Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 96 (2016) 1294 – 1303

Procedia Computer Science

20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems

Selecting Learning Algorithms for Simultaneous Identification of Depression and Comorbid Disorders

Blessing Ojeme[a*], Audrey Mbogho[b]

[a]*CAIR, Department of Computer Science, University of Cape Town, Rondebosch 7701, and CSIR, South Africa*
[b]*Department of Mathematics and Physics, Pwani University, Kilifi, Kenya*

Abstract

Depression is a serious worldwide public health problem, and its diagnosis still remains a challenge in the medical community. The difficulties of detecting depression are largely due to its high comorbid factor. Given the reciprocal relationship between depression and physical illness, mental health professionals have called for a diagnostic approach that identifies and evaluates each disorder, concurrently. This paper reports the findings of a study based on data collected in Nigeria to investigate the simultaneous identification of depression and co-occuring physical illness using a multi-dimensional Bayesian network classification approach. The predictive model would be useful to clinicians of all categories in Nigeria in overcoming the challenges of depression diagnosis caused by its frequent co-occurence with physical illnesses. The benefits of this approach are demonstrated with anonymised multi-dimensional depression dataset comprising 1090 instances, 22 symptoms, and two class attributes. The results, are also described.

Peer-review under responsibility of KES International

Keywords: Machine learning; Multi-dimensional classification; Depression; Physical illnesses; Predictions; Medical diagnosis; Comorbidity.

1. Introduction

Depression is an important disease to study because of its virulence. Its detection, though highly relevant, still remains a major challenge in the medical community in Nigeria and other developing countries despite the high

* Corresponding author. Tel.: +27725574409.
E-mail address: bojeme@cs.uct.ac.za

Peer-review under responsibility of KES International
doi:10.1016/j.procs.2016.08.174

accuracy values being achieved by various classification algorithms on different depression datasets in the literature. Fahrer and Grassi [1] and Golberge [2] noted that depression is frequently comorbid with a variety of physical illnesses, which are often undetected. The prevalence rates of depression in comorbidity with physical illnesses, according to the studies, is over 22%, with varying rates for specific illnesses. A breakdown of the dataset for this study, shown in Table 2, is a replication of these important findings. Depression comorbidity refers to the occurrence of depressive disorders and other disorders (physical or mental) in the same person, either simultaneously or sequentially [2]. Depression comorbidity has become a major concern for mental health (MH) researchers, MH professionals and multi-purpose clinicians (MPC) because of its close link with compromised quality of lives, morbidity and mortality[3]. Similarities in symptoms of depression and physical illnesses make its diagnosis complicated [4]. Compared to the volume and variety of machine learning (ML) techniques used for the diagnosis of depression, there is a gap in literature on how the same techniques could be used to simultaneously identify depression and comorbid physical illnesses. Methodological limitations, such as lack of access to adequate primary depression datasets in comorbid population [5] and lack of combined expertise in both MH and computer science [6], have been identified as some of the reasons for this gap. Evidently, there is a need for a more comprehensive multi-dimensional classification (MDC) learning approach that exploits the correlations, identifies and evaluates depression and co-occuring physical illnesses, concurrently. As noted by Batal et al [7], in MDC problems, the class variables usually exhibit conditional dependence relations among themselves. Such dependencies must be modelled in order to learn about not just what is possible, but what is probable. Probability theory provides the framework for considering such multiple possible outcomes [8]. This study investigates the needed MDC learning with a probabilistic framework, which given symptoms, can be used to compute the probabilities of the presence of various diseases. Specifically, the study has, as its contribution, the construction and description of a multi-dimensional predictive model, which uses a multi-dimensional Bayesian network classifier (MBC) that provides reliable and clinician-interpretable diagnostic results with respect to simultaneously determining the presence of depression and physical illnesses from symptoms, correctly.

The rest of the paper is organized as follows: as a follow-up to the introduction in sub-section 1.1, the relationship between machine learning, multi-class classification, multi-label classification and multi-dimensional classification is briefly discussed, followed by a description of the methods of handling the tasks that fall within them. Section 1.2 describes the identification of depression as a multi-dimensional problem and its proposed solution using multi-dimensional Bayesian network classifier. In section 2, we discuss the dataset for the study and the method of data collection. Section 3 describes our experimental design and the environment where it was performed. Section 4 presents the results of our experiment and its analysis while section 5 outlines the evaluation metrics for our model. In section 6, we discuss a few studies that are related to ours while section 7 concludes the study with a plan for future work.

