Potential biomarkers for predicting of depression in diabetes mellitus

Xiuli Song
clinical psychology

Qiang Zheng
school of computer

Rui Zhang
information center

Miye Wang
information center

Wei Deng
Sichuan University West China Hospital

Qiang Wang
Sichuan University West China Hospital

Wanjun Guo
Sichuan University West China Hospital

Tao Li
Sichuan University West China Hospital

xiaohong ma (lz1757@163.com)
Sichuan University West China Hospital

Research article

Keywords: Diabetes Mellitus, Depression, Support Vector Machine

Posted Date: April 22nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-23813/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published at Frontiers in Psychiatry on November 29th, 2021. See the published version at https://doi.org/10.3389/fpsyt.2021.731220.
Abstract

Background

To identify the potential biomarkers for predicting depression in diabetes mellitus using the support vector machine technique to analyze routine biochemical tests and vital signs between two groups: subjects with both diabetes mellitus and depression, and subjects with diabetes mellitus alone.

Methods

Electronic medical records upon admission and biochemical tests and vital signs of 135 patients with both diabetes mellitus and depression and 178 patients with diabetes mellitus alone were identified for this retrospective study. After the covariate regression analysis on age and sex, the two groups were classified by the recursive feature elimination-based support vector machine and the biomarkers were also identified by 10-fold cross validation. Specifically, the training data, evaluation data, and testing data were split for ranking the parameters, determine the optimal parameters, and assess classification performance.

Results

The experimental results identified 12 predictive biomarkers with classification accuracy of 74%. The 12 biomarkers are hydroxybutyrate, magnesium, hydroxybutyrate dehydrogenase, creatine kinase, total protein, high-density lipoprotein cholesterol, cholesterol, absolute value of the lymphocyte, blood urea nitrogen, chlorine, platelet count, and glutamyltranspeptidase. Receiver operating characteristic curve analysis was also used with area under the curve being 0.79.

Conclusions

Some biochemical parameters may be potential biomarkers to predict depression among the subjects with diabetes mellitus.

Background

Diabetes mellitus is a chronic illness affecting about 347 million people worldwide in 2017, and this number is expected to increase more than half by 2035 [1, 2]. The disease will also lead to emotional distress other than physical symptoms and impose psychosocial impacts on life quality, which complicates its management.

Depression and diabetes mellitus are common comorbid conditions [3]. A meta-analysis reported that patients with diabetes mellitus more than doubled the odds of developing depression [3]. Another study described that depression was highly prevalent, affecting approximately 26% of the patients with diabetes mellitus [4]. In addition, depression was found to be associated with a greater number of complications of diabetes mellitus [5]. Furthermore, depression itself is a disabiling disease and imposes...
a significant impact on life quality by undermining physical health [6] and impairing cognitive functions [7]. Therefore, it is not surprising that diabetes mellitus comorbidity with depression is associated with higher morbidity and mortality rates, decreased compliance with treatment, poorer functionality, poor glycemic control, and more expenditure on use to health services [7–12]. A prospective study involving more than 4,000 patients having diabetes mellitus with comorbidity of depression reported a higher risk of developing macrovascular complications, even when variables such as the type of treatment and the existed history of complications before the study were controlled [13]. This highlights the severity of diabetes mellitus in comorbidity with depression and the need to treat both conditions concurrently.

Comorbid depression in diabetes mellitus might be considered not as the result of mental problem only, but more important, as an early sign of a multi-systemic disorder. Thus, medical monitoring is an important component of case assessment. The diagnosis of depression mainly depends on doctors’ clinical experience and scale. The lack of objective indicators, the strong subjective consciousness of doctors and patients, and the avoidance or denial in some symptoms due to patients’ insufficient understanding of the disease interfere with the accuracy of scale score; and this may affect the correct diagnosis of the disease [14–16]. Therefore, it is particularly important to identify objective indicators of depression diagnosis and establish scientific diagnostic methods. Nonetheless, very few approaches have been proposed to facilitate early prediction of depression in patients having diabetes mellitus because objective indicators of laboratory examinations are rare.

