Biologically Interpretable Deep Learning for Metabolomics

Predicting Depression with Biological Insight

Tom Kitak[1]

Supervisor(s): Marcel J.T. Reinders[1]**, Gennady V. Roshchupkin** [2]

[1]**EEMCS, Delft University of Technology, The Netherlands**
[2]**Department of Epidemiology; Department of Radiology and Nuclear Medicine, Erasmus MC, The Netherlands**

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 23, 2024

Name of the student: Tom Kitak
Final project course: CSE3000 Research Project
Thesis committee: Marcel J.T. Reinders, Gennady V. Roshchupkin, Neil Yorke-Smith

Abstract

Depression, a leading cause of disability worldwide, is challenging to diagnose due to its reliance on subjective clinical evaluations. Metabolomics, which analyzes small molecules to reflect physiological and pathological states, holds promise for enhancing the diagnosis and identifying biomarkers for depression, potentially leading to better understanding and treatment options. Despite the complexity of metabolomics data, deep learning methods have not been extensively explored due to issues with interpretability, which are crucial for gaining insights into biological mechanisms. This study evaluates a biologically interpretable deep neural network, MetaboNet, trained on metabolomics data, for predicting depression and identifying key metabolites and biochemical pathways relevant to the condition. Our results demonstrate that MetaboNet outperforms logistic regression, though the overall classification performance remains modest. Notably, the classification results revealed sex-related differences, with better performance observed in females. Our findings do not support the capability of MetaboNet to identify biologically relevant individual metabolites. However, MetaboNet shows promise in identifying biochemical sub-pathways and super-pathways relevant to depression, which are validated by existing literature.

1 Introduction

Metabolomics is the extensive analysis of metabolites, which are small molecules such as amino acids, lipids, bile acids, carbohydrates, and organic acids [1]. Within the omics field, metabolomics is recognized as the closest to the phenotype because it reflects information from various omics layers [2–4]. Metabolomics offers a comprehensive snapshot of an organism's metabolic state by profiling and measuring a wide range of metabolites, making it possible to identify metabolic signatures linked to various physiological and pathological conditions [5]. Due to these capabilities, metabolomics is a rapidly emerging field [1], with broad implications and potential applications in prognosis, therapy, diagnosis, and the development of personalized medicine, particularly because of its direct applicability to biomarker discovery [6, 7]. Biomarkers are quantifiable attributes of an individual that potentially signify risk factors for a disease or outcome, or serve as indicators of the advancement of a disease or alterations associated with treatment [8].

One of the major public health issues that could benefit from advancements in metabolomics is depression [9–13]. Depression is a prevalent and serious psychological condition marked by feelings of sadness and a lack of interest, which may potentially result in suicidal tendencies [12]. The World Health Organization ranks depression as the leading cause of disability worldwide [14]. In the current medical practice, the diagnosis and treatment of depression rely on subjective evaluations of various symptoms that represent different underlying characteristics [15]. Mental health specialists typically conduct routine medical and behavioural assessments through interviews to diagnose depression [16]. However, this diagnostic approach is not economical, and obtaining an accurate diagnosis necessitates highly qualified medical professionals, more time to perform, and a psychometric evaluation based on interviews [17]. Consequently, there is a pressing need to establish an objective and practical standard for diagnosing depression [9] and metabolomics biomarkers may be a promising approach as many studies have shown altered metabolomics profiles in people with depression [9–13, 17–19]. Furthermore, metabolic disorders are viewed as contributing factors to depression [11]. Identifying new biomarkers would not only lead to more accurate diagnoses but also enhance our knowledge of the underlying mechanisms of the disease, which is useful for better therapeutic management of depression [9].

Utilization of machine learning (ML) in metabolomics may significantly improve diagnosis, identify more targets for therapies, and enable more accurate predictions of disease outcomes by potentially better addressing the inherent challenges of metabolomics data [1]. Metabolomics data is characterized by linear and nonlinear correlations among metabolites, as well as challenges like missing values, batch effects during quantification, data noise, and the challenge of high-dimensional data [1]. In recent years, the most commonly applied ML algorithms in metabolomics research are random forest, support vector machines, and logistic regression (LR) [2]. Despite deep learning (DL) outperforming traditional algorithms in scenarios involving high-dimensional, large-scale, and particularly complex data [20, 21], its application in metabolomics is relatively recent and still emerging, especially when compared to other omics fields [2]. This slow adoption may be attributed to challenges in interpretability and explainability, which are crucial for gaining insights into biological mechanisms [20, 21], and the significant computational power required [2].

To address the research gap in depression and the challenges associated with deep learning, we propose an interpretable deep neural network called MetaboNet for predicting depression from metabolomics data. MetaboNet incorporates prior biological knowledge about biochemical pathways to establish meaningful connections. Consequently, the neural network is memory efficient, interpretable, and provides biological insights for its predictions. The model architecture is based on the GenNet architecture described by van Hilten et al. [22].

The main objective of this study is to assess the performance of MetaboNet in comparison to LR, a widely used machine learning model in metabolomics [2], for classifying individuals with depression based on metabolomics data. Additionally, this research aims to validate the effectiveness of MetaboNet and LR in identifying biologically important metabolites for predicting depression. For MetaboNet, we will further analyze its biologically interpretable architecture to assess whether the model identifies biologically relevant biochemical sub-pathways and super-pathways important for depression. The identified important metabolites and bio-

chemical pathways will be cross-referenced with existing literature to ensure their biological relevance with depression, thereby validating the models' potential to discover new biological insights for depression.

2 Methodology

2.1 Metabolomics Dataset and Prior Knowledge

Our study utilized the Rotterdam study [23]. The dataset comprises 597 females, of which 62 are considered depressed, and 470 males, of which 24 are considered depressed. In total, there are 86 individuals with depression and 982 healthy control subjects.

Depression Measurement

The measurement of depression was conducted on individuals who completed the Center for Epidemiologic Studies Depression scale (CES-D). The CES-D is a widely used self-report scale designed to measure depressive symptomatology in the general population [24]. A score $\geq$ 16 was used to indicate depression [25, 26].

Metabolite Measurements

The Rotterdam study [23] analyzed blood metabolites from 1082 participants using the untargeted Metabolon HD4 platform, which encompasses 1387 metabolites across various biochemical pathways, including lipids, amino acids, xenobiotics, nucleotides, cofactors and vitamins, peptides, carbohydrates, energy-related metabolites, and uncharacterized metabolites. Before analysis, the data underwent preprocessing. Initially, 14 samples missingness exceeding five times the standard deviation (SD) of the mean missingness in all samples were excluded. Additionally, metabolites with missingness greater than five times the SD of the mean missingness in metabolites and a coefficient of variance exceeding 30% in internal control samples (NIST Standard Reference Material) were excluded. The remaining metabolites (N = 1111) were log2 transformed. In addition, metabolites that had missingness greater than thirty percent were eliminated, leaving 991 metabolites in 1068 samples. Finally, missing values were imputed using the K-nearest neighbor method, which has demonstrated robust performance across various evaluation schemes [27].

