

## Decision tree-based expert system for adverse drug reaction detection using fuzzy logic and genetic algorithm

Ayman M. Mansour\*

Department of Communication, Electronics and Computer Engineering, Tafila Technical University, Tafila, Jordan

Received: 07-January-2018; Revised: 01-February-2018; Accepted: 26-April-2018

©2018 ACCENTS

### Abstract

*Early detection of unknown adverse drug reactions (ADRs) could save patient lives and prevent unnecessary hospitalizations. Current surveillance systems are not ideal for rapidly identifying rare unknown ADRs. Current methods largely rely on passive spontaneous reports, which suffer from serious underreporting, latency, and inconsistent reporting. A more effective system is needed as the electronic patient records become more and more easily accessible in various health organizations such as hospitals, medical centers and insurance companies. These data provide a new source of information that has great potentials to detect ADR signals much earlier. In this paper, we have developed a methodology that uses both decision tree and fuzzy logic to generate a decision model. The developed model is equipped with a fuzzy inference engine, which enables it to find the causal relationship between a drug and a potential ADR. This could assist healthcare professionals to early detect previously unknown ADRs. Optimizing fuzzy rule weights and fuzzy sets parameters using genetic algorithm has been embedded in the proposed system to achieve excellent performance and improve the accuracy of the developed model. To evaluate the performance of the system, we have implemented the system using Weka and FuzzyJess software packages, and generated simulation results. To conduct the experiments, clinical information on 280 patients treated at the Detroit Veterans Affairs Medical Center was used. Two physicians on the team independently reviewed the experiment results. Kappa statistics show excellent agreement between the physicians and the developed model.*

### Keywords

*Adverse drug reactions, Fuzzy logic, Decision tree, Genetic algorithm.*

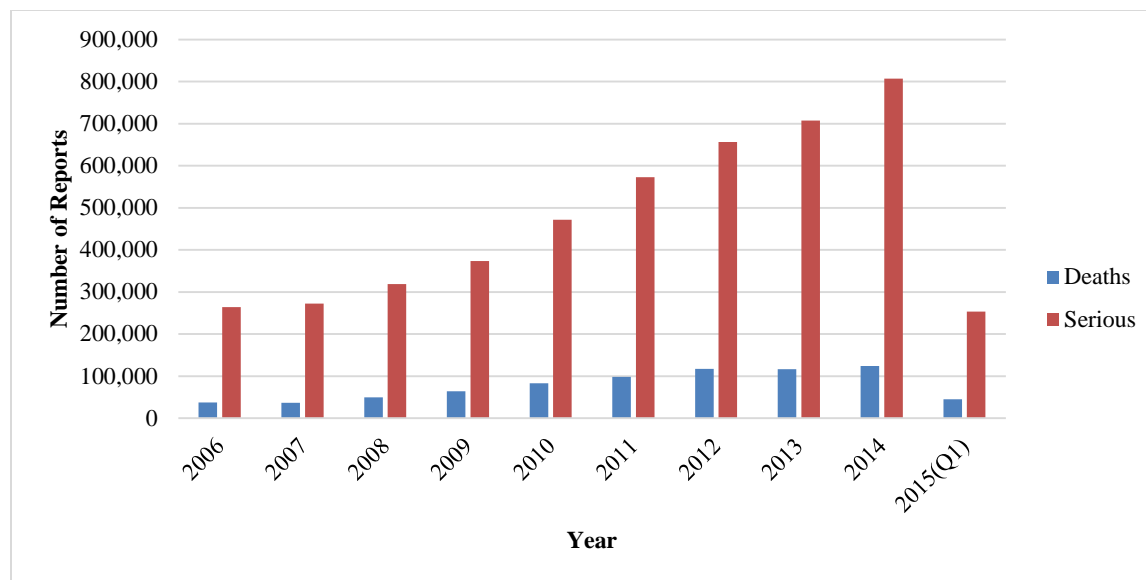
### 1.Introduction

ADR was defined as “An appreciably harmful or unpleasant reaction. It results from an intervention related to the use of a medicinal product. Adverse effects usually predict hazard from future administration and warrant prevention, or specific treatment. It contributes to alternate the dosage regimen or withdrawal of the product” [1]. ADRs are a major public health problem in the USA [1, 2]. In 2014, for instance, adverse events reporting system which is managed by the USA that is food and drug administration (FDA) showed that 123,927 deaths were attributed to serious ADR and 807,270 serious cases were reported, which included, among others, hospitalization, life-threatening, and/or disability [2]. Figure 1 [2] illustrates the patient outcome(s) for ADRs since the year 2006 until the first quarter of 2015.

A complete understanding of the safe use of drugs is not possible at the time when the drug is developed or marketed. At that time, the safety information is only limited to a few thousand people in a typical clinical trials. For example, people were not aware of the risk of heart attacks associated with the use of rofecoxib until five years later after it was launched to the market.

Detecting ADR signal pairs is technically a complex problem [3-10]. This is the case if we realistically assume that there does not exist a set of rules that are readily acceptable to all human experts (e.g., physicians, epidemiologists and pharmacists). The parameters used in identifying the signal pairs are really a vague, subjective measure rather than an objective measure. Furthermore, human experts often disagree one another owing to their knowledge and experiences and there is no “ground truth” to indicate which physician is right or wrong.

\*Author for correspondence



**Figure 1** Patient outcome(s) for ADRs since the year 2006 until the first quarter of 2015 [2]

Given the limited information available when the drug is marketed, post-marketing study has become increasingly important. Post-marketing surveillance is the process of identifying, reporting, and responding to the issues occurred while taking medication [11-18]. This method is the principal method used for monitoring the safety of marketed drugs nowadays [19]. The responding includes actions that can be taken to improve product safety and protect the public health, such as labeling changes, safety alerts or product withdrawals [3, 12]. Even if the report does suggest labeling changes, the information provided will be kept for further investigated especially when more information became available. Once the reports are studied and evaluated, the data generated can help to identify ADR with certain medications and investigate these ADRs to provide clear indicators that can be used to identify other ADR resulted from other medications. To date, many methods have been adopted into post-marketing studies, including ADR case review, comparative observational study, ADRs spontaneous reporting, and data mining algorithms.

The most important source of adverse event information to the FDA in USA is the MedWatch program [4]. In this program, health care providers and patients can submit an adverse event report via several mechanisms, including an online report form, phone, fax and mail. Detection of signals in MedWatch is limited due to the low percentage of adverse events which are rare events (less than 10%). The quality of the reports and the accuracy of the

information written in the reports can also affect the detection performance of the adverse events. Some time, for an instant, the health providers don't appreciate clinical finding or adverse events until the occurrence became wider in the patient population. Also, we need to remember that MedWatch reporting is voluntary job and many suspected adverse events are not reported based on the physician's point of view [9, 20].

Other spontaneous reporting systems, such as yellow card in the UK, suffer from low reporting rates, typically less than 10%. Underreporting of ADRs is a common problem in spontaneous surveillance programs which can delay adverse event detection and underestimate a problem size [13-18]. Knowing the factors that may confirm an adverse event will assist health providers and physicians in establishing ways to correct underreporting. Different reasons that need to be analyzed in order to improve the quality of reporting [21]. There are huge numbers of reports that are available in the databases. This makes it difficult for physicians to analyze the reports of adverse reactions. They have little time especially most of the reported events have a low likelihood of a causal relation.

Several data mining algorithms have been well described in the literature; some are based on simple analysis, e.g., the reporting odds ratios and the proportional reporting ratios [22, 23].

Data mining algorithms are being used to explore spontaneous reporting databases for adverse reaction signal pairs. The finding of data mining algorithms has been compared with those comes from classical reporting methods in [5]. Most adverse events identified by both methods highlight in the product labeling. Classical reporting methods identify four potentially unexpected serious adverse events which may lead to label changing and close monitoring. The other finding of that paper that none of these adverse events has been identified using the data mining algorithms. This is not helpful since it cannot detect or enhance the classical reporting methods, surveillance in this particular setting. Data mining algorithm's performance may be enhanced by selecting the most appropriate Pharmacovigilance tools that are designed specifically for each situation [24]. Also in [25] five data mining algorithms are used for identifying possible ADRs from spontaneous reports information. The study concludes that the detected drug-ADR signal pairs vary between different methodologies and this makes the data mining algorithms unreliable.

The availability of huge amounts of reported events, including false-positive signals as a result of the existence of confounding will produce unhelpful hypotheses. This may affect the capability of data mining algorithms to detect true positive signals of real causal associations. This will lead to serious consequences that delay the detection of the signal pairs.

Data mining algorithms had not been generally accepted by health providers, physicians or pharmacists. This is due to the three main reasons which are as follows [3]:

1. The apparent complexity of its mathematics and the hidden strategy of detecting the signal pairs deter those unfamiliar with statistics.
2. Even with the increasing availability of epidemiological and pharmacoepidemiological databases, background information for calculating prior probabilities is still either unclear or unavailable.
3. The pre-market data from clinical trials is usually not available for the public in order to be used for estimating the prior probability. The data has been kept confidential between the drug companies and regulatory authorities. Even when they are available, they may not be in a suitable format.

Since no consensus exists regarding the use of data mining algorithms, the use of such methods in

Pharmacovigilance is still limited and not embraced by health providers. Data mining algorithm may draw attention to more "surprising" drug-event signal pairs. In fact, the unknown features of the data (i.e., the event background incidences) and the underreporting problem of the events from health providers to the regulatory authorities will have direct effect the outcome of data mining algorithms [26].

