Journal of
Clinical Medicine

MDPI

Article

Using an Interpretable Amino Acid-Based Machine Learning Method to Enhance the Diagnosis of Major Depressive Disorder

Cyrus Su Hui Ho [1,*,†], Trevor Wei Kiat Tan [2,3,4,5,6,†], Howard Cai Hao Khoe [7], Yee Ling Chan [8], Gabrielle Wann Nii Tay [1] and Tong Boon Tang [8]

1 Department of Psychological Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117543, Singapore; gtwn@nus.edu.sg
2 Centre for Sleep and Cognition, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117543, Singapore; trevor.tan@u.nus.edu
3 Centre for Translational MR Research, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117543, Singapore
4 Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583, Singapore
5 N.1 Institute for Health & Institute for Digital Medicine (WisDM), National University of Singapore, Singapore 117456, Singapore
6 Integrative Sciences and Engineering Programme (ISEP), National University of Singapore, Singapore 119077, Singapore
7 Singapore Psychiatry Residency, National Healthcare Group, Singapore 308433, Singapore; howard.khoe@mohh.com.sg
8 Centre for Intelligent Signal and Imaging Research (CISIR), Universiti Teknologi PETRONAS (UTP), Seri Iskandar 32610, Perak, Malaysia; yee_20001812@utp.edu.my (Y.L.C.); tongboon.tang@utp.edu.my (T.B.T.)
* Correspondence: pcmhsh@nus.edu.sg
† These authors contributed equally to the work.

Abstract: Background: Major depressive disorder (MDD) is a leading cause of disability worldwide. At present, however, there are no established biomarkers that have been validated for diagnosing and treating MDD. This study sought to assess the diagnostic and predictive potential of the differences in serum amino acid concentration levels between MDD patients and healthy controls (HCs), integrating them into interpretable machine learning models. **Methods:** In total, 70 MDD patients and 70 HCs matched in age, gender, and ethnicity were recruited for the study. Serum amino acid profiling was conducted by means of chromatography-mass spectrometry. A total of 21 metabolites were analysed, with 17 from a preset amino acid panel and the remaining 4 from a preset kynurenine panel. Logistic regression was applied to differentiate MDD patients from HCs. **Results:** The best-performing model utilised both feature selection and hyperparameter optimisation and yielded a moderate area under the receiver operating curve (AUC) classification value of 0.76 on the testing data. The top five metabolites identified as potential biomarkers for MDD were 3-hydroxy-kynurenine, valine, kynurenine, glutamic acid, and xanthurenic acid. **Conclusions:** Our study highlights the potential of using an interpretable machine learning analysis model based on amino acids to aid and increase the diagnostic accuracy of MDD in clinical practice.

Keywords: major depressive disorder; diagnosis; biomarker; metabolomics; amino acids; machine learning

check for updates

Citation: Ho, C.S.H.; Tan, T.W.K.; Khoe, H.C.H.; Chan, Y.L.; Tay, G.W.N.; Tang, T.B. Using an Interpretable Amino Acid-Based Machine Learning Method to Enhance the Diagnosis of Major Depressive Disorder. *J. Clin. Med.* **2024**, *13*, 1222. https://doi.org/10.3390/jcm13051222

Academic Editor: Domenico De Berardis

Received: 8 January 2024
Revised: 13 February 2024
Accepted: 19 February 2024
Published: 21 February 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

According to the World Health Organization [1], major depressive disorder (MDD) is a leading cause of disability worldwide, affecting over 280 million people of all ages and nationalities at present. MDD causes 50 million person-years of disability yearly, and depressive disorders are the single largest contributor to non-fatal health loss. With the

significant socioeconomic and healthcare burden that MDD imposes, prompt assessment and treatment are essential. Currently, practitioners in the field commonly diagnose MDD in patients by assessing patients' self-reported symptoms against the designated key symptoms outlined in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). Individuals are diagnosed with depression if they report experiencing a depressed mood or loss of interest in activities for at least two weeks in duration, as well as four or more of the other eight symptoms (including weight and sleep disturbances, fatigue, worthlessness, poor concentration, psychomotor disturbances, and suicidal ideation). Practitioners also rely on their professional clinical expertise and past experiences to diagnose, which can be challenging, especially when patients are not forthcoming about their symptoms. These subjective diagnostic methods, coupled with the complex nature of MDD given its wide range of presenting symptoms, have led to the underdiagnosis and inadequate management of the disease.

To date, no definitive biomarkers have been identified for the diagnosis, prognosis, and therapeutic management of MDD [2]. Despite the existence of large-scale genome-wide association studies that aim to identify loci linked to depression, their success in distinguishing and prognosticating the disease remains equivocal [3]. However, advancements in genomics and metabolomics remain at the forefront of psychiatry biomarker discovery, which may, in the future, guide the prognosis and treatment response of patients afflicted with psychiatric disorders [4]. Amino acids, for example, have been steadily gaining traction as potential biomarkers for MDD in recent years [2,5–7]. Depression can be attributed to a lack of monoamine neurotransmitters, including serotonin, dopamine, and norepinephrine, according to the well-established monoamine hypothesis. This was substantiated by the fact that antidepressants that target monoamine depletion increase neurotransmitter levels, thus improving the function of the monoaminergic system [8] and, in turn, improving clinical symptoms. In addition, it has been reported that depressed patients have decreased levels of the amino acid precursors tryptophan and tyrosine, from which serotonin and norepinephrine are synthesised [9].

Baranyi et al. [10] showed that branched-chain amino acids (BCAA) isoleucine, leucine, and valine had decreased levels in depressed subjects, with the concentration of these amino acids inversely associated with the Hamilton Depression Rating Scale (HAM-D) scores. Kynurenine, 3-hydroxykynurenine, and kynurenate concentrations were shown to be negatively associated with suicidal ideation, while citrate and alanine were positively associated with suicidal ideation [11]. Amino acids such as GABA were associated with the severity of depression as well [11]. These findings suggest that amino acids do, indeed, harness a predictive potential in depression symptom manifestation, disease progression and predicting treatment response.

A study by Ding and colleagues found that patients with MDD had lower levels of leucine and higher levels of alanine, serine, and proline in their blood serum compared to healthy controls (HCs) [12]. These findings are corroborated by Hung et al., which show that patients with MDD in complete remission presented significantly lower serum levels of metabolites related to pyruvate metabolism when compared to pre-remission levels, which is linked to the metabolism of amino acids such as proline and alanine, amongst others [13]. Further, in a double-blind ketamine and placebo crossover randomised controlled trial involving the participation of unmedicated MDD patients and HCs, the plasma metabolic profiling conducted by Moaddel et al. found that treatment with intravenous ketamine resulted in a slight increase in kynurenine, a metabolite of tryptophan, in the blood serum of patients around 4 h after treatment [14]. Given that low levels of tryptophan and kynurenine are commonly observed in MDD [15,16], these results reinforce the therapeutic role of ketamine in the treatment of MDD [17] and the potential of amino acids to serve as biomarkers for MDD.