Biochemical Pathways

The prior knowledge pathway annotations were obtained using the Metabolon HD4 platform. The biochemical super-pathways of the metabolites are categorized as follows: lipids (390 metabolites), amino acids (197 metabolites), xenobiotics (99 metabolites), nucleotides (32 metabolites), cofactors and vitamins (27 metabolites), peptides (24 metabolites), carbohydrates (17 metabolites), energy-related metabolites (8 metabolites), partially characterizable molecules (14 metabolites), and uncharacterized metabolites (183 metabolites). There were further 99 sub-pathway groups used for metabolites. The complete table of metabolites, sub-pathways and super-pathways can be found in the *annotations.txt* file in MetaboNet repository [28].

Dataset Visualization

We used t-Distributed Stochastic Neighbor Embedding (t-SNE) [29] in order to visualize the high-dimensional data in 2 dimensions. This non-linear technique was chosen because of its ability to effectively preserve the local structure of the data, potentially allowing for the identification of clusters and patterns that are not easily detectable in higher dimensions. Random seed 0 was used to ensure reproducibility.

2.2 Machine Learning Used for Classification

MetaboNet uses prior biological knowledge about biochemical pathways to establish meaningful connections as portrayed graphically in Figure 1. Metabolites are connected to the first hidden layer, where neurons represent sub-pathways. These sub-pathways are then connected to super-pathways in the second hidden layer. To achieve a fair interpretation of connection weights, each layer is preceded by batch normalization, without scaling and centring, to standardize the input to zero mean and unit standard deviation. The sigmoid function was used as the final activation function for binary classification. This architecture makes the network inherently interpretable and lightweight, with fewer learnable parameters compared to a fully connected neural network. MetaboNet is implemented using PyTorch (version 2.3.0.) [30] and the codebase is available on GitHub [28].

The baseline model used was LR, one of the most commonly used models in metabolomics [2]. It was implemented with PyTorch using a linear layer with a sigmoid function.

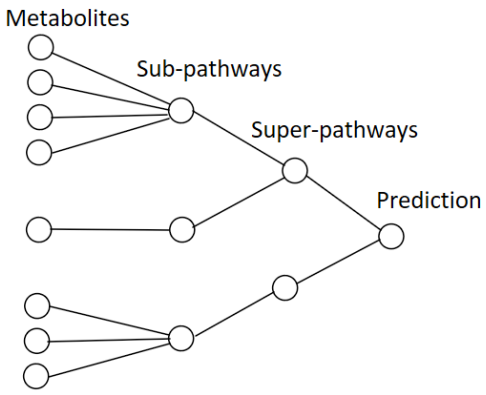

Figure 1: MetaboNet is constructed using prior biological knowledge about biochemical sub-pathways and super-pathways to establish meaningful connections. This architecture makes the network inherently interpretable and lightweight.

2.3 Machine Learning Model Development and Evaluation Workflow

The workflow for model development was the same for both models to ensure fair comparability. The model development workflow followed the general methodology described by Galal et al. [2]. The initial workflow optimized the hyperparameters, and then with those optimized hyperparameters, we evaluated the performance of the model in the evaluation workflow.

To ensure reproducibility, account for randomness, and achieve more robust results, we ran workflows multiple times with different random seeds. By varying the random seeds, we ensured diverse train-validation-test splits. This approach not only captures the overall performance but also provides insights into the models' stability across different random runs.

The tuning metric was the Matthews correlation coefficient (MCC) [31], as it accounts for all four categories of the confusion matrix, providing a balanced measure even with an unbalanced dataset.

Using a batch size of 32, every model was trained on a single GPU Nvidia GeForce GTX 2080 Ti.

Hyperparameter Optimization Workflow

Firstly, the dataset was split into 60% training, 20% validation, and 20% test sets, ensuring an equal proportion of both classes in all sets. The splits were performed using the *train_test_split* function from the scikit-learn Python library (version 1.5.0) [32]. Next, hyperparameter grid search was achieved through stratified 5-fold cross-validation on the training set, utilizing the *StratifiedKFold* and *ParameterGrid* classes from the scikit-learn library. We trained the model for 100 epochs. To obtain more stable models, we decided to determine the best model based on the average of the top 10 best-performing epoch model states, and then average their weights. The best-performing model from the hyperparameter grid search was evaluated on an independent validation set. Hyperparameter optimization workflow was run 5 times with seeds 1 to 5 and the best-performing hyperparameters on validation set were selected to be final hyperparameters.

Hyperparameter Grid

The hyperparameters optimized for both models were positive class weight, learning rate, L1 regularization penalty, and scheduler. For MetaboNet we also optimized the activation function. Options for the positive class weight were 8, 12, and 16, with 12 chosen as the middle value because it represents the ratio of the count of negative class samples to positive class samples. Learning rates considered were 0.01 and 0.001. L1 regularization penalty options were 0.1, 0.01, 0.001, and 0. We tested three PyTorch schedulers: *StepLR*, with a step size of 30 and gamma of 0.1, *ReduceLROnPlateau* with a patience of 20 and factor of 0.1, and *CosineAnnealingLR* with T_max set to 100. We evaluated three activation functions for MetaboNet: hyperbolic tangent, ReLU, and PReLU. The hyperbolic tangent function was coupled with Xavier uniform weight initialization, while ReLU and PReLU were combined with Kaiming normal weight initialization.

Evaluation Workflow

The dataset was split into 60% training, 20% validation, and 20% test sets. Next, the final model was trained on the training set using the optimized hyperparameters and validated on the validation set. We ran the training for 100 epochs, and the 10 best-performing epoch model states on the validation set were averaged to obtain the final model. The final model was then evaluated on an independent test set. This evaluation workflow was repeated 100 times with seeds 1 to 100 to get a dataset of classification performances.

2.4 Evaluation Metrics

In the context of our binary classification task with significant class imbalance, we decided to use the following metrics: MCC, F1 score, area under the precision-recall curve (AUC-PR), and area under the receiver operating characteristic curve (AUC-ROC). Collectively, these metrics provide a comprehensive view of classification performance.

MCC [31] is a robust metric that considers all four quadrants of the confusion matrix and is generally regarded as a balanced measure that can be used even if the classes are of different sizes. The F1 score [33] is the harmonic mean of precision and recall, and it provides a balanced measure of a model's performance by considering both false positives and false negatives. The PR-AUC [34] plots precision against recall for different threshold values. The area under this curve provides an aggregate measure of performance across all thresholds, focusing on the positive class . The AUC-ROC [33] metric measures the model's ability to distinguish between classes. It plots the true positive rate against the false positive rate at various threshold settings. The area under this curve is a single scalar value that summarizes the model's performance across all thresholds.

The metrics were implemented using the scikit-learn Python library (version 1.5.0) [32].

2.5 Statistical Tests

The relative classification performance across model evaluation workflow runs of MetaboNet and LR was assessed with paired t-tests since the train-test splits were the same, making the samples dependent. An independent t-test was employed to evaluate the relative classification performance across sexes, comparing the performance metrics between male and female groups.

