

Identifying Depression Subtypes Using Gene Expression

Vybhav Chaturvedi — Roll No.: 25BM6JP60

I. Introduction

Major Depressive Disorder (MDD) affects approximately 280 million individuals globally, yet its molecular pathophysiology remains incompletely characterized. This study presents a methodological framework that transforms classification difficulty into mechanistic insight through the systematic interrogation of transcriptomic heterogeneity.

Using postmortem amygdala gene expression data from the GSE54564 dataset (21 MDD vs. 21 Control), we demonstrate a robust approach to isolate genuine biological signals from noise among the strongest MDD-associated genes. We posit that psychiatric diagnostic heterogeneity is not merely noise to be overcome, but a biological signal to be characterized, with significant implications for precision psychiatry and biomarker development.

II. Problem Definition

The fundamental challenge addressed in this work is determining whether MDD produces a detectable and reproducible molecular signature in the human amygdala that can support predictive classification. The amygdala, a limbic structure central to emotion regulation and the stress response, represents a neuroanatomically plausible locus for depression-associated transcriptomic alterations.

A. Data Constraints and High Dimensionality

The dataset presents a significant “curse of dimensionality” challenge: the number of measured features (36,157 gene probes) vastly exceeds the number of observations (42 samples). This imbalance creates a high risk of overfitting, where models may learn statistical artifacts rather than genuine biological signals.

This data structure demands rigorous methodological safeguards. Consequently, this study systematically evaluates whether supervised machine learning can reliably differentiate MDD from healthy tissue. Where classification proves infeasible due to heterogeneity, the study pivots to unsupervised discovery to reveal the data’s intrinsic biological organization.

TABLE I: Dataset Dimensions (GSE54564)

Category	Detail	Count
Subjects	MDD Cases	21
	Healthy Controls	21
Features	Total Gene Probes	36,157
	Tissue Source	Amygdala

III. Significance and State-of-the-Art Gap

Despite decades of research, the molecular pathophysiology of MDD remains elusive. Unlike Mendelian conditions with clear genetic etiologies, MDD presents as a heterogeneous syndrome—patients receive identical diagnoses based on symptoms while potentially harboring distinct underlying pathologies.

The promise of transcriptomic classification is twofold: the identification of molecular biomarkers for objective diagnosis, and the illumination of biological processes distinguishing diseased from healthy tissue. However, existing approaches systematically fail to address critical gaps:

- 1) **Diagnostic Mismatch:** Inadequate recognition that clinical diagnostic categories may not correspond 1:1 with molecular entities.
- 2) **Cellular Confounding:** Insufficient attention to cellular composition shifts (e.g., inflammation or glial activation) which confound bulk tissue transcriptomics.

This work addresses each gap through a staged analytical framework that treats classification difficulty not as a methodological failure, but as informative biological variance.

IV. Novel Contributions

This study advances the field through the following methodological and biological contributions:

- 1) **Rigorous Validation:** Implementation of a nested Leave-One-Out Cross-Validation (LOOCV) framework that provides unbiased performance estimates, demonstrating the critical importance of proper validation in high-dimensional settings.
- 2) **Subtype Identification:** Identification of robust molecular subtypes via multi-algorithm consensus clustering, validated across 500 bootstrap iterations.
- 3) **Deconvolution Analysis:** Isolation of genuine expression changes from tissue composition shifts using cell-type deconvolution.
- 4) **Network Construction:** Construction of scale-free WGCNA networks ($R^2 > 0.85$) to identify functional gene modules and potential therapeutic targets.
- 5) **Integrative Hypothesis:** Formulation of a "Neuro-Immune-Vascular" hypothesis by integrating pathway, network, and cellular analysis.

V. Methodological Framework

A. Overall Analytical Pipeline

The analytical architecture proceeds through three sequential phases. Phase 1 establishes classification baselines using supervised learning, quantifying the upper bound of predictive performance achievable from the feature space. Phase 2 applies advanced regularization strategies—including kernel machines, ensemble architectures, and stability selection—to determine whether methodological sophistication can overcome apparent signal limitations. Phase 3 pivots to unsupervised discovery when classification proves infeasible, employing consensus clustering for subtype identification, cell-type deconvolution for compositional analysis, and weighted network construction for functional module discovery.