Recently, machine learning algorithms have been widely used in the medical sciences. It was reported that machine learning algorithms in combination with smartphone-based data will be a new approach to classify affective states accurately in bipolar disorder [17]. In addition, machine learning methods may be used to predict treatment effect of electroconvulsive therapy (ECT) [18], cognitive behavioral therapy (CBT) [19], and clozapine [20]; or to help diagnostic clarification [21]. According to KIM et al., comprehensive machine-learning methods that adopt supervised classification and appropriate feature selection methods that have interaction with the classifier show particular advantages in predicting complicated disorders with multi-facet etiology such as depression [22]. Support Vector Machine (SVM) is a method of machine learning and is of great significance in accurately identifying depression among patients with diabetes mellitus in clinical practice. This method provides insights for understanding the underlying pathological mechanisms of depression.

Previous studies have reported a high accuracy of over 80% in differentiating patients with depression from healthy controls, using machine learning methods to analyze heart rate variability (HRV) and/or protein markers [22, 23]. Nevertheless, the existing extraction procedures of parameters are usually complex. For example, Danni Kuang et al. [23] need to examine the 64 features of HRV in the Ewing test including the different states—resting, valsalva, deep breathing, and standing states. By contrast, our study was much simpler in that only easy-to-obtain routine biochemical tests and vital signs of patients were needed. By SVM, the best executing classification system can be set up with a small number of parameters that are selected from a variety of biochemical tests and vital signs.
To address this need, we proposed using SVM to identify potential prediction biomarkers for depression in patients with diabetes mellitus.

Methods

Data Acquisition

Biochemical tests and vital signs were obtained from electronic medical records of admissions in West China Hospital of Sichuan University between January 1, 2011 and October 31, 2016. A total of 313 patients were divided into two groups: 135 with both diabetes mellitus and depression (comorbidity group), and 178 with diabetes mellitus alone (DM group). Specifically, the DM group was diagnosed using the ICD-10 categories E10.x - E14.x, and the depression in comorbidity group was diagnosed using the ICD-10 categories F32.x and F33.x. To avoid confounding, patients with other diseases or of non-Han ethnicities were excluded. Each department had different biochemical parameters checked as appropriate, and we analyzed the same biochemical parameters for both groups (Table 2). Written informed consent had been obtained from all patients, and the Institutional Ethics Committee of Sichuan University approved this study.
## Table 2
The 52 biochemical tests and 5 vital signs.

<table>
<thead>
<tr>
<th>52 biochemical tests and 5 vital signs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>
Data Processing

To detect whether biochemical tests and vital signs can function as markers for predicting depression in diabetes mellitus, a RFE-SVM algorithm was adopted to identify the markers and assess the classification performance (Fig. 1).

Before applying the machine learning method to identify predictive markers, covariate regression analysis was performed because age and sex both differed significantly between the DM group and the comorbidity group ($P < 0.05$) (Table 1). After covariate regression analysis, the experimental data were split into training data, evaluation data, and testing data with the proportion of 1/2, 1/4, 1/4 to obtain feature ranking, determine the optimal features, and assess the classification performance. Specifically, the implementation of the machine learning can be summarized as follows:

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic of 313 patients having both diabetes mellitus and depression and having diabetes mellitus alone.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>DM group (n = 178)</th>
<th>Comorbidity group (n = 135)</th>
<th>Statistics</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male (n = 114)</td>
<td>Male (n = 42)</td>
<td>27.97</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Age</td>
<td>54.59 ± 9.64</td>
<td>57.26 ± 8.14</td>
<td>-2.68</td>
<td>0.01</td>
</tr>
</tbody>
</table>

- Determine the feature ranking by recursive feature elimination-based SVM on the training data. The experiments were repeated 1,000 times with 10-fold cross validation.
- Train a SVM classification model on the training data using the liblinear toolbox, and determine the most predictive features using the evaluation data based on the feature ranking obtained above. The feature that ranked No. 1 was first used to train the model, and the performance was evaluated by the evaluation data. Then, the feature that ranked No. 2 was combined to train the model and to compare the performance with the previous one. If the performance of the latter classifier was worse than the former, the feature that ranked No. 2 would be removed. In this way, only the features that could increase the classification accuracy were remained, and finally we obtained 12 biomarkers (Fig. 2).
- Train the classification model on the training data with the selected 12 biomarkers, and assess the performance on the testing data by the measurements of accuracy, AUC, sensitivity, and specificity.