1.1. Machine Learning Algorithms for Multi-Label Classification Problems

ML, a subfield of artificial intelligence, provides the techniques that are able to exploit the correlations among large datasets and classify them into a manageable and easily human-interpretable format. ML tasks can be broadly divided into five main classes: Association, Classification, Regression, Clustering and Optimization tasks [9]. Classification, which is the primary concern of this paper, is the task of learning a model using a set of previously classified instances and applying the obtained model to a set of previously unseen examples.

In medical diagnosis, ML techniques have proved to be useful in classification tasks. Data of previously diagnosed patients, called training instances, are used to build a classification model, which is then used to classify previously unseen data. The resulting classification is evaluated for accuracy. As depicted in Figure. 1, classification problems can be categorized according to the number of class labels that can be assigned to a particular input instance. Multi-label classification (MLC) is the supervised ML problem where the classification algorithm learns from data instances, and each instance has multiple class-labels, but each class-label has binary values [10]. In MLC, instance may have different sets of labels. This is different from the traditional Single-label classification (SLC), or multi-class classification (MCC), which is concerned with learning from data instances that are associated with a single label from a set of disjoint labels, and involves only a single nominal target variable [11] [12]. In MLC each class label $|Y_j| = 2$ (i.e. only two classes) for all $j = 1,\ldots\ldots,d$ (i.e. binary classification where a label is either relevant (1)

or not (0). In MDC, each class variable $|Y_j| = K_j$ for any positive integer K_j [12] (d is the number of class variables and K is the number of values each of these variables may take). This is illustrated in Table 1.

Table 1. Class label divisions

	K = 2	K > 2
d = 1	Binary	Multi-class
d > 1	Multi-label	Multi-dimensional

MDC (also called multi-target classification (MTC) or multi-output classification (MOC)) being a generalization of MLC, is concerned with learning from examples, in which each data instance has multiple target variables, and each variable takes multiple values [13]. MDC has found application in a wide variety of real-world domains, including medical diagnosis, bioinformatics, robotics, text, vision, and audio processing[13]. In medical diagnosis, for instance, MDC deals with a situation where a patient may be suffering from multiple diseases and assumes each instance is associated with d discrete-value class variables $Y_1 \ldots\ldots Y_d$. The aim is to learn a function that assigns to each instance represented by its feature vector (symptoms) $X = x_1 \ldots\ldots x_m$, the most probable assignment of the class variables (target or output variable) $y = y_1 \ldots\ldots y_d$. That is, MDC is a function that maps X into Y or $(x_1 \ldots\ldots x_m) \rightarrow (y_1 \ldots\ldots y_d)$, with x and y being vectors. This large number of possible combination increases the complexity of MDC problems, making them more difficult to solve than SLC problems, where one such variable is associated with a data instance.

To solve this problem of class label combination and its resultant computational complexity, Read et al [12] and Tsoumakas and Katakis [14] in their works, proposed two methods: problem transformation (PT) and algorithm adaptation (AA) . some of the commonly used algorithms in the two methods are shown in Figure 1 while their strengths and weaknesses are well discussed in the work of Madjarov et al. [15].