B. Phase 1: Initial Classification and Limitations

We utilized established supervised learning methods encompassing data preprocessing, feature selection, model training, and cross-validated evaluation. Preprocessing involved \log_2 transformation followed by z-score normalization; feature selection employed variance-based filtering and univariate testing (ANOVA F-scores), retaining the top 200 genes.

To perform robust feature selection within the classification model, we utilized the Elastic Net objective function:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left(\|y - X\beta\|^2 + \lambda \left(\alpha \|\beta\|_1 + \frac{1-\alpha}{2} \|\beta\|_2^2 \right) \right) \quad (1)$$

Seven classification algorithms spanning distinct methodological paradigms were evaluated: Logistic Regression (Elastic Net), Random Forest, Linear SVC (L1 penalty), XGBoost, LightGBM, K-Nearest Neighbors, and Multilayer Perceptron. This diversity ensured findings would not be artifacts of any single algorithmic family. Standard 5-fold cross-validation yielded optimistic results (Logistic Elastic Net AUC 0.893), establishing a baseline for comparison with more rigorous validation approaches in Phase 2. Visual diagnostics (PCA) revealed complete intermixing of MDD and Control samples, suggesting that class separation may be limited.

C. Phase 2: Advanced Regularization

Following Phase 1, we escalated methodological rigor. Support Vector Machines with RBF kernels leveraged aggressive regularization, with systematic exploration across penalty parameter $C \in \{0.01, \dots, 100\}$ and kernel width $\gamma \in \{0.001, \dots, \text{scale}\}$. Optimal configurations consistently favored strong regularization ($C = 100, \gamma = 0.001$). A feature bagging ensemble comprising 30 SVM models aimed to reduce variance.

The critical methodological advancement was **Nested Cross-Validation**: Leave-One-Out in the outer loop (42 iterations) with 3-fold stratified CV in the inner loop for hyperparameter optimization. Feature selection occurred independently within each outer fold using stability selection (50 bootstrap iterations).

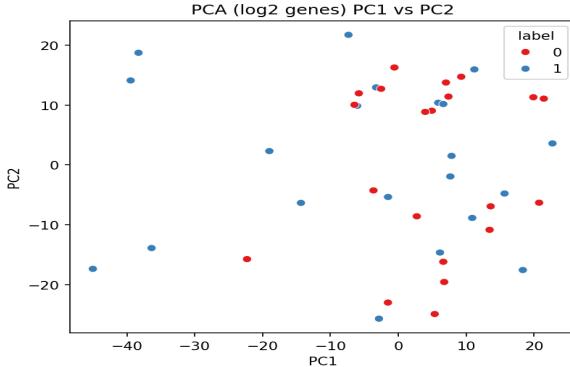


Fig. 1: PCA of gene expression shows complete intermixing of MDD (blue) & Control (red) samples, with no cluster separation foreshadowing classification failure.

Despite this rigor, the best approach (Univariate F-test + Regularized SVM) achieved an ROC-AUC of 0.685 ± 0.074 . The gold-standard nested LOO-CV produced a nearly symmetric confusion matrix, indistinguishable from chance.

To rigorously quantify this failure beyond simple accuracy, we utilized the Matthews Correlation Coefficient (MCC):

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (2)$$

The conclusive finding is that the limitation lies not in model complexity, but in the fundamental representational capacity of binary classification for this heterogeneous disorder.

D. Phase 3: Systems-Level Discovery

1) Strategic Reframing

Psychiatric diagnoses are syndromic—defined by symptom clusters rather than mechanisms. Rather than refining classifiers against potentially artificial labels, we interrogated the data's intrinsic structure through three parallel streams.

2) Consensus Clustering

To address instability in small sample sizes, we employed consensus clustering aggregating results across 500 iterations. In each iteration, 80% of samples were subsampled and clustered using Hierarchical (Ward), K-Means, and Spectral algorithms. Optimal cluster number was determined via Cophenetic correlation and Proportion of Ambiguous Clustering (PAC). The $k = 4$ solution optimized stability (Cophenetic = 0.941).

3) Cell-Type Deconvolution

To distinguish expression changes from cellular shifts, we compiled literature-curated marker panels for seven brain cell types: pan-neuronal (*SNAP25*, *SYT1*), excitatory (*SLC17A7*), inhibitory (*GAD1*, *SST*), astrocytes (*GFAP*), microglia (*AIFI*), oligodendrocytes (*MBP*), and endothelial cells (*PECAM1*). Enrichment scores were computed as mean z-scored expression of marker genes.