**Statistical analysis**
Statistical analysis was performed using SPSS 20.0. We conducted the test of normality of variances using K-S method and applied log operation to conform to the normal distribution. Two-sample t test and chi-squared test were used for comparison between groups. Statistical significance was set at $P<0.05$ for both tests.

**Results**

In this retrospective study, medical records upon admission of 313 patients were analyzed. Demographic characteristics of the DM group ($n = 178$) and the comorbidity group ($n = 135$) were summarized (Table 1). The two groups differed significantly in age and sex with in comorbidity group had older patients and more women (Table 1).

The two groups differed significantly in the 12 biomarkers of hydroxybutyrate, magnesium, creatine kinase, total protein, high-density lipoprotein cholesterol, cholesterol, absolute value of the lymphocyte, blood urea nitrogen, chlorine, platelet count, glutamyltranspeptidase, and hydroxybutyrate dehydrogenase, with $P<0.05$ (except Hydroxybutyrate Dehydrogenase) (Table 3). The performance of classification of both groups reached 75% for sensitivity, 72% for specificity, 74% for accuracy, and 0.79 for AUC based on ROC analysis (Fig. 3).
Table 3
Biomarkers of experimental results of 313 patients having both diabetes mellitus and depression and having diabetes mellitus alone.

<table>
<thead>
<tr>
<th></th>
<th>DM group (n = 178)</th>
<th>Comorbidity group (n = 135)</th>
<th>Statistics</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydroxybutyrate</td>
<td>0.27 ± 0.59</td>
<td>0.17 ± 0.22</td>
<td>0.12</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Magnesium</td>
<td>0.84 ± 0.11</td>
<td>0.88 ± 0.11</td>
<td>0.06</td>
<td>0.04</td>
</tr>
<tr>
<td>Hydroxybutyrate Dehydrogenase</td>
<td>129.49 ± 28.92</td>
<td>126.50 ± 42.39</td>
<td>-1.74</td>
<td>0.08*</td>
</tr>
<tr>
<td>Creatine Kinase</td>
<td>80.75 ± 37.69</td>
<td>80.94 ± 108.09</td>
<td>-3.00</td>
<td>0.003*</td>
</tr>
<tr>
<td>Total Protein</td>
<td>69.22 ± 6.32</td>
<td>66.47 ± 4.46</td>
<td>0.06</td>
<td>0.04</td>
</tr>
<tr>
<td>High-density Lipoprotein Cholesterol</td>
<td>1.25 ± 0.32</td>
<td>1.40 ± 0.37</td>
<td>-2.74</td>
<td>0.006*</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>4.22 ± 0.79</td>
<td>4.70 ± 0.94</td>
<td>0.06</td>
<td>0.04</td>
</tr>
<tr>
<td>Absolute Value of the Lymphocyte</td>
<td>1.68 ± 0.52</td>
<td>1.67 ± 0.57</td>
<td>0.07</td>
<td>0.01</td>
</tr>
<tr>
<td>Blood Urea Nitrogen</td>
<td>5.98 ± 1.82</td>
<td>5.19 ± 1.97</td>
<td>-3.72</td>
<td>&lt; 0.001*</td>
</tr>
<tr>
<td>Chlorine</td>
<td>105.12 ± 3.40</td>
<td>104.99 ± 4.10</td>
<td>0.07</td>
<td>0.003</td>
</tr>
<tr>
<td>Platelet Count</td>
<td>150.01 ± 57.76</td>
<td>174.28 ± 58.86</td>
<td>0.11</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Glutamyltranspeptidase</td>
<td>36.59 ± 57.55</td>
<td>26.81 ± 20.28</td>
<td>0.12</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

Note: * the results by log operation

Discussion
In this retrospective study, we found 12 important depression biomarkers using SVM. These biomarkers are hydroxybutyrate, magnesium, hydroxybutyrate dehydrogenase, creatine kinase, total protein, high-density lipoprotein cholesterol, cholesterol, absolute value of the lymphocyte, blood urea nitrogen, chlorine, platelet count, and glutamyltranspeptidase, which differentiate depression in patients with diabetes mellitus at an overall classification accuracy of 74%. Twelve identified factors imply that modulation of the inflammatory, immune, energy metabolism, and lipid metabolism pathways were mainly involved in the pathophysiology process of depression in patients with diabetes mellitus.