A decision-tree based model is an effective supervised technique to implement the classification methods in high-dimensional data has been developed in this paper [27]. The developed system is equipped with an intelligent decision maker that uses a fuzzy rule-based reasoning capability. Fuzzy logic is used to represent, interpret, and compute vague and/or subjective information which is very common in medicine [28-30]. Fuzzy logic is commonly accepted by physicians more than other intelligent techniques because the fuzzy rules are extracted from domain experts, i.e. physicians. The fuzzy inference system is implemented using the freeware FuzzyJess [31].

The developed system is equipped with a fuzzy inference engine in order to be able to find a causal relationship between a drug and an adverse reaction. The parameters used in identifying the signal pairs are really a vague, subjective measure rather than an objective measure. Up to our knowledge, no set of rules was clearly mentioned in the literature that took advantage of this vagueness. Furthermore, physicians often disagree one another owing to their knowledge and experience levels and usually there is no "ground truth" to indicate which physician is right or wrong. The developed fuzzy inference engine also uses the real patient databases. Such databases have not been used before in the literature for such purposes. Databases in current systems are mainly used to know the medical history of patients by only their physicians, improving the quality of provided services and reducing the costs of medical errors. These databases provide valuable information about patients, including age, sex, medication took by the patients, symptoms appears in patients, laboratory test results, and procedure followed by the physician at the visit time. The proposed ADR signal pairs detection methodology is based on five cues: temporal association, rechallenge, dechallenge, abnormality in laboratory tests and other explanation. The cues represent the higher-level information that is obtained from the patients' elementary data. The detection rules that use the above cues were acquired

through the joint efforts of the engineering and medical team members.

Optimizing fuzzy rule weights and fuzzy sets parameters using genetic algorithm (GA) have been used in this paper to improve the initial performance of decision trees. The final decision tree performance has been enhanced with the achieved optimized fuzzy rule weights and fuzzy sets parameters. Using Waikato environment for knowledge analysis (WEKA) software [32], the collected and pre-processed training set are randomly chosen from real patients' dataset, and are used to build the decision tree model. After that, the developed model is tested on another portion of the data unseen in the training stage for model evaluation.

The rest of the paper is organized in the following manner: section 2 explains the framework of the proposed method. Results have been presented in section 3. Section 4 shows the result discussion and analysis. Finally the concluding remarks were given in section 5.

## 2. Materials and methods

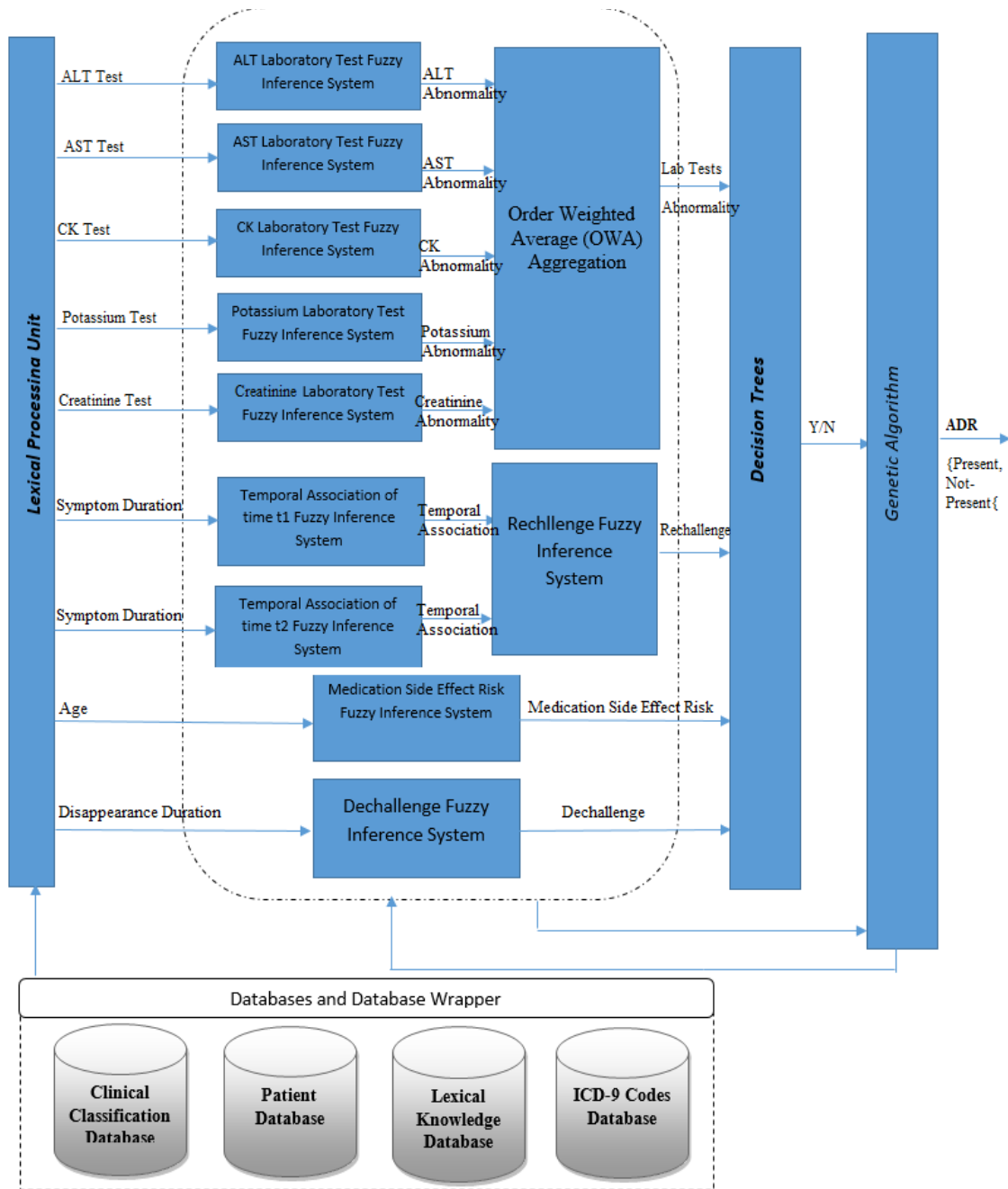
Figure 2 shows decision tree-based ADRs detection system. ADRs detection is characterized by a number of factors related to physician's experience.

In this paper ADRs detection is based on temporal association, dechallenge and rechallenge of a medication of interest, abnormality of a laboratory test, medication side effect risk, medications, and morbidities. To get some of those cues, we used the international classification of diseases, ninth edition, clinical modification (ICD-9) and physicians' current procedural terminology codes (CPT). These are two widely used coding standards in the USA. Using these codes, every clinical condition of the patients, symptoms that appear one patient and any treatment required or done by a health provider has a unique code. The ICD-9 code provides codes to classify diseases and a wide variety of signs, symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or disease. Every health condition is assigned a unique category and given a unique code. For example, if a patient is diagnosed with Hepatitis C, he/she will be given the ICD-9 code "070.51". If the diagnosis is for something acute, something that goes away with treatment like a rash

or the flu, then the ICD-9 code will be less important because the illness or condition will go away. However, if the patient is diagnosed with a chronic or lifelong problem, like heart disease or diabetes, the ICD-9 code will be more important and will affect his future medical care.

Since different ICD-9 codes may represent the same (or similar) diagnoses, we clustered the ICD-9 codes into a manageable number of categories. This was done based on the clinical classification system (CCS) for the ICD-9 fact sheet. CCS is developed by the Agency for Healthcare Research and Quality (AHRQ) [33]. The CCS can group over 13,600 ICD-9 codes into 285 mutually exclusive and clinically meaningful categories. The clinical classification system makes it easier to quickly understand patterns of diagnoses and procedures. The physicians can easily analyze patient cases using CCS and assign them labels.

Temporal association describes the symptom duration which is the time between taking a drug and the appearance of symptoms. It should be noted that in the case of a potential ADR, exposure to a drug should always precede the effect (symptom). This distinction is important because the effect might result from entirely different causes (e.g., underlying diseases or reception of another medication). Dechallenge refers to the relationship between the time of discontinuity of a drug and abatement of a symptom which is disappearance duration. We cannot directly evaluate dechallenge of a pair since the drug stop date is usually unavailable in electronic health databases. However, we can indirectly assess the existence of dechallenge of a pair if a symptom occurs after the drug start date and another drug in the same class was prescribed after the appearance of the symptom. This is because the physicians often stop a drug and prescribe another drug in the same class to avoid the apparent adverse effect found in a patient. Rechallenge depicts the relationship between re-introduction of the drug discontinued before and recurrence of a symptom. Rechallenge is determined by the temporal associations of the two consecutive occurrences of the same pair one after taking the medication and the other one after the reintroduction of the medication.



**Figure 2** General architecture of the developed system

Functionally speaking, the proposed system architecture consists of four components, namely fuzzy inference system unit, decision tree unit, GA unit and database wrapper. The proposed system algorithm is shown in *Figure 3*.

Abnormality of a laboratory test shows the degree of elevation of a laboratory test result. The abnormality will be calculated for transaminases aspartate

aminotransferase (AST), alanine aminotransferase (ALT), creatine kinase (CK), potassium, and creatinine, which are common tests. Medication side effect risk gives the side effect risk degree that could be on a patient based on his/her age [34]. The side effect risk of a medication in a healthy adult is not that dangerous as in elderly patients. Medication factor shows the overall view of medications took by the patients, including medication name, class and quantity.

