



UWS Academic Portal

An efficient feature selection based Bayesian and rough set approach for intrusion detection

Prasad, Mahendra; Tripathi, Sachin; Dahal, Keshav

Published in:
Applied Soft Computing

DOI:
[10.1016/j.asoc.2019.105980](https://doi.org/10.1016/j.asoc.2019.105980)

Published: 01/02/2020

Document Version
Peer reviewed version

[Link to publication on the UWS Academic Portal](#)

Citation for published version (APA):

Prasad, M., Tripathi, S., & Dahal, K. (2020). An efficient feature selection based Bayesian and rough set approach for intrusion detection. *Applied Soft Computing*, 87, [105980].
<https://doi.org/10.1016/j.asoc.2019.105980>

General rights

Copyright and moral rights for the publications made accessible in the UWS Academic Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact pure@uws.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

An efficient feature selection based Bayesian and Rough set approach for intrusion detection

Mahendra Prasad^{a,*}, Sachin Tripathi^a, Keshav Dahal^b

^a*Department of Computer Science and Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, India*

^b*School of Engineering and Computing, University of West of Scotland, UK*

Abstract

The exponential growth of network size leads to increase attacks and intrusions. Detection of these attacks from the network has turned into a noteworthy issue of security. An intrusion detection system is an important approach to achieves high detection rate. A high dimensional dataset increase complexities of detection systems. In this paper, we have designed a novel intelligent system that comprises the feature selection with a hybrid approach of the Rough set theory and the Bayes theorem. The proposed feature selection computed core features and ranked them based on estimated probability. In a decision system, an object may belong to a single or multiple decision, and a feature contains a set of objects that occurrences compute an estimated probability. The rough set theory is being applied to classify information into lower and upper approximations. Uncertain information is distinguished using rough set approximations and solved by the Bayes theorem. In this research work, it has also been highlighted the quantitative realism of recently generated dataset and compared to publicly available datasets. This approach reduces false alarm rate, computational complexity, training complexity and increases detection rate. Comparisons with relevant classifiers are also tabled that show proposed method performs better than existing classifiers.

*Corresponding author

Email addresses: je.mahendra@gmail.com (Mahendra Prasad), var_1285@yahoo.com (Sachin Tripathi), Keshav.Dahal@uws.ac.uk (Keshav Dahal)

Keywords: Intrusion detection system, CICIDS2017 dataset evaluation, Dataset realism, Feature selection, Rough set theory, Bayes theorem.

1. Introduction

A revolution of ability to exchange information increases security challenges of the communication system. Network security policies prevent accessible network resources from unauthorized access. An Intrusion Detection System (IDS) introduces the detecting, preventing, and repelling of unauthorized access to the communication. The main focus of IDS is to increase the detection rate and decrease the false alarm rate. There are many IDSs designed with the reduction of false alarm rate but failed to perform due to new attacks. These are dynamically changing nature or attack methods. They generate a huge number of traffic connections and extremely imbalanced dataset. Denial-of-Service (DoS) and Distributed DoS (DDoS) are example of flooding attack that shutdown system [1]. Flooding attacks create a problem such as spread malicious packets, exhaust bandwidth, memory overflow, and overloading controller. Botnets are a collection of bots (i.e., hosts) to make a network that remotely controlled by bot-master [2]. These bots increase malicious packets and infected activities like spreading spam, conducting DDoS, etc.

Traditional IDSs are based on authentication, encryption, and decryption mechanism. Firewall is an example of traditional security of first-line defense mechanism that enables by the installation of a computer program on host site. This system is not powerful enough and flexible to protect the network, and due to poor detection rate, it is not sufficient to defend hence attackers easily bypass the system. Anti-virus software is an example of a second-line defense mechanism that has considered limitations of first line defense. Virus detection and prevention software is a signature-based system that works on the only stored signature in the database. [The main limitation of second-line defense mechanism is unable to detect an unseen signature of attacks \[3\].](#)

The IDS is classified into three categories namely misuse detection system,

28 anomaly detection system and hybrid detection system. Misuse detection sys-
29 tem analyzes the packet features and compares with the stored signature from
30 the database. It has many limitations such as (1) unavailability of new attack
31 signature in the database (lack of awareness), (2) a huge number of imbalanced
32 data generated by an attack that increases the database volume, (3) search com-
33 plexity of attack signature. In anomaly detection system, an attack creates a
34 similar profile to the normal profile. It analyzes the packet features and com-
35 putes the deviation whenever any deviation observes from the baseline profile
36 assumes as an attack [4, 5, 6]. A hybrid detection system is used to perform a
37 sequence of misuse and anomaly detection operation. This system continues the
38 detection process whenever it fails to detect then check any deviation from the
39 baseline profile. In all these detection systems, the hybrid system has succeeded
40 to reduce the false alarm rate and deals with limitations [7].

41 We compute qualitative realism of IDS dataset using fuzzy model [8]. Cana-
42 dian Institute for Cybersecurity has recently (in July 2017) generated an IDS
43 dataset named as CICIDS2017 [9] and defined attacks characteristics. It con-
44 tains significant features with many insignificant features that increase the vol-
45 ume of the dataset and computational complexity of the system. Features need
46 to remove them which either marginally or do not contribute to better accu-
47 racy and speed of IDS [3]. These reductions will make better IDS performance
48 in terms of system accuracy and complexities. We introduce a probabilistic
49 method to estimate the probability or importance of the feature. It computes
50 core features and sequences them based on estimated probabilities. This feature
51 selection method is effective that makes system computationally efficient and
52 reduces complexities.

53 The Rough Set Theory (RST) is a mathematical tool to discover hidden
54 patterns of data. It solves many problems such as (1) prediction of missing
55 data, (2) generate sets of decision rules, (3) redundant data elimination, (4)
56 evaluation of significance data, (5) dependency calculation [10], etc. The for-
57 mal approximation of conventional set (crisp set) has termed as a lower and
58 upper approximation in the RST. It provides classification of packets activities

59 in the form of lower and upper approximation [11, 12]. The Bayes theorem
60 is a probabilistic approach that handle ambiguous (boundary region) and un-
61 seen signatures using prior probability of available data. It gains quantitative
62 support from the probabilistic theorem due to their available options. The pro-
63 posed Bayesian rough set method easily resolves uncertainty [13], classification
64 with unseen data, and multi-decision problems. Our major contributions are
65 enumerated below.

- 66 1. Data collection for IDS from the real-world-enterprise network is not an
67 easy task. It is very important to know the quality realism of dataset
68 before deploying it to the system. Fuzzy Logic System (FLS) provides a
69 model [8] for quality evaluation of any IDS dataset. We have measured
70 the quality realism of publicly available CICIDS2017 dataset using FLS
71 based on provided information [9]. This work has also shown qualitative
72 realism of related datasets.
- 73 2. This paper elaborates an efficient feature selection method that examine
74 core features [14] and ranked them using estimated probability which in
75 Section 6.2. It is an efficient learning method that can compute estimated
76 probabilities in an epoch. The higher probability is more significant and
77 zeros ineffective features. These ineffective features increase the only vol-
78 ume of the dataset those are not essential information.
- 79 3. Our proposed hybrid system enlarge the detection capacity and decrease
80 the false alarm rate. This system is a combination of two well-known
81 machine learning techniques which are the RST and Bayes theorem. It
82 improves learning efficiency by facilitating a function based on the rough-
83 set approximations [15] that forward direction to the Bayes theorem for
84 uncertainty [16]. This process detects benign and abnormal packets or
85 abnormality type (if present).

86 The rest of paper is organized as follows. Section 2 reviews recent litera-
87 tures and Section 3 provides introduction of publicly available IDS datasets.
88 An evaluation of the realism of datasets and comparative results in Section 4.

89 Section 5 describes machine learning techniques as preliminaries, while Section
90 6 elaborates proposed method. Section 7 illustrates core features and feature
91 ranking using a suggested method and compared to rough topology approach
92 that provides the same result. Experiments are conducted and examine the
93 performance of the proposed method in Section 8. Finally, Section 9 concludes
94 the work as well as future direction.

95 **2. Related work**

96 The high dimensional dataset contains much information as well as leads
97 the many problems [17] such as computational complexity, time complexity,
98 system learning complexity, consumes system resources, alert delays, etc. These
99 problems directly affect IDSs performance. Manzoor et al. [3] proposed a system
100 based on feature reduction for intrusion detection. They were ranked of features
101 using information gain and correlation. The artificial neural network was applied
102 for classification into non-attack and attacks that took more training time. They
103 report that their method achieved the detection rate of normal connections
104 98.8% and maximum detection rate of a attack type 93.8%. Feature ranking
105 method increased the detection rate of different attack class like R2L and U2R
106 but the precision of U2R shown 42.9%. They reported a marginal improvement
107 in accuracy of only DoS class of attack.

108 Zhu et al. [18] proposed a multi-objective approach for feature selection
109 based on population evaluation targeted special domination approach and pre-
110 defined search. Their IDS applied in the cloud computing scenario when the
111 classifier found abnormal data then put an alert to the firewall to block the con-
112 nection. They were handled multi-objective feature selection and other notice-
113 able difficulties using an Improved NSGA-III algorithm. The results obtained
114 using their method shown better in detection rate, training time, and test time
115 for selected features compared to NSGA-II, NSGA-III, and All features. The
116 only approach that had better total detection rate.

117 Ambusaidi et al. [19] proposed a selection of optimal features based on

118 mutual information. They applied the Least Squared Support Vector Machine
119 (LSSVM) for intrusion detection. The SVM is suitable for binary classification
120 then it deals with the problem having more than two classes using two popular
121 techniques namely “One-vs-One” and “One-vs-All”. They tested their method
122 on different datasets and report that their method achieved the best detection
123 rate for R2L and U2R attacks with rates of 88.38% and 22.11% respectively.
124 Overall, the detection rate shown 78.86% which achieved the best rate among
125 recent methods.

126 Aburomman et al. [20] proposed an ensemble method performed by a com-
127 bination of classifiers. It was an ensemble of SVM, kNN, and PSO that used
128 meta optimizer named as Local-Unimodal-Sampling to find better parameter.
129 They were trained and tested on 12 experts and combined them into the ensem-
130 ble. They report that their expert methods obtained accuracies in the range
131 from 87.44% to 91.67%. The somewhat poorer results were obtained for normal
132 category classified the accuracy as low as 68.95% whenever the best accuracy
133 was obtained of LUS based method as high as 92.90%. Well selected training
134 data, good choice of RBF, and a wide range of variation of selected parameters
135 are some factors of high accuracies of base classifiers. Therefore, the ensem-
136 ble method can handle high volume dataset, but it is a difficult task to decide
137 configurations.

