

A peer-reviewed version of this preprint was published in PeerJ on 13 June 2016.

[View the peer-reviewed version](https://peerj.com/articles/cs-65) (peerj.com/articles/cs-65), which is the preferred citable publication unless you specifically need to cite this preprint.

Chamikara MAP, Galappaththi A, Yapa RD, Nawarathna RD, Kodituwakku SR, Gunatilake J, Jayathilake AAC, Liyanage L. 2016. Fuzzy based binary feature profiling for modus operandi analysis. PeerJ Computer Science 2:e65
<https://doi.org/10.7717/peerj-cs.65>

Fuzzy based binary feature profiling for modus operandi analysis

Mahawaga Arachchige Pathum Chamikara, Akalanka Galappaththi, Roshan D Yapa, Ruwan D Nawarathna, Saluka Ranasinghe Kodituwakku, Jagath Gunatilake, Aththanapola Arachchilage Chathranee Anumitha Jayathilake, Liwan H Liyanage

It is a well-known fact that some criminals follow perpetual methods of operations, known as modus operandi (MO) which is commonly used to describe the habits in committing something especially in the context of criminal investigations. These modus operandi are then used in relating criminals to other crimes where the suspect has not yet been recognized. This paper presents a method which is focused on identifying the perpetual modus operandi of criminals by analyzing their previous convictions. The method involves in generating a feature matrix for a particular suspect based on the flow of events. Then, based on the feature matrix, two representative modus operandi are generated: complete modus operandi and dynamic modus operandi. These two representative modus operandi will be compared with the flow of events of the crime in order to investigate and relate a particular criminal. This comparison uses several operations to generate two other outputs: completeness probability and deviation probability. These two outcomes are used as inputs to a fuzzy inference system to generate a score value which is used in providing a measurement for the similarity between the suspect and the crime at hand. The method was evaluated using actual crime data and four other open data sets. Then ROC analysis was performed to justify the validity and the generalizability of the proposed method. In addition, comparison with five other classification algorithms showed that the proposed method performs competitively with other related methods.

1 **Fuzzy based binary feature profiling for modus operandi**
2 **analysis**
3

4

5

6

7 M. A. P. Chamikara^{1,2*}, A. Galappaththi¹, Y.P.R.D. Yapa^{1,2}, R.D. Nawarathna^{1,2}, S. R.
8 Kodituwakku^{1,2}, J. Gunatilake^{1,3}, A.A.C.A. Jayathilake⁴, L.H. Liyanage⁵

9

10

11 ¹Postgraduate Institute of Science, University of Peradeniya, Sri Lanka,

12 ²Department of Statistics and Computer Science, University of Peradeniya, Sri Lanka,

13 ³Department of Geology, University of Peradeniya, Sri Lanka,

14 ⁴Department of Mathematics, University of Peradeniya, Sri Lanka,

15 ⁵School of Computing, Engineering and Mathematics, University of Western Sydney, Australia.

16

17

18

19

20

21 Corresponding author:

22

23 M.A.P. Chamikara,

24 No.22, Nikaketiya, Menikhinna, 20000, Sri Lanka

25 Email address: pathumchamikara@gmail.com

26

27

28 **Abstract**

29 It is a well-known fact that some criminals follow perpetual methods of operations, known as
30 modus operandi (MO) which is commonly used to describe the habits in committing something
31 especially in the context of criminal investigations. These modus operandi are then used in
32 relating criminals to other crimes where the suspect has not yet been recognized. This paper
33 presents a method which is focused on identifying the perpetual modus operandi of criminals
34 by analyzing their previous convictions. The method involves in generating a feature matrix for
35 a particular suspect based on the flow of events. Then, based on the feature matrix, two
36 representative modus operandi are generated: complete modus operandi and dynamic modus
37 operandi. These two representative modus operandi will be compared with the flow of events
38 of the crime in order to investigate and relate a particular criminal. This comparison uses
39 several operations to generate two other outputs: completeness probability and deviation
40 probability. These two outcomes are used as inputs to a fuzzy inference system to generate a
41 score value which is used in providing a measurement for the similarity between the suspect
42 and the crime at hand. The method was evaluated using actual crime data and four other open
43 data sets. Then ROC analysis was performed to justify the validity and the generalizability of the
44 proposed method. In addition, comparison with five other classification algorithms showed that
45 the proposed method performs competitively with other related methods.

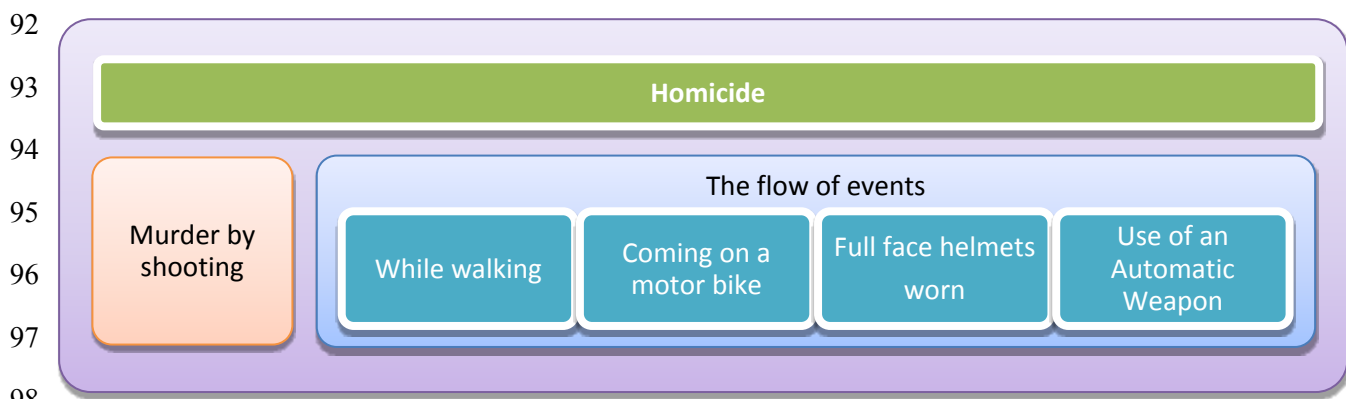
46 **Introduction**

47 Scientists have long played a role in examining deviant behavior in society. "Deviance
48 behaviour" is a term used by scientists to refer to some form of "rule-breaking" behaviour [1]. It
49 can be the behaviour of violating a social norm or the law. Criminal behaviour is also a form of
50 deviance, one that is defined as the breaking of legal rules. Nevertheless, there is a difference
51 between deviance and crime. Deviance involves breaking a norm and evoking a negative
52 reaction from others. Crime is a deviance that breaks a law, which is a norm stipulated and
53 enforced by government bodies [1]. However, crimes affect the society negatively. Therefore,
54 law enforcement authorities take necessary actions to mitigate crime in an environment where
55 high crime frequencies are observed each year. In this exercise the application of technology for
56 crime analysis is being widened in the world. Locard's Exchange principle states that every
57 contact of the perpetrators of a crime scene leaves a trace. The perpetrators will both bring
58 something into the scene and leave with something from the scene [2]. However, the cognitive
59 abilities of criminals will always make them minimize their risks of apprehension by conducting
60 the perfect crime and maximizing their gain [3]. Modus operandi or method of operation such
61 as preparation actions, crime methods and weapons are frequently used in criminal profiling
62 because the past crime trends show that, after criminals get used to a certain method of
63 operation, they try to use the same modus operandi in committing his/her next crime [4].

64 The criminals develop a set of actions during the performance of a series of crimes which we
65 refer to as "modus operandi" (MO). MO is developed with the crimes he/she commits and the
66 nature of trying to stick with the developed MO that has worked throughout the previous

67 crimes [5]. In any criminal career, the MO happens to evolve, no matter what the
68 circumstances. Also, it is a common behaviour that serial offenders tend to exhibit significant
69 behaviour known as his/her signature. Therefore, MO of criminals plays a major role in
70 investigating crimes [5]. It is a known fact that features such as criminal signature are used in
71 crime investigation in almost all the police departments around the world. Sri Lanka police also
72 uses MOs of suspects to identify the criminals who have conducted crimes. Currently Sri Lanka
73 Police uses a manual crime recording and investigation system. This manual system has many
74 problems such as data redundancy, inefficiency, tediousness, inability to support crime
75 investigation and many other problems associated with a conventional manual system. To
76 overcome these problems, a web-based framework was proposed with geographical
77 information support containing a centralized database for crime data storage and retrieval,
78 named SL-CIDSS: Sri Lanka Crime Investigation and Decision Support System [6].

79 According to the penal code of Sri Lanka first enacted in 1882 and amended subsequently
80 several times in later years [7], Sri Lanka police classifies crimes into two categories: Grave
81 crimes and Minor offences. Until 2014, grave crimes were classified under 21 crime categories
82 and in 2015 another 5 new crime categories were introduced, making it 26 categories of grave
83 crime types. Kidnapping, Fraud or mischief causing damage greater than 25000 rupees,
84 Burglary, Grievous hurt, Hurt by sharp weapon, Homicide, Rape, Robbery, Cheating by trust,
85 Theft are 10 of the most frequent crime types. To identify the patterns involved in crimes, a
86 collection of subtypes were identified under these 26 crime types. These subtypes have been
87 created mainly for the purpose of modus operandi analysis. Most frequent behaviors of
88 criminals/crimes are considered as crime subtypes. When a crime is logged in the Grave Crime
89 Record (GCR) book, it is classified under one of the 26 main categories. But, under the section
90 of “nature of crime” in the GCR book, the police officers record the flow of the crime incident
91 including the subtypes.



99 **Figure 1.** Relationship between main crime type, subtypes and crime flows

100 A subtype is a sub category of one of the main crime types. This flow of crime is written in the
101 GCR book in the nature of the crime section as a description. For investigation, the nature of
102 the crimes is broken into subtypes and flows according to their frequency of occurrence and
103 uniqueness. These sub categorizations have been introduced mainly to minimize the broadness

104 of main type and to improve clarity. Fig.1. depicts the relationship of the subtypes and flows
105 where there can be a flow of events to a crime recorded as one of the 26 main crime types. For
106 the simplicity and easy handling of data, the investigators have provided subtype codes and
107 flow codes. The flow of events provides a modus operandi which is most of the time unique to
108 an offender. Each subtype is provided with a code under the main type, to make the crime
109 investigation process easier. For example, ROB/S001 denotes a subtype that is Highway
110 robbery; here ROB denotes the main type under which the corresponding subtype appears. In
111 this case, it is Robbery. Crime types are further subdivided into sub types to make the analysis
112 and processing simpler. In this manner, crime subtypes and flows have been identified under all
113 the 26 crime types. The space for adding more subtypes and flows under these crime types
114 exists. A new subtype or a flow is introduced to a particular main crime, if the same subtype or
115 the flow happens to persist for a prolonged time.