Additionally, the Mann-Whitney U test was utilized to assess the t-SNE visualization, specifically to determine the similarity between the dimensions of the depressed and control groups. This non-parametric test was chosen due to its robustness against non-normality.

We considered the threshold for statistical significance to be 0.05 for all statistical tests, and they were implemented using the SciPy Python library (version 1.13.1) [35].

2.6 Interpreting Machine Learning Models to Identify Important Metabolites and Biochemical Pathways

We analyzed MetaboNet and LR models to identify the most important metabolites for classification prediction. The MetaboNet architecture facilitated a further examination of the importance of biochemical pathways. To ensure robust relative importance, we averaged the relative importance values obtained from 100 different model development workflows. We identified the top seven metabolites from both models and cross-referenced them with existing literature for validation. Additionally, we compared the top four MetaboNet sub-pathways and the most significant superpathway with findings from previous studies to confirm their relevance.

MetaboNet architecture, defined by prior biological knowledge about biochemical pathways, ensures innate interpretability, with each node and connection representing a specific biological entity. The weights learned between layers reflect the influence of metabolites on sub-pathways or the impact of super-pathways on prediction, where stronger weights indicate greater importance. The relative importance in MetaboNet is determined by multiplying the weights along the path from the end node to the input metabolite. At each input, a value representing its contribution is obtained. To get the relative feature importance, the absolute values of weights are normalized by the absolute value of the sum over the weights on that layer. The relative importance is then summed according to the groups of the next layer to obtain a relative importance estimate for each node in the network.

In the LR model, we determined the most important metabolites by comparing the absolute values of the coefficients, since the inputs were already normalized.

It is important to note that in our context when comparing relative importance, it is more akin to effect size or odds ratio than to statistical significance [22].

3 Results and Discussion

3.1 The Depression and Control Groups Show Significant Similarities in t-SNE Data Visualization

To gain insight into the separability of the data, we employed the t-SNE visualization technique, as illustrated in Figure 2. Additionally, the Mann-Whitney U test was conducted to assess the differences between the depression and control groups within the two-dimensional space generated by t-SNE. The results indicate a substantial overlap between the distributions of the depression and control groups.

The mean and standard deviation of t-SNE dimensions are summarized in table 1, along with Mann-Whitney U test p-values. For Dimension 1, the p-value value was 0.33, while for Dimension 2, the p-value was 0.34. Both p-values exceed the 0.05 significance threshold, indicating no statistically significant differences between the groups in either dimension. These results suggest that the distributions of the depression and control groups overlap substantially, implying a lack of clear separability in the current dataset. This substantial overlap is consistent with the visual observation from the t-SNE plot, where no distinct clustering of the groups was evident. The inability to differentiate between depressed individuals and controls based on this analysis suggests that the current dataset might lack features that provide meaningful discrimination.

3.2 Modest Classification Performance With MetaboNet Outperforming Logistic Regression

Our objective was to evaluate the classification performance in correctly identifying depressed and healthy individuals using the metabolomics dataset with MetaboNet and LR models, and then investigate their relative performance. The performance of both models is modest; however, MetaboNet shows a pattern of outperforming LR.

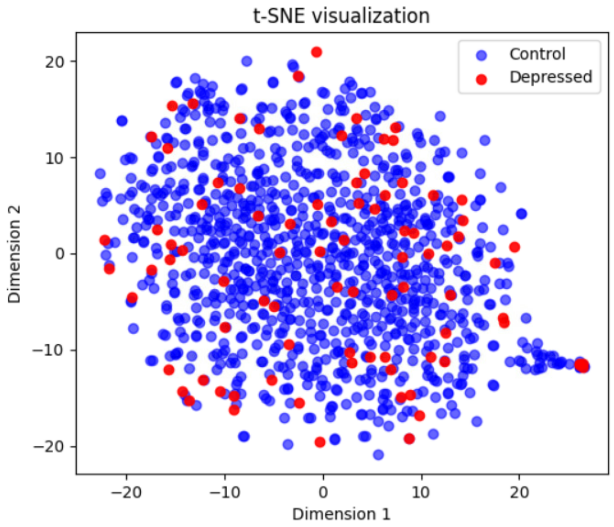

Figure 2: Significant overlap can be observed between the depression and control groups in the two-dimensional t-SNE visualization. There are no obvious patterns or distinct clusters that would suggest that the two groups can be easily separated.

Dimension	Depressed	Control	p-values
Dimension 1	0.82 ± 12.0	-0.44 ± 10.5	0.33
Dimension 2	-1.52 ± 9.9	-0.55 ± 9.0	0.34

Table 1: Mean and standard deviation of t-SNE dimensionality reduction and Mann-Whitney U test p-value for depressed and control groups. There are no statistically significant differences between the groups in either dimension, as indicated by the two p-values exceeding the 0.05 significance threshold. These findings indicate that there is a significant overlap between the depression and control group distributions.

To obtain robust results and account for randomness, we ran the evaluation workflow for both models using 100 different random seeds. The metrics we used to evaluate the efficacy of the model classification were MCC, F1 score, PR-AUC, and ROC-AUC. The mean and standard deviation results are presented in Table 2. We validated the relative performance of both models using paired t-tests, with results also shown in Table 2. Additionally, we plotted the averaged confusion matrix, which is displayed in Figure 3.

Metric	MetaboNet	LR	p-value
MCC	0.171 ± 0.066	0.148 ± 0.078	0.023
F1 score	0.236 ± 0.043	0.225 ± 0.068	0.173
PR-AUC	0.189 ± 0.060	0.160 ± 0.044	$2 \cdot 10^{-4}$
ROC-AUC	0.702 ± 0.054	0.650 ± 0.060	$5 \cdot 10^{-9}$

Table 2: Comparison of performance metrics (mean $\pm$ standard deviation) and paired t-test p-values between MetaboNet and LR depression classification. MetaboNet shows a pattern of outperforming LR. The performance of both models remains modest. Random baselines are: 0 for MCC, 0.08 for F1 score and PR-AUC, and 0.5 for ROC-AUC.

The performance metrics mean and standard deviation for

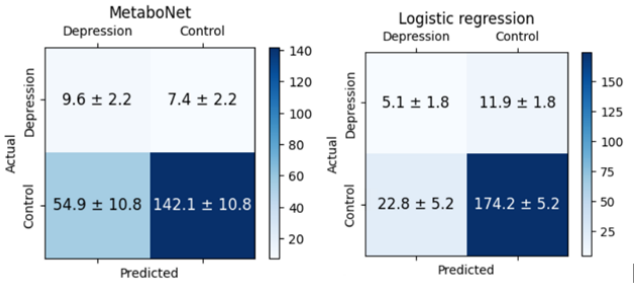

Figure 3: Averaged confusion matrices for MetaboNet (left) and LR (right) depression classification, with values shown as mean ± standard deviation. MetaboNet shows better identification of depressed individuals with more true positives and fewer false negatives. In contrast, LR correctly classifies more healthy individuals (true negatives) and has fewer false positives.