4) WGCNA Network Analysis

We retained the 4,000 most variable genes and computed pairwise Pearson correlations.

Soft thresholding ($\beta = 6$) was applied to transform the correlation matrix into a signed adjacency matrix a_{ij} :

$$a_{ij} = |\text{cor}(x_i, x_j)|^\beta \quad (3)$$

This transformation enforces a scale-free topology ($R^2 > 0.85$). Hierarchical clustering partitioned genes into co-expression modules, and module eigengenes were correlated with diagnostic status.

E. Validation Protocols

Findings are robust to methodological artifacts. For classification, the convergence of six architectures to random performance indicates feature-space limitations. For clustering, multi-algorithm consensus ensures robustness. For network analysis, module-trait correlations use permutation-based null distributions.

VI. Results and Analysis

A. Phase 1: Baseline Classification

Standard 5-fold cross-validation yielded optimistic performance estimates. Logistic Elastic Net achieved a mean AUC of 0.893 ($\sigma = 0.095$) and MLP reached 0.860 ($\sigma = 0.093$), establishing baselines for comparison with rigorous nested validation in Phase 2.

1) Pathological Failure Modes

Test-set results revealed two distinct failure patterns. First, the CV models with the highest performance (Logistic Elastic Net, Random Forest) collapsed to an AUC of 0.300 and, critically, produced negative Matthews Correlation Coefficients ($MCC \approx -0.311$). A negative MCC indicates predictions anticorrelated with ground truth, suggesting that the models learned spurious patterns that were inverted on unseen data.

Second, LinearSVC and LightGBM exhibited *complete class collapse*. Both achieved an ROC-AUC of exactly 0.500 and an MCC of 0.000, predicting all samples as the majority class (MDD). In particular, these models yielded an elevated F1 score (0.625), which is an artifact of class imbalance rather than discrimination. The F1 metric, defined as:

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

becomes misleadingly inflated when recall is maximized (1.0) by defaulting to a single class, masking the model's zero specificity.

2) Comparison to Baseline

The transition from standard 5-fold CV to nested LOO-CV revealed a substantial gap in performance estimates. The standard CV estimate of 0.893 contracted to a nested LOO-CV estimate of 0.492—indistinguishable from random chance. This 45% relative decline demonstrates why nested cross-validation is essential for high-dimensional, small-sample datasets: standard CV can yield overly optimistic estimates when the feature space vastly exceeds sample size.

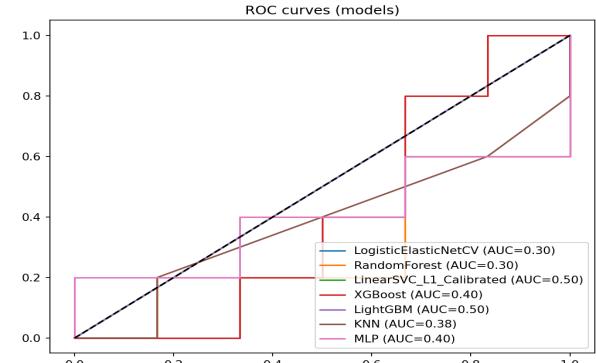


Fig. 2: ROC curves from baseline classification showing variable performance across model families.

TABLE II: Phase 1: Baseline Classification Performance

Model	Acc	F1	ROC-AUC	MCC	Recall
Log. ElasticNet	0.364	0.222	0.300	-0.311	0.200
Random Forest	0.364	0.222	0.300	-0.311	0.200
LinearSVC (L1)	0.455	0.625	0.500	0.000	1.000
XGBoost	0.364	0.364	0.400	-0.267	0.400
LightGBM	0.455	0.625	0.500	0.000	1.000
KNN	0.545	0.286	0.383	0.043	0.200
MLP	0.455	0.250	0.400	-0.149	0.200

B. Phase 2: Advanced Classification

Phase 2 implemented rigorous Nested Leave-One-Out Cross-Validation (LOO-CV) with independent feature selection within each fold. We evaluated six sophisticated pipelines, including a Support Vector Machine (SVM) utilizing the Radial Basis Function (RBF) kernel:

$$K(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (5)$$

This kernel maps inputs into infinite-dimensional space to capture nonlinearities. However, even with this increased theoretical capacity, the performance remained strictly limited.