116 This paper proposes a novel method of criminal profiling using the modus operandi which can
117 be used to identify associations between crimes and chronic criminals. The method is based on
118 a new technique named, “binary feature vector profiling”. Key relationships between a criminal
119 and the conducted crimes are analyzed using binary feature profiling and association rule
120 mining techniques. Due to the impreciseness and vagueness of these extracted attributes, a
121 fuzzy inference system is used in making the final decision. The newly proposed method was
122 adapted into a classification algorithm in order to test its accuracy. An actual crime data set
123 which was obtained from Sri Lanka Police was used in testing the performance of the newly
124 proposed method and it was compared against five well established classification algorithms.
125 Comparisons were done using Friedman’s rank tests. The results confirmed that the proposed
126 method produced competitive results compared to the other five classification algorithms.

127 The rest of the paper is organized as follows. Related work section presents a summary of the
128 work that has been done on modus operandi analysis as well as a brief discussion on crime
129 investigation using link analysis and association mining in general. Materials and Methods
130 section discusses the main steps of the newly proposed algorithm. Next, Results and Discussion
131 section provides a validation and performance evaluation of the newly proposed method along
132 with a performance comparison with five other classification algorithms. Finally, some
133 concluding remarks and future enhancements are outlined in the conclusion section.

134 **Related work**

135 Literature shows many methods which have been developed in the area of automated crime
136 investigation. Our major concern has been laid upon the research carried out on crime
137 investigation using association mining as our research considers on developing a model to find
138 the associations between the criminals and the crimes depending on the modes operandi. C
139 Bennell and DV Canter [8] have proposed a method to use statistical models to test directly the
140 police practice of utilizing modus operandi to link crimes to a common offender. The results
141 indicated that certain features such as the distance between burglary locations, lead to high
142 levels of predictive accuracy. Craig Bennell, et. al. [9] have tried to determine if readily

143 available information about commercial and residential serial burglaries, in the form of the
144 offender's modus operandi, provides a statistically significant basis for accurately linking crimes
145 committed by the same offenders. Benoit Leclerc, et al. [10] have reviewed the theoretical,
146 empirical, and practical implications related to the modus operandi of sexual offenders against
147 children. They have presented the rational choice perspective in criminology followed by
148 descriptive studies aimed specifically at providing information on modus operandi of sexual
149 offenders against children. Clustering crimes, finding links between crimes, profiling offenders
150 and criminal network detection are some of the common areas where data mining is applied in
151 crime analysis [11]. Association analysis, classification and prediction, cluster analysis, and
152 outlier analysis are some of the traditional data mining techniques which can be used to
153 identify patterns in structured data. Offender profiling is a methodology which is used in
154 profiling unknown criminals or offenders. The purpose of offender profiling is to identify the
155 socio-demographic characteristics of an offender based on information available at the crime
156 scene [12]. Association rule mining discovers the items in databases which occur frequently and
157 present them as rules. Since this method is often used in market basket analysis to find which
158 products are bought with what other products, it can also be used to find associated crimes
159 conducted with what other crimes. Here, the rules are mainly evaluated by the two probability
160 measures, support and confidence [13]. Association rule mining can also be used to identify the
161 environmental factors that affect crimes using the geographical references [14]. Incident
162 association mining and entity association mining are two applications of association rule
163 mining. Incident association mining can be used to find the crimes committed by the same
164 offender and then the unresolved crimes can be linked to find the offender who committed
165 them. Therefore, this technique is normally used to solve serial crimes like serial sexual offenses
166 and serial homicide [15].

167 Similarity-based association mining and outlier-based association mining are two approaches
168 used in incident association mining. Similarity-based association mining is used mainly to
169 compare the features of the crime with the criminal's behavioral patterns which are referred as
170 modus operandi or behavioral signature. In outlier-based association mining, crime associations
171 will be created on the fact that both the crime and the criminal have the possibility of having
172 some distinctive feature or a deviant behavior [16]. Entity association mining/link analysis is the
173 task of finding and charting associations between crime entities such as persons, weapons, and
174 organizations. The purpose of this technique is to find out how crime entities that appear to be
175 unrelated at the surface, are actually linked to each other [15]. Link analysis is also used as one
176 the most applicable methods in social network analysis [17] in finding crime groups, gate
177 keepers and leaders [18].

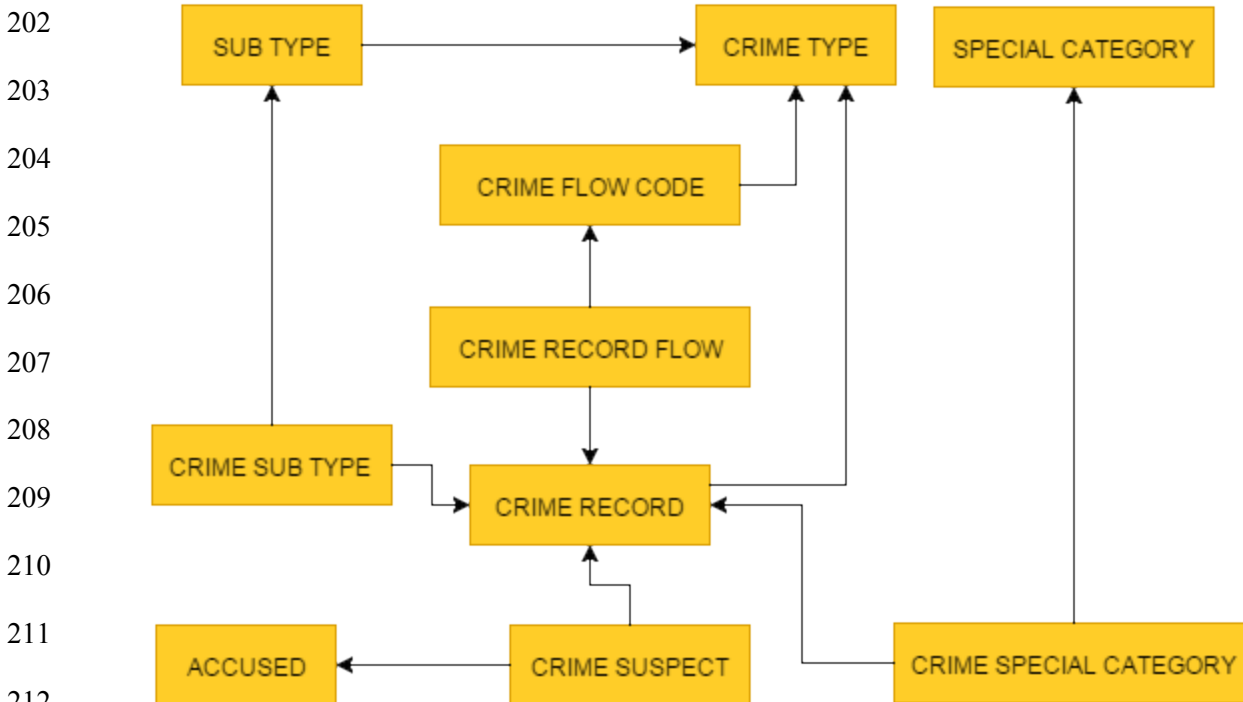
178 Attribution can be used to link crimes to offenders. If two offences in different places involve
179 the same specific type, those may be readily attributed to the same offender [11]. There are
180 three types of link analysis approaches, namely Heuristic-based, Statistical-based and Template-
181 based [15]. Sequential pattern mining is also a similar technique to association rule mining. This
182 method discovers frequently occurring items from a set of transactions occurred at different
183 times [19]. Deviation detection detects data that deviates significantly from the rest of the data
184 which is analyzed. This is also called outlier detection, and is used in fraud detection [19]. In

185 classification, the data points will be assigned to a set of predefined classes of data by
186 identifying a set of common properties among them. This technique is often used to predict
187 crime trends. Classification needs a reasonably complete set of training and testing data since a
188 high degree of missing data would limit the prediction accuracy [19]. Classification comes under
189 supervised learning method [15] which includes methods such as Bayesian models, decision
190 trees, artificial neural networks [20] and support vector machines. String comparison
191 techniques are used to detect the similarity between the records. Classification algorithms
192 compare the database record pairs and determine the similarity among them. This concept can
193 be used to avoid deceptive offender profiles. Information of offenders such as name, address,
194 etc. might be deceptive and therefore the crime database might contain multiple records of the
195 same offender. This makes the process of identification of their true identity difficult [19].

196 **Materials and Methods**

197 This section provides a description about the materials and methods used in developing the
198 fuzzy based binary feature profiling for modus operandi analysis. First, an overview about how
199 SL-CIDSS captures the logics of modus operandi is explained. Then a detailed description about
200 the steps of the newly proposed algorithm is explained.

201



213 **Figure 2.** *Crime flow entity arrangement of SL-CIDSS*

214

215 Figure 2 shows how SL-CIDSS database captures the crime types and subtypes. A crime record
216 has a crime record flow. Typically, a crime is committed by a criminal and a particular accused
217 might commit one or more crimes. A CRIME RECORD can be of one the 26 crime types. A
218 particular CRIME RECORD will be considered under one main CRIME TYPE with the highest
219 precedence in the order of seriousness. For example, a crime incident that includes a murder
220 and a robbery will be categorized as a murder though a robbery has also taken place. But in the
221 nature of crime section, all crimes followed by the main type will be stated. Therefore the
222 CRIME RECORD FLOW captures all the steps of the crime as a sequence of steps recorded. The
223 crime flows that have been previously registered are mapped under CRIME FLOW CODE. Also, a
224 particular CRIME RECORD instance can contain multiple SUB TYPES which are recorded as
225 CRIME SUB TYPE. The SPECIAL CATEGORY captures the crimes with special features such as
226 crimes occurring at the same location or retail shop. A crime may involve several special
227 categories which are saved in the CRIME SPECIAL CATEGORY. The ACCUSED entity records the
228 information of suspects and accused and they are related to crime through the CRIME SUSPECT
229 entity.

230 As the first step of the newly employed method, a feature matrix is generated, resulting in a
231 binary matrix representing the crime flows. This binary feature matrix is composed of the
232 binary patterns generated on previous convictions of a particular criminal/suspect. As the MOs
233 are represented in binary all the analyses are conducted on this binary feature matrix. This
234 binary form of the feature matrix provides a provision to direct application of computer
235 algorithms with methods such as Apriori based association rule mining as the matrix is already
236 in the binary format. The reduced complexity of the binary feature matrices provides an easy
237 manipulation over the categorical and continuous valued features. Figure 3 shows the steps of
238 the proposed MO analysis algorithm.