MetaboNet and LR are summarized in Table 2. MetaboNet achieved an MCC of 0.171 ± 0.066, higher than LR's MCC of 0.148 ± 0.078 ($p = 0.023$), indicating a statistically significant difference. The F1 score for MetaboNet was 0.236 ± 0.043, compared to 0.225 ± 0.068 for LR ($p = 0.173$), showing no statistically significant difference. PR-AUC for MetaboNet was 0.189 ± 0.060, outperforming LR's 0.160 ± 0.044 ($p = 2 \cdot 10^{-4}$), with statistical significance. ROC-AUC also demonstrated MetaboNet's superior performance at 0.702 ± 0.054 compared to LR's 0.650 ± 0.060 ($p = 5 \cdot 10^{-9}$).

Both models demonstrate the ability to classify depression more effectively than random prediction, though the improvement is modest. To provide context, MCC ranges from -1 to 1, where 1 indicates perfect prediction, 0 indicates random prediction, and -1 indicates complete disagreement between prediction and observation [31]. For the F1 score and PR-AUC, values range from 0 to 1, with 0.08 representing the random prediction score, reflecting the proportion of the depression class in the dataset. The AUC-ROC has a random baseline of 0.5, with scores ranging from 0 (worst) to 1 (best).

The modest performance of both models could be partially attributed to several factors, including the limited size of the dataset, the high dimensionality of the data, and the significant class imbalance, with only 8% of the dataset consisting of depressed individuals. These factors likely constrain the models' ability to generalize and accurately classify depressed individuals versus healthy controls.

The results demonstrate that MetaboNet outperforms LR in several key metrics, namely MCC, PR-AUC, and ROC-AUC, in a statistically significant manner. Although MetaboNet also outperformed LR in terms of the F1 score, this difference did not reach statistical significance at the 0.05 level.

The superior performance of MetaboNet over LR can potentially be explained by the different architecture of the models. LR is a linear model, while MetaboNet is capable of modeling non-linear relationships. Given that metabolites are known to exhibit non-linear relationships [1], MetaboNet's ability to capture these complexities likely contributes to its better performance. Moreover, the high dimensionality of the data favors deep learning models like MetaboNet, which are known to handle high-dimensional data more effectively than traditional machine learning models [20, 21].

Figure 3 presents the confusion matrices for MetaboNet and LR. MetaboNet demonstrates a superior capability in identifying depressed individuals, as evidenced by a higher number of true positives and a lower number of false negatives compared to LR. Conversely, LR shows a higher number of correctly classified healthy individuals (true negatives) and fewer false positives than MetaboNet.

These results suggest that MetaboNet is more effective in detecting depression cases, reducing the likelihood of missing depressed individuals. However, this comes at the cost of a higher rate of false positives, indicating a tendency to misclassify healthy individuals as depressed.

Overall, these results suggest that while both models exhibit limited performance on the current dataset, MetaboNet outperforms LR, highlighting a promising direction for improving depression diagnostics. Future work should focus on acquiring larger datasets with more depressed samples to enhance model training. Additionally, advanced feature selection techniques and dimensionality reduction methods should be explored to better manage the high dimensionality of the data. Finally, applying these models to larger and more diverse cohorts will be essential for validating their effectiveness in real-world clinical settings.

3.3 Superior Female Classification Performance

To evaluate potential biases in our depression classification models related to the sex covariate, we examined how well the MetaboNet and LR models, trained on metabolomics data, classified samples within each sex category. Our analysis revealed a statistically significant trend of better classification performance for females across several performance metrics in both models.

The classification results from running the evaluation workflow with 100 different random seeds were grouped and analyzed based on sex. The performance of the models for each sex was assessed using MCC, F1 score, PR-AUC, and ROC-AUC metrics. Independent t-tests were employed to validate the relative performance differences between the two sex groups.

The performance metrics for MetaboNet and LR by sex are summarized in Table 3. For MetaboNet, statistically significant differences were observed between males and females in MCC, F1 score, and PR-AUC, with females showing better performance in these metrics. However, no significant difference was found in ROC-AUC between males and females. For LR, female classification performed better across all metrics. The F1 score and PR-AUC showed statistically significant differences, while MCC and ROC-AUC had p-values slightly above the 0.05 threshold, at 0.08 and 0.06, respectively.

The results indicate a general trend of better performance for females compared to males across several metrics, suggesting that the models may more effectively capture patterns related to depression in females. However, it is important to recognize that the ROC-AUC metric, which did not show sta-

	MetaboNet			Logistic regression		
Metric	Male	Female	p-value	Male	Female	p-value
MCC	0.14 ± 0.11	0.18 ± 0.08	$2 \cdot 10^{-3}$	0.12 ± 0.14	0.15 ± 0.11	0.08
F1 score	0.17 ± 0.08	0.27 ± 0.06	$5 \cdot 10^{-21}$	0.17 ± 0.13	0.25 ± 0.10	$3 \cdot 10^{-7}$
PR-AUC	0.13 ± 0.09	0.24 ± 0.09	$8 \cdot 10^{-16}$	0.13 ± 0.12	0.20 ± 0.07	$2 \cdot 10^{-6}$
ROC-AUC	0.70 ± 0.13	0.69 ± 0.07	0.51	0.62 ± 0.16	0.65 ± 0.07	0.06

Table 3: Comparison of performance metrics (mean ± standard deviation) and independent t-test p-values by sex for MetaboNet and LR depression classification. Female classification performance is significantly better than male classification performance in F1 score and PR-AUC for both models, and in MCC score for MetaboNet. This indicates a strong trend of sex-related differences in the models' performance.

tistically significant differences in classification performance between sexes, has a limitation relevant to our context: it can be overly optimistic for imbalanced datasets where the majority class is well classified [33]. Therefore, the ROC-AUC metric could be misleading because the majority class, or control group, is well classified, as can be seen in Figure 3.

The better female classification performance trend could potentially be explained by the fact that there are more depressed females than males in our dataset, with 73% of the depressed samples being female. There could also be different metabolomic profiles of depression for males and females, which means that the smaller dataset for males may fail to generalize and learn about the male depression metabolomic profile. There is evidence supporting significant differences in metabolomic profiles between males and females [36].

A potential problem could be that the models may not always be using biologically relevant metabolites for depression but instead using metabolites that predict sex, thereby acting as a proxy for predicting depression since women are more likely to be depressed in our dataset and the general population [37]. Thus, the lack of accounting for a sex covariate in our classification is a limitation of our study.

Overall, our analysis revealed sex-related differences in the performance of depression classification models, with better classification performance for females. Future work should focus on obtaining unbiased datasets and incorporating sex covariates, as well as potentially other important covariates, to improve the generalizability and accuracy of predictive models.

3.4 Metabolite and Pathway Importance Analysis Validates MetaboNet's Ability to Identify Biologically Relevant Pathways

Our objective was to validate the effectiveness of MetaboNet and LR models' identification of biologically important metabolites for predicting depression. For MetaboNet, we further analyzed the biologically interpretable structure of the model to derive insights about the importance of biochemical sub-pathways and super-pathways, as illustrated in Figure 4. We did not find evidence of either model accurately identifying the correct important metabolites, however, MetaboNet showed compelling evidence it is capable of correctly identifying important biochemical pathways. The most compelling evidence, based on our relative importance score and validation from previous work, highlighted the significance of food-related metabolites for depression.

