Molecular Neurobiology (2019) 56:7822–7835
https://doi.org/10.1007/s12035-019-01647-0

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]

Received: 28 March 2019 / Accepted: 10 May 2019 / Published online: 23 May 2019

Abstract

Major depressive disorder (MDD) is characterized by signaling aberrations in interleukin (IL)-6, IL-10, beta-endorphins as well as μ (MOR) and κ (KOR) opioid receptors. Here we examined whether these biomarkers may aid in the classification of unknown subjects into the target class MDD. The aforementioned biomarkers were assayed in 60 first-episode, drug-naïve depressed patients and 30 controls. We used 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. PCA shows that both groups were well separated using the first three PCs, while correlation loadings show that all five 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 = 110.2); all biomarkers showed a significant discrimination and modeling power, while 100% of the patients were authenticated as MDD cases with a specificity of 93.3%. 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 to (a) authenticate patients as belonging to the claimed target class and identify other subjects as outsiders, members of another class, or aliens; and (b) acquire knowledge through learning from the data by constructing a biomarker fingerprint of the target class.

Keywords Supervised learning · Major depression · Cytokines · Inflammation · Neuro-immune · Opioids

✉ Michael Maes
dr.michaelmaes@hotmail.com

Hussein Kadhem Al-Hakeim
headm2010@yahoo.com

Suhaer Zeki Al-Fadhel
suhaeralfadel@yahoo.com

Arafat Hussein Al-Dujaili
arafat.aldujaili@uokufa.edu.iq

Andre Carvalho
andre.carvalho@camh.ca

[1] Department of Chemistry, College of Science, University of Kufa, Kufa, Iraq

[2] Department of Clinical Laboratory Sciences, College of Pharmacy, University of Kufa, Kufa, Iraq

[3] Faculty of Medicine, University of Kufa, Kufa, Iraq

[4] Department of Psychiatry, University of Toronto, Toronto, ON, Canada

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

[6] Research Affairs, Chulalongkorn University, Bangkok, Thailand

[7] Department of Psychiatry, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

[8] Department of Psychiatry, Medical University Plovdiv, Plovdiv, Bulgaria

[9] IMPACT Strategic Research Center, Barwon Health, Deakin University, Geelong, VIC, Australia

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 rate 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, while no validated biomarkers have been incorporated as part of the diagnostic criteria for MDD [8–10].

Immune-inflammatory mediators play a relevant ethiopathological role in MDD and may affect neuroprogressive pathways [10]. Endogenous opioids and their cognate receptors in the central nervous system are important regulators involved in the neurobiology of mood and they are involved in the pathophysiology of MDD [11, 12]. Recently, we reported highly significant aberrations in the opioid system in individuals with MDD relative to controls including elevated serum β-endorphin, κ-opioid receptor (KOR), and 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].

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 $p = 0.05$) 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, specificity, positive predictive value as well as the area under the receiver operating curve (ROC), which summarizes an assay's overall discriminatory performance. 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, with implications for psychiatric nosology [17].

Moreover, a combination or panel of different biomarkers 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) are often used to build prediction models [18, 19]. These 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 modeling procedure applied to a training or calibration set is used in order to construct a new biomarker model, which is subsequently validated on unknown subjects who were not previously considered during model building. Modeling procedures performed without this validation step are likely to result in an overfitted or overtrained model. Methods employed to validate the models and to compute figures of merit comprize 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 splitting the study group into two sets, namely the training (calibration) and test (validation) sets. Cross-validation may be used to optimize the internal parameters of SVM and to assess its accuracy, while a LDA model should be validated using a train-test splitting method. Figure 1 shows the procedure comprising a training (calibration) and a validation (test) stage.

Nevertheless, there are a number of methodological challenges and limitations with those machine learning techniques. 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]. SVM employs decision rules that are challenging to interpret [19], and the technique is prone to overfitting [31]. As a 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 modeling 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]. By comparing the subjects' distances to the model subspaces at a predefined

