Development of a novel neuro-immune and opioid-associated fingerprint with a cross-validated ability to identify and authenticate unknown patients with major depression: far beyond differentiation, discrimination and classification.

Hussein Kadhem Al-Hakeim (1), Suhaer Zeki Al-Fadhel (2), Arafat Hussein Al-Dujaili (3), Andre Carvalho (4, 5), Sira Sriswasdi (6), Michael Maes (7,8,9).

1) Department of Chemistry, College of Science, University of Kufa, Iraq. E-mail:

headm2010@yahoo.com

2) Department of Clinical Laboratory Sciences, College of Pharmacy, University of Kufa, Iraq. E-mail: suhaeralfadel@yahoo.com

3) Senior Clinical Psychiatrist at the Faculty of Medicine, University of Kufa, Iraq. E-mail:

arafat.aldujaili@uokufa.edu.iq

4) Department of Psychiatry, University of Toronto, Toronto, ON, Canada;

5) Centre for Addiction and Mental Health (CAMH), Toronto, ON, Canada;

andre.carvalho@camh.ca

6) Research Affairs, Chulalongkorn University, Bangkok, Thailand

sirasris@gmail.com

7) Department of Psychiatry, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand.

8) Department of Psychiatry, Medical University Plovdiv, Plovdiv, Bulgaria.

9) IMPACT Research Center, Deakin University, Geelong, Australia

dr.michaelmaes@hotmail.com

2

Corresponding author:

Prof. Dr. Michael Maes, M.D., Ph.D.

IMPACT Strategic Research Center

Barwon Health

Deakin University

Geelong, Vic

Australia

E-mail: dr.michaelmaes@hotmail.com

Abstract

Rationale: Major depressive disorder (MDD) is characterized by signaling aberrations in interleukin (IL)-6, IL-10, beta-endorphins as well as mu (MOR) and kappa (KOR) opioid receptors. Here we examined whether these biomarkers may aid in the classification of unknown subjects into the target class MDD.

Methods: The aforementioned biomarkers were assayed in 60 first-episode, drug-naïve depressed patients and 30 controls. We analyzed the data using joint principal component analysis (PCA) performed on all subjects to check whether subjects cluster by classes; support vector machine (SVM) with 10-fold validation; and linear discriminant analysis (LDA) and SIMCA performed on calibration and validation sets and we computed the figures of merit and learnt from the data.

Results: PCA shows that both groups were well separated using the first three PCs, while correlation loadings show that all 5 biomarkers have discriminatory value. SVM and LDA yielded an accuracy of 100% in validation samples. Using SIMCA there was a highly significant discrimination of both groups (model-to-model distance=87.5); all biomarkers showed a significant discrimination and modeling power, while 10% of the patients were identified as outsiders and no aliens could be identified.

Discussion: We have delineated that MDD is a distinct class with respect to neuro-immune and opioid biomarkers and that future unknown subjects can be authenticated as having MDD using this SIMCA fingerprint. Precision psychiatry should employ SIMCA a) to authenticate patients as belonging to the claimed target class and identify other subjects as outsiders, members of another class or aliens; and b) to acquire knowledge through learning from the data by constructing a biomarker fingerprint of the target class.

Introduction

Major Depressive Disorder (MDD) is one of the most prevalent psychiatric disorders worldwide. It has been estimated that 322 million people worldwide live with depression and its prevalence varies by WHO Region, from a low end of 2.6% among males in the Western Pacific Region to 5.9% among females in the African Region [1]. A large body of evidence supports the view that activation of immune-inflammatory pathways may contribute to the pathophysiology of MDD [2, 3]. Furthermore, immune-inflammatory biomarkers have been widely investigated in MDD [4, 5] and some of those have been proposed as candidate biomarkers for both the diagnosis and prediction of treatment response in MDD [6, 7]. In routine psychiatric practice, the diagnosis of MDD mainly relies on the clinical assessment and subjective evaluation of depressive symptoms, whilst no validated biomarkers have been incorporated as part of the diagnostic criteria for MDD [8-10].

Immune-inflammatory mediators play a relevant etio-pathological role in MDD and may affect other pathways including neurotransmission systems (serotonin, glutamate) and neuroprogressive pathways [10]. Endogenous opioids and their cognate receptors in the central nervous system are important regulators involved in the neurobiology of a variety of psychiatric illnesses including MDD [11]. A growing body of research indicates that the endogenous opioid system is directly involved in mood regulation and is dysregulated in MDD [12].

Recently, we reported highly significant aberrations in the opioid system in individuals with MDD relative to controls including elevated serum β-endorphin and κ-opioid receptor (KOR) levels as well as a decrease in μ-opioid receptor (MOR) levels [13]. Furthermore, these alterations in the opioid system were strongly associated with increased cytokine levels, including IL-6, a pro-inflammatory cytokine, and IL-10 an anti-inflammatory cytokine [13].

6

Those inter-group differences are often mis-interpreted as evidence that these abnormalities actually comprise biomarkers for MDD. Consequently, different types of biomarkers have been proposed, including state and trait biomarkers, staging biomarkers, biomarkers of treatment response, etc. [14]. Nevertheless, it is clear that statistical significances (at p=0.05, for example) of biomarker differences between the target class (MDD) and controls do not validate those assays for routine clinical use as predictive biomarkers. In far fewer studies, figures of merit are computed including sensitivity (the proportion of individuals with disease who are correctly identified as such), specificity (the proportion of individuals without disease who are correctly identified as such), positive predictive value (the proportion of individuals with positive test results who in fact have the disease), as well as the area under the receiver operating curve (ROC) which summarizes an assay's overall discriminatory performance. Those figures are then employed to denote that those biomarkers are highly sensitive and specific or predictive for the target class (e.g. MDD or endophenotypes, treatment responder) versus controls. A sensitivity of around 50-65% and a specificity of 96% is commonly used as denoting a good diagnostic performance [15], whereas in some cases a sensitivity of 80% and a specificity of 51% (for the Hypomania-Checklist-32) is already considered to provide enough evidence to support such a claim [16, 17]. These figures of merit are often prematurely used on other study groups not used to compute the diagnostic performance to propose new theories, e.g. the bipolar spectrum theory applied to fibromyalgia, with implications for psychiatric nosology [17].

Moreover, a combination or panel of different biomarkers (e.g. obtained through high-throughput omics-based assays) in a new predictive model is likely to offer a better prediction than the use of single biomarkers. Toward this end, machine learning techniques such as support vector machine (SVM) and linear discriminant analysis (LDA) may be used to build prediction models

[18, 19]. These type of techniques are known as supervised pattern recognition or supervised machine learning methods whereby diagnostic classes or groups are predefined a priori [18, 20-25) [the latter references are frequently used in the Introduction and Methods sections and therefore we do not always repeat these citations].

However, any new (i.e. "candidate") biomarker set should also be validated by taking into account the reliability of the biomarker model to forecast the target class [26, 27]. Toward this end, a modelling procedure applied to a training or calibration set is used in order to construct the new biomarker model, which is subsequently validated on unknown subjects who were not previously considered during model building. Modelling procedures performed without this validation step are likely to result in an overfitted or overtrained model. Several methods may be employed to validate the models and to compute figures of merit. These include bootstrapping, which generates new data sets by resampling existing data points with replacement, cross-validation, which partitions a current data set into equal portions, and split the study group into two sets, namely the training (calibration) and test (validation) sets. For example, cross-validation may be used to optimize the internal parameters of SVM and to assess its accuracy, while a new LDA model should be validated using a train-test splitting method. **Figure 1** shows this procedure comprising a training (calibration) stage and a validation (test) stage.

Nevertheless, there are a number of methodological challenges and limitations with machine learning techniques such as SVM and LDA. For example, LDA entails restrictive assumptions including multivariate normality and homogeneity of variance and can be biased due to outliers and multicollinearity [28, 29]. Moreover, LDA modeling explicitly maximizes the differences between classes which may lead to overfitting [30]. Also, SVM employs decision rules that are challenging to interpret [19], and the technique is prone to over-fitting [31]. As a

8

consequence, the knowledge acquisition provided by LDA and SVM models is limited to interpretation of discriminant scores, support vectors and SVM weights. Most importantly, because LDA and SVM operate based on the assumption that all subjects must belong to one of the pre-specified classes [32], subjects not belonging to any of the presumed classes (outsiders) cannot be identified as such.

