Int. J. Functional Informatics and Personalised Medicine, Vol. X, No. X, 2010

Mining of Patient Data: Towards Better Treatment Strategies for Depression

Maja Hadzic*

Curtin University of Technology,
GPO Box U1987, Perth,
WA 6845, Australia
E-mail: m.hadzic@curtin.edu.au

Fedja Hadzic

Curtin University of Technology,
GPO Box U1987, Perth,
WA 6845, Australia
E-mail: f.hadzic@curtin.edu.au

Tharam S. Dillon

Curtin University of Technology,
GPO Box U1987, Perth,
WA 6845, Australia
E-mail: t.dillon@curtin.edu.au

*Corresponding author

Abstract: An intelligent system based on data-mining technologies that can be used to assist in the prevention and treatment of depression is described. The system integrates three different kinds of patient data as well as the data describing mental health of therapists and their interaction with the patients. The system allows for the different data to be analysed in a conjoint manner using both traditional data-mining techniques and tree-mining techniques. Interesting patterns can emerge in this way to explain various processes and dynamics involved in the onset, treatment and management of depression, and help practitioners develop better prevention and treatment strategies.

Keywords: depression; depression treatments; data mining; XML-mining; data analysis; personalized care; personalized treatment.

Reference to this paper should be made as follows: Hadzic, M., Hadzic, F., and Dillon, T.S. (2010) ‘Mining of Patient Data: Towards Better Treatment Strategies for Depression’, *Int. J. Functional Informatics and Personalised Medicine*, Vol. X, No. X, pp. X–X.

Biographical notes: Dr Maja Hadzic is a Postdoctoral Research Fellow at Curtin University, Perth, Australia. She has a multi-disciplinary background and capacity to work across different domains. Dr Hadzic has successfully completed her PhD in Health Information Systems from Curtin University, and a Master’s degree in Biochemistry from the University of Antwerp, Belgium. She has made significant contribution in the area of use frontier information technologies in the studies of human disease and mental health. Her work has resulted in a number of valuable publications including book, book chapters, journal papers, conference papers and invited talks at the leading international conferences.

Dr Fedja Hadzic received his PhD from Curtin University of Technology in 2008. His PhD thesis is entitled: Advances in Knowledge Learning Methodologies and their Applications. He is currently a Research Fellow at the Digital Ecosystems and Business Intelligence Institute of Curtin University of Technology. He has contributed in a number of fields of data mining and knowledge discovery, and published his work in several refereed conferences and journals. His research interests include data mining and AI in general, with focus on tree mining, graph mining, neural networks, knowledge matching and ontology learning.

Professor Tharam Dillon is internationally recognised for his research on Semantic Web, Web services, knowledge discovery, and data mining, neural networks, intelligent systems, object-oriented systems, communications, fault tolerant systems, and distributed protocol engineering. He has published 12 books, over 700 research papers as book chapters, journal articles, and refereed conference papers. His research has received over 4,500 citations with a Hurst index of 31 (Google scholar). His research has made significant contributions to a number of application areas including logistics, banking and finance, electrical power systems, telecommunications and management.

1 Introduction

Depression is rapidly emerging as one of the major problems of our society. The World Health Organization predicted that depression would be the world's leading cause of disability by 2020 (Lopez & Murray, 1998). Depression is often a causal factor in many chronic conditions such as diabetes, hypertension, and HIV/AIDS, resulting in a higher cost to the health system (Horvitz-Lennon *et al*. 2006).

We are noticing a spread of a depression epidemic throughout the whole world. Usually, an epidemic, such as a swine flue epidemic, has a pathogen associated with it. But there is no pathogen involved with the depression epidemic. The precise causes of depression have not been identified yet. Nevertheless, it is clear that a large number of various biological, psychological and social factors and their interactions are responsible for the onset of depression (Patel *et al*. 2007). The interactions, relationships and social factors play an important role in the onset of depression.

On one side, we have a very unique case of depression in each individual. No two cases of depression are the same because individuals differ in their background and previous experiences, in their knowledge, in their thought patterns, in their life circumstances, and similar. While two different cases of depression may resemble to some degree, there is a need to approach each patient individually as a unique case.

On the other side, experts are familiar with the fact that certain types of behaviours are specific for certain patient profiles. For example, a study by Jinich *et al*. (1998) showed that over one quarter of gay and bisexual men reported a history of childhood sexual abuse. Through the collection of personal information and analysis of this information in a collective manner, we would be able to derive patterns specific to certain patient profiles. Human efforts are inadequate to carry out this enormous task while powerful data analysis techniques, such as data mining, are able to collectively analyse the large amount of patient information.

The solution presented in this paper aims to establish a balance between the two extremes, namely, the need to approach every case of depression with a fresh method due to its unique nature, and the need to analyse all information about depression simultaneously and collectively. This solution involves collection of personal information from both patients and therapists with help of daily diaries, SQR20 questionnaires and sensors. The patient information will not only be used for development of personalized therapies but also merged with the information from other patients for the purpose of collective data analysis and deriving patterns. The results of this data mining step will be new knowledge and increased understanding of depression. This information will provide the evidence on basis of which highly personalized and effective therapies can be developed. Nevertheless, the development of highly personalized therapies does not only require knowledge about depression and its treatments, but it also requires clear understanding of, and insight, into a patient’s case combined with discernment abilities of the therapist and the wisdom to make right choices. By no means are we at a stage where data mining programs have replaced humans.

This paper is structured as follows. In Section 2 we discuss some major obstacles in treating depression. Section 3 represents a framework of the integrated system designed to collect and mine patient’s information. Section 4 illustrates use of the system in specific scenarios. Significance of the system is discussed in Section 5. The paper is concluded in Section 6.

2 Obstacles in Treating Depression

In this section, we discuss the main obstacles in development of effective treatment approaches for depression:

- multivariate causes of depression

- variable effect of medications
- variable effect of psychotherapies
- incompetent therapists
- lack of a patient monitoring system outside hospitals

2.1. Multivariate causes of depression

Despite major medical advances, the identification of the factors responsible for the onset of depression and their interactions still remains unsolved and is therefore a very active research focus today (Craddock & Jones, 2001). Researchers now converge on the opinion that depression is caused by a number of interacting biological, psychological and social factors (Patel *et al*. 2007). *Biological factors* include genetics, substance abuse, brain defects or injury, prenatal damage, poor nutrition, exposure to toxins, etc. Severe psychological trauma such as emotional, physical, or sexual abuse, loss such as the loss of a parent, neglect, difficult temperament and similar have been identified as major *psychological factors*. There is also clear evidence that *social factors* including various family, school and community factors such as death of a family member, a dysfunctional family life, academic failure, bullying, transitions, exposure to violence etc. may cause depression in a large number of cases.

As the exact causes of depression are unclear, development of precise treatment strategies is difficult at this stage. The system we present in this paper will help systematic data collection and analysis, gain new knowledge with the help of powerful data mining techniques and provide evidence for the development of efficient and effective depression treatment therapies.

2.2. Variable effects of medications

A number of medications are on the market but a large body of scientific evidence exists to show that their effectiveness is variable (Pacher & Kecskemeti, 2004; Check 2004; Friedman & Leon 2007; Werneke *et al*. 2006). Specifically, the negative effect on cardiovascular health is evident (Pacher & Kecskemeti, 2004). Additionally, sexual dysfunction (Werneke *et al*. 2006) is another side effect of some of the antidepressants which, in its turn, may create additional stress and new problems in the life of the patient. Suicidal thoughts are the most common and dangerous side effect of the antidepressants. Under certain conditions some of antidepressants even promote suicidal thought rather than the intended opposite (Check 2004; Friedman & Leon 2007).