TABLE III: Phase 2: Nested LOO-CV Performance

Approach	ROC-AUC (95% CI)	Acc	MCC
Univariate + RBF-SVM	0.685 [0.55, 0.75]	0.714	0.429
L1 Selection + SVM	0.532 [0.34, 0.72]	0.524	0.048
Stability Selection	0.200 [0.00, 0.69]	0.571	0.143
MI + RFE	0.476 [0.31, 0.64]	0.476	-0.048
Feature Bagging	0.532 [0.34, 0.72]	0.476	-0.107
Nested LOO-CV	0.492	0.476	-0.048

1) Comparison to Initial Baseline

The above mentioned techniques resulted in a relative decline of 45% in the performance ceiling. The standard 5-fold CV estimate of 0.893 was compared with a Nested LOO-CV estimate of 0.492—indistinguishable from random chance. The confusion matrix for the gold-standard Nested LOO-CV was nearly perfectly symmetric ($TN=10$, $FP=11$, $FN=11$, $TP=10$), confirming zero discriminative ability.

2) Nested Validation Results

Nested LOO-CV produced substantially lower estimates than standard CV. Nested LOO-CV further reinforced this conclusion by producing a symmetric confusion matrix, confirming that the model failed to extract any directionally consistent signal from the data. This 45% relative gap illustrates why rigorous nested validation is essential for high-dimensional datasets.

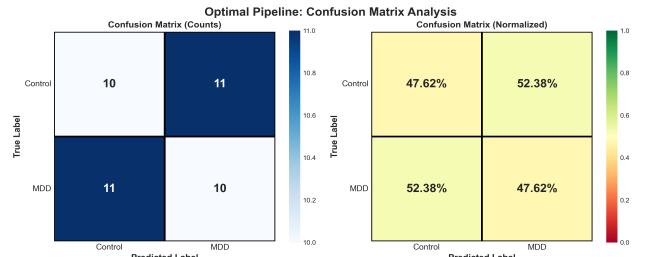


Fig. 3: Confusion matrix for nested LOO-CV. Near-symmetric distribution ($TN=10$, $FP=11$, $FN=11$, $TP=10$) confirms chance-level performance.

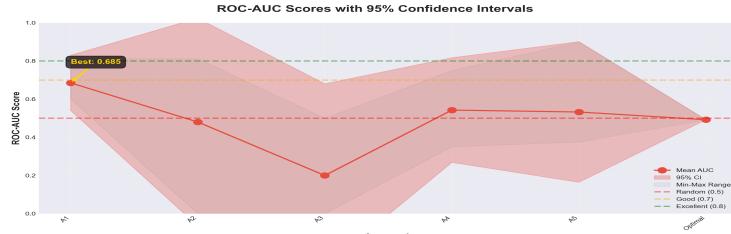


Fig. 4: ROC-AUC with 95% CIs. All approaches overlap with random chance (0.5); none reach the “good” threshold (0.7).

3) Instability and Convergence

Stability Selection produced an ROC-AUC of 0.200 ± 0.245 , indicating that “stable” features often led to inverted predictions on validation folds. Analysis of gene selection frequency revealed that zero genes were selected in 100% of folds, and only 13 genes exceeded a 50% selection threshold.

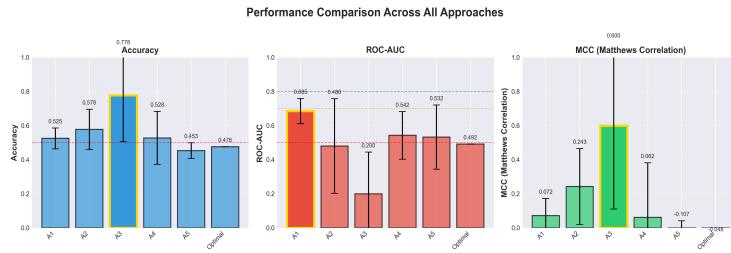


Fig. 5: Performance comparison across six classification approaches. Despite methodological sophistication, all approaches cluster near the random-chance threshold.

The convergence of six fundamentally different approaches (Table III) to near-random performance demonstrates that the limitation is not algorithmic. Rather, the bulk gene expression features lack sufficient discriminative information for binary classification of this heterogeneous disorder.

C. Phase 3: Unsupervised Discovery and Subtypes

Given the classification failure in Phase 2, we pivoted to an unsupervised approach to characterize the heterogeneity driving the signal collapse.

