Open Access Library Journal
2026, Volume 13, e14447
ISSN Online: 2333-9721
ISSN Print: 2333-9705

Enhanced Multimodal Transformer for Treatment-Resistant Depression Prediction Using Synthetic fMRI, Genomic, and Clinical Data

Rocco de Filippis[1*], Abdullah Al Foysal[2]

[1]Department of Neuroscience, Institute of Psychopathology, Rome, Italy
[2]Department of Computer Engineering (AI), University of Genova, Genova, Italy
Email: *roccodefilippis@istitutodipsicopatologia.it, niloyhasanfoysal440@gmail.com

How to cite this paper: de Filippis, R. and Al Foysal, A. (2026) Enhanced Multimodal Transformer for Treatment-Resistant Depression Prediction Using Synthetic fMRI, Genomic, and Clinical Data. *Open Access Library Journal*, **13**: e14447.
https://doi.org/10.4236/oalib.1114447

Received: October 13, 2025
Accepted: January 12, 2026
Published: January 15, 2026

Abstract

Treatment-Resistant Depression (TRD) remains one of the most challenging subtypes of major depressive disorder, affecting approximately one-third of patients and leading to significant morbidity, healthcare costs, and reduced quality of life. Predicting TRD onset and progression is complex, as it requires integrating heterogeneous biomarkers spanning neuroimaging, genomics, and clinical history. This study presents an Enhanced Multimodal Transformer (EMT) designed to fuse functional magnetic resonance imaging (fMRI), single-nucleotide polymorphism (SNP) profiles, and structured clinical variables into a unified predictive framework. The architecture employs modality-specific encoders patch-based embeddings with positional encodings for fMRI, attention-weighted embeddings for SNP data, and normalized dense projections for clinical features-followed by the introduction of modality tokens to enable cross-modal information exchange within a Transformer encoder. To validate the architecture in a controlled setting, we generated a balanced, clinically inspired synthetic dataset with distinct activation patterns in brain regions (prefrontal cortex, amygdala, anterior cingulate), SNP distributions with predictive loci, and clinically relevant severity profiles. Model training achieved rapid convergence with early stopping at eight epochs. Evaluation demonstrated a perfect Area Under the ROC Curve (AUC = 1.00) and average precision of 1.00, indicating complete separation in probability space. However, accuracy at a fixed 0.5 decision threshold was limited (50%), reflecting probability compression into distinct but narrow ranges (≈0.32 for non-TRD, ≈0.41 for TRD). Feature analysis revealed that clinical severity, specific SNP clusters, and region-specific fMRI activations dominated predictive importance. These results

provide a proof-of-concept that Transformer-based multimodal fusion can capture complex, cross-domain patterns in TRD, supporting its potential for precision psychiatry. Future work will extend to real-world datasets and incorporate probability calibration to improve threshold-based classification performance in clinical settings.

Subject Areas

Computational Psychiatry, Artificial Intelligence

Keywords

Treatment-Resistant Depression, Multimodal Fusion, Transformer Architecture, fMRI Analysis, SNP Genomics, Clinical Features, Probability Calibration, Computational Psychiatry, Synthetic Data Generation, Machine Learning in Mental Health

1. Introduction

Treatment-Resistant Depression (TRD) is a severe and persistent form of major depressive disorder (MDD) that affects an estimated 20% - 30% of patients, even after multiple adequate trials of antidepressant medications [1]-[7]. Individuals with TRD often experience chronic symptoms, high relapse rates, functional impairment, and elevated suicide risk, making it a major public health and economic burden [8]-[13]. Accurate early prediction of TRD is crucial for guiding personalized treatment strategies, optimizing resource allocation, and reducing the risk of long-term disability [14]-[18]. The pathophysiology of TRD is multifactorial, involving complex interactions between neurobiological, genetic, and psychosocial factors [19]-[22]. Neuroimaging studies, particularly functional magnetic resonance imaging (fMRI), have revealed abnormal activation patterns in key brain regions such as the prefrontal cortex, amygdala, and anterior cingulate cortex [23]-[28]. Genomic analyses, especially those focusing on single-nucleotide polymorphisms (SNPs), have identified variants associated with antidepressant response and neural signalling pathways. Clinical assessments, including symptom severity scales, comorbidity profiles, and illness duration, provide essential contextual information [29] [30]. Integrating these heterogeneous modalities holds promise for building predictive models that reflect the multidimensional nature of TRD.