The mechanism by which most antidepressants actually function to relieve depression is not well understood (Schafer 1999). The incomplete understanding of mechanisms points to the fact that the antidepressants are being developed having a limited set of factors in mind. This may be a reason why the unwanted side effects of antidepressants are so frequent and numerous.

Most commonly the antidepressants contain an active ingredient that regulates levels of neurotransmitters in the brain (Schafer 1999). But what happens in the cases where the abnormal levels of neurotransmitters in the brain are a consequence of some external factors, and not the real cause of depression? For example, a patient may develop depression due to an environmental cause such as relationship breakdown. When a doctor prescribes antidepressants, the patient may be in the danger of developing life-long addiction to this medication. The ‘numbing’ effect of medications may prevent the patient from learning how to deal with life issues by finding short-term solutions in medications when difficulties arise. In other case, where the neurotransmitter levels get naturally restored, continued drug intake has the potential to disturb natural levels of neurotransmitters and to influence the mental health of patients in a negative way. We currently have no means to detect when this natural restoration in the brain takes place. A patient may continue taking medication while this, actually, may be the very reason for the mental instability.

The system we present in this paper will provide the means of systematic and consistent collection of data regarding patients’ mental wellbeing (with the help of daily diaries and SQR20 questionnaires) as well as physical wellbeing (with the help of sensors). This will introduce greater control in the treatment process and enable quick interventions when needed.

2.3. Variable effects of psychotherapies

Experts have classified the psychotherapies under two major groups: hedonic and eudaimonic. Hedonic approaches seek solutions to mental health problems through encouragement of expression of positive emotions and suppression of negative emotions. Eudaimonic approaches emphasise that mere control of positive and negative emotions is not enough in achieving happiness. Eudaimonic approaches are focused on developing and establishing a meaningful life through cultivating autonomy, competence and relatedness (Ryan & Deci 2001) and viewing this meaningful life as the source of joy and happiness.

Lots of current hedonic psychotherapies have failed to provide real solutions. Ryan *et al.* (2008) claimed that the focus on maximizing pleasure and avoiding pain, and the associated prescriptions are dead-end routes to wellness and are often accompanied by selfishness, materialism and objectified sexuality. The authors warn about the misleading nature of hedonic thinking.

The principle of cultivating positive emotions as proposed by Fredrickson (2000) must be viewed and evaluated with hedonic/eudaimonic understanding. If a person practises relaxation therapies (such as imagery, muscle or mediation exercises) and/or activities that will increase pleasant activities, but continue with the destructive lifestyle and habits, the relaxation therapies will not work for a long time. The feel-good approaches are temporarily solutions, or simply, a kind of solution that deals with symptoms rather than causes. It is like a weeding process where the roots still remain in the soil. Moreover, some short-term pleasures can bear long-term negative consequences. Encouraging pleasure-filled lifestyle to mentally ill people can sometimes have negative consequences because the patients usually have distorted understanding of the world and are not able to choose profitable activities for themselves.

Personality affects how we view events and these events, in turn, determine the kind of emotions we develop (Diener & Lucas, 2000). In the case of negative emotions, the focus should not be on emotion expression or emotion suppression but on changing the negative aspect of personality. An approach that focuses on changing the personality rather than changing the emotions is based on the eudaimonic approach. The ultimate goal should be to free the person from any kind of medication-dependency and arrive at the personality where negative emotions will not be developed in the first place, or developed but positively dealt with. Change of personality, often described by positive personal transformation, should be the goal of the most mental health therapies.

Psychotherapists need to embrace eudaimonic thinking and focus on the quality of a person's life and the processes involved in living well. For example, cognitive behavioural and interpersonal therapies that focus on finding positive meaning in life, teaching optimistic view on life, developing healthy lifestyle and habits, encouraging the personal growth etc. are more profitable than therapies with short-term effects. Seeking a temporarily relief from negative emotions is not a solution as the core problem with the personality still remains.

A system developed to collect detailed patient information will help therapists carry out effective therapies in a more controlled way. The eudaimonic-based therapies require psychotherapists to closely work with patients on developing healthy lifestyle and habits, and pursuit of healthy meaning and purpose in life. A computer-based system designed to closely monitor patients' wellbeing can assist in achieving this goal.

2.4. Incompetent therapists

Some evidence suggests that a large number of therapists experience mental health problems and are, consequently, not providing their patients with adequate services (Deutsch 1985, Quattrochi-Tubin *et al*. 1982, Farber 1990).

A very frequent phenomena among psychotherapists is 'burnout'. One of the most prevalent types are the therapists who have lost interest in their work as a result of the burnout (Farber 1990). If these therapists continue to work with patients, it is quite likely that their patients will not receive the best treatments. If therapists do not have enough time and energy to invest in their patients, they may often seek easy solutions such as prescription of medications in the situations where cognitive behavioural therapies would be more profitable. Prescription of medications is often the easiest option for a therapist, but may not be the best possible treatment option for a patient.

Burnout of psychotherapists is a situations in which both parties loose: the patient is not receiving desirable treatment and the therapist is in danger of developing mental health problems her/himself. Quattrochi-Tubin *et al*. (1982) examine mental health of counsellors and service workers, and report that the higher levels of burnout are related with greater job dissatisfaction, greater use of alcohol and prescription drugs and higher illness rates among psychotehrapists.

A study by (Deutsch 1985) showed that a large number of psychotherapist suffer from mental problems themselves. The study involved therapists from mental health centres, psychiatric hospitals, university counselling centres, and similar. The results of this study showed that half of the therapists experienced relationship difficulties and depression themselves, and significant proportion of them were in a therapy. Additionally, large proportion of the therapists was ill and missed work days during the preceding six months. These are quite concerning facts. Is there a place for such a paradox in our society where depression is becoming the leading case of disability (Lopez & Murray, 1998)? Are the people who are suffering depression themselves eligible to guide others in overcoming depression? Is it possible for the people who are feeling depressed to speak enthusiasm into the lives of others? These are quite important questions, and it is our duty to address them appropriately for the sake of patients and our society.

The system we propose will not only document patient’s state but also therapist’s mental health, her/his responses and interaction with the patient. This will provide evidence that will help a third party, such as a committee of experts, evaluate therapist’s actions and performance. This feedback will be used to encourage therapist towards better performance or, in cases where mental health of the therapist is at risk, the therapist will be replaced. This feedback loop will benefit both parties: patients, as they will receive better care, as well as the therapists, as they will be challenged to achieve greater success or saved from greater damage to their mental health.

In conclusion, the computer system will be designed to collect evidence with the purpose of not only serving one particular subclass of a community, such as serving only therapists, but serving wider community members including patients, researchers and general public with the goal of benefiting the whole society.

2.5. Lack of effective patient monitoring and quality interaction with the patient outside the hospital

A step-functional quantum drop in the care of patients is evident once patients leave hospitals. Patients usually receive maximum (24 hours per day) support and care in the hospitals. When they leave the hospital, they are left on their own. The monitoring is then limited to the scheduled appointments with regular intervals varying from 1 week to 1 month apart. Due to the lack of efficient monitoring, support and care outside the hospitals, the patient frequently returns to the hospital. The associated stigma often causes loss of self-esteem in majority of patients, damages their wellbeing and makes their condition even worse (Link *et al*. 2001). Additionally, the treatment cost increases and additional pressure is placed on the mental health budget. It is important to monitor patients once they leave hospitals, and the system described in this paper will help us achieve this goal.

Secondly, we notice a problem of patients who behave normally in a hospital but once found in real-life circumstances, they may start exhibiting unusual behaviors. To be able to provide a complete overview on progress of patients, it is required to design a patient monitoring system that can be used outside the hospitals and can be easily integrated within the real-life environment of patients.

