. (2009). An Improved Decision Tree Classification Algorithm Based on ID3 and the Application in Score Analysis. *Procceding of the 21st Annual International Conference On Chinese Control And Decision Conference,* China, 1931–1934.

Mohammed, M. N., & Sulaiman, N. (2012). Intrusion Detection System Based on SVM for WLAN. *Procedia Technology*, 313-317.

Mukherjee, S., & Sharma, N. (2012). Intrusion Detection using Naive Bayes Classifier with Feature Reduction. *Procedia Technology*, 119-128.

Neha, G. R., Dharmaraj, R. P. (2015). Implementation of Network Intrusion Detection System using Variant of Decision Tree Algorithm. *Procceding of the International Conference of Nascent Technologies in the Engineering Field, IEEE,* India, 1-5.

Nguyen, H. S. (1998). Discretization Problem for Rough Sets Methods. Lecture Notes in Computer Science, Vol: 1424. *Procceding of the 1st International Conference on Rough Sets and Current Trend in Computing* (pp. 545-552).

Ohrn, A. (1999). *Discernibility and Rough Sets in Medicine: Tools and Applications* (Doctoral dissertation, Norwegian University of Science and Technology). Retrieved from https://wiki.eecs.yorku.ca/course_archive/2011-12/F/4403/_media/ohrn_thesis.pdf (last accessed January, 2013).

Pang, N. T., & Kumar, V. (2000). *Interestingness Measures for Association Patterns: A Perspective.* (Report No TR00-036). University of Minnesota: KDD 2000 Workshop on Postprocessing in Machine Learning and Data Mining.

Panigrahi, A., & Patra, M. R. (2018). Fuzzy Rough Set Based Network Intrusion Detection with Wrapper Subset Evaluator. *International Journal of Engineering Science Invention (IJESI), 7*(2), 51–57.

Pawlak, Z. (1991). Rough Sets, Theoretical Aspects of Reasoning about Data. *Kluwer Academic Publisher.*

Pawlak, Z. (1991). Rough Sets: Theoretical Aspects of Reasoning about Data. *System Theory, Knowledge Engineering and Problem Solving.* Kluwer Academic Publishers. Retrieved from https://www.springer.com/gp/book/9780792314721 (last accessed January, 2014).

Pawlak, Z. (1997). Rough Set Approach to Knowledge-Based Decision Support, *European of Operation Research, 99*(1).

Pawlak, Z., & Skowron, A. (2007). Rough Sets and Boolean Reasoning. *Information Science*, 177, 41-73.

Peddabachigari, S., Abraham, A., Gransen, C., & Thomas, J. (2007). Modeling Intrusion Detection System using Hybrid Intelligent Systems. *Network and Computer Applications*, *30*(1), 114-132.

Rahm, E. (2000). Data Cleaning: Problems and Current Approaches. *IEEE Data Engineering Bulletin*, 3-13. Retrieved from https://betterevaluation.org/sites/default/files/data_cleaning.pdf (last accessed March, 2013).

Rampure, V., & Tiwari, A. (2015). A Rough Set Based Feature Selection on KDD CUP 99 Dataset. *Procceding of the Conference of IEEE Region 10 Conference* 2006, China, *8*(1), 149–156.

Repalle, S. A., & Kolluru, V. R. (2017). Intrusion Detection System using AI and Machine Learning Algorithm. *International Research Journal of Engineering and Technology, 4*(12), 1709-1715.

Rizvi, R. S. H., & Keole, R. R. (2015). A Review on Intrusion Detection System. *International Journal of Advance Research in Computer Science and Management Studies*, *3*(3), 22–28.

Sabri, F. N., Norwawi, N. M., & Seman, K. (2011). Identifying False Alarm Rates for Intrusion Detection System with Data Mining. *International Journal of Computer Science and Network Security*, *11*(4), 95-55.

Sadek, R. A., Soliman, M. S., & Elsayed, H. S. (2013). Effective Anomaly Intrusion Detection System based on Neural Network with Indicator Variable and Rough set Reduction. *International Journal of Computer Science Issue*, *10*(6).