138 Eesa et al. [4] introduced feature selection based on optimization technique
139 and decision tree for classification. They applied the Cuttlefish optimization
140 technique that found an optimal set of features. Their work pushed fitness in
141 saturation stage and produced a set of optimal features that took more time to
142 compute. They were trained and tested their method on part of training and test
143 dataset that maintained well proportionate categories ratio. They were obtained
144 best results with less than or equal to 20 features. One ambiguity shown in their
145 method that better detection rate, accuracy and fitness on selected 10 features
146 while better false positive rate with 30 features.

147 Singh et al. [21] suggested an extreme leaning to reduced computational
148 complexity, false alerts and increased the detection rate. Their proposed algo-

149 rithm used alpha profiling to reduced the time complexity by discarding the
150 insignificance features and reduced the volume of the training dataset used beta
151 profiling. In a short period, extreme learning approach can deal with high vol-
152 ume dataset. They report that their method achieved 99.07% detection rate
153 and 1.74% false positive rate for normal category and 99.14% detection rate
154 and 1.49% false positive rate for DoS attack. While beta profiling reduced only
155 up to 7.66% size of training samples.

156 Elhag et al. [5] introduced a fuzzy genetic system within pairwise learning
157 to improve detection. They were adopted pairwise learning for multi-class clas-
158 sification using “One-vs-One” binarization technique and divide-and-conquer
159 applied to partition the problem that made fast execution. The method was
160 executed on non-redundant samples that divided into training and test set. It
161 carried out a standard holdout based validation methodology. They report that
162 their method achieved detection rate range from 65.38% to 99.81% and precision
163 range from 23.25 to 99.84% of test set of different categories.

164 These work were performed on high volume KDD’99 dataset that contains
165 many redundant data [22]. The former IDSs are efficient, but they are limited
166 to the information which they were trained. This dataset suffers from traffic
167 diversity, new attack behaviors, packet anonymity, feature set, and meta-data.
168 Hence, the testing IDS approaches against new dataset does not provide effective
169 performance [23]. It indicates that the proposed method is suitable IDS that
170 efficient feature selection method reduces complexities. This proposed method
171 is executed on new IDS dataset that performance has been retained and results
172 are encouraging. The next section provides comparative details of recent IDS
173 (CICIDS2017) dataset.

174 **3. IDS datasets**

175 Since 1998, more than ten IDS datasets are publicly available that unreliable
176 to use [9]. The main reason behind this situation is the exponential growth of
177 network traffic and volumes. Regarding dynamic changes in network structures

178 and attack characteristics these datasets are outdated. Moreover, these out-
179 dated datasets do not aware of recent attacks. There is a brief introduction of
180 online available IDS datasets. The KDD'99 IDS dataset was created by DARPA
181 (Defense Advanced Research Project Agency) in 1998 named as DARPA dataset
182 [22]. It contains benign that collected from real network whenever attacks gen-
183 erated from testbed. In 1999, a set of new attacks merged with the DARPA
184 dataset and named as KDD cup 99 (in short KDD'99). A lot of redundant data
185 are present in the KDD'99 dataset [22].

186 An IDS dataset was released by the Kyoto University, Japan in 2006 named
187 as Kyoto2006. It was generated through established of honeypots in normal
188 traffic. The Kyoto2009 dataset is upgraded dataset of Kyoto2006 that avoid
189 manual labeling process. It contains a set of attack that was collected from the
190 real network. Two different environments have used to generate a set of normal
191 and a set of attack activities [24]. Another IDS dataset (ISCX2012 [25]) was
192 generated through real network configuration and collected packets activities
193 in normal and abnormal form. It describes α and β profiles wherever α pro-
194 file defines multistage scenarios of attacks, and β profile defines mathematical
195 distributions of the entity which contain preconditions and postconditions. In
196 the latest era of network technologies, most of the network traffic has engaged
197 with HTTPS protocol whenever many datasets were not simulated on HTTPS.
198 The CAIDA dataset collected from multiple sources like specific events orga-
199 nized with skilled participants and published it for research. The LBNL dataset
200 recorded at a medium size network that generated with full header network traf-
201 fic. It has suffered from massive anonymity and payload. Evaluation of CAIDA
202 and LBNL datasets are difficult due to proper labeling [25].

203 The DEFCON dataset collected from the exclusive procedure by conduct-
204 ing hacking and anti-hacking competition events with skilled participants. It
205 is a very restrictive competition environment, and different from the real net-
206 work [25] where network traffics are busy with intrusions or attacks, and alarms
207 (alerts). In 2013, Australian Defense Force Academy published Linux based IDS
208 Dataset named as ADFA-LD [23]. It was generated normal and abnormal pro-

Table 1: Evaluation framework of IDS dataset

Parameters	KDD'99	DEFCON	CAIDA	LBNL	Kyoto	ISCX2012	ADFA2013	CICIDS2017
Network	yes	no	yes	yes	yes	yes	yes	yes
Traffic	no	no	yes	yes	no	no	yes	yes
Interaction	yes	yes	no	no	yes	yes	yes	yes
Capture	yes	yes	no	no	yes	yes	yes	yes
Protocol	http	yes	yes	–	yes	yes	yes	yes
	https	no	no	–	no	yes	no	yes
	SSH	yes	yes	–	yes	yes	yes	yes
	FTP	yes	no	–	no	yes	yes	yes
	Email	yes	no	–	no	yes	yes	yes
Attack Diversity	Browser	no	no	no	–	yes	yes	yes
	Bforce	yes	no	no	–	yes	yes	yes
	DoS	yes	no	yes	–	yes	yes	no
	Scan	yes	yes	yes	yes	yes	yes	no
	Bdoor	no	yes	no	–	yes	yes	yes
	DNS	no	no	yes	–	yes	no	no
	other	yes	yes	yes	–	yes	yes	yes
Heterogeneity	no	no	no	no	no	yes	–	yes
Meta data	yes	no	yes	no	yes	yes	yes	yes
Features set	yes	no	no	no	yes	no	no	yes
Label	yes	no	no	no	yes	yes	yes	yes

209 files using Linux based system. This dataset generation process was simulated
210 through host-based IDS over handmade testbed. Unavailability of important
211 information and lack of access from a real network are the main shortcomings.
212 Next-Generation IDS DataSet (NGIDS-DS) has published and shown maximum
213 quality realism [8]. It has generated through IXIA hardware and provides a
214 combination of normal and abnormal samples that includes maximum attacks,
215 cyber traffic, and ground truth.

216 Recently, CICIDS2017 has generated by Canadian Institute for Cyberse-
217 curity that short out limitations of datasets. On the given information of the
218 dataset, our quality evaluation has defined in Section 4 and the result has queued
219 it in the maximum possible quality datasets. This reliable dataset contains a
220 set of attack that reflects the real world criteria. Its generation process has con-
221 sidered characteristics of new attacks and dynamic nature of network structure.

222 Table 1 shows the evaluation framework of IDS datasets based on different net-
 223 work parameters such as network configuration, complete traffic, interaction of
 224 network, all traffic capture or recorded on the storage server, heterogeneity, set
 225 of protocols, attack diversity, meta data, features set and labeled dataset [9].

Table 2: Data distributions

Sub-datasets	Class	Data samples		Total
		Training	Testing	
Tuesday	BENIGN	388868	43206	445909
	FTP Patator	7131	807	
	SSH Patator	5319	578	
Wednesday	BENIGN	396088	43943	692703
	Dos slowloris	5208	588	
	Dos slowhttpstest	4897	602	
	Dos Hulk	207982	23091	
	Dos GoldenEye	9258	1035	
	Heartbleed	11	11	
Thursday Morning	BENIGN	151397	16789	170366
	Web Attack-Brute Force	1365	142	
	Web Attack-XSS	563	89	
	Web Attack-sql injection	21	17	
Thursday AfterNoon	BENIGN	259742	28824	288602
	Infiltration	36	36	
Friday Morning	BENIGN	170166	18901	191033
	Bot	1764	202	
FridayAfterNoon-DDoS	BENIGN	165537	18373	225745
	DDoS	37633	4202	
FridayAfterNoon-PortScan	BENIGN	114843	12694	286467
	PortScan	142977	15953	

226 Table 2 shows the distribution of samples in training and testing that the
 227 last column contains total samples. Training and testing samples have randomly
 228 divided except Wednesday (Heartbleed), Thursday-Morning (Web Attack-sql

229 injection) and Thursday-Afternoon (Infiltration) due to attacks contain very
 230 less number of samples. This is based on the Holdout validation method, it
 231 randomly disjoint the original dataset into two subsets as training and test set
 232 [5]. Initially, it is randomly divided into a training set (90%) and test set (10%)
 233 that is based on KDD'99 dataset [22]. Latterly, the reduction of redundant data
 234 on selected features compressed training set that maintains a training set (66%)
 235 and test set (34%). The CICIDS2017 dataset comprises eight sub-datasets which
 236 are named by their data generation day and time. Monday sub-dataset contains
 237 only benign samples and other sub-datasets benign with attacks.

238 4. Evaluation of realism of dataset

239 We have evaluated quality realism of CICIDS2017 dataset using FLS. The
 240 FLS is based on Sugeno fuzzy model that evaluates quality realism of IDS
 241 dataset [8]. It exists with four parts namely fuzzification, inference system,
 242 rules, and defuzzification. Dataset realism defined by crisp-set of input data,
 243 linguistic terms, generation environment, rules, inference system, membership
 244 functions, etc. The FLS functions are similar to the Sugeno fuzzy model those
 245 require input set and relationship set. As the definition of FLS, it represents
 246 sets $P = \{p_1, p_2, p_3, p_4, p_5, p_6\}$ and $Q = \{q_1, q_2\}$, and their membership function
 247 $F_1(p_u)$ and $F_2(q_v)$ are defined in Table 3, and Table 4 respectively. Input set P
 248 contains the possibility of realistic factors and input set Q contains the dataset
 249 generation environments.

Table 3: Crisp input set P

Elements	Description	$F_1(p_u)$
p_1	Complete capture of audit logs of computer operating system and network packets	1/6
p_2	Maximum number of possible attacks included	1/6
p_3	Current attack behaviors	1/6
p_4	Real world normal traffic dynamic with operation timings and industry complexity	1/6
p_5	Maintenance of cyber environment during complete capture	1/6
p_6	Ground truth information included to assist labeling process	1/6

250 Input set P is assigning membership function $F_1(p_u)$ with the help of prede-
 251 fined value where given maximum realism probability is 1, and u contains many
 252 factors. In Table 3, maximum probability is being divided by number of factors
 253 ($\frac{1}{6}$; i.e., each factor membership value is 0.16). In Table 4, membership value
 254 $F_2(q_v)$ has predefined and dataset generation process is denoted by $F_2(q_v)$ where
 255 v is index of data generation environment. Although, the realism probability
 256 has given maximum one for real network and a half for synthesis network or
 257 testbed.