- 239
- 240 **Step 1:** Generate the feature matrix.
 - 241 **Step 2:** Generate the dynamic MOs (DMO) of the criminals.
 - 242 **Step 3:** Generate the complete MO profile (CMOP) of the criminals.
 - 243 **Step 4:** Find deviation probability (DP) of CMOP from the crime MO under consideration (UMO).
 - 244 **Step 5:** Find completeness probability of UMO against DMO.
 - 245 **Step 6:** Use the two values obtained from step 4 and 5 as inputs of a fuzzy Inference system to
246 obtain the final similarity value (out of 100).
 - 247 **Step 7:** Classify the UMO under the class with highest similarity score for validation.

248 **Figure 3.** Steps of the newly employed algorithm

249

250 **Generating the binary feature matrix**

251 Table 1 shows how the feature matrix is generated and provides the way to generate modus
 252 operandi of criminals as binary sequences. According to the table, events of the crime scene are
 253 observed starting from its crime type. After a particular crime type is identified, the feature
 254 matrix is updated with ones for each subtype and flow code that is available in the crime or
 255 suspect modus operandi. The matrix will be filled by zeros in places which the modus operandi
 256 does not have any contact with. The column names to the feature matrix are generated in such
 257 a way that it covers the collection of main types, sub types, crime flows and special categories
 258 at hand. For example, if we consider the list of crime types, subtypes, crime flows and the
 259 special category in Table 1, it results in a feature matrix of a 21-bit vector shown in the last two
 260 columns.

261 **Table 1.** *An instance of a feature selection for the feature matrix generation*

Main Semantic	Crime flow element code	Description	Suspect 1	Suspect 2
Crime types	HB	House Breaking	0	1
	HK	Hurt by Knife	0	0
	RB	Robbery	1	0
	TH	Theft	0	0
Sub types	ABD/S003	Abduction from the legal guardian	1	0
	ABD/S004	Abducting to marry	0	0
	ABD/S005	Abducting for sexual harassment	0	0
	BGL/S004	Use of stealth	0	1
	BGL/S011	Burglary in business places	0	0
	ROB/S001	Organized vehicle robbery	1	0
Crime Flows	BGL/F001	Entering from the window	0	1
	BGL/F002	Entering from the Fanlights	0	0
	BGL/F003	Removing grills	0	1
	BGL/F004	Breaking glasses	0	0
	ROB/F001	Showing identity cards	1	0

	ROB/F003	Wearing uniforms	1	0
	ROB/F004	Robbery using identity cards, uniforms and chains	0	1
	ROB/F009	Seizing inmates	0	0
	ROB/F010	Appearing as CID officers	0	0
Special Category	Retailer 1	Attacking/ robbing retailer 1's stores	0	0
	Retailer 2	Attacking/ robbing retailer 2's stores	0	0

262 In this manner we can produce binary MO patterns based on the crimes committed by different
263 criminals as shown in the last two columns of Table 1. According to Table 1, Accused 1 has
264 committed a robbery with the subtypes, ABD/S003 (an abduction of a child from the legal
265 guardian), ROB/S001 (an organized vehicle robbery) and the flows, ROB/F001 (Identity cards
266 have been shown), ROB/F003 (accused has been wearing uniforms). Accused 2 has committed a
267 house breaking with the sub type BGL/S011 (use of stealth), and the flows, BGL/F001 (Entering
268 from the window), BGL/F003 (Removing Grills).

269 In this manner, depending on the complete crime MO under consideration, it may generate
270 modus operandi of different lengths. According to the full crime MO, criminal based MOs can
271 be generated and taken into a full feature matrix of binary patterns. *ct*, *st*, *fl* and *sc* in Table 2
272 represents the abbreviations for crime type, sub type, flow and special category respectively.

273 **Table 2.** Feature matrix generated using the selected modus operandi attributes in Table 4.

	ct1	ct2	ct3	ct4	st1	st2	st3	st4	st5	st6	fl1	fl2	fl3	fl4	fl5	fl6	fl7	fl8	fl9	sc1	sc2
Accused 1	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0
Accused 2	1	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	1	0	0	0	0

274

275 **Generating the dynamic MOs (DMOs) of the criminals**

276 Dynamic MO is a binary feature vector which is generated on bit patterns of the feature matrix
277 of a particular criminal. The main purpose of the DMO is to obtain a criminal specific crime flow
278 which captures the crime patterns which are frequently followed by a particular criminal. It
279 was named as the dynamic modus operandi as it is subject to change when the new crime flows
280 are added to the feature matrix. Therefore, this addresses the changing nature of the patterns
281 used by the criminals in committing crimes. First, a frequency threshold is generated using
282 characteristic features of the sub matrix at hand which is the matrix of all crimes committed by
283 the same criminal under consideration. If the corresponding accused has convicted
284 for/committed four crimes, the four bit patterns related to those four crimes will be available in

285 the feature matrix. The matrix shown in Table 3 is an example to a situation of a feature matrix
 286 generated on the previous convictions of a criminal. For simplicity let's consider a feature
 287 matrix of 10 columns.

288 **Table 3.** Feature matrix generated on the four previous convictions of a criminal.

A	B	C	D	E	F	G	H	I	J
1	0	0	0	1	1	1	0	1	0
1	0	0	1	0	1	1	0	1	0
1	0	0	0	1	1	0	1	1	0
1	0	0	1	0	1	1	1	1	0

289
 290 If we consider A to J of Table 3 as crime flow features of the corresponding MOs, we can
 291 understand that in the first MO the criminal happens to have conducted a flow of A-E-F-G-I.
 292 The second criminal has followed a crime flow of A-D-F-G-I-J. Likewise the other two flows are,
 293 A-E-F-H-I and A-D-F-G-H-I respectively.

294 The dynamic MOs (DMO) are generated for all the suspects/accused included in the feature
 295 matrix. The DMO of a particular criminal is generated using the idea of Apriori method [21].
 296 Apriori method is used to find the crime entities with the frequency threshold (frt) which is
 297 generated according to Equation 3. Apriori is a method to find frequent item sets in
 298 transactions in association rule mining [21]. A demonstration of the generation of D in Equation
 299 1 on the properties of feature matrix is shown in Table 4.

$$D = \left\{ d \mid d = \sum_{i=1}^n y_i \right\} \quad (1)$$

$$M_D = l_D + \frac{h}{f_D} \left(\frac{2}{n} - c \right) \quad (2)$$

$$frt = \frac{M_D}{n} \quad (3)$$

300 Where,

301 y_i = cells in each column

302 M_D = Median of D,

303 l_D = lower class boundary of the model class,

304 f_D = frequency of the median class,

305 $n = \sum f$ = number of values or total frequencies,


306 c = cumulative frequency of the median class

307 h = class interval size.


308

309 **Table 4.** Column-wise addition of the feature matrix of the suspect under consideration

A	B	C	D	E	F	G	H	I	J
1	0	0	0	1	1	1	0	1	0
1	0	0	1	0	1	1	0	1	0
1	0	0	0	1	1	0	1	1	0
1	0	0	1	0	1	1	1	1	0
4	0	0	2	2	4	3	2	4	0



 Summation



310

311 The column-wise addition of the matrix shown in Table 4 gives 4, 0, 0, 2, 2, 4, 3, 2, 4 and 0. The
 312 unique numbers are selected from the resulting vector which results in 4, 0, 2 and 3. Since set D
 313 has no repeated elements, *D* carries only 4,0,2 and 3, where 4,0,2 and 3 are the unique
 314 numbers obtained from the summation operation. The median of D is then divided by the
 315 number of instances (rows) in the matrix as the frt, which is $2.5/4 = 0.625$ for the above case.
 316 Therefore, frt will range within 0 to 1. This value provides an insight to a fair threshold value for
 317 the Apriori method to generate the dynamic modus operandi with the most frequent elements.
 318 frt is used as the frequency threshold in finding the lengthiest MO with a probability of 0.625
 319 because this value suggests that there is a moderate possibility of one feature having 0.625
 320 probability in each of MO. This results in a dynamic modus operandi (DMO) as shown in
 321 Equation 5, because the only transaction of crime attributes which provides a support of 0.625
 322 is $\sigma(A,F,G,I)$ as shown in Equation 4.

$$s = \frac{\sigma(A,F,G,I)}{|T|} = \frac{3}{4} = 0.75 \tag{4}$$

323

$$DMO = [1 \ 0 \ 0 \ 0 \ 0 \ 1 \ 1 \ 0 \ 1 \ 0] \tag{5}$$


324

325 **Generating the complete MO profile (CMOP) of the criminals**

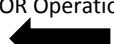
326
 327 The complete MO profile (CMOP) is obtained by the OR operation between the bits of each
 328 column of the feature matrix of the corresponding criminal. CMOP guarantees the provision of a
 329 composite crime flow by considering all of the previous crime flow entities of a particular criminal. For
 330 example, the complete profile for the feature matrix shown in Table 3 is obtained as shown in
 331 Table 5.

332 **Table 5.** OR operation on the columns to obtain the complete MO profile.

A	B	C	D	E	F	G	H	I	J
1	0	0	0	1	1	1	0	1	0
1	0	0	1	0	1	1	0	1	0
1	0	0	0	1	1	0	1	1	0
1	0	0	1	0	1	1	1	1	0
1	0	0	1	1	1	1	1	1	0



 OR Operation



333 Therefore, $CMOP = [1\ 0\ 0\ 1\ 1\ 1\ 1\ 1\ 1\ 0]$. CMOP contains 1s for each place for which a
334 particular crime flow entity has taken place at least once.

335 **Finding deviation probability (DP) of CMOP from the crime MO under** 336 **consideration (UMO)**

337 First the deviation of each individual flow (IFD) in the criminals feature matrix is obtained
338 according to Equation 6. This probability value is used to obtain a measurement as to what
339 extent of information is available in the UMO, extra to what is already available in the CMOP of
340 a particular criminal. Let's assume that the bit pattern to be compared with the criminal's
341 modus operandi profile under consideration is $UMO = [1\ 0\ 0\ 0\ 1\ 1\ 1\ 0\ 0\ 1]$. Here a measurement
342 is taken called deviation probability (DP) which provides the probability of 1s which are
343 available in UMO but not in CMOP. This provides a measurement about features available in
344 UMO but not in CMOP. Therefore,

345 The deviation probability, DP can be given as,

$$346 \quad DP = \frac{\sum_{i=1}^n x_i - y_i}{n}, \text{ for } x_i = 1 + y_i; i = 1, 2, \dots, n \quad (6)$$

347 Where,

348 $x_i = \text{elements of the UMO}$

349 $y_i = \text{elements of the CMOP}$

350 If we consider the feature matrix on Table 7,

$$351 \quad Deviation = [1\ 0\ 0\ 0\ 1\ 1\ 1\ 0\ 0\ 1] - [1\ 0\ 0\ 1\ 1\ 1\ 1\ 1\ 1\ 0]$$

$$352 \quad Deviation = [0\ 0\ 0\ -1\ 0\ 0\ 0\ -1\ -1\ 1] \quad (7)$$

353 Define $AD = 1$, where AD is the number of positive 1s.

354 Therefore, $DP = 1/10 = 0.1$

355 As it appears in Equation 7, it produces positive 1s for the places with the features available in
356 UMO but not in CMOP. Therefore, DP provides a notion of how much extra information
357 available in UMO compared to CMOP. The higher the DP, higher the amount of extra
358 information available in UMO. Hence, a DP value close to 0 indicates the absence of extra
359 features in UMO.