Nevertheless, it is important to recognise that the aetiology of MDD is multifactorial, complex, and poorly understood [18,19]. This contributed to the advent of precision psychiatry, which aims to translate research findings into providing individualised clinical

care to patients with psychiatric disorders [20]. In the field, multiple data types are often integrated with machine learning and artificial intelligence algorithms, which serve to identify complex patterns from observational datasets [21,22] and leverage such learned insight into estimating predictive outcomes for new, unknown data and/or events in the future [20]. Previous studies incorporating machine learning methods have shown that changes in peripheral amino acid levels can differentiate depressed subjects and HCs with high reliability [5]. In one study, levels of GABA, dopamine, kynurenine, and tyramine accurately differentiated depressed subjects from healthy subjects at approximately 96.8% and 95.3% in the training and testing set, respectively, and separated subjects with unipolar and bipolar depression precisely [23]. With the key role played by biomarker usage based on machine learning and artificial intelligence methods in precision psychiatry [24], the first aim of this study was to demonstrate that combining machine learning with serum amino acid levels can be used to achieve a reasonable prediction performance of MDD diagnosis.

Interpretable machine learning models have the potential to generate knowledge via hypothesis generation [25,26]. Complex machine learning models such as neural networks, support vector machines, and ensemble models can learn complex patterns in data for high predictive performance. However, the disadvantage of such complex models is that they are less interpretable than simpler models. Such simple and interpretable machine learning models can quantify feature importance to generate knowledge about which features were most responsible for guiding the predictive decision of the model, thereby generating new hypotheses [27,28]. Therefore, the second aim of this study was to identify which amino acids are most important in diagnosing MDD using an interpretable machine learning model. The hypothesis for this study is that amino acids, when integrated with machine learning, can enhance the diagnostic accuracy of MDD in a manner that can be easily understood and interpreted.

2. Materials and Methods

2.1. Sample Size and Participants

For the sample size, reference was made to a study by Pan and colleagues [29], which compared 50 healthy controls and 50 unmedicated MDD patients and reported the effect size of glutamic acid concentration (Cohen's *d* value of 0.91). For a two-tailed Welch's *t*-test to achieve 80% power and 0.05 probability of type I errors, assuming a normal distribution, this study derived a minimum number of 26 subjects per group to detect differences in amino acid concentrations between groups. In addition to the limitations of study grant funding, this contributed to the final study sample size of 140.

In total, 140 participants were recruited for this cross-sectional study with 70 MDD patients and 70 HCs matched for sex, age, and ethnicity. All participants were aged between 21 and 50 years, were English-speaking, and right-handed. The patients were recruited from the outpatient psychiatry clinics of a university hospital in Singapore. They were diagnosed with MDD by their psychiatrist following the DSM-5 criteria. Patients with other significant psychiatric disorder comorbidities, such as bipolar depression, schizophrenia, and substance use disorder, were excluded from the study. The HCs were recruited from the community by word-of-mouth and matched with MDD patients accordingly. All subjects were excluded if they had a neurological disorder or medical conditions that could affect the central nervous system. Each subject's depressive symptoms and disease severity were assessed using the 17-item version of the Hamilton Rating Scale for Depression (HAM-D 17), with scores of 8 to 16 indicating mild depression, 17 to 23 indicating moderate depression, and 24 or higher indicating severe depression.

All study details were fully explained to the participants, and their written, informed consent was obtained. Recruited subjects provided their sociodemographic data, and patients' clinical information was obtained from them and verified with computer records; all subjects also completed written questionnaires, had their blood drawn during the study visit and had their collected data de-identified. It was difficult to standardise the collection

time and lifestyle characteristics, such as the subjects' nutrition and sleep, as subjects' samples were collected once they enrolled on the study.

2.2. Blood Collection and Metabolite Analysis

Blood sample collection and serum amino acid profiling were conducted using procedures employed in previous studies [27,28]. In total, 21 total metabolites were analysed: 17 from a preset amino acid panel (glycine, alanine, serine, proline, valine, leucine, isoleucine, methionine, histidine, phenylalanine, tyrosine, aspartic acid, glutamic acid, ornithine, citrulline, arginine, and tryptophan) and four from the preset kynurenine panel (kynurenine, kynurenic acid, xanthurenic acid, and 3-Hydroxykynurenine).

2.3. Classification Algorithm

Logistic regression was selected because of its interpretability in understanding which features are the most important for the prediction task [5,30–34], compared to more complex models such as deep neural networks [34,35]. This is a crucial characteristic of the logistic regression model for this research, given that the main objective of this paper is to determine which amino acids are most significant for MDD prediction. Furthermore, logistic regression is a well-established algorithm in the literature that differentiates MDD and healthy control subjects using metabolomic data [12,29,36]. L1-regularized (i.e., LASSO-regularized) logistic regression was implemented using the LogisticRegression function from the Scikit-learn library [37].

2.4. Cross-Validation Framework

Nested cross-validation was used to avoid an over-optimistic prediction performance and to provide an unbiased estimate for a small sample size while simultaneously allowing for hyperparameter optimisation and feature selection [6,7]. The nested cross-validation method comprises two loops: firstly, the outer loop to assess the prediction performance on unseen data in the test set, and secondly, the inner loop to train the model on the training set and to select the best-performing trained model based on the prediction performance on the validation set. To avoid an over-optimistic prediction performance on the test set, hyperparameter optimisation and feature selection are performed only on data within the inner loop. As such, data within the test set will remain unseen by the trained model.

Amino acid data from 140 subjects (70 MDD subjects and 70 HCs) were split into a 9-fold model development set and a 1-fold test set for the outer loop (see Figure 1). The inner loop comprises only data from the 9-fold model development set, split further into a 9-fold training set and a 1-fold validation set. Within the inner loop, minimum-maximum scaling was fit to the training data and subsequently applied to the training and validation data. This normalises feature values to the range [0, 1]. Minimum-maximum scaling was selected to preserve the original distribution shape and thus preserve the information embedded within the data [38,39]. This allows feature selection to be performed later based on the original distribution of the data rather than an altered distribution.

A filter-based feature selection method in the form of a one-way analysis of variance (ANOVA) was performed, whereby the ANOVA F-value was utilised to rank the features. The higher the ANOVA F-value, the more important a feature is deemed to differentiate between the MDD and healthy control groups. Hyperparameter optimisation was implemented. To be clear, the only hyperparameter that was optimised was C, the inverse of L1 regularisation strength for the logistic regression model.