Soft independent modelling of class analogies (SIMCA) is another supervised classification method whereby separate models are built for each class (thus one model for MDD and one for controls) using principal component analysis (PCA) [18, 19, 32, 33]. Thus, SIMCA is a class-modeling technique that constructs confidence envelopes around the models of the predefined classes, which, as a consequence, comprise similar subjects with common features or characteristics [18, 19]. Consequently, SIMCA is applied to the class models whereby subjects of a test set or unknown subjects are projected to the class models. Consequently, the orthogonal distances of the subjects to the class models are computed (S_i) as well as their leverages (H_i; the distance of the subject's projection to the model centre). By comparing the subjects' distances to the model subspaces at a predefined significance interval (e.g. alpha=0.05), subjects are allocated to the model classes and identified as belonging to the target class (MDD class members), alternative class members (healthy controls), both classes (hybrids) or none of the classes (outsiders). The latter consists of strangers (subjects with high S_i and H_i values), and subjects with either high S_i or high H_i values [25]. Moreover, healthy controls that intrude into the MDD target class are identified as aliens.

The advantages of SIMCA are that this method a) does not force a subject to belong to a specific class; b) allows the verification of authenticity of target class members, i.e. whether a subject with MDD actually belongs to the claimed MDD class or alternatively is a hybrid, stranger

or an alien; and c) identify new (unknown) subjects as belonging to one of these modeled classes (25, 32]. Another important advantage of SIMCA is that the number of observations does not have to be higher than the number of variables as in other pattern recognition methods [25]. However, SIMCA has never been used to assess the ability of neuro-immune and opioid biomarkers to identify and authenticate MDD patients.

Hence, the present study was conducted to examine whether a biomarker set based on serum levels of IL-6, IL-10, β-endorphins, KOR and MOR may be used to classify (SVM and LDA), identify (SIMCA) and authenticate (SIMCA) unknown subjects into the target class of patients with MDD. Based on this example and our experience with these methods since 1990 [18, 19, 23] the second aim of this paper is to propose new criteria to define new models (decision rules) developed to classify, identify and authenticate patients with MDD (and other neuro-psychiatric and medical disorders).

Subjects and Methods

Participants

The present study recruited 60 drug-free male participants with MDD aged 14–70 year and 30 age-matched healthy males as a control group. The samples were collected at "The Psychiatry Unit", Al-Hakeem General Hospital and at a private psychiatric clinic run by an assistant professor in psychiatry, Najaf Governorate-Iraq during the period of January to July 2017. The diagnosis was made using criteria of the 4th edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) [34]. Severity of depressive symptoms was assessed using the 24-item Hamilton Depression Rating Scale (HDRS) one or two days before blood was drawn and only MDD patients with a total HDRS score >21 were enrolled in the present study. Patients were evaluated by a full

medical history. We excluded subjects with systemic disease that may affect immune parameters, including diabetes type 1, autoimmune disorders, psoriasis, neuro-inflammatory disorders, inflammatory bowel disorder, COPD, and chronic kidney disease. We also excluded MDD patients who were medicated, and subjects with other-axis I diagnosis including schizophrenia, psycho-organic disorders and substance abuse. To eliminate any effects of overt inflammation from other disorders, serum C-reactive protein (CRP) was evaluated in all samples and we excluded subjects with CRP values >6 mg/L. Written informed consent was obtained from all participants, according to the guidelines laid down in the current version of the Declaration of Helsinki, after approval from the ethics committee (IRB) of the College of Science, University of Kufa, Iraq (229-1/2017).

Methods

Five milliliters of venous blood samples were drawn, utilizing disposable needle and plastic syringes, from patients and controls. The samples were transferred into a clean plain tube. Haemolyzed samples were discarded. Blood was left at room temperature for 15 min for clotting, centrifuged 3000 rpm for 10 min, and then serum was separated and transported into two Eppindroff tubes to be stored at -80 °C until analyzed. Serum CRP was measured using a kit supplied by Spinreact®, Spain. The test is based on the principle of the latex agglutination. Commercial ELISA sandwich kits were used to measure serum KOR and MOR (MyBioSource, Inc. CA, USA) and β-endorphin, IL-6, and IL-10 (CUSABIO Co., China). The procedures were followed exactly without modifications according to manufacturer's instructions. The intra-assay coefficient of variation (CV) (precision within an assay) were < 7.0%.

Statistics

The measurements of the 5 biomarkers, namely IL-6, IL-10, β-endorphin, KOR, MOR and the z unit weighted composite score zMOR-zKOR, used here are presented in another study [35]. We selected those variables because they are not affected by smoking, age or body mass index (BMI). Differences in the 5 biomarkers among both classes are displayed as mean (SE) values computed on the z-scores of the 5 biomarkers. We used multivariate general linear model (GLM) analyses to examine the effects of diagnosis (MDD versus controls) on the biomarkers (while adjusting for age, sex and BMI), and tests for between-subjects effects (univariate GLM) were employed to assess the effects of significant explanatory variables on each biomarker.

Use of principal component analysis (PCA) and correlation loadings.

We used PCA as an unsupervised learning method and performed a single, joint PCA on the 5 biomarkers in MDD and controls combined in order to visualize the distribution of both classes in the multivariate space [25]. A standard deviation weighting process, 20-fold cross-validation scheme, and singular value decomposition were used. For the visualization of PC scores in different 2D and 3D dimension combinations (PC1 and PC2, PC1 and PC2 and PC3, PC1 and PC3, PC2 and PC3, etc), the two groups were differentiated by marker colours and shapes. In the same 2D PC plots, we display Hotelling's T² ellipse (alpha=0.05%) to highlight outliers that may influence the model. We set outlier limits based on 0.05% F-residuals and Hotelling's T2 and computed the percentage of variance explained by the consecutively extracted PCs. Various variances including the ratio of calibrated versus validated residual variance, residual variance increased limits, and Q-residuals [25] were also checked. Correlation loadings for each of the biomarkers are shown in the displayed PC dimensions. This plot comprises 2 ellipses, one

indicating 100% explained variance (outer ellipse) and another indicating 50% of the explained variance (the inner ellipse).

Support Vector Machine (SVM)

SVM Classification is a supervised pattern recognition method that is useful as a tool of supervised classification and learning and this method is commonly employed as a data mining method. We applied SVM with linear kernel (linear SVM) and radial basis function kernel (RBF SVM) using the Unscrambler [25]. The input variables were normalized using a standard deviation weighting process and the model was validated using a 10-fold cross-validation scheme. Support vectors, which define the optimal decision boundary separating the MDD and control classes, are selected from original samples located close to the decision boundary [25]. Various values of the capacity factor C are examined to optimize the classification performance. Figures of merit are a) the confusion matrix showing the classification results in the form of predicted classes versus actual classes; and b) the accuracy of the classification that is the percentage of correctly classified cases in the calibration and validation samples. Classification results are shown in a 2D scatter plot with the two best biomarkers as the axes.

Linear discriminant analysis (LDA)

LDA is another supervised pattern recognition method that is commonly used to classify objects. LDA develops a model or decision rule, which is determined by the probability distribution within the classes and which can be employed for allocating new subjects to the most probable class. As with SVM, the most important figures of merit are the confusion matrix and the accuracy of the classification (prediction rate). The loadings of the input variables (biomarkers) on

13

the discriminant scores may be used to extract the features that discriminate the classes. We also display a discrimination plot, which visualizes the LDA results of the training samples or the validating samples where subjects are projected onto the trained LDA model. The axes show the canonical discriminant components whereby subjects (colour and shape-coded according to the predefined class) located close to zero on an axis are associated with that class.

Soft Independent Modeling of Class Analogy or Statistical Isolinear Multiple Component Analysis (SIMCA)

As described in the Introduction, SIMCA is a class modeling technique whereby the predefined classes are modeled separately (one model per class) using PCA. In the training sets (containing both MDD and controls), the number of PCs used to build is delineated by cross-validation and thus may differ between the classes. As such, the class PC models describe the analogous features and similarities between the subjects in the model classes. New subjects (test set) are classified into the class for which they display the best similarity thereby assessing how adequately the model can fit these subjects. As such, SIMCA assigns new subjects (identification) to computed boundaries between classes modeled by PCA. The decision rule to delineate those boundaries is based on two distances, namely Si (the subject to model distance which reflects how far the subject is located from the target class) and Hi or leverage of one subject to the model centre, reflecting how different the subject is from the other subjects in that class [25]. Critical distance limits are calculated for both Si and Hi and these are used for classification purposes by accepting target class members using F tests at a false negative ratio of $\alpha=0.05$ (or 0.01). Using the subjects in the test set, subjects are projected into both class models and the computed distances and their critical limits determine subject assignation.

Figures of merit comprise a) the model-to-model distance which indicates how different the models are with regard to each other. A distance > 3 indicates that the models can be adequately distinguished and a distance between 0 or 1-2 indicates that there are no significant differences between both classes [18, 23, 25, 33]. Large model-to-model distances (e.g. > 20 and >50) imply strongly separated models. b) The modeling power of the input variables reflects the contribution of that feature to the modeled classes. Values close to 1 indicate a strong impact of that biomarker on the modeled class, whereas values < 0.300 indicate low modeling power. Features with a low modeling power may be eliminated from the final model (feature selection). c) The discrimination power reflects the power of the features to discriminate both class models and therefore the impact of the biomarkers to classify objects. Features with low discriminatory power may be eliminated from the model. d) Identification of the subjects as members of a class, hybrids (belonging to two classes), outsiders (not belonging to any class), strangers or aliens and authentication of subjects as belonging to the target class [25, 32].