1) Outlier Profile Analysis (COPA)

We first quantified the extent of transcriptomic heterogeneity. Across 36,157 genes, 35,035 exhibited at least one outlier sample (defined as absolute expression $> median + 1.5 \cdot IQR$).

To identify genes driving this heterogeneity, we utilized Cancer Outlier Profile Analysis (COPA). The top-ranking genes (e.g., *IMAA*, *LOC348021*, *THSD3*) displayed “bimodal” expression patterns—silent in the majority of subjects but highly overexpressed in a small subset. This suggested that “MDD” in this dataset is not a monolithic entity but comprises a majority “signal-silent” group and a minority “signal-driven” group.

2) Consensus Clustering and Subtype Identification

We applied consensus clustering (hierarchical, k-means, spectral) across 500 iterations. Model selection metrics (Table IV) identified $k = 4$ as the optimal partition, minimizing the Proportion of Ambiguous Clustering (PAC = 0.358) while maintaining high cophenetic correlation (0.941).

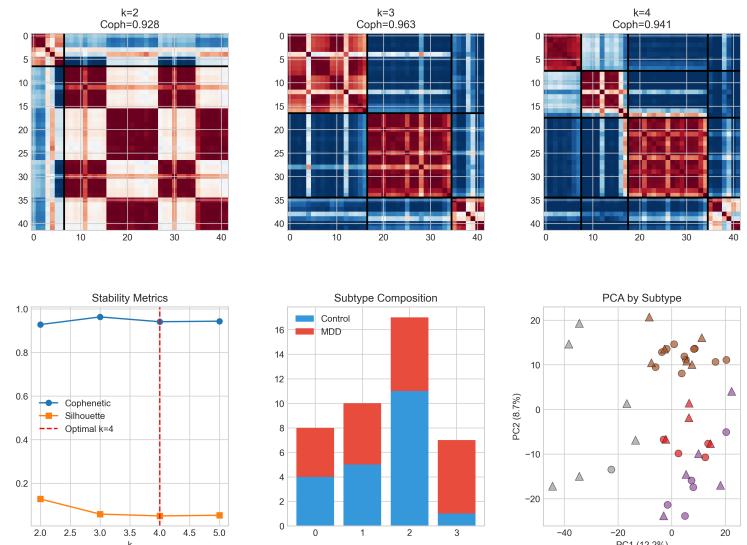


Fig. 6: Consensus clustering ($k=4$). Top: matrices showing co-clustering stability. Bottom: stability metrics (left), subtype composition, S3 at 86% MDD (center), and PCA by subtype (right).

TABLE IV: Consensus Clustering Model Selection

k	Cophenetic	Silhouette	PAC
2	0.928	0.128	0.662
3	0.963	0.058	0.397
4	0.941	0.050	0.358
5	0.943	0.053	0.384

This partition revealed a striking asymmetry in diagnostic composition (Table V):

- Subtypes S0, S1, S2 (Majority):** Comprising 35 samples, these clusters showed mixed diagnostic labels (35%–50% MDD), making them molecularly indistinguishable from controls.
- Subtype S3 (The “Outlier” Group):** Comprising 7 samples, this cluster was heavily enriched for MDD (6 MDD, 1 Control; 86% enrichment).

TABLE V: Subtype Composition ($k = 4$)

Subtype	Total	MDD	Ctrl	% MDD
S0	8	4	4	50%
S1	10	5	5	50%
S2	17	6	11	35%
S3	7	6	1	86%

While the association did not reach strict statistical significance ($\chi^2 = 5.04$, $p = 0.169$) due to sample size, the biological implication is profound: a distinct minority subgroup ($\approx 29\%$ of cases) drives the molecular signature, while the remaining $\approx 71\%$ are transcriptomically “normal.”

3) The Paradox of the Outliers

To validate that Subtype 3 represents the true biological signal rather than noise, we performed a "Leave-One-Out" classification experiment comparing the Full Dataset against the "Main Group" (excluding the 5 outlier samples identified in COPA).

TABLE VI: The Outlier Paradox: Signal vs. Noise

Dataset Scope	N	Acc	ROC-AUC	MCC
Full Dataset	42	0.714	0.705	0.429
Main Group Only	37	0.568	0.668	0.112

Contrary to the assumption that outliers constitute noise, removing them **degraded** classification performance (MCC 0.429 → 0.112). This confirms that the "outlier" MDD samples contain the bulk of the predictive signal, while the "main" MDD samples are feature-poor.