Table 4: Crisp input set Q

Elements	Linguistic terms	Generation environments	$F_2(q_v)$
q_1	Good	Real network	1
q_2	Average	Testbed or synthesis network	1/2

258 Table 5 lists comparative information of publicly available IDS datasets for
 259 elements in an input set P. The NGIDS-DS dataset has followed the maximum
 260 elements of input set P and shown highest realism probability. We have assessed
 261 for CICIDS2017 dataset that provided input data elements $P = \{p_1, p_3, p_4, p_6\}$
 262 whenever rest of information has not provided in the literature [9]. In this
 263 dataset, it has used maximum possible packets through different operating sys-
 264 tems in dynamic timings with the current behavior of attacks. For labeling, each
 265 generated flow has included Source IP, Source port, Destination IP, Destination
 266 port and protocol of attacks [9].

$$\xi_i = \alpha[F_1(x_u)] + \gamma[F_2(y_v)] + \zeta \quad (1)$$

267

$$w_i = \text{AndMethod}[F_1(p_u), F_2(q_v)] \quad (2)$$

268 In the FLS, the numerical R defines using Eq. 1, 2 and 3 that output of rule
 269 ξ_i is weighted to strength w_i . The FLS defines scaling of output parameters
 270 with α, γ , and ζ where number of rule (η) is directly proportional to scaling

Table 5: Existing crisp input set of P elements

Datasets	p_1	p_2	p_3	p_4	p_5	p_6
KDD'99[8]	yes	no	no	no	NIP	yes
ISCX 2012[8]	yes	no	yes	no	NIP	yes
ADFA-LD[8]	no	no	yes	no	NIP	yes
Kyoto [8]	no	no	no	no	NIP	yes
DEFCON [8]	yes	no	no	no	NIP	yes
LBNL, CAIDA [8]	no	no	no	yes	NIP	no
NGIDS-DS[8]	yes	yes	yes	yes	yes	yes
CICIDS 2017 [9]	yes	NIP	yes	yes	NIP	yes

Note: NIP 'no information provided'.

271 parameters as $\alpha = \gamma = \zeta = \eta$.

$$numericalR = \frac{\sum_{i=1}^{\eta} w_i \xi_i}{\sum_{i=1}^{\eta} w_i} \quad (3)$$

272 The final R is computed with the help of numerical R, and expected maximum R (i.e., $6*0.16+6*1+6=12.96$) using Eq. 1 where the numerator is numerical R (i.e., 7.97) and denominator is expected maximum R (i.e., 12.96) [8] gives result (i.e., 0.615) that gives normalized membership as $0 \leq R \leq 1$. Table 273 6 describes IDS datasets probability realism that has evaluated using the FLS. 274 275 Evaluation of each dataset identified by some observed rules and generation environment that are mentioned in set P and Q. Based on some observed rules 276 277 compute a value of ξ_i and w_i using Eq. 1 and 2 respectively. An Eq. 3 computes 278 numerical R using the value of ξ_i and w_i that gives final R as a normalized value 279 280 of numerical R to the expected maximum value. 281

282 To map of final R into fuzzy linguistic terms hence five terms of fuzzy logic 283 related to overlapping ranges [8] of the quantitative realism respectively, i.e., $0 \leq$ 284 $R < 0.10 \rightarrow Verylow$, $0.08 < R \leq 0.30 \rightarrow Low$, $0.28 < R \leq 0.60 \rightarrow Medium$,

Table 6: Realism of existing IDS datasets

Datasets	η	Observed rules	ξ_i	w_i	Numerical R	Final R
KDD'99 [8]	2	$(p_1 \text{ AND } q_1)$	$\xi_1=4.32$	$w_1=0.16$	4.32	0.33
		$(p_6 \text{ AND } q_1)$	$\xi_2=4.32$	$w_2=0.16$		
ISCX 2012 [8]	3	$(p_1 \text{ AND } q_1)$	$\xi_1=6.48$	$w_1=0.16$	6.18	0.47
		$(p_3 \text{ AND } q_2)$	$\xi_2=4.98$	$w_2=0.08$		
		$(p_6 \text{ AND } q_1)$	$\xi_3=6.48$	$w_3=0.16$		
ADFA-LD [8]	2	$(p_3 \text{ AND } q_2)$	$\xi_1=3.32$	$w_1=0.08$	3.98	0.30
		$(p_6 \text{ AND } q_1)$	$\xi_2=4.32$	$w_2=0.16$		
Kyoto [8]	1	$(p_6 \text{ AND } q_1)$	$\xi_1=2.16$	$w_1=0.16$	2.16	0.16
DEFCON [8]	1	$(p_1 \text{ AND } q_1)$	$\xi_1=2.16$	$w_1=0.16$	2.16	0.16
LBNL, CAIDA [8]	1	$(p_4 \text{ AND } q_1)$	$\xi_1=2.16$	$w_1=0.16$	2.16	0.16
NGIDS-DS [8]	6	$(p_1 \text{ AND } q_1)$	$\xi_1=12.96$	$w_1=0.16$	12.36	0.95
		$(p_2 \text{ AND } q_1)$	$\xi_2=12.96$	$w_2=0.16$		
		$(p_3 \text{ AND } q_2)$	$\xi_3=9.96$	$w_3=0.08$		
		$(p_4 \text{ AND } q_2)$	$\xi_4=9.96$	$w_4=0.08$		
		$(p_5 \text{ AND } q_1)$	$\xi_5=12.96$	$w_5=0.16$		
		$(p_6 \text{ AND } q_1)$	$\xi_6=12.96$	$w_6=0.16$		
CICIDS2017	4	$(p_1 \text{ AND } q_1)$	$\xi_1=8.64$	$w_1=0.16$	7.97	0.61
		$(p_3 \text{ AND } q_2)$	$\xi_3=6.64$	$w_3=0.08$		
		$(p_4 \text{ AND } q_2)$	$\xi_4=6.64$	$w_4=0.08$		
		$(p_6 \text{ AND } q_1)$	$\xi_6=8.64$	$w_6=0.16$		

285 $0.58 < R \leq 0.97 \rightarrow High$, and $0.95 < R \leq 1 \rightarrow Veryhigh$. The NGIDS-DS
286 dataset qualitative realism probability is higher than CICIDS2017 whenever a
287 fuzzy linguistic term is same (High) for both IDS dataset. Figure 1 has shown
288 a comparative quality realism in fuzzy terms of IDS datasets. However, these
289 datasets are realized in three categories such as Low, Medium and High. In this
290 work, we have assessed the quality realism of the CICIDS2017 that maintains
291 the high category.

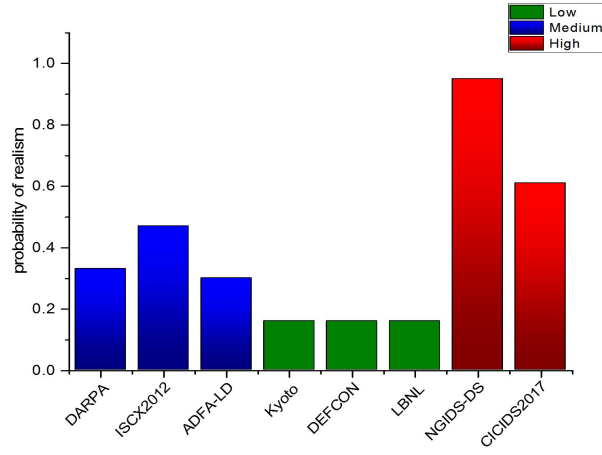


Figure 1: Quality of realism of datasets

292 **5. Preliminaries**

293 *5.1. Rough set theory*

294 The RST offers an easy process to generate a set of decision rules. This
 295 set suitable for distributed processing and obtained result straightforward. A
 296 set of significant features of the relation (R) consider as CORE that contains
 297 meaningful information of the relation opposite of this concept is reduct. Let
 298 decision system $DS = [U, A \cup \{d\}]$ where, U is a nonempty set of finite objects,
 299 A is conditional attributes, and d is decision attribute (i.e., $d \notin A$). Let infor-
 300 mation system (IS) = {U,A} then $\{a : U \rightarrow V_a, \forall a \in A\}$ and V_a is set of domain
 301 values of feature ‘a’. Indiscernible relation $IND(R)$ for any $R \subseteq A$ associated
 302 with equivalence relation [11, 12, 26]. Equivalence classes are the equivalence
 303 to relation partition of the universe (U) into a family of a disjoint subset.

304 Figure 2 shows an association between a pair of the set namely lower-
 305 approximation and upper-approximation. These sets define objects (samples)
 306 present in the dataset by target the class. A set contains samples that surely
 307 belongs to the target class as lower-approximation and probably belongs to
 308 the target class as upper-approximation [27, 28]. These sets are also suitable
 309 for multi-class classification using “One-vs-Remain” and “One-vs-One” binary

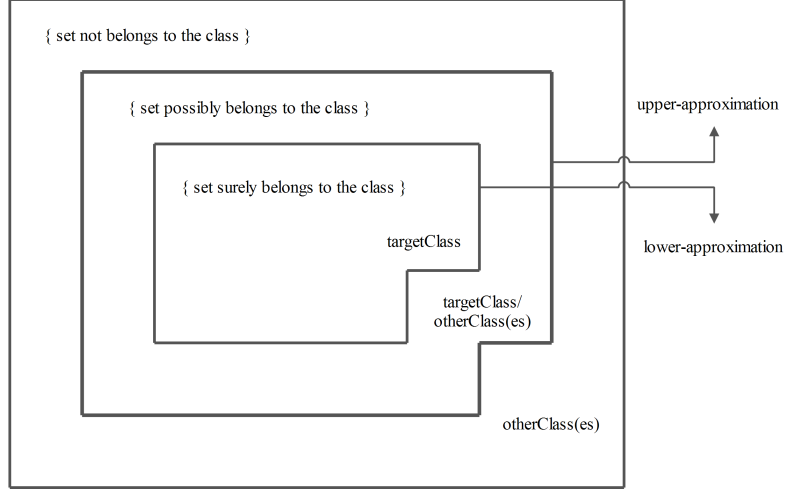


Figure 2: Set approximations

310 technique [20].

$$IND(R) = \{(x, y) \in U^2 | (\forall a \in R)(a(x) = a(y))\} \quad (4)$$

311 Where, the equivalence class included x that is denoted by $R(x)$. The set of
 312 equivalence class included by $IND(R)$ that is also denoted by $U/IND(R)$ or in
 313 short $U/I(R)$. Let X be a set of the class ($X \subseteq U$) that is used to determine
 314 lower and upper approximation by $U/I(R)$.

$$R_*(X) = \{R(x) \subseteq X | x \in U\} \quad (5)$$

$$R^*(X) = \{R(x) \cap X \neq \phi | x \in U\} \quad (6)$$

315 Boundary region of X $\{B_R(X) = R^*(X) - R_*(X)\}$ when the pair $\{R_*(X), R^*(X)\}$
 316 gives result $R_*(X) \neq R^*(X)$ indicates uncertain information [29].