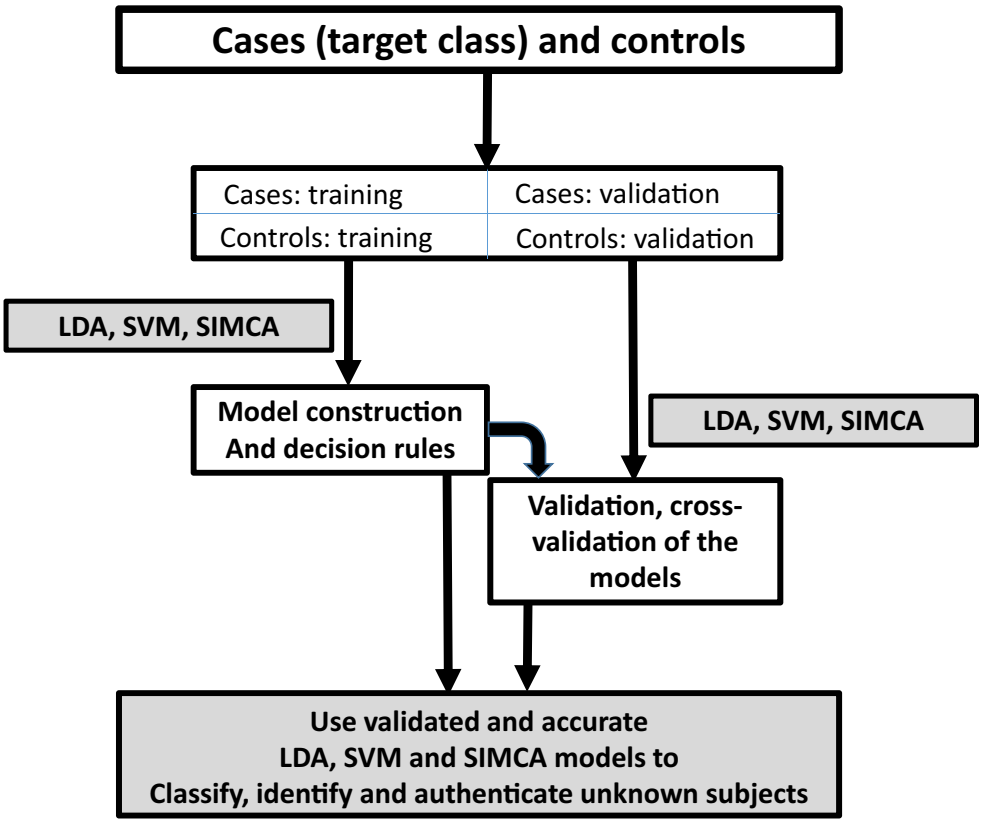

Fig. 1 Procedure to build a model (decision rule) to classify unknown subjects. The procedure entails two stages, namely a training (calibration) stage and a test (validation) stage. In the training stage, supervised pattern recognition methods are applied to biomarker data to construct models that best 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

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), an alternative class (healthy controls), both classes (hybrids), or none of the classes (outsiders) [25]. The advantages of SIMCA are that this method (a) does not force a subject to belong to a specific class, (b) allows to verify authenticity (check whether a subject with MDD actually belongs to the claimed MDD class), and (c) identify new (unknown) subjects as belonging to the modeled target class [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. Of the many immune biomarkers we [8] and others [7–9] assayed in major depression, this study selected to examine two cytokines with a good diagnostic performance for MDD [13]. 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) and 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 and 30 healthy males aged 14–71 years. Only males were recruited to reduce sex-related variability in biomarker data. 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) 1 or 2 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). The body mass index (BMI) was computed as weight (in kg) divided by length (in meter) squared.

Methods

Five milliliters of venous blood samples was drawn, utilizing disposable needle and plastic syringes, from patients and controls. The samples were transferred into a clean plain tube. Hemolyzed 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 Eppendorf tubes to be stored at − 80 °C until analyzed. Serum CRP was measured using a kit supplied by Spinreact®, Spain. Biomarkers were assayed [35] using Commercial ELISA sandwich kits, namely for KOR and MOR (MyBioSource, Inc. CA, USA) and β-endorphin, IL-6, and IL-10 (CUSABIO Co., China). We followed exactly according to the manufacturer's instructions. All intra-assay coefficients of variation (CV) were < 7.0%. Biomarkers including KOR, MOR, and β-endorphin were adjusted for the putative effects of smoking using regression analyses. In addition, we computed a z unit weighted composite score summing up the z scores of both KOR and MOR values (indexed as KOR-MOR).

Statistics

Differences in the biomarkers among depression and controls are displayed as mean (SE) values computed on the z scores of the biomarkers. We employed multivariate general linear model (GLM) analyses to examine the effects of diagnosis (MDD versus controls) on the biomarkers (while adjusting for age 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 and Correlation Loadings

We used PCA as an unsupervised learning method and performed a single, joint PCA on the five biomarkers in all subjects combined (MDD and controls) 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, we used different 2D and 3D dimension combinations (PC1 and PC2, PC1 and PC2 and PC3, PC1 and PC3, PC2 and PC3, etc.), while the two groups were differentiated by marker colors and shapes. In the same 2D PC plots, we display Hotelling's T^2 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 T^2 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 two ellipses, one indicating 100% explained variance (outer ellipse) and another indicating 50% of the explained variance (the inner ellipse).

Support Vector Machine

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 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 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 (color 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 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]. In the training sets (containing both MDD and controls), the number of PCs used to build the models 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. Consequently, SIMCA is applied to the class models whereby subjects of a test set or unknown subjects are projected to the class models for which they display the best similarity. As such, SIMCA assigns new subjects (identification) to the 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 center, 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).

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 less than 3 indicates that there are no significant differences between both classes [18, 23, 25, 33]. Large model-to-model distances (e.g., > 20 and > 40) 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 subjects as members of a class, hybrids (belonging to two classes), or outsiders, either strangers (not belonging to any class) or aliens (healthy controls that intrude into the MDD target class).

We used a training (50% of the MDD and 50% of the control subjects) set to estimate predictive models of MDD and control classes and a test set (the remaining 50% of the MDD and controls samples) to validate the models. Gross outliers are eliminated, such as 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. Moreover, influential outliers and subjects with more extreme F-residual and Hotelling's T^2 values are eliminated if they show biomarker values that are not relevant to the model. Toward this end, we inspect influence and stability plots, sample residual vs samples and Hotelling's T^2 vs samples plots, and through inspection of residual values and leverages are eliminated. In this study, we used two SIMCA plots: (a) the Si/S0 vs Hi plot which shows the residual standard deviation (relative distance of the subjects to the class model) versus Hi scores for a given class as well as the class limits at $\alpha = 0.05$ or 0.01. The class membership limits allow to classify subjects as outsiders, strangers, or aliens. (b) The discriminatory power plot which displays the discriminatory power of the biomarkers to discriminate the class models.