We have performed two different SIMCA analyses, namely a) we employed all MDD patients and all controls to model both classes and assessed the model-to-model distance and modeling and discriminatory power of the features. This analysis does not aim to make a predictive model of the target groups, but rather examines the separation between the PCA models. Therefore, only gross or extreme outliers are eliminated, namely outliers that completely influence the PCA model for example when one outlier determines the variance in one PC leaving most of the variance in the following PC. Nevertheless, subjects with high residual and leverage values are used to construct the models as long as they are not influential and, therefore, subjects who are different from the other subjects are not per se eliminated from this analysis. b) Secondly, we used a training (50% of the MDD and 50% of the control subjects) set to estimate predictive models of

the MDD and control classes and a test (the remaining 50% of the MDD and controls samples) set to validate the models. The PCA models of both classes are constructed by eliminating outliers as detected in score, influence and stability plots, sample residual vs samples and Hotelling's T2 vs samples plots and through inspection of residual values and leverages. Subjects with increased leverages and residual values may be eliminated from the model when they are influential outliers and also subjects with more extreme F-residual values and Hotelling's T^2 values will be eliminated if they show biomarker values that are not relevant to the model.

In this study we used different SIMCA plots. a) SIMCA classification as visualized using Cooman's plot which is a Si to Si plot that shows the subject to model distances plotted for both classes along orthogonal axes. b) The Si vs Hi plot displays Si (subject to model distance) and Hi (leverage) for a given class as well as the class limits at α=0.05. The Si/S0 vs Hi plot shows the residual standard deviation (relative distance of the subjects to the class model) as well as Hi. The class membership limits allow to classify subjects: when they fall within the class limits, subjects belong to the target class and are therefore authenticated, whereas when they fall outside the limits they are outsiders, either strangers or members of another class. This plot also allows to detect aliens, namely members of another class that intrude into the target class critical limits. c) The discriminatory power plot which displays the discriminatory power for all biomarkers and thus shows their power to discriminate the class models.

Results.

GLM analysis

Figure 2 shows the mean (SE) values of the z scores of the 5 biomarkers in both MDD patients and controls. Multivariate GLM analysis (adjusted for age, BMI and smoking) shows that

16

there is a significant effect of diagnosis (F=66.51, df=1/5/80, p<0.001) and that all 5 biomarkers are significantly increased in MDD as compared with controls (all at p<0.001). Age (F=1.73, df=5/80, p=0.138), BMI (F=1.28, df=5/80, p=0.282) and smoking (F=1.68, df=5/80, p=0.148) are not significant,

PCA

Figure 3 shows the PC score plot, namely PC1 vs PC2, which visualizes the actual subject distribution in the 2D space made up by both PCs. PC1 and PC2 explains together 72% of the variance, while PC3 explains 14% of the variance. Thus, the three first PCs explain a large part of the variance (86%) and, therefore, the separation and loadings of the biomarkers on the PCs can be interpreted accurately. Both classes group together with MDD patients clustering at the right-hand side of the plot, whereas controls cluster together at the left-hand side. There is no overlap between the two groups and all subjects are well separated, although there are no large boundaries (streets) between both groups. The PC1 vs PC3 plot shows a similar distribution pattern (not shown). **Figures 4a and 4b** show the correlation loadings of the 5 biomarkers on PC1 vs PC2, and PC1 vs PC3, showing that all 5 biomarkers contribute to the differentiation of both classes.

SVM

The SVM with linear kernel delineated 8 support vectors (including 4 controls and 4 MDD patients). The confusion matrix shows that all 30 controls and all 60 MDD patients were correctly classified and that the accuracy of the classification is 100% both before and after 10-fold cross-validation. **Figure 5** shows a plot of the classification results with IL-10 and zKOR-zMOR, which yield the best differentiation of the two classes, as input variables. In order to examine the ability

of the model to predict class membership of new subjects, we have rerun the SVM analysis on a training (50% of the patients and the controls) and a validation (the remaining 50% of patients and controls) set. Seven support vectors were selected to construct the model (3 controls and 4 MDD patients). Projecting the validation set on the SVM model showed an accuracy of 100%. Again, the two classes are best separated when zKOR-zMOR and IL-10 were used as inputs.

LDA

LDA was performed on training (50% of MDD and 50% of controls) and test (remaining 50% of MDD and controls) sets. The confusion matrices of both the training and validation samples showed that all subjects were correctly classified either as controls or MDD patients. **Figure 6** shows the LDA discrimination plot for the subjects in the training set. Both classes are well separated and are located relatively close to zero on the corresponding axes.

SIMCA

In order to construct a first SIMCA model on all MDD patients and controls, we deleted three healthy controls (statistical outliers) from the model, while there was no evidence that any of the MDD patients should be omitted from the model. MDD and controls were each modeled using 4 PCs. We found that the model-to-model distance was 87.5 indicating a strong separation of both classes. All controls were indentified as belonging to the control class, whereas all MDD patients were identified as belonging to the MDD model, except one MDD patient who was identified as an outsider. This subject showed a leverage that was quite similar to the other MDD members, while the Si/S0 ratio was higher than the other MDD members. **Figure 7** shows the discrimination plot with discriminatory power of all biomarkers in separating MDD from controls with in

descending order of discriminatory power: IL-10, zKOR_zMOR, KOR, β-endorphin and IL-6. The modeling power of all biomarkers was highly significant in controls (all > 0.95) and MDD patients (all > 0.65). Consequently, no features were excluded to compute the final model.

A second model was constructed on a calibration set (50% of MDD patients and 50% of controls), whilst the remaining subjects were used as the validation set. During the training step, we eliminated one control as a statistical outlier (same subject that was also excluded from the first SIMCA model). MDD and controls were each modeled using 4 PCs. The distance between the models was 131.7 and inspection of membership classification showed that 3 MDD patients were indentified as outsiders. None of the controls intruded the MDD hyperspace, while one control was identified as an outsider because the Si/S0 value was somewhat increased, while its leverage fell within the normal range. **Figure 8** shows the Si/S0 vs Hi plot and the 27 MDD patients that were authenticated as well as the three MDD patients that were identified as outsiders. As such the sensitivity of the model in authenticating MDD subjects is 90% with a specificity of 100%.

Discussion

The first major finding of this study is that patients with MDD are highly significantly separated from healthy controls using neuro-immune and opioid biomarkers as input variables in different supervised learning techniques. These data show that interrelated aberrations in pro-inflammatory (IL-6) and immune-regulatory cytokines and the opioid system are useful as a decision rule to classify patients with MDD and, thus, that an interrelated upregulation of those pathways is involved in the pathophysiology of MDD [13]. Using SVM and LDA, we developed and validated biomarker decision rules that achieved 100% accuracy, and therefore these models may be employed to classify unknown subjects into MDD or control classes. While LDA has been

used since the 1990s to classify patients with MDD using biomarkers [18, 23], SVM was only recently used to classify MDD patients based on brain imaging, metabolomics and EEG-based functional connectivity data (e.g. [36]). However, the figures of merit of the most recent SVM studies show often a reasonable sensitivity (78% to 90%) but lower sensitivity (32.0-79.7%) [36, 37]. In other machine learning studies, such as EEG-based functional activity and NMR-based glucose-lipid signaling, much better figures of merit were obtained [38, 39].

The second major finding of this study is that SIMCA is a better tool to learn from the data thereby obtaining a more precise identification of the subjects as compared to LDA and SVM. At first sight it could be concluded that LDA and SVM are more accurate than SIMCA as the former methods correctly classified more objects. An increased alpha error (high number of wrongly rejected patients) is a known problem of SIMCA [40] and, consequently, some authors concluded that SIMCA is not as adequate as other supervised techniques [41].

Nevertheless, as stated in the Introduction, there are a number of problems with LDA and SVM, the most notably being that these techniques force subjects to be classified as cases or controls. SIMCA, on the other hand, improves precision of the identification process by detecting statistical outliers, hybrids, outsiders, aliens and strangers. This is an important benefit as there are often (especially in neuro-psychiatric research) subjects that may be classified as hybrids (in the case of overlapping classes), while other subjects do not belong to any class (outsiders). Also, subjects belonging to the target class but showing a high leverage may be detected by SIMCA through inspection of Cooman's and Si/S0 vs Hi plots. The presence of outsiders, hybrids and strangers is likely to decrease the success rate of LDA and SVM decision rules when predicting class membership of unknown subjects. Finally, SIMCA may also authenticate cases belonging to the target class as real members of the claimed class. As such, we detected that in our study sample

90% of the MDD patients in the test set could be established as real members and that three MDD patients showed higher S_i values (wrongly rejected), but normal leverage, while no aliens could be detected (or no controls intruded the MDD hyperspace). As such, our SIMCA model is useful to identify and authenticate new unknown cases. **Table 1** shows a comparison of the different supervised methods used in our study and their ability to differentiate, discriminate, classify, validate, identify and authenticate subjects with MDD and controls.