317 5.2. Bayes theorem

318 The probability of event governs by the Bayes theorem that provides support
 319 to the learning algorithm. It combines the prior knowledge with observed data

320 and computes the final probability [30].

$$P(C_r/A_1, A_2, \dots, A_F) = \prod_{j=1}^F P(A_j/C_r) * P(C_r) \quad (7)$$

321 Where, $0 < j \leq F$, F is the number of features and r is $0 < r \leq \text{number_of_class}$,
322 A_j is a conditional attribute. This theorem supports the class of uncertain and
323 unseen object that is much more likely to the class [31, 32].

324 6. Proposed method

325 This section expands our key contributions mainly feature selection and the
326 Bayesian rough set method those follow data normalization. Figure 3 depicts
327 the sequence of work that starts with data processing and ends with producing
328 the result. In between these, work has partitioned into three phases that are
329 namely, dataset quality realism evaluation, feature selection or ranking using
330 estimated probabilities, and classification of samples. Subsequently, this paper
331 has discussed in three phases that are first phase in Section 4, second in Section
332 6.2, and last in Section 6.3.

333 6.1. Data normalization

334 Data normalization is an essential step for transferring symbolic to a numer-
335 ical value. Each value requires to scale that into well proportionate range. This
336 process helps to eliminate greater deviations and also biases of features. Nor-
337 malization process applies to both training and testing dataset with the same
338 minimum and same maximum value [19].

$$x_{ij} = \text{round} \left[\left(\frac{\chi_{ij} - \min(\chi_j)}{\max(\chi_j) - \min(\chi_j)} \right) * 100 \right] \quad (8)$$

339 Where, x_{ij} is normalized value of χ_{ij} that range is in integer form from 0 to
340 100. $\min(\chi_j)$ represents the minimum value of the j^{th} feature and $\max(\chi_j)$ is
341 the maximum value of the j^{th} feature.

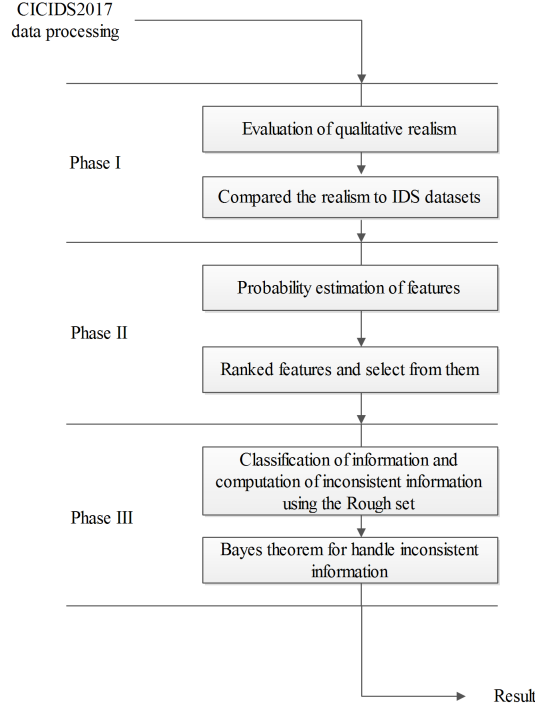


Figure 3: Diagrammatic representation of sequence of work

342 6.2. Feature selection

343 In this section, we have suggested a probabilistic approach for feature selection
 344 named Feature Probability Estimation (FPE) which computes a probability
 345 of features. This process figures out core features and ranks them based on estimated
 346 probability. It describes significant features based on the occurrence
 347 of objects to class(es). FPE computes occurrence of objects of the feature and
 348 gives an estimated probability ($prob_j$). This process has illustrated in Section
 349 7.2 on standard data of chikungunya disease with many symptoms.

$$\mu_{vj} = \frac{1}{M} \sum_{c=1}^M 1|y_{vj} \in d_c \quad (9)$$

350 Where, $d = \{\text{set of class or decision}\}$ which size is M , $y_j = \text{distinct}(x_j)$, $N_j =$
 351 $\text{count}(y_j)$, $1 \leq v \leq N_j$, and $0 < \mu_{vj} \leq 1$. y_{vj} represents distinct object of j^{th}

352 feature and N_j represents total distinct object of j^{th} feature.

$$\rho_j = \begin{cases} \frac{1}{n_j} \sum_{v=1}^{N_j} \mu_{vj}, & \text{if } \mu_{vj} < 1 \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

353 Where, $n_j = \text{count}(y_j) | \mu_{vj} < 1$, $z_j = \sum_{v=1}^{N_j} 1 | \mu_{vj} = 1$, which indicates $N_j =$
 354 $n_j + z_j$.

$$\text{prob}_j = 1 - \begin{cases} 1, & \text{if } \rho_j = 0 \\ \frac{1}{2} \left(\frac{z_j}{N_j} + \rho_j \right), & \text{otherwise} \end{cases} \quad (11)$$

355 Where, prob_j is estimated probability of j^{th} feature. Eq. 9, 10, and 11 are
 356 applied to compute occurrences of objects and facilitate prob_j .

Table 7: Selected features

Sub-dataset	Features
Tuesday	f36, f31, f5, f84, f83, f82, f81, f78, f74, f33, f79, f26, f49, f17, f73, f28, f18, f61, f19, f80, f47, f77, f34, f23, f46, f48, f60, f15, f45, f29, f59, f24, f20, f35, f13, f25, f16, f70, f11, f72
Wednesday	f6, f74, f45, f70, f11, f18, f73, f14, f36, f62, f41, f75, f71, f69, f42, f10, f9, f13, f72, f12, f16, f79, f61, f19, f78, f82, f60, f15, f21, f46, f26, f17, f77, f80, f58, f34, f76, f3, f35, f32
Thursday Morning	f36, f31, f5, f84, f83, f81, f82, f18, f26, f79, f74, f14, f33, f28, f73, f34, f78, f35, f77, f45, f80, f30, f25, f24, f29, f21, f23, f16, f75, f62, f41, f69, f9, f72, f42, f12, f60, f15, f71, f10
Thursday AfterNoon	f36, f5, f18, f26, f61, f19, f14, f31, f3, f20, f44, f17, f21, f45, f72, f12, f74, f73, f43, f49, f22, f33, f28, f13, f58, f23, f46, f47, f59, f78, f29, f76, f16, f34, f75, f83, f79, f81, f35, f30
Friday Morning	f36, f31, f82, f26, f18, f79, f33, f84, f78, f29, f77, f24, f80, f61, f19, f28, f23, f14, f83, f81, f34, f45, f35, f32, f30, f27, f25, f8, f74, f20, f21, f17, f62, f41, f42, f73, f71, f69, f10, f9
FridayAfterNoon-DDoS	f36, f74, f16, f13, f60, f15, f42, f70, f11, f72, f12, f33, f71, f10, f75, f14, f5, f78, f62, f41, f69, f9, f73, f45, f35, f32, f34, f18, f31, f76, f21, f79, f61, f19, f28, f46, f47, f77, f3, f17
FridayAfterNoon-PortScan	f41, f74, f71, f11, f61, f40, f81, f73, f70, f9, f69, f10, f77, f17, f72, f33, f78, f20, f44, f15, f68, f8, f25, f13, f12, f59, f14, f76, f79, f60, f18, f19, f46, f28, f58, f45, f16, f47, f48, f57

357 Table 7 contains 40 selected features and shows the rank of features that
 358 are evaluated using feature selection (FPE) method. A set of selected features
 359 are different from other sets of features that obtained from given information
 360 in decision systems. The feature comes in first is more significant to upcoming
 361 feature and so on. The Bayesian Rough Set Training Data Set (BRSTDS) is a
 362 set of non-redundant samples on selected features. Table 8 contains information

Table 8: Bayesian rough set training dataset on selected features

Sub-datasets	Training dataset		
	BRSTDS	Avg freq	Total samples
Tuesday	73845	5.43	401318
Wednesday	144756	4.31	623442
Thursday Morning	22320	6.87	153329
Thursday AfterNoon	99547	2.61	259777
Friday Morning	30213	5.69	171930
FridayAfterNoon-DDoS	53944	3.77	203170
FridayAfterNoon-PortScan	22941	11.24	257820

363 in three columns that are BRSTDS, average frequency, and total samples. An
 364 average frequency is average of occurrences of BRSTDS from total samples. A
 365 multiplication result of the BRSTDS and average frequency as a total sample
 366 is present in training set. The BRSTDS training samples are 447566 instead of
 367 2070786 that imply a great reduction. Our proposed method has trained on the
 368 BRSTDS that reduced the computational and training complexity.

369 6.3. Bayesian-rough set

370 The Bayesian-Rough Set (BRS) has designed with approximations and prior
 371 knowledge of the decision system. Lower approximation and upper approxima-
 372 tion which provides boundary region where prior knowledge moves for strong
 373 decision [26]. This method has categorized samples into three categories namely
 374 normal (eB), intermediary ($intB$), abnormal (nB).

- 375 • eB : It surely belongs to the target class. i.e., $eB = R_*(X)$
- 376 • nB : It does not belong to the target class. i.e., $nB = U - R^*(X)$
- 377 • $intB$: It may belong to the target class. i.e., $intB = B_R(X)$

378 The BRS categories samples by target one class at a time and repeat until
 379 remaining. It is based on classification technique “One-vs-Remain” for multi-

380 class and “One-vs-One” when binary class problem [20]. This hybrid approach
 381 of mathematical approximations and probabilistic theorem increase detection
 382 capacity.

Algorithm 1 Bayesian-rough set

Input: data samples

Output: confusion_matrix

- 1: data processing
 - 2: normalization of data using Eq. 8.
 - 3: compute core features and ranked them using Eq. 9, 10, and 11.
 - 4: select best 40 features.
 - 5: reduction of redundant samples from the training set.
 - 6: Rough set group samples into $eB, nB, intB$ categories.
 - 7: Bayes theorem for $intB$ and unseen samples (using Eq. 7).
 - 8: each test sample contributes in confusion matrix $C_{M,M}$.
 - 9: **if** $predicated_class = testdata_class$ **then**
 - 10: increment in confusion matrix at $C_{r,r}$
 - 11: **else**
 - 12: increment in confusion matrix at $C_{r,t}$
 - 13: **end if**
 - 14: return confusion_matrix
-

383 Algorithm 1 starts with reading dataset and analyzes the number of samples,
 384 features, a format of data, data deviation, etc. Subsequently, data normaliza-
 385 tion that scales data and makes them in a proper format (remove imbalance
 386 data). Then, evaluate CORE features using FPE and ranked them based on
 387 the estimated probability. Reduction of features and redundant samples from
 388 the training set that reduce the computational complexity. Finally, classify be-
 389 nign and attacks and then in the repetitive process attack types. This system
 390 performance has assessed using confusion matrix on different parameters.