360 **Finding completeness probability (CP) of UMO against DMO**

361 For the same feature matrix which was considered in Table 3, the CP is obtained according to
362 Equation 8. Here, the UMO is compared with DMO to obtain a probability of, to what extent

363 the features available under the CP is available in UMO. Therefore, it is derived by the
 364 percentage of attributes which are present in both UMO and DMO.

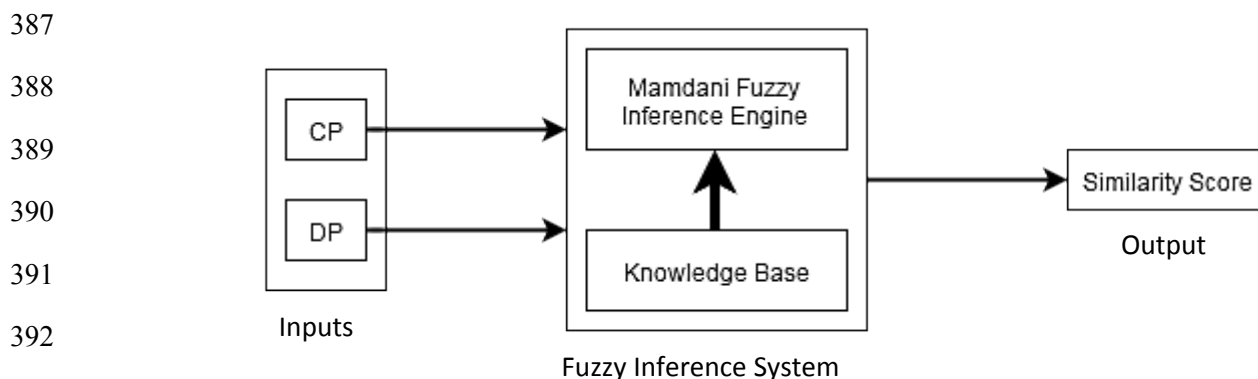
365 Let $DMO = \{x_i\}_{i=1}^n$ and $UMO = \{y_j\}_{j=1}^n$ be two binary sequences.

366 Define, $z_k = \begin{cases} 1; & x_i = y_j \\ 0; & otherwise \end{cases}$. Then, $CP = \frac{\sum_{k=1}^n z_k}{n}$ is the completeness probability. (8)

367 For example, if we consider $DMO = [1\ 0\ 0\ 0\ 0\ 1\ 1\ 0\ 1\ 0]$, then for the
 368 $UMO = [1\ 0\ 0\ 0\ 1\ 1\ 1\ 0\ 0\ 1]$ a CP of $3/10 = 0.3$ is generated as in the 1st, 6th and 7th positions
 369 of UMO where there are ones in both DMO and UMO. Higher the CP value, the more the UMO
 370 is composed of crime flow entities which are available in the DMO. Therefore, a CP value close
 371 to 1 indicates that the completeness of UMO compared to DMO is 100%.

372 Building a fuzzy inference system to obtain the final similarity score

373 The vagueness of the two measurements CP and DP generates a difficulty in calculating a similarity score
 374 using crisp logic. Therefore, fuzziness was introduced to CP and DP in order to generate a fuzzy inference
 375 system. Two values CP and DP were fed into a Mamdani inference [22] based fuzzy system which
 376 accepts two inputs and provides a score for the similarity between the DMO and UMO. Figure 4 shows a
 377 block diagram of the proposed fuzzy inference system. Mamdani fuzzy inference was proposed as an
 378 attempt to solve a control problem by a set of linguistic rules obtained from experienced human
 379 operators [22]. First, the rule base of the fuzzy controller was defined by observing the variations of CP
 380 and DP. The membership functions of the inputs and outputs were then adjusted in such a way that, the
 381 parameters which seem to be wrong can be fine-tuned, which is a common practice in defining fuzzy
 382 inference systems [23]. Therefore, the input and output space of the two inputs CP and DP and the
 383 output were partitioned into 3 subsets. Namely, LOW, MODERATE and HIGH. Center of gravity was used
 384 as the defuzzification strategy of the fuzzy controller. Mamdani fuzzy inference was especially selected
 385 for the similarity score generation procedure, for the highly intuitive knowledge base it offers due to the
 386 fact that both antecedents and the consequents of the rules are expressed as linguistic constraints [24].



394 **Figure 4.** Block diagram of the proposed fuzzy inference system.

395 Figures 5 and 6 show the fuzzy inputs of the Fuzzy Inference System (FIS) which correspond to
 396 CP and DP values respectively. Figure 7 depicts the fuzzy output of the FIS. As the Figures 5, 6

397 and 7 depict, all the different levels of membership functions under each input and the output
 398 are selected to be triangular and trapezoidal functions as triangular or trapezoidal shapes are
 399 simple to implement and computationally efficient [25].

400

401

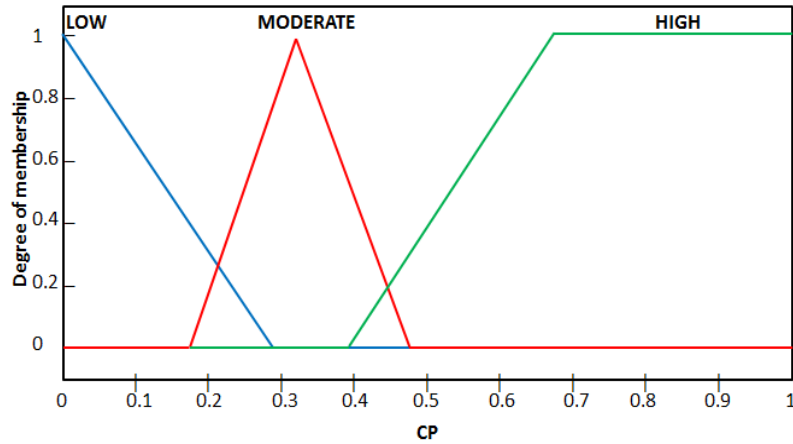
402

403

404

405

406



407

Figure 5. Input fuzzy variable 1: CP

408

409

410

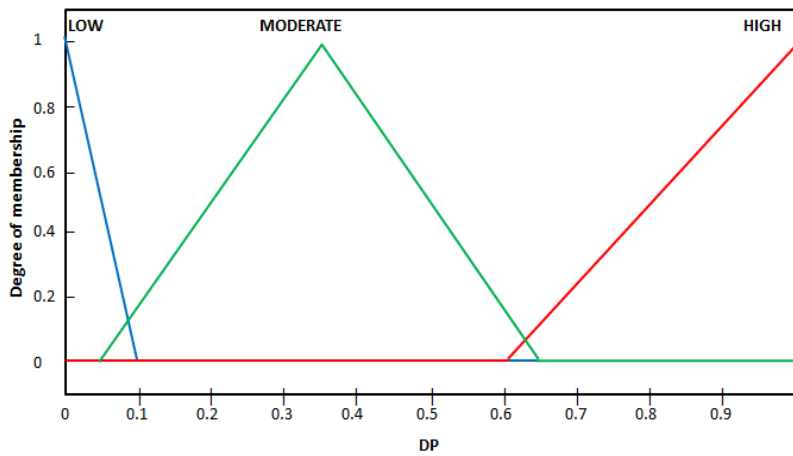
411

412

413

414

415



416

Figure 6. Input fuzzy variable 2: DP

417

418

419

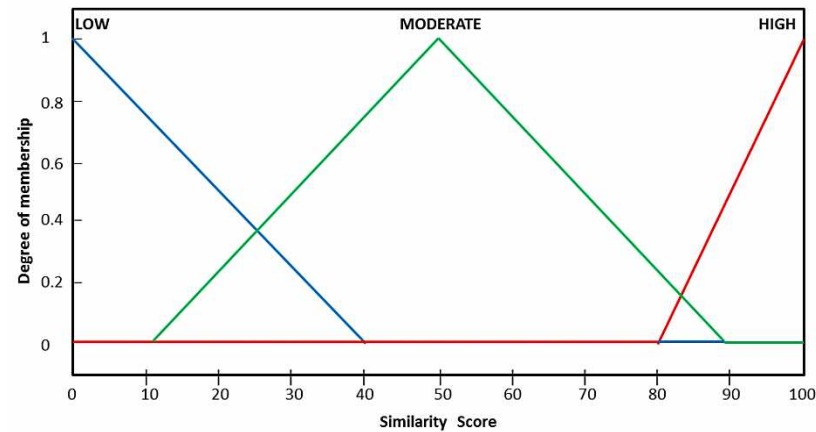
420

421

422

423

424



425

Figure 7. Output fuzzy variable: similarity score.

426 As shown in Figure 7, the universe of discourse of similarity score (fuzzy output) ranges from 0
 427 to 100. The defuzzifier score which is generated from the FIS is considered as the measurement
 428 for how close the modus operandi under consideration is to a particular suspect's profile. A
 429 higher score value close to 100, which is generated from the FIS provides a good indication
 430 about a high similarity between the modus operandi of the crime and suspect under
 431 consideration.

432 The fuzzy rule derivation of the fuzzy controller is heuristic in nature. According to the
 433 calculations of the two inputs, for higher values of CP, close to 1 and lower values of DP close to
 434 0, positively affect the final similarity score. Therefore, the rule base of the fuzzy model is
 435 generated accordingly. Our rule base provides a non-sparse rule composition of 9
 436 combinations as illustrated in Figure 8.

437

DP \ CP	LOW	MODERATE	HIGH
LOW	MODERATE	LOW	MODERATE
MODERATE	HIGH	MODERATE	LOW
HIGH	HIGH	MODERATE	LOW

438

439

440

441

442

Figure 8. Fuzzy rule set of the rule base of the inference system.

443 The rule surface of the fuzzy controller depicted in Figure 9, shows the variation of the score
 444 value with the changes of the two inputs CP and DP. According to the figure it's perfectly
 445 visible, for higher values of CP (close to 1) and for lower values of DP (close to 0), the fuzzy
 446 controller generates higher values for the score which are close to 100.