To obtain robust results, both models were trained, validated, and tested across 100 different random seeds, and we then averaged the relative importance values of the models. We cross-referenced the most important metabolites, and biochemical pathways with existing literature for validation.

Important Metabolites Lack Validation From the Literature

The analysis of the top seven metabolites most important for the prediction models MetaboNet and LR revealed one common metabolite: *methionine sulfone*. The other metabolites identified by MetaboNet include *catechol sulfate*, *tiglylcarnitine (C5:1-DC)*, *beta-hydroxyisovalerate*, *N,N,N-trimethyl-5-aminovalerate*, *N-acetyl-2-aminoadipate*, and *isovalerylcarnitine (C5)*. Conversely, the top metabolites identified by LR are *sphinganine-1-phosphate*, *3-methoxytyramine sulfate*, *N-acetylasparagine*, *glucuronate*, *2-O-methylascorbic acid*, and *N-stearoyl-sphinganine (d18:0/18:0)*. To our knowledge, none of these metabolites have been validated by other studies.

The poor overlap in the top seven most important metabolites between the MetaboNet and LR models could suggest that these models capture different aspects of the metabolic profile relevant to the prediction task. This discrepancy may be attributed to the distinct architectures of the models; MetaboNet, a more complex model, can capture non-linear relationships, whereas LR is constrained to linear associations. Additionally, the high dimensionality and correlation within metabolomics data, as demonstrated in Figure 5, suggest that multiple combinations of top metabolites could produce effective predictive results. The lack of validation from the current body of research could partly be attributed to the limited number of studies on the association between depression and metabolites. Our findings underscore the need for further research in this area.

Top MetaboNet Sub-pathways Validated by Literature

The outer ring in Figure 4 illustrates the relative importance of biochemical sub-pathways in MetaboNet. The top four sub-pathways are *food component/plant* (6%), *leucine, isoleucine and valine metabolism* (6%), *benzoate metabolism* (4%), and *fatty acid dicarboxylate* (3%).

Van der Spek et al. [38] conducted a large-scale metabolome-wide association study and discovered that cir-

culating metabolites modulated by food are linked to depression. They identified metabolites either directly derived from food or produced through host and gut microbial metabolism of food-derived products as particularly significant [38]. This finding is consistent with our results, which highlight the *food component/plant* pathway as the most important sub-pathway. Growing evidence indicates that nutrition influences mood through the modulation of gut microbiota [39, 40]. Specifically, the consumption of red and/or processed meat, refined grains, sweets, and high-fat products is associated with an increased risk of depression [41], whereas a high-quality diet rich in antioxidants, whole grains, fish, and fresh fruits and vegetables is linked to improved gut health [40] and a decreased risk of depression. These results suggest potential targets for diet-based therapies aimed at treating depression. This interpretation underscores the importance of dietary choices in mental health management and highlights the potential of dietary interventions as a non-pharmacological approach to depression management. However, it is important to note that while these findings are promising, further research is needed to fully understand the mechanisms involved and to develop effective dietary guidelines for depression management.

Leucine, *isoleucine*, and *valine* are branched-chain amino acids that were found to be significantly associated with depression [42]. Baranyi et al. [42] study found that these branched-chain amino acids are significantly decreased in patients with major depression compared to healthy controls. This reduction may lead to lower activation of the mammalian target of rapamycin (mTOR) pathway, which is crucial for cell growth and energy metabolism [42]. Dysregulation of mTOR is associated with depressive symptoms, suggesting that branched-chain amino acid levels could serve as biomarkers for depression and potential targets for novel antidepressant therapies [42]. Furthermore, Whipp et al. [43] found a significant negative association between *valine* and depression, while *leucine* also showed a significant negative association. No significant association was found for *isoleucine*. These findings imply that lower levels of certain BCAAs may be linked to higher rates of depression, highlighting the need for further research into their role in depressive disorders. However, it is important to note that the data from Whipp et al. [43] was collected during adolescence and young adulthood, which might limit its relevance to our study, as our dataset does not focus on a specific age range. A limitation of our work is that age was not included as a covariate, and there appears to be a difference between juvenile-onset and adult-onset depression [44]. Therefore, while our findings underscore the potential of BCAAs as biomarkers for depression, they also indicate the necessity for age-specific analyses in future research to fully understand their role across different age groups.

The *benzoate* pathway exhibits potential therapeutic effects on depression, as demonstrated by Cheng et al. [45]. Their findings indicate that benzoate drugs may reduce brain inflammation [45]. Both in vitro and in vivo studies reveal that benzoate achieves its anti-inflammatory effects by inhibiting microglial activity [45]. Our research corroborates the significance of *benzoate metabolism* in the context of depression, highlighting the promise of benzoate-based drugs as a treatment option. Nevertheless, further research is essential to elucidate the underlying mechanisms and to determine the long-term efficacy and safety of benzoate-based therapies.

Prince et al. [46] found that metabolites associated with *dicarboxylated fatty acids* were consistently linked with abnormal scores on a psychological self-report symptom scale for depression. *Fatty acid dicarboxylate* is a diet-derived metabolite [47], which further underscores the influence of dietary factors in depression [40]. This suggests that dietary interventions targeting specific fatty acids could potentially modulate depressive symptoms, offering a novel approach for prevention and treatment. However, the precise mechanisms through which these metabolites affect mood regulation remain unclear, and further research is needed to elucidate these pathways.

Top MetaboNet Super-pathway Validated by Literature

The inner ring of Figure 4 illustrates the relative importance of biochemical super-pathways within MetaboNet. The analysis indicates that the lipid super-pathway is the most prominent, accounting for 35%.

A comprehensive meta-analysis conducted by Bot et al. [48] reveals a distinct profile of circulating lipid metabolites linked to depression. Furthermore, lipids have been suggested as biomarkers for depression by Parekh et al. [49]. These findings underscore the significant role that specific lipid metabolites may play in the pathophysiology of depression, thereby validating the importance of the lipid super-pathway in our MetaboNet model. Given that lipids constitute the largest super-pathway group in our dataset, comprising 39% of the total, their prominence in our predictive model was expected. Although these results are promising, specific recommendations should be approached with caution due to the complexity and extensive nature of lipid metabolism. Nonetheless, this insight points to a valuable direction for future research.

Overall, our findings provide evidence that MetaboNet is capable of accurately identifying biologically important biochemical sub-pathways and super-pathways. However, we did not find evidence supporting the accuracy of MetaboNet in identifying specific biologically relevant metabolites for depression. Similarly, LR also failed to identify metabolites that have been highlighted in previous studies. However, it is important to note that the body of research investigating the association between metabolomics and depression is limited. The most compelling evidence we have found for a single pathway, based on our relative importance and existing literature, is related to food components. However, experimental validation is necessary to confirm their biological relevance and clinical significance. A notable limitation of our study is the unbalanced, correlated, and high-dimensional nature of the dataset. This characteristic may impact the robustness and generalizability of our findings. Further research with more balanced and comprehensive datasets is necessary to validate and expand upon these results.