391 **7. Method verification**

392 In this section, we have computed core features and ranked them using FPE
 393 and compared to rough topology [14] which show the same result. The following
 394 table presents the standard symptoms of chikungunya disease such as joint-pain,
 395 headache, nausea, and temperature. In this table, columns represent attributes
 396 and rows represent patients or sufferers (S denotes sufferer).

Table 9: Symptoms of chikungunya disease

Patients	Joint-pain (Jp)	Headache (Hd)	Nausea (Na)	Temperature (Tp)	Chikungunya
S_1	yes	yes	yes	high	yes
S_2	yes	no	no	high	no
S_3	yes	no	no	high	yes
S_4	no	no	no	very high	no
S_5	no	yes	yes	high	no
S_6	yes	yes	no	very high	yes
S_7	yes	yes	no	normal	no
S_8	yes	yes	no	very high	yes

397 This table having symptoms as conditional attributes and disease having
 398 chikungunya as decision attribute. A conditional attribute joint-pain gener-
 399 ate two equivalence classes of patients $[\{S_1, S_2, S_3, S_6, S_7, S_8\}$ and $\{S_4, S_5\}]$,
 400 joint-pain and headache generate equivalence classes $[\{S_1, S_6, S_7, S_8\}$, $\{S_2, S_3\}$,
 401 $\{S_4\}$, and $\{S_5\}]$, joint-pain, headache and nausea generate equivalence classes
 402 $[\{S_1\}$, $\{S_2, S_3\}$, $\{S_4\}$, $\{S_5\}$, and $\{S_6, S_7, S_8\}]$, joint-pain, headache, nausea and
 403 temperature generate equivalence classes $[\{S_1\}$, $\{S_2, S_3\}$, $\{S_4\}$, $\{S_5\}$, $\{S_6, S_8\}$,
 404 and $\{S_7\}]$. On the basis of symptoms given in Table 9, we compute approx-
 405 imations of patients having chikungunya disease such as lower-approximation
 406 $\{S_1, S_6, S_8\}$, upper-approximation $\{S_1, S_2, S_3, S_6, S_8\}$, and the boundary region
 407 $\{S_2, S_3\}$. For given information, patient S_2 and S_3 symptoms are not able to
 408 exactly classified. Then, it is possible that evaluated decision can be obtain
 409 using subset of conditional attributes (symptoms).

410 *7.1. Rough topology*

411 **Case 1:** Let the set of patient having chikungunya is $X = \{S_1, S_3, S_6, S_8\}$ and
 412 U represents the equivalence relation R w.r.t conditional attributes. The family
 413 of equivalence classes are $U/I(R) = [\{S_1\}, \{S_2, S_3\}, \{S_4\}, \{S_5\}, \{S_6, S_8\}, \{S_8\}]$.
 414 Approximations are lower-approximation as $R_*(X) = \{S_1, S_6, S_8\}$ and upper-
 415 approximation as $R^*(X) = \{S_1, S_2, S_3, S_6, S_8\}$. Rough topology on U w.r.t X
 416 is $\tau_R = [U, \phi, \{S_1, S_6, S_8\}, \{S_1, S_2, S_3, S_6, S_8\}, \{S_2, S_3\}]$ that defines $\beta_R = [U,$
 417 $\{S_1, S_6, S_8\}, \{S_2, S_3\}]$. We remove conditional attribute from the information
 418 system one by one and check significance (role) of attributes, if any changes
 419 occur in τ and β then it indicates significance of the attribute.

- 420 (i) Remove joint-pain and compute equivalence families $U/I(R-Jp) = [\{S_1, S_5\},$
 421 $\{S_2, S_3\}, \{S_4\}, \{S_6, S_8\}, \{S_7\}]$; $(R-Jp)_*(X) = \{S_6, S_8\}$; $(R-Jp)^*(X) =$
 422 $\{S_1, S_2, S_3, S_5, S_6, S_8\}$; $\tau_{(R-Jp)} = [U, \phi, \{S_6, S_8\}, \{S_1, S_2, S_3, S_5, S_6, S_8\},$
 423 $\{S_1, S_2, S_3, S_5\}]$; $\beta_{(R-Jp)} = [U, \{S_6, S_8\}, \{S_1, S_2, S_3, S_5\}]$ i.e., $\tau_{(R-Jp)} \neq$
 424 τ_R and $\beta_{(R-Jp)} \neq \beta_R$ that indicate as significance attribute.
- 425 (ii) Remove headache and compute $U/I(R-Hd) = [\{S_1\}, \{S_2, S_3\}, \{S_4\},$
 426 $\{S_5\}, \{S_6, S_8\}, \{S_7\}]$; $(R-Hd)_*(X) = \{S_1, S_6, S_8\}$; $(R-Hd)^*(X) =$
 427 $\{S_1, S_2, S_3, S_6, S_8\}$; $\tau_{(R-Hd)} = [U, \phi, \{S_1, S_6, S_8\}, \{S_1, S_2, S_3, S_6, S_8\}, \{S_2, S_3\}]$;
 428 $\beta_{(R-Hd)} = [U, \{S_1, S_6, S_8\}, \{S_2, S_3\}]$ i.e., $\tau_{(R-Hd)} = \tau_R$ and $\beta_{(R-Hd)} =$
 429 β_R that indicate as insignificance attribute.
- 430 (iii) Remove nausea and compute $U/I(R-Na) = [\{S_1\}, \{S_2, S_3\}, \{S_4\}, \{S_5\},$
 431 $\{S_6, S_8\}, \{S_7\}]$; $(R-Na)_*(X) = \{S_1, S_6, S_8\}$; $(R-Na)^*(X) = \{S_1, S_2, S_3, S_6, S_8\}$;
 432 $\tau_{(R-Na)} = [U, \phi, \{S_1, S_6, S_8\}, \{S_1, S_2, S_3, S_6, S_8\}, \{S_2, S_3\}]$; $\beta_{(R-Na)} = [U,$
 433 $\{S_1, S_6, S_8\}, \{S_2, S_3\}]$ i.e., $\tau_{(R-Na)} = \tau_R$ and $\beta_{(R-Na)} = \beta_R$ that indicate
 434 insignificance attribute.
- 435 (iv) Remove temperature and compute $U/I(R-Tp) = [\{S_1\}, \{S_2, S_3\}, \{S_4\},$
 436 $\{S_5\}, \{S_6, S_7, S_8\}]$; $(R-Tp)_*(X) = \{S_1\}$; $(R-Tp)^*(X) = \{S_1, S_2, S_3, S_6, S_7, S_8\}$;
 437 $\tau_{(R-Tp)} = [U, \phi, \{S_1\}, \{S_1, S_2, S_3, S_6, S_7, S_8\}, \{S_2, S_3, S_6, S_7, S_8\}]$; $\beta_{(R-Tp)}$
 438 $= [U, \{S_1\}, \{S_2, S_3, S_6, S_7, S_8\}]$ i.e., $\tau_{(R-Tp)} \neq \tau_R$ and $\beta_{(R-Tp)} \neq \beta_R$ that
 439 indicate as significance attribute.

440 **Case 2:** Let the set of patient not having chikungunya is $X = \{S_2, S_4, S_5, S_7\}$
 441 then $U/I(R) = [\{S_1\}, \{S_2, S_3\}, \{S_4\}, \{S_5\}, \{S_6, S_8\}, \{S_7\}]$. Approximations are
 442 as lower-approximation $R_*(X) = \{S_4, S_5, S_7\}$ and upper-approximation $R^*(X)$
 443 $= \{S_2, S_3, S_4, S_5, S_7\}$. Rough topology $\tau_R = [U, \phi, \{S_4, S_5, S_7\}, \{S_2, S_3, S_4, S_5, S_7\},$
 444 $\{S_2, S_3\}]$ and $\beta_R = [U, \{S_4, S_5, S_7\}, \{S_2, S_3\}]$.

445 (i) Remove joint-pain and compute $U/I(R - Jp) = [\{S_1, S_5\}, \{S_2, S_3\}, \{S_4\},$
 446 $\{S_6, S_8\}, \{S_7\}]$; $(R - Jp)_*(X) = \{S_4, S_7\}$; $(R - Jp)^*(X) = \{S_1, S_2, S_3,$
 447 $S_4, S_5, S_7\}$; $\tau_{(R-Jp)} = [U, \phi, \{S_4, S_7\}, \{S_1, S_2, S_3, S_4, S_5, S_7\}, \{S_1, S_2, S_3, S_5\}]$;
 448 $\beta_{(R-Jp)} = [U, \{S_4, S_7\}, \{S_1, S_2, S_3, S_5\}]$ i.e., $\tau_R \neq \tau_{(R-Jp)}$ and $\beta_R \neq$
 449 $\beta_{(R-Jp)}$ that indicate as significance attribute.

450 (ii) Remove headache and compute $U/I(R - Hd) = [\{S_1\}, \{S_2, S_3\}, \{S_4\},$
 451 $\{S_5\}, \{S_6, S_8\}, \{S_7\}]$; $(R - Hd)_*(X) = \{S_4, S_5, S_7\}$; $(R - Hd)^*(X) =$
 452 $\{S_2, S_3, S_4, S_5, S_7\}$; $\tau_{(R-Hd)} = [U, \phi, \{S_4, S_5, S_7\}, \{S_2, S_3, S_4, S_5, S_7\},$
 453 $\{S_2, S_3\}]$; $\beta_{(R-Hd)} = [U, \{S_4, S_5, S_7\}, \{S_2, S_3\}]$ i.e., $\tau_R = \tau_{(R-Hd)}$ and β_R
 454 $= \beta_{(R-Hd)}$ that indicate as insignificance attribute.

455 (iii) Remove nausea and compute $U/I(R - Na) = [\{S_1\}, \{S_2, S_3\}, \{S_4\}, \{S_5\},$
 456 $\{S_6, S_8\}, \{S_7\}]$; $(R - Na)_*(X) = \{S_4, S_5, S_7\}$; $(R - Na)^*(X) = \{S_2, S_3, S_4,$
 457 $S_5, S_7\}$; $\tau_{(R-Na)} = [U, \phi, \{S_4, S_5, S_7\}, \{S_2, S_3, S_4, S_5, S_7\}, \{S_2, S_3\}]$;
 458 $\beta_{(R-Na)} = [U, \{S_4, S_5, S_7\}, \{S_2, S_3\}]$ i.e., $\tau_R = \tau_{(R-Na)}$ and $\beta_R = \beta_{(R-Na)}$
 459 that indicate insignificance attribute.