Grid search (via GridSearchCV function from Scikit-learn [37]) was implemented to perform an exhaustive search over both features and C (i.e., the inverse of L1 regularisation strength hyperparameter values). The search range of C spanned 11 values, including [1, 11, 21, 31, 41, 51, 61, 71, 81, 91, 101]. Features were added one at a time, and the prediction performance on the validation set was computed for each combination of features and L1 regularisation value. Once all combinations of features and L1 regularisation values have been computed, this process undergoes 9 more repetitions (10 repetitions in total for

each of the 10 folds within the model development set to take turns being the validation set). The combination of features and L1 regularisation value with the highest prediction performance on the validation set is then selected as the best model. The prediction performance of the best model is then evaluated based on the prediction performance of the test data in the outer loop.

The entire process (comprising minimum-maximum scaling, feature selection, hyperparameter optimisation, and the 10 repetitions to obtain the validation set prediction performance) then undergoes 9 more repetitions (10 repetitions in total for each of the 10 folds in the outer loop to take turns being the test set).

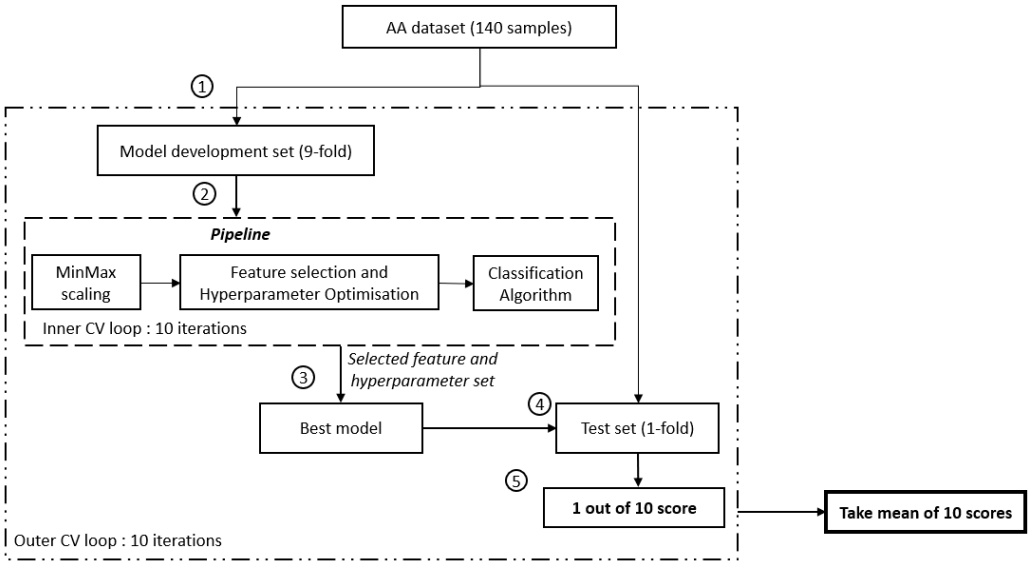

Figure 1. Flowchart of the study. ① Split of the model development and evaluation set, ② model development set will be fed into the pipeline and further search for the best feature set using GridSearchCV. SelectKbest with ANOVA F-value f_classif (min feature = 1, max_feature = 21, step = 1) is used to select the feature, ③ establish the best mode using the selected feature set, ④ evaluate the developed model using an unseen test set, and ⑤ determine the performance score of 1 iteration.

2.5. Performance Metrics

Four performance metrics are presented in this paper: area under the receiver operating characteristic curve (AUC), accuracy, precision, and recall. AUC was used as the main overall performance metric due to its suitability for medical diagnostic problems. Therefore, the best model was selected based on the AUC prediction performance on the validation set during cross-validation.

3. Results

3.1. Sample Characteristics

HCs and patients with MDD did not differ significantly in terms of age, sex, or ethnicity ($p > 0.05$; see Table 1).

They did, however, differ in relation to several factors, including years of education, incidence of familial mental illness, history of traumatic experiences, perceived level of social support, and HAM-D 17 scores. MDD patients were more likely to have a history of mental illness in their families ($p = 0.032$) and reported lower social support ($p < 0.001$) when compared to HCs. As anticipated, individuals with MDD obtained significantly higher scores on the HAM-D relative to HCs ($p < 0.001$), with 70% being moderately to severely depressed.

Table 1. Demographic and clinical characteristics of subjects.

	MDD (n = 70)	HC (n = 70)	p Value *
Age (years)	28.3 (SD 7.2)	28.2 (SD 7.3)	0.926
Sex			1.000
Male	16 (22.9%)	16 (22.9%)	
Female	54 (77.1%)	54 (77.1%)	
Ethnicity			1.000
Chinese	45 (64.3%)	45 (64.3%)	
Malay	15 (21.4%)	15 (21.4%)	
Indian	9 (12.9%)	9 (12.9%)	
Eurasian	1 (1.4%)	1 (1.4%)	
Education (years)	14.5 (SD 1.8)	15.6 (SD 1.2)	**<0.001**
Perceived social support			
Poor	17 (24.3%)	0 (0.0%)	**<0.001**
Average	44 (62.9%)	18 (25.7%)	
Good	9 (12.9%)	52 (74.3%)	
HAM-D 17 score	19.8 (SD 5.4)	1.9 (SD 2.5)	**<0.001**
Mild (8–16)	21 (30.0%)	4 (5.7%)	
Moderate (17–23)	30 (42.9%)	0	
Severe (≥24)	19 (27.1%)	0	
Family psychiatric history	30 (42.9%)	17 (24.3%)	**0.032**
History of trauma	35 (50%)	14 (20.0%)	**<0.001**
Past admission to a psychiatric ward	16 (22.9%)		
Past suicide attempt	32 (45.7%)		
Pharmacotherapy	60 (85.7%)		

* p-values ≤ 0.05 are in bold.

3.2. Logistic Regression Model Classification Performance

The logistic regression model with feature selection and hyperparameter optimisation classified MDD patients and healthy controls with an average performance across test sets in the outer loop with an AUC of 0.76 (±0.16), accuracy of 68.6% (±15.7%), precision of 71.2% (±18.7%), and recall of 65.7% (±21.4%). Additionally, an average of 14.6 features (±1.56) were selected across the outer loop. The outer loop AUC result is visualised in Figure 2.

The performance of the logistic regression model with feature selection and hyperparameter optimisation can be compared against two benchmark versions of the logistic regression model. Namely, a version without feature selection but with hyperparameter optimisation and another version without both feature selection and hyperparameter optimisation in Table 2.