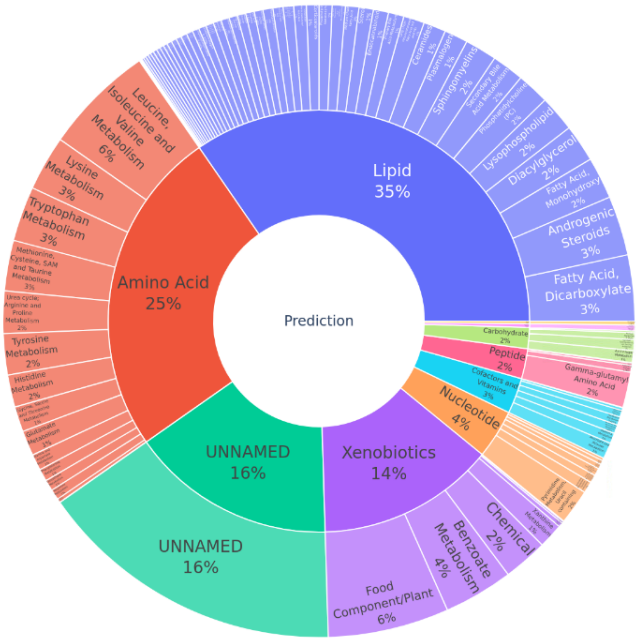

Figure 4: MetaboNet biochemical pathways relative importance sunburst. The relative importance of each pathway was obtained from the learned weights of the MetaboNet models. The outer ring is sub-pathways and the inner ring is super-pathways. The specific metabolites are omitted for clarity. Lipids, with 35%, can be seen as having the highest relative importance on the super-pathway level.

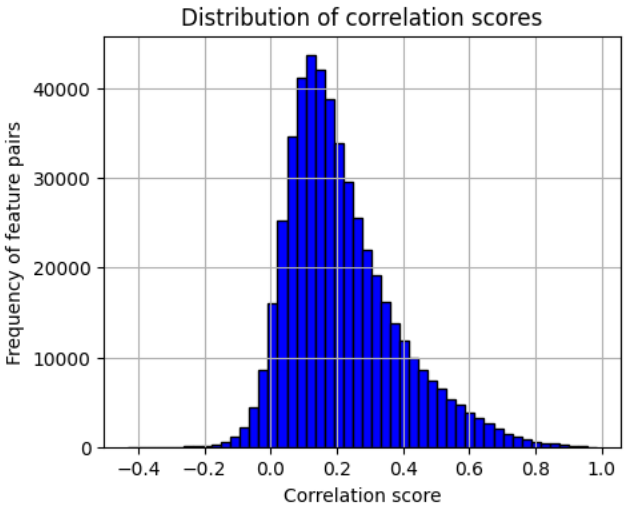

Figure 5: Display of the distribution of correlation scores between the metabolites. The x-axis represents the correlation scores ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation), while the y-axis shows the number of feature pairs that fall within each correlation score bin. Overall, we can see that the metabolites display a lot of positive correlations.

4 Responsible Research

Ensuring responsible research practices is integral to the scientific process, promoting both ethical standards and reproducibility. This section provides a comprehensive reflection on the ethical considerations and reproducibility challenges encountered during this research, along with the measures taken to address them.

Code availability: To ensure transparency and facilitate reproducibility, all code used in this research is publicly available on GitHub [28]. By making our code accessible, we provide a detailed view of the methods employed, allowing other researchers to verify and build upon our work.

Data availability: Although the dataset used in this study is not publicly available due to privacy regulations and informed consent constraints, it can be requested from Erasmus MC. Requests are subject to approval based on a predefined protocol to protect participant privacy.

Data handling procedures: To ensure proper handling of the dataset, the author read, signed, and adhered to the Rotterdam Study Data User Agreement. By following these guidelines, the author ensured that all data handling procedures were conducted in accordance with ethical research standards and privacy regulations. The dataset was used exclusively for the agreed-upon research purposes. Additionally, the author followed established security protocols to ensure the data's integrity and confidentiality.

Addressing biases: The dataset used in this research presents certain biases that need to be acknowledged. Firstly, there is a class imbalance, with a minority class of individuals diagnosed with depression. Secondly, the data was collected from a single geographical location, which could limit the applicability of the findings to other populations. Lastly, the dataset comprises a higher proportion of women, which restricts its generalizability to men. However, addressing these biases comprehensively was beyond the scope of this research.

Reproducibility: Ensuring the reproducibility of research findings is a fundamental aspect of responsible research. To this end, the methodology was extensively described and we utilized random seeds for all non-deterministic components of our algorithms and documented the hardware and tooling used.

Competing interests: There are no commercial or financial conflicts of interest to declare in this research.

Reliability of results: Due to the high variance in model performances, we accounted for the effects of random seed selection by running our algorithms with a large number of different random seeds. This extensive testing ensures that our results are not unduly influenced by any particular seed choice. By using a broad range of seeds, we minimized the potential bias that could arise from selecting a specific subset of seeds that might favor our proposed model over the baseline. This approach enhances the reliability and validity of our findings.

5 Conclusion and Future Work

In this study, we assessed the performance of MetaboNet and logistic regression (LR) in classifying individuals with depression based on metabolomics data. Both models demonstrated modest performance; however, MetaboNet showed a promising trend of outperforming LR across several key metrics. Notably, the classification results revealed sex-related

differences, with better performance observed in females.

We also aimed to validate whether the most important metabolites identified by LR and the most important metabolites, sub-pathways, and super-pathway identified by MetaboNet align with existing literature regarding their biological relevance to depression. Our findings did not support the capability of either model to identify biologically relevant individual metabolites. Nonetheless, MetaboNet exhibited promise in identifying sub- and super-pathways that are relevant to depression as documented in existing literature.

This research underscores the necessity for developing unbiased, diverse, and large datasets for studying depression. While MetaboNet can suggest important factors in predicting depression, experimental validation is necessary to confirm their biological relevance and clinical significance.

References

[1] E. Barberis, S. Khoso, A. Sica, M. Falasca, A. Gennari, F. Dondero, A. Afantitis, and M. Manfredi, "Precision medicine approaches with metabolomics and artificial intelligence," *International Journal of Molecular Sciences*, vol. 23, no. 19, 2022. [Online]. Available: https://www.mdpi.com/1422-0067/23/19/11269

[2] A. Galal, M. Talal, and A. Moustafa, "Applications of machine learning in metabolomics: Disease modeling and classification," *Frontiers in genetics*, vol. 13, p. 1017340, 2022.

[3] Y. Hasin, M. Seldin, and A. Lusis, "Multi-omics approaches to disease," *Genome biology*, vol. 18, no. 1, pp. 1–15, May 2017.

[4] B. B. Misra, C. Langefeld, M. Olivier, and L. A. Cox, "Integrated omics: tools, advances and future approaches," *Journal of molecular endocrinology*, vol. 62, no. 1, pp. R21–R45, 2019.