**Training Phase:**

Retrieve training patients' demographic data (i.e. laboratory test results, age, medications, etc.)  
 Extract medical cues using fuzzy inference systems which represent the higher level information (dechallenge, rechallenge, medication side effect risk, Laboratory test Abnormality)  
 Classify medical cues as present or not present ADR using decision tree  
 Optimize fuzzy rules weights and fuzzy sets parameters using genetic algorithm and update the constructed decision tree  
 End by returning the optimized fuzzy rules weights and fuzzy sets parameters

**Testing Phase:**

Retrieve Testing patients' demographic data (i.e. laboratory test results, age, medications, etc.)  
 Extract medical cues using fuzzy inference systems which represent the higher level information (dechallenge, rechallenge, medication side effect risk, Laboratory test abnormality)  
 Classify medical cues as present or not present ADR using decision tree  
 End by labelling the medical cases as present or not present ADR

**Figure 3** Proposed system algorithm

The morbidity factor presents all the symptoms appear in a patient and have been verified by his physician. The symptoms are recorded at the time of physician visit as ICD-9 codes.

The medication plays a key role in identifying ADR since an adverse event can change by changing the medication. Sometimes the resulted adverse events are pleasant as in the case of some out of counter medications such as pain relievers and fever reducers. However, sometimes, the resulted adverse events are dangerous and can cause a great harm to the human body. In this paper a strategy of classifying drugs available in the market is used and has been embedded in our framework. The medications were catalogued according to the anatomical therapeutic chemical (ATC) classification [35]. This system is recommended by the WHO for drug utilization studies.

In the ATC classification system, the active substances are divided into different groups according to the organ or the system on which they act and their therapeutic, pharmacological and chemical properties. For example, in the ATC system medications captopril and enalapril, which are angiotensin converting enzyme (ACE) inhibitor medications are given the code C09AA.

**2.1 Fuzzy inference system unit**

Based on the above-mentioned factors, we defined input variables of Fuzzy Inference Engines, as shown in *Table 1*. Each variable is fuzzified by input fuzzy sets whose names, types, and parameters are specified in *Table 1*. The mathematical type definitions are given in *Table 2*. Likewise, we define output variables and their fuzzy sets for the engine (*Table 3*). The linguistic concepts, most of which are

inherently vague, are represented and manipulated by fuzzy sets using the theoretical tools provided by fuzzy set theory. They enable us to express and deal with various relations and functions that involve linguistic concepts. The construction of fuzzy sets involves a specific knowledge domain of interest, two experts in the adverse drug reaction domain and a knowledge engineer. The knowledge of, interest is elicited from the experts by the knowledge engineer. In the first stage the knowledge engineer attempts to elicit the knowledge in terms of propositions expressed in natural languages. In the second stage, he attempts to determine the meaning of each linguistic term employed in these propositions. During the second stage the functions representing fuzzy sets and operations are constructed. A direct method of construction has been used in this research. In direct methods, experts are expected to give answers to questions of various kinds that explicitly lead to construction of membership functions.

Fuzzy inference engine is equipped with the fuzzy rules that link the input variables to the output variables (*Table 4*). The initial rules and formulas are provided by the physicians on the team. There is a total of 40 rules developed by the physicians and is used in this paper. In this paper, we assign a confidence level, a value in  $[0, 1]$ , to each rule based on the experience of the agent in that rule. The confidence levels affect the contribution of the rules which consequently affects the calculation of the output of the fuzzy inference engines. The inferred output of each rule is scaled by the confidence level value via algebraic product before aggregating the output of individual rules. A reference patient case is evaluated by the proposed system using the rules stored in the local patient database.

**Table 1** Input variables and their fuzzy set names, types and parameters

Output variable	Fuzzy set name	Fuzzy set type	Fuzzy set parameters		
			a	b	c
Temporal association	Unlikely	Triangular	-	0	0.5
	Possible		0	0.5	1
	Likely		0.5	1	-
Dechallenge	Unlikely	Triangular	-	0	0.5
	Possible		0	0.5	1
	Likely		0.5	1	-
Rechallenge	Unlikely	Triangular	-	0	0.5
	Possible		0	0.5	1
	Likely		0.5	1	-
AST abnormality	Very Low	Triangular	-	0	0.25
	Low		0	0.25	0.5
	Medium		0.25	0.50	0.75
	High		0.50	0.75	1
	Very High		0.75	1	-
ALT abnormality	Very Low	Bell	0	0.13	3
	Low		0.25	0.13	3
	Medium		0.5	0.13	3
	High		0.75	0.13	3
	Very High		1	0.13	3
CK abnormality	Very Low	Bell	0	0.17	3
	Low		0.33	0.17	3
	High		0.67	0.17	3
	Very High		1	0.17	3
Potassium abnormality	Low	Gaussian	0	0.21	-
	Medium		0.5	0.21	-
	High		1	0.21	-
Creatinine abnormality	Low	Bell	0	0.5	2.5
	High		1	0.5	2.5
Medication side effect risk	Very Low	Gaussian	0	0.14	-
	Low		0.33	0.14	-
	High		0.67	0.14	-
	Very High		1	0.14	-

**Table 2** Definitions of the fuzzy sets

Fuzzy set type	Fuzzy set definition
Triangular	$\mu_T(x) = \begin{cases} -\frac{1}{a-b}(a-x), & a \leq x \leq b \\ \frac{1}{c-b}(c-x), & b \leq x \leq c \\ 0, & \text{otherwise} \end{cases}$
Bell	$\mu_B(x) = \frac{1}{1 + \left \frac{a-x}{b}\right ^{2c}}, \quad c > 0$
Gaussian	$\mu_G(x) = e^{-\frac{(a-x)^2}{2b^2}}$

**Table 3** Output variables and their fuzzy set names, types and parameters

Input variable	Fuzzy set name	Fuzzy set type	Fuzzy set parameters		
			A	B	C
Symptom appearance	Short	Triangular	-	0	15
	Medium		0	15	30
	High		15	30	-



Input variable	Fuzzy set name	Fuzzy set type	Fuzzy set parameters		
			A	B	C
Disappearance duration	Short	Triangular	-	0	15
	Medium		0	15	30
	High		15	30	-
AST test	Very Low	Triangular	-	60	92.5
	Low		60	92.5	125
	Medium		92.5	125	157.5
	High		125	157.5	190
	Very High		157.5	190	-
ALT test	Very Low	Bell	70	16	3
	Low		102.5	16	3
	Medium		135	16	3
	High		167.5	16	3
	Very High		200	16	3
CK test	Very Low	Bell	150	75	3
	Low		300	75	3
	High		450	75	3
	Very High		600	75	3
Potassium test	Low	Gaussian	5	0.42	-
	Medium		6	0.42	-
	High		7	0.42	-
Creatinine test	Low	Bell	1.5	0.75	2.5
	High		3	0.75	2.5
Age	Very Young	Gaussian	35	8	-
	Young		50	8	-
	Old		75	8	-
	Very Old		90	8	-

**Table 4** Fuzzy rules that link input and output variables

Relationship between	Rules	Confidence Level
Symptom appearance and temporal association	<ul style="list-style-type: none"> <li>If Symptom Appearance is <i>Short</i> Then Temporal Association is <i>Likely</i>.</li> <li>If Symptom Appearance is <i>Medium</i> Then Temporal Association is <i>Possible</i>.</li> <li>If Symptom Appearance is <i>Long</i> Then Temporal Association is <i>Unlikely</i>.</li> </ul>	CL <sub>1</sub> CL <sub>2</sub> CL <sub>3</sub>
Disappearance duration and dechallenge	<ul style="list-style-type: none"> <li>If Disappearance Duration is <i>Short</i> Then Dechallenge is <i>Likely</i>.</li> <li>If Disappearance Duration is <i>Medium</i> Then Dechallenge is <i>Possible</i>.</li> <li>If Disappearance Duration is <i>Long</i> Then Dechallenge is <i>Unlikely</i>.</li> <li>If Symptom does disappear after withdrawal of the drug then Dechallenge is <i>Unlikely</i>.</li> </ul>	CL <sub>4</sub> CL <sub>5</sub> CL <sub>6</sub> CL <sub>7</sub>
Temporal association and rechallenge	<ul style="list-style-type: none"> <li>If Temporal Association of time <math>t_1</math> is <i>Likely</i> and Temporal Association of time <math>t_2</math> is <i>Likely</i> Then Rechallenge is <i>Likely</i>.</li> </ul>	CL <sub>8</sub>
	<ul style="list-style-type: none"> <li>If Temporal Association of time <math>t_1</math> is <i>Likely</i> and Temporal Association of time <math>t_2</math> is <i>Possible</i> Then Rechallenge is <i>Likely</i>.</li> </ul>	CL <sub>9</sub>
	<ul style="list-style-type: none"> <li>If Temporal Association of time <math>t_1</math> is <i>Likely</i> and Temporal Association of time <math>t_2</math> is <i>Unlikely</i> Then Rechallenge is <i>Possible</i>.</li> </ul>	CL <sub>10</sub>
	<ul style="list-style-type: none"> <li>If Temporal Association of time <math>t_1</math> is <i>Possible</i> and Temporal Association of time <math>t_2</math> is <i>Likely</i> Then Rechallenge is <i>Likely</i>.</li> </ul>	CL <sub>11</sub>
	<ul style="list-style-type: none"> <li>If Temporal Association of time <math>t_1</math> is <i>Possible</i> and Temporal Association of time <math>t_2</math> is <i>Possible</i> Then Rechallenge is <i>Possible</i>.</li> </ul>	CL <sub>12</sub>
	<ul style="list-style-type: none"> <li>If Temporal Association of time <math>t_1</math> is <i>Possible</i> and Temporal Association of time <math>t_2</math> is <i>Unlikely</i> Then Rechallenge is <i>Possible</i>.</li> </ul>	CL <sub>13</sub>
	<ul style="list-style-type: none"> <li>If Temporal Association of time <math>t_1</math> is <i>Unlikely</i> and Temporal Association of time <math>t_2</math> is <i>Likely</i> Then Rechallenge is <i>possible</i>.</li> </ul>	CL <sub>14</sub>
	<ul style="list-style-type: none"> <li>If Temporal Association of time <math>t_1</math> is <i>Unlikely</i> and Temporal Association of time <math>t_2</math> is <i>Possible</i> Then Rechallenge is <i>Possible</i>.</li> </ul>	CL <sub>15</sub>
	<ul style="list-style-type: none"> <li>If Temporal Association of time <math>t_1</math> is <i>Unlikely</i> and Temporal Association of time <math>t_2</math> is <i>Unlikely</i> Then Rechallenge is <i>Unlikely</i>.</li> </ul>	CL <sub>16</sub>