We found three biomarkers involved in inflammatory and immune pathway including magnesium, absolute value of the lymphocyte, and glutamyltranspeptidase. Depression often coexists with diabetes, metabolic disorders and other diseases, and is linked to inflammatory and oxidative stress [24]. The research found there is a link between depression and insulin resistance [25]. Diabetes can cause a rise in blood sugar and insulin levels and has an effect on inflammation that may contribute to depression. Recent studies have shown that oxidative stress may enhance induction of HO-1 expression, which may
result in insulin resistance and insuficiency [26, 27]. It is clear that increased oxidative stress may lead to insulin resistance and impose an impact on insulin secretion in patients having depressive disorder [27]. One study demonstrated that reducing inflammation through non-drug treatments such as psychological interventions, physical exercises, and meditation can play a role in preventing depression [28]. Magnesium has received great concern over its potential role in the pathophysiology of depression [29–31]. Lymphocytes are produced by lymphoid organs and constitute an important component of immune response. Previous studies indicated a decrease in lymphocyte counts among depressive patients [32], which was in agreement with our findings. One explanation is that inflammatory or chronic stress-induced cellular immunosuppression would cause elevated neutrophils and leukocytes and a relatively reduced lymphocyte counts [32–34]. Glutathione (GSH) is an important substance that protects cells from oxidative stress, and its synthesis requires the participation of Glutamyltranspeptidase [35, 36]. In addition, some researchers reported Glutamyltranspeptidase deficiency in human resulted specific symptoms such as abnormal behavior, mental retardation, and absence seizure [37, 38]. Emerging evidence showed that antidepressant treatments decrease inflammatory and improve mitochondrial dysfunction in patients with depression [39, 40].

We also found five biomarkers potentially related to energy metabolism. These biomarkers are hydroxybutyrate, hydroxybutyrate dehydrogenase, creatine kinase, total protein, and blood urea nitrogen. Hydroxybutyrate is a product of ketone body metabolism pathway. A previous study reported that synthesis and degradation of ketone bodies influenced immensely the pathophysiologic process of depression [41]. Hydroxybutyrate might be helpful for screening depression and predicting its progress [41]. Creatine kinase (CK) activity was reported to increase in the prefrontal cortex, hippocampus, and striatum of rats, and CK levels were increased in the serum of a patient with depression after antidepressant treatment [42, 43]. The normal role of CK is to catalyze the reversible transfer of the phosphoryl group from phosphocreatine to adenosine diphosphate (ADP), and through this process ATP used as energy by cells is generated [44]. The final product of protein metabolism is urea [45]. Hu et al. found that 10% of 260 hemodialysis patients had a diagnosis of depression using the Diagnostic and Statistical Manual of Mental Disorders, 4th edition. They also found that patients with lower monthly income, shorter duration of hemodialysis, and lower levels of blood urea nitrogen were more likely to have a diagnosis of depression. They considered that depression symptoms were usually associated with poor appetite and poor nutrition in hemodialysis patients with depression [46]. We observed lower concentrations of total protein in patients with both diabetes mellitus and depression compared to patients with diabetes mellitus alone, the result is consistent with the research by Peng et al. [27]. The above results suggested that blood biochemical parameters, including urea nitrogen, lactate dehydrogenase, alanine transaminase, uric acid, and total protein, were significantly different between depression patients and healthy controls, and that multiple biochemical parameters in combination may improve the diagnostic effectiveness of depression and the comprehensive management for depressive patients.