[5] E. A. Coler, W. Chen, A. V. Melnik, J. T. Morton, and A. A. Aksenov, "Metabolomics in the era of artificial intelligence," *Microbiota and Host*, vol. 2, no. 1, p. e230017, 2024. [Online]. Available: https://mah.bioscientifica.com/view/journals/mah/2/1/MAH-23-0017.xml

[6] N. J. Shah, S. Sureshkumar, and D. G. Shewade, "Metabolomics: a tool ahead for understanding molecular mechanisms of drugs and diseases," *Indian Journal of Clinical Biochemistry*, vol. 30, no. 3, pp. 247–254, 2015.

[7] A. V. Aderemi, A. O. Ayeleso, O. O. Oyedapo, and E. Mukwevho, "Metabolomics: A scoping review of its role as a tool for disease biomarker discovery in selected non-communicable diseases," *Metabolites*, vol. 11, no. 7, 2021. [Online]. Available: https://www.mdpi.com/2218-1989/11/7/418

[8] R. Perlis, "Translating biomarkers to clinical practice," *Molecular psychiatry*, vol. 16, no. 11, pp. 1076–1087, 2011. [Online]. Available: https://doi-org.tudelft.idm.oclc.org/10.1038/mp.2011.63

[9] Y. Wang, X. Cai, Y. Ma, Y. Yang, C.-W. Pan, X. Zhu, and C. Ke, "Metabolomics on depression: A comparison of clinical and animal research," *Journal of Affective Disorders*, vol. 349, pp. 559–568, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165032724000636

[10] J. Duan and P. Xie, "The potential for metabolomics in the study and treatment of major depressive disorder and related conditions," *Expert Review of Proteomics*, vol. 17, no. 4, pp. 309–322, 2020, pMID: 32516008. [Online]. Available: https://doi.org/10.1080/14789450.2020.1772059

[11] M. Liu, W. Ma, Y. He, Z. Sun, and J. Yang, "Recent progress in mass spectrometry-based metabolomics in major depressive disorder research," *Molecules*, vol. 28, no. 21, 2023. [Online]. Available: https://www.mdpi.com/1420-3049/28/21/7430

[12] X.-j. Guo, P. Wu, X. Jia, Y.-m. Dong, C.-m. Zhao, N.-n. Chen, Z.-y. Zhang, Y.-t. Miao, K.-m. Yun, C.-r. Gao *et al.*, "Mapping the structure of depression biomarker research: A bibliometric analysis," *Frontiers in Psychiatry*, vol. 13, p. 943996, 2022.

[13] L. N. G. Costa, B. A. Carneiro, G. S. Alves, D. H. L. Silva, D. F. Guimaraes, L. S. Souza, I. D. Bandeira, G. Beanes, A. M. Scippa, L. C. Quarantini *et al.*, "Metabolomics of major depressive disorder: A systematic review of clinical studies," *Cureus*, vol. 14, no. 3, 2022.

[14] World Health Organization, "Depression and other common mental disorders: Global health estimates," Geneva, 2017, licence: CC BY-NC-SA 3.0 IGO. [Online]. Available: https://iris.who.int/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf

[15] H. D. Schmidt, R. C. Shelton, and R. S. Duman, "Functional biomarkers of depression: diagnosis, treatment, and pathophysiology," *Neuropsychopharmacology*, vol. 36, no. 12, pp. 2375–2394, 2011.

[16] S. M. Bentley, G. L. Pagalilauan, and S. A. Simpson, "Major depression," *Medical Clinics of North America*, vol. 98, no. 5, pp. 981–1005, 2014, psychiatric Diagnosis and Management in Primary Care. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0025712514001011

[17] N. Kawamura, K. Shinoda, H. Sato, K. Sasaki, M. Suzuki, K. Yamaki, T. Fujimori, H. Yamamoto, D. Osei-Hyiaman, and Y. Ohashi, "Plasma metabolome analysis of patients with major depressive disorder," *Psychiatry and clinical neurosciences*, vol. 72, no. 5, pp. 349–361, 2018.

[18] C. R. Brydges, S. Bhattacharyya, S. M. Dehkordi, Y. Milaneschi, B. Penninx, R. Jansen, B. S. Kristal, X. Han, M. Arnold, G. Kastenmüller, M. Bekhbat, H. S. Mayberg, W. E. Craighead, A. J. Rush, O. Fiehn, B. W. Dunlop, and R. Kaddurah-Daouk, "Metabolomic and inflammatory signatures of symptom dimensions in major depression," *Brain, Behavior, and*

Immunity, vol. 102, pp. 42–52, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0889159122000290

[19] H. U. Zacharias, J. Hertel, H. Johar, M. Pietzner, K. Lukaschek, S. Atasoy, S. Kunze, H. Völzke, M. Nauck, N. Friedrich *et al.*, "A metabolome-wide association study in the general population reveals decreased levels of serum laurylcarnitine in people with depression," *Molecular psychiatry*, vol. 26, no. 12, pp. 7372–7383, 2021.

[20] Y. Xu, L. Cao, Y. Chen, Z. Zhang, W. Liu, H. Li, C. Ding, J. Pu, K. Qian, and W. Xu, "Integrating machine learning in metabolomics: A path to enhanced diagnostics and data interpretation," *Small Methods*, p. 2400305, 2024.

[21] P. Sen, S. Lamichhane, V. B. Mathema, A. McGlinchey, A. M. Dickens, S. Khoomrung, and M. Orešič, "Deep learning meets metabolomics: a methodological perspective," *Briefings in Bioinformatics*, vol. 22, no. 2, pp. 1531–1542, 09 2020. [Online]. Available: https://doi.org/10.1093/bib/bbaa204

[22] A. van Hilten, S. A. Kushner, M. Kayser, M. A. Ikram, H. H. Adams, C. C. Klaver, W. J. Niessen, and G. V. Roshchupkin, "Gennet framework: interpretable deep learning for predicting phenotypes from genetic data," *Communications biology*, vol. 4, no. 1, p. 1094, 2021.

[23] M. A. Ikram, B. C. Kieboom, W. P. Brouwer, G. Brusselle, L. Chaker, M. Ghanbari, A. Goedegebure, M. K. Ikram, M. Kavousi, R. J. de Knegt *et al.*, "The rotterdam study. design update and major findings between 2020 and 2024," *European Journal of Epidemiology*, pp. 1–24, 2024.

[24] L. S. Radloff, "The ces-d scale: A self-report depression scale for research in the general population," *Applied psychological measurement*, vol. 1, no. 3, pp. 385–401, 1977.

[25] L. Radloff, "Scale: A self-report depression scale for research in the general population." *J Clin Exp Neuropsychol*, vol. 19, pp. 340–356, 1997.

[26] P. M. Lewinsohn, J. R. Seeley, R. E. Roberts, and N. B. Allen, "Center for epidemiologic studies depression scale (ces-d) as a screening instrument for depression among community-residing older adults." *Psychology and aging*, vol. 12, no. 2, p. 277, 1997.

[27] S. Ahmed, "Metabolomics data preprocessing in the rotterdam study," 2024, unpublished.