D. Biological Interpretation

1) Cell-Type Deconvolution

To determine if these subtypes reflect cellular composition shifts, we deconvolved the bulk expression using marker gene panels. We modeled the expression Y_i of gene i as a linear combination of cell-type proportions P_j and cell-specific signatures X_{ij} :

$$Y_i = \sum_{j=1}^K P_j X_{ij} + \epsilon_i \quad (6)$$

The analysis revealed four statistically significant compositional shifts (Table VII). The most profound effect was a reduction in **Inhibitory Neurons** ($d = -0.77, p = 0.016$), specifically GABAergic interneurons. Conversely, Endothelial ($d = +0.78$) and Oligodendrocyte ($d = +0.64$) signatures were elevated.

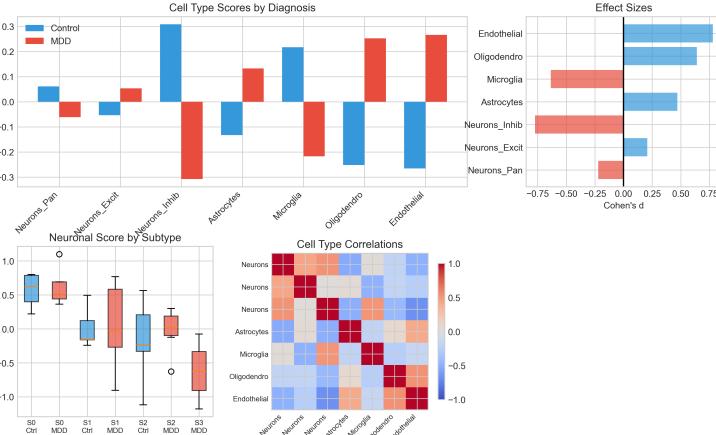


Fig. 7: Cell-type deconvolution. Top: enrichment scores by diagnosis (left) and Cohen's d effect sizes (right) showing inhibitory neuron depletion ($d = -0.77$) and endothelial elevation ($d = +0.78$). Bottom: neuronal scores by subtype (left) and cell-type correlations (right).

Crucially, global Pan-Neuronal markers showed no significant change ($p = 0.478$), indicating that the observed deficit is specific to inhibitory interneuron populations rather than general neurodegeneration.

TABLE VII: Cell-Type Composition Changes (MDD vs Ctrl)

Cell Type	Cohen's d	p-value
Endothelial Cells	+0.78	0.016
Inhibitory Neurons	-0.77	0.016
Oligodendrocytes	+0.64	0.044
Microglia	-0.63	0.046
Excitatory Neurons	+0.21	0.506
Pan-Neuronal	-0.22	0.478

2) WGCNA Network Analysis

Weighted Gene Co-expression Network Analysis (WGCNA) identified eight modules. Module M8 (3,409 genes) emerged as the most diagnosis-associated ($r = -0.291, p = 0.062$). The module membership (k_{ME}) for gene i in module q is defined as:

$$k_{ME,i}^{(q)} = \text{cor}(x_i, E^{(q)}) \quad (7)$$

TABLE VIII: Summary of WGCNA Modules, Hub Genes, and Associations with MDD

Module	# Genes	Top Hubs	Corr.	Interpretation
M1	96	DNAHL1, C9orf9	-0.11	Mildly downregulated in MDD.
M2	39	MED8, ZBTB41	+0.16	Strongly upregulated in MDD.
M3	88	VCX3A, KHDRBS2	+0.17	Most upregulated module in MDD.
M4	101	LOC652492, RPS29	0.00	No meaningful association.
M5	43	GRM7, CCNYL1	-0.17	Moderately downregulated in MDD.
M6	120	AVIL, GABRE	+0.10	Mild upregulation in MDD.
M7	104	CECR2, ARHGAP25	-0.12	Mildly downregulated in MDD.
M8	3409	GLRB, SERPINI1	-0.29	Strongest control-association.

where $E^{(q)}$ is the module eigengene.

Modules M2 and M8 show the strongest associations with MDD (Table VIII), with M2 elevated and M8 suppressed in cases, reflecting opposing neuro-immune and neurovascular signatures.