Relationship between	Rules	Confidence Level
AST Test and AST abnormality	• If AST Test is <i>Very Low</i> Then AST Abnormality is <i>Very Low</i> .	CL <sub>17</sub>
	• If AST Test is <i>Low</i> Then AST Abnormality is <i>Low</i> .	CL <sub>18</sub>
	• If AST Test is <i>Medium</i> Then AST Abnormality is <i>Medium</i> .	CL <sub>19</sub>
	• If AST Test is <i>High</i> Then AST Abnormality is <i>High</i> .	CL <sub>20</sub>
	• If AST Test is <i>Very High</i> Then AST Abnormality is <i>Very High</i> .	CL <sub>21</sub>
ALT Test and ALT abnormality	• If ALT Test is <i>Very Low</i> Then ALT Abnormality is <i>Very Low</i> .	CL <sub>22</sub>
	• If ALT Test is <i>Low</i> Then ALT Abnormality is <i>Low</i> .	CL <sub>23</sub>
	• If ALT Test is <i>Medium</i> Then ALT Abnormality is <i>Medium</i> .	CL <sub>24</sub>
	• If ALT Test is <i>High</i> Then ALT Abnormality is <i>High</i> .	CL <sub>25</sub>
	• If ALT Test is <i>Very High</i> Then ALT Abnormality is <i>Very High</i> .	CL <sub>26</sub>
CK Test and CK abnormality	• If CK Test is <i>Very Low</i> Then CK Abnormality is <i>Very Low</i> .	CL <sub>27</sub>
	• If CK Test is <i>Low</i> Then CK Abnormality is <i>Low</i> .	CL <sub>28</sub>
	• If CK Test is <i>High</i> Then CK Abnormality in is <i>High</i> .	CL <sub>29</sub>
	• If CK Test is <i>Very High</i> Then CK Abnormality is <i>Very High</i> .	CL <sub>30</sub>
Potassium test and potassium abnormality	• If Potassium Test is <i>Low</i> Then Potassium Abnormality <i>Low</i> .	CL <sub>31</sub>
	• If Potassium Test is <i>Medium</i> Then Potassium Abnormality is <i>Medium</i> .	CL <sub>32</sub>
	• If Potassium Test is <i>High</i> Then Potassium Abnormality is <i>High</i> .	CL <sub>33</sub>
Creatinine test and creatinine abnormality	• If Creatinine Test is <i>Low</i> Then Creatinine Abnormality is <i>Low</i> .	CL <sub>34</sub>
	• If Creatinine Test is <i>High</i> Then Creatinine Abnormality is <i>High</i> .	CL <sub>35</sub>
Age and medication side effect risk	• If Age is <i>Very Young</i> Then Medication Side Effect is <i>Very Low</i> .	CL <sub>36</sub>
	• If Age is <i>Young</i> Then Medication Side Effect is <i>Low</i> .	CL <sub>37</sub>
	• If Age is <i>Old</i> Then Medication Side Effect is <i>High</i> .	CL <sub>38</sub>
	• If Age is <i>Very Old</i> Then Medication Side Effect is <i>Very High</i> .	CL <sub>39</sub>
		CL <sub>40</sub>

Fuzzy inference engine evaluates the rules using the min-max fuzzy inference operations. The resulting aggregated fuzzy set is converted to numerical values for the output variables by Defuzzifier that uses the center of gravity scheme.

The ordered weighted averaging (OWA) operators provide a parameterized class of mean type aggregation operators such as the max, arithmetic average, median and mean, are members of this class. The OWA operator is a mean or averaging operator. This is a reflection of the fact that the operator is commutative, monotonic, bounded and idempotent. Different families of OWA operators can be used by choosing a different manifestation of the weighting factor.

The aggregated laboratory abnormality is calculated as a linear combination of the corresponding sub abnormalities. The aggregated laboratory abnormality is computed as follows:

Laboratory Abnormality = ALT Abnormality x w1 + AST Abnormality x w2 + CK Abnormality x w3 + Potassium Abnormality x w4 + Creatinine Abnormality x w5.

Where w1 + w2 + w3 + w4 + w5 =1 (1)

The weights control the importance of the sub abnormalities. In ADR problem some laboratory abnormalities are more important than the other based on the studied medication.

The agents are supported with ability of mapping free text terms of unique concepts. This was done by lexical processing unit. It uses the lexical variants generator (LVG) program provided by the National Library of Medicine [36]. LVG is the most powerful solution for lexical variations at the individual word level. This unit allows the agents to deal with inflectional variants, spelling variants, acronyms and abbreviations, expansions, derivational variants, synonyms as well as combinations of these. The *Lexical Processing Unit uses Lexical Knowledge Base*. This knowledge base typically contains linguistic knowledge, such as word meanings, the syntactic patterns in which they occur, and special usage and idiosyncratic information, organized around the words in the language.

## 2.2 Decision tree unit

ADR classification process as {Present, Not Present} involves four phases: data gathering phase, data pre-processing phase, the learning phase and the recognition phase. In data gathering, the training and test set will be obtained from medical databases. The

second phase is to pre-process the experimental data, including data cleaning, sampling, creating new records (attributes) and record selection. The relatively less correlated and redundant records in the database will be removed in this phase. In the learning phase, the target is to build a model.

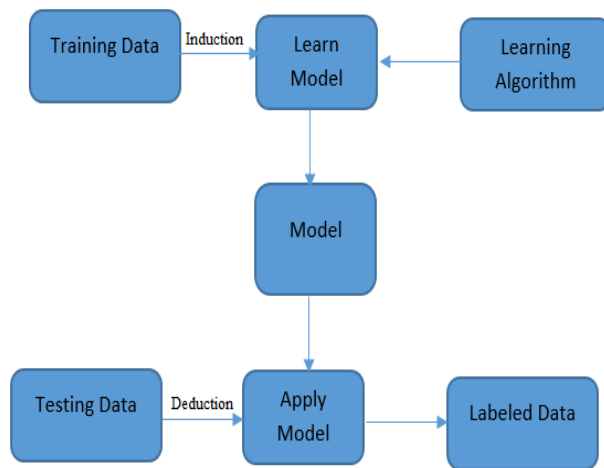
The last step is using the remainder of pre-processed data to test the model. A test set is used to determine the accuracy of the model. Usually, the given dataset is divided into training and test sets, with training set used to build the model and test set used to validate it (Figure 4).

A decision tree partition the input space of the dataset into mutually exclusive regions by giving each region a label [27, 37, 38]. The decision tree that consists of a root and internal nodes grows from a root node, by determining the best split that partition the region at internal nodes into disjoint smaller subsets and proceed down to the leaf node (terminal nodes) labelled as-present or not present. In order to do the split an error function that quantifies the performance of a node  $t$  in separating data from different classes. The used error function is impurity function. The best known impurity function for splitting is entropy function and Gini index. By using the impurity function  $\phi$ , the impurity measure of a given node  $t$  as in (2).

$$E(t) = \phi(p_1, p_2, \dots, p_j) = -\sum_{j=1}^J p_j \log_2 p_j \quad (2)$$

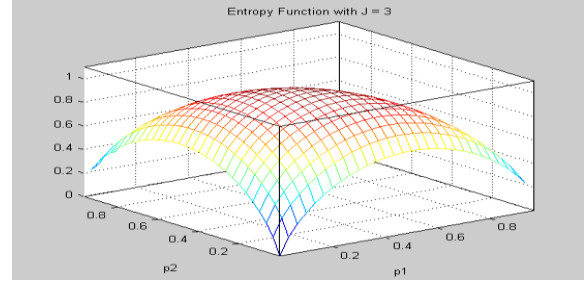
where  $p_j$

=  $P(j|t)$ , the probability (relative frequency) of class  $j$  at node  $t$



**Figure 4** The learning process: training and testing

The entropy function for  $j=3$  is shown in Figure 5.



**Figure 5** Entropy function with  $J=3$

The entropy measures homogeneity of a node. The maximum entropy value ( $\log n_c$ ) when records are evenly distributed among all classes, implying least information and minimum value of entropy (0.0) when all records belong to one class, implying most information.