A second major disadvantage of LDA and SVM is that knowledge acquisition through interpretation of the models learned from the data is limited. The discriminant scores and SVM weights of the models are difficult to interpret and in addition are prone to multicollinearity especially when there are strong intercorrelations between the input variables [19, 29, 42]. In contrast, evaluation of SIMCA results allows a more profound knowledge acquisition including feature extraction of the PCA models, the model-to-model distance, and the modeling and discriminatory power of the features. Thus, by constructing envelopes around the data points of both classes using PCA, one can describe the common properties or similarities of the subjects in a class. The model-to-model distance allows to evaluate the degree of separation of both classes [18, 23] whereby a distance > 3 indicates that both classes are well separated. This may address important and heavily debated issues including the quantitative versus qualitative theories of melancholia versus simple major depression [18, 23], deficit versus non-deficit schizophrenia [19] or Myalgic Encephalomyelitis versus chronic fatigue syndrome [24]. Since SIMCA constructs multi-dimensional envelopes (ellipses) around the data points of cases and controls, a large distance between both models indicates that those models occupy distinct spaces in the multivariate hyperspace and, thus, that they are qualitiatively distinct [18]. Another relevant question is whether a critical limit can be proposed to ensure a good accuracy of the SIMCA model

21

in predicting class memberships based on the model-to-model distance. Based on our experience with SIMCAs performed on most of our data sets since 1990, we would recommend very high values, e.g. > 20 and >50 although these large distances are rarely obtained in neuro-psychiatric research using current biomarkers, which most often tend to overlap between cases and controls. For example, Kanchanatawan et al. (2018) [19] observed that cases with deficit schizophrenia were well-separated from non-deficit schizophrenia using biomarkers and cognitive tests, however these differences were insufficient to use the model to identify unknown subjects (model-to-model distances were around 4-15). Other important learning points are the modeling and discriminatory powers which in addition may be used to reduce the number of features when the modeling and discriminatory power are below a certain limit. In the present study we found that all 5 biomarkers were very useful to model the MDD and control classes, and that IL-10, the KOR-MOR ratio and KOR were top-3 features discriminating both classes.

Nevertheless, inspection of the Si/S0 vs Hi plot shows that some MDD patients are too far away (high Si distance) from the class limits (wrongly rejected patients), whereas part of the controls show a tendency to come closer to the MDD class limits. One method to further improve authentication of these subjects is to add more predictors to our decision rule. Such variables should not only comprise peripheral biomarkers, but also brain imaging, EEG, metabolomics (see above) and clinical data [8]. Hence, using a combination of biomarkers and clinical data (severity of illness, severity of symptom dimensions) can considerably improve the detection and prediction of "pathway phenotypes" or pathway-related diagnostic classes [8]. Likewise, the results of the present study suggest that our biomarkers could be added to selected clinical diagnostic criteria to learn from these combined data and pathway-phenotypes. Other examples of newly developed pathway-phenotypes are tryptophan catabolites combined with somatization symptoms (using

SIMCA); pro-inflammatory cytokines combined with post-exertional malaise in chronic fatigue syndrome (using SIMCA); IgA responses to tryptophan catabolites combined with cognitive disorders and negative symptoms in schizophrenia (using SIMCA); and antioxidant levels combined with staging features in affective disorders (using Partial Least Squares path modeling) [8, 19, 24, 43].

Nevertheless, our SIMCA, LDA and SVM decision rules and the SVM classifiers developed by other authors have a number of limitations. Firstly, we have a priori selected our 5 biomarkers not to be affected by age, BMI and smoking and performed our study in males only. We also examined only drug-naïve patients and used many exclusion criteria and therefore our biomarker set may not have full generalizibility because MDD patients often exhibit many comorbidities, including anxiety, substance use and medical disorders. Moreover, in the clinical scenario most patients take antidepressants, which are known to increase the production of IL-10 [44] and therefore our decision rule may be less accurate in medicated patients. Furthermore, decision rules should also be validated against other psychiatric and medical diagnoses to delineate the accuracy of the model with regard to other disorders. Another question is whether such a decision rule is at all needed in MDD patients with a known history of depression because psychiatrists can make the diagnosis based on clinical and staging features. Therefore, the benefits of our developed decision rule are restricted to the following important matters: a) the model is useful to identify and authenticate drug-naïve patients in a first phase of illness when the clinical diagnosis of MDD may be less evident; and b) knowledge acquisition to further decipher the pathophysiology of depression and pathway phenotypes.

Based on the above and our previous machine learning publications we would recommend using 4 different approaches to construct new models aimed at predicting memberships of

unknown subjects. **Table 2** lists the different methods that may be used towards this purpose and our subjective criteria defining accuracy of the models. In a preliminary analysis, analyses of variance should show significant differences between MDD patients and controls, while the area under the boostrapped ROC curve (2000 bootstraps) should be > 0.900. Secondly, a joint PCA performed on all subjects should be performed to check whether the subjects cluster by classes. If there is an overlap between the classes on the PCs explaining the highest variance the separation of the classes using machine learning is probably insufficient to pass the stage of cross-validation, especially when using SIMCA. Thirdly, SVM should be applied and the cross validated model should have an accuracy > 95.0%. Fourthly, LDA can be applied to training and validation sets and its accuracy should be > 95.0%. Finally, SIMCA is helpful to identify and authenticate unknown subjects when > 90% of the validation subjects of the target group are authenticated and when there are <10% aliens.

The incorporation of biomarkers for the identification and biological validation of mental disorders including MDD and the construction of new pathway-phenotypes is an awaited achievement of the emerging field of precision psychiatry, which may radically change the way this speciality is conceived and practiced in the near future [8, 24, 45]. More precisely, the results of the current study indicate that the incorporation of peripheral biomarkers in a SIMCA model may aid in the authentication of patients with MDD. In addition, we provide further evidence that the use of machine learning approaches, and especially SIMCA, is a pre-requisite for this endeavor [24, 46]. In addition, our new findings open relevant research directions. First, our model deserves further validation in larger and more heterogeneous samples comprising "real world" individuals with several co-occurring medical and mental disorders as well as otherwise healthier subjects. In addition, our SIMCA findings provide important suggestions for the use and interpretation of "big

data" approaches [46] which could be further refined in future studies. Finally, we are aware that our effort provides a relevant yet preliminary step in the development of precision psychiatry, which ultimately aims to integrate a wider array of data pertaining to individual variations in genes, environment and a lifestyle to diagnose and treat mental disorders using an individualized data-driven approach [45].

Funding

Conflict of interest

The authors have no conflict of interest with any commercial or other association in connection with the submitted article.

Author's contributions

All the contributing authors have participated in preparation of the manuscript.

References

1. World Health Organization (WHO), 2017. Depression and Other Common Mental Disorders. WHO reference number: WHO/MSD/MER/2017.2.
2. Maes M. (1995) Evidence for an immune response in major depression: a review and hypothesis. Prog Neuropsychopharmacol Biol Psychiatry 19:11-38.
3. de Melo LGP, Nunes SOV, Anderson G, Vargas HO, Barbosa DS, Galecki P, Carvalho AF, Maes M. (2017) Shared metabolic and immune-inflammatory, oxidative and nitrosative stress pathways in the metabolic syndrome and mood disorders. Prog Neuropsychopharmacol Biol Psychiatry 78:34-50.
4. Maes M. (1999) Major depression and activation of the inflammatory response system. Adv Exp Med Biol 461:25-46.

5. Köhler CA, Freitas TH, Maes M, de Andrade NQ, Liu CS, Fernandes BS, Stubbs B, Solmi M, Veronese N, Herrmann N, Raison CL, Miller BJ, Lanctôt KL, Carvalho AF. (2017) Peripheral cytokine and chemokine alterations in depression: a meta-analysis of 82 studies. Acta Psychiatr Scand 135(5):373-387.

6. Al-Hakeim HK, Twayej AJ, Al-Dujaili AH. (2018) Reduction in serum IL-1β, IL-6, and IL-18 levels and Beck Depression Inventory-II score by combined sertraline and ketoprofen administration in major depressive disorder: A clinical trial. Neurol Psychiatry Brain Res 30:148-153.

7. Köhler CA, Freitas TH, Stubbs B, Maes M, Solmi M, Veronese N, de Andrade NQ, Morris G, Fernandes BS, Brunoni AR, Herrmann N, Raison CL, Miller BJ, Lanctôt KL, Carvalho AF. (2018) Peripheral alterations in cytokine and chemokine levels after antidepressant drug treatment for major depressive disorder: systematic review and meta-analysis. Mol Neurobiol 55(5):4195-4206.