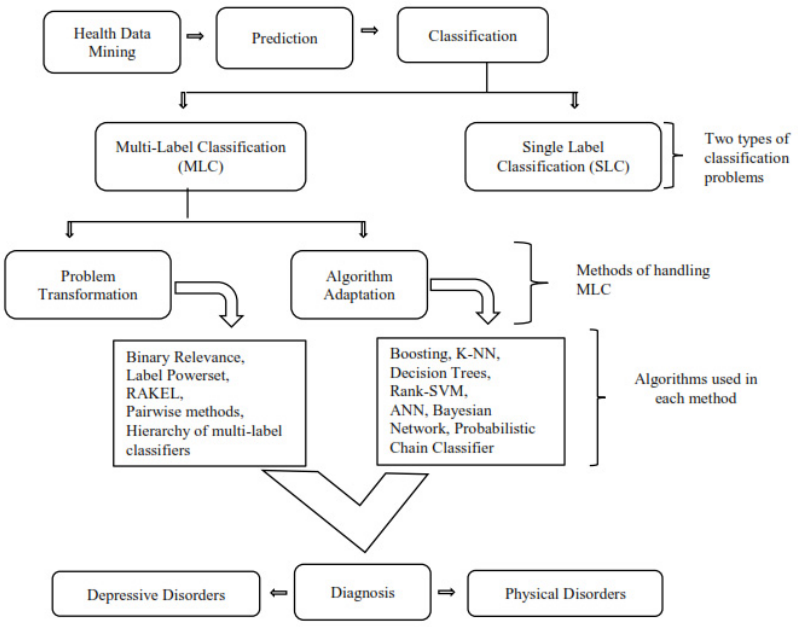

Fig. 1. Framework of the study

1.2. Depression Diagnosis as a Multi-dimensional Classification Problem

Classification problems in which instances (described by a number of features or symptoms) are assigned to multiple classes (diseases) simultaneously are solved by MDC approach [13]. Given that a depression patient may also be suffering from other diseases, the identification of depression and co-occuring physical illnesses (classes or targets) from symptoms (features or attributes) is an MDC task [16]. Here, the multiple diseases, which are simultaneously determined, are used to explain the symptoms [17]. As noted earlier, the class variables usually exhibit conditional dependence relations among themselves. Such dependencies must be modelled in order to learn about not just what is possible, but what is probable. Probability theory provides the framework for considering such possible multiple outcomes. Having been extensively used as classifiers in the machine learning field due to its ability to maximize the accuracy of each class variable at the same time, Bayesian networks, an appealing tool for probabilistic MDC, is proposed for this study. Bayesian networks compactly represent the joint probability distribution of class and feature variables P(x, y). However, traditional Bayesian network classifier cannot handle such an MDC problem since its class variable is restricted to one [18].

Multi-dimensionality has been introduced in Bayesian network classifiers to provide for modelling of the MDC problem and ensuring interactions among all variables [19]. Multidimensional Bayesian network classifiers (MBC) are the generalization of the traditional single-class-oriented Bayesian network classifiers to domains with more than one class variable. MBC models the probabilistic relationships between the variables by directed acyclic graphs, partitioning the set of class and feature variables into three different subgraphs: class subgraph representing the dependence relationships between class variables; bridge subgraph representing the dependence relationships between class and feature variables, and feature subgraph representing the dependence relationships between feature variables. In learning the MBC from data, several approaches have been offered, including Tree-Tree method [19], Polytree-Polytree method [18], Pure Filter and Pure Wrapper methods [13] and Class-Bridge decomposable MBC method [20]. These approaches can be used to solve two problems identified by Alessandro et at. [21] and Corani et al. [22]: learning from data the structure of the MBC model and performing predictive inference on the learned model in order to classify instances.

Inspired by the work of Borchani et al [20], this study uses the class bridge decomposable MBC, extending it to detecting depression and comorbid physical illness from a primary depression dataset. The MBC is defined thus: An MBC is a Bayesian network B = (G, Θ) where the structure G = (V, A) has a restricted topology. The set of n vertices V is partitioned into two sets: $V_c = \{C_1,\ldots\ldots,C_d\}$, d ≥1, of the variables and $V_x = \{X_1,\ldots\ldots,X_m\}$, m ≥1, of feature variables (d + m = n). The set of arcs A is partitioned into three sets A_c, A_x and A_{cx}, with the following properties:

- $A_c \subseteq V_c * V_c$ is composed of the arcs between the class variables having a subgraph $G_c = (V_c, A_c)$ called class subgraph of G induced by V_c
- $A_x \subseteq V_x * V_x$ is composed of the arcs between the feature variables having a subgraph $G_x = (V_x, A_x)$ called feature subgraph of G induced by V_x.
- $A_{cx} \subseteq V_c * V_x$ is composed of the arcs from the class variables to the feature variables having a subgraph $G_{cx} = V, A_{cx}$ called bridge subgraph of G connecting class and feature variables

The Bayesian network B of an MBC codifies a joint probability distribution P(x, c) which factorises according to

$$PB(x, c) = \prod_{c \in C} P(c|pa(C)) \prod_{X \in X} P(x|pa(X)) \qquad (1)$$

where pa(c) represents the parent of C and pa(X) represents the parent of X in the structure of the Bayesian network. In essence, a multi-dimensional classifier actually serves to find the most probable explanation (MPE) or the maximum a posteriori (MAP) of a given data instance with feature vector X= $\{x_1,\ldots.., x_m\}$, which is the mode of the joint probability distribution of the classes C given the features X [21]. This is computed to get:

$$c^* = \{c_1^*,\ldots,c_n^*\} = \arg_{c_1,\ldots,c_n} \max p(C_1 = c_1,\ldots,C_n = C_n/x) \qquad (2)$$

where C_i is the array of class variables, c^* is the actual class and x is the feature.