460 (iv) Remove temperature and compute $U/I(R - Tp) = [\{S_1\}, \{S_2, S_3\}, \{S_4\},$
 461 $\{S_5\}, \{S_6, S_7, S_8\}]$; $(R - Tp)_*(X) = \{S_4, S_5\}$; $(R - Tp)^*(X) = \{S_2, S_3, S_4,$
 462 $S_5, S_6, S_7, S_8\}$; $\tau_{(R-Tp)} = [U, \phi, \{S_4, S_5\}, \{S_2, S_3, S_4, S_5, S_6, S_7, S_8\},$
 463 $\{S_2, S_3, S_6, S_7, S_8\}]$; $\beta_{(R-Tp)} = [U, \{S_4, S_5\}, \{S_2, S_3, S_6, S_7, S_8\}]$ i.e.,
 464 $\tau_R \neq \tau_{(R-Tp)}$ and $\beta_R \neq \beta_{(R-Tp)}$ that indicate as significance attribute.

465 Above assessments in both cases provide the same outcome as **CORE(R)={Jp, Tp}**.

466 7.2. Feature probability estimation

467 It computes probability of available features in decision system using Eq. 9, 10
 468 and 11. Table 9 contains binary class $d = \{\text{yes}, \text{no}\}$ where *yes* denotes patient
 469 having chikungunya and *no* denotes patient not having chikungunya.

- 470 (i) Probability of joint-pain (Jp), $d=\{\text{yes,no}\}$; $M=2$; $y_{Jp}=\{\text{yes,no}\}$; $N_{Jp} = 2$;
471 $\mu_{yes,Jp} = 1$; $\mu_{no,Jp} = 1/2$; $n_{Jp} = 1$; $\rho_{Jp} = 1/2$; $z_{Jp} = 1$; $prob_{Jp} = 0.50$.
- 472 (ii) Probability of headache (Hd), $d=\{\text{yes,no}\}$; $M=2$; $y_{Hd}=\{\text{yes,no}\}$; $N_{Hd} = 2$;
473 $\mu_{yes,Hd} = 1$; $\mu_{no,Hd} = 1$; $n_{Hd} = 0$; $\rho_{Hd} = 0$; $z_{Hd} = 2$; $prob_{Hd} = 0$.
- 474 (iii) Probability of nausea (Na), $d=\{\text{yes,no}\}$; $M=2$; $y_{Na}=\{\text{yes,no}\}$; $N_{Na} = 2$;
475 $\mu_{yes,Na} = 1$; $\mu_{no,Na} = 1$; $n_{Na} = 0$; $\rho_{Na} = 0$; $z_{Na} = 2$; $prob_{Na} = 0$.
- 476 (iv) Probability of temperature (Tp), $d=\{\text{yes,no}\}$; $M=2$; $y_{Tp}=\{\text{high, very high,}$
477 $\text{normal}\}$; $N_{Tp} = 3$; $\mu_{high,Tp} = 1$; $\mu_{veryhigh,Tp} = 1$; $\mu_{normal,Tp} = 1/2$;
478 $n_{Tp} = 1$; $\rho_{Tp} = 1/2$; $z_{Tp} = 2$; $prob_{Tp} = 0.416$.

479 Evaluated probabilities point out attributes a headache and nausea are insignif-
480 icant; therefore **CORE(R)={Jp,Tp}** and higher probability assumes as more
481 informative.

482 8. Experiments

483 8.1. Performance measures

484 The outcome of work has collected in the form of confusion matrix that every
485 cell contributes to measuring statistical parameters. A quantitative representa-
486 tion of the matrix in rows and columns carried out the information. Row gives
487 predicted samples for the class and column gives test samples for the class. The
488 diagonal value of matrix shows the correct classification of the class [31].

$$TPR = \frac{TP}{TP + FN} \quad (12)$$

$$FPR = \frac{FP}{FP + TN} \quad (13)$$

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (15)$$

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (16)$$

489 Confusion matrix gives quantitative symbols as True Positive (TP), True
 490 Negative (TN), False Positive (FP), and False Negative (FN). TP is a correct
 491 prediction of the classifier as “normal” whenever actual test sample is “normal”.
 492 TN is also a correct prediction of the classifier as “attack” whenever actual test
 493 sample is “attack”. FN is an incorrect prediction of the classifier as “attack”
 494 whenever the actual test sample is “normal”. FP is also an incorrect prediction
 495 of the classifier as “normal” whenever actual test sample is “attack” [3]. These
 496 collected quantities have been used for the computation of statical parameters.
 497 TP Rate (TPR) also known as sensitivity or Recall or Detection Rate (DR) that
 498 is a proportion of actual positive cases and correct cases. FP Rate (FPR) or
 499 False Alarm Rate (FAR) is a proportion of false positive and identified attacks.
 500 Precision is a proportion of correctly identified and identified cases. Accuracy
 501 is a proportion of correct prediction by a classifier and the sum of both (correct
 502 and incorrect) prediction by the classifier. F-measure is a harmonic mean of
 503 precision and recall. It measures system performance with weights of precision
 504 and recall [19]. The overall performance of the system can compute using the
 505 sum of related measures [33] and related equations.

506 8.2. Complexity analysis

507 The effects of proposed feature selection are analyzed in detail as time and
 508 space complexity. This method reduces time complexity as well as the space
 509 complexity. Without feature selection, the number of training samples and the
 510 size of training set are high while every test sample has to be tested every
 511 training samples. The reduction of features reduces the memory requirement
 512 during computation [21]. So the number of connections of computations can be
 513 defined as the following equations.

$$TimeComplexity = TR_n * TT_n \quad (17)$$

$$SpaceComplexity = TR_n * nFeatures \quad (18)$$

514 These equations discuss literature [21], where TR_n represents a total number
 515 of training samples, TT_n is a total number of test samples, and $nFeatures$ is
 516 a total number of features. As the training set multiplied by the number of
 517 features has an effect on space complexity.

Table 10: Complexities comparison

Sub-dataset name	Time complexity		Space complexity	
	EMTDS	BRSTDS	EMTDS	BRSTDS
Tuesday	37281476064	3292822395	26754560	2953800
Wednesday	89969261400	10027248120	41562240	5790240
Thursday Morning	5442065025	380265840	10222000	892800
Thursday Afternoon	15617011800	2872926420	17316160	3981880
Friday Morning	6842527450	577158939	11462000	1208520
Friday Afternoon DDoS	9555122724	1217785800	13544720	2157760
Friday Afternoon PortScan	15386769216	657190827	17188080	917640

518 We have executed our proposed method on the compressed dataset (BRSTDS)
 519 and compared complexities to existing method training dataset named as EMTDS.
 520 Existing methods have validated on fourfold manner with 80 traffic features
 521 [9]. Table 10 contains time complexity and space complexity of both existing
 522 methods and the proposed method where time complexity represents the total
 523 number of comparison and space complexity represents memory required during
 524 computation. This table shows that our proposed method reduces both time
 525 and space complexity.

526 8.3. Performance analysis

527 A proposed BRS classifier is a mathematical approach which provides a high
 528 detection rate and low false alarm rate. Existing conventional classification
 529 systems are not much effective in classification than the proposed classification

530 system. The time complexity of the rough set based classifier for learning is
 531 less than other classification techniques such as Random Forest (RF) and ID3.
 532 High volume CICIDS2017 dataset contains 85 attributes (features) and millions
 533 of samples. Reduction of insignificant features and redundant samples reduce
 534 system complexities.

Table 11: Confusion matrix for test dataset Tuesday

Class	BENIGN	FTP-Patator	SSH-Patator
BENIGN	42076	0	1
FTP-Patator	1130	807	311
SSH-Patator	0	0	266

Table 12: Statistical parameters for test dataset Tuesday

Parameters	BENIGN	FTP-Patator	SSH-Patator
TP	42076	807	266
TN	1384	42343	44013
FP	1	1441	0
FN	1130	0	312
TPR(%)	97.38	100	46.02
FPR(%)	0.07	3.29	0
Precision(%)	99.99	35.9	100
Accuracy(%)	97.46	99.77	99.3

Table 13: Confusion matrix for test dataset Wednesday

Class	BENIGN	DoS slowloris	DoS slowhttptest	DoS Hulk	DoS GoldenEye	Heartbleed
BENIGN	43380	56	3	242	6	0
DoS slowloris	6	509	0	0	0	0
DoS slowhttptest	11	0	570	0	1	0
DoS Hulk	544	23	29	22841	13	0
DoS GoldenEye	2	0	0	8	1015	0
Heartbleed	0	0	0	0	0	11

Table 14: Statistical parameters for test dataset Wednesday

Parameters	BENIGN	DoS slowloris	DoS slowhttptest	DoS Hulk	DoS GoldenEye	Heartbleed
TP	43380	509	570	2281	1015	11
TN	25020	68676	68656	45570	68225	69259
FP	307	6	12	609	10	0
FN	563	79	32	250	20	0
TPR(%)	98.72	86.56	94.68	98.92	98.07	100
FPR(%)	1.21	0.009	0.02	1.32	0.05	0
Precision(%)	99.3	98.83	97.94	97.4	99.02	100
Accuracy(%)	98.74	99.88	99.94	98.76	99.96	100

Table 15: Confusion matrix for test dataset Thursday-Morning

Class	BENIGN	Web Attack -Brute Force	Web Attack-XSS	Web Attack-sql Injection
BENIGN	16749	12	7	8
Web Attack -Brute Force	23	80	41	0
Web Attack-XSS	16	50	41	0
Web Attack-sql Injection	1	0	0	9

Table 16: Statistical parameters for test dataset Thursday-Morning

Parameters	BENIGN	Web Attack -Brute Force	Web Attack-XSS	Web Attack-sql Injection
TP	16749	80	41	9
TN	221	16831	16882	17019
FP	27	64	66	1
FN	40	62	48	8
TPR(%)	99.76	56.34	46.07	52.94
FPR(%)	10.89	0.38	0.39	0.006
Precision(%)	99.84	55.56	38.32	90
Accuracy(%)	99.6	99.26	99.33	99.95

535 Statistical parameters are computed from the confusion matrix by using
536 diagonal as TP, a column of the matrix represents actual samples of the class,
537 and a row of the matrix predicted samples of the class. As an example from
538 Table 13 for “benign”, TP is 43380, TN as sum of rows and columns values
539 except benign (i.e., 25020 is sum of 509, 0, 0, 0, 0; 0, 570, 0, 1, 0; 23, 29, 22841,
540 13, 0; 0, 0, 8, 1015, 0; 0, 0, 0, 0, 11), FN is sum of benign column except
541 diagonal (i.e., 563 is sum of 6, 11, 544, 2, 0), and FP as sum of benign row