8. Maes M, Rief W. (2012) Diagnostic classifications in depression and somatization should include biomarkers, such as disorders in the tryptophan catabolite (TRYCAT) pathway. Psychiatry Res 196(2-3):243-9.

9. Gururajan A, Clarke G, Dinan TG, Cryan JF. (2016) Molecular biomarkers of depression. Neurosci Biobehav Rev 64:101-133.

10. Hacimusalar Y, Eşel E. (2018) Suggested biomarkers for major depressive disorder. Noro Psikiyatr Ars 55(3): 280-290.

11. Crowley NA, Kash TL. (2015) Kappa opioid receptor signaling in the brain: circuitry and implications for treatment. Prog Neuropsychopharmacol Biol Psychiatry 62: 51-60.

12. Carlezon WA Jr, Krystal AD. (2016) Kappa-opioid antagonists for psychiatric disorders: from bench to clinical trials. Depress Anxiety 33(10):895-906.

13. Al-Fadhel SZ, Al-Hakeim HK, Al-Dujaili AH, Maes M. (2019) IL-10 is associated with increased mu-opioid receptor levels in major depressive disorder. Eur Psychiatry 21(57):46-51.

14. Berk M, Kapczinski F, Andreazza AC, Dean OM, Giorlando F, Maes M, Yücel M, Gama CS, Dodd S, Dean B, Magalhães PV, Amminger P, McGorry P, Malhi GS. (2011) Pathways underlying neuroprogression in bipolar disorder: focus on inflammation, oxidative stress and neurotrophic factors. Neurosci Biobehav Rev 35(3):804-17.

15. Carroll BJ. (1982) Clinical applications of the dexamethasone suppression test for endogenous depression. Pharmacopsychiatria 15(1):19-25.

16. Angst J, Adolfsson R, Benazzi F, Gamma A, Hantouche E, Meyer TD, Skeppar P, Vieta E, Scott J. (2005) The HCL-32: towards a self-assessment tool for hypomanic symptoms in outpatients. J Affect Disord 88(2):217-33.

17. Alciati A, Sarzi-Puttini P, Batticciotto A, Torta R, Gesuele F, Atzeni F, Angst J. (2012) Overactive lifestyle in patients with fibromyalgia as a core feature of bipolar spectrum disorder. Clin Exp Rheumatol 30(6 Suppl 74):122-8.

18. Maes M, Schotte C, Maes L, Cosyns P. (1990) Clinical subtypes of unipolar depression: Part II. Quantitative and qualitative clinical differences between the vital and nonvital depression groups. Psychiatry Res 34(1):43-57.

19. Kanchanatawan B, Sriswasdi S, Thika S, Sirivichayakul S, Carvalho AF, Geffard M, Kubera M, Maes M. (2018) Deficit schizophrenia is a discrete diagnostic category defined by neuro-immune and neurocognitive features: results of supervised machine learning. Metab Brain Dis 33(4):1053-1067.

20. Wold S, Sjostrom M, 1977. SIMCA: A Method for Analysing Chemical Data in Terms of Similarity and Analogy. Chemometrics: Theory and Application, cahter 12, ACS Symposium Series No 52, Washington, pp 243-282.

21. Harman HH. Modern Factor analysis, 3rd edition, revised. University of Chicago Press, 1976.

22. Derde M-P, Coomans D, Massart DL. (1984) SIMCA (soft independent modeling of class analogy) demonstrated with characterization and classification of Italian olive oil. J Assoc Official Analytical Chem 67:721-726.

23. Maes M, Maes L, Schotte C, Cosyns P. (1992) A clinical and biological validation of the DSM-III melancholia diagnosis in men: results of pattern recognition methods. J Psychiatr Res 26(3):183-96.

24. Maes M, Twisk FN, Johnson C. (2012) Myalgic Encephalomyelitis (ME), Chronic Fatigue Syndrome (CFS), and Chronic Fatigue (CF) are distinguished accurately: results of supervised learning techniques applied on clinical and inflammatory data. Psychiatry Res 200(2-3):754-60.

25. CAMO, 2019. The Unscrambler Appendices: Method References. 2019. As assessed 19-3-2019. www.camo.com/helpdocs/The_Unscrambler_Method_References.pdf.

26. Ripley BD. Pattern Recognition and Neural Networks. Cambridge University Press. (2008) Paperback. ISBN-13: 978-0521717700.

27. Brownlee J. Supervised and Unsupervised Machine Learning Algorithms; Understand Machine Learning Algorithms. March 16, 2016. As assessed https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/. As accessed online 19 March 2019.

28. Green, S.B. Salkind, N. J. & Akey, T. M. (2008). Using SPSS for Windows and Macintosh: Analyzing and understanding data. New Jersey: Prentice Hall.

29. Çokluk-Bökeoğlu Ö, & Büyüköztürk Ş. (2008). Discriminant function analysis: Concept and application. Eğitim araştırmaları dergisi, (33), 73-92.

30. Qiao Z, Zhou L, Huang JZ. (2008) Effective Linear Discriminant Analysis for High Dimensional, Low Sample Size Data In Proceeding of the World Congress on Engineering (2), Citeseer, 2-4.

31. Cawley GC, Talbot NCL. (2010) On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. J Machine Learning Res 11;2079-2107.

32. Alamprese C, Casiraghi E. (2015) Application of FT-NIR and FT-IR spectroscopy to fish fillet authentication. Lebensmittel-Wissenschaft und-Technologie 63(1):720-725

33. Derde M-P, Massart DL. (1982) Extraction of information from large data sets by pattern recognition. Fresmius Zeitschrift for Analytische Chemie 313:484-495.

34. American Psychiatric Association. Diagnostic and statistical manual of mental disorders.4. Washington, DC: Author; 2000, text revision.

35. Al-Hakeim HK, Al-Fadhel SZ, Al-Dujaili AH, Maes M. (2019) Depression and immune activation are strongly associated with an increased kappa versus mu opioid receptor ratio. (In press).

36. Chu Y, Zhao X, Zou Y, Xu W, Han J, Zhao Y. (2018) A Decoding Scheme for Incomplete Motor Imagery EEG With Deep Belief Network. Front Neurosci. 12:680.

37. Yang J, Zhang M, Ahn H, Zhang Q, Jin TB, Li I, Nemesure M, Joshi N, Jiang H, Miller JM, Ogden RT, Petkova E, Milak MS, Sublette ME, Sullivan GM, Trivedi MH, Weissman M, McGrath PJ, Fava M, Kurian BT, Pizzagalli DA, Cooper CM, McInnis M, Oquendo

MA, Mann JJ, Parsey RV, DeLorenzo C. (2018) Development and evaluation of a multimodal marker of major depressive disorder. Hum Brain Mapp 39(11):4420-4439.

38. Mumtaz W, Ali SSA, Yasin MAM, Malik AS. (2018) A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD). Med Biol Eng Comput 56(2):233-246.

39. Zheng H, Zheng P, Zhao L, Jia J, Tang S, Xu P, Xie P, Gao H. (2017) Predictive diagnosis of major depression using NMR-based metabolomics and least-squares support vector machine. Clin Chim Acta 464:223-227.

40. De Maesschalk A, Candolfi A, massart DL, Heuerding S. Decision criteria for soft independent modelling of class analogy applied to near infrared data. Chemometrics Intelligent laboratory Systems 1999; 47 (1): 65-77.

41. Racz A, Gere A, Bajusz D, Heberger K. Is soft independent modeling of class analogies a reasonable choice for supervised pattern recognition? RSC Advances 2019; 16, in press.

42. Naes T, Mevik B-H. (2001) Understanding the collinearity problem in regression and discriminant analysis. J Chemometrics 15(4):413-426.

43. Maes M, Landucci Bonifacio K, Morelli NR, Vargas HO, Barbosa DS, Carvalho AF, Nunes SOV. (2019) Major differences in neurooxidative and neuronitrosative stress pathways between major depressive disorder and types I and II bipolar disorder. Mol Neurobiol 56(1):141-156.

44. Maes M, Song C, Lin AH, Bonaccorso S, Kenis G, De Jongh R, Bosmans E, Scharpé S. (1999) Negative immunoregulatory effects of antidepressants: inhibition of interferon-gamma and stimulation of interleukin-10 secretion. Neuropsychopharmacology 20(4):370-9.

45. Fernandes BS, Williams LM, Steiner J, Leboyer M, Carvalho AF, Berk M. The new field of 'precision psychiatry'. BMC Med. 2017;15(1):80.