Similarly, the impurity measure of a tree  $T$  can be expressed as

$$E(T) = \sum_{t \in \Psi} \frac{n_t}{n} E(t) \quad (3)$$

Where  $\Psi$  is the set of terminal nodes in the tree  $T$ ,

$n_t$  = number of records at child  $t$ ,

$n$  = number of records at the terminal node.

The information gain is calculated as in (4)

$$\begin{aligned} \text{GAIN}_{\text{split}} &= E(P) - E(T) \\ &= E(P) - \sum_{i=1}^k \frac{n_i}{n} E(i) \end{aligned} \quad (4)$$

Where Parent Node  $P$  (non-leaf node, node with partition) is split into  $k$  partitions (children),  $n_i$  is the number of records in partition  $i$ ,  $n$  = number of records at the terminal node.

The information gain measures reduction in Entropy achieved because of the split. Choose the split that achieves most reductions (maximizes GAIN) due to the disadvantage that information gain tends to prefer splits that result in a large number of partitions, each being small but pure. GainRATIO is designed to overcome the disadvantage of Information Gain by adjusting information gain by the entropy of the partitioning.

$$\text{GainRATIO} = \frac{\text{GAIN}_{\text{split}}}{-\sum_{i=1}^k \frac{n_i}{n} \log_2 \frac{n_i}{n}} \quad (5)$$

Where,  $n_i$  is the number of records in partition  $i$

Four attributes with high information gain ranking are chosen to develop the decision tree model: challenge, rechallenge, a laboratory test abnormality, and medication risk factor. The algorithm for DT induction is shown in Figure 6.

**Initialization**

Partition the training set space into {Present ADR, Not-present ADR}  
 Calculate the total Entropy of the system  
 Calculate the  $\text{Gain}_{\text{split}}$  for each one of the five attributes {Dechallenge, Rechallenge, Lab Test Abnormality, Medication side effect risk}  
 Choose the attribute with the highest information gain as the root node

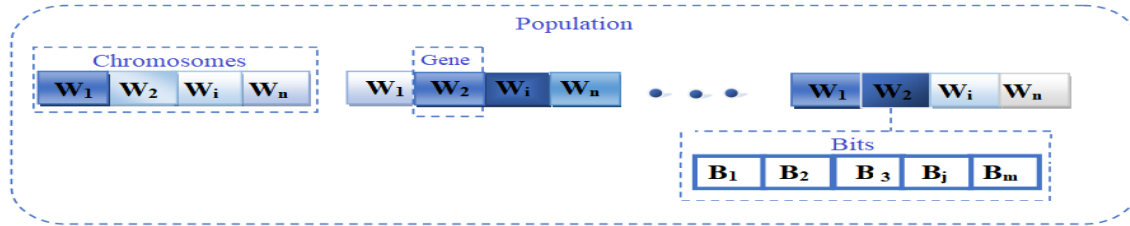
**Expand the Tree starting from the previously selected node****Repeat until all records are classified**

Eliminate the previous selected node from the list of attributes  
 Partition attributes values of the remaining attributes in the space into {Present, Not-present}  
 Calculate the Entropy of the attributes  
 Calculate the  $\text{Gain}_{\text{split}}$  of the new list of attributes  
 Choose the attribute with the highest information gain as the next node

**End by return Tree****Figure 6** The decision tree induction algorithm**2.3 Genetic algorithm unit**

GA is one of the most popular derivative free optimization technique which is based on the principles of evolution and natural genetics [38]. The GA starts by encoding each point in the parameter space into concatenated binary strings in which each concatenated value composed of set binary bits using binary coding techniques. Different binary encoding techniques are available such that Excess-3 code, binary coded decimal (BCD) and gray code. The

resulted binary bits is called a chromosome. A set of chromosomes in the solution space is called a population. Chromosome consists of a set of genes that contains information about the key parameters in a candidate solution. *Figure 7* shows a set of chromosomes of a population, where  $n$  weight being used as genes. Each gene is composed of  $m$  binary bits.

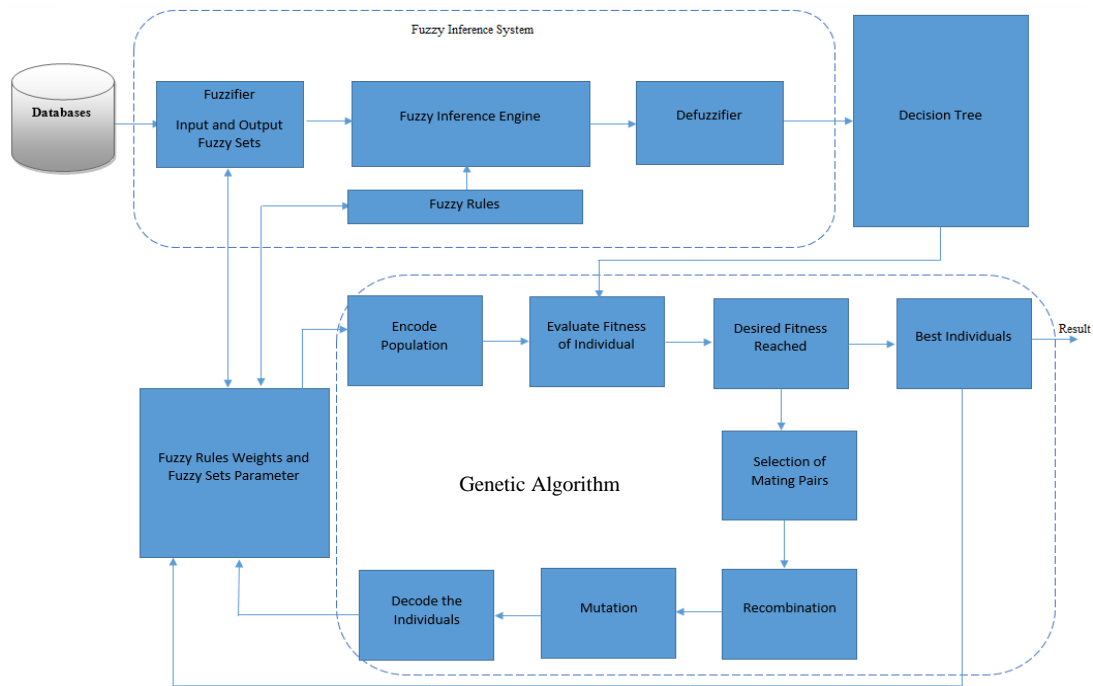
**Figure 7** Block diagram of population structure

The second step after creating a population of chromosomes is to calculate the fitness value of each member of the population. The fitness value  $f_i$  of the  $i^{\text{th}}$  weight parameter is the objective function evaluated at this weight set. The fitness function is chosen to be the root mean squared differences between the correct decision specified by physician  $T$  and the decision given by the decision tree  $\hat{T}$ . The root mean squared error (RMSE) is given by (6). It is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the actual class labeled by a physician,

$$\text{Objective Function} = \text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (T - \hat{T})^2}{n}} \quad (6)$$

Where  $n$  is the number of records of training data

By this definition, the lower the fitness, the better the developed model and a fitness of zero means that the model achieve the desired behavior for all inputs. As long as the fitness measure ranks the individuals accurately based on their performance, the exact form of the fitness is irrelevant to the working of the algorithm. The proposed GA operates as illustrated in *Figure 8*.



**Figure 8** Block diagram of the developed GA methodology

GAs makes multiple way searches by creating a population of candidate solutions instead of just one single test solution. A starts by constructing a new population using genetic operation such as crossover and mutation through an iterative process until some convergence criteria are met.

The resulted new population will be decoded back to its original format. The process is:

1. **Evaluation:** Sort the population based on chromosomes scores (fitness).
2. **Selection:** Choose the best chromosomes to generate the next population (natural selection). 75% of the sorted population will be kept in the new population.

3. **Crossover:** Merge the chromosomes by mixing their genes. Repeat the crossover operation until the new population is fully generated.
4. **Mutation:** Change some chromosomes arbitrarily. Usually, around 1% of the crossover chromosomes will go through the mutation process. Mutation process prevents any single bit from converging to a value throughout the entire population.

A new generation is created by repeating the selection, recombination and mutation processes until all chromosomes in the new population replace the initial population. *Figure 9* shows the algorithm used to determine the optimal solution.

#### Initialization

Set Parameters and choose encode method

Create Initial Population (Initial fuzzy rule weights and fuzzy sets parameters)

**For (iteration number= max number of iterations) and (Fitness=Max Fitness)**

    Fitness calculation

    Parent selection

    Crossover

    Mutation

    Elitism

**END\_For**

    Decode individual with maximum fitness

**End by return best solution: Output Fittest Fuzzy Rue Weights and Fuzzy Sets Parameters**

**Figure 9** GA-based decision tree optimization algorithm

We retrieved the patients who received at least one of the 8 drugs of our interest in Veterans Affairs Medical Center in Detroit during the time period from January 1, 2005 to December 31, 2008. These 8 drugs represent the first targets in studying ADRs. These drugs are statin drugs and ACE drugs. A statin is a type of drug that helps patients lowers their cholesterol. An ACE is a type of drug that treats high blood pressure. The interested drugs include 6 statin drugs (i.e., rosuvastatin, atorvastatin, fluvastatin, lovastatin, pravastatin, and simvastatin) and 2 ACE drugs (i.e., captopril, and enalapril).