Table 17: Confusion matrix and statistical parameters for test dataset Thursday-Afternoon

			Parameters	BENIGN	Infiltration
			TP	28824	34
			TN	34	28824
Class	BENIGN	Infiltration	FP	2	0
BENIGN	28824	2	FN	0	2
Infiltration	0	34	TPR(%)	100	94.44
			FPR(%)	5.56	0
			Precision(%)	99.99	100
			Accuracy(%)	99.99	99.99

Table 18: Confusion matrix and statistical parameters for test dataset Friday-Morning

			Parameters	BENIGN	Bot
			TP	16453	202
			TN	202	16453
Class	BENIGN	Bot	FP	0	2448
BENIGN	16453	0	FN	2448	0
Bot	2448	202	TPR(%)	87.05	100
			FPR(%)	0	12.95
			Precision(%)	100	7.62
			Accuracy(%)	87.19	87.19

542 except diagonal (i.e., 307 is sum of 56, 3, 242, 6, 0). TP and TN have correctly
 543 classified samples whose values indicate the system performance. The high value
 544 of TP and TN are favorable for the system. An opposite of this, FP and FN
 545 have incorrectly classified samples whose values indicate negative performance.
 546 A higher value of correctly classified samples and lower value of incorrectly
 547 classified samples increase system performance. FP is falsely predicted attacks
 548 as normal that allow malicious packets to enter into the system. FN is also
 549 falsely predicted normal as attacks that generate alert and increases system

Table 19: Confusion matrix and statistical parameters for test dataset Friday-Afternoon-DDoS

			Parameters	BENIGN	DDoS
			TP	115591	4198
			TN	4198	15591
Class	BENIGN	DDoS	FP	4	2782
BENIGN	15591	4	FN	2782	4
DDoS	2782	4198	TPR(%)	84.86	99.9
			FPR(%)	0.095	15.14
			Precision(%)	99.74	60.14
			Accuracy(%)	87.66	87.66

Table 20: Confusion matrix and statistical parameters for test dataset Friday-Afternoon-PortScan

			Parameters	BENIGN	PortScan
			TP	12143	15953
			TN	15953	12143
Class	BENIGN	PortScan	FP	0	551
BENIGN	12143	0	FN	551	0
PortScan	551	15953	TPR(%)	95.66	100
			FPR(%)	0	4.34
			Precision(%)	100	99.66
			Accuracy(%)	98.08	98.08

Table 21: Overall performance of BRS on selected features

TP	TN	FP	FN	TPR	FPR	Precision	Accuracy	F-Measure
221752	577465	8331	8331	0.96379	0.01422	0.96379	0.97958	0.96379

550 overhead. Table 11 shows confusion matrix of test dataset Tuesday and Table
 551 12 computes performance of the system on various statistical parameters. It has
 552 shown SSH-Patator attack predicted as FTP-Patator attack. For Wednesday

Table 22: Performance of classifiers

Classifier	No. of features	Precision	TPR	F-Measure
KNN	80	0.96	0.96	0.96
RF	80	0.98	0.97	0.97
ID3	80	0.98	0.98	0.98
Adaboost	80	0.77	0.84	0.80
MLP	80	0.77	0.83	0.79
QDA	80	0.97	0.88	0.92
BRS	40	0.96	0.96	0.96

553 dataset, confusion matrix and performance of system are in Table 13 and Table
554 14 respectively. This dataset is higher volume than listed datasets whenever
555 it is efficiently classified. Table 15 and 16 show performance of the system
556 respectively as confusion matrix and statistical measures. It also predicted
557 Web-Attack-Brute-Force to Web-Attack-XSS attack. Table 17 gives confusion
558 matrix and performance of Thursday-AfterNoon dataset that shows minimum
559 incorrectly predicted samples and maximum correctly predicted samples. These
560 results on various statistical parameters indicate better system performance.
561 Only Table 18, 19 and 20 are shown a negative predication of benign as attacks.

562 This evaluation is shown characteristics of the dataset that is divided into
563 multi-attack and single attack dataset. In multi-attack, most of the attacks
564 predicted as an attack and decrease only quantitative measure of system per-
565 formance. These do not increase false alarm and system overhead. In a single
566 attack, some benign samples are wrongly predicted attacks that decrease the
567 system performance. Table 21 provides the overall performance of the system
568 that has computed using the sum of related measures. Table 22 summaries
569 comparative results of the BRS to other related algorithms. It has trained
570 on 40 features and non-redundant samples (BRSTDS) rather than 80 features.
571 The BRSTDS is much smaller (only 22%) than original training dataset. This
572 approach provides better performance to many algorithms regarding precision,

573 detection rate (or TPR), and f-measure. Moreover, it gives a low false alarm
574 rate (or FPR) and high accuracy.

575 **9. Conclusion**

576 This research work proposed a new intrusion detection system that works
577 on the best subset of features. We have first shown a detailed analysis and
578 qualitative realism of a recently generated IDS dataset. The method uses to
579 extracts significant features using the probabilistic method and ranked them.
580 Such selection process considers the core features of the dataset and selects
581 best features where selected features built huge redundant samples and remove
582 them from the training set that reduces training complexity. Finally, the BRS
583 categorized samples into three categories namely normal, intermediary, and ab-
584 normal (abnormality type) using the rough set. The Bayes theorem computed
585 strong decision for intermediary or unseen samples using occurrence frequencies
586 of samples. This system is trained and tested on seven different subsets of CI-
587 CIDS2017 dataset. It can be seen from experimental results that the method
588 outperformed other methods for normal and attacks. Overall, the method has
589 shown that reduce training complexity and increase system accuracy. Our pro-
590 posed method was tested and found encouraging results. The implication of this
591 system is a demonstration of the fact that feature selection is an important phe-
592 nomenon to reduce dimensionality and system complexities. This system shows
593 better performance while it reduces 50% features thereby influencing design the
594 systems with less complexities. The prime importance of the proposed method
595 can be used to provide network, organizational and social area security where
596 intruders are more active. This study can inspire researchers from the area of
597 network security, data science, and machine learning to utilize their work and
598 propose more challenging current problems.

599 The present method seems convincing but it has some limitations like pre-
600 processing work have done manually, decide optimal subset of features, range
601 of estimated probability of significant feature other than null probability. Al-

602 though realism evaluation of CICIDS2017 has shown the first time. There is still
603 no method of feature selection using estimated probability as feature probability
604 estimation method. The Bayesian rough set for uncertainty and classification
605 has applied first time for IDS. The present method can extend to improve listed
606 limitations and can develop probability based feature ranking method for an un-
607 supervised intrusion detection. An unsupervised dataset also leads to the same
608 problems as supervised dataset. It is a difficult task to design a feature selection
609 system for increasing the detection rate and decrease the training complexity for
610 the unsupervised dataset as labels are usually expensive to acquire. However,
611 unsupervised datasets are free from labeling cost and unsupervised methods are
612 applied more broadly than supervised in intelligence systems.

613 **Appendix A. Tutorial**

614 **State 1 :** Realism evaluation of dataset is explained in Section 4.

615 **State 2 :** Feature ranking or selection process is discussed in Section 6.2 and
616 illustrated in Section 7.2.

617 **State 3 :** The BRS is discussed in Section 6.3 and Algorithm 1 provides a
618 stepwise execution of proposed method.

619 Table A.23 contains a small example of network traffic connection which
620 having attacks and benign information. It has only four features namely dura-
621 tion, protocol, PSH, and URG which contains numerical value while it having
622 two types of attack such as Bot and DDoS. This table contains finite number
623 of objects which are $X = \{x_1, x_2, x_3, \dots, x_{12}\}$. The objective of this small ex-
624 ample to understand the classification method of BRS. We categorize samples
625 into three categories as $eB = R_*(X)$, $nB = U - R^*(X)$, $intB = B_R(X)$ using
626 “One-vs-Remain” and “One-vs-One” classification techniques.

627 (i) targetClass is $\{Benign\}$ and otherClass is $\{Bot, DDoS\}$. The network
628 traffic having *Benign* connections $X = \{x_1, x_2, x_{11}, x_{12}\}$ and U represent
629 the equivalence relation. The family of equivalence classes are $U/I(R) = [\{x_1\},$
630 $\{x_2\}, \{x_3, x_7\}, \{x_4, x_{12}\}, \{x_5\}, \{x_6\}, \{x_8\}, \{x_9\}, \{x_{10}\}, \{x_{11}\}]$. We com-

Table A.23: An example of normalized information of network traffic connections

U	duration	protocol	PSH	URG	Label
x_1	2	2	0	0	Benign
x_2	3	2	0	0	Benign
x_3	3	3	1	0	Bot
x_4	1	2	0	1	Bot
x_5	4	4	1	0	Bot
x_6	4	3	1	0	Bot
x_7	3	3	1	0	DDoS
x_8	4	3	1	0	DDoS
x_9	3	4	1	0	DDoS
x_{10}	3	2	1	0	DDoS
x_{11}	1	1	0	1	Benign
x_{12}	1	2	0	1	Benign

631 compute $R_*(X) = \{x_1, x_2, x_{11}\}$, $R^*(X) = \{x_1, x_2, x_4, x_{11}, x_{12}\}$, $B_R(X) = \{x_4,$
632 $x_{12}\}$. Then, $eB = \{x_1, x_2, x_{11}\}$, $nB = \{x_3, x_5, x_6, x_7, x_8, x_9, x_{10}\}$, $intB = \{x_4,$
633 $x_{12}\}$.

634 (ii) targetClass is $\{Bot\}$ and otherClass as $\{DDoS\}$. The network traffic hav-
635 ing *Bot* attack connections $X = \{x_3, x_4, x_5, x_6\}$ and family of equivalence
636 classes are $U/I(R - \text{“Benign”}) = [\{x_3, x_7\}, \{x_5\}, \{x_6\}, \{x_8\}, \{x_9\}, \{x_{10}\}]$.
637 We compute $R_*(X) = \{x_5, x_6\}$, $R^*(X) = \{x_3, x_5, x_6, x_7\}$, $B_R(X) = \{x_3,$
638 $x_7\}$. Then, $eB = \{x_5, x_6\}$, $nB = \{x_8, x_9, x_{10}\}$, $intB = \{x_3, x_7\}$.

639 (iii) The Bayes theorem classifies $intB = \{ \{x_4, x_{12}\}, \{x_3, x_7\} \}$ and unseen
640 traffic connections (using Eq. 7).

641 This classification method follows a binary structure that classifies network traf-
642 fic connections (or samples) as binary classifier. The binary structure method
643 took less processing time to other methods.