46. Passos IC, Mwangi B, Kapczinski F. Big data analytics and machine learning: 2015 and beyond. Lancet Psychiatry 2016;3(1):13-15.
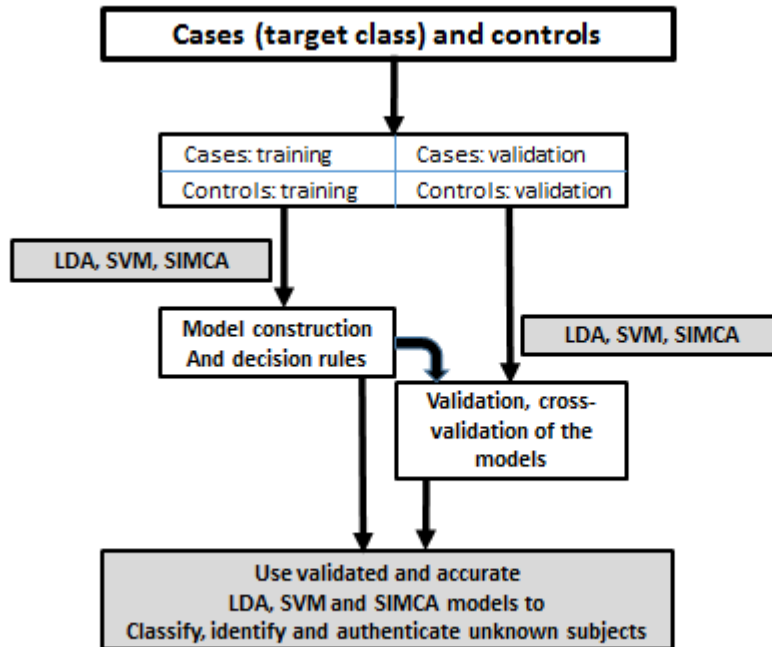
28

Figure 1. Procedure to build a model (decision rule) to classify unknown subjects.

The procedure entails two stages, namely a training (calibration) stage and a test (validation) stage. In the training stage, supervised pattern recognition methods are applied to biomarker data to construct models that best distinguish samples according to predefined classes. In the validation stage statistical tests are employed to evaluate the accuracy of the trained models by projecting cases and controls to the models. Accurate models can then be used to classify unknown subjects.

29
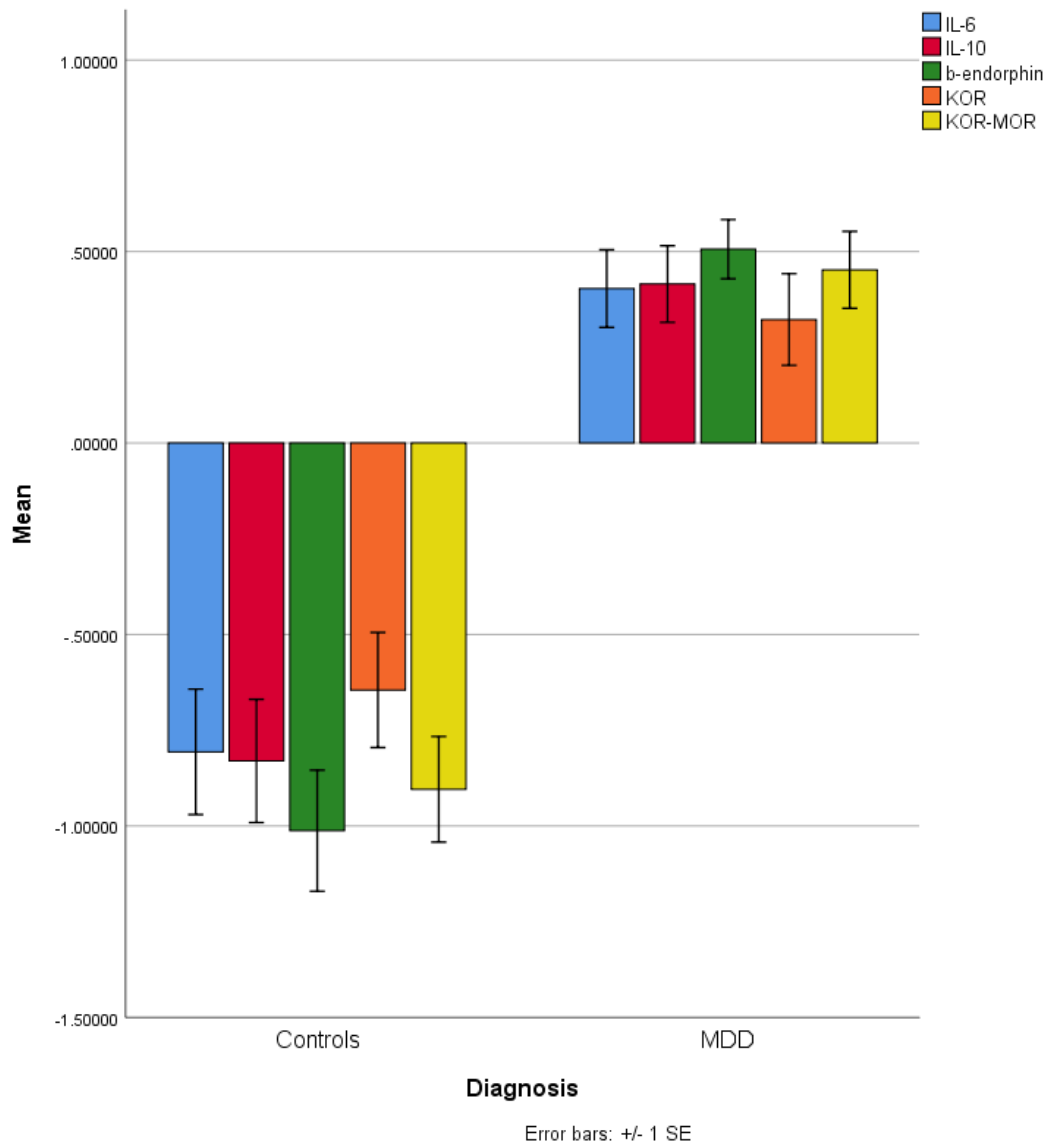
Figure 2 Mean (SE) values of the z scores of the 5 biomarkers in both MDD (major depressed) patients and controls. Mean values of all biomarkers are signifcantly higher in MDD than in controls.
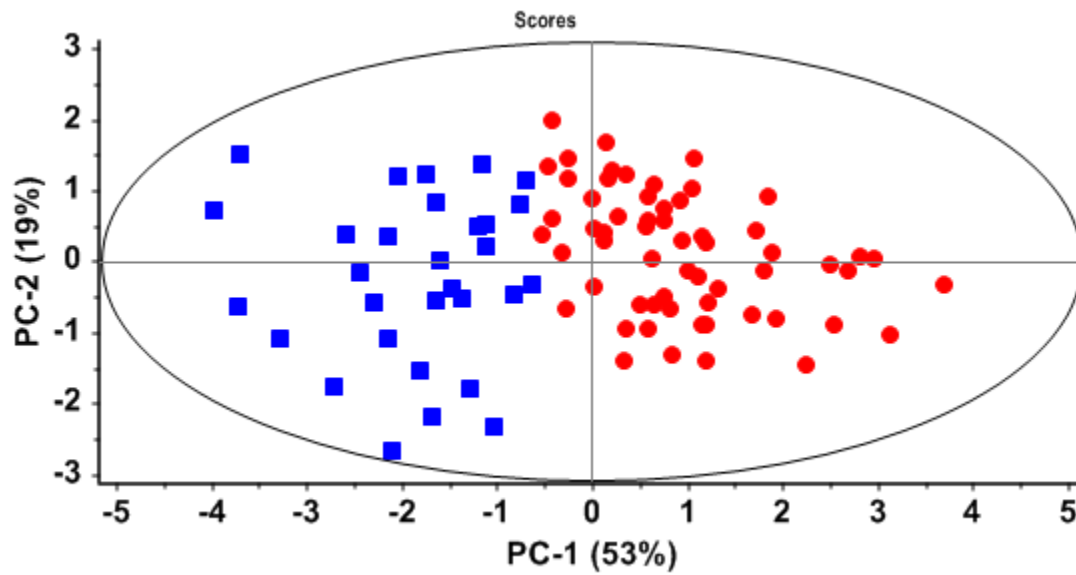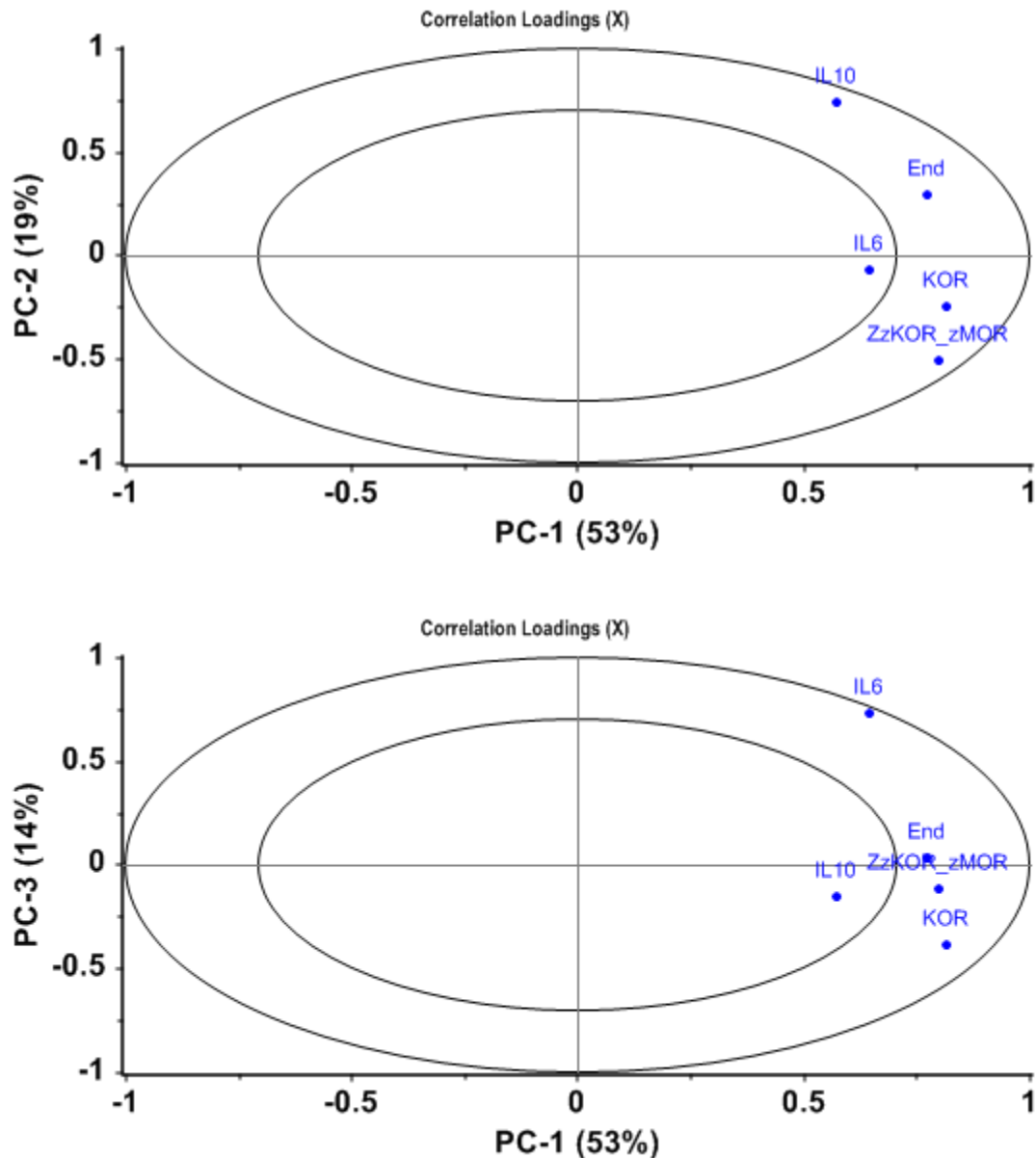
Figure 3 Principal Component (PC) score plot.