The retrieved patient data includes dispensing of drug, office visits, symptoms experienced, and laboratory testing. For each event certain details were obtained. The total number of retrieved patients was 20,000 (19,102 males and 898 females). Their average age was 68.0. All the data stored in a Microsoft access database.

The retrieved patient data includes dispensing of drug, office visits, symptoms experienced, and laboratory testing. For each event certain details were obtained. The data for dispensing of drug includes the name of the drug, quantity of the drug dispensed, dose of the drug, drug start date, and the number of refills. The office visits data include treatment regimens, treatment start dates and stop dates. The symptoms experienced data includes the symptoms appearance date, the symptoms ICD-9 codes and the ICD-9 code description. The laboratory testing data includes the names of the laboratory tests, laboratory test dates, laboratory test normal ranges and laboratory test results.

The database had five tables, each of which contained one of the five types of information: (1) demographic data, (2) clinic visit data, (3) diagnostic data, (4) drug data, and (5) laboratory data.

The 20,000 patients clustered into three groups. The 1st group is for the patients who took only inhibitor drugs and they are 3,414 patients. The 2nd group is for the patients who took both drug classes, statin and inhibitor, and it contains 7,711 patients.

The 3rd group is for the patients who took only Statin drugs and it contains 8,875 patients. For the detection, evaluation experiments, we have selected randomly 200 patients out of the 20,000 patients (20% of the total population).

To improve the accuracy and decrease the training cost of the developed system, outlier and missing values in the data are removed. This decreases the population of 186 patients.

We have selected 1% patients from each group. From the 1st, 2nd, 3rd group we have selected 46, 60, and 80 patients respectively. The selection of 1% samples from each group was based on systematic sampling. The systematic sampling is a method of selecting samples from a larger population, according to a random starting point and a fixed, periodic interval. Typically, every "nth" member is selected from the total population to be included in the sampled population. The "nth" member is selected by dividing the total number of members in the general population by the desired number of members of the sampled population. For example, for selecting a random group of 88 patients from a population of 8,800 using systematic sampling, you would simply select every 100th person, since  $8,800/88 = 100$ . Systematic sampling can be considered random, as long as the periodic sampling interval is determined beforehand and the sampling starting point is random [39].

Two physicians were participated in this study. They were asked to independently review each of the 280 patient cases and checked whether the patient case suffer from ADR or not by assigning a {present, not present}. In this evaluation, patient cases were retrieved one by one from databases using a visual basic program done for that purpose (*Figure 10*).

If the best-fitting instances are selected with GA, this might lead to over-fitting. So, in order to make sure that there is no over fitting, 1/4 of the data was held (70 cases out of 280 cases) and will be used in testing the proposed algorithm. In all the experiments the GA operates with the configuration shown in *Table 5*.

Patient ID  AGE  48 SEX  MALE

**PATIENT ICD9 TABLE**

ICD9 CODE	ICD9 DESCRIPTION	DIAG
305.90	OTHER DRUG ABUSE-UNSPEC	1/17/2001
303.93	ALCOH DEP NEC/NOS-REMISS	1/18/2001
311.	Depressive Disorder NOS (ICD-9-CM)	11/4/2002
401.9	HYPERTENSION NOS	10/11/2004
389.18	SENSORINEURAL HEARING LO	1/28/2005
389.8	OTHER SPECIFIED FORMS OF	1/28/2005
362.56	EPIRETINAL MEMBRANE	4/25/2005
362.11	HYPERTENSIVE RETINOPATHY	4/25/2005
070.51	Hepatitis C	5/10/2005
304.20	Cocaine Dependence, Nos	7/14/2005
304.00	Opioid Dependence (ICD-9-CM)	7/2/2007

Record: 1 of 12 No Filter Search

**PATIENT DRUG TABLE**

DRUG NAME	START DATE	SCH
ATENOLOL 50MG/CHLORTHALID	9/5/2008	DAILY
LISINAPRIL 40MG TAB	9/5/2008	DAILY
CEPHALEXIN 500MG CAP	8/11/2008	Q6H
GAUZE PAD 4IN X 8IN 12-PLY N	8/11/2008	QD(DAILY)
IBUPROFEN 600MG TAB	8/11/2008	Q6H WF PRN
KERLIX 4.5IN NONSTERILE	8/11/2008	AS DIRECTED
TAPE,MICROPORE 1IN 3M #153	8/11/2008	QD(DAILY)
NAPROXEN NA 550MG TAB	7/6/2007	Q12H PRN
MOXIFLOXACIN HCL 400MG TAE	4/14/2008	QD(DAILY)
LORATADINE 10MG TAB	4/14/2008	QD(DAILY)
DM 10/GUAIFENESIN 100MG/5M	4/14/2008	Q6H

Record: 1 of 18 No Filter Search

**PATIENT LAB TABLE**

LAB DATE	TEST NAME	RESULT
APR 21,2005 02:22:21	ALT	64
APR 21,2005 02:22:21	AST	66
APR 21,2005 02:22:21	CKMB	3.04
APR 21,2005 02:22:21	CK	337
APR 21,2005 02:22:21	CREATININE	1.2
APR 21,2005 02:22:21	POTASSIUM	4.1
APR 21,2005 06:59:25	CKMB	1.79
APR 21,2005 06:59:25	CK	223
APR 28,2005 07:58:36	ALT	62
APR 28,2005 07:58:36	AST	62
APR 28,2005 07:58:36	CREATININE	1.1
APR 28,2005 07:58:36	POTASSIUM	3.4
APR 6,2005 14:27:56	ALT	40
APR 6,2005 14:27:56	AST	42
APR 6,2005 14:27:56	CREATININE	0.9

Record: 1 of 81 No Filter Search

**PATIENT VISIT TABLE**

CPT C	CPT DESCRIPTION	VISIT DATE	
104840	MISCELLANEOUS DME SUPPLY	JUL 16,2004 23:31	311.
104840	MISCELLANEOUS DME SUPPLY	JUL 29,2004 23:31	719.47
99243	OFFICE CONSULTATION FOR A	JAN 4,2005 15:19	300.9
99285	EMERGENCY DEPARTMENT VIS	APR 21,2005 01:31:03	786.50
99281	EMERGENCY DEPARTMENT VIS	APR 20,2005 08:20	490.
99284	EMERGENCY DEPARTMENT VIS	APR 28,2005 09:14:32	558.9
101513	INFUSION, NORMAL SALINE SO	APR 28,2005 09:14:32	558.9
101464	INJECTION, TRIMETHOBENZAMI	APR 28,2005 09:14:32	558.9
105861	ALCOHOL AND/OR DRUG SERV	JUL 2,2007 15:00	304.01
105861	ALCOHOL AND/OR DRUG SERV	JUL 13,2007 08:00	304.01

Record: 1 of 10 No Filter Search

Figure 10 User interface for showing patient data

Table 5 GA setting

Parameter	Value
Population size	20
Variable range	[1,J], J is a number of attributes
Maximum generation	200
Crossover points	2 points
Crossover probability	0.75
Mutation probability	0.005
Elitism	Yes
Selection method	Uniform selection

### 3.Results

Java was selected as the development language and J2SDK version 1.6.0\_22 was used in the Java environment. Access database 2007 was adopted for the development of the database. Java database connectivity (JDBC), an application programming interface for the Java programming language, was used to access the database. JDBC could wrap a structured query language (SQL) statement, send it to the database, and retrieve the desired.

In this study, Weka is used [32] to construct decision trees, according to the training set with the standard algorithm J48 [37, 38]. Weka is a popular suite of

machine learning software written in Java, developed at the University of Waikato, New Zealand. Weka is a free software tool available under the GNU General Public License. It contains a collection of visualization tools and algorithms for data analysis and predictive modelling that support data pre-processing, clustering, classification, regression, visualization, and feature selection. Weka has a powerful graphical user interface that supports its functionality.

Once the features have been extracted and grouped into a feature vector, classification takes place, where each patient case is classified in one of the two classes: present or not present. J48 (C4.5) is an algorithm used to generate a decision tree. It has been developed by John Ross Quinlan. C4.5 is an extension of Quinlan's earlier ID3 algorithm [40]. The decision trees generated by C4.5 are normally used for classification.

It has been ranked first in the top 10 algorithms in data mining algorithm [41]. Figure 11 shows Weka J48 algorithm results GUI. Figure 12 is the Weka constructed decision tree from the Fuzzy Inference system optimized by GA.