447

448

449

450

451

452

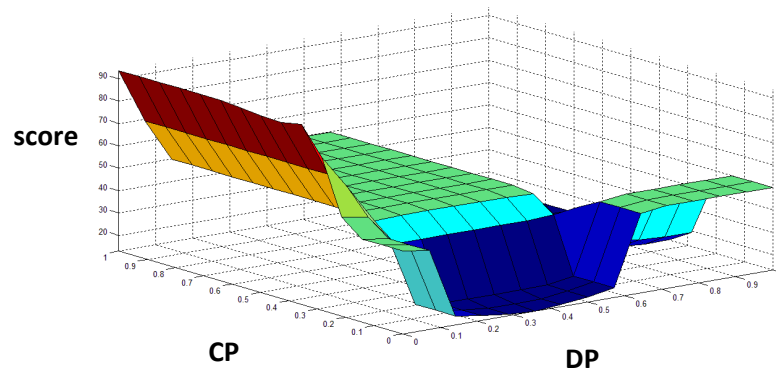


Figure 9. Rule surface of the fuzzy controller.

453 **Classification of the UMO under the class with highest similarity score**

454 When the algorithm is used to find associations between modus operandi of criminals/ suspects
455 with modus operandi of crimes, the similarity score which is generated from the newly
456 proposed method can be used directly. The similarity score provides a measurement to how
457 much an accused is related to a particular criminal in the sense of marks which varies from 0 to
458 100. A mark which is close to 100 would suggest that the criminal has a very high tendency to
459 have committed the crime which is under investigation. The similarity scores generated are
460 used to classify a particular modus operandi to a most probable suspect with the highest
461 similarity score. Therefore, the new method provides the flexibility of being used as a
462 classification algorithm or to directly provide a similarity score between a suspect and a crime.

463 The proposed method was developed by using MATLAB 7.12.0 (R2011a) [26]. All the necessary
464 implementations were conducted manually using the MATLAB Script editor [27] apart from the
465 FIS which was implemented using the MATLAB fuzzy toolbox [28]. The five classification
466 algorithms which were used for the performance comparison were selected to be five of the
467 classification algorithms which are already packaged with the WEKA 3.6.12 tool [29].

468 **Results and Discussion**

469 The method was tested with a crime data set obtained from Sri Lanka Police. Figure 10 shows
470 the crime frequencies in Sri Lanka by the crime types from 2005 to 2011. It shows only 21 crime
471 types because the new 5 crime types were introduced in 2015. 4th column denoting House
472 Breaking and Theft shows the highest number of occurrences. 14: Theft of property, 10:
473 Robbery, 13: Cheating/ Misappropriation, 6: Hurt by Knife, 7: Homicide, 8: Rape/ Incest, 5:
474 Grievous Hurt, 3: Mischief over Rs. 5000/=, 1: Abduction/Kidnapping comes next. For the
475 validation of the algorithm, 7 crime types out of these 10 types were selected for the testing
476 data. They are, House Breaking and Theft, Theft of property, Robbery, Homicide, Rape/ Incest,
477 Grievous Hurt, Abduction/Kidnapping. 31 crime flows were selected which are common to the
478 seven selected crime types. The data set is also composed of 8 sub types and 2 special
479 categories. Altogether the data set consisted of 67 instances in which each instance is
480 composed of 48 attribute values. The data set is distributed over 20 classes (suspects).

481

482

483

484

485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511

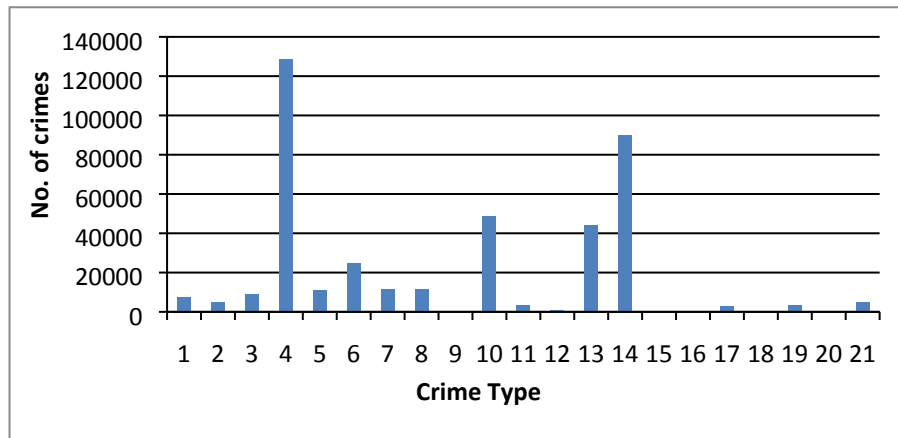


Figure 10. Frequency of different crime types from year 2005-2011

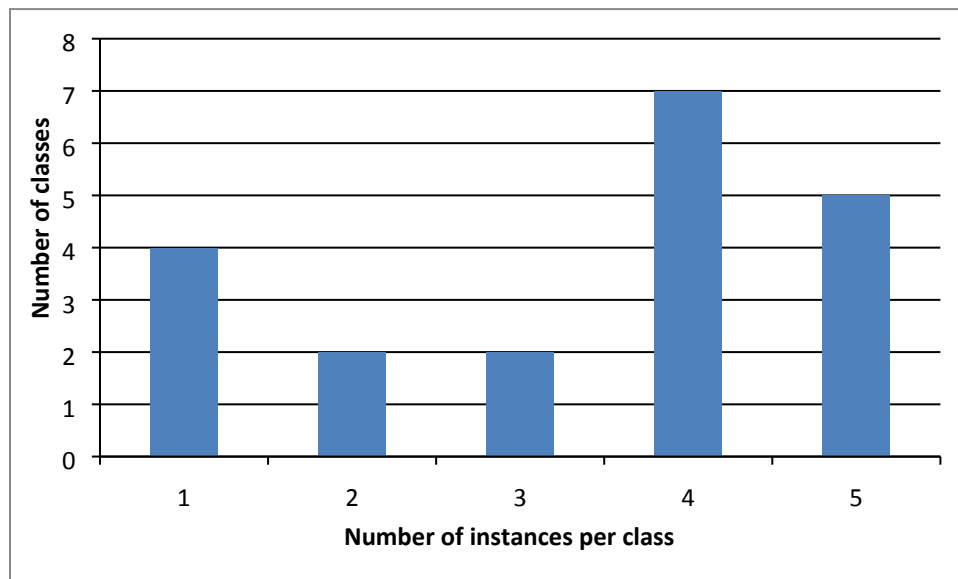
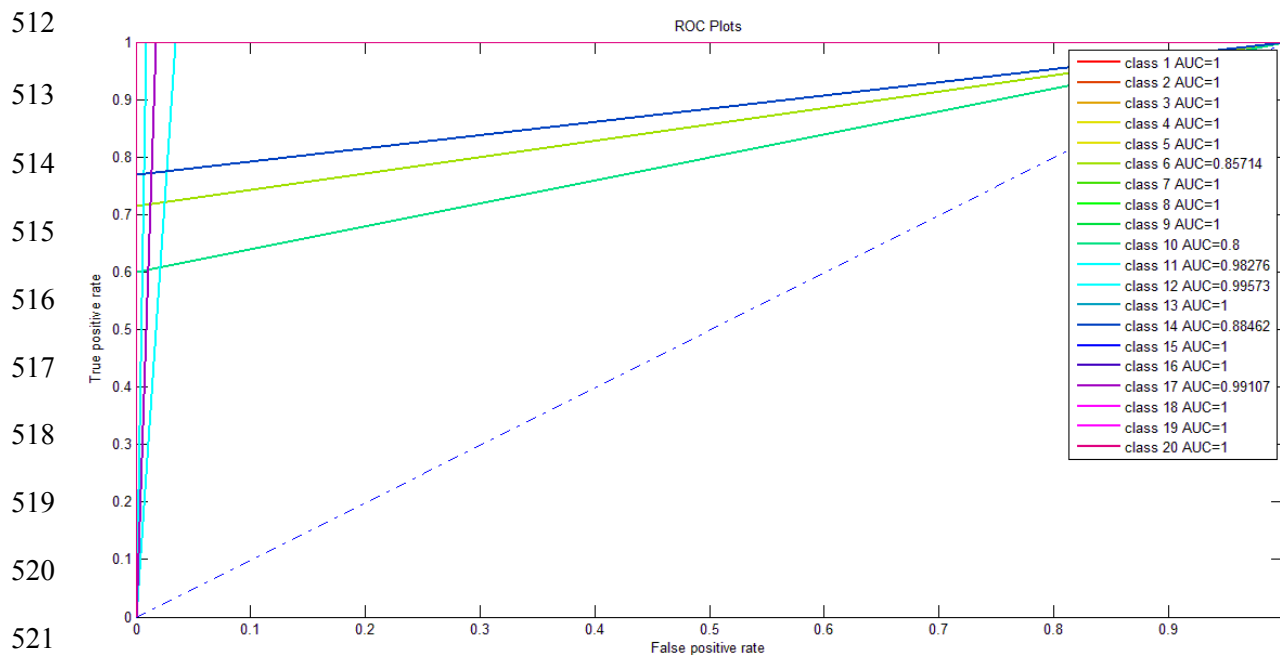


Figure 11. Distribution of modus operandi instances over the classes of the dataset

All the tests were performed in a Windows computer with Intel (R) Core (IM) i7-2670QM CPU of 2.20 GHz and a RAM of 8GB. The histogram of the instance distribution over the suspects (classes) is shown in Figure 11.

10 fold cross validation [30] was used on the data set for a fair testing procedure. In 10-fold cross validation, the data set is divided into 10 subsets, and the holdout method is repeated 10 times. Each time, one of the 10 subsets is used as the test set and the other 9 subsets are put together to form a training set. Then the average error across all 10 trials was computed [30].



522 **Figure 12.** ROC curves returned by the newly proposed method on the 20 classes of crime dataset

523 The test results of modus operandi classifications in Area Under Curve (AUC) [31], Root Mean
 524 Squared Error (RMSE) and time elapsed for the classification are shown in Table 6. A Receiver
 525 Operating Characteristic (ROC) curve is a two dimensional graphical illustration of the trade-off
 526 between the true positive rate (sensitivity) and false positive rate (1-specificity). Figure 12
 527 depicts the ROC curve plotted on the classification results obtained by the newly proposed
 528 method on the crime data set. In the particular instance which is shown in Figure 12, all the
 529 ROC curves related to the Crime data set are plotted well over the diagonal line and all of them
 530 have returned AUC values which are either equal to 1 or very close to 1, providing a very good
 531 classification.

532 The sole intention of this research was to find out relationships among the modus operandi of
 533 criminals with the modus operandi found in the crime scenes to find the associations. To
 534 prepare the data set which was used under this research, a crime data set of around 3000
 535 instances was considered. Due to limitations of the real crime data set, it was quite a complex
 536 task to prepare a data set with a collection of sufficient modus operandi where each instance
 537 happens to have a considerable flow of crime flows. Therefore, only a sample of 67 instances
 538 could be filtered from the population. As the number of instances was around 67, it can be
 539 assumed to be an under-represented data sample. Another reason for the data set to become
 540 under- represented was the challenge in finding classes/criminals with more than one crime
 541 committed. The actual crime data set which is used for the testing purposes is imbalanced as it
 542 is apparent in figure 11. For example the data set is composed of 4 single instance classes while

543 there are seven classes with four instances each. With a classification data set like this, there is
 544 a very high tendency of getting biased results.