Third problem is that most of the patients seek help only when depression reaches a well-advanced stage. We are often missing data from the early onset of depression. This kind of data may be crucial in providing a better insight in the causes of depression. Medical practitioners, such as general practitioners, may be able to detect first signs of depression in patients and suggest use of a patient monitoring system. By putting a patient monitoring system in place, we will be able to obtain data from the early onset of depression.

3 Monitoring System to Assist the Treatment

Various methods and approaches have been used to collect and analyze mental health data and derive collective knowledge. Statistical analysis such as, principal components analysis (McHorney *et al*. 1993), multivariate analysis (Cooper-Patrick *et al*. 1994) and meta-analysis of related literature (DiMatteo *et al*. 2000) have been able to generate some preliminary knowledge. Nevertheless, statistical analysis are increasingly being replaced by data mining due to automation, greater scalability and reliability, and cost and time efficiency of data mining techniques. A study of 667,000 patients using HealthMiner has illustrated the importance of data mining techniques in revealing novel clinical disease associations (Mullins *et al*. 2006). A limited number of studies have reported application of data mining techniques within mental health domain. These studies mainly focus on a specific aspect of this complex domain. For example, Bayesian neural networks have been used to measure safety of antipsychotic drugs (Coulter *et al*. 2001). A study by Soini *et al*. (2008) have applied a greedy Bayesian algorithm to predict forensic admission among the mentally ill.

In this paper we propose a system that will systematically collect patient data and analyze them with the help of data mining algorithms with the purpose of improving patient treatment. To the best of our knowledge, such system does not exist yet. The system has three clear components (see Figure 1):

(1) *patient monitoring systems* that involves collecting patients' data in the form of daily diaries and SQR 20 questionnaires as well as the sensor data which measure bodily activities in patients.

(2) *therapist monitoring system* collects the data describing interaction between therapists and patients. Additionally, SQR 20 questionnaires are used to collect therapists' mental health data.

(3) *knowledge deriving system* analyses the collected data and derives interesting patterns from them.

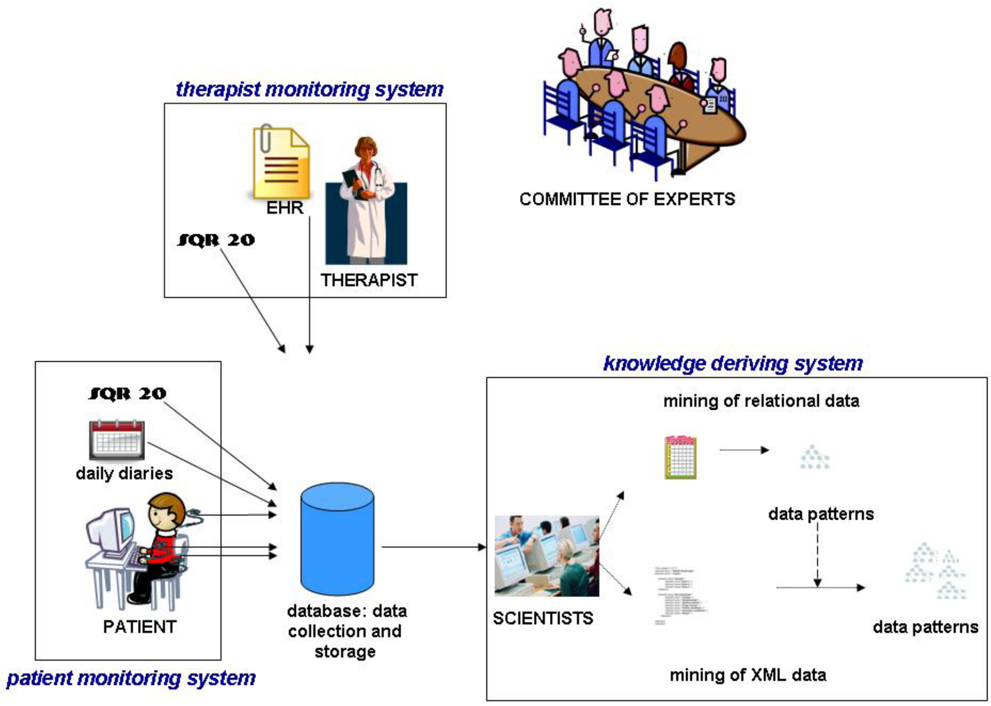

Fig 1. The integrated data collection and data mining system. The system has three components: therapist monitoring system, patient monitoring system and knowledge deriving system.

3.1 Patient monitoring systems

The patient monitoring system collects three types of data concerning mental wellbeing of patients.

The first type includes daily diaries which have forms of tables and are stored as relational data. Various activities a patient engages in herself/himself during the day are represented as attributes in the tables. The patient enters the time during which this specific activity was carried out. The entered time represents the value of this attribute for that specific day.

The second type of data includes online SRQ20 questionnaire (Harpham *et al*. 2003) filled in weekly by the patients. This questionnaire consists of 20 questions and has been specifically designed to evaluate mental health. Each question represents an attribute and the patient's answer determines the value of this attribute for that specific week.

The third type of data involves sensor networks which provide a way for active and continuous monitoring of physical functions in patients. The physical functions are often related to mental functions, and sensor data can serve as objective inputs, especially in the cases where self-reports of patients cannot be trusted. Sensors collect patient data and monitor patients' bodily functions, and include devices such as Continuous Glucose Monitoring Systems (CGMS), heart rate sensor, respiration sensor, and similar. Additionally, brain activity can be regularly measured with the help of practitioners. Electro EncephaloGraphy (EEG) can be used for this purpose. For the sensor data, there is a set of predefined attributes for which the sensors provide values for a particular patient (instance). As such, the raw sensor data are in relational table format where for each attribute considered as important, a value is recorded which forms part of a single instance that is associated with a predetermined time fragment of the day.

The daily diaries, SQR20 and sensor data are streamed to a dedicated database to be further used by therapists and scientists. The Committee of Experts, who will evaluate performance of therapists and progress of patients, will also have access to this data.

3.2 Therapist monitoring system

Due to the high levels of burnout among psychotherapists and the large number of psychotherapists with mental health problems, it is increasingly becoming important to introduce greater control over treatment processes. This is not only important from patients who are in risk of receiving compromised care, but also for therapists who may be in danger of developing serious mental health problems. Committee of Experts will oversee and evaluate the performance of each therapist, her/his mental health and the progress of patients under her/his care.

The therapist monitoring system collects two types of data.

The first type of data is of the form of an Electronic Health Record (EHR). The therapist notes her/his interaction with the patient, prescribed treatment, progress of the treatment, and similar. These data have a combined relational/semi-structured (XML) form. The relational data come from the part of EHR where therapist needs to assign values to certain attributes, such as assigning certain age to the 'patients' age' attribute. Semi-structured data originates from the fields in EHR where therapists need to enter specific notes. The therapists' notes (unstructured text) will be structured under a specific XML tag referring to a specific topic.

The second type of data includes online SRQ20 questionnaire (Harpham *et al*. 2003) filled in weekly by the therapists. This questionnaire consists of 20 questions and has been specifically designed to evaluate mental health. Each question represents an attribute and the therapists's answer determines the value of this attribute for that specific week.

Both the EHR and SRQ20 data are streamed to a dedicated database to be accessed by the Committee of Experts and scientists, and used in further evaluation and analysis.

3.3 Knowledge deriving system

Mining relational data can discover some novel and interesting patterns, but may be limited in the sense that it is not integrated with existing knowledge of the domain, or does not express the discovered knowledge in a meaningful and ready to use way. In other words, the patterns discovered from such data will need lots of validation and classification of the found results to be able to incorporate it with existing knowledge or expectations of the domain. We are therefore using a template that allows one to represent the available data in a more meaningful and organized way in order to perform integrated and complementary analysis of the data. We choose to use an XML based template as it enables one to represent domain information in a more meaningful and specialized way.