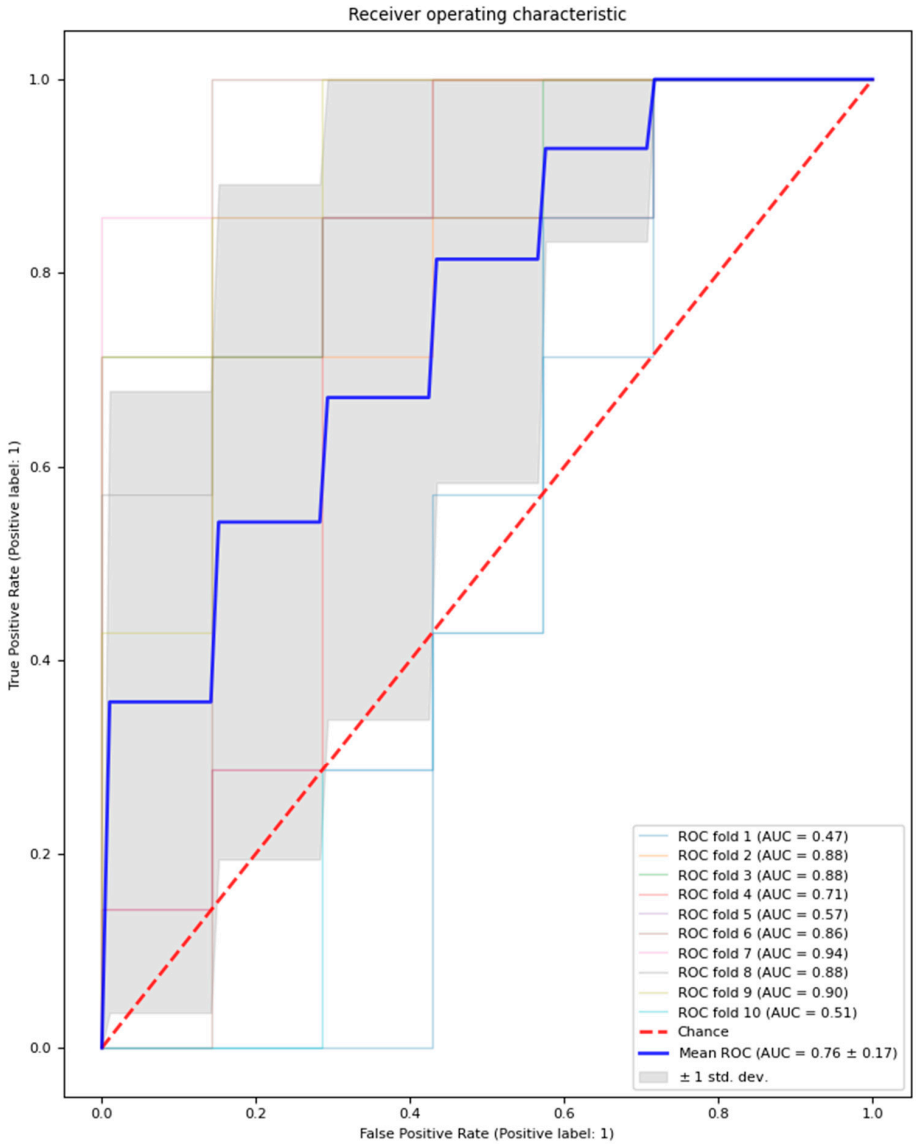

Figure 2. Outer loop AUC result for the model with feature selection and hyperparameter optimisation. Mean AUC results across 10 folds are represented by a bolded blue line; the chance AUC result is represented by a red dotted line.

Table 2. Validation and test set performance results for three variations of logistic regression models: with feature selection, without feature selection, without feature selection and hyperparameter optimisation.

	Validation Set Performance	**Test Set Performance**				
Type of Logistic Regression Model	**AUC**	**AUC**	**Accuracy**	**Precision**	**Recall**	**Number of Features Selected**
With feature selection and with hyperparameter optimisation	0.74 ± 0.03	0.76 ± 0.16	68.6 ± 15.7	71.2 ± 18.7	65.7 ± 21.4	14.6 ± 1.56
No feature selection and with hyperparameter optimisation	0.73 ± 0.03	0.72 ± 0.17	67.9 ± 14.0	70.6 ± 17.3	65.7 ± 19.4	21.0 ± 0.00
No feature selection and no hyperparameter optimisation	0.71 ± 0.04	0.73 ± 0.17	65.0 ± 14.8	66.6 ± 20.1	60.0 ± 20.0	21.0 ± 0.00

Additionally, we evaluated the performance of the logistic regression model with feature selection and hyperparameter optimisation by computing binary classification positive likelihood ratios (LR+) and negative likelihood ratios (LR−). This test was implemented using the class_likelihood_ratios function from Scikit-learn [34] The positive class is the MDD class, and the negative class is the HC class. We note that across the 10 outer folds, there are 2 out of 10 instances where there are 0 false positives. In these two instances, the LR+ was infinitely high, undefined, and returned as a NaN (not a number) value. However, to report a mean non-NaN LR+ value, these two instances with zero false positives are ignored, and the mean is taken across the remaining eight outer folds. The mean LR+ across the remaining 8 non-NaN outer folds was 2.37, and the mean LR− across the 10 outer folds was 0.49.

3.3. Logistic Regression Model-Selected Features

The top five amino acids identified by our best-performing logistic regression model (the model with both feature selection and hyperparameter optimisation) in order of decreasing ranking are as follows: 3-hydroxy-kynurenine, valine, kynurenine, glutamic acid, and xanthurenic acid. The ranking of all the amino acids used in this study can be interpreted from Figure 3, whereby the lower the average ranking, the more important the amino acid.

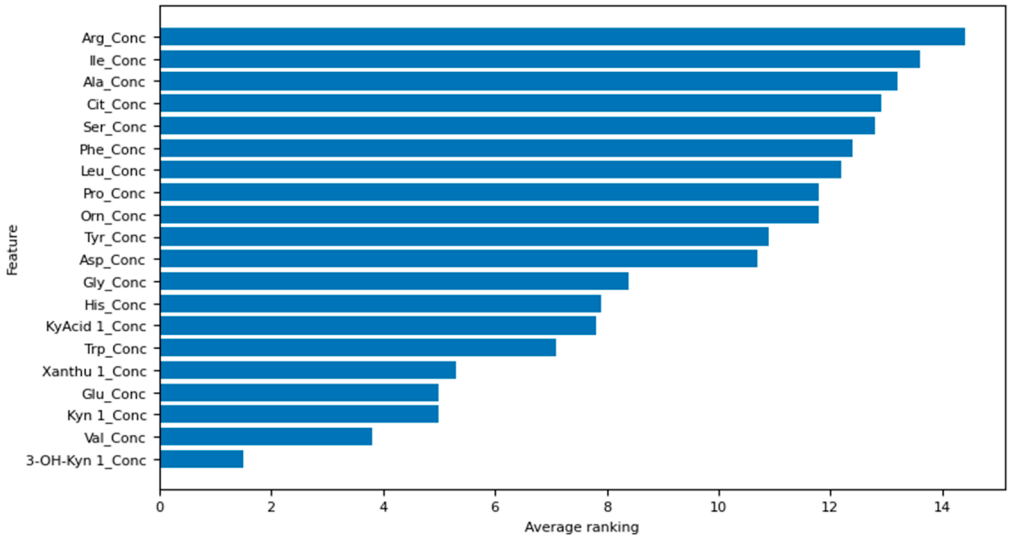

Figure 3. Feature selection results for the model with both feature selection and hyperparameter optimisation.