Results

GLM Analysis

Figure 2 shows the mean (SE) values of the z scores of the five biomarkers in both MDD patients and controls. Multivariate GLM analysis (adjusted for age and BMI) shows that there is a significant effect of diagnosis ($F = 64.13$, $df = 5/81$, $p < 0.001$), that all five biomarkers are significantly increased in MDD as compared with controls (all at $p < 0.001$) and that age ($F = 1.44$, $df = 5/81$, $p = 0.219$) and BMI ($F = 1.28$, $df = 5/81$, $p = 0.281$) are not significant. There were no significant differences in age between controls (mean ± SD = 30.3 ± 8.8 years) and MDD patients (32.3 ± 11.0 years) ($F = 0.69$, $df = 1/88$, $p = 0.410$). There were no significant differences in BMI between controls (mean ± SD = 26.2 ± 2.8 kg/m^2) and MDD patients (24.7 ± 3.7 kg/m^2) ($F = 3.81$, $df = 1/88$, $p = 0.054$).

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 75% of the variance, while PC3 explains 14% of

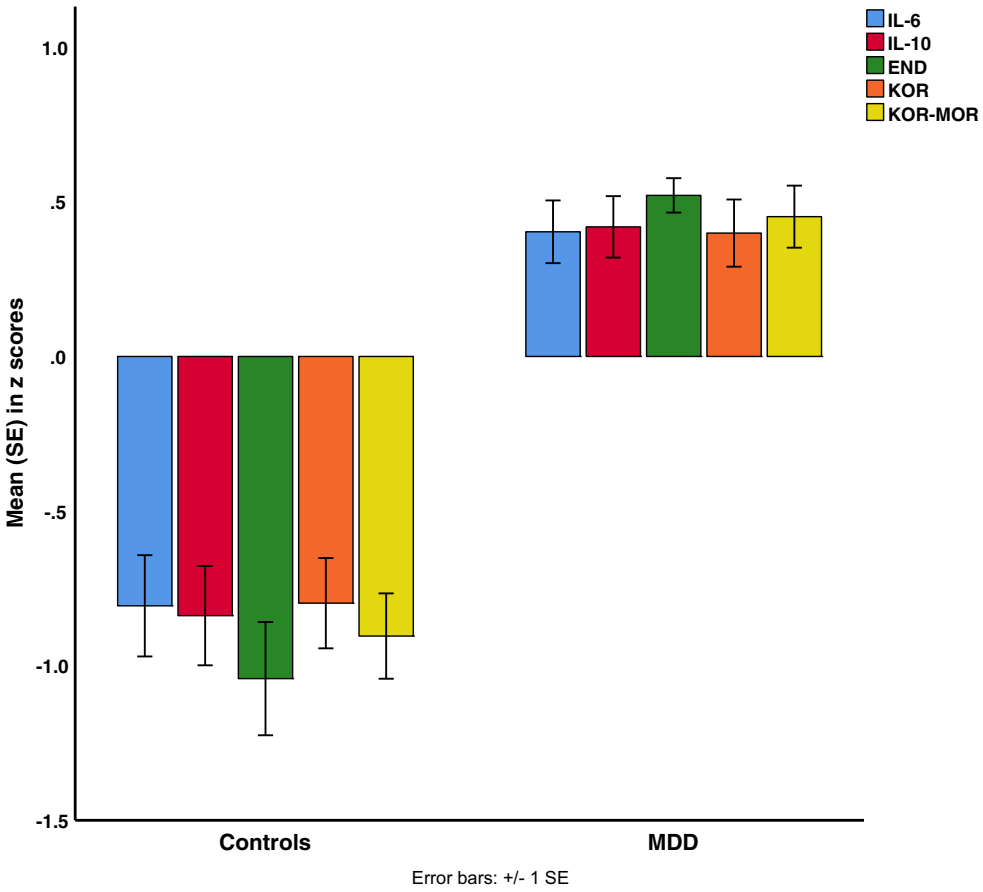

Fig. 2 Mean (SE) values of the z scores of the five biomarkers in both MDD (major depressed) patients and controls. Mean values of all biomarkers are significantly higher in MDD than in controls

the variance. Thus, the three first PCs explain a large part of the variance (89%) 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 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 4 and 5 show the correlation loadings of the five biomarkers on PC1 vs PC2, and PC1 vs PC3, showing that all five biomarkers contribute to the differentiation of both classes.