Therefore the main data mining process will occur using the techniques capable of mining tree-structured documents such as XML, whereas the techniques for relational data will be mainly used for pre-processing purposes where the aim is to integrate the related data and add useful information to appropriate place in the XML template. In other words, the raw data will be mined with the main purpose of preparing the instance(s) in XML format so that we mine the collective information in a much more meaningful and guided way. Having the integration of all the information together in one template will allow one to detect interesting associations, not just within one specific case, but across different cases, and thereby have the potential to substantially extend the current body of knowledge. Furthermore, it may already be the case that the complementary information resides in an organization in both relational and semi-structured form such as XML. Combining this information into a single source will enrich the information content, which will potentially lead to the discovery of more useful knowledge patterns. Some preliminaries studies presented in (Pan *et al*. 2008), have demonstrated the usefulness of conjointly mining of XML and relational data sources which leads to the discovery of knowledge patterns that could not have been attained if each of the sources was mined separately.

Another advantage of presenting the information in XML is that it will be easier for the domain expert to analyse. While interesting patterns can be extracted from the raw data, placing these patterns in the predefined XML template will allow for a better post analysis of the patterns. It will allow one to associate the discovered patterns with the domain knowledge.

3.3.1 Data pre-processing

Common pre-processing for data mining will take place where we will convert the raw data into predefined categories that exist in the XML template. The pre-processing will first occur at the large set of data that is collected and analysed to discover common knowledge model(s), outlying behaviour etc. The aim is to automate this process as much as possible so that future data is automatically processed into suitable format to be integrated into the pre-defined XML template.

In the pre-processing stage of the data, common techniques such as normalization, discretization, handling of missing values, anomaly detection, and attribute relevance analysis will take place. The normalization and discretization of attribute values is a common pre-processing step, to either make the data suitable for a data mining method used, and/or to make the found generalizations more comprehensible and easier to analyse. The normalization will be performed on any numerical or continuous attribute values, using the min-max normalization method. This method preserves all relationships of the data values exactly. Normalization will be a pre-requisite for the application of a neural network based technique for clustering discussed later in Section 3.3.2. For the application of association rule mining techniques (also discussed in the next section) the continuous attributes will be discretized to ensure that some associations can be formed from such attributes. Attributes with many unique values may not appear as a part of a detected association. Their occurrence is unlikely to satisfy the minimum occurrence (support) threshold used within the general association rule mining framework. By assigning the specific values to a group, the likelihood of a group being frequent increases, and hence the continuous attributes are more likely to become a part of an association rule. Furthermore, the rules with discretized values are usually shorter and easier to comprehend and can lead to improved predictive accuracy (Liu *et al*. 2002). Since in our pre-processing phase the attribute relevance analysis is dealt with separately from discretization (as explained later in this section), we will use the *entropy MLDP* (minimum description length principle) measure. It was empirically suggested in (Liu *et al*. 2002) that this measure is generally the first to consider for discretization purposes.

Missing values may occur when the value for a specific attribute was not collected, or it has been determined to be anomalous, or there was some interference, or error that does not allow us to have certain value for this specific attribute. Also, it may well be possible that there are attributes for which the patient does not want us to collect information because s/he regards it as a violation to her/his privacy. All these factors may contribute to the fact that some attributes may not have values for certain instances. We will use an inference based technique that will fill in missing values based on the most probable value as indicated by the knowledge model learned from the analysed data. A knowledge model will generally represent a set of rules whether in a decision tree fashion or as association rules. The attribute values of the record with a missing value will be matched against the available rules to find the closest rule that describes that particular instance. The value of the attribute that is missing from the record is then set to be the value of the attribute as indicated by the rule. Hence, the approach we use has a combination of the characteristics from the *k*-nearest neighbour technique for missing value imputation (Batista & Monard 2003) and general inference based techniques (Han & Kamber 2006). The difference is that we will neither have to search through the whole datasets as is the case when *k*-nearest neighbor approach is adopted, nor will we build knowledge models to predict the values for each of the attributes that have missing values in the dataset. Rather we will use the the data mining techniques for association rule mining in general context. The record (or groups of records with similar characteristics) with a missing attribute value will be matched against the association rules with the aim to find a rule that most specifically describes the instance. This rule may indicate the association between the missing attribute value and the remaining characteristics of the record, based on which the missing value is filled in. However, it may be possible that the attribute for which the record has a missing value does not appear in the association rule that most specifically describes that particular record. In this case, we will investigate the value of the attribute in the records whose characteristics are also captured by that rule, and assign the most probable value.

It is important to note here, that this method for missing value imputation will be applied in collective studies, i.e. taking the information about each patient as one instance or record and studying the characteristics of the patients as a whole. However, in cases where we want to analyze the change in behavior of patients over time, care needs to be taken not to bias the missing value imputation toward the overall trend of the population under study. In fact, before the general analysis takes place we may also generalize the characteristics of a single patient before using its case as a record in the database reflecting the whole. In these cases, we will need to consider whether the missing value is available in one of the other recorded instances (i.e. sensor data or SQR questionnaires) for that single patient. The importance of this has been demonstrated by a study presented in (Engels & Diehr 2003) which has evaluated a few methods for the imputation of missing longitudinal data. It was shown that methods which do not consider information specific to a particular person will have worse performance. They suggest several strategies based on whether there are record(s) within the longitudinal data before and after the record(s) with missing values. In such cases, we will also adopt a more statistically based approach based on the distributions of singular patients' values. Measures such as *last & next*, *row mean* and *row median* as suggested in (Engels & Diehr 2003), will be used to fill missing values in patients' longitudinal data.

The first strategy that will be employed for **anomaly detection** is to incorporate domain knowledge in form of rules, and to detect any records that do no conform to those set of rules. We will use stand-alone rule optimization technique (Hadzic & Dillon 2008) that takes any rules as input and uses new data to refine and adapt the rules according to future cases. With these characteristics, the method can also be used to detect any records that are contradictory to the initial rule set supplied by the end-user. The method will automatically detect any records that do not conform to the rule such as records from daily diaries where the hours spend on all activities exceed 24 hours. The degree of the deviation from the general rule(s) can be indicated and the exact point of contradiction indicated. This enables the detection of anomalous records and will enable for automatic fixing of data.

In many situations, the set of rules may not be known according to which all anomalies would be detected. Furthermore, the errors that can be found in data are unpredictable and often change with time. Hence it is highly unlikely that all the rules can be incorporated into an anomaly detection technique so that all anomalies in data are detected, as we only have incomplete knowledge about every entity in the real world. These types of anomalies are the most difficult to find and even more complicated to correct because there are no rules (or constraints) which are violated by the records. Hence, we need an additional verification technique to detect any remaining anomalies in the datasets.

A data object that deviates from the norm is referred to as an **outlier**. Outliers in the dataset often correspond to a small percentage of the records. One would hope that inaccurate entries also correspond to a small percentage of records and that they sufficiently deviate from the norm so that they can be revealed as being anomalous. To find these hard to detect anomalies we will make use outlier detection strategy presented in (Hadzic *et al*. 2007). It is a clustering method that automatically learns the norm of behavior from the presented dataset, and detects any records that do not conform to this norm as outliers (i.e. potential anomalies). In some cases the learned rules of the norm could also cover only a small subset of data which indicates that the data objects covered by those rules are suspected of being outliers. By comparing the detected outlying instances or rules with the model exhibiting normal behavior, the method enables one to more confidently distinguish anomalies from true exceptions. For example, if we have a group of 10 depressive patients of a specific depression type and we have 20 attributes of identical values among those patients, then a patient who suffers from the same depression type and shows only 8 out of 20 attributes with identical values, can represent a true exceptional case. Outliers are detected elegantly by forming abstractions of common behavior and thereby easily isolating uncommon behavior.