[28] T. Kitak, "Metabonet," https://github.com/tom-kitak/MetaboNet, 2024.

[29] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.

[30] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, 2019, pp. 8024–8035.

[31] D. Chicco and G. Jurman, "The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation," *BMC genomics*, vol. 21, pp. 1–13, 2020.

[32] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[33] L. A. Jeni, J. F. Cohn, and F. De La Torre, "Facing imbalanced data–recommendations for the use of performance metrics," in *2013 Humaine association conference on affective computing and intelligent interaction*. IEEE, 2013, pp. 245–251.

[34] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets," *PloS one*, vol. 10, no. 3, p. e0118432, 2015.

[35] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright *et al.*, "Scipy 1.0: fundamental algorithms for scientific computing in python," *Nature methods*, vol. 17, no. 3, pp. 261–272, 2020.

[36] K. Mittelstrass, J. S. Ried, Z. Yu, J. Krumsiek, C. Gieger, C. Prehn, W. Roemisch-Margl, A. Polonikov, A. Peters, F. J. Theis *et al.*, "Discovery of sexual dimorphisms in metabolic and genetic biomarkers," *PLoS genetics*, vol. 7, no. 8, p. e1002215, 2011.

[37] D. Otten, A. N. Tibubos, G. Schomerus, E. Brähler, H. Binder, J. Kruse, K.-H. Ladwig, P. S. Wild, H. J. Grabe, and M. E. Beutel, "Similarities and differences of mental health in women and men: a systematic review of findings in three large german cohorts," *Frontiers in Public Health*, vol. 9, p. 553071, 2021.

[38] A. van der Spek, I. D. Stewart, B. Kühnel, M. Pietzner, T. Alshehri, F. Gauß, P. G. Hysi, S. MahmoudianDehkordi, A. Heinken, A. I. Luik *et al.*, "Circulating metabolites modulated by diet are associated with depression," *Molecular Psychiatry*, vol. 28, no. 9, pp. 3874–3887, 2023.

[39] J. Firth, J. E. Gangwisch, A. Borsini, R. E. Wootton, and E. A. Mayer, "Food and mood: how do diet and nutrition affect mental wellbeing?" *bmj*, vol. 369, 2020.

[40] T. S. Ghosh, S. Rampelli, I. B. Jeffery, A. Santoro, M. Neto, M. Capri, E. Giampieri, A. Jennings, M. Candela, S. Turroni *et al.*, "Mediterranean diet intervention alters the gut microbiome in older people reducing frailty and improving health status: the nu-age 1-year dietary intervention across five european countries," *Gut*, vol. 69, no. 7, pp. 1218–1228, 2020.

[41] Y. Li, M.-R. Lv, Y.-J. Wei, L. Sun, J.-X. Zhang, H.-G. Zhang, and B. Li, "Dietary patterns and depression risk: a meta-analysis," *Psychiatry research*, vol. 253, pp. 373–382, 2017.

[42] A. Baranyi, O. Amouzadeh-Ghadikolai, D. von Lewinski, H.-B. Rothenhäusler, S. Theokas, C. Robier, H. Mangge, G. Reicht, P. Hlade, and A. Meinitzer, "Branched-chain amino acids as new biomarkers of major depression-a novel neurobiology of mood disorder," *PloS one*, vol. 11, no. 8, p. e0160542, 2016.

[43] A. M. Whipp, M. Heinonen-Guzejev, K. H. Pietiläinen, I. van Kamp, and J. Kaprio, "Branched-chain amino acids linked to depression in young adults," *Frontiers in Neuroscience*, vol. 16, p. 935858, 2022.

[44] S. R. Jaffee, T. E. Moffitt, A. Caspi, E. Fombonne, R. Poulton, and J. Martin, "Differences in Early Childhood Risk Factors for Juvenile-Onset and Adult-Onset Depression," *Archives of General Psychiatry*, vol. 59, no. 3, pp. 215–222, 03 2002. [Online]. Available: https://doi.org/10.1001/archpsyc.59.3.215

[45] Y.-J. Cheng, C.-H. Lin, and H.-Y. Lane, "Ketamine, benzoate, and sarcosine for treating depression," *Neuropharmacology*, vol. 223, p. 109351, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0028390822004105

[46] N. Prince, M. Stav, M. Cote, S. H. Chu, C. M. Vyas, O. I. Okereke, N. Palacios, A. A. Litonjua, P. Vokonas, D. Sparrow, A. Spiro, J. A. Lasky-Su, and R. S. Kelly, "Metabolomics and self-reported depression, anxiety, and phobic symptoms in the va normative aging study," *Metabolites*, vol. 13, no. 7, 2023. [Online]. Available: https://www.mdpi.com/2218-1989/13/7/851

[47] H. M. Melo, L. E. Santos, and S. T. Ferreira, "Diet-derived fatty acids, brain inflammation, and mental health," *Frontiers in neuroscience*, vol. 13, p. 439762, 2019.

[48] M. Bot, Y. Milaneschi, T. Al-Shehri, N. Amin, S. Garmaeva, G. L. Onderwater, R. Pool, C. S. Thesing, L. S. Vijfhuizen, N. Vogelzangs, I. C. Arts, A. Demirkan, C. van Duijn, M. van Greevenbroek, C. J. van der Kallen, S. Köhler, L. Ligthart, A. M. van den Maagdenberg, D. O. Mook-Kanamori, R. de Mutsert, H. Tiemeier, M. T. Schram, C. D. Stehouwer, G. M. Terwindt, K. Willems van Dijk, J. Fu, A. Zhernakova, M. Beekman, P. E. Slagboom, D. I. Boomsma, B. W. Penninx, M. Beekman, H. Suchiman, J. Deelen, N. Amin, J. Beulens, J. van der Bom, N. Bomer, A. Demirkan, J. van Hilten, J. Meessen, R. Pool, M. Moed, J. Fu, G. Onderwater, F. Rutters, C. So-Osman, W. van der Flier, A. van der Heijden, A. van der Spek, F. Asselbergs, E. Boersma, P. Elders, J. Geleijnse, M. Ikram, M. Kloppenburg, I. Meulenbelt, S. Mooijaart, R. Nelissen, M. Netea, B. Penninx, C. Stehouwer, C. Teunissen, G. Terwindt, L. 't Hart, A. van den Maagdenberg, P. van der Harst, I. van der Horst, C. van der Kallen, M. van Greevenbroek, W. van Spil, C. Wijmenga, A. Zwinderman, A. Zhernikova, J. Jukema, and N. Sattar, "Metabolomics profile in depression: A pooled analysis of 230 metabolic markers in 5283 cases with depression and 10,145 controls," *Biological Psychiatry*, vol. 87, no. 5, pp. 409–418, 2020, mechanisms of Major Depression. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0006322319316282

[49] A. Parekh, D. Smeeth, Y. Milner, and S. Thuret, "The role of lipid biomarkers in major depression," in *Healthcare*, vol. 5, no. 1. MDPI, 2017, p. 5. [Online]. Available: https://www.mdpi.com/2227-9032/5/1/5