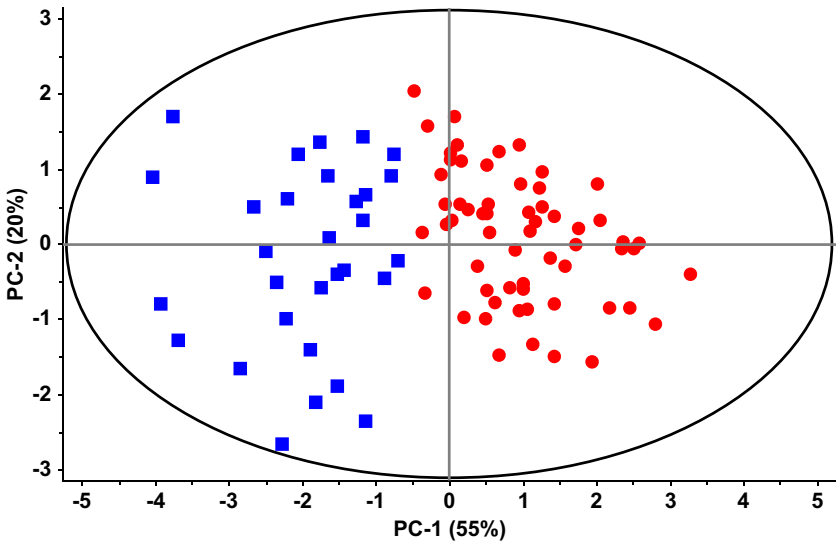

Fig. 3 Principal component (PC) score plot. This plot shows a 2D display of PC1 vs PC2 visualizing the subject distribution in a 2D space, indicating that the five biomarkers allow a clear differentiation of major depressed patients (red circles) vs controls (blue squares)

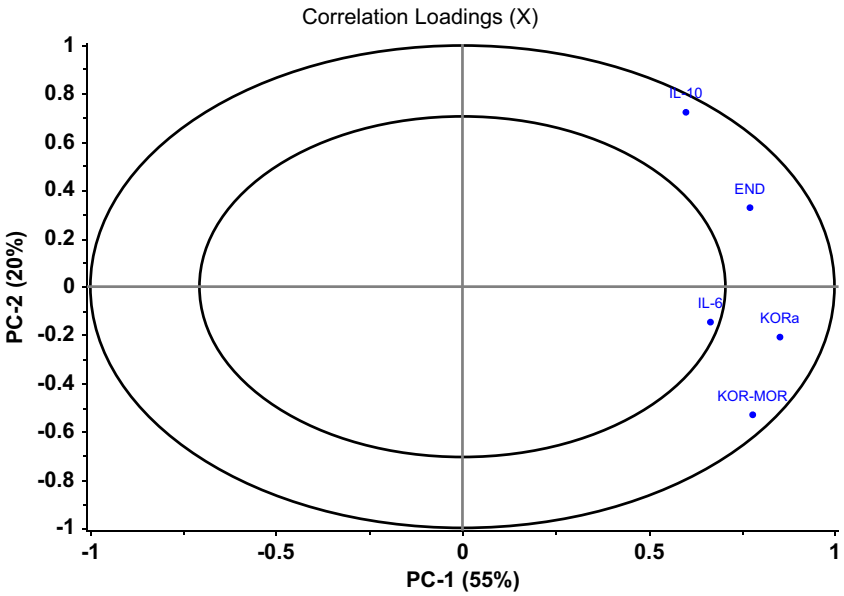

Fig. 4 Correlation loadings of the five biomarkers on principal component (PC)1 vs PC2. This figure shows that four variables are located between both ellipses, namely interleukin (IL-10), β-endorphin, κ opioid receptor (KOR), and KOR-MOR (μ-opioid receptors), and that the three opioid biomarkers are important is separating both classes along PC1, while IL-10 adds to the differentiation via its loading on PC1 but also PC2

SVM

SVM with linear kernel delineated eight support vectors (including four controls and four 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 6 shows a plot of the classification results with IL-10 and KOR-MOR, 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 (three controls and four MDD patients). Projecting the validation set on the SVM model showed an accuracy of 100%. Again, the two classes are best separated when KOR-MOR 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 97.8% of all subjects were correctly classified either as controls or MDD patients. Figure 7 shows the LDA discrimination plot for the all subjects. Both classes are well separated and are located relatively close to zero on the corresponding axes.

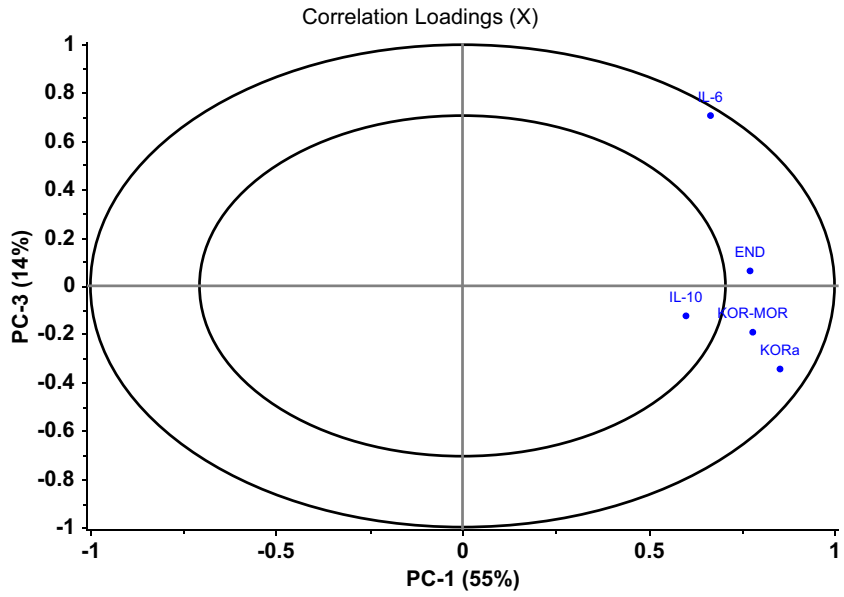

Fig. 5 Correlation loadings of the five biomarkers on principal component (PC)1 vs PC3. This figure shows that IL-6 loads highly on PC3 and contributes to the separation of both classes through its loadings on PC1 but also PC3. Those five biomarkers are close together in the plot suggesting that they show significant and positive intercorrelations