4. Discussion

In this study, depressed patients were found to have significantly lower levels of education, were more likely to have a family history of mental illness, experienced trauma and reported lower social support than HCs, which concurred with findings from the existing literature. Research has shown that lower education levels are correlated with diminished mental well-being and an increased susceptibility to the development of psychiatric disorders [40]. A positive family history of depression is associated with a personal history of depression, and it is the most significant risk factor for developing a depressive disorder [41]. Traumatic experiences are also associated with an increased risk for MDD [42]. Individuals who perceived less social support were at a greater risk for developing depressive symptoms [43].

While many relevant works implement traditional statistical techniques to differentiate MDD and healthy control subjects [44–48], there are substantially fewer studies on using amino acid data in machine learning for the predictive classification of MDD and HCs. Despite many machine learning studies aimed at distinguishing between individuals with

MDD and HCs, effectively applying machine learning techniques to diagnose MDD in the clinical setting still poses significant challenges. This difficulty could be explained by the fact that conventional statistical methods excel at detecting relationships at the population level. In contrast, machine learning is adept at identifying generalisable patterns in the data, enabling predictions at the individual level [49]. In our study, which deals with data from multiple amino acids, machine learning techniques were generally considered more suitable due to their ability to discern complex relationships between multiple variables [50].

In this study, the logistic regression model with feature selection and hyperparameter optimisation classified MDD and healthy control subjects with an average performance across test sets in the outer loop with an AUC of 0.759. A study conducted by Hung et al. [13], which also aimed to develop a predictive model to discriminate between MDD patients and HCs with as high a performance score as possible, achieved a test set performance AUC of 0.784 using the linear support vector machine (SVM) classifier to discriminate between MDD patients and HCs. This disparity in model performances could be attributed to the utilisation of the logistic regression machine learning model in this study, whose secondary aim was to construct an amino acid feature importance ranking. Hence, it is unsurprising that a simpler logistic regression model would slightly underperform the more complex linear SVM model. Furthermore, it is worth mentioning that this study includes a more comprehensive set of 21 amino acids, while Hung et al.'s study focused only on 8 metabolites.

Another related study by Zheng et al. [23] achieved an outstanding test set performance AUC of 0.96 using the least-squares SVM model. While amino acids were used as part of the input data for the model in Zheng et al.'s study, they did not exclusively focus on amino acids; other metabolites, such as glucose and lipids, were also included. This could partly explain the substantial discrepancy in model performance between Zheng et al.'s model and the mode proposed by our study. Furthermore, Zheng et al.'s study included a small sample size of 126 subjects; however, it had a single split of the training and test sets. At such small sample sizes (like this study's sample size), it is crucial to note that a single random training-test set split may lead to highly over-optimistic results. Therefore, this could also partly explain the difference in model performances between this study and Zheng et al.'s study. Also, like Hung et al.'s study, Zheng et al.'s study did not include any feature importance rankings of metabolites.

The third related study by Liu et al. [51] was able to identify blood metabolite markers which distinguished melancholic depressed patients (n = 90) from HCs (n = 97) with 80% accuracy after employing the ensemble feature selection framework (EFSF). Liu et al. analysed 228 metabolites (of which a union of 76 metabolites from the four feature selection methods was used as input). Of the 76 metabolites, 56 were mapped to the Encyclopedia of Genes and Genomes (KEGG). 11 of the 56 metabolites analysed were amino acids, and 48 of 56 metabolites were increased in melancholic depressed patients versus HCs. While Liu et al. could discriminate between HCs and depressed subjects at a higher rate of 80%, they used a more complex machine learning model known as the ensemble feature selection framework (EFSF), which affected how interpretable the model was.

Regarding this study's amino acid and metabolite ranking, the top five were determined to be 3-hydroxy-kynurenine, valine, kynurenine, glutamic acid, and xanthurenic acid. These compounds are implicated in neurobiological mechanisms that potentially contribute to the development of depression, rendering them clinically significant. In depression, the kynurenine pathway is thought to be activated, and available tryptophan is diverted away from serotonin production and into further degradation, leading to an increase in the production of 3-hydroxykynurenine (3-HK). 3-HK is considered neurotoxic due to its ability to generate reactive oxygen species, alter mitochondrial function, and induce DNA damage [52], which may be associated with depression [53]. 3-HK is then converted to xanthurenic acid by the enzyme kynurenine aminotransferase, and increased xanthurenic acid levels have also been found in depressed patients relative to controls [54]. Glutamic acid may contribute to depression through its action as an N-methyl-D-aspartate

(NMDA) receptor agonist [55], with hypofunctioning NMDA receptors causing cognitive deficits and excessive activation causing excitotoxicity and neurodegeneration [56]. The blood serum concentration of valine has been found to increase in MDD patients and has been shown to be highly correlated with the other BCAAs associated with depression [10,57,58]. It has also been suggested that increased blood serum valine concentration can produce depressive symptoms by decreasing brain serotonin (5-HT) function; it inhibits the transport of 5-HT precursor, L-tryptophan, across the blood-brain barrier [59].

Our study contributes to the limited body of literature on the feasibility of using amino acids as biomarkers for MDD. Given the observed changes in amino acid concentrations and their correlations with MDD, our results have shown that the logistic regression model incorporating feature selection and hyperparameter optimisation can differentiate between MDD patients and HCs with reasonable accuracy. This highlights another main strength of our study, which utilises machine learning to identify specific amino acids that are of significance in the context of MDD.

However, it would be important to acknowledge that the clinical implications of our findings are limited by our study's relatively small sample size. Furthermore, since our study population consisted mostly of Asian individuals with a mean age of approximately 27 years and most of the patients were receiving prescribed medication, it is unlikely that our findings can be extrapolated to the wider population. Further studies should focus on recruiting a larger sample with different age groups who are unmedicated. In addition, not all MDD patients in our study are medication naive, so our findings could be attributed to treatment effects. Pertinently, combining multimodal data for classification can overcome the limitations of using only a single data source. Therefore, future investigations integrating clinical, omics (including genetic, transcriptomic and proteomic), and neuroimaging data would be crucial for enhancing the diagnostic utility of amino acids as biomarkers of MDD in clinical practice. An assumption of the logistic regression model is the absence of perfect multicollinearity among features [60]. Figure S1 shows the pairwise Pearson correlation coefficient value between pairs of amino acid features used to train the logistic regression model. We note that only 1 out of 210 pairwise entries (%) of the correlation matrix lower triangle was a close-to-perfect correlation (i.e., r equal to or greater than 0.95). Therefore, most features did not have a close-to-perfect correlation with one another, and we consider this assumption to be largely held. This pairwise correlation of 0.96 was between isoleucine and leucine concentrations. Isoleucine and leucine were retained as features because they contain biological significance in MDD patients, as shown by Baranyi and colleagues [10]. Hence, it is unclear if isoleucine or leucine should be removed. Additionally, the mean test AUC performance (across 10 outer folds) numerically declined when the logistic regression model (with feature selection and hyperparameter optimisation) either dropped isoleucine concentration or leucine concentration features, from 0.76 ± 0.17 to 0.73 ± 0.17, and 0.73 ± 0.17, respectively (Table S1). Consequently, including both isoleucine and leucine concentration features leads to a more robust prediction performance.