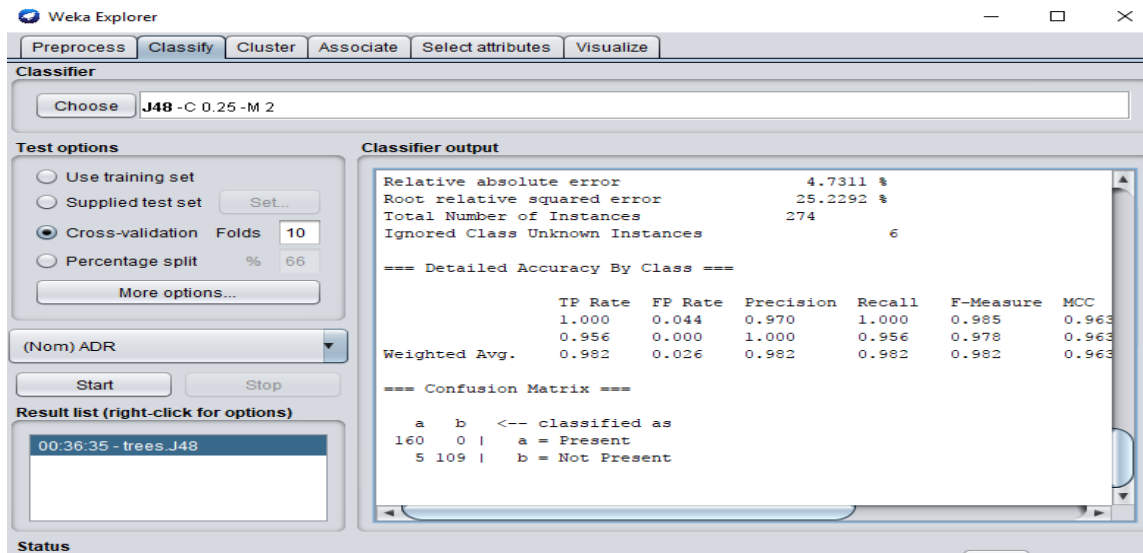


Figure 11 Weka J48 algorithm results

```

==== Run information ====
Scheme:   weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: ADR-fin
Instances: 280
Attributes: 6
          Patient ID
          Challenge
          Dechallenge
          Lab Abnormality
          Medication Side Effect Risk
          ADR

Test mode: 10-fold cross-validation
==== Classifier model (full training set) ====
J48 pruned tree

Challenge = likely
| Lab Abnormality = very high : Present (48.0)
| | Lab Abnormality = high
| | | Dechallenge = likely
| | | | Medication Side Effect Risk = high: Present (6.0)
| | | | Medication Side Effect Risk = Very High: Present (0.0)
| | | | Medication Side Effect Risk = Low: Not Present (6.0)
| | | | Medication Side Effect Risk = Very Low: Present (0.0)
| | | Dechallenge = Possible: Present (12.0)
| | | Dechallenge = Unlikely: Present (0.0)
| | Lab Abnormality = high
| | | Medication Side Effect Risk = high: Present (0.0)
| | | Medication Side Effect Risk = Very High: Present (12.0)
| | | Medication Side Effect Risk = Low: Present (0.0)
| | | Medication Side Effect Risk = Very Low: Not Present (12.0)
| | Lab Abnormality = low
| | | Medication Side Effect Risk = high
| | | | Dechallenge = likely: Not Present (6.0)
| | | | Dechallenge = Possible: Present (6.0)
| | | | Dechallenge = Unlikely: Present (0.0)
| | | Medication Side Effect Risk = Very High
| | | | Dechallenge = likely: Present (6.0)
| | | | Dechallenge = Possible: Present (0.0)
| | | | Dechallenge = Unlikely: Not Present (6.0)
| | | Medication Side Effect Risk = Low: Not Present (12.0)
| | | Medication Side Effect Risk = Very Low: Not Present (12.0)
| | Lab Abnormality = Very Low
| | | Medication Side Effect Risk = high: Present (10.0)
| | | Medication Side Effect Risk = Very High: Present (10.0)
| | | Medication Side Effect Risk = Low: Present (10.0)
| | | Medication Side Effect Risk = Very Low: Not Present (10.0)
Challenge = Possible
| Lab Abnormality = very high : Present (12.0)
| Lab Abnormality = high : Present (12.0)
| Lab Abnormality = high : Present (0.0)
| Lab Abnormality = low: Present (11.0)
| Lab Abnormality = Very Low
| | Medication Side Effect Risk = high: Present (5.0)
| | Medication Side Effect Risk = Very High: Present (0.0)
| | Medication Side Effect Risk = Low: Not Present (5.0)
| | Medication Side Effect Risk = Very Low: Present (0.0)
Challenge = Unlikely: Not Present (45.0)

Number of Leaves :    32
Size of the tree : 43

```

Figure 12 ADRs classification model



#### 4. Discussion and analysis

We examined the agreement between the results generated by the developed system and the one by the physicians.

**Table 6** Confusion matrix

Actual class	Predicted class		
	Class=Present		Class=Not-Present
	Class=Present	True Positive (TP)	False Negative (FN)
	Class=Not-Present	False Positive (FP)	True Negative(TN)

The performance measurements used in this paper were recall, precision, classifier  $F_1$  rating, accuracy, pre-processing time, and running time. Pre-processing time includes dataset pre-processing time, and classifier training time. Running time indicates the classifier testing time.

They are defined as follows:

Recall (R) is the ratio of the relevant data among the retrieved. Precision (P) is the ratio of the accurate data between the retrieved data. Their formulas are given as follows:

$$\text{Recall(R)} = \frac{T_P}{T_P + F_P} \text{ if } TP + FN > 0, \text{ otherwise undefined.} \quad (7)$$

$$\text{Precision(P)} = \frac{T_P}{T_P + F_N} \text{ if } TP + FN > 0, \text{ otherwise undefined.} \quad (8)$$

The classifier  $F_1$  rating is the harmonic mean of the classifier recall and the precision. It is given as

$$F_1 = \frac{2 \cdot P \cdot R}{P + R} \quad (9)$$

Where  $R$  represents the recall and  $P$  represents the precision.

Accuracy, which indicates the fraction of correctly classified samples among all the samples, obtained by:

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (10)$$

Dataset pre-processing time represents the time needed to read the dataset and convert it to an accepted format. Classifier training time represents the time needed by the classifier. Classifier testing time represents the time to classify the testing long string values.

Table 7 shows the resulted ADRs classification model evaluation results. The proposed DT model shows high accuracy on ADR classification (up to 98.17%) on the test data. To further validate the results, K-fold cross validation was used. In K-fold the training set will be randomly split into K that has

We constructed the confusion matrix for each class (present or not -present). The confusion matrix has the form shown in Table 6.

approximately the same size. Then the decision tree will be trained using (K-2) subsets. One of the two remaining subsets will be used for validation and the last for testing. This process will be repeated  $K$  times, while a different subset is used for testing and validation.

**Table 7** Overall performance results (training and validation set)

Total number of instances	210
Correctly classified instances	207
	98.5714 %
Incorrectly classified instances	3
	1.4286 %
Kappa statistic	0.9703
Mean absolute error	0.0273
Root mean squared error	0.1199
Relative absolute error	5.6443%
Root relative squared error	24.3688 %

Another performance indicated by confusion matrix is shown in Table 8. This confusion matrix was built based on data testing. We constructed the confusion matrix for each class (Present, not Present). The confusion matrix has the form shown in Table 8.

**Table 8** Confusion matrix (training and validation set)

	Present	Not present
Present	124	0
Not Present	3	83

The performance measurement results are shown in Table 9.

**Table 9** ADRs classification results (training and validation set)

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0.035	0.976	1	0.988	Present
0.965	0.000	1	0.965	0.982	Not Present

Using the held 70 cases not previously used in the training or cross validation. The achieved results are shown in *Table 10* and *Table 11*.

**Table 10** Overall performance results (testing set)

Total number of instances	70
Correctly classified instances	68
	97.1429 %
Incorrectly classified instances	2
	2.8571 %
Kappa statistic	0.9405
Mean absolute error	0.0359

**Table 12** ADRs classification results (testing set)

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0.069	0.953	1	0.976	Present
0.931	0.000	1	0.931	0.964	Not-Present

The final performance of decision tree has been improved compared with the initial performance of decision tree shown in *Table 13*.

**Table 13** Initial decision tree performance results

Total number of instances	210
Correctly classified instances	124
	59.0476 %
Incorrectly classified instances	86
	40.9523 %
Kappa statistic	0
Mean absolute error	0.4135
Root mean squared error	0.643
Relative absolute error	90.5263%
Root relative squared error	123.9239 %

We have utilized Kappa statistic to estimate the levels of agreement. The Kappa coefficient is an estimate of the agreement between two raters. Kappa scores range between 1 (complete agreement) and 0. If the two raters randomly assign their ratings, however, they would sometimes agree just by chance. Kappa gives us a numerical rating of the degree to which this occurs. The calculation is based on the difference between how much agreement is actually

**Table 14** Statistics for weighted Kappa, ASE and 95% confidence limits

	Weighted Kappa coefficient	ASE	95% confidence limits
Physician 1 vs. developed System	0.89	0.044	(0.89 , 0.92)
Physician 2 vs. developed System	0.91	0.034	(0.86 , 0.93)
Physician 1 vs. Physician 2	0.88	0.051	(0.85 , 0.89)

These coefficients suggest good to excellent agreement between the proposed model and the physicians. According to this experiment, the system showed a superior performance and it was able to solve the problem efficiently.

Root mean squared error	0.1572
Relative absolute error	7.4114 %
Root relative squared error	31.9186 %

**Table 11** Confusion matrix (testing set)

	Present	Not present
Present	41	0
Not Present	2	27

The performance measurement result is shown in *Table 12*.

present (“observed” agreement) compared to how much agreement would be expected to be present by chance alone (“expected” agreement). Not everyone would agree on the interpretation of Kappa. However, a commonly cited scale [42] suggested that there is excellent agreement for values of Kappa greater than 0.75, poor agreement for values less than 0.4 and fair to good agreement for values between 0.40 and 0.75.

*Table 10* summarizes the number of matches between the decision of the system and the two physicians.

The estimate of agreements is as follows: Kappa = 0.89 for physician 1 and the model; Kappa = 0.91 for physician 2 and the model; Kappa = 0.88 for physician 1 and physician 2. These coefficients suggest excellent agreement between the system and the physicians. The asymptotic standard error (ASE) is also computed, as well as 95% confidence bounds. The results of these two methods are shown in *Table 14*.