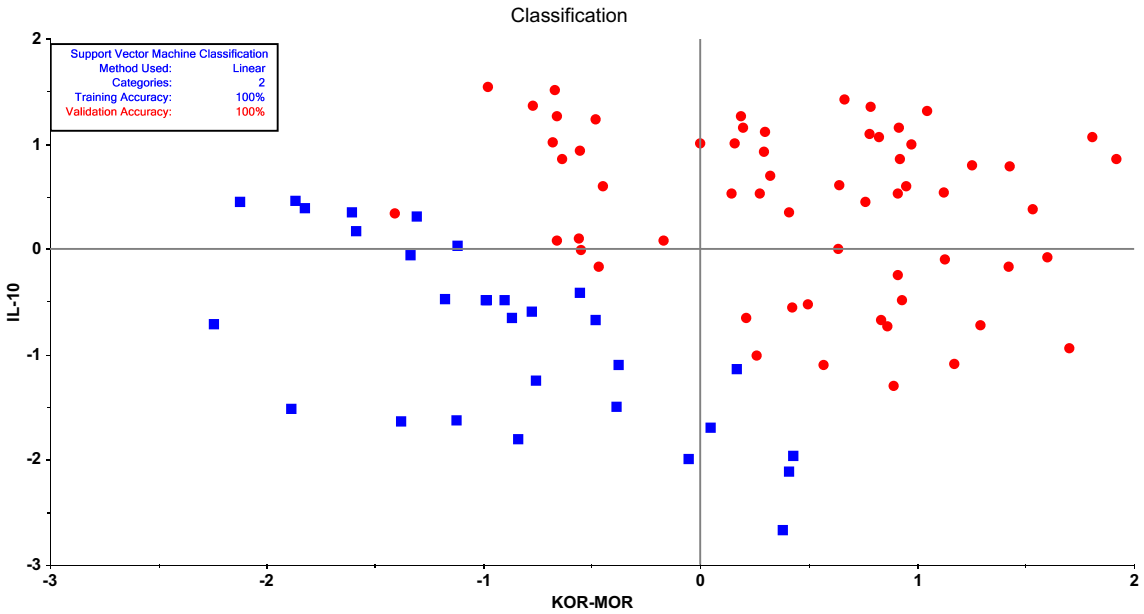

Fig. 6 Plot of the classification results obtained by support vector machine. IL-10, interleukin-10; KOR-MOR, z unit weighted composite score computed as sum of z score of κ-opioid receptor (KOR) levels and z score of μ-opioid receptor (MOR) levels

SIMCA

First we construct a SIMCA model on a calibration set and used the remaining subjects as validation set. During the training step, we omitted two healthy controls (statistical outliers) from the model but no MDD patients. MDD and controls were each modeled using four PCs. We found that the model-to-model distance was 110.2 indicating a strong separation of both classes. Figure 8 shows the discrimination plot with discriminatory power of all biomarkers in separating MDD from controls with in descending order of discriminatory power: IL-10, KOR, KOR-MOR, β-endorphin, and IL-6. The modeling power of all biomarkers was highly significant in controls (all > 0.95) and MDD patients (all > 0.71). Consequently, no features were excluded to compute the final model. Figure 9 shows the Si/S0 vs Hi plot with the subjects distances to the MDD model. All MDD patients were authenticated as belonging to the target MDD class, while there was one alien (a control) intruding in the MDD critical class limits. Figure 10 shows none of the MDD patients intruded in the control

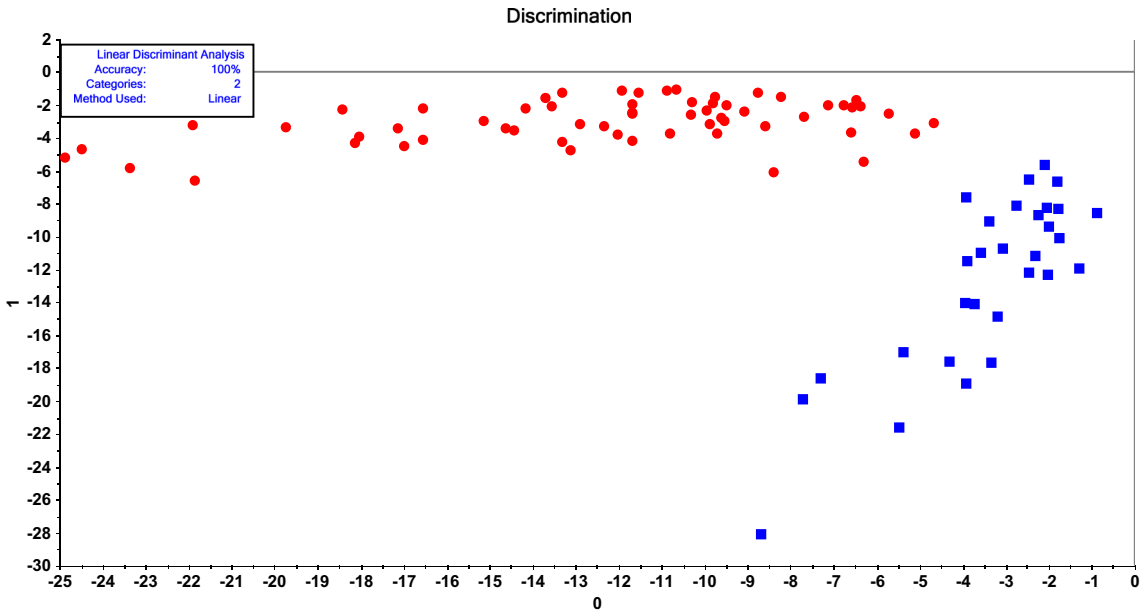

Fig. 7 Linear discriminant analysis plot. The decision rule is computed on all subjects and shows a 100% accuracy rate