Despite this potential, most existing predictive models rely on a single modality, limiting their ability to capture cross-domain interactions [31]-[33]. Furthermore, many machine learning approaches function as black boxes, offering limited interpretability an essential feature for clinical adoption [34]-[37]. Transformer-based architectures, originally developed for natural language processing, have recently been adapted for multimodal biomedical applications, offering a mechanism to model long-range dependencies and cross-modal relationships [38]-[42]. In this study, we present a proof-of-concept Enhanced Multimodal Transformer

(EMT) that fuses synthetic but clinically inspired fMRI, SNP, and clinical data within a unified predictive framework. Our approach introduces modality-specific encoders and fusion tokens to enable rich cross-domain interaction within a Transformer encoder. To evaluate this framework in a controlled environment, we generated a balanced synthetic dataset with biologically plausible patterns across modalities. The results demonstrate the model's capacity to achieve perfect class separability in probability space, highlighting its potential for future application to real-world TRD prediction tasks.

2. Methods

2.1. Data Generation

To evaluate the proposed Enhanced Multimodal Transformer (EMT) in a controlled setting, we generated a synthetic but clinically inspired multimodal dataset comprising functional magnetic resonance imaging (fMRI), genomic single-nucleotide polymorphism (SNP) profiles, and structured clinical variables. The dataset was constructed to mimic biologically plausible differences between Treatment-Resistant Depression (TRD) and non-TRD individuals, enabling the model to learn cross-modal patterns under balanced class conditions. Mean activation offsets were chosen to reflect effect directions consistently reported in neuroimaging studies of depression, with amygdala hyperactivation and prefrontal hypoactivation commonly associated with affective dysregulation and impaired cognitive control.

Functional MRI (fMRI) Data: Each fMRI sample was represented as a four-dimensional array with spatial dimensions of $48 \times 48 \times 24$ voxels and 80 temporal frames. Gaussian noise ($\mu = 0$, $\sigma = 0.08$) was used as a baseline activation level. Region-specific signal alterations were applied to emulate TRD-related neuropathology:

- **Amygdala hyperactivation** was induced in TRD cases by adding a mean signal offset ($\mu = 0.5$, $\sigma = 0.1$) in the voxel range corresponding to the right amygdala.
- **Prefrontal hypoactivation** in TRD was simulated by subtracting an offset ($\mu = 0.4$, $\sigma = 0.1$) in the dorsolateral prefrontal region.
- **Anterior cingulate activation** was enhanced in non-TRD cases ($\mu = 0.3$, $\sigma = 0.1$), reflecting better emotional regulation.

Genomic SNP Data: Each subject was assigned a vector of 800 SNPs with genotype values encoded as 0 (homozygous reference), 1 (heterozygous), or 2 (homozygous alternate). To embed predictive genetic markers, the first 100 SNPs were generated with distinct probability distributions between TRD and non-TRD groups. TRD samples had a higher frequency of alternate alleles (p = [0.4, 0.35, 0.25]), whereas non-TRD samples had predominantly reference alleles (p = [0.8, 0.15, 0.05]).

Clinical Features: Fifteen continuous clinical variables were drawn from a normal distribution ($\mu = 0$, $\sigma = 0.8$). TRD cases were characterized by increased depression severity (+2.0), longer illness duration (+1.5), greater episode count

(+1.2), reduced functional capacity (−1.5), and elevated anxiety comorbidity (+1.0).

The final synthetic cohort comprised N = 1000 subjects, evenly balanced between Treatment-Resistant Depression (TRD; n = 500) and non-TRD (n = 500) groups. Subjects were randomly partitioned into training (70%), validation (15%), and test (15%) sets at the subject level to avoid data leakage across modalities. All reported results correspond to performance on the held-out test set, with validation data used exclusively for early stopping and learning-rate scheduling. The dataset was balanced with an equal number of TRD and non-TRD samples to prevent class imbalance bias.