545 Table 6 shows the results returned by the fuzzy based binary feature profiling which was
 546 conducted on the actual crime data set. As shown in the table, there is an increase in the
 547 accuracy when the input data set undergoes oversampling. Since the maximum number of
 548 instances available under one suspect is equal to 5, under-sampling does not provide a good
 549 accuracy. Oversampling and under-sampling are two concepts which are used in overcoming
 550 class imbalance problems in input data sets. Oversampling and under-sampling are two
 551 different categories of resampling approaches, where in oversampling the small classes are
 552 incorporated with repeated instances to make them reach a size close to larger classes, whereas
 553 in under-sampling, the number of instances is decreased in such a way that the number of
 554 instances reach a size close to the smaller classes [32]. The results prove that the new
 555 algorithm works well for a balanced data set as the new method is proved to have increased
 556 performance when the data set is subjected to an oversampling greater than or equal to 5
 557 which is the highest number of instances under a particular class.

558 **Table 6.** Results returned by the fuzzy based binary feature profiling for modus operandi analysis on
 559 actual data.

Data set (Number)	Oversampling or Under-sampling value	AUC	Root Mean Squared Error	Average time elapsed
1	2	0.5417	0.6986	0.0015
2	3	0.5562	0.4969	0.0011
3	4	0.5965	0.4303	0.0014
4	N/A	0.6937	0.7	0.001
5	5	0.6612	0.4126	0.0011
6	6	0.7063	0.2959	0.0011
7	10	0.8033	0.1396	0.0012
8	20	0.9339	0.0941	0.0013
9	30	0.9661	0.0853	0.0014
10	40	0.9637	0.0551	0.0015
11	50	0.9756	0.0578	0.0016
12	60	0.9626	0.0638	0.0018
13	70	0.9365	0.0792	0.0019
14	80	0.9391	0.0784	0.0023
15	90	0.9671	0.0752	0.0029

560 Figure 13 shows the change in AUC with the increase of sampling which starts from under-sampling of 2
 561 and goes on to an over sampling of 90. According to the plot it can be observed that the ROC values are
 562 increased when the oversampling is increased.

563

564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589

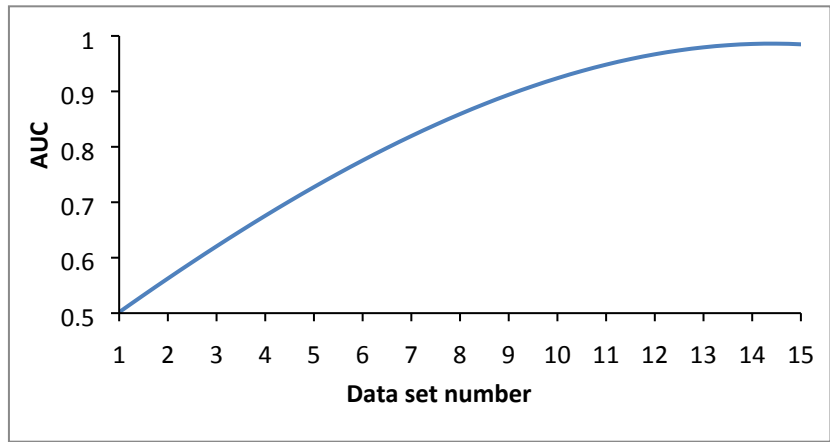


Figure 13. Change of ROC values with oversampling

Figure 14 depicts the change of RMSE with the increase of sampling. The plot clearly illustrates that the RMSE values are decreased with the oversampling.

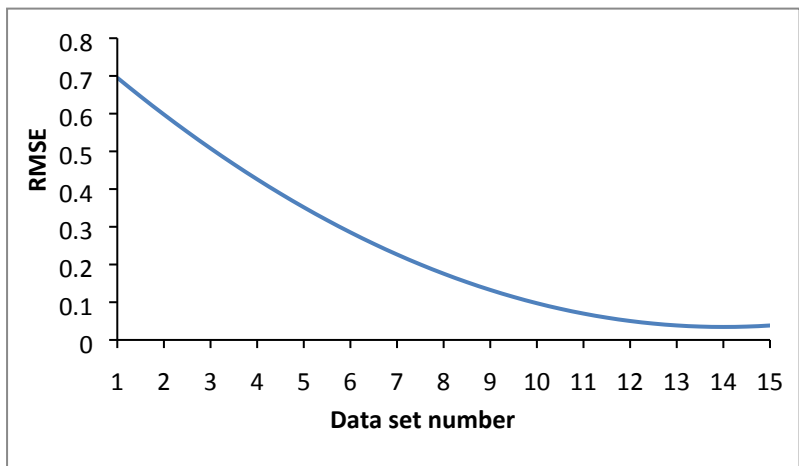


Figure 14. Change of RMSE value with oversampling

The execution time of the algorithm was 0.001s when there is no oversampling or under-sampling. The maximum execution time is 0.0031 when there is an oversampling of 90. According to the plot shown in Figure 15, it is clear that there is an increase of execution time as the oversampling size increases. Even though, the overall execution is under 3 milliseconds.

590
591
592
593
594
595
596
597
598
599

600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619

620
621

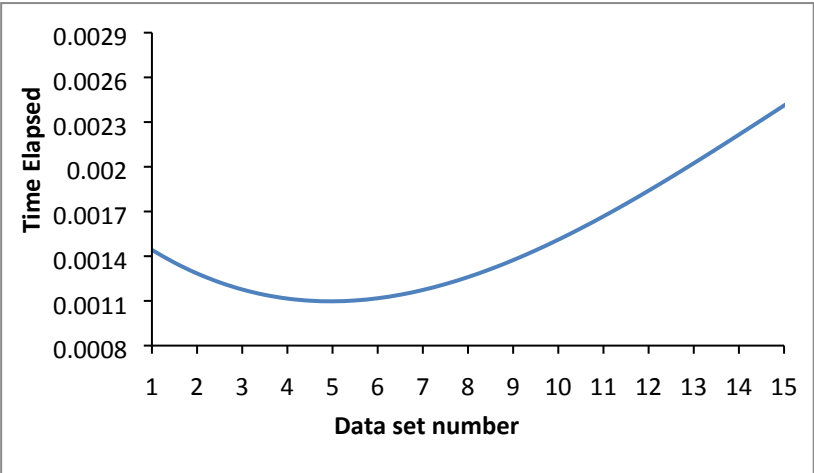


Figure 15. Change of time elapsed for the 15 data sets.

Overview of the classification algorithms used for the comparison

The newly proposed method was then tested against four other open classification data sets and the performance was evaluated against the results obtained with five well-known classification techniques, thereby assessing the quality of the newly proposed method. It is a known fact that there is no single algorithm which can be categorized as the best to solve any problem. Different classification algorithms may perform differently in different situations [33]. Therefore, five of the well-established classification algorithms were used for the performance evaluation of the newly proposed method. The five classification techniques used for this purpose include, Logistic Regression, J48 Decision Tree, Radial Basis Function Network, Multi-Layer Perceptron, Naive Bayes Classifier. Logistic Regression learns conditional probability distribution. Relating qualitative variables to other variables through a logistic cumulative distribution functional form is logistic regression [34]. J48 is an open source java implementation of the C4.5 decision tree algorithm [35]. A decision tree consists of internal nodes that specify tests on individual input variables or attributes that split the data into smaller subsets, and a series of leaf nodes assigning a class to each of the observations in the resulting segments. C4.5 algorithm constructs decision trees using the concept of information entropy [36]. Neural networks are flexible in being modeled virtually for any non-linear association between input variables and target variables [37]. Both Radial basis function networks and Multilayer perceptron (MLP) networks are neural networks [38]. Bayesian classifiers assign the most likely class to a given example described by its feature vector [39].

622 **Table 7.** *Description of the classification data sets for performance comparison*

Data set	Description	Number of Instances	No of Attributes
Dermatology Data Set [40]	This database has been created on a dermatology test carried out on skin samples which have been taken for the evaluation of 22 histopathological features. The values of the histopathological features have been determined by an analysis of the samples under a microscope. In the dataset constructed for this domain, the family history feature has the value 1 if any of these diseases has been observed in the family, and 0 otherwise. Every other feature (clinical and histopathological) was given a degree in the range of 0 to 3. Here, 0 indicates that the feature was not present, 3 indicates the largest amount possible, and 1, 2 indicate the relative intermediate values.	336	33
Balance Scale Data Set [41]	This data set has been generated to model psychological experimental results. Each example is classified as having the balance scale tip to the right, tip to the left, or be balanced. The attributes are the left weight, the left distance, the right weight, and the right distance. The correct way to find the class is the greater of (left-distance * left-weight) and (right-distance * right-weight). If they are equal, it is balanced. There are 3 classes (L,B,R), five levels of Left-Weight (1,2,3,4,5), five levels of Left-Distance (1,2,3,4,5), five levels of Right-weight (1,2,3,4,5) and five levels of Right-Distance (1,2,3,4,5).	625	4
Balloons Data Set [42]	This data set has been generated using an experiment of stretching a collection of balloons carried out on a group of adults and children [43]. In the data set, Inflated is true if (color=yellow and size = small) or (age=adult and act=stretch). In the data set there are two main output classes, namely T if inflated and F if not inflated, two colors yellow and purple, two sizes, large and small, two act types, stretch and dip, and two age groups, adult and child.	20	4
Car Evaluation Data Set [44]	Car Evaluation Database has been derived from a simple hierarchical decision model originally developed for the demonstration of DEX by M. Bohanec and V. Rajkovic [45]. The Car Evaluation Database contains examples with information that is directly related to CAR. They are buying, maint, doors, persons, lug_boot and safety. The attribute buying is the buying price which is considered to have four levels v-high, high, med, low. Maint is the price of the maintenance which contains the four levels, v-high, high, med, low. Doors have the four levels 2, 3, 4, 5-more. Person	1728	6