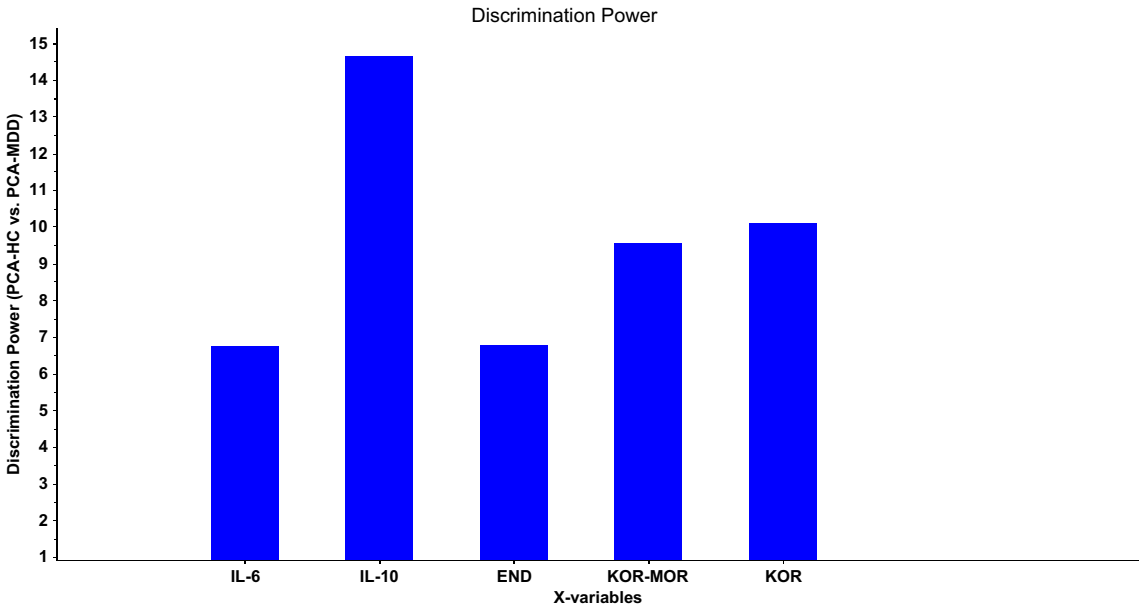

Fig. 8 Results of SIMCA showing the discrimination plot. This plot shows the discrimination power of the five biomarkers used in the current study. IL, interleukin; KOR, κ-opioid receptor; KOR-MOR, z unit weighted composite score computed as sum of z scores of κ-opioid receptor (KOR) levels and z scores of μ-opioid receptors (MOR)

hyperspace, while all controls were correctly authenticated. As such, the sensitivity of the model in authenticating MDD subjects is 100% with a specificity of 93.3%.

Discussion

The first major finding of this study is that, using different supervised learning techniques, patients with MDD are highly significantly separated from healthy controls using neuro-immune and opioid biomarkers as input variables. These data show that aberrations in cytokines with pro-inflammatory and immune-regulatory activities and opioid biomarkers 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 a very high accuracy and, therefore, these models may be employed to classify unknown subjects into MDD or control classes. While LDA