This plot shows a 2D display of PC1 vs PC2 visualizing the subject distribution in a 2D space, indicating that the 5 biomarkers allow a clear differentiation of major depressed patients (red circles) vs controls (blue squares).

Figures 4a and 4b Correlation loadings of the 5 biomarkers on principal component (PC)1 vs

PC2.

Figure 4a shows that 4 variables are located between both ellipses, namely interleukin (IL-10), β-

endorphin, κ opoid receptor (KOR) and zKOR-zMOR (μ opoid receptors) and that the three opioid

biomarkers are important is separating both classes along PC1, whilst IL-10 adds to the

differentiation via its loading on PC1 but also on PC2. **Figure 4b** shows that IL-6 loads highly on

32

PC3 and contributes to the separation of both classes through its loadings on PC1 but also on PC3.

Those 5 biomarkers are close together in the plot suggesting that they show significant and positive

intercorrelations.

zKOR-zMOR: z unit weighted composite score computed as z score of κ opioid receptor levels –

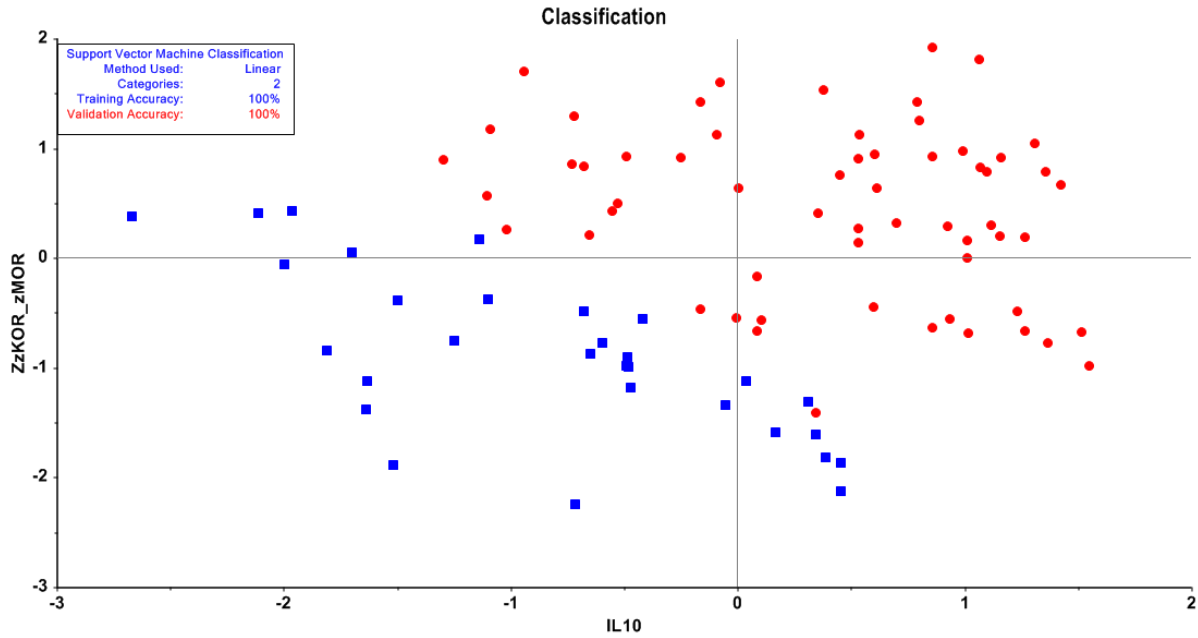z scores of μ opioid receptor levels.

Figure 5 Plot of the classification results obtained by support vector machine.

IL-10: interleukin-10; zKOR-zMOR: z unit weighted composite score computed as z score of κ

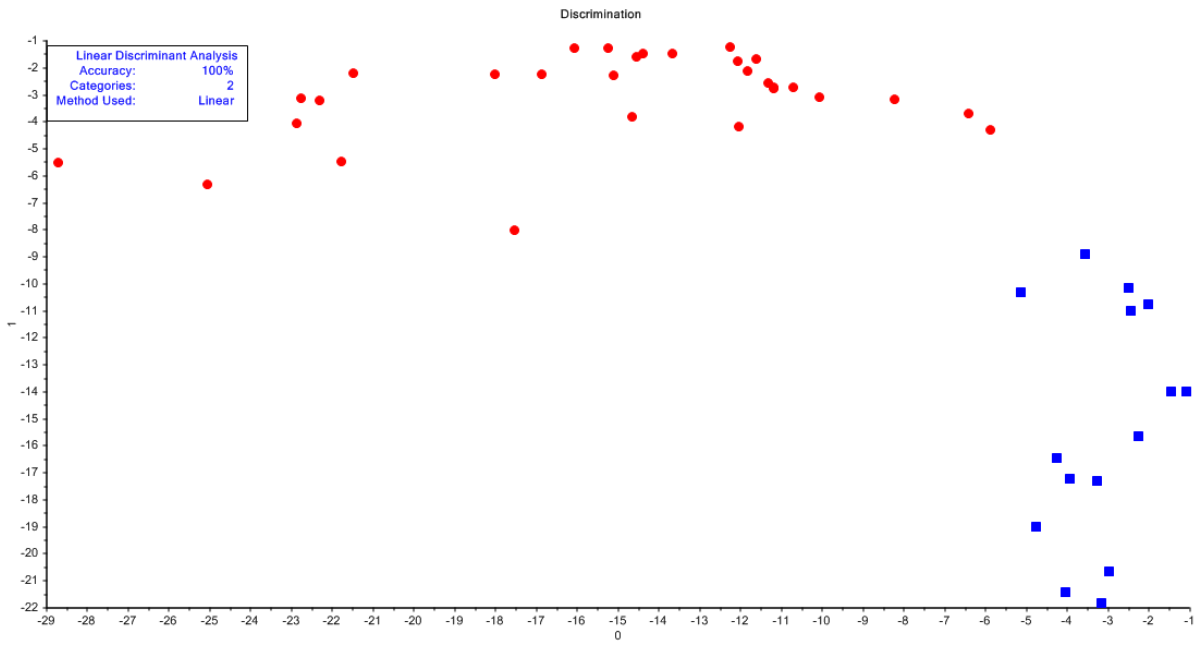opioid receptor levels – z scores of μ opioid receptor levels.