In summary, our study highlights the potential of using amino acid-based machine learning analysis to predict and diagnose MDD due to its acceptable accuracy, interpretability, and stability. This approach can potentially enhance the reliability of MDD diagnosis in clinical practice, hence warranting further research in this area.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jcm13051222/s1, Figure S1: Pearson correlation coefficient (r) results between pairs of amino acid concentration features. Table S1: Validation and test set performance results for three variations of logistic regression models (all 3 with feature selection and with hyperparameter optimisation): original model, model without isoleucine concentration feature, and model without leucine concentration feature.

Author Contributions: Conceptualization, C.S.H.H. and T.B.T.; methodology, C.S.H.H. and T.B.T.; data collection, G.W.N.T.; data cleaning and preparation, T.W.K.T., G.W.N.T. and Y.L.C.; formal

analysis, T.W.K.T. and G.W.N.T.; writing—original draft preparation, C.S.H.H., T.W.K.T., G.W.N.T. and H.C.H.K.; writing—review and editing, C.S.H.H. and T.B.T.; supervision, C.S.H.H. and T.B.T.; project administration, G.W.N.T.; funding acquisition, C.S.H.H. and T.B.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National University Health System Singapore Seed Fund, August 2018 [grant number: NR19MRF015], as well as the Ministry of Higher Education, Malaysia, under the Higher Institution Centre of Excellence scheme awarded to the Centre for Intelligent Signal and Imaging Research, Universiti Teknologi PETRONAS [cost centre: 015MA0-050].

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Domain Specific Review Board of the National Healthcare Group, Singapore (protocol number 2019/00141) on 2 May 2019.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data supporting this study's findings are available upon request from co-first author Cyrus Su Hui Ho.

Acknowledgments: The authors would like to thank all study participants.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimates. 2017. Available online: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf?sequence=1 (accessed on 7 January 2024).
2. Costa, L.N.F.G.; Carneiro, B.A.; Alves, G.S.; Silva, D.H.L.; Guimaraes, D.F.; Souza, L.S.; Bandeira, I.D.; Beanes, G.; Scippa, A.M.; Quarantini, L.C. Metabolomics of Major Depressive Disorder: A Systematic Review of Clinical Studies. *Cureus* **2022**, *14*, e23009. [CrossRef]
3. Ormel, J.; Hartman, C.A.; Snieder, H. The genetics of depression: Successful genome-wide association studies introduce new challenges. *Transl. Psychiatry* **2019**, *9*, 114. [CrossRef] [PubMed]
4. Gandal, M.J.; Leppa, V.; Won, H.; Parikshak, N.N.; Geschwind, D.H. The road to precision psychiatry: Translating genetics into disease mechanisms. *Nat. Neurosci.* **2016**, *19*, 1397–1407. [CrossRef] [PubMed]
5. Altmann, A.; Tolosi, L.; Sander, O.; Lengauer, T. Permutation importance: A corrected feature importance measure. *Bioinformatics* **2010**, *26*, 1340–1347. [CrossRef] [PubMed]
6. Abdulaal, M.J.; Casson, A.J.; Gaydecki, P. Critical Analysis of Cross-Validation Methods and Their Impact on Neural Networks Performance Inflation in Electroencephalography Analysis. *IEEE Can. J. Electr. Comput. Eng.* **2021**, *44*, 75–82. [CrossRef]
7. Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A.J. Machine learning algorithm validation with a limited sample size. *PLoS ONE* **2019**, *14*, e0224365. [CrossRef] [PubMed]
8. Morilak, D.A.; Frazer, A. Antidepressants and brain monoaminergic systems: A dimensional approach to understanding their behavioural effects in depression and anxiety disorders. *Int. J. Neuropsychopharmacol.* **2004**, *7*, 193–218. [CrossRef] [PubMed]
9. Jenkins, T.A.; Nguyen, J.C.; Polglaze, K.E.; Bertrand, P.P. Influence of Tryptophan and Serotonin on Mood and Cognition with a Possible Role of the Gut-Brain Axis. *Nutrients* **2016**, *8*, 56. [CrossRef] [PubMed]
10. Baranyi, A.; Amouzadeh-Ghadikolai, O.; von Lewinski, D.; Rothenhausler, H.B.; Theokas, S.; Robier, C.; Mangge, H.; Reicht, G.; Hlade, P.; Meinitzer, A. Branched-Chain Amino Acids as New Biomarkers of Major Depression—A Novel Neurobiology of Mood Disorder. *PLoS ONE* **2016**, *11*, e0160542. [CrossRef]
11. Setoyama, D.; Kato, T.A.; Hashimoto, R.; Kunugi, H.; Hattori, K.; Hayakawa, K.; Sato-Kasai, M.; Shimokawa, N.; Kaneko, S.; Yoshida, S.; et al. Plasma Metabolites Predict Severity of Depression and Suicidal Ideation in Psychiatric Patients—A Multicenter Pilot Analysis. *PLoS ONE* **2016**, *11*, e0165267. [CrossRef]
12. Ding, X.; Yang, S.; Li, W.; Liu, Y.; Li, Z.; Zhang, Y.; Li, L.; Liu, S. The potential biomarker panels for identification of Major Depressive Disorder (MDD) patients with and without early life stress (ELS) by metabonomic analysis. *PLoS ONE* **2014**, *9*, e97479. [CrossRef]
13. Hung, C.I.; Lin, G.; Chiang, M.H.; Chiu, C.Y. Metabolomics-based discrimination of patients with remitted depression from healthy controls using (1)H-NMR spectroscopy. *Sci. Rep.* **2021**, *11*, 15608. [CrossRef] [PubMed]
14. Moaddel, R.; Shardell, M.; Khadeer, M.; Lovett, J.; Kadriu, B.; Ravichandran, S.; Morris, P.J.; Yuan, P.; Thomas, C.J.; Gould, T.D.; et al. Plasma metabolomic profiling of a ketamine and placebo crossover trial of major depressive disorder and healthy control subjects. *Psychopharmacology* **2018**, *235*, 3017–3030. [CrossRef] [PubMed]
15. Ciocan, D.; Cassard, A.M.; Becquemont, L.; Verstuyft, C.; Voican, C.S.; El Asmar, K.; Colle, R.; David, D.; Trabado, S.; Feve, B.; et al. Blood microbiota and metabolomic signature of major depression before and after antidepressant treatment: A prospective case-control study. *J. Psychiatry Neurosci.* **2021**, *46*, E358–E368. [CrossRef]