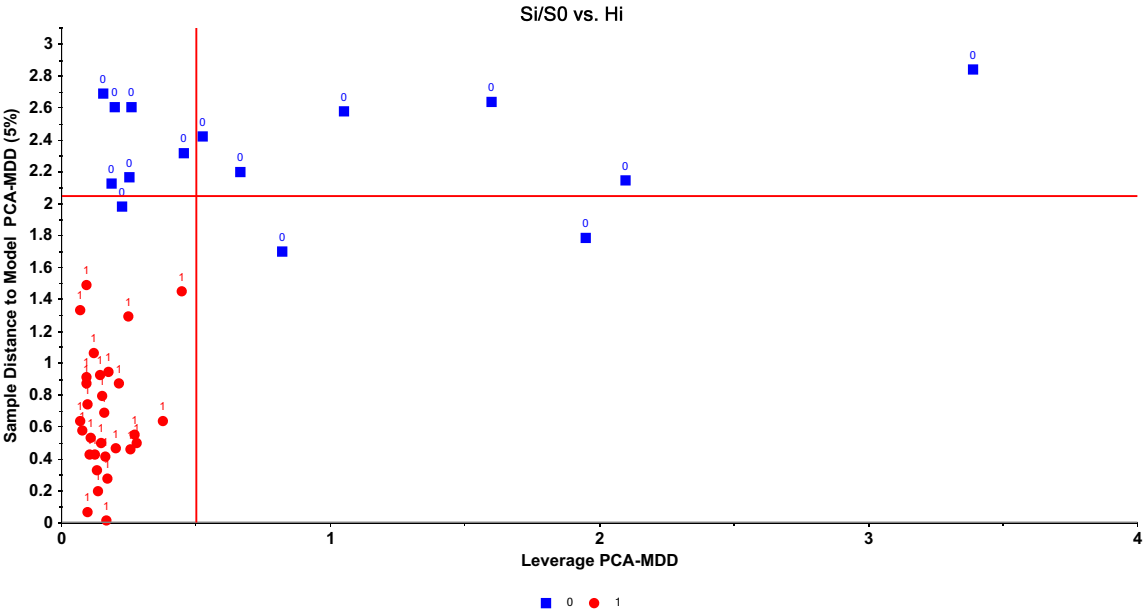

Fig. 9 Results of SIMCA showing the Si/S0 vs Hi plot with the subject's distances to the depression (MDD) model built by principal component analysis (PCA). Test subjects (validation set) are projected into the model which was computed based on subjects in the training set

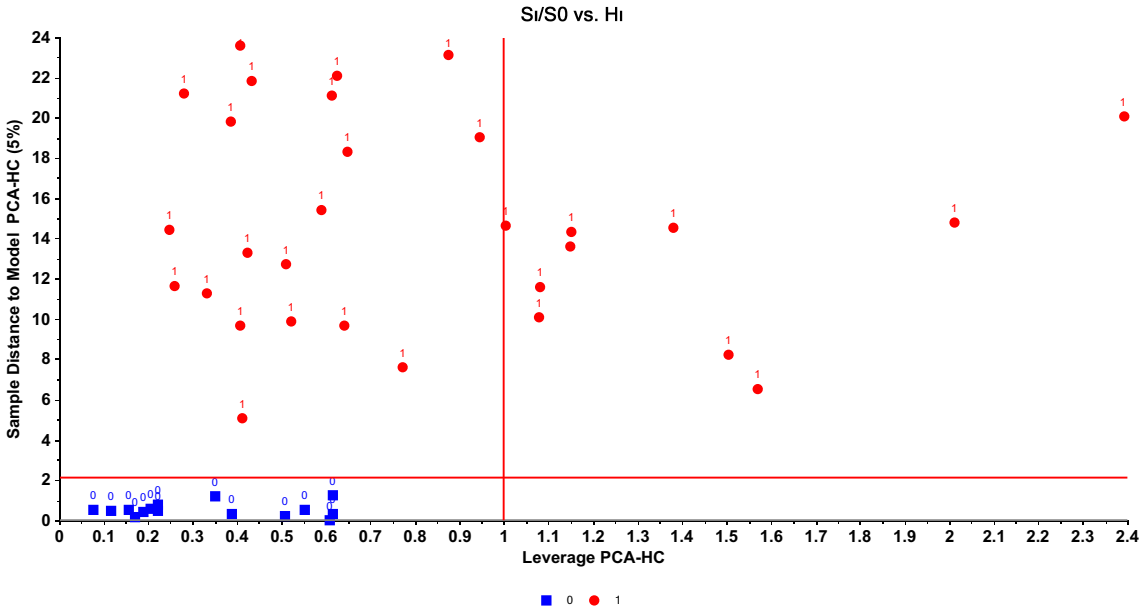

Fig. 10 Results of SIMCA showing the Si/S0 vs Hi plot with the subject's distances to the normal control (HC) model built by principal component analysis (PCA). Test subjects (validation set) are projected into the model which was computed based on subjects in the training set

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 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 SVM is more accurate than SIMCA as the former method correctly classified all subjects. Some authors concluded that SIMCA is not as adequate as other supervised techniques as the accuracy is often less accurate [40, 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 all MDD patients in the test set could be authenticated as real members of the MDD class, while one alien was detected (one control intruding the MDD hyperspace). As such, our SIMCA model is in theory (see below) 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. 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

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 cutoff 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

indicates that those models occupy distinct spaces in the multivariate hyperspace and, thus, that they are qualitatively distinct [18]. Another relevant question is whether a critical limit can be proposed to ensure a good accuracy of the SIMCA model 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 higher values, e.g., > 20 or 40, although these larger distances are rarely obtained in neuro-psychiatric research using current biomarkers, which most often tend to overlap between cases and controls. 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 five biomarkers were useful to model the MDD and control classes, and that IL-10, KOR-MOR, and KOR were top 3 features discriminating both classes. Nevertheless, inspection of the Si/S0 vs Hi plot shows that one control intruded the MDD class critical limits. One method to further improve the accuracy of our SIMCA model is to add more predictors to the decision rule. Such variables should not only comprise peripheral blood 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 any classifiers in general have a number of limitations. Firstly, we examined only drug-naïve patients and used many exclusion criteria and therefore our biomarker set may not have full generalizability because MDD patients often exhibit many comorbidities, including 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 will 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 SIMCA model are restricted (a) to its use in drug-free MDD patients in a first episode of illness, (b) to acquire knowledge to decipher the pathophysiology of depression, and (c) to delineate novel pathway phenotypes.

Based on the above and our previous machine learning publications, we would recommend using four different approaches to construct new models aimed at predicting memberships of unknown subjects. Table 2 lists the different methods that may be used toward 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, separation of the classes using machine learning will probably be insufficient to pass the validation stage, especially using SIMCA. Thirdly, SVM should be applied and the cross-validated model should have an accuracy >

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 distance (e.g., > 20) Authenticates > 90% of target class members ($\alpha = 0.05$) with < 10.0% aliens

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.