The degree to which a data object is considered as outlying can be measured using a distance measure between the data object and rule(s) exhibiting normal behavior. The number of attributes in which they differ and the degree of difference in attribute value(s) is taken into consideration. Hence, a user can provide a parameter to the method to indicate the degree of allowed deviation from the norm before an object is considered as outlying, i.e. a potential anomaly. Once any outlying data objects are extracted, we can analyse the specific attributes in which their values are different to verify whether the instance is anomalous, and if so, to correct the instance at this point. In our example of a depressive patient who differs from a group of 10 depressive patients in values of 12 different attributes, the values of the 12 attributes need to be critically evaluated. This needs to be done in consultation with a domain expert rather than just discarding the anomalous records.

In **attribute relevance analysis**, the aim is to identify the most prominent attributes that can represent the mental health situation or state of a patient. While there will be lots of aspects collected from the data, in some specific application aims, not all of them will be relevant. For example, when evaluating effect of television on mental health of patients, attributes such as patients’ ‘name’, ‘address’, ‘height’ are irrelevant. To measure the attribute relevance with respect to some aspect being monitored or learned about, we will be using the Zhou and Dillon’s Symmetrical Tau attribute relevance measure (Zhou & Dillon 1991). The Symmetrical Tau is based on the measure of association from the statistics area, and allows one to measure the capability of a set of attributes or their Boolean combinations in predicting the class or value of another attribute (or sets of attributes). It will be used in 2 ways: (a) the set of relevant attributes will be identified prior to the application of data mining and pattern recognition, and (b) the learned knowledge model will be further improved by removing attributes with no predictive capability for a specific rule. The first approach will disallow the interference of irrelevant attributes in the learning process, which will result in better application of data mining and pattern recognition. The learned knowledge model will be more accurate and specialized toward the aim or the behaviour being monitored. The second approach will further refine this knowledge model by simplifying the rules that indicate particular causes of the behaviour being monitored. For example, in this way we may be able to find causes of insomnia which is a frequent phenomenon in patients who suffer from depression. Overall, by using these strategies the generalization capability of the knowledge model will be increased and it will be more reliable as the means for predicting future alarming situations.

3.3.2 Data mining

Clustering techniques can be used to group common instances together and assign virtual labels allowing us to place these kinds of instances in the specific place in the XML template or link them to an already identified case. In this setting there is not a class to be predicted but rather the aim is to group the data objects (patients) with similar characteristics together and/or to find associations between the data objects. For example, as there exists some evidence for relationship between depression and Internet addiction (Young & Rogers 1998), we may want to group patients who spend a great amount of time on Internet.

For this problem we will use the unsupervised BRAINNE method proposed in (Sestito & Dillon 1994) that extracts symbolic rules from a trained Self-Organizing Map (SOM) which is one of the most popular clustering techniques. SOM is an unsupervised neural network that effectively creates spatially organized “internal representations” of the features and abstractions detected in the input space. SOM groups frequently occurring data object characteristics together in the form of clusters. A number of learning parameters determine the type of abstractions that will be learned and the types of clusters formed. The parameters can be adjusted in such a way so that the clusters formed from the data objects can be either very specific or general**.** In our example of people who spend significant amount of time on Internet, we can make some specific clusters within this group such as people (1) whose work requires spending significant amount of time on Internet, (2) who enjoy social networks, (3) who are addicted to computer games, (4) who are addicted to cyber sex, etc. Establishing more specific clusters will help gain precise insight into relationships between a specific activity and mental health. The reason for choosing to use the BRAINNE method for clustering purposes, as opposed to other clustering techniques, is because the data from the domain under study is expected to be noisy by nature. BRAINNE is a neural network based technique, and neural networks are known to perform better in noisy environments in comparison with other machine learning techniques. In addition, neural networks effectively handle numerical attributes and can be used when little is known about the attributes and their relationships (Han & Kamber 2006). Furthermore, BRAINNE will automatically form clusters based on the distribution of data object characteristics, as opposed to pre-specifying the number of clusters to be formed, as is the case in many partitioning methods, such as *k*-modes (Huang 1998, Chaturvedi *et al*. 2001). This is a crucial property as one cannot predict the number of groups in this domain where the distribution of characteristics is governed by the complex human nature. Last but not least, the BRAINNE method has been demonstrated to be very effective in terms of accuracy, coverage rate and generalization power (Sestitto & Dillon 1994). The acquired knowledge is represented in symbolic form, which allows for further analysis to confirm and extend the current body of knowledge when applicable.

Clustering techniques are also useful for outlier detection purposes. It was mentioned earlier that we aim to use outlier detection and analysis technique (Hadzic *et al*. 2007) for removal of anomalies from the data. The method is capable of distinguishing anomalies from true exceptional cases from the domain. Hence, in our case, the detection of outliers is not only important for cleaning of anomalous data, but is also useful in revealing those rare cases and the types of patients that do not fit any expected model of behaviour. These rare cases often contain relationships and patterns that have not been identified before, because they did not occur in enough examined cases to be considered as an extension to the current body of knowledge. Hence if the outliers are detected and verified to be correct reflections of the real world situation, the relationships/patterns contained in those outlying objects are of crucial importance to the domain. Within the method described in (Hadzic *et al*. 2007) it was necessary to enable a detailed comparison of the characteristics of the different groups detected by the clustering method. This was done for outlier analysis purposes while it has a wider importance in the proposed system, as will be discussed in 2 case studies presented in Section 4. Hence, it will not only be part of the outlier detection and analysis technique within our system, but will also be used to accompany the clustering techniques for a more detailed analysis of the groups. For example, the therapists will often want to match a profile of a new patient with the profiles of existing patients to obtain additional information regarding the correct diagnosis, and/or any exceptional characteristics calling for a different treatment.

A therapist may also desire to compare the data of the same patient that was taken at different times. We will therefore describe the process of comparisons of the records and/or the groups of records in more detail. In the explanation that follows the term record can refer to both a record from a relational database, or an XML document instance.

Let us denote the set of groups formed through the clustering technique as $G = \{g_1, g_2, \ldots, g_k\}$ ($k = |G|$). Each group g_i has a weight associated with it, which reflects the number of records that belong to that group. Hence each group is represented by a set of attribute values denoted as $g_iA = \{g_ia_1, g_ia_2, \ldots, g_ia_m\}$ where m is equal to the number of attributes in the database. The Euclidean Distance (*ED*) is commonly used in the clustering techniques, and we will also use it to indicate the differences among groups, denoted as $ED(g_i,g_j)$. Hence, for any two or more given groups the user can be provided with the information regarding the number of attributes that they differ in and the the exact attribute value(s) in which they differ. For example, let us say that a group g_3, is closest to the group g_4, and it differs in the value of a single attribute a_2,.then $ED(g_3,g_4) = 1$ and $ED(g_3, g_i) > 1$ ($\forall\ i = (1, \ldots, k)$ and $i \neq 3$ and $i \neq 4$)). If the user is looking for a group that is closest to group g_3, then the method will return the group g_4, and indicate that there is a difference between g_3a_2 and g_4a_2. The search will be user controlled to specify the amount of allowed difference as well as the specific attributes in which the groups are allowed/disallowed to differ in. This information can be of great use to the therapists as will be discussed in case studies 1 and 2 in Section 4.

Association rule mining techniques will be used in combination with the BRAINNE method to even further detect the characteristics within the formed groups. Using this method enables us to detect the characteristics of the monitored patients that are frequently associated together, and this information serves as another basis for categorization of the cases. For example, spending significant amount of time on Internet and lack of physical excises are two characteristics often associated with each other.