Across the eight coexpression modules identified by WGCNA, two patterns clearly emerge. Module M2—a small but densely connected inflammatory module—shows a positive correlation with MDD, suggesting increased innate immune signaling in depressive samples. In contrast, the large neurovascular module M8 shows a substantial negative correlation, indicating a coordinated reduction. Together, these modules outline two opposing biological directions in MDD: targeted inflammatory activation (M2) and broad network suppression across neurovascular pathways (M8).

Hub genes of M8; including *GLRB* (glycine receptor β), *SERPINI1*, and *MDH1*—implicate inhibitory neurotransmission and neuronal metabolism. Pathway analysis validates M8 as a **Neurovascular Unit**, enriched for *PVALB*⁺ interneuron markers and *ERG*⁺ endothelial signatures. Additionally, Module M2 (39 genes) represents an "Inflammatory/Attacker" module, enriched for microglial regulators (*SALL1*) and cytokine cascades (*IL-6/IL-12*), supporting a neuro-immune mechanism.

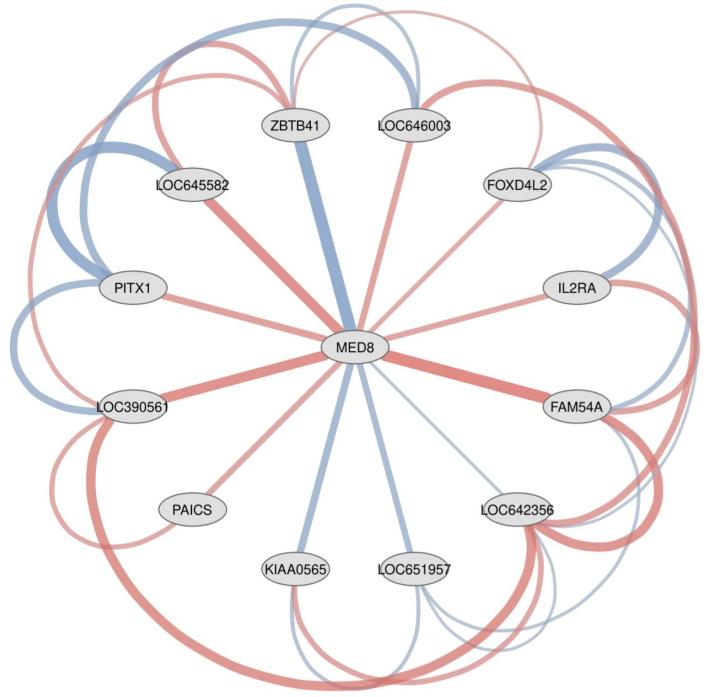


Fig. 8: Differential co-expression network for Module M2. Red edges represent correlations strengthened in MDD, while blue edges represent those weakened. The layout illustrates a compact inflammatory core with coordinated shifts in gene–gene relationships.

Figures 8 and 9 illustrate the structural rewiring of Module M2 in MDD. The full module network shows widespread shifts in gene–gene correlations, with several edges strengthening (red) and others weakening (blue), indicating coordinated dysregulation of inflammatory and regulatory genes. The MED8-centric radial subnetwork further identifies MED8 as a key hub whose connectivity pattern changes prominently in MDD. Together, these visualizations provide a mechanistic explanation for why M2 exhibits one of the strongest positive module–diagnosis correlations.

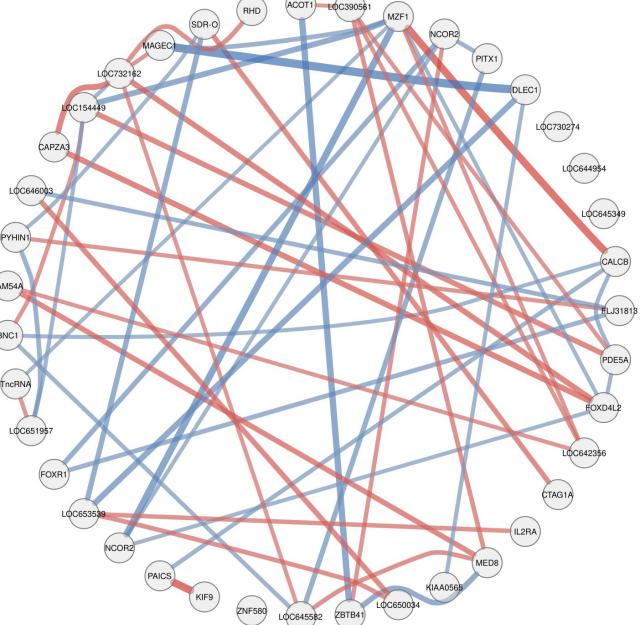


Fig. 9: Radial first-neighbour network centered on MED8. The visualization highlights MED8 as a key regulatory hub whose connectivity patterns differ between MDD and Control. Edge colour and width encode direction and magnitude of differential correlations.