In prediction tasks, if the target variables (Y) consist of continuous/numeric variables, then the task at hand is multi-target regression. Similarly, if the target variables (Y) consist of discrete/nominal variables (as shown in Table 2), then the task is called multi-target classification.

Table 2. Depression dataset with multiple target variables

Descriptive variables							Target variables	
	Age (X_1)	Sex (X_2)	Sad mood (X_3)	Suicidal (X_4)	……	Employment status (X_{22})	Depression disorders (Y_1)	Physical illness (Y_2)
1	34	F	Yes	Yes		No	Severe depression	Heart disease
2	58	M	Yes	Yes		PT	Moderate depression	Stroke
3	44	F	No	No		FT	No depression	?
4	21	M	Yes	Yes		FT	?	?
5	15	F	Yes	No		PT	?	Peptic ulcer
. . . .	. . .	. . .	. . .	. . .	. .	. . .	. . .	. .
1090	42	M	Yes	No		PT	Mild depression	hypertension

2. Data Collection and Description of Dataset

Due to the centrality of data collection in the data analysis stage, much attention was given to the process. The researcher employed the use of workshops (in the form of seminars), semi-structured interviews and record extraction from previously diagnosed cases for the collection of data. The dataset (shown in Table 2) consisted of anonymised records of 1090 data instances, (484 male and 606 female cases from 12 to 92 years old with a mean age of 41.95 and standard deviation of 16.07). It had 22 main attributes (age, sex, sad mood, suicidal, loss of pleasure, insomnia, hypersomnia, loss of appetite, psychomotor agitation, psychomotor retardation, loss of energy, feeling of worthlessness, lack of thinking, indecisiveness, recurrent thought of death, impaired function, weight gain, weight loss, stressful life events, financial pressure, depression in family and employment status) and two class attributes (depression diagnosis and depression comorbid with) collected in a comorbid population, from the University of Benin Teaching Hospital (UBTH) and primary care centres in Nigeria. All depression cases selected fulfilled clinical criteria for depression as defined in DSM-5 [23] and ICD-10 [24]. Further validation of the data was done by a group of eight mental health professionals, made up of three Psychiatrists, one Child/Adolescent Psychiatrist, two Clinical Psychologists, one Clinical Social Worker and one Nurse Psychotherapists. A control group (No depression) of 36 cases were also collected. All ethical clearances for data collection were obtained from the appropriate authorities.

Table 3 shows the distribution of depression cases in co-occurence with physical illnesses. The data clearly show that persons diagnosed with moderate and severe depression are much more likely to suffer also from comorb physical illnesses. The 29 comorb physical illnesses case included: cerebrovascular accident (cva), hypertension (htn), benign prostatic hypertrophy (bph), diabetes, human immunodeficiency virus (hiv), non hodgkin's lymphoma (nhl), liver disease (ld), chronic liver disease (cld), erectile dysfunction (ed), poor sight (ps), tuberculosis (tb), pneumonia, cannabis, infertility, esophageal stricture (es), peptic ulcer disease (pud), malaria, toxic diffuse goitre (tdg), appendicitis, heart disease, alcohol use disorder (aud), cataract, arthritis, hypertensive heart disease (hhd), enuresis, diabetes mellitus (dm), hydrocele, b essential tremors (bet), epilepsy, and seizure.

As noted earlier, there are some correlations among depression and physical illnesses. Many symptoms of depression are similar to those of physical illnesses, including depressed mood, loss of pleasure, helplessness, fatigue, weight loss, weight gain, hypersomnia, psychomotor retardation, decreased concentration, and cognitive

impairment[1]. Due to the difficulties this poses to the diagnosis of depression in patients with physical illnesses, this study adopts the inclusive approach suggested by Fahrer and Grassi [1], where all of the symptoms are included in the assessment for depression, irrespective of the fact that these symptoms may be attributable to physical illnesses.