16. Liu, D.; Ray, B.; Neavin, D.R.; Zhang, J.; Athreya, A.P.; Biernacka, J.M.; Bobo, W.V.; Hall-Flavin, D.K.; Skime, M.K.; Zhu, H.; et al. Beta-defensin 1, aryl hydrocarbon receptor and plasma kynurenine in major depressive disorder: Metabolomics-informed genomics. *Transl. Psychiatry* **2018**, *8*, 10. [CrossRef]
17. Shin, C.; Kim, Y.K. Ketamine in Major Depressive Disorder: Mechanisms and Future Perspectives. *Psychiatry Investig.* **2020**, *17*, 181–192. [CrossRef] [PubMed]
18. Yang, Y.; Chen, J.; Liu, C.; Fang, L.; Liu, Z.; Guo, J.; Cheng, K.; Zhou, C.; Zhan, Y.; Melgiri, N.D.; et al. The Extrinsic Coagulation Pathway: A Biomarker for Suicidal Behavior in Major Depressive Disorder. *Sci. Rep.* **2016**, *6*, 32882. [CrossRef]
19. Cipriani, A.; Zhou, X.; Del Giovane, C.; Hetrick, S.E.; Qin, B.; Whittington, C.; Coghill, D.; Zhang, Y.; Hazell, P.; Leucht, S.; et al. Comparative efficacy and tolerability of antidepressants for major depressive disorder in children and adolescents: A network meta-analysis. *Lancet* **2016**, *388*, 881–890. [CrossRef]
20. Lin, E.; Lin, C.H.; Lane, H.Y. Precision Psychiatry Applications with Pharmacogenomics: Artificial Intelligence and Machine Learning Approaches. *Int. J. Mol. Sci.* **2020**, *21*, 969. [CrossRef]
21. Bzdok, D.; Meyer-Lindenberg, A. Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **2018**, *3*, 223–230. [CrossRef]
22. Davatzikos, C. Machine learning in neuroimaging: Progress and challenges. *NeuroImage* **2019**, *197*, 652–656. [CrossRef]
23. Zheng, H.; Zheng, P.; Zhao, L.; Jia, J.; Tang, S.; Xu, P.; Xie, P.; Gao, H. Predictive diagnosis of major depression using NMR-based metabolomics and least-squares support vector machine. *Clin. Chim. Acta* **2017**, *464*, 223–227. [CrossRef]
24. Fernandes, B.S.; Williams, L.M.; Steiner, J.; Leboyer, M.; Carvalho, A.F.; Berk, M. The new field of 'precision psychiatry'. *BMC Med.* **2017**, *15*, 80. [CrossRef]
25. Weissler, E.H.; Naumann, T.; Andersson, T.; Ranganath, R.; Elemento, O.; Luo, Y.; Freitag, D.F.; Benoit, J.; Hughes, M.C.; Khan, F.; et al. The role of machine learning in clinical research: Transforming the future of evidence generation. *Trials* **2021**, *22*, 537. [CrossRef] [PubMed]
26. Panch, T.; Szolovits, P.; Atun, R. Artificial intelligence, machine learning and health systems. *J. Glob. Health* **2018**, *8*, 020303. [CrossRef] [PubMed]
27. Friederich, P.; Krenn, M.; Tamblyn, I.; Aspuru-Guzik, A. Scientific intuition inspired by machine learning-generated hypotheses. *Mach. Learn. Sci. Technol.* **2021**, *2*, 025027. [CrossRef]
28. Au, Q.; Herbinger, J.; Stachl, C.; Bischl, B.; Casalicchio, G. Grouped feature importance and combined features effect plot. *Data Min. Knowl. Discov.* **2022**, *36*, 1401–1450. [CrossRef]
29. Pan, J.X.; Xia, J.J.; Deng, F.L.; Liang, W.W.; Wu, J.; Yin, B.M.; Dong, M.X.; Chen, J.J.; Ye, F.; Wang, H.Y.; et al. Diagnosis of major depressive disorder based on changes in multiple plasma neurotransmitters: A targeted metabolomics study. *Transl. Psychiatry* **2018**, *8*, 130. [CrossRef] [PubMed]
30. Stiglic, G.; Kocbek, P.; Fijacko, N.; Zitnik, M.; Verbert, K.; Cilar, L. Interpretability of machine learning-based prediction models in healthcare. *WIREs Data Min. Knowl. Discov.* **2020**, *10*, e1379. [CrossRef]
31. Ning, Y.; Li, S.; Ong, M.E.H.; Xie, F.; Chakraborty, B.; Ting, D.S.W.; Liu, N. A novel interpretable machine learning system to generate clinical risk scores: An application for predicting early mortality or unplanned readmission in a retrospective cohort study. *PLoS Digit. Health* **2022**, *1*, e0000062. [CrossRef]
32. Kyeong, S.; Shin, J. Two-stage credit scoring using Bayesian approach. *J. Big Data* **2022**, *9*, 106. [CrossRef]
33. Shevade, S.K.; Keerthi, S.S. A simple and efficient algorithm for gene selection using sparse logistic regression. *Bioinformatics* **2003**, *19*, 2246–2253. [CrossRef]
34. Choi, E.; Bahadori, M.T.; Kulas, J.A.; Schuetz, A.; Stewart, W.F.; Sun, J. RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism. In Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 3512–3520.
35. Sheu, Y.H. Illuminating the Black Box: Interpreting Deep Neural Network Models for Psychiatric Research. *Front. Psychiatry* **2020**, *11*, 551299. [CrossRef]
36. Xu, H.B.; Fang, L.; Hu, Z.C.; Chen, Y.C.; Chen, J.J.; Li, F.F.; Lu, J.; Mu, J.; Xie, P. Potential clinical utility of plasma amino acid profiling in the detection of major depressive disorder. *Psychiatry Res.* **2012**, *200*, 1054–1057. [CrossRef]
37. Pedregosa, F.; Varoquaux, G.L.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
38. Borah, S.; Gogoi, P.; Gohain, P.; Boro, C.; Muchahari, M.K. Machine Learning for Detection of Obstructive Sleep Apnoea. *Smart Intell. Comput. Appl.* **2022**, *2*, 243–251. [CrossRef]
39. Shahid, F.; Zameer, A.; Muneeb, M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos Solitons Fractals* **2020**, *140*, 110212. [CrossRef]
40. Niemeyer, H.; Bieda, A.; Michalak, J.; Schneider, S.; Margraf, J. Education and mental health: Do psychosocial resources matter? *SSM Popul. Health* **2019**, *7*, 100392. [CrossRef] [PubMed]
41. Zalar, B.; Blatnik, A.; Maver, A.; Klemenc-Ketiš, Z.; Peterlin, B. Family History as an Important Factor for Stratifying Participants in Genetic Studies of Major Depression. *Balk. J. Med. Genet.* **2018**, *21*, 5–12. [CrossRef] [PubMed]
42. Thorp, J.G.; Gerring, Z.F.; Colodro-Conde, L.; Byrne, E.M.; Medland, S.E.; Middeldorp, C.M.; Derks, E.M. The association between trauma exposure, polygenic risk and individual depression symptoms. *Psychiatry Res.* **2023**, *321*, 115101. [CrossRef] [PubMed]