Another aspect of association mining that will be necessary in our work is that of finding of associations that occur across transactions. As the information is collected for a predefined fragment of time, a time attribute will be associated with each record/transaction. We expect that interesting association exist between the attribute values over time and can indicate important aspects such as the mood/behaviour change over time. The work in (Feng *et al*. 2001) has provided a framework to detect the associations that occur across transactions which is important for domains where dimensional attributes can be associated with each transaction or record. Hence in our domain the dimensional attributes will correspond to time and possibly location, while others may be identified during the analysis process.

The aim will be to discover the associations that can be used to predict the state of the patient over time. Hence, historical data will be analysed where it will be possible to identify patterns and chain effects in the behaviour or environmental factors that lead to a particular state or undesired mental state. To take all these factors into account it is necessary to mine for associations that occur across transactions, as is the case in the inter-transactional association rule mining method presented in (Feng *et al*. 2001). This involves one in defining augmented transactions, concatenating transactions corresponding to different times, as well as contiguous time intervals using a moving window cover in time intervals. The extracted inter-transactional association rules can then be used to predict the potential alarming state of the patient when the part of the rule (association) occurs in real time. This information allows one to react on time to prevent a negative event from occurring, and resorting to particular resolution strategies before the event actually occurs.

It is common that **XML** is used for exchanging data between heterogeneous data sources. It was mentioned earlier, that within our system the XML format will be used to represent the combined information from the different sources and to semantically annotate and contextualize the available information. The data objects in an XML document are hierarchically organized, and we therefore require the use of data mining technologies capable of discovering useful associations among tree-structured data objects. An XML-enabled framework for mining of association rules in XML repositories was first presented in (Feng *et al*. 2003) where the rules extracted are more powerful than traditional ones in expressing association relationships, at both the structural and semantic levels. To extract such rules the prerequisite and the most difficult task is to find all the frequent subtrees from an XML database. This is known as the frequent subtree mining problem and can be defined as: Given a tree database T_{db} an minimum support threshold (σ) find all subtrees that occur at least σ times in T_{db}. Since many of the different types of information will be stored in the XML database, it will be necessary to mine all different subtree types using different support definitions. This will allow for a detailed analysis of the available information. For this purpose we plan to use the available algorithms developed within the Tree Model Guided (TMG) candidate subtree enumeration framework (Tan *et al*. 2006, 2008; Hadzic *et al*. 2010). A preliminary study of the application of the frequent subtree mining algorithms to mental health data has been provided in (Hadzic *et al*. 2008, 2009).

4. Case Studies

4.1. Case study 1: assigning a treatment to a particular patient

Clustering techniques can be used to cluster patients in different groups according to common characteristics. Furthermore, prediction techniques can be used to first discover a knowledge model according to which the patients are classified, and then use this model to predict the class of a new patient. If the classes are formed according common symptoms, then a new patient with similar symptoms can be classified under this group. The therapist can view treatments and effects of these treatments on patients with similar symptoms, and factor in this knowledge when choosing treatment for a new patient.

Once the therapist has chosen a treatment approach for a new patient, s/he can also examine the group of patients who underwent the same treatment and evaluate their progress. In some situations, the information obtained from other patients with similar symptoms and patients undergoing similar treatments can have extreme value.

Optionally, the therapist can match the profile of a new patient against existing patients' profiles in the databases. In some cases, the therapist can obtain some additional information that will help her/him correctly diagnose a patient. This is specially the case in patients who are not willing to disclose some information that may be crucial in understanding the case of this particular patient. This information may include some sensitive information necessary for correct diagnosis. Matching behavioral patterns of these patients against the existing records in the database may reveal this crucial information. The therapist needs then to sensibly approach the patient and examine the case. In other cases, the therapist can view all the patients with similar profile and choose the best possible treatment for this particular patients' profile. This will be enabled by the clustering and the group comparison process described earlier in Section 3.3.2. The difference here is that the clusters will first be formed form the existing data. The characteristics (i.e. attribute values) of the profile of a new patient will be considered as a new group g_{new}, to be compared with existing groups. Hence, given the set of groups $G = \{g_1, g_2, \ldots, g_k\}$ ($k = |G|$), if $ED(g_i, g_{new}) = 0 \;\exists\; i = (1, \ldots, k)$, then the therapist may consider applying the same treatment that was applied in the cases of the group g_i. Otherwise, if $ED(g_i,g_{new}) > 0 \;\forall\, i = (1, \ldots, k)$, the therapist will search for the most similar group(s) and will investigate the characteristics (i.e. attribute values) in which the profile of the new patient differs to the existing cases in the database. The therapist will then also consider the importance of the differing attributes in order to determine whether it is safe to apply the same treatment.

As mentioned previously, the new knowledge derived with the help of data mining techniques can be helpful in gaining insight but is not sufficient evidence for the therapists to base their decision upon it. The clear understanding of, and the insight into the patient's case, as well as wisdom and discernment abilities of the therapist are other key ingredients in the development of effective therapies.

4.2. Case Study 2: monitoring patient's response to a specific drug

The data obtained through daily diaries, SQR20 questionaries and sensor data can be combined to evaluate effects of a specific drug on patients. Frequently, same drugs have different effects on different people. For this reason, it may be more useful to compare data of the same patient at different times than to compare data from different patients.

Firstly, data before and after administering of drugs will be compared (pre- vs. post-administration data from Figure 2). Secondly, we will compare data at the early and at the later stage of treatment (early vs. later administration data from Figure 2). The second comparison is done for the reason that some drugs may have beneficial effect at the start, but be damaging over a prolonged period of time. This scenario corresponds to the case where a predefined fragment of time is associated with each transaction. As mentioned earlier in Section 3.3.2, the method presented in (Feng *et al.* 2001) deals with the problem of inter-transactional association rule mining, and it is one of the techniques used in this case. Hence, the dimensional attribute will be the pre-defined fragment of time which corresponds to a stage of treatment. The extracted inter-transactional association rules could then indicate positive/negative effects of medication as that patient progresses through the different stages of treatment.

It is also required to examine the effects of drugs on a patients' behaviours. A patient may report positive effects but these effects must be evident and recognizable in the patient's activities as well. Drugs should make patients feel good, but most importantly, drugs should affect patients' lives in a positive manner. For example, opium may make someone feel good temporarily, but there is a danger that this person will exchange natural feeling of life satisfaction for the drug-induced feelings of satisfactions, and consequently abandon all daily responsibilities. Monitoring of patients' behaviours is a crucial aspect in evaluating effects of drugs, especially in eudaimonic-based therapies. To aid in this process, we will also use the clustering and the group comparison process described in Section 3.3.2. The groups will be formed from the records of a single patient that were taken at different times of the treatment. In contrast to the aim in case study 1, the therapist will perform the group analysis with a stronger focus on the attributes where the change is expected or desired/undesired. The therapist can then investigate the difference among the groups in terms of the differing attribute values to monitor the behavioral change in a patient.