Figure 6 Linear Discriminant Analysis plot.

The decision rule is computed on a training set, which shows a 100% accuracy rate in the training
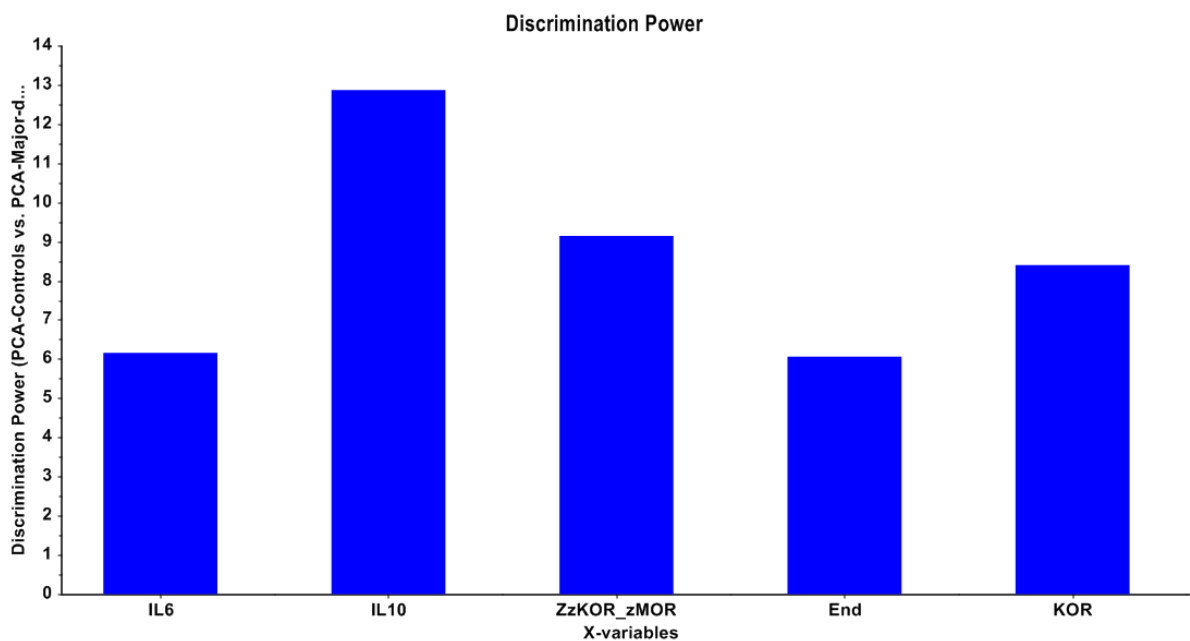
and test sets.

Figure 7 Results of SIMCA showing the discrimination plot.

This plot shows the discrimination power of the five biomarkers used in the current study.

IL: interleukin

KOR: κ opioid receptor

zKOR-zMOR: z unit weighted composite score computed as z score of κ opioid receptor levels –
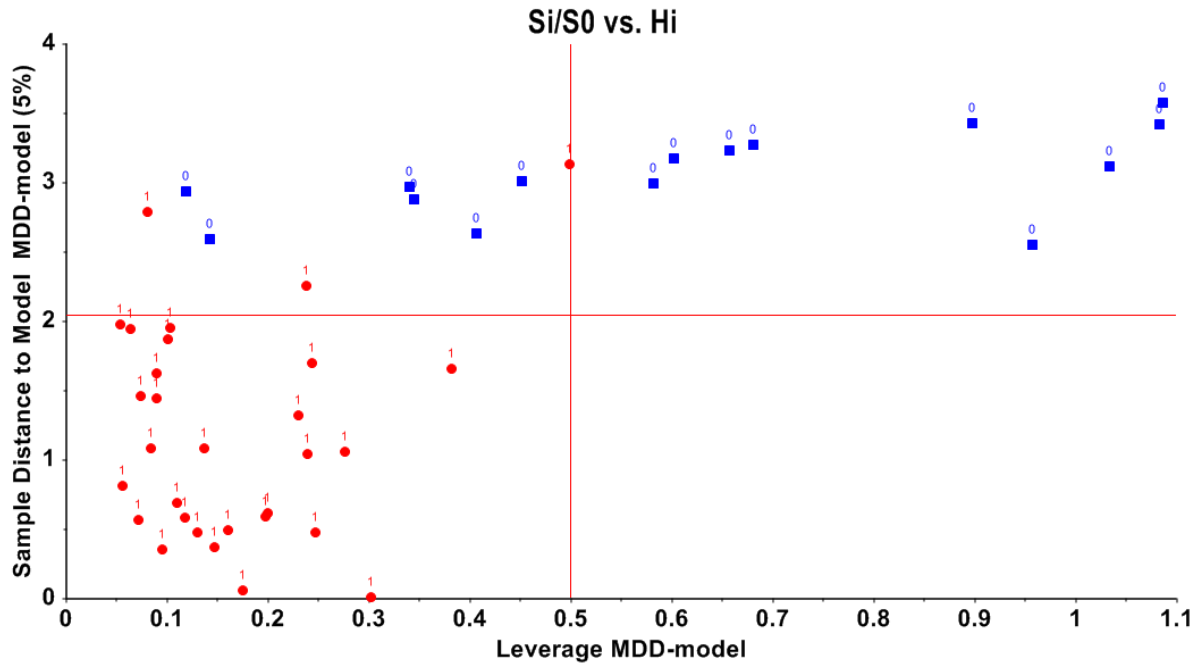
z scores of μ opioid receptors.

Figure 8 Results of SIMCA showing the Si/S0 vs Hi plot.

Test subjects (validation set) are projected into the model which was computed based on subjects

in the training set.

Table 1. Supervised learning methods used on our data for differentiation, discrimination, classification, validation, identification and authentication of subjects with major depression and controls.

Supervised learning and classification	Differentiation; Group mean differences	Discrimination SVM and LDA	Discrimination with validation of SVM and LDA	Identification and authentication: SIMCA
Predefined classes	Cases vs controls	Cases vs controls	Cases vs controls in training and test sets, bootstrapping or cross-validation	PCA models of cases vs controls in training and test sets using critical limits of the classes
Main aims	Differences between groups	Membership allocation	Validated models to allocate subjects to the class models and predict membership of unknown subjects	Validated multivariate models to: * authenticate cases * identify new subjects as members, outsiders, aliens or strangers
Models (decision rules)	GLM (t-test, ANOVA)	SVM: support vectors LDA: discriminant scores	Projection of validation sets to SVM and LDA models	Projection of validation sets to SIMCA PCA models at a predefined α level
Figures of merit	Effect size ROC curve with diagnostic performance	Accuracy of models Diagnostic performance of training sets	Accuracy of models after validation Diagnostic performance of a validation set	Model similarities Model-to-model distance Discrimination power Modeling power Classification accuracy

Membership prediction of new subjects	Using bootstrapped ROC curves and based on a cut-off value obtained by ROC curves; subjects are classified as cases or controls	No validation	Using validated LDA and SVM models; new subjects are classified as cases or controls	Using validated SIMCA models, new MDD patients are authenticated and new subjects are identified as members, aliens, strangers or outsiders
Learning from the data and knowledge acquisition from the models	Quantitative differences in some biomarkers	Features separating cases and controls, using loadings on LDA scores and weights in SVM (but: difficult to interpret)	Validated features (difficult to interpret)	Features of cases and controls Modeling power of features Qualitative differences between the models Ranking of most important features Non relevant features
Overall strength (our subjective 10 point score)	1	2	7	10

40

Table 2. Our proposed techniques that can be used to classify, identify and authenticate unknown subjects into predefined classes. The table also shows our proposed (subjective) accuracy limits for the figures of merit

	Methods to be used	Validation techniques	Proposed Limits for figures of merit
1	ROC curve	Bootstrapped area ROC curve	> 0.95
2	Joint PCA on all subjects	Visualization of all subjects and their classes in 2D/3D spaces	Clear separation between both classes
3	SVM	10-fold cross-validation	Accuracy > 95%
4	LDA	Training and test set	Accuracy > 95%
5	SIMCA	Training and test set	Large model-to-model distances (e.g. >20 or >50) Should authenticate >90% of target class members ($\alpha=0.05$) with <10.0% aliens