43. Wang, D.-F.; Zhou, Y.-N.; Liu, Y.-H.; Hao, Y.-Z.; Zhang, J.-H.; Liu, T.-Q.; Ma, Y.-J. Social support and depressive symptoms: Exploring stigma and self-efficacy in a moderated mediation model. *BMC Psychiatry* **2022**, *22*, 117. [CrossRef] [PubMed]
44. Liu, X.; Zheng, P.; Zhao, X.; Zhang, Y.; Hu, C.; Li, J.; Zhao, J.; Zhou, J.; Xie, P.; Xu, G. Discovery and validation of plasma biomarkers for major depressive disorder classification based on liquid chromatography-mass spectrometry. *J. Proteome Res.* **2015**, *14*, 2322–2330. [CrossRef]
45. Islam, M.R.; Ali, S.; Karmoker, J.R.; Kadir, M.F.; Ahmed, M.U.; Nahar, Z.; Islam, S.M.A.; Islam, M.S.; Hasnat, A.; Islam, M.S. Evaluation of serum amino acids and non-enzymatic antioxidants in drug-naive first-episode major depressive disorder. *BMC Psychiatry* **2020**, *20*, 333. [CrossRef] [PubMed]
46. Kawamura, N.; Shinoda, K.; Sato, H.; Sasaki, K.; Suzuki, M.; Yamaki, K.; Fujimori, T.; Yamamoto, H.; Osei-Hyiaman, D.; Ohashi, Y. Plasma metabolome analysis of patients with major depressive disorder. *Psychiatry Clin. Neurosci.* **2018**, *72*, 349–361. [CrossRef] [PubMed]
47. Ong, S.K.; Husain, S.F.; Wee, H.N.; Ching, J.; Kovalik, J.P.; Cheng, M.S.; Schwarz, H.; Tang, T.B.; Ho, C.S. Integration of the Cortical Haemodynamic Response Measured by Functional Near-Infrared Spectroscopy and Amino Acid Analysis to Aid in the Diagnosis of Major Depressive Disorder. *Diagnostics* **2021**, *11*, 1978. [CrossRef] [PubMed]
48. Ogawa, S.; Koga, N.; Hattori, K.; Matsuo, J.; Ota, M.; Hori, H.; Sasayama, D.; Teraishi, T.; Ishida, I.; Yoshida, F.; et al. Plasma amino acid profile in major depressive disorder: Analyses in two independent case-control sample sets. *J. Psychiatr. Res.* **2018**, *96*, 23–32. [CrossRef] [PubMed]
49. Bzdok, D.; Altman, N.; Krzywinski, M. Statistics versus machine learning. *Nat. Methods* **2018**, *15*, 233–234. [CrossRef]
50. Ley, C.; Martin, R.K.; Pareek, A.; Groll, A.; Seil, R.; Tischer, T. Machine learning and conventional statistics: Making sense of the differences. *Knee Surg. Sports Traumatol. Arthrosc.* **2022**, *30*, 753–757. [CrossRef]
51. Liu, Y.; Yieh, L.; Yang, T.; Drinkenburg, W.; Peeters, P.; Steckler, T.; Narayan, V.A.; Wittenberg, G.; Ye, J. Metabolomic biosignature differentiates melancholic depressive patients from healthy controls. *BMC Genom.* **2016**, *17*, 669. [CrossRef]
52. Okuda, S.; Nishiyama, N.; Saito, H.; Katsuki, H. 3-Hydroxykynurenine, an endogenous oxidative stress generator, causes neuronal cell death with apoptotic features and region selectivity. *J. Neurochem.* **1998**, *70*, 299–307. [CrossRef] [PubMed]
53. Liu, H.; Ding, L.; Zhang, H.; Mellor, D.; Wu, H.; Zhao, D.; Wu, C.; Lin, Z.; Yuan, J.; Peng, D. The Metabolic Factor Kynurenic Acid of Kynurenine Pathway Predicts Major Depressive Disorder. *Front. Psychiatry* **2018**, *9*, 552. [CrossRef]
54. Hoes, M.J.; Sijben, N. The clinical significance of disordered renal excretion of xanthurenic acid in depressive patients. *Psychopharmacology* **1981**, *75*, 346–349. [CrossRef] [PubMed]
55. Blanke, M.; VanDongen, A. Activation Mechanisms of the NMDA Receptor. In *Biology of the NMDA Receptor*; VanDongen, A., Ed.; CRC Press/Taylor & Francis: Boca Raton, FL, USA, 2009.
56. Newcomer, J.W.; Farber, N.B.; Olney, J.W. NMDA receptor function, memory, and brain aging. *Dialogues Clin. Neurosci.* **2000**, *2*, 219–232. [CrossRef] [PubMed]
57. Bot, M.; Milaneschi, Y.; Al-Shehri, T.; Amin, N.; Garmaeva, S.; Onderwater, G.L.J.; Pool, R.; Thesing, C.S.; Vijfhuizen, L.S.; Vogelzangs, N.; et al. Metabolomics Profile in Depression: A Pooled Analysis of 230 Metabolic Markers in 5283 Cases with Depression and 10,145 Controls. *Biol. Psychiatry* **2020**, *87*, 409–418. [CrossRef] [PubMed]
58. Koochakpoor, G.; Salari-Moghaddam, A.; Keshteli, A.H.; Afshar, H.; Esmaillzadeh, A.; Adibi, P. Dietary intake of branched-chain amino acids in relation to depression, anxiety and psychological distress. *Nutr. J.* **2021**, *20*, 11. [CrossRef]
59. Williamson, D.J.; McTavish, S.F.; Park, S.B.; Cowen, P.J. Effect of valine on 5-HT-mediated prolactin release in healthy volunteers, and on mood in remitted depressed patients. *Br. J. Psychiatry* **1995**, *167*, 238–242. [CrossRef]
60. Harris, J.K. Primer on binary logistic regression. *Fam. Med. Community Health* **2021**, *9* (Suppl. S1), e001290. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.