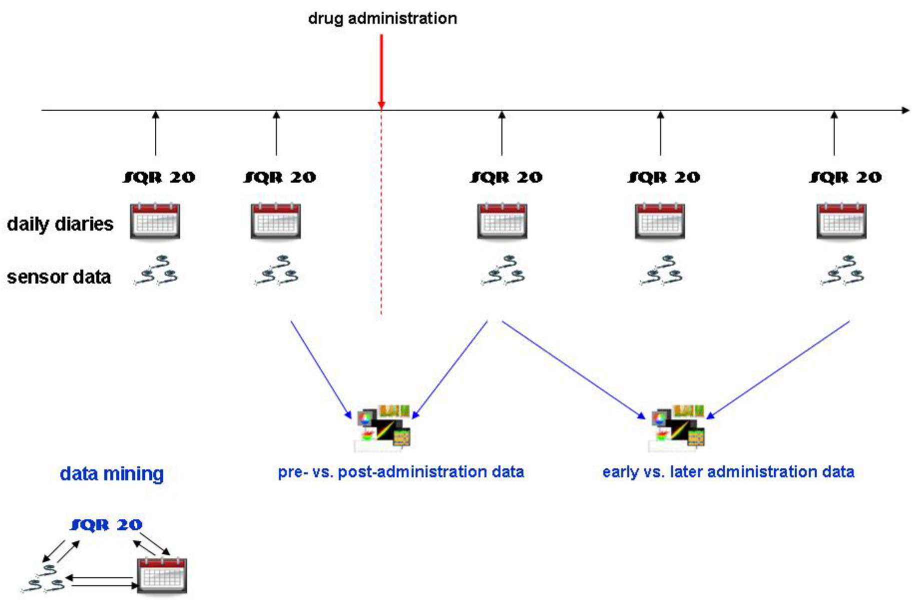

Fig 2. Examining effect of a drug on a patient

4.3. Case study 3: identifying healthy and unhealthy behaviour patterns

There is a relationship between mental health and behavioural patterns. For example, depression has been linked to internet addiction (Young & Rogers 1998), and physical exercise to good mental health (Taylor *et al*. 1985). Linking of the data describing daily activities of patients (daily diaries) with data describing mental health of patients (SQR20 data) and bodily functions in patients (sensor data), will enable identifying behavioural patterns that have positive or negative effects on wellbeing of patients.

4.4. Case study 4: gaining insight in the causes of depression

With the help of data mining techniques, it is possible to form groups of patients experiencing a special type of depression (such as postnatal depression) and then examine patterns specific to this group. It has been hypothesized that depression is caused by multiple factors of biological, psychological and social nature. Lets say that all biological factors are b_1, b_2, b_3 ….b_n, all psychological factors are p_1, p_2, p_3 ….p_n, and all social factors are s_1, s_2, s_3 ….s_n. By applying data mining techniques, it will be possible to collect patients' data describing each of these three dimensions, and systematically and collectively analyse this data to reveal common patterns specific to this particular depression type. For example, we will be able to say that pattern of causal factors (b_7, b_{13}, b_{14}, p_3, p_4, p_{10}, p_{22}, s_5, s_6, s_{12}) is specific to a particular depression type such as postnatal depression.

5. Significance of the System

The developed system will help all parties involved: patients, therapists and scientists.

Patients will be able to receive the best possible treatments. Having a monitoring system placed in an everyday environment of a patient, combined with data mining techniques will enable therapists to closely examine and interact with the patient, and provide specialized feedbacks to the patient. The collective knowledge derived with the help of data mining techniques can also be used to provide decision support for treatments of new patients. Monitoring mental health of therapists and their performance will help increase overall efficiency of the treatment process.

The proposed system will enable monitoring patients' condition when they get released from hospitals. Patients will be able to submit SRQ20 and daily diaries from home, and if needed, sensor measurements will be requested. This will be done in addition to regular patient follow-ups. Such process will introduce a greater control into the monitoring and treatment process and, in most cases, allow the doctors to provide useful feedback to patients and timely intervention in critical situations. Moreover, general practitioners can identify milder cases of depression, and suggest putting the monitoring system in the place. The proposed monitoring system will enable collection of data in the early stages of mental illness, and provide a basis for development of early intervention strategies.

Researchers will use the data to reveal patterns of interest, but in such a way that privacy of individuals remains inviolate. The results will be used to provide insight in possible causes of depression, assess the progress and effectiveness of a patients' treatment and predict possible problems.

6. Conclusion

The main obstacles in effective treatment of depression are (1) multivariate causes of depression, (2) variable effect of medications, (3) variable effect of psychotherapies, (4) incompetent therapists and (5) lack of a patient monitoring system outside hospitals.

The information technologies have the power to positively transform the way patients are being treated, and help us advance with knowledge more rapidly. We are designing a system that will help address the identified issues through the implementation of cutting-edge information technologies to enable systematic data collection and analysis. The data describing patients' activities, bodily functions and feelings as well as the data describing mental health of therapists will be collected and collectively mined to reveal interesting patterns.

The revealed knowledge will be profitable to all parties involved: patients, therapists and scientists. Patients will be able to receive highly personalized treatments, the therapists will be assisted in making evidence-based decisions, and the scientist will be able to pursue new knowledge revealing true causes of depression whilst developing more effective treatment approaches.

7. References

Batista, G. & Monard, M., 2003, 'An analysis of four missing data treatment methods for supervised learning', *Applied Artificial Intelligence*, vol. 17, no. 5-6, pp. 519-533.

Chaturvedi, A., Green, P., & Carroll, J. 2001, 'K-modes clustering', *Journal of Classification*, vol. 18, no. 1, pp.35-55.

Check, E. 2004, 'Antidepressants: Bitter pills', *Nature*, vol. 431, no. 7005, pp. 122-124.

Cooper-Patrick, L., Crum, R.M., & Ford, D.E. 1994, 'Characteristics of patients with major depression who received care in general medical and specialty mental health settings', *Medical Care*, vol. 32, no. 1, pp. 15-24.

Coulter, D.M., Bate, A., Meyboom, R.H., Lindquist, M., & Edwards, I.R. 2001, 'Antipsychotic drugs and heart muscle disorder in international pharmacovigilance: data mining study', *British Medical Journal*, vol. 322, no. 7296, pp. 1207-1209.

Craddock, N., & Jones, I. 2001, 'Molecular genetics of bipolar disorder', *The British Journal of Psychiatry*, vol. 178, no. 41, pp. 128-133.

Deutsch, C.J. 1985, 'A survey of therapists' personal problems and treatments', *Professional Psychology: Research and Practice*, vol. 16, no. 2, pp. 305-315.

Diener, E., & Lucas, R.E. 2000, 'Subjective emotional wellbeing', in *Handbook of Emotions* (2^{nd} ed.), eds. M. Lewis, & J. M. Haviland, Guilford, New York, pp. 325-337.

DiMatteo, M.R., Lepper, H.S., & Croghan, T.W. 2000, 'Depression is a risk factor for noncompliance with medical treatment: meta-analysis of the effects of anxiety and depression on patient adherence', *Archives of Internal Medicine*, vol.160, no. 14, pp. 2101-2107.

Engels, J.M. & Diehr, P. 2003, 'Imputation of missing longitudinal data: a comparison of methods', *Journal of Clinical Epidemiology*, vol, 56, no. 10, pp. 968-976.

Farber, B. A. 1990, 'Burnout in Psychotherapists: Incidence, Types and Trends', *Psychotherapy in Private Practice*, vol. 8, no. 1, pp. 35-44.

Fayyad, U., & Irani, K. 1996, 'Discretizing continuous attributes while learning bayesian networks', in *Proceedings 13^{th} International Conference on Machine Learning*, Bari, Italy, pp. 157-165.

Feng, L., Dillon, T.S., & Liu, J. 2001, 'Inter-transactional association rules for multi-dimensional contexts for prediction and their application to studying meterological data', *Data & Knowledge Engineering*, vol. 37, no. 1, pp. 85-115.

Feng, L., Dillon, T. S., Weigand, H., & Chang, E. 2003, 'An XML-enabled association rule framework', in *Proceedings of the 14th International Conference on Database and Expert Systems Applications*, Prague, Czech Republic, pp. 88-97.

Fredrickson, B.L. 2000, 'Cultivating Positive Emotions to Optimize Health and Wellbeing', *Prevention & Treatment*, vol. 3, no. 1, viewed 2 March 2009, <http://www.unc.edu./peplab/publications/cultivating.pdf>.