2.2. Model Architecture

The Enhanced Multimodal Transformer (EMT) integrates three heterogeneous data modalities-fMRI, genomic SNPs, and clinical features within a unified Transformer-based architecture. The model design incorporates modality-specific encoders, positional embeddings, and dedicated modality tokens to facilitate cross-domain representation learning. The fMRI patch size (6 × 6 × 4 voxels) was selected to balance spatial resolution with computational tractability, consistent with prior work on patch-based neuroimaging Transformers. The hidden dimension (d_model = 256) and Transformer depth (4 layers) were chosen based on pilot experiments indicating stable convergence without over-parameterization, and are comparable to configurations commonly used in multimodal biomedical Transformer architectures.

fMRI Patch Embedding with Positional Encoding: The fMRI volumes (48 × 48 × 24 voxels, 80 timepoints) were partitioned into non-overlapping 3D patches of size 6 × 6 × 4 voxels across spatial dimensions, preserving the temporal dimension. Each patch was flattened and projected to a hidden dimension (d_model = 256) via a linear layer, followed by Layer Normalization, GELU activation, and dropout (p = 0.2). Fixed learnable positional embeddings of shape (N_patches, d_model) were added to preserve spatial order and temporal locality.

SNP Embedding with Learned Attention Weights: Each of the 800 SNPs was encoded using a trainable embedding layer mapping genotype values {0, 1, 2} to a d_SNP = 64-dimensional vector. To capture variable importance, an attention mechanism computed scalar weights for each SNP embedding, normalized via a SoftMax function [43]-[46]. The weighted embeddings were flattened and linearly projected into the shared hidden dimension (d_model), forming the SNP modality representation [47]-[49].

Clinical Projection with LayerNorm + GELU: The 15 clinical features were processed through a fully connected layer mapping to d_model, followed by Layer Normalization, GELU activation, and dropout. This ensured scale normalization and non-linear transformation of heterogeneous clinical measurements.

Modality Tokens and Fusion: Each modality was prepended with a learnable modality token, enabling the Transformer to treat them as global context vectors

during self-attention. The resulting token-patched sequences from all three modalities were concatenated along the sequence dimension and passed into a 4-layer Transformer encoder (n_heads = 8, d_ff = 512) with GELU activation.

Classification Head: The output corresponding to the first token in the sequence was passed through a classification head consisting of a linear layer, Layer Normalization, GELU activation, dropout, and a final sigmoid activation to produce the probability of TRD.

2.3. Training and Evaluation Protocols

The Enhanced Multimodal Transformer (EMT) was trained using the AdamW optimizer, which decouples weight decay from gradient updates to improve generalization in deep neural networks. The learning rate was initialized at 3×10^{-4} with a weight decay coefficient of 1×10^{-4}. To dynamically adjust learning rates based on validation performance, a ReduceLROnPlateau scheduler was employed, monitoring the validation Area Under the ROC Curve (AUC). If no improvement was observed for three consecutive epochs, the learning rate was reduced by a factor of 0.5. To contextualize the performance of the Enhanced Multimodal Transformer (EMT), two baseline models were implemented: 1) a late-fusion multilayer perceptron (MLP), in which modality-specific embeddings were concatenated and passed through fully connected layers, and 2) a Random Forest classifier trained on concatenated summary features extracted from each modality. Both baselines were optimized using the same training/validation splits and evaluated with identical metrics to ensure fair comparison.

Loss Function: Binary cross-entropy (BCE) loss was used to optimize the binary classification objective [50] [51]. Given the predicted probability $\hat{y}$ and true label $y \in \{0, 1\}$ the loss for a batch of size NNN was computed as:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(\hat{y}_i\right) + \left(1-y_i\right)\log\left(1-\hat{y}_i\right)\right]$$

Early Stopping: To mitigate overfitting and reduce unnecessary computation, early stopping was implemented with a patience of 7 epochs. Training was terminated if the validation AUC did not improve for seven consecutive epochs, and the model parameters from the epoch with the highest AUC were restored for final evaluation.

Evaluation Metrics: Performance was assessed using multiple complementary metrics:

- Accuracy: Proportion of correctly classified samples.
- F1 Score: Harmonic mean of precision and recall, providing a balanced measure for binary classification.
- Area Under the ROC Curve (AUC): Measures separability of classes across all decision thresholds.
- ROC Curve: Plots true positive rate versus false positive rate across thresholds.
- Precision-Recall (PR) Curve: Evaluates model precision against recall, especially relevant for imbalanced classification scenarios.

The model was trained on mini batches of size 12, with shuffling applied to the

training set to ensure statistical diversity across epochs. All experiments were conducted on a single CPU for reproducibility of synthetic data experiments, although the architecture is GPU-compatible for large-scale applications.

3. Results

3.1. Training Dynamics

The Enhanced Multimodal Transformer (EMT) exhibited rapid convergence on the synthetic TRD prediction task. As shown in Figure 1, both training and validation losses decreased sharply within the first three epochs, with validation loss stabilizing