Importantly, in this study we proposed an evidence-based approach [45, 46], which was (cross) validated by statistical validation techniques. 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, 18, 19, 24, 43]. Moreover, 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 and that the use of machine learning approaches, and especially SIMCA, is a pre-requisite for this endeavor. As such, the approach of the American Psychiatric Association in the DSM, which lacks statistical and biological validation, and the concept of trans-diagnostic phenotypes of the Research Domain Criteria (RDoC) [47, 48] miss the point that classifications and the development of phenomenological dimensions should be based on results of (un)supervised pattern recognition methods, which should include (cross-)validation techniques. 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. 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 [49].

Author's Contributions All the contributing authors have participated in preparation of the manuscript.

Compliance with Ethical Standards

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

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 Neuro-Psychopharmacol 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 Neuro-Psychopharmacol 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 et al (2017) Peripheral cytokine and chemokine alterations in depression: a meta-analysis of 82 studies. Acta Psychiatr Scand 135(5):373–387
6. Maes M, Carvalho AF (2018) The compensatory immune-regulatory reflex system (CIRS) in depression and bipolar disorder. Mol Neurobiol 55(12):8885–8903
7. Köhler CA, Freitas TH, Stubbs B, Maes M, Solmi M, Veronese N, de Andrade NQ, Morris G et al (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):243–249
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 Neuro-Psychopharmacol 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 et al (2011) Pathways underlying neuroprogression in bipolar disorder: focus on inflammation, oxidative stress and neurotrophic factors. Neurosci Biobehav Rev 35(3):804–817
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 et al (2005) The HCL-32: towards a self-assessment tool for hypomanic symptoms in outpatients. J Affect Disord 88(2):217–233
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):122–128
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 neuroimmune 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 (1976) Modern factor analysis, 3rd edn revised. University of Chicago Press, Chicago
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 Off Anal 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):183–196
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):754–760
25. CAMO (2019) The Unscrambler appendices: method references. 2019. Accessed 19-3-2019. www.camo.com/helpdocs/The_Unscrambler_Method_References.pdf
26. Ripley BD (2008) Pattern recognition and neural networks. Cambridge University Press, Cambridge Paperback. ISBN-13: 978-0521717700.
27. Brownlee J (2016) Supervised and unsupervised machine learning algorithms; understand machine learning algorithms. As assessed https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/. Accessed online 19 Mar 2019.
28. Green SB, Salkind NJ, Akey TM (2008) Using SPSS for Windows and Macintosh: analyzing and understanding data. Prentice Hall, New Jersey
29. Çokluk-Bökeoğlu Ö, Büyüköztürk Ş (2008) Discriminant function analysis: concept and application. Eurasian J Educ Res (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, pp 2–4
31. Cawley GC, Talbot NCL (2010) On over-fitting in model selection and subsequent selection bias in performance evaluation. J Mach Learn Res 11:2079–2107
32. Alamprese C, Casiraghi E (2015) Application of FT-NIR and FT-IR spectroscopy to fish fillet authentication. Lebensm Wiss Technol 63(1):720–725
33. Derde M-P, Massart DL (1982) Extraction of information from large data sets by pattern recognition. Fresmius'Z Anal Chem 313:484–495
34. American Psychiatric Association (2000) Diagnostic and statistical manual of mental disorders, vol 4. Author, Washington, DC text revision
35. Al-Hakeim HK, Al-Fadhel SZ, Al-Dujaili AH, Maes M (2019) In major depression, increased serum dynorphin and kappa opioid receptor levels are positively associated with mu opioid receptor levels and immune activation and are attenuated by nicotine dependence. Preprints 2019, 2019040176. https://doi.org/10.20944/preprints201904.0176.v1
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 et al (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 (1999) Decision criteria for soft independent modelling of class analogy applied to near infrared data. Chemom Intell Lab Syst 47(1):65–77
41. Racz A, Gere A, Bajusz D, Heberger K (2019) Is soft independent modeling of class analogies a reasonable choice for supervised pattern recognition. RSC Advances 8(1):10–21
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, Moraes JB, Congio A, Bonifacio KL, Barbosa DS, Vargas HO, Michelin AP, Carvalho AF et al (2019) Development of a novel staging model for affective disorders using partial least squares bootstrapping: effects of lipid-associated antioxidant defenses and neuro-oxidative stress. Mol Neurobiol. https://doi.org/10.1007/s12035-019-1552-z
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):370–379
45. Stojanov D, Korf J, Jonge P, Popov G (2011) The possibility of evidence-based psychiatry: depression as a case. Clin Epigenetics 2(1):7–15
46. Stojanov D, Machamer PK, Schaffner KF, Rivera-Hernández R (2012) The challenge of psychiatric nosology and diagnosis. J Eval Clin Pract 18(3):704–709
47. Zachar P, Stoyanov DS, Aragona M, Jablensky A (eds) (2014) Alternative perspectives on psychiatric validation: DSM, ICD, RDoC, and beyond. OUP, Oxford
48. Kendler KS, Parnas J (eds) (2015) Philosophical issues in psychiatry: explanation, phenomenology, and nosology. JHU Press, Baltimore
49. Fernandes BS, Williams LM, Steiner J, Leboyer M, Carvalho AF, Berk M (2017) The new field of 'precision psychiatry'. BMC Med 15(1):80