Friedman, R.A., & Leon, A.C. 2007, 'Expanding the Black Box-Depression, Antidepressants, and the Risk of Suicide', *The New England Journal of Medicine*, vol. 356, no. 23, pp. 2343-2346.

Hadzic, F., & Dillon, T.S., 2008, 'Human-like Rule Optimization for Continuous Domains', In *Biomedical Engineering Systems and Technologies, Communications in Computer and Information Science*, eds. A. Fred, J. Filipe, & H. Gamboa, Springer, vol. 25, pp. 330-343.

Hadzic, F., Dillon, T.S, & Tan, H. 2007, 'Outlier detection strategy using the Self-Organizing Map', in *Knowledge Discovery and Data Mining: Challenges and Realities with Real World Data*, eds. X.H. Zhou & I. Davidson, IGI Global, Hershey, Pennsylvania, pp. 224 -243.

Hadzic, M., Hadzic, F., & Dillon, TS. 2008, 'Tree Mining in Mental Health Domain', in *Proceedings of the Hawaii International Conference on System Sciences*, Hawaii, USA, pp. 230.

Hadzic, M., Hadzic, F., & Dillon, T. 2009, 'Domain Driven Data Mining for the Mental Health Domain', in *Data Mining for Business Applications*, eds. P.S. Yu, C. Zhang, & H. Zhang, Springer, New York, pp. 127-141.

Hadzic, F., Tan, H., & Dillon, T.S. 2010, 'Tree Model Guided Algorithm for Mining Unordered Embedded Subtrees', *Web Intelligence and Agent Systems: An International Journal (WIAS)*, IOS Press, vol. 8, no. 4.

Han, J. & Kamber, M., 2006, *Data Mining: concepts and techniques* (2^{nd} ed.), Elsevier, Morgan Kaufmann Publishers, San Francisco, CA, USA.

Harpham, T., Reichenheim, M., Oser, R., Thomas, E., Hamid, N., Jaswal S., Ludermir, A., &

Aidoo, M. 2003, ‘Measuring mental health in a cost-effective manner’, *Health Policy and Planning*, vol. 18, no. 3, pp. 344-349.

Horvitz-Lennon, M., Kilbourne, A.M., & Pincus, H.A. 2006, ‘From silos to bridges: meeting the general health care needs of adults with severe mental illnesses’, *Health Affairs*, vol. 25, no. 3, pp. 659-669.

Huang, J.1998, ‘Extensions to the k-means algorithm for clustering large data sets with categorical values’, *Data Mining and Knowledge Discovery*, vol. 2, no. 3, pp. 283-304.

Jinich, S., Paul, J.P., Stall, R., Acree, M., Kegeles, S., Hoff, C., & Coates, T.J. 1998, ‘Childhood sexual abuse and HIV risk-taking behaviour among gay and bisexual man’, *AIDS and Behaviour*, vol. 2, no. 1, pp. 41-51.

Link, B.G., Struening, E.L., Neese-Todd, S., Asmussen, S., & Phelan, J.C. 2001, ‘Stigma as a barrier to recovery: The consequences of stigma for the self-esteem of people with mental illnesses’, *Psychiatric Services*, vol. 52, no. 12, pp. 1621-1626.

Liu, H., Hussain, F., Tan, C.L., & Dash, M. 2002, ‘Discretization: An Enabling Technique’, *Data Mining and Knowledge Discovery*, vol. 6, no. 4, pp. 393-423.

Lopez, A.D., & Murray, C.C.J.L. 1998, ‘The global burden of disease, 1990-2020’, *Nature Medicine*, vol. 4, no. 11, pp. 1241-1243.

McHorney, C.A., Ware, J.E. Jr, & Raczek, A.E. 1993, ‘The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs’, *Medical Care*, vol. 31, no. 3, pp. 247-263.

Mullins, I.M., Siadaty, M.S., Lyman, J., Scully, K., Garrett, C.T., Miller, W.G., Muller, R., Robson, B., Apte, C., Weiss, S., Rigoutsos, I., Platt, D., Cohen, S., & Knaus, W.A. 2006, ‘Data mining and clinical data repositories: Insights from a 667,000 patient data set’, *Computers in Biology and Medicine*, vol. 36, no. 12, pp. 1351-1377.

Pacher, P., & Kecskemeti, V. 2004, ‘Cardiovascular Side Effects of New Antidepressants and Antipsychotics: New Drugs, old Concerns?’, *Current Pharmaceutical Design*, vol. 10, no. 20, pp. 2463-2475.

Pan, Q.H., Hadzic, F., & Dillon, T.S., 2008, ‘Conjoint Data Mining of Structured and Semi-structured Data’, in *Proceedings of the 4th International Conference on the Semantics, Knowledge and Grid*, Beijing, China, pp.87–94.

Patel, V., Flisher, A.J., Hetrick, S., & McGorry, P. 2007, ‘Mental health of young people: a global public-health challenge’, *Lancet*, vol. 369, no. 9569, pp. 1302-1313.

Quattrochi-Tubin, S.J., Jones, J.W., & Breedlove, V. 1982, ‘Syndrome in geriatric counsellors and service workers’, *Activities, Adaptation and Aging*, vol. 3, no. 1, pp. 65-76.

Ryan, R.M., & Deci, E.L. 2001, ‘On Happiness and Human Potentials: A Review of Research on Hedonic and Eudaimonic Wellbeing’, *Annual Review of Psychology*, vol. 52, pp. 141-166.

Ryan, R.M., Huta, V., & Deci, E.L. 2008, ‘Living well: a self-determination theory perspective on eudemonia‘, *Journal of Happiness Studies*, vol. 9, no. 1, pp. 139-170.

Schafer, W.R. 1999, ‘How Do Antidepressants Work? Prospects for Genetic Analysis of Drug Mechanisms’, *Cell*, vol. 98, no. 5, pp. 551-554.

Sestito, S., & Dillon, T.S. 1994, *Automated knowledge acquisition*. Sydney: Prentice Hall.

Soini, E.J.O., Rissanen, T., Tiihonen, J., Eronen, M., Hodgins, S., & Ryynanen, O.-P., 2008, ‘Predicting Forensic Admission among the Mentally Ill: A Bayesian Approach’, in *Proceedings of the 21st IEEE International Symposium on Computer-Based Medical Systems*, Jyvaskyla, Finland, pp.242-247.

Tan, H., Dillon, T.S., Hadzic, F., Feng, L., & Chang, E. 2006, ‘IMB3-Miner: Mining Induced/Embedded Subtrees by Constraining the Level of Embedding’, in *Proceedings of the 10th Pacific-Asia Conference on Knowledge Discovery and Data Mining,* Singapore, pp. 450-461

Tan, H., Hadzic, F., Dillon, T.S., Feng, L., & Chang, E. 2008, ‘Tree Model Guided Candidate Generation for Mining Frequent Subtrees from XML’, *ACM Transactions on Knowledge Discovery from Data (TKDD),* vol. 2, no. 2, pp. 1-43.

Taylor, C.B., Sallis, J.F., & Needle, R. 1985, ‘The relation of physical activity and exercise to mental health’, *Public Health Report*, vol. 100, no. 2, pp. 195-202.

Werneke, U., Northey, S., & Bhugra, D. 2006, ‘Antidepressants and sexual dysfunction’, *Acta Psychiatrica Scandinavica*, vol. 114, no. 6, pp. 384-397.

Young, K.S., & Rodgers, R.C. 1998, ‘The Relationship Between Depression and Internet Addiction’, *CyberPsychology & Behavior*, vol. 1, no. 1, pp. 25-28.

Zhou, X.-J.M., & Dillon, T.S. 1991, ‘A statistical-heuristic feature selection criterion for decision tree induction’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 834-841.