## 5. Conclusion

Several thousands of drugs are currently available on the market. A complete understanding of the safe use of drugs is not possible at the time when the drug is developed or marketed. Early detection of unknown

adverse drug reactions could save lives and prevent unnecessary hospitalizations. The majority of current methods largely rely on spontaneous reports which suffer from serious underreporting, latency, and inconsistent reporting. Thus, they are not ideal for rapidly detect adverse drug reactions. In this paper, we have developed a system to detect adverse drug reaction using both fuzzy logic and decision tree approaches. Fuzzy logic is used to extract higher level information from patients' medical data. The extracted information is used by decision tree approach to construct a decision model.

Using real data of patients who were treated at the Veterans Affairs Medical Center in Detroit during the period between 2005 and 2008, the decision model has been constructed, tested and verified. The performance results show that the proposed system can detect adverse drug reaction with 97.14% accuracy. The experiment results were evaluated independently by two physicians on our team. Kappa statistics has been to evaluate the system results. The kappa coefficients show excellent agreement between the decision of physicians and the proposed model. In future work, other decision tree approaches such as random forests will be considered as they normally perform better than decision trees. Type 2 fuzzy sets or interval type 2 fuzzy can also be used instead of type 1 fuzzy sets in order to get a more precise decision and hence enhance the detection performance.

### Acknowledgment

We would like to thank Ellison Floyd for retrieving the de-identified patient information from Detroit Veterans Affairs Medical Center's databases.

### Conflicts of interest

The author has no conflicts of interest to declare.

### References

- [1] Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. *The Lancet*. 2000; 356(9237):1255-9.
- [2] <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm070461.htm>. Accessed 26 January 2018.
- [3] Stephens MD, Talbot J, Waller P. Stephens' detection of new adverse drug reactions. John Wiley & Sons; 2004.
- [4] Kaufman G. Adverse drug reactions: classification, susceptibility and reporting. *Nursing Standard* (2014+). 2016; 30(50):53-63.
- [5] Svendsen K, Halvorsen KH, Vorren S, Samdal H, Garcia B. Adverse drug reaction reporting: how can drug consumption information add to analyses using spontaneous reports?. *European Journal of Clinical Pharmacology*. 2018; 74(4):497-504.
- [6] Schmiedl S, Rottenkolber M, Szymanski J, Drewelow B, Siegmund W, Hippus M, et al. Preventable ADRs leading to hospitalization-results of a long-term prospective safety study with 6,427 ADR cases focusing on elderly patients. *Expert Opinion on Drug Safety*. 2018; 17(2):125-37.
- [7] Dormann H, Criegee-Rieck M, Neubert A, Egger T, Geise A, Krebs S, et al. Lack of awareness of community-acquired adverse drug reactions upon hospital admission. *Drug Safety*. 2003; 26(5):353-62.
- [8] Zheng Y, Lan C, Peng H, Li J. Using constrained information entropy to detect rare adverse drug reactions from medical forums. In annual international conference of the engineering in medicine and biology society 2016 (pp. 2460-3). IEEE.
- [9] Cazacu I, Stroe R, Dondera R, Mogosan C, Haramburu F, Fourier-Réglat A, et al. Adverse drug reactions of analgesic medicines: analysis of the Romanian pharmacovigilance database. *Fundamental & Clinical Pharmacology*. 2018; 32(3):330-6.
- [10] Stricker BH, Psaty BM. Detection, verification, and quantification of adverse drug reactions. *BMJ*. 2004; 329(7456):44-7.
- [11] Klein E, Bourdette D. Postmarketing adverse drug reactions: a duty to report?. *Neurology: Clinical Practice*. 2013; 3(4):288-94.
- [12] U.S. food and drug administration. <https://open.fda.gov/data/faers/>. Accessed 26 January 2018.
- [13] U.K. Yellow Card ADR reporting system. <http://yellowcard.mhra.gov.uk/>. Accessed 26 January 2018.
- [14] Canadian ADR reporting system. <http://www.hc-sc.gc.ca/dhp-mps/medeff/databasdon/index-eng.php>. Accessed 26 January 2018.
- [15] Central drugs standard control organization ADR reporting system of India. <http://www.cdsco.nic.in/forms>. Accessed 26 January 2018.
- [16] Maggo SD, Chua EW, Chin P, Cree S, Pearson J, Doogue M, et al. A New Zealand platform to enable genetic investigation of adverse drug reactions. *The New Zealand Medical Journal*. 2017; 130(1466):62-9.
- [17] Miller D. FDA's MedWatch program monitors drug safety, but it depends on nurses to report adverse events. *ONS Connect*. 2015; 30(3):49.
- [18] Rehan HS, Chopra D, Holani SN, Mishra R. An observational study to compare the contents and quality of information furnished in CDSCO ADR reporting form, yellow card, medwatch and blue form by the healthcare professionals. *International Journal of Risk & Safety in Medicine*. 2014; 26(1):1-8.
- [19] Gupta A, Woosley R, Crk I, Sarnikar S. An information technology architecture for drug effectiveness reporting and post-marketing surveillance. *International Journal of Healthcare Information Systems and Informatics (IJHISI)*. 2007; 2(3):65-80.

- [20] McGee S. Evidence-based physical diagnosis E-Book. Elsevier Health Sciences; 2016.
- [21] Kelly M, Kaye KI, Davis SR, Shenfield GM. Factors influencing adverse drug reaction reporting in New South Wales teaching hospitals. *Journal of Pharmacy Practice and Research*. 2004; 34(1):32-5.
- [22] Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology and Drug Safety*. 2001; 10(6):483-6.
- [23] Abin D, Mahajan TC, Bhoj MS, Bagde S, Rajeswari K. Causal association mining for detection of adverse drug reactions. In *international conference on computing communication control and automation 2015* (pp. 382-5). IEEE.
- [24] Harpaz R, Chase HS, Friedman C. Mining multi-item drug adverse effect associations in spontaneous reporting systems. In *BMC bioinformatics 2010* (p. S7). BioMed Central.
- [25] Kubota K, Koide D, Hirai T. Comparison of data mining methodologies using Japanese spontaneous reports. *Pharmacoepidemiology and Drug Safety*. 2004; 13(6):387-94.
- [26] Lilienfeld DE. A challenge to the data miners. *Pharmacoepidemiology and Drug Safety*. 2004; 13(12):881-4.
- [27] Giannakopoulos I, Tsoumakos D, Koziris N. A decision tree based approach towards adaptive modeling of big data applications. In *international conference on Big Data*. 2017 (pp. 163-72). IEEE.
- [28] Belohlavek R, Klir GJ. *Concepts and fuzzy logic*. MIT Press; 2011.
- [29] Singhai R. *Introduction to fuzzy logic* paperback. PHI Learning; 2013.
- [30] Zadeh LA. Fuzzy sets. *Information and control*. 1965; 8(3): 338-53.
- [31] Orchard R. Fuzzy reasoning in JESS: the fuzzyj toolkit and fuzzyjess. *International conference on enterprise information systems 2001* (pp. 533-42). National Research Council, Canada.
- [32] University of Waikato. Weka Software. <https://www.cs.waikato.ac.nz/ml/weka/>. Accessed 26 January 2018.
- [33] Doc N. Department of Health & Human Services. 2004.
- [34] Mansour AM. A multi-agent intelligent system for monitoring health conditions of elderly people. *International Journal of Electrical, Robotics, Electronics and Communications Engineering*. 2014; 8(6):993-8.
- [35] U.S. food and drug administration, National Drug Code Directory, USA. <http://www.fda.gov/Drugs/InformationOnDrugs/ucm142438.htm>. Accessed 26 January 2018.
- [36] U.S. National Library of Medicine. <http://www.nlm.nih.gov>. Accessed 26 January 2018.
- [37] Olson D. *Data mining models*. Business Expert Press; 2016.
- [38] Witten IH, Frank E, Hall MA, Pal CJ. *Data mining: practical machine learning tools and techniques*. Morgan Kaufmann; 2016.
- [39] Granger CW, Siklos PL. Systematic sampling, temporal aggregation, seasonal adjustment, and cointegration theory and evidence. *Journal of Econometrics*. 1995; 66(1-2):357-69.
- [40] Quinlan JR. C4. 5: programs for machine learning. Elsevier; 2014.
- [41] Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, et al. Top 10 algorithms in data mining. *Knowledge and Information Systems*. 2008; 14(1):1-37.
- [42] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33(1):159-74.



**Dr. Ayman M Mansour** received his Ph.D. degree in Electrical Engineering from Wayne State University in 2012. Dr. Mansour received his M.Sc degree in Electrical Engineering from University of Jordan, Jordan, in 2006 and his B.Sc degree in Electrical and Electronics Engineering from University of Sharjah, UAE, in 2004. He graduated top of his class in both Bachelor and Master. Currently, Dr. Mansour is an Assistant Professor in the Department of Communication and Computer Engineering, Tafila Technical University, Jordan. He is also the director of the Energy Research Center at Tafila Technical University. His areas of research include Communication Systems, Multi-agent Systems, Fuzzy Systems, Data Mining and Intelligent Systems. He conducted several researches in his area of interest. Dr. Mansour is a member of IEEE, Michigan Society of Professional Engineers, IEEE Honor Society (HKN), Society of Automotive Engineers (SAE), Tau Beta Pi Honor Society, Sigma Xi and Golden Key Honor Society  
Email: mansour@ttu.edu.jo