Table 3. Description of dataset

S/N	Depression Diagnosis	Frequency	Percent	No. of cases with Comorbidity	Percent	No. of cases without Comorbidity	Percent
1	Mild	47	4.3	6	0.6	41	3.8
2	Moderate	530	48.6	95	8.7	435	40
3	Severe	477	43.8	77	7.1	400	36.7
4	Not depressed	36	3.3	Nil	0	36	3.3
	Total	**1090**	**100**	**178**	**16.4**	**912**	**83.8**

3. Experimental Design

Since each classifier has its own inductive bias, testing out multiple classifiers and selecting the best model can be a good idea. The experiment was conducted using Bayesian classifier chains (BCC), probabilistic classifier chains, (PCC), Super class classifier (SCC), Bagging, ensemble of classifier chains (ECC), pruned sets and classifier chains (CC). In order to make our experiments reproducible, the algorithms were implemented in multi-label extension to Weka (Meka, version 1.9.0) [25] using default parameters of the Explorer panel in the graphical user interface (GUI). Meka is an open source ML framework which provides an extensible support for developing, running and evaluating multi-label and multi-target classifiers-the main focus of the study. The data shown in Table 3 were pre-processed and reformatted to yield a suitable input for the Meka tool and our classifier.

4. Results and Analysis of the Performance of the Proposed Technique

Using the experimental set up in section 3, and the dataset described in Table 3, the result is displayed in Table 4. From the obtained results we can conclude that, for this problem, our classifier showed a noticeable performance in considered performance metrics (Hamming score, Exact-match and Hamming loss) for this dataset, and hence may be used to establish baseline performance.

Table 4. Results of the methods used

Method	Hamming score	Exact-match	Hamming loss
BCC	**0.91**	**0.826**	**0.09**
PCC	**0.91**	**0.826**	**0.09**
SCC	0.921	0.854	0.079
Bagging	0.923	0.852	0.077
ECC	0.908	0.822	0.092
Pruned sets	0.921	0.921	0.079
CC	0.91	0.826	0.09

The findings of this study need to be interpreted in light of the methodological limitations on the performance of the model. The most important limitation was the small size of the dataset (N = 1090). Another limitation that could

have degraded the model performance was the presence of irrelevant, redundant and noisy features, which are common occurrences in medical datasets [26]. For a given instance size, there is a maximum number of features above which the performance of a classifier will degrade rather than improve. The MDC algorithms were evaluated on their ability to predict the presence of depression and physical illnesses based on previously diagnosed cases. Some of the predictors were found to be very good. For instance, the Hamming score for the BCC method (0.91) is comparable to the results obtained in the study by Fernandez-gonzalez et al [27] and Ibrahim [28]. Fernandez-gonzalez et al [27] achieved a Hamming score of of 0.801 and Exact-match of 0.25 while Ibrahim [28], achieved the highest Hamming score of 0.77 (significantly less than ours). The probabilistic methods are marked in boldface in Table 4 for visualization purpose

5. Evaluation Procedure

In this work we utilized the Hamming score, Hamming Loss and Exact-match to evaluate the performance of depression and physical illnesses diagnosis. Hamming score gives the accuracy for each class correctly predicted, averaged across all classes. Hamming loss gives the percentage of data predicted incorrectly on average (The best performance is reached when hamming loss is equal to 0. The smaller the value of hamming loss is, the better the performance is). Exact -match (also called global accuracy) gives the percentage of test dataset predicted exactly same as in the training dataset. To obtain a fair accuracy estimation, the dataset used to train the classifier must be independent from the dataset used to test it [29]. Meka provided the platform that ensured this independence by using a 10-fold cross-validation procedure to partition the dataset into 10 disjoint subset. In other words, Meka helped to ensure that our limited and unbalanced dataset got a good balance between the size and representation of the training and test sets. And since the joint probability distribution p(x, y) was unknown and the accuracy of the class variables estimated from data, the partitioning of the dataset also helped to reduce the variance of the estimates and improved the estimation of the generalization performance of the classifier algorithms [30]. Each subset was utilized once as a test set and nine times as part of a training set in order to ensure a valid and robust results. This technique was used for all the algorithms under study and the average results are reported.