Additionally, we found some other biomarkers that may be related to lipid metabolism, such as cholesterol and high-density lipoprotein cholesterol. One of the characteristics of depression is loss of
appetite. Previous studies suggested that LDL-c increase is mostly determined by the severe loss of body fat [47, 48]. Higher level of cholesterol was observed in patients with depression than in controls [27]. In the same way, increased levels of cholesterol were found to be associated with comorbidity of diabetes mellitus and depression in our study.

Changes of creatine kinase, cholesterol, total protein, and high-density lipoprotein cholesterol etc. in blood are not specific to depression and may be present in other psychiatric disorders such as eating disorders [47], schizophrenia [49, 50], and / or bipolar disorder [51, 52]. Researchers suggested that a single biomarker often lacks in sensitivity and specificity [27] and thus may not well distinguish depression from other diseases. Monitoring changes in multiple factor levels will provide a more comprehensive and accurate assessment, which can help us better understand the disease status and characteristics of specific diseases. Although the model of multiple biomarkers is more conducive for the diagnosis of diseases, it is usually used in the diagnosis of cancer instead of nervous system diseases [53, 54]. Our study is advantageous in that laboratory biochemical indexes are routine examinations in clinical settings, which could be obtained with minimal invasiveness, maximal convenience, and low cost, thus having a great potential for wider clinical access and more efficient population screening. Due to the inconsistency of biochemical test results between the two groups, different test items were deleted. The lack of biochemical tests as variables in SVM learning affected accuracy, which is one limitation of the present study. Second, the parameters chosen retrospectively instead of consecutively were inadequate and included only those that were clinically applicable. This may have caused an enrollment bias and an erroneous classification by the algorithm. This is one of the major methodological limitations of the present study, which should be remedied in future investigations using a prospective and consecutive design.

Conclusions

(1) SVM can facilitate clinical diagnosis of depression in patients with diabetes mellitus using commonly available laboratory parameters. (2) Twelve potential biomarkers were identified for depression diagnosis in patients with diabetes mellitus.

Abbreviations

ECT: Electroconvulsive therapy; CBT: Cognitive behavioral therapy; SVM: Support Vector Machine; HRV: Heart rate variability; DM: Diabetes mellitus; CK: Creatine kinase; ADP: Adenosine diphosphate; ATP: Adenosine Triphosphate.

Declarations

Acknowledgements
The authors thank Prof. Dongtao Lin of Sichuan University for copyediting this manuscript and all individuals who have participated in this study.

Authors’ contributions

Author XL designed the study and completed the original draft of the manuscript. Data analysis was performed by authors XL and QZ. Authors RZ and MY managed and extracted data. WD, QW, WJ, and TL managed the literature analyses. Corresponding author XH revised the draft of the manuscript. All authors contributed to and have approved the final manuscript.

Funding

This research was partly funded by National Natural Science Foundation of China [Grant No. 81671344]. The data is owned by the investigators and National Natural Science Foundation of China is not involved in conduction and monitoring of the study, in the data analysis or in the publication process.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Ethical approval was obtained from the Institutional Ethics Committee of Sichuan University. Written informed consent was obtained from each participant.

Consent for publication

Not applicable.

Competing interests

We declared that there was no conflict of interests in this study.

Author details

a Psychiatric Laboratory and Department of Psychiatry, West China Hospital, Sichuan University, Chengdu 610041, P. R. China

b Clinical psychology, Yantai Affiliated Hospital of Binzhou Medical University, Yantai 264005, P. R. China

c School of Computer and Control Engineering, Yantai University, Yantai 264005, P. R. China

d Information Center, West China Hospital, Sichuan University, Chengdu 610041, P. R. China
References


50. Meng XD, Cao X, Li T, Li JP. Creatine kinase (CK) and its association with aggressive behavior in patients with schizophrenia. Schizophrenia research 2018.


Figures
Figure 1
The flowchart of data processing.

Running RFE-SVM 1000 times with 10-fold cross-validation for feature ranking

Determine the optimal feature number

Test the performance of the selected features

Figure 2
The procedure of feature selection on the evaluation data.
Figure 3

ROC curve analysis with AUC value.