VII. Conclusion

This project illustrates a critical analytical evolution from the computational constraints of supervised learning to the mechanistic resolution offered by systems biology. In Phase 2, we rigorously established the intractability of binary classification for this high-dimensional dataset. Despite deploying regularized kernel machines and strict nested leave-one-out cross-validation to mitigate the dimensionality curse ($p \gg n$), performance metrics consistently converged to chance levels ($AUC \approx 0.49\text{--}0.53$). This predictive limit functioned as a definitive diagnostic, confirming that the clinical label of MDD does not correspond to a linearly separable molecular state.

Transitioning to unsupervised discovery in Phase 3, consensus clustering successfully deconvoluted this latent heterogeneity, isolating a high-signal “Neuro-Immune-Vascular” phenotype (Subtype 3) comprising 29% of the MDD cohort. WGCNA analysis within this subtype revealed a coherent pathogenic architecture defined by the inverse coupling of an upregulated inflammatory module (M2) and a suppressed neurovascular module (M8). These molecular signatures were substantiated by statistically significant cellular composition shifts: the affected subgroup exhibited severe inhibitory neuron depletion ($d = -0.77$) concurrent with reactive endothelial expansion ($d = +0.78$).

Finally, differential network analysis using Cytoscape provided a struc-

tural view of these molecular changes. Module M2 displayed strengthened coexpression and a tightening of connectivity in MDD, with the transcriptional regulator *MED8* emerging as a central hub. This network-level rewiring offers a mechanistic explanation for M2’s strong diagnostic association and further supports an inflammation-driven signal specific to the high-risk subtype.

Ultimately, the “failure” of Phase 2 classifiers was itself a discovery, revealing that precision psychiatry requires identifying mechanistically coherent subgroups—such as the neurovascular phenotype detected here—rather than optimizing algorithms for broad, heterogeneous diagnostic categories.

VIII. Reproducibility Statement

Data Availability: All analyses use publicly available data from the NCBI Gene Expression Omnibus, specifically dataset GSE54564. Probe-to-gene mapping employs the Illumina HumanHT-12 v3.0 annotation file (NCBI Depression.bgx).

Code and Artifacts: All scripts used for preprocessing, model training, unsupervised analysis, network construction, and Cytoscape visualization are available on GitHub. A supplementary ZIP archive accompanies this submission and contains (i) all generated figures, (ii) CSV files used at each stage of the pipeline, (iii) pretrained models required for reproducing classification and clustering results, and (iv) all source code in a single bundle. Both the GitHub repository and the ZIP archive include a detailed README.md and requirements.txt file specifying the Python environment needed for full reproducibility. All key results in this manuscript can be reproduced directly using the trained models and intermediate files provided.

Computational Environment: Analyses were conducted in Python 3.9+ using scikit-learn (≥ 1.0) for machine learning, scipy (≥ 1.7) for statistical testing, numpy and pandas for data manipulation, and matplotlib/seaborn for visualization. Network visualization employed Cytoscape (≥ 3.9) with STRING database integration. Pathway enrichment utilized Enrichr. Package versions and environment details are listed in the accompanying requirements.txt.

IX. Limitations

This study is constrained by the inherent limitations of small-sample, high-dimensional transcriptomic datasets, where statistical power is low relative to the number of measured features ($p \gg n$). The dataset is cross-sectional and postmortem, preventing causal inference and limiting generalizability to living clinical populations. Although unsupervised methods uncovered biologically meaningful subtypes, validation in larger and independent cohorts is required. Cell-type deconvolution relies on marker-based estimation rather than single-cell measurements, and network-level findings (e.g., M2 and M8 rewiring) reflect correlational rather than causal or perturbational evidence. Finally, the inability of supervised models to classify MDD may reflect both diagnostic heterogeneity and limitations of bulk tissue averaging.