Hamming score is stated thus:

$$\frac{\text{set of correct classes}}{\text{union of predicted and true classes}}$$

Mathematically, Hamming loss is stated thus:

$$H_s = \frac{1}{N}\sum_{i=1}^{N} \left| \frac{T_i \cap P_i}{T_i \cup P_i} \right| \qquad (3)$$

where N is the total number of instances in the data, T_i is the set of true labels for the ith instance and P_i is the set of predicted labels by the classifier for the ith instance.

Exact match is expressed:

$$E_s = \frac{1}{N}\sum_{i=1}^{N} \delta_{T_i}^{P_i} \qquad (4)$$

where $\delta_{T_i}^{P_i}$ is a function that outputs 1 if $T_i = P_i$ and 0 otherwise.

6. Related Work

A number of studies on the possibility of automating conventional depression diagnosis with machine learning algorithms have been done in the past. Some of these studies that are related to the identification of depression disorders and co-occuring physical illnesses are reviewed.

Bielza et al [13] proposed a class-bridge (CB) decomposable MBC for a multi-dimensional classification problem. When the learning algorithms were experimented with three datasets (Emotions dataset, Scene dataset and Yeast dataset) taken from the literature on multi-label classification problems, the results were encouraging, outperforming state-of-the-art algorithms for multi-label classification. The study used accuracy and F1 measure as performance evaluation measures.

In Rodriguez et al [31], the authors used a multi-objective learning approach and multidimensional J/K dependencies Bayesian classifier for predicting the progression of multiple sclerosis (MS) in patients that had been recently diagnosed. The dataset used consisted of DNA and clinical information of 605 (86 features and 2- class variables) unrelated Dutch Caucasian patients selected from natural history literature. Using a k-fold cross-validation evaluation method, the algorithms returned different tradeoff solutions to the multi-dimensional classification problem with different accuracy (59%, 64%, 77%, 80%, and 90%) values for the different class variables.

Kessler et al [32] presented a report of machine learning models developed from self-reports about incident episode characteristics and comorbidities among respondents with lifetime major depressive disorder (MDD) in the World Health Organization World Mental Health (WMH) survey. The machine learning predicted accuracy was compared with observed scores assessed 10–12 years after baseline and that of conventional logistic regression models. Area under the receiver operating characteristic (ROC) curve based on machine learning (0.63 for high chronicity and 0.71–0.76 for the other prospective outcomes) was consistently higher than for the logistic models (0.62–0.70) despite the latter models including more predictors.

The closest to this study was by Lueken et al [33].The study investigated the impact of comorbid depression in panic disorder (PD with agoraphobia (AG) on the neural correlates of fear conditioning. Using tree ensemble in leave-one-out cross-validation framework, 59 PD/AG patients including 26 (44%) with a comorbid depressive disorder (PD/AG + DEP) underwent functional magnetic resonance imaging (fMRI). The results showed that comorbidity status was correctly predicted in 79% of patients (sensitivity: 73%; specificity: 85%) based on brain activation during fear conditioning (corrected for potential confounders: accuracy: 73%; sensitivity: 77%; specificity: 70%). These findings demonstrate the relevance of comorbidity patterns when investigating neurobiological substrates of anxiety disorders. The limitation of the study was the unavailability of primary depressed patients; only medication-free patients were used.

However, these studies do not have a primary depression dataset with comorbid physical illnesses; do not use cross-validation; and do not evaluate the models with metrics described in this study. Though a different approach was used to address a similar problem, with a different dataset for evaluation in Lueken et al [33], our MBC model outperformed their method by obtaining a higher true detection rate.

7. Conclusions and Future Work

We have demonstrated that the class-bridge MBC predictive model has the potential to offer good performance. Though the results do not show significant boost over non-probabilistic methods in terms of prediction performance, it showed a noticeable performance in considered performance metrics, which would be useful to both mental health professionals and clinicians in practice and facilitate the adoption of objective, effective and reliable computer-based diagnosis tool. Given a large amount of suitable data with relevant features, the technique used in this study is general-purpose and can be used for MDC problems with other diseases.

A number of areas can be identified for further research, some relating to methodological advances, and others to the application of the methods used in this study to other aspects of depression detection. In future work, we will explore the possibility of improving the model performance by sourcing for more data and using unsupervised method such as the principal component analysis (PCA)[34] and latent Semantic Indexing (LSI) [35] to ensure that optimal subset of features is being used for training and prediction, since irrelevant and redundant features in dataset can result in reduced classification accuracy and increased computational burden in practice. Finally, of practical

interest to us would be to further investigate whether our learning algorithms can be as useful in other application domains.