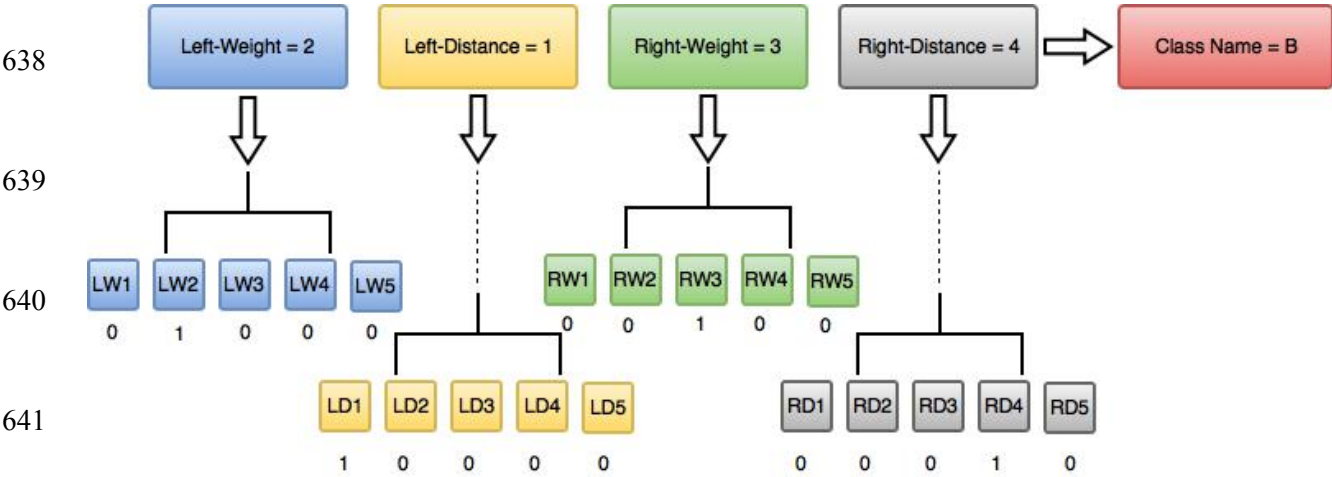
	(capacity in terms of persons to carry), lug_boot (the size of luggage boot) and safety (estimated safety of the car) have 3 levels each.		
--	---	--	--

623 As the newly proposed method accepts only binary input variables, the data sets which are
 624 used for the analysis must be preprocessed into the acceptable format. For example, the
 625 balance scale data set is composed of 4 attributes. Table 8 shows the attributes and their
 626 information of the balance scale data set.

627 **Table 8.** Attribute information of the balance data set

Attribute	Number of Categories	Categories
Class Name	3	L, B, R
Left-Weight	5	1,2,3,4,5
Left-Distance	5	1,2,3,4,5
Right-Weight	5	1,2,3,4,5
Right-Distance	5	1,2,3,4,5

628 Therefore, the data set was adjusted as shown in figure 16, prior to using it with the proposed
 629 method. Each category of a particular attribute is represented by a dummy variable. For
 630 example, Left-Weight attribute results in 5 attributes in the preprocessed data set and each
 631 attribute is represented using 5 binary variables as LW1, LW2, LW3, LW4 and LW5 where the
 632 presence of the attribute denotes 1 and 0 otherwise. As depicted in Figure 16, if Left-Weight
 633 has a value of 2 in an instance it results in 1 for the corresponding derived attribute that is LW2.
 634 Therefore, if there is an instance where Left-Weight=2, Left-Distance=1, Right-Weight=3 and
 635 Right-Distance=4, Class Name=B, it is represented as LW1=0,LW2=1, LW3=0, LW4=0, LW5=0,
 636 LD1=1, LD2=0, LD3=0, LD4=0, LD5=0, RW1=0, RW2=0, RW3=1, RW4=0, RW5=0, RD1=0, RD2=0,
 637 RD3=0, RD4=1, RD5=0, Class Name=B.



642 **Figure 16.** Schematic diagram for pre-processing of the balance dataset in such a way that it matches
 643 the format of inputs of the newly proposed method

644 The pre-processed data is then fed to the newly proposed algorithm and the five other
 645 algorithms. Performances were compared based on AUC, root mean squared error (RMSE) and
 646 the processing time for model generation. 10 fold cross validation was used under each test for
 647 fair testing procedure. Then the receiver operating characteristic curves of the results were
 648 analysed. Table 9 shows the average time elapsed for the model generation by each algorithm
 649 in seconds. For simplicity newly proposed modes operandi analysis algorithm has been
 650 acronymed as BFPM (Binary feature profiling methodology).

651 As all the data sets which were used for the tests are composed of multi classes, weighted
 652 average AUC was used, where each target class is weighted according to its prevalence as given
 653 in Equation 8. Weighted average was used in order to prevent target classes with smaller
 654 instance counts from adversely affecting the results [46].

$$655 \quad AUC_{weighted} = \sum_{\forall c_i \in C} AUC(c_i) \times p(c_i) \quad (8)$$

656 Table 9 shows the weighted average AUC values obtained for each data set under each
 657 classification algorithm.

658 **Table 9.** Weighted average AUC values obtained by the algorithms on classifying the data sets

	BFPM	Logistic Regression	J48	Radial Basis Function Network	Multi-Layer Perceptron	Naive Bayes Classifier
Dermatology data set	1	0.999	0.975	0.986	0.998	0.998
Balance scale data set	0.7945	0.976	0.811	0.968	0.977	0.971
Balloons Data Set	1	1	1	1	1	1
Car evaluation data set	0.8087	0.99	0.976	0.974	1	0.976

659

660 The Friedman's rank test results returned on the AUC test data which are shown in Table 10,
 661 indicates that the new model provides a better performance than J48 and RBF Networks for the
 662 4 data sets tested.

663

664 **Table 10.** *Friedman’s mean rank values returned on the data available in Table 9.*

Method	Mean Rank
BFPM	2.88
Logistic Regression	4.63
J48	2.50
RBF Networks	2.63
MLP	4.75
naïve Bayes Classifier	3.63

665
 666 The Friedman’s test returns a significance of 0.170 for the AUC values. This proves that there is
 667 no significant difference between the AUC values of the six methods.

668 Table 11 shows the RMSE (Root Mean Squared Error) values obtained on the 10-fold cross
 669 validation results obtained by each algorithm. Friedman’s rank test was conducted on the RMSE
 670 values obtained by the six algorithms. The test returned the results shown in Table 12.
 671 According to the mean rank values, the new model provides a better accuracy than J48, RBF
 672 Networks and Naive Bayes Classifier in the means of RMSE for the four data sets tested.

673 **Table 11.** *RMSE values returned by each algorithm on the classification of the four data sets*

	BFPM	Logistic Regression	J48	Radial Basis Function Network	Multi-Layer Perceptron	Naive Bayes Classifier
Dermatology data set	0.0881	0.0748	0.1163	0.1091	0.0764	0.0876
Balance scale data set	0.3491	0.2092	0.3699	0.2464	0.2055	0.2793
Balloons Data Set	0	0	0	0.001	0.0264	0.2385
Car evaluation data set	0.1766	0.152	0.1718	0.1983	0.044	0.2162

674

675

676

677

678 **Table 12.** Friedman’s rank test values returned on the data available in Table 11

Method	Mean Rank
BFPM	3.75
LogisticRegression	1.75
J48	4.25
RBFNetworks	4.25
MLP	2.25
NaiveBayesClassifier	4.75

679

680 The significance of 0.123 proves that there is no significant difference between the RMSE values
 681 returned by the six methods.

682 **Table 13.** Average processing time for each algorithm on the classification of the four data sets.

	BFPM	Logistic Regression	J48	Radial Basis Function Network	Multi-Layer Perceptron	Naive Bayes Classifier
Dermatology data set	0.0027	0.39	0.08	0.38	2.73	0.05
Balance scale data set	0.0030	0.03	0.08	0.22	0.48	0
Balloons Data Set	0	0	0	0	0.02	0
Car evaluation data set	0.0048	0.42	0.03	0.27	13.4	0.02

683

684 Friedman’s rank test run on the data set available in Table 13 returned the mean rank values
 685 which are shown in Table 14. Friedman’s rank test is a nonparametric test analogous to a
 686 standard one-way repeated-measures analysis of variance [47]. According to the mean rank
 687 values, new model has got the lowest mean rank, proving the conclusion that it is the most
 688 efficient method out of all the five other classification algorithms.

689 **Table 14.** Mean rank values returned by the Friedman's rank test on the time values available in Table
690 13.

Method	Mean Rank
BFPM	1.75
LogisticRegression	4.00
J48	3.25
RBFNetworks	4.00
MLP	6.00
NaiveBayesClassifier	2.00

691
692 For the time values returned by the six methods, the Friedman's statistical test returns 0.006
693 which in turn proves that there is a significant difference between the time elapsed by the six
694 methods.

695 The Wilcoxon Signed Ranks multiple comparison test returns P values greater than 0.05 proving
696 that the null hypothesis is true for all the three datasets, AUC, RMSE and time.

697 Friedman's rank test results for the three measurements, time elapsed, RMSE and AUC,
698 concludes that the newly proposed method provides some acceptable results against the five
699 well established classification algorithms.

700 **Conclusion**

701 The studies of modus operandi help crime investigation by letting the police officers link
702 criminals/suspects to crimes which are unresolved also to solve new crimes. Though there are
703 many descriptive studies available under modus operandi analysis, a very little amount of work
704 is available under computer science. Many of these methods have been derived using the
705 methods based on link analysis. But, the accuracy of these methods is always compromised due
706 to the cognitive biases of the criminals.

707 A novel Fuzzy based Binary Feature Profiling method (BFPM) to find associations between
708 crimes and criminals, using modus operandi is introduced. The newly proposed method
709 subjects not only the properties of the present, but also the properties of his/her previous
710 convictions. The concept of dynamic modus operandi which is available in the proposed
711 method considers all the modus operandi of his/her previous convictions to provide a fair
712 rectification to the errors which result due to the human cognition. Dynamic MO uses frequent
713 item set mining to result in a generalized binary feature vector. Complete MO profile also
714 encapsulates past modus operandi of a particular criminal by aggregating the modus operandi
715 of all of his/her previous convictions to one binary feature vector. This feature also guarantees
716 a usage of criminal's past crime record with more generalizability. Completeness probability

717 measures how much information is available in the new crime which is not available in the
718 complete MO profile. Therefore, this measurement provides the capability of measuring how
719 much extra amount of information is carried by the MO of the new crime. The deviation
720 probability provide a notion about how much the new MO deviates from the most frequent
721 crime flows which are available in the dynamic MO of a particular criminal. The vagueness and
722 the impreciseness prompted the fact that it is not possible to use crisp logic to generate the
723 similarity score. Therefore a fuzzy inference system was modeled to generate the similarity
724 score.

725 Due to the under-represented and imbalanced properties of the actual data set, the new
726 method has returned a lower performance when it is proposed to the data set without any pre-
727 processing on the data set. However, with the introduction of over sampling, the method
728 returns a very good performance, allowing one to arrive at the conclusion that the method
729 could provide acceptable results for a balanced data set. The method generated favorable
730 results in providing a good similarity measurement to suggest the connections between crimes
731 and criminals. Fuzzy controller of the new approach guarantees to resemble the human
732 reasoning process by confirming the usage of human operator knowledge to deal with
733 nonlinearity of the actual situation. The newly proposed method was then adapted into a
734 classification algorithm for the validation and comparison with other classification algorithms.
735 The comparison of the new method with the well-established classification algorithms
736 confirmed the generalizability of the new method. A noticeable feature of the newly proposed
737 method, over the other classification algorithms was the very low amount of time elapsed for
738 the model generation. Compared to other algorithms, the new method consumes the least
739 amount of time. While the new method fluctuates above and below the performance of other
740 classification algorithms, it has showed always a performance greater than J48 and RBF
741 Networks for the data sets which were used for the tests.