Sengupta, N., Sen, J., Sil, J., & Saha, M. (2013). Designing of Online Intrusion Detection System using Rough Set Theory and Q-Learning Algorithm. *Neurocomputing*, 161–168.

Shen, J., Wang, J., & Ai, H. (2012). An Improved Artificial Immune System-Based Network Intrusion Detection by Using Rough Set. *Communications and Network of Scientific Research*, 41–47.

Singh, B. D., Choudhary, N., & Samota, J. (2013). Analysis of Data Mining Classification with Decision Tree Technique. *Global Journal of Computer Science and Technology Software and Data Engineering*. *13*(13), 1-7. Retrieved from https://pdfs.semanticscholar.org/a0e8/f11b9610b45d9e114fcb8c092c9ced2c3f6a.pdf (last accessed January, 2014).

Software, M. (2018). MedCalc: Easy-To-Use Statistical Software. Retrieved from https://www.medcalc.org/manual/roc-curves.php (last accessed September, 2018).

Srikant, R., & Agrawal, R. (1995). Mining Generalized Association Rules. *Procceding of the Conference of 21st Very large Scale Data Bases*, Switzerland, 407-419.

Sujendran, R., & Arunachalam, M. (2015). Hybrid Fuzzy Adaptive Wiener Filtering with Optimization for Intrusion Detection. *Electronic & Telecommunication Research Institute, 37*(3).

Suresh, C. S., & Anima, N., (2012). Hybridization of Rough Set and Differential Evolution Technique for Optimal Feature Selection. *Procceding of the International Conference on Information Systems Design and Intelligent Applications 2012*, India, 453-460.

Vadim, K. (2018). Overview of Different Approaches to Solving Problems of Data Mining. *Procedia Computer Science*, 234–239.

Wu, H. C., & Huang, S. H. S. (2010). Neural Networks Based Detection of Stepping Stone Intrusion. *Expert Systems with Applications*, *37*(2), 1431-1437.

Wu, S. X., & Banzhaf, W. (2010). The Use of Computational Intelligence in Intrusion Detection Systems: A Review. *Applied Soft Computing*, 1-35.

Wu, S., & Yen, E. (2009). Data Mining Based Intrusion Detectors. *Expert Systems with Applications*, 36(3), 5605-5612.

Yang, Y., Webb, G. I. (2009). Discretization for Naïve Bayes Learning: Managing Discretization Bias and Variance. *Machine Learning*, 39–74.

Ye, C. Z., Yang, J., Geng, D., Zhou, Y., & Chen, N. Y. (2002). Fuzzy Rules to Predict Degree of Malignancy in Brain Glioma. *Medical Biology Computer Engineering*, *40*(2), 145-152.

Yu, B., Byres, E., & Howey, C. (2001). Monitoring Controller's "DNA Sequence" for System Security. *Procceding of the Conference of Emerging Technologies, Instrumentation Systems and Automation Society*, Houston, 1-10.

Zainal, A., Hamid H. Jebur, Mohd, A. M. (2015). Enhancing Rough Set Theory Attributes Selection of Kdd Cup 1999. *Theoretical and Applied Information Technology*, *76*(3), 393-400.

Zhang, X., Jia, L., Shi, H., Tang, Z., & Wang, X. (2012). The Application of Machine Learning Methods to Intrusion Detection. *Proceeding of the Spring Congress on Engineering and Technology 2012*, China, 1-6.

Zimmermann, H. (2001). Fuzzy Set Theory and Its Applications. *Kluwer Academic Publisher*. 4th Edition. Retrieved from https://cours.etsmtl.ca/sys843/REFS/Books/ZimmermannFuzzySetTheory2001.pdf (last accessed January, 2014).

Zweig, M. H., & Campbell, G. (1993). Receiver Operating Characteristic (ROC) Plots a Fundamental Evaluation Tool in Clinical Medicine. *Journal of Clinic Chem*ical, 561–577.