Reference

1. Fahrer R, Creed F, Grassi L. Prevalence, Pathogenesis, and Diagnosis of Depressive Disorders in the Medically Ill. In: World Psychiatric Association Educational Programme on Depressive Disorders - Physical Illness and Depression [Internet]. 2008. p. 143. Available from: http://www.wpanet.org/uploads/Education/Educational_Programs/depressive-disorders-volume2.pdf
2. Goldberg D. The detection and treatment of depression in the physically ill. World Psychiatry [Internet]. 2010;9(1):16–20. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2816927&tool=pmcentrez&rendertype=abstract
3. Teesson M, Proudfoot H. Comorbid mental disorders and substance use disorders: epidemiology, prevention and treatment [Internet]. And MT, Proudfoot H, editors. National Drug and Alcohol Research Centre University of New South Wales Sydney, Australia. National Drug and Alcohol Research Centre University of New South Wales Sydney, Australia; 2003. 148 p. Available from: http://health.gov.au/internet/main/publishing.nsf/Content/health-pubhlth-publicat-document-mono_comorbid-cnt.htm
4. Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. Lancet. 2007;370(9590):851–8.
5. National Institute on Drug Abuse. Comorbidity: Addiction and Other Mental Illnesses. US Department of Health and Human Services, National Institute of Health. Available at: https://www.drugabuse.gov/sites/default/files/rrcomorbidity.pdf. US Department of Health and Human Services, National Institute of Health. Aaavailable at: https://www.drugabuse.gov/sites/default/files/rrcomorbidity.pdf; 2010.
6. Chattopadhyay S. A neuro-fuzzy approach for the diagnosis of depression. Elsevier Appl Comput Informatics [Internet]. King Saud University; 2014;In Press(February):19. Available from: http://www.sciencedirect.com/science/article/pii/S2210832714000027
7. Batal I, Hong C, Hauskrecht M. An efficient probabilistic framework for multi-dimensional classification. In: Proceedings of the 22nd ACM international conference on Conference on information & knowledge management - CIKM '13 [Internet]. 2013. p. 2417–22. Available from: http://dl.acm.org/citation.cfm?doid=2505515.2505594
8. Koller D, Friedman N. Probabilistic Graphical Models: Principles and Techniques. In: The MIT Press. MIT Press; 2009. p. 16.
9. Carvalho, André C P L F de AAF. A Tutoria on Multi-Label Classification Techniques. Found Comput Intell Springer-Verlag [Internet]. 2009;5:177–95. Available from: http://www.cs.kent.ac.uk/people/staff/aaf/pub_papers.dir/Found-Comp-Intel-bk-ch-2009-Carvalho.pdf
10. Read J, Pfahringer B, Holmes G, Frank E. Classifier chains for multi-label classification. Mach Learn. 2011;85(3):333–59.
11. Cherman E, Monard M, Metz J. Multi-label problem transformation methods: a case study. CLEI Electron J [Internet]. 2011;14(1):1–10. Available from: http://www.scielo.edu.uy/scielo.php?pid=S0717-50002011000100005&script=sci_arttext&tlng=es
12. Read J, Bielza C, Larrañaga P. Multi-Dimensional Classification with Super-Classes. IEEE Trans Knowl Data Eng. 2014;26(7):1720–33.
13. Bielza C, Li G, Larrañaga P. Multi-dimensional classification with Bayesian networks. Int J Approx Reason. 2011;52(6):705–27.
14. Tsoumakas G, Katakis I. Multi-Label Classification: An Overview. Int J Data Warehous Min [Internet]. 2007;3(3):309–19. Available from: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-60566-058-5.ch021
15. Madjarov G, Kocev D, Gjorgjevikj D, D??eroski S. An extensive experimental comparison of methods for multi-label learning. Pattern Recognit. 2012;45(9):3084–104.
16. Tsoumakas G, Buntine W, Batagel V, Cunningham P, Lazaric A, Caetano T. Learning From Multi-label Data. In: Ecml-Pkdd. 2009. p. 5–15.
17. Keren B, Kalech M, Rokach L. Model-Based Diagnosis with Multi-Label Classification. 22nd Int Work