742 The method only provides the capability to process the categorical data sets. If there are any
743 continuous variables in the data set, the values must be preprocessed into categories before
744 further processing. The method can be further extended to directly accept the continuous
745 attributes. As the center of gravity method is used for the defuzzification process, further
746 optimizations can be done by simplifying the defuzzification procedure. Adapting the fuzzy
747 inference engine to a Sugeno [48] type and converting the defuzzification method to a more
748 computationally efficient method such as the weighted average [49] method would provide a
749 less complex computation. This would result in even less processing time when the
750 sophistication of the data set rises.

751

752

753

754

References

- [1] Holdaway, S., *Issues in Sociology: Crime and Deviance.*: Nelson Thornes Ltd, 1993.
- [2] Chisum, W.J. , Turvey, B., "Evidence dynamics: Locard's exchange principle & crime reconstruction.," *Journal of Behavioral Profiling*, vol. 1, no. 1, pp. 1-15, 2000.
- [3] Paternoster, R., Bachman, R., *Explaining Criminals and Crime: Essays in Contemporary Criminological Theory*, Ronet Bachman Raymond Paternoster, Ed.: Roxbury Publishing Company, 2001.
- [4] Palmiotto, M.J., "Crime Pattern Analysis: An Investigative Tool," *Critical Issues in Criminal Investigation*, vol. 2, pp. 59-69, 1988.
- [5] Douglas, J. E., Munn, C., "Modus operandi and the signature aspects of violent crime," in *Crime Classificatio Manual*, 2nd ed.: John Wiley & Sons, 2006, pp. 19-30.
- [6] Chamikara, M.A.P., Galappaththi, A., Yapa, Y.P.R.D., Nawarathna, R.D., Kodituwakku, S.R., Gunathilake, J., Liyanage, L.H., "A Crime Data Analysis Framework with Geographical Information Support for Intelligence Led Policing," *Manuscript submitted to PeerJ*, July 2015.
- [7] The 'Lectric Law Library. The 'Lectric Law Library. [Online]. HYPERLINK "http://www.lectlaw.com/files/int20.htm" <http://www.lectlaw.com/files/int20.htm>
- [8] Bennell, C., Canter, D. V., "Linking commercial burglaries by modus operandi: Tests using regression and ROC analysis," *Science & Justice*, vol. 42, no. 3, pp. 153-164, 2002.
- [9] Bennell, C., Jones, N. J., "Between a ROC and a hard place: A method for linking serial burglaries by modus operandi," *Journal of Investigative Psychology and Offender Profiling*, vol. 2, no. 1, pp. 23-41, 2005.
- [10] Leclerc, B., Proulx, J., Beauregard, E., "Examining the modus operandi of sexual offenders against children and its practical implications," *Aggression and violent behavior*, vol. 14, no. 1, pp. 5-12, 2009.
- [11] Oatley, G., Ewart, B., "Data mining and crime analysis," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 2, pp. 147-153, 2011.
- [12] (2011) Offender Profiling. [Online]. HYPERLINK "http://www.liv.ac.uk/psychology/ccir/op.html" <http://www.liv.ac.uk/psychology/ccir/op.html>
- [13] Agrawal, R., Imielinski, J., Swami, A., "Mining Association rule between sets of items in large databases," in *Proceedings of the ACM SIGMOD International Conference of Management of Data*, New York, 1993, pp. 207-216.

- [14] Koperski, K., Han, J., "Discovery of spatial association rules in geographic information databases," in *Proceeding of the 4th International Symposium on Spatial Databases*, 1995, pp. 47-67.
- [15] Chen, H., *Intelligence and security informatics for international security*, 1st ed.: Springer US, 2006, vol. 10.
- [16] Lin, S., & Brown, D. E., "An outlier-based data association method for linking criminal incidents," *Decision Support Systems*, vol. 41, no. 3, pp. 604-615, 2006.
- [17] Berry, M. J., Linoff, G., *Data Mining Techniques: For Marketing, Sales, and Customer Support*, 3rd ed.: Wiley, 2011.
- [18] Chen, H., Zeng, D., Atabakhsh, H., Wyzga, W., & Schroeder, J., "COPLINK : Managing Law Enforcement Data and Knowledge," *Communications of the ACM*, vol. 46, no. 1, 2003.
- [19] Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., Chau, M., "Crime Data Mining: A general framework and some examples.," *Computer*, vol. 37, no. 0018-9162, pp. 50-56, April 2014.
- [20] Chen, H., "Machine learning for information retrieval: neural," *Journal of the American Society for Information Science*, vol. 46, no. 3, pp. 194-216, 1995.
- [21] Adamo, J. M., *Data mining for association rules and sequential patterns, Sequential and Parallel Algorithms*, 1st ed. New York: Springer Science & Business Media, 2001.
- [22] Mamdani, E.H., Assilina, S., "An experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Man-Machine Studies*, vol. 7, no. 1, pp. 1-13, 1975.
- [23] Godjevac, J., *Neuro-fuzzy Controllers: Design and Application.*: PPUR presses polytechniques, 1997.
- [24] Zadeh, L.A., "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets and Systems*, vol. 90, pp. 117-117, 1997.
- [25] MathWorks, Inc. (1994-2015) MathWorks. [Online]. HYPERLINK
"http://in.mathworks.com/help/fuzzy/foundations-of-fuzzy-logic.html"
<http://in.mathworks.com/help/fuzzy/foundations-of-fuzzy-logic.html>
- [26] MathWorks. (1994-2015) MathWorks. [Online]. HYPERLINK "https://in.mathworks.com/"
<https://in.mathworks.com/>
- [27] MathWorks. (1994-2015) MathWorks Documentation. [Online]. HYPERLINK
"http://in.mathworks.com/help/matlab/ref/edit.html"
<http://in.mathworks.com/help/matlab/ref/edit.html>
- [28] MathWorks. (1994-2015) MathWorks Fuzzy Logic Toolbox. [Online]. HYPERLINK
"http://in.mathworks.com/help/matlab/ref/edit.html"

<http://in.mathworks.com/help/matlab/ref/edit.html>

- [29] Machine Learning Group at the University of Waikato. WEKA. [Online]. HYPERLINK "http://www.cs.waikato.ac.nz/ml/weka/" <http://www.cs.waikato.ac.nz/ml/weka/>
- [30] Refaeilzadeh, P., Tang, L., Liu, H., "Cross-validation," in *Encyclopedia of database systems*, M. Tamer Özsu Ling Liu, Ed.: Springer US, 2009, pp. 532-538.
- [31] Hanley, J. A., McNeil, B. J., "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29-36, 1982.
- [32] Estabrooks, A., Jo, T., Japkowicz, N., "A multiple resampling method for learning from imbalanced data sets," *Computational Intelligence*, vol. 20, no. 1, pp. 18-36, 2004.
- [33] Wolpert, D.H., "The Lack of a Priori Distinctions Between Learning Algorithms," *Neural Computation*, vol. 8, pp. 1341-1390, 1996.
- [34] Chang, Y. C. I., Lin, S. C., "Synergy of Logistic Regression and Support Vector Machine in Multiple-class Classification," in *Intelligent Data Engineering and Automated Learning - IDEAL 2004*, vol. 5, Exeter, UK, 2004, pp. 132-141.
- [35] Machine Learning Group at the University of Waikato. Class J48. [Online]. HYPERLINK "http://weka.sourceforge.net/doc.dev/weka/classifiers/trees/J48.html" <http://weka.sourceforge.net/doc.dev/weka/classifiers/trees/J48.html>
- [36] J.R. Quinlan, "C4.5 programs for machine learning," *Machine Learning*, vol. 16, no. 3, pp. 235-240, 1993.
- [37] Bishop, C.M., *Neural networks for pattern recognition*. Oxford, UK: Oxford University Press, 1995.
- [38] Jayawardena, A. W., Fernando, D. A. K., Zhou, M.C., "Comparison of Multilayer Peceptron and Radial Basis Function Networks as Tools for Flood Forecasting," in *Destructive Water: Water- Caused Natural Disasters, their Abatement and Control*, California, 1996, pp. 173-181.
- [39] Rish, I., "An empirical study of the naive Bayes classifier," , vol. 3, New York, 2001, pp. 41-46.
- [40] Ilter, N., Guvenir, H.A. (1998) UCI Machine Learning Repository. [Online]. HYPERLINK "https://archive.ics.uci.edu/ml/datasets/Dermatology" <https://archive.ics.uci.edu/ml/datasets/Dermatology>
- [41] Hume, T. (1994) UCI Machine Learning Repository. [Online]. HYPERLINK "https://archive.ics.uci.edu/ml/datasets/Balance+Scale" <https://archive.ics.uci.edu/ml/datasets/Balance+Scale>

- [42] Pazzani, M. (1991) UCI Machine Learning Repository. [Online]. HYPERLINK
"https://archive.ics.uci.edu/ml/datasets/Balloons" <https://archive.ics.uci.edu/ml/datasets/Balloons>
- [43] M. Pazzani, "The influence of prior knowledge on concept acquisition: Experimental and computational results," *Journal of Experimental Psychology: Learning, Memory & Cognition*, vol. 17, no. 3, pp. 416-432, 1991.
- [44] Bohanec, M., Zupan, B. (1997) UCI Machine Learning Repository. [Online]. HYPERLINK
"https://archive.ics.uci.edu/ml/datasets/Car+Evaluation"
<https://archive.ics.uci.edu/ml/datasets/Car+Evaluation>
- [45] Bohanec, M., Rajkovic, V., "Expert system for decision making," *Sistemica*, vol. 1, no. 1, pp. 145-157, 1990.
- [46] Hempstalk, K., Frank, E., "Discriminating Against New Classes: One-class versus Multi-class Classification," in *AI 2008: Advances in Artificial Intelligence: 21st Australasian Joint Conference on Artificial Intelligence*, Auckland, 2008, pp. 325-336.
- [47] Howell, D.C., *Fundamental Statistics For The Behavioral Sciences Focuses*, 8th ed. Belmont: Wadsworth, Cengage Learning, 2013.
- [48] Takagi, T., Sugeno, M., "Fuzzy identification of systems and its applications to modeling and control," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 1, pp. 116-132, 1985.
- [49] Wu, D., Mendel, J. M., "Aggregation using the linguistic weighted average and interval type-2 fuzzy sets," *Fuzzy Systems, IEEE Transactions on*, vol. 15, no. 6, pp. 1145-1161.