644 **Appendix B. Features of CICIDS2017**

Feature	Feature name	Type
f1	Flow ID	IP address
f2	Source IP	IP address
f3	Source Port	integer
f4	Destination IP	IP address
f5	Destination Port	integer
f6	Protocol	integer/string
f7	Time stamp	time
f8	Flow duration	integer
f9	Total forward packet	integer
f10	Total backward packet	integer
f11	Total length of forward packet	integer
f12	Total length of backward packet	integer
f13	Forward packet length max	integer
f14	Forward packet length min	integer
f15	Forward packet length mean	real
f16	Forward packet length std.	real
f17	Backward packet length max	integer
f18	Backward packet length min	integer
f19	Backward packet length mean	real
f20	Backward packet length std.	real
f21	Flow bytes/s	real
f22	Flow packets/s	real
f23	Flow IAT mean	real
f24	Flow IAT std.	real
f25	Flow IAT max	integer
f26	Flow IAT min	integer
f27	Forward IAT total	integer
f28	Forward IAT mean	real

Feature	Feature name	Type
f29	Forward IAT std.	real
f30	Forward IAT max	integer
f31	Forward IAT min	integer
f32	Backward IAT total	integer
f33	Backward IAT mean	real
f34	Backward IAT std.	real
f35	Backward IAT max	integer
f36	Backward IAT min	integer
f37	Forward PSH flags	binary
f38	Backward PSH flags	binary
f39	Forward URG flags	binary
f40	Backward URG flags	binary
f41	Forward header length	integer
f42	Backward header length	integer
f43	Forward packets/s	real
f44	Backward packets/s	real
f45	Min packet length	integer
f46	Max packet length	integer
f47	Packet length mean	real
f48	Packet length std.	real
f49	Packet length variance	real
f50	FIN flag count	binary
f51	SYN flag count	binary
f52	RST flag count	binary
f53	PSH flag count	binary
f54	ACK flag count	binary
f55	URG flag count	binary
f56	CWE flag count	binary
f57	ECE flag count	binary

Feature	Feature name	Type
f58	Down/Up ratio	integer
f59	Average packet size	real
f60	Average forward segment size	real
f61	Average backward segment size	real
f62	Forward header length	integer
f63	Forward average bytes/bulk	binary
f64	Forward average packets/bulk	binary
f65	Forward average bulk rate	binary
f66	Backward average bytes/bulk	binary
f67	Backward average packets/bulk	binary
f68	Backward average bulk rate	binary
f69	Subflow forward packets	integer
f70	Subflow forward bytes	integer
f71	subflow backward packets	integer
f72	subflow backward bytes	integer
f73	Init win bytes forward	integer
f74	Init win bytes backward	integer
f75	act data packet forward	integer
f76	min segment size forward	integer
f77	Active mean	real
f78	Active std.	real
f79	Active max	integer
f80	Active min	integer
f81	Idle mean	real
f82	Idle std.	real
f83	Idle max	integer
f84	Idle min	integer
f85	Label	string

645 **References**

- 646 [1] Z. Tan, A. Jamdagni, X. He, P. Nanda, R. P. Liu, J. Hu, Detection of denial-
647 of-service attacks based on computer vision techniques, *IEEE transactions*
648 *on computers* 64 (9) (2015) 2519–2533.
- 649 [2] J. Á. Cid-Fuentes, C. Szabo, K. Falkner, An adaptive framework for the
650 detection of novel botnets, *Computers & Security* 79 (2018) 148–161.
- 651 [3] I. Manzoor, N. Kumar, et al., A feature reduced intrusion detection system
652 using ann classifier, *Expert Systems with Applications* 88 (2017) 249–257.
- 653 [4] A. S. Eesa, Z. Orman, A. M. A. Brifcani, A novel feature-selection approach
654 based on the cuttlefish optimization algorithm for intrusion detection sys-
655 tems, *Expert Systems with Applications* 42 (5) (2015) 2670–2679.
- 656 [5] S. Elhag, A. Fernández, A. Bawakid, S. Alshomrani, F. Herrera, On the
657 combination of genetic fuzzy systems and pairwise learning for improving
658 detection rates on intrusion detection systems, *Expert Systems with Ap-*
659 *plications* 42 (1) (2015) 193–202.
- 660 [6] G. Gowrison, K. Ramar, K. Muneeswaran, T. Revathi, Minimal complexity
661 attack classification intrusion detection system, *Applied Soft Computing*
662 13 (2) (2013) 921–927.
- 663 [7] F. Kuang, W. Xu, S. Zhang, A novel hybrid kpca and svm with ga model
664 for intrusion detection, *Applied Soft Computing* 18 (2014) 178–184.
- 665 [8] W. Haider, J. Hu, J. Slay, B. Turnbull, Y. Xie, Generating realistic intrusion
666 detection system dataset based on fuzzy qualitative modeling, *Journal of*
667 *Network and Computer Applications* 87 (2017) 185–192.
- 668 [9] I. Sharafaldin, A. H. Lashkari, A. A. Ghorbani, Toward generating a new in-
669 trusion detection dataset and intrusion traffic characterization., in: *ICISSP*,
670 2018, pp. 108–116.

- 671 [10] M. S. Raza, U. Qamar, Feature selection using rough set-based direct de-
672 pendency calculation by avoiding the positive region, *International Journal*
673 *of Approximate Reasoning* 92 (2018) 175–197.
- 674 [11] S. Ghosh, P. S. Prasad, C. R. Rao, An efficient gaussian kernel based
675 fuzzy-rough set approach for feature selection, in: *International Workshop*
676 *on Multi-disciplinary Trends in Artificial Intelligence*, Springer, 2016, pp.
677 38–49.
- 678 [12] S. Ghosh, P. S. Prasad, C. R. Rao, Third order backward elimination ap-
679 proach for fuzzy-rough set based feature selection, in: *International Con-*
680 *ference on Pattern Recognition and Machine Intelligence*, Springer, 2017,
681 pp. 254–262.
- 682 [13] L. Sturlaugson, J. W. Sheppard, Uncertain and negative evidence in contin-
683 uous time bayesian networks, *International Journal of Approximate Rea-*
684 *soning* 70 (2016) 99–122.
- 685 [14] M. L. Thivagar, C. Richard, N. R. Paul, Mathematical innovations of a
686 modern topology in medical events, *International journal of information*
687 *science* 2 (4) (2012) 33–36.
- 688 [15] C. Luo, T. Li, H. Chen, H. Fujita, Z. Yi, Incremental rough set approach
689 for hierarchical multicriteria classification, *Information Sciences* 429 (2018)
690 72–87.
- 691 [16] M. Martinez, P. L. D. Leon, D. Keeley, Bayesian classification of falls risk,
692 *Gait & Posture* 67 (2019) 99–103.
- 693 [17] B. Selvakumar, K. Muneeswaran, Firefly algorithm based feature selection
694 for network intrusion detection, *Computers & Security* 81 (2019) 148–155.
- 695 [18] Y. Zhu, J. Liang, J. Chen, Z. Ming, An improved nsga-iii algorithm for
696 feature selection used in intrusion detection, *Knowledge-Based Systems*
697 116 (2017) 74–85.

- 698 [19] M. A. Ambusaidi, X. He, P. Nanda, Z. Tan, Building an intrusion detection
699 system using a filter-based feature selection algorithm, *IEEE transactions*
700 *on computers* 65 (10) (2016) 2986–2998.
- 701 [20] A. A. Aburomman, M. B. I. Reaz, A novel svm-knn-pso ensemble method
702 for intrusion detection system, *Applied Soft Computing* 38 (2016) 360–372.
- 703 [21] R. Singh, H. Kumar, R. Singla, An intrusion detection system using net-
704 work traffic profiling and online sequential extreme learning machine, *Ex-*
705 *pert Systems with Applications* 42 (22) (2015) 8609–8624.
- 706 [22] M. Tavallae, E. Bagheri, W. Lu, A. A. Ghorbani, A detailed analysis
707 of the kdd cup 99 data set, in: *Computational Intelligence for Security*
708 *and Defense Applications, 2009. CISDA 2009. IEEE Symposium on, IEEE,*
709 *2009*, pp. 1–6.
- 710 [23] G. Creech, J. Hu, Generation of a new ids test dataset: Time to retire the
711 kdd collection, in: *Wireless Communications and Networking Conference*
712 *(WCNC), 2013 IEEE, IEEE, 2013*, pp. 4487–4492.
- 713 [24] J. Song, H. Takakura, Y. Okabe, M. Eto, D. Inoue, K. Nakao, Statistical
714 analysis of honeypot data and building of kyoto 2006+ dataset for nids
715 evaluation, in: *Proceedings of the First Workshop on Building Analysis*
716 *Datasets and Gathering Experience Returns for Security, ACM, 2011*, pp.
717 29–36.
- 718 [25] A. Shiravi, H. Shiravi, M. Tavallae, A. A. Ghorbani, Toward developing
719 a systematic approach to generate benchmark datasets for intrusion detec-
720 tion, *computers & security* 31 (3) (2012) 357–374.
- 721 [26] L. Feng, S. Xu, F. Wang, S. Liu, H. Qiao, Rough extreme learning machine:
722 A new classification method based on uncertainty measure, *Neurocomput-*
723 *ing* 325 (2019) 269–282.
- 724 [27] E.-S. M. El-Alfy, M. A. Alshammari, Towards scalable rough set based at-
725 tribute subset selection for intrusion detection using parallel genetic algo-

- 726 rithm in mapreduce, *Simulation Modelling Practice and Theory* 64 (2016)
727 18–29.
- 728 [28] M. Abdel-Basset, M. Mohamed, The role of single valued neutrosophic
729 sets and rough sets in smart city: imperfect and incomplete information
730 systems, *Measurement* 124 (2018) 47–55.
- 731 [29] Y.-C. Hu, Flow-based tolerance rough sets for pattern classification, *Ap-
732 plied Soft Computing* 27 (2015) 322–331.
- 733 [30] H. Zhang, J. Zhou, D. Miao, C. Gao, Bayesian rough set model: A further
734 investigation, *International journal of approximate reasoning* 53 (4) (2012)
735 541–557.
- 736 [31] W. Hadi, Q. A. Al-Radaideh, S. Alhawari, Integrating associative rule-
737 based classification with naïve bayes for text classification, *Applied Soft
738 Computing* 69 (2018) 344–356.
- 739 [32] N. S. Harzevili, S. H. Alizadeh, Mixture of latent multinomial naive bayes
740 classifier, *Applied Soft Computing* 69 (2018) 516–527.
- 741 [33] Y. Li, J. Xia, S. Zhang, J. Yan, X. Ai, K. Dai, An efficient intrusion de-
742 tection system based on support vector machines and gradually feature
743 removal method, *Expert Systems with Applications* 39 (1) (2012) 424–430.