Princ Diagnosis, DX. 2011;
18. de Waal P, van der Gaag L. Inference and Learning Multi-dimensional Bayesian Network Classifiers. Vol. 4724, Lecture Notes in Artificial Intelligence. 2007.
19. van der Gaag L, de Waal P. Multi-dimensional Bayesian Network Classifiers. In: Proceedings of the Third european workshop in probabilistic graphical models. 2006. p. 107–14.
20. Borchani H, Bielza C, Larrañaga P. Learning CB-decomposable multi- dimensional Bayesian network classifiers. Proc Fifth Eur Work shop Probabilistic Graph Model. 2010;25–32.
21. Alessandro A, Corani G, Mauá D, Gabaglio S. An ensemble of Bayesian networks for multilabel classification. IJCAI Int Jt Conf Artif Intell. 2013;(200020):1220–5.
22. Corani G, Antonucci A, Mau D. Trading off Speed and Accuracy in Multilabel Classification. Probabilistic Graph Model [Internet]. 2014;8754(1):145–59. Available from: <Go to ISI>://WOS:000358253800010
23. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders fifth edition. 5th ed. Washington, DC: American Psychiatric Publishing; 2013. 947 p.
24. World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders. Vol. 10, Mental Health Division, World Health Organization. 1992. 1-267.
25. Read J, Reutemann P, Pfahringer B, Holmes G. MEKA: A Multi-label / Multi-target Extension to WEKA. J Mach Learn Reasearch Available http//jmlr.org/papers/volume17/12-164/12-164.pdf. 2016;17(21):1–5.
26. Xu J, Liu J, Yin J, Sun C. A multi-label feature extraction algorithm via maximizing feature variance and feature-label dependence simultaneously. Knowledge-Based Syst [Internet]. Elsevier B.V.; 2016;98:172–84. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0950705116000526
27. Fernandez-gonzalez P, Bielza C, Larranaga P. Multidimensional classifiers for neuroanatomical data. ICML Workshop on Statistics, Machine Learning and Neuroscience (Stamlins 2015), Jul 2015, Lille, France. 2015. <hal-01225249. 2015. p. 7.
28. Ibrahim N. Data Mining Model to Predict Fosamax Adverse Events. Int J Comput Inf Technol. 2014;03(05):934–41.
29. Witten I, Frank E, Hall M. Data Mining : Practical Machine Learning Tools and Techniques. Third Edit. Morgan KaufmannPublishers; 2011. 665 p.
30. Juan Fernandez del pozo, Pedro, Larranaga C bielza. Technical Report on Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms. Computational Intelligence Group Universidad Politécnica de Madrid 1; 2013.
31. Rodriguez JD, Perez A, Arteta D, Tejedor D, Lozano JA. Using Multidimensional Bayesian Network Classifiers to Assist the Treatment of Multiple Sclerosis. Syst Man, Cybern Part C Appl Rev IEEE Trans [Internet]. 2012;42(6):1705–15. Available from: http://ieeexplore.ieee.org/ielx5/5326/6330018/06392464.pdf?tp=&arnumber=6392464&isnumber=6330018
32. Kessler R, van Loo H, Wardenaar K, Bossarte R, Brenner L, Cai T, et al. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. Mol Psychiatry [Internet]. 2016;(October 2015):1–6. Available from: http://www.nature.com/doifinder/10.1038/mp.2015.198
33. Lueken U, Straube B, Yang Y, Hahn T, Beesdo-Baum K, Wittchen HU, et al. Separating depressive comorbidity from panic disorder: A combined functional magnetic resonance imaging and machine learning approach. J Affect Disord. 2015;184:182–92.
34. Kumar D, Singh R, Kumar A, Sharma N. An Adaptive Method of PCA for Minimization of Classification Error Using Naïve Bayes Classifier. Procedia Comput Sci [Internet]. Elsevier Masson SAS; 2015;70:9–15. Available from: http://www.sciencedirect.com/science/article/pii/S1877050915031828
35. Yu K, Yu S, Tresp V. Multi-label informed latent semantic indexing. In: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, pages 258–265, Salvador, Brazil, 2005 ACM Press [Internet]. 2005. p. 258–65. Available from: http://portal.acm.org/citation.cfm?doid=1076034.1076080\nhttp://dl.acm.org/citation.cfm?id=1076080