Connections of Reduced Performance Health Data for Severe Persistent Uncontrolled Allergic Asthma Treated by Omazulimab

Stefanos Matsopoulos and Valentina Plekhanova
Department of Computing, Enineering and Technology, University of Sunderland, Sunderland, U.K.

Keywords: Association Rules, Asthma, Data Mining, Spirometry Data.

Abstract: An application of association rules mining method for the discovery of associations in abnormal quantitative health data for inadequately controlled severe allergic Asthma treated by Omazulimab is presented. To the best of authors' knowledge, no formal approaches have ever been used for extraction of association rules among dysfunctional elements in Spirometry datasets. Initially is provided an explanation of the procedures used for diagnosing inadequately controlled severe allergic Asthma. Following this, it is conducted critical evaluation of well-known 'association rule' mining techniques, in order to identify the one with the best utility for discovery of associations among abnormal elements of Spirometry datasets. Apriori Algorithm is applied to real-life Spirometry datasets to illustrate the contribution of application of association rule mining techniques. This revealed the existence of association rules among dysfunctional Spirometry elements for this disease. Moreover it has been identified that this disease is provoked by association of Spirometry elements that do not function properly as these are provided by Spirometer. This is translated in human factors as a dysfunction of small and medium airways of patients'. Furthermore Spirometry element FEV1, is not as valuable parameter as the European Medical Agency supports. Finally it has been observed that Omazulimab treatment improves respiratory function and makes the connection among associated elements weaker.

1 INTRODUCTION

Asthma is a disease that affects a large portion of global population. Although asthma is separated into different categories according to its severity, with the most dangerous type being 'poorly controlled severe allergic Asthma'. This type of Asthma may have mortal effects if not treated, or if patients do not get the proper treatment in time, which occurs only after hospitalization (Bousquet et al, 2007).

Immunoglobulin E (IgE) is a responsible factor for severe allergic Asthma. During the last decade a treatment has been discovered, called Omazulimab, which is able to stabilize IgE. The appropriate dosage is calculated by a practitioner and its use leads to the reduction of exacerbations and hospitalization (Nowak, 2006). Furthermore this treatment can improve the patients' Quality of Life and reduce the risk of effects caused by Asthma (Nowak, 2006).

The difference between inadequately controlled severe allergic Asthma with other types of Asthma is its severity. It has plenty of exacerbations and it is not treated with the same medication as other types of Asthma. The use of Inhaled Corticosteroids (ICS) and Long-Acting-b2-Agonistics (LABAs) are not effective on this type of Asthma. It has to be treated with Omazulimab which has the ability to reduce high number of free-IgE that patients suffer from (Tzortzaki et al, 2012).

For an individual to be evaluated by a practitioner as a patient who requires treatment with Omalizumab, there are minimum medical requirements that have to be met. These will be discussed below. An important requirement that has to be fulfilled is an abnormal FEV1 value (air exhaled in the first second) (Tzortzaki et al, 2012).

Spirometry is a significant examination for the measurement of lung function. It is used to measure patients' respiratory health and is helpful in assessing conditions such as Chronic Obstructive Pulmonary Disease and Asthma (Stout et al, 2012).

Spirometry provides an outcome which consists of several numerical parameters. Each parameter represents functionality of a different part of the human respiratory. Also this examination is used to

Matsopoulos S. and Plekhanova V..
Connections of Reduced Performance Health Data for Severe Persistent Uncontrolled Allergic Asthma Treated by Omazulimab.
DOI: 10.5220/0004763802760286
In *Proceedings of the International Conference on Health Informatics* (HEALTHINF-2014), pages 276-286
ISBN: 978-989-758-010-9

measure response to treatment of conditions which Spirometry detects (Stout et al, 2012).

In general results above 80% predicted, according to the ideal, are considered normal, while results under 80% are considered abnormal and will be discussed below.

Spirometry data are collected and used for the discovery of associations of abnormal Spirometry results/values. This paper is based on the analysis of data through Data Mining Techniques. These techniques require quantitative-numerical data for concluding to a result. These are given by Spirometry examinations, which are the only examinations that provided quantitative data during treatment. They are used for representation of patients' clinical image and after appropriate analysis with 'data mining' techniques will lead to associations of abnormal Spirometry parameters (Nowak, 2006).

Except Spirometry data analysis, further analysis of health examinations by practitioner has to be taken into consideration when dealing with inadequately controlled severe allergic Asthma. This action leads to higher diagnostic accuracy. Forthwith the problem will be approached and examined only by a quantitative point of view and not overview of all health examinations. The reason for approaching this subject by a different point of view is because it has been found that Omalizumab treatment is not effective on all patients for unclear reasons (Slavin et al, 2009). After appropriate management and analysis of Spirometry data, this paper focuses on 'hidden' information which exists beneath Spirometry examination results and cannot be analysed in detail without an appropriate technique. In order to understand the nature of the problem, relevant information and techniques that are being used in traditional diagnosis and treatment procedures are required to be briefly introduced for the disease.

The structure of this paper is as follows. In the second section key aspects of severe allergic Asthma, with poorly controlled symptoms are analysed as are the methods used for diagnosing and evaluating Spirometry results.

Moreover in the third section there is a critical evaluation of data mining techniques. In the fourth section characteristics of the sample and simulation of the technique used are presented. Finally the last section focuses on recommendations for appropriate use of this technique.

The following research questions will be addressed in this paper:

Research Question 1: Are there any 'data mining' techniques that can find 'association rules' in Spirometry Data?

Research Question 2: Is there a 'data mining' technique that could be useful in the discovery of associations among Spirometry parameters for severe allergic Asthma inadequately controlled?

2 BACKGROUND OF INFORMATION MANAGEMENT IN ASTHMA DISEASE: KEY ASPECTS

As defined by multiple authors Asthma is an inflammatory disorder of the lung-function and its airways. It involves multiple inflammatory cells and the interaction of a big amount of different mediators of the lungs (Holgate et al, 2009).

Asthma is divided into mild and moderate, that will not be further analysed in this paper and severe persistent allergic Asthma, with inadequately controlled symptoms. (Korn et al, 2009).

2.1 Severe Allergic Asthma Inadequately Controlled: Key Aspects

Severe allergic Asthma with inadequately controlled symptoms can be characterized by the presence of IgE antibodies against allergens that are common such as animal dander or house dust (Slavin et al, 2009).

This can lead to lack of oxygen and incorrect lung function leading to hospitalisation or increased risk of death. (Nowak, 2006). Furthermore there are multiple exacerbations during a small period of time that can decrease the patients' Quality of Life.

The use of Omazulimab treatment, which contains Anti-Immunoglobulin E (anti-IgE), is useful for patients' who suffer from this type of Asthma (Tzortzaki et al, 2012).

Though it is unclear why the use of Omalizumab is not effective on all patients (Slavin et al, 2009). The relationship between clinical symptoms and free IgE has not been studied adequately. Consequently it is difficult to predict the characteristics of the patients that will be positively affected by the treatment (Holgate et al, 2009).

2.2 Minimum Medical Requirements to Initiate Omazulimab Treatment

For an individual to be taken under consideration for

Omazulimab treatment some minimum medical requirements have to be met. Patients have to be older than 12 years old and have a positive skin prick test or have reactivity to a perennial aeroallergen (Korn et al, 2009). Moreover lung function and especially FEV1 has to be lower than 80% as supported by the European Medical Agency (Korn et al, 2009). Additionally the number of severe asthma exacerbations that individuals may suffer from, no matter what amounts of ICS and LABAs is being used, has to be observed. Finally patients must have a treatable IgE over 30 IU/ml, which is measured before treatment starts.

Nevertheless the most important factor is the practitioners' evaluation on the patients' overall health and their need for treatment.

2.3 Utility of Spirometry in Severe Allergic Asthma Inadequately Controlled

As discussed above, one of the minimum requirements for receiving Omazulimab treatment is maintaining a FEV1 percentage less than 80% of the 'normal' measurement. 'Normal' is calculated by a Spirometer based on age, gender, height and weight (Stout et al, 2012). Every 2 to 4 weeks, depending on patient's dosage a Spirometry examination is performed (Holgate et al, 2009). With this examination a practitioner will be able to evaluate the patient's improvement. Furthermore Spirometry is a 'mirror-indicator' of a patients' respiratory function and cannot be examined by any other quantitative test (Stout et al, 2012).

2.4 Traditional Procedures on Diagnosis of Improvement during Treatment

Traditionally practitioners do not have a specific technique which can be used for accurate identification and evaluation of patients' health. Instead there is a procedure that has to be followed. After evaluation and comparison of different examinations a decision is made based only on the practitioners' opinion. There is no relevant technique for an in-depth analysis of the examination results. Forthwith, studies show that after a number of exacerbations (Nowak, 2006) Spirometry examination is the most important factor for the evaluation of respiratory performance.

Due to the fact that IgE is examined only before treatment starts, Spirometry is the only examination that can reveal the patients' lung function and overall respiratory health (Bousquet et al, 2007). Also based on previous research, (Tzortzaki et al, 2012) the only factors that are important and are taken into consideration are Forced Vital Capacity (FVC), Forced Expiratory Volume 1 (FEV1) and Peak Expiratory Flow (PEF). However not all authors take into account PEF measurements, for the evaluation of patients' respiratory function (Tzortzaki et al, 2012). Although the research of this paper will not only focus to only to these three elements. Contrariwise in this paper PEF factor and all factors that are provided by a Spirometry examination will be considered.

A diagnosis is made by a practitioner and improvement of respiratory performance is evaluated. This occurs after a procedure that entails the evaluation of a patients' health, the number of exacerbation and the percentage of FEV1 and FVC according to the 'normal' measurement which is provided by a spirometer.

Previous research has shown (Korn et al, 2009) that secure and accurate measurement of results can be assessed after at least 16 weeks of treatment. On the other hand only 12 weeks of treatment can create a secure overview of treatments' outcomes (Slavin et al, 2009).

2.5 The Proposed Scenario for Examination of Spirometry Elements

In this paper we consider the following scenario for examination of Spirometry elements: each patient has a number for Spirometry examinations taken on different dates. Each Spirometry examination consists of elements which measure parts of the respiratory system. Elements such as FEV1 represent the percentage of function of different part of human respiratory, according to the 'ideal' function that each individual should have. The 'ideal' is created based on each individual personal characteristics, which will be analysed below. All abnormal (parameters) elements, which will be analysed below, will be added to the proposed technique. After simulation of the technique to the Spirometry data a combination of abnormal elements (parameters) will be revealed.

This procedure approaches each patient separately, as a different database. Each patient provides an association of dysfunctional elements. Following a statistical analysis is taking place which leads to associations of abnormal Spirometry elements that are met to most of patients. The following issues could be addressed: patients could have different number of days for examinations; different Spirometry elements but some could be the same as in the previous examination days.

3 ASSOCIATION RULES AND SPIROMETRY DATA

Quantitative data that are relevant to measurement of Spirometry elements, are used for a consideration of examination results. They are divided into before and during a treatment period. The results are then further analysed in order to make the final judgment on the patients' health improvement. This work examines Spirometry data to find if there is an association among Spirometry abnormal parameters. Identified associations could be used for better diagnosis and/or Asthma prediction.

This section presents well known techniques of 'data mining' which are used in analysis of associations from the perspective of application to Spirometry data. 'Association analysis' is used for discovering interesting or uncovered relationships hidden in data sets. 'Association rules' are used for data of multiple nature/types and each one of them has different criteria for significance of outcomes (Lee et al, 2005), (Guang-Yuana et al, 2011).

Let us define the following key aspects: let $I = \{i_1, i_2, i_3, \ldots, i_n\}$ be the set of all Spirometry elements, e.g. Forced Expiratory Flow when 50% and 25% of total air are blown (MEF 50, MEF 25 respectively), Forced Expiratory Flow between 25 percent and 75 percent of the Vital Capacity (MMEF 75/25), FEV1 (i.e. items) in a Spirometry dataset for one patient during examination period of time and $T = \{ t_1, t_2, t_3, \ldots, t_m\}$ be the set of examination results relevant to Spirometry data elements for one patient (i.e. "transactions"). Each Spirometry examination ti contains a subset of k-items (i.e. a k- itemset) chosen from I.

We are interested in associations of abnormal Spirometry elements and we seek to find association rules that define an implication $A \rightarrow B$, where A is MEF 50 and B is MEF 25. A and B are disjoint itemsets, i.e., $A \cap B = \emptyset$.

The strength of an 'association rule' can be measured in terms of its 'support' and 'confidence'. 'Support' determines how often a rule is applicable to a given i.e., patient Spirometry dataset, while 'confidence' determines how frequently items in B appear in Spirometry dates that contain A. A 'support' measure in Spirometry data identifies rules that have very low occurrence on examination results. For this reason, 'support' can be used to eliminate uninteresting rules in Spirometry data. On the other hand 'confidence' can be defined as the probability of element/item A to be found at the same time in combination with element/item B, where A and B are Spirometry elements such as MEF 50 and MEF 25 respectively. Both 'support' and 'confidence' measures have to be equal or bigger than the one that 'user' has specified as minimum. Otherwise items and combinations of them are not taken into account (Lee et al, 2005). Furthermore 'confidence' measures the reliability of the inference/implication made by a rule. For a given 'association rule', the higher the 'confidence', the more likely it is for an elements/items such as Spirometry elements to be present in a patient Spirometry dataset, that contains elements such as MEF 50 and MEF 25. That is how, 'confidence' determines how frequently items in transactions/Spirometry dates, appear in a patient Spirometry dataset that contains dysfunctional elements.

Subsequently, three 'association rules' techniques are further analysed. The reason for this selection is because they are the most well-known and used 'association rules' techniques (Cokpinar and Gundem, 2012). These are: Apriori Algorithm, Frequent Pattern (FP) Growth and Sampling algorithm. Each one has advantages and disadvantages, which are further discussed, although it is highly important to adopt the nature/type of Spirometry data and the nature of the problem for the production of a significant outcome.

3.1 Apriori Algorithm

Apriori Algorithm is a procedure that scans the frequent Itemset of Spirometry data to reveal associations that satisfy minimum 'support'. Consequently this algorithm generates new Candidates and re-scans the initial database to find more associations among elements/items at one at a time/scan (Yu et al, 2010). Finally it terminates when no more associations/combinations can be found among elements. For example it checks only for element A if it satisfies minimum 'support' then for B, C and so on. In its second scan it examines for elements that have passed minimum 'support' in the first step only. It generate combinations of A and B, A and C and it continues for all elements of the sample, where A, B and C are elements such as MEF 50, MEF 25 and MMEF 75/25 respectively. In a new candidate combinations of two elements that satisfy minimum ‘support’ are stored. In its next scan it searches for combinations of 3 elements and stores them again. This procedure continues until no more combinations of abnormal elements that satisfy minimum ‘support’ can be found. Consequently all Itemsets that do not satisfy minimum ‘confidence’ are abandoned. Its final outcome is a set of combined Items that fulfil minimum requirements.

One of its great advantages is that it has great attention to detail and increases accuracy accordingly to sample size. Improved accuracy can be achieved to bigger samples. (Ykhlef, 2011).

One of the drawbacks of the Apriori Algorithm that has been observed by researchers is (Umarani and Punithavalli, 2011) the need of an entire Dataset scan for each item added on the Itemset. This results in slow implementation and high execution time. Moreover because of its delay more electricity is needed and because of technical equipment such as Computers and storage space there is an increase of cost (Liu et al, 2012). Another problem that is encountered is the need for additional storage space and this is due to the creation of new Itemsets. Storing is significantly important for future analysis and evaluation through Apriori. As a result, acquisition of technical equipment (Computer, Storage space) will lead to an additional cost and may be unaffordable. A basic influential factor is the size of the Itemset that is used for analysis (Nahar et al, 2012). This problem is also encountered if there is a small ‘support’, set by the ‘user’, which will lead to the creation of multiple new Itemsets (Umarani and Punithavalli, 2011). Consequently, the smaller the sample the faster and cheaper the implementation will be.

3.2 Frequent Pattern Growth Algorithm

In Frequent Pattern Growth Algorithm the procedure to solve the problem is divided into two steps (Umarani and Punithavalli, 2011). In the first step the database is scanned only once and an FP-tree is created. In each branch of FP-tree Items /Spirometry elements from dataset/Spirometry examination results from the dataset are added and stored by their names. In every new element a new sub-branch is created and stored with a Transaction ID (TiD). A unique ‘path’ of each transaction is also created. The size of FP-tree is at least as a big as the original Database.

In the second step it recursively mines all the patterns from the FP-tree that was built in the first step and concludes on the result. It follows the prefixed path based on a specific search e.g. path that contains a specific element as FEV1 does. This leads to results that satisfy ‘support’ by exclusion of TiD that do not satisfy ‘support’ and ‘confidence’.

The creation of smaller structures (FP-tree), if not the same size as the entire database, is analysed (Chen et al, 2011) is one of the advantages of FP-Growth Algorithm. Also it creates a methodology that does not demand the creation of multiple Candidate Itemsets. Finally it reduces the search space that is needed because of its methodology. Therefore only two patterns of development are needed, with smaller execution time (Lin et al, 2011). This results in faster analysis of Spirometry examination of patients who have large amount of examinations.

One of its drawbacks is that it needs the same amount of storage and memory capacity as Apriori does (Liu et al, 2012). It also has high cost because of high demand of technical equipment (Computer, Storage space, Electricity). Moreover there are crossing authors opinions about Aprioris’ implementation speed (Ke et al, 2013), (Umarani and Punithavalli, 2011). Meanwhile it has high efficiency; its complexity in each step (Duneja and Sachan, 2012) results in the presence of a great barrier for individuals/practitioners who lack the experience in modelling aspects; to calculate findings and execute proper analysis of data.

Finally every new Spirometry dataset of each patient is added to the previous database. Some patients may have thousands of Spirometry examinations which could lead to a massive Database. As it has been discussed, FP-growth is based on the development of a FP-tree. This leads to bigger samples which have reduced mining performance (Lin et al, 2011). Thereafter there are some patients with a lot of Spirometry datasets. Lower mining performance which can lead to lower significance of results is not desirable for nature of Spirometry Data and Health research that needs high accuracy.

3.3 Sampling Algorithm

Sampling algorithm does not have a specific structure as Apriori and FP-growth algorithms do. Instead it is based on a different algorithm, it is based on the nature of Data. By appropriate algorithm it generates an accurate sample size for implementation initially. It provides an outcome significance and accuracy of sample size. If it is not appropriate it is modified and repeats this step. After identification of appropriate sample size, the final outcome is provided. Analysis of the sample takes place by other Association Rules techniques such as Apriori or FP Growth Algorithm (Chen et al, 2011). As is revealed by its name it uses a sample of the entire database and analyses it (Umarani and Punithavalli, 2010).

Its methodology is divided into a number of steps that are depended on the algorithms that will be used for implementation.

Initially the selection of the sample takes place during step one (Umarani and Punithavalli, 2010). The selection of the sample that leads to accurate results differs in each type of data. The most efficient technique can be revealed only after its implementation. A random selection from the database can be found if it satisfies 'support' by implementing it by giving a slightly lower minimum 'support' than the one that the 'user' specified (Chen et al, 2011). If it is not satisfied then more Data are being added to the sample and the sampling technique is re-implemented. The procedure terminates when the sample satisfies minimum 'support' of the 'user'.

Consequently the sample is analysed by an 'association rule', like Apriori or FP-growth and results are provided after implementation (Ackan et al, 2008).

Advantages that have been identified are the decreased use of storage space, the low cost because of the lower demand of technical equipment (computer, storage space). Moreover it is implemented faster than the total analysis of the database because less Data have to be analysed (Umarani and Punithavalli, 2010).

Each algorithm has particular strengths and weaknesses, relevant to the nature/type of data that are being used. The same rule applies for the Sampling Algorithm on Spirometry data. Although the Sampling algorithm has disadvantages on its own thus the possibility of data that the sample is consisted to be similar. This could result in faulty outcomes of the analysis. Moreover quality and accuracy of outcome depends on the database size. Databases of bigger size have higher accuracy and better quality than smaller ones. This is not appropriate and effective on nature/type of Spirometry data because some patients have plenty of examinations and others have few.

3.4 Critical Comparison of Techniques in Relation to Data Needs

Apriori and FP-Growth have high efficiency, quality of outcome and pay attention to detail. These advantages do not apply to the Sampling algorithm. On the other hand they have high cost and long implementation time, which can lead to delayed results.

For an outcome of high significance medical data need to have accuracy of results. As a result, the most important factor for the selection of the most appropriate algorithm is based on its accuracy, quality and attention to detail of the data. Each sample method needs to be implemented in order to find the most suitable for Spirometry type of data. Based on these requirements, Sampling Algorithm is excluded because it lacks of these requirements. Moreover, the need for mathematical and data analytical knowledge for implementing the sample algorithm process, are additional reasons for not selecting this algorithm for the analysis of this type of data.

Likewise FP-growth will not be selected for analysis of Spirometry data because of its high complexity during implementation. Furthermore exceptional mathematical and 'data mining' knowledge is required by the 'user'. Also it has similar cost as Apriori.

Concluding based on the above comparisons, the best suited algorithm that will be used for analysis of our Data is the Apriori algorithm. The reason for this decision is that it fulfils limitations such as accuracy and quality of outcome in better significance than the FP Growth and Sampling algorithm. Moreover the ability to analyse big Datasets, based on size, it affects positively the outcome (Amato et al, 2011). The larger the sample the more accurate the outcome. Also it is important to the 'user' because multiple stored data (Transactions) from previous years are needed for analysis. So as time passes, the sample becomes bigger. Furthermore it can be adopted by Spirometry Data with no limitations.

The Apriori algorithm is used as a guide for finding associations among elements of Spirometry data that are dysfunctional/abnormal (<80%) such as MEF 50 and MEF 25. Additionally 'support' measure helps to prune candidate Itemsets discovered during frequent Itemset generation on Spirometry dates/transactions.

4 APPLICATION OF APRIORI ALGORITHM TO REAL-LIFE SPIROMETRY DATA SAMPLE

4.1 Sample Characteristics of Patients

All Data were collected from the General Hospital of Ioannina 'Hatzikosta' in Greece. The sample that was chosen for analysis consists of 20 patients; 13 of which are women and 7 are men. All patients' are residents in the region of Epirus and are under treatment with Omazulimab in the General Hospital 'Hatzikosta'. This sample of 20 patients is consisted by 559 Spirometry examinations that were taken for analysis; 12 of which (2, 14% of total sample) have not been taken into consideration for analysis due to the lack of total Spirometry elements indications. This was caused by mechanic failure of the Spirometer. Moreover examinations before the 12^{th} week of treatment are not being taken into consideration as explained below (10, 7% of total sample). Furthermore each patient had a different number of examinations. The average number of examinations per patients is almost 28 (27, 9 is the exact number) with the minimum number of examinations being 9 per patient and maximum being 60 per patient.

Moreover the average IgE level of our sample is 535, 65 with lowest being 140 and the highest 1476. Also the average age of our patients was 64 (63, 8) with youngest being 48 years of age and oldest being 78. Finally the average treatment period of sample was almost 21 months (21, 15 is the exact number) with shortest being 7 and longest period being 40 months per patient. Although averages have been computed for the total sample, a portion of our patients (10 out of 20) are still under treatment and others have completed the treatment. Finally the examinations that have been collected are from July of 2001 until July of 2012. Finally 10 out of 20 patients have completed the Omazulimab treatment. So 'after treatment' period examinations were not further analysed for higher significance of results.

4.2 Process Followed for Analysis of Spirometry Data based on Study and Disease's Requirements

Data were collected from 20 patients in three periods: Spirometry examination before Omazulimab's treatment, after 12 weeks of treatment and after treatments' completion. The first period consists of Spirometry examinations that have been taken before Omazulimab's treatment, 31, 3% of the sample (175 out of 559) as shown in Table 1. The percentage of each element, on each patient who has not proper function is presented.

The second period is composed by Spirometry examinations taken after 12 weeks of treatment until the completion of the treatment and represents 51, 8% of total sample. The effectiveness of Omazulimab is visible and proves the previous researches (Slavin et al, 2009). In the second table it is presented 'during treatment' period and shows the percentage of examinations that each Spirometry element is abnormal (>80%).

The last period consists of Spirometry examinations taken after the treatments' completion. Although this period is not taken into consideration for analysis due to lack of examinations, as only 50% of the sample (10 out of 20 patients) has completed the treatment. Something that is translated to 6, 08 % of total sample examinations (34 out of 559).

Each element given by Spirometry as healthy and functional has to have at least 80% function performance for the 'ideal' measurement (National Heart Lung, and Blood Institute, 2007). This is given for each element separately by the Spirometer. Only elements that fulfilled this requirement were taken into consideration. The identification of associations among elements that are not fully functional and are responsible for this disease is being researched.

Information about each patient is presented by a dataset, from where associations of elements occurred after analysis of Spirometry examinations. Each date of a Spirometry examination was set as a Transaction and elements in it were set as Items. Below 80% were taken into consideration (set as 1) and above 80% were not (set as 0). Following this allocation each patients' Spirometry examinations (dataset) have been imported and analyzed with Apriori. The process that has been followed is briefly described in section 3.1. After this analysis combination of abnormal Spirometry parameters for each patient has been discovered. Each patient has been analysed twice as different Datasets for before and during treatment period were used.

The provided technique has only been used as an additional technique for practitioner to make more accurate diagnose and evaluation patients' health on top of other Medical findings.

'Support' measurement has been set to 66, 6% for having significance of result, as abnormal elements have to be more of a half times in examinations sample. Additionally 'confidence' measurement has been set to 100% because the

Table 1: 'Before' treatment period.

Patients Number	VC IN	VC EX	FEV 1	FVC	FEV 1 % VC MAX	MEF 75	MEF 50	MEF 25	PEF	MMEF 75/25	FVC IN	Out of
1	22.72	22.72	63.63	9.09	22.72	14	100	100	22.72	100	22.72	22
2	0	0	0	0	0	0	100	100	0	100	0	2
3	4.76	38.09	28.57	28.57	4.76	66.66	100	85.71	57.14	100	0	21
4	37.5	37.5	37.5	37.5	12.5	50	75	100	25	75	37.5	8
5	50	50	50	50	25	50	75	100	50	75	50	4
6	78.94	94.73	100	94.73	31.57	100	100	100	100	100	78.94	19
7	12.5	0	0	0	0	50	100	100	25	100	12.5	8
8	50	0	100	0	0	50	100	100	50	100	50	2
9	0	0	0	0	0	0	100	100	0	100	0	1
10	0	0	33.33	0	0	33.33	100	100	0	100	0	3
11	38.46	30.76	38.46	26.92	19.23	65.38	61.53	80.76	42.3	69.23	38.46	26
12	0	0	0	0	0	0	100	100	0	100	0	1
13	0	0	0	0	0	77.77	66.66	55.55	66.66	66.66	11.11	9
14	50	50	50	50	0	100	100	100	100	100	25	4
15	50	50	50	25	0	25	50	50	0	75	50	4
16	5	5	5	0	0	30	80	70	15	85	5	20
17	100	100	100	100	100	0	100	100	100	100	100	3
18	66.66	33.33	33.33	33.33	0	33.33	66.66	100	33.33	66.66	66.66	3
19	100	100	100	100	40	100	100	100	100	100	100	5
20	10	20	10	20	0	80	100	100	70	100	10	10

Table 2: 'During' treatment period.

Patient Number	VC IN	VC EX	FEV1	FVC	FEV1 % VC MAX	MEF 75	MEF 50	MEF 25	PEF	MMEF 75/25	FVC IN	Out of
1	10	20	30	0	0	50	100	100	30	100	20	10
2	0	0	0	0	5	15	70	100	0	90	5	20
3	0	20	20	20	0	40	100	100	60	100	0	5
4	100	100	100	100	9.09	100	100	100	27.27	100	100	11
5	14.29	14.29	14.29	14.29	0	14.29	100	100	14.29	100	14.29	7
6	100	100	100	100	0	100	100	100	100	100	100	6
7	0	0	0	0	0	0	100	100	0	100	0	1
8	54.17	54.17	87.50	45.83	0	66.67	100	100	8.33	100	54.17	24
9	23.53	0	0	0	0	0	23.53	88.24	0	58.82	23.53	17
10	6.25	3.1	9.38	3.13	0	21.88	46.88	96.88	18.75	71.88	12.5	32
11	0	0	5.56	0	0	38.89	77.78	88.89	0	88.89	0	18
12	0	0	0	0	0	100	100	100	0	100	0	5
13	0	0	0	0	0	100	100	100	83.33	100	0	6
14	0	31.,03	27.59	13.79	0	100	100	100	93.10	100	10.34	29
15	0	0	0	0	0	0	0	100	0	0	0	4
16	0	2.78	0	0	0	44.44	94.44	100	33.33	97.22	0	36
17	100	100	100	100	0	100	100	100	100	100	81.82	11
18	37.5	31.25	50	25.00	0	93.75	100	100	0	100	43.75	16
19	86.67	100	100	86.67	86.67	100	100	100	100	100	86.67	15
20	5.88	0	5.88	0	0	70.59	100	100	29.41	100	5.88	17

nature of data demands high accuracy.

Although 20 associations occurred (one of each patient/Dataset) only patients that met minimum ‘support’ and ‘confidence’ measurement criteria were statistically evaluated. And separately for each period which led to associations of Spirometry elements that are mostly met among patients. Patients that failed to meet the required minimum criteria were not considered in the process of statistical analysis for achieving improved accuracy and significance of final results.

Finally a comparison of associations/ combinations, before and during treatment periods has been made. This led to an outcome which revealed the amount of associations of ‘before treatment’ period that have been healed by the use of Omazulimab in ‘during treatment’ period.

4.3 Discussion of Results from Patients’ Spirometry Data Analysis

It has been found that in ‘before’ treatment period there were three patients who although satisfied the ‘support’ indicator through the full process of analysis, the association of Spirometry parameters/elements did not fulfil ‘confidence’ indicator. This resulted in the exclusion from the second step of analysis which is the Statistical analysis of the sample. Also one patient had only one element that fulfilled ‘support’ indicator. This concluded to an end of his/her analysis because there were no associations to be found. Additionally the patient has also been excluded from second step of analysis.

An association of elements that has been found to 50% (8 out of 16) of sample that satisfied ‘support’ and ‘confidence’ with no other elements on it, is consisted by the elements MEF 50, MEF 25 and MMEF 75/25. From which 75% (6 out of 8) had 100% Support. Also this combination has been found in associations with other elements to 93, 75% (15 out of 16) of sample. Also it has satisfied Support and Confidence limitations.

In ‘during’ treatment period two patients had only one element that fulfilled ‘support’ indicator. This concluded to end of analysis because there were no associations to be examined. Also they were excluded from second step of Statistical analysis.

The same association/combination of elements has been found in ‘before’ and in ‘during’ treatment period. It is consisted by MEF 50, MEF 25 and MMEF 75/25 has been found to 44, 45% (8 out of 18) of sample. It also satisfies ‘support’ and ‘confidence’ limitation. Although it has been found at the same height as ‘before’ treatment period 37, 5% (3 out of 8) of patients are different. Furthermore it has been found that associations that have 100% ‘support’ have been reduced to 62, 5% (5 out of 8). Furthermore this combination is met in associations with other elements to 94, 44% (17 out of 18) of sample that satisfied ‘support’ and ‘confidence’.

Another significant finding that has been revealed from this work is the importance of FEV1 for diagnosis of this disease. It has been revealed that although FEV1 is an element that is a prerequisite medical requirement for beginning of treatment, as also a significant measurement for evaluation of it, it is not as dysfunctional as it had been initially supported. The reason of this outcome is because it has not fulfilled ‘support’ indicator to most of patients neither ‘before’ nor ‘during’ treatment period. Although it has been met to 25% (4 out of 16) of associations on ‘before’ treatment period and in 22, 22% (4 out of 18) to ‘during’ treatment period, it has not significant appearance. It was expected to be found to all datasets and to be highly significant as it is one of the minimum Medical requirements to begin Omazulimab treatment according to European Medical Agency (Tzortzaki et al, 2012)

Furthermore 100% ‘confidence’ had been demanded because of Data nature. It has been indicated by some factors that there is a possibility for some Spirometry elements/parameters with lower than 100% ‘confidence’ to fulfil needs of analysis. It has been found to some elements and there is high possibility to indicate individuals that suffer from severe allergic Asthma inadequately controlled and it is not reducing significance of finding.

4.4 Critical Evaluation of the Outcomes

After implementation of Apriori Algorithm and analysis of sample, we identified an association of elements. It is the same in ‘before’ and ‘during’ treatment period. The association that is mostly on both periods is consisted by MEF 50, MEF 25 and MMEF 75/25 elements. Importance of results is being provided in more detail in section 3, where ‘Support’ and ‘Confidence’ measures are discussed in the context of their application to Spirometry data. The meaning of this finding in health value is that when a patient has low flow speed of air moving out of his lungs at the time of 50% of FVC is blown it has also low flow of air blown when 25% of FVC is blown and his/her small and medium airways are not as functional as they had to be based on the ‘ideal’.

Patients who fulfil other medical requirements are indicated by this association of elements as patients who require treatment with Omazulimab. The reason is that they suffer or will suffer in future by severe allergic Asthma inadequately controlled.

It has been found the same association of elements to both periods and revealed that treatment has a positive effect on this association of elements. Level of dependence among these elements became weaker. Elements are not that close correlated after 12 weeks of treatment or more. Furthermore it has been proved that treatment has positive reaction except medical view point also by Spirometry Data point of view.

Encounter to less than 25% of sample, on both periods, of FEV1 has reduced its utility as a minimum medical requirement for start of Omazulimab treatment.

Furthermore associations of elements that provoke severe allergic Asthma inadequately controlled are MEF50, MEF25 and MMEF75/25. This association/combination of Spirometry elements it has been found as a ‘stand-alone’ combination. Although it has been met in associations with other elements in both ‘before’ and ‘during’ treatment periods.

In addition it has been revealed that Omazulimab treatment has positive reaction upon this association as it makes it weaker. Furthermore ‘confidence’ of 100% there is a possibility to be lower to some Spirometry elements. Although this does not affect significance and utility of the result. Finally it has been revealed that FEV1 element has lower significance on disease recognition as also to evaluation of it as it has been concerned by authors in previous section.

5 RECOMMENDATIONS

Initially it has to be clearly understood that the proposed technique is only to be used as an add-on tool for better evaluation of health. Even though it produces an outcome of associations that can forecast and evaluate patient’s health, other health examinations have to be taken under consideration. Additionally the final decision for patient’s health and treatment must be taken after practitioner’s overall evaluation. This occurs after combination of multiple examinations needed as also requirements that have to be filled.

Through this research both the research questions that have been set for investigation have been answered positively. It has been proved that there are data mining techniques that can find ‘association rules’ in Spirometry datasets. Moreover it has been investigated and proved that Apriori algorithm could be used as most appropriate ‘data mining’ technique to discover associations among abnormal Spirometry data. Detailed discussion on how and why this algorithm is the most suitable for this type of data are provided in section 3.

Additionally the proposed tool examines and evaluates patient’s health from Data perspective. There have to be as many as possible Spirometry examinations for increased accuracy. A reason for this is because the algorithms’ process takes into consideration from 2 to millions of examinations for analysis. Moreover as it has been discussed respiratory health is affected by multiple factors such as weather, daytime, age, weight, etc. With a small amount of Data the significance may be reduced but this instance can be overcome by techniques for small data samples such as Kernel’s methods and Support Vector Machine (SVM).

Additionally in regular periods, when a significant amount of Spirometry examinations are gathered, process of comparison of Data from different patients has to be carried out. This is because factors such as climate, global temperature and multiple factors that affect respiratory and cannot be measured, change. This could result in change of associations that provoke severe allergic hardly controlled Asthma. For avoidance of radical changes in the association found it has to be checked in regular periods.

Furthermore because of different environmental characteristics around globe, technique may have to be undertaken to samples from different places of the world. This is for avoiding misleading guidance from technique caused on environmental changes. These and some other aspects could be taken into account and addressed in future work.

Even though this subject it has been approached by a specific perspective it has been revealed that it is ‘open’ to be adopted and analysed by different perspectives.

REFERENCES

Ackan, H., Astashyn, A. and Bronnimann, H., 2008. Deterministic algorithms for sampling count data. In *Data & Knowledge Engineering*, Volume 64, pp 405–418.

Amato, F., Fasolino, A. R., Mazzeo, A., Moscato, V., Picariello, A., Romano, S., and Tramontana, P., 2011.

Ensuring semantic interoperability for e-health applications, In Complex, Intelligent and Software Intensive Systems (CISIS), International Conference on IEEE, ISBN: 978-1-61284-709-2, June 30 - July 2, Seoul, Korea, pp 315-320.

Bousquet, J., Rabe, K., Humbert, M., Chung, K.F., Berger, W., Fox, H., Ayre, G., Chen, H., Thomas, K., Blogg, M. and Holgate, S., .2007. Predicting and evaluating response to omalizumab in patients with severe allergic asthma. In *Respiratory Medicine*, Volume 101, pp 1483-1492.

Chen, C., Horng, S.J. and Huang, C. P., 2011. Locality sensitive hashing for Sampling-based algorithms in association rule mining. In *Expert Systems with Applications*, Volume 38, pp 12388–12397.

Cokpinar, S. and Gundem, T.I., 2012. Positive and negative association rule mining on XML data streams in database as a service concept. In *Expert Systems with Applications*, Volume 39, pp 7503–7511.

Duneja, E. and Sachan, A. K., 2012. A Survey on Frequent Itemset Mining with Association Rules. In *Computer Applications*, Volume 46, pp 18-24.

Guang-Yuana, L., Dan-Yanga, C. and Jian-Wei, G., 2011. Association Rules Mining with Multiple Constraints. In *Procedia Engineering*, Volume 15, pp 1678 – 1683.

Holgate, S., Buhl, R., Bousquet, J., Smith, N., Panahloo, Z. and Jimenez, P., 2009. The use of omalizumab in the treatment of severe allergic asthma: A clinical experience update. In *Respiratory Medicine*, Volume 103, pp 1098-1113.

Ke, J., Zhan, Y., Chen, X., and Wang, M., 2013. The retrieval of motion event by associations of temporal Frequent Pattern growth. In *Future Generation Computer Systems,* Volume 29, pp 442-450.

Korn, S., Thielen, A., Seyfried, S., Taube, C., Kornmann, O. and Buhl, R., 2009. Omalizumab in patients with databases. In *Computer and Information Science*, Volume 23, pp 1-6.

Lee, Y. C., Hong, T. P. and Lin, W. Y., 2005. Mining association rules with multiple minimum supports using maximum constraints. In *Approximate Reasoning*, Volume 40, pp 44–54.

Lin, K. C., Lia, I.E. and Chen, Z. S., 2011. An improved frequent pattern growth method for mining association rules. In *Expert Systems with Applications*, Volume 38, pp 5154–5161.

Liu, X., Zhai, K. and Pedrycz, W., 2012. An improved association rules mining method. In *Expert Systems with Applications*, Volume 39, pp 1362–1374.

National Heart, Lung, and Blood Institute, 2007. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. In *National Asthma Education and Prevention Program*, 28 August.

Nahar, J., Imama, T., Tickle, K. S. and Chen, Y. P. P., 2012. Association rule mining to detect factors which contribute to heart disease in males and females Expert. In *Systems with Applications*, in press.

Nowak, D., 2006. Management of asthma with anti-immunoglobulin E: A review of clinical trials of omalizumab. In *Respiratory Medicine*, Volume 100, pp 1907-1917.

Slavin, R., Ferioli, C., Tannenbaum, S., Martin, C., Blogg, M. and Lowe, P., 2009. Asthma symptom re-emergence after omalizumab withdrawal correlates well with increasing IgE and decreasing pharmacokinetic concentrations. In *Allergy and Clinical Immunology*, Volume 123, pp 107-113.

Stout, J. W., Smith, K., Zhou, C., Solomon, C., Dozor, A. J., Garrison, M. M. and Mangione-Smith, R., 2012. Learning from a Distance: Effectiveness of Online Spirometry Training in Improving Asthma Care. In *Academic Pediatrics*, Volume 12, pp 88-95.

Tzortzaki, E., Georgiou, A., Kampas, D., Lemessios, M., Markatos, M., Adamidi, T., Samara, K., Skoula, G., Damianaki, A., Schiza, S., Tzanakis, N., and Siafakas, N., 2012. Long-term omalizumab treatment in severe allergic asthma: The South-Eastern Mediterranean “real-life” experience. In *Pulmonary Pharmacology & Therapeutics*, Volume 25, pp 77-82.

Umarani, V. and Punithavalli, M., 2010. Sampling based Association Rules Mining- A Recent Overview. In *Computer Science and Engineering*, Volume 2, pp 314-318.

Umarani, V. and Punithavalli, M., 2011. An Empirical Analysis over the Four Different Methods of Progressive Sampling-Based Association Rule Mining. In *Scientific Research*, Volume 66, pp 620-630.

Ykhlef, M., 2011. A Quantum Swarm Evolutionary Algorithm for mining association rules in large.

Yu, K.M., Zhou, J., Hong,T.P. and Zhou, J.L., 2010. A load-balanced distributed parallel mining algorithm. In *Expert Systems with Applications*, Volume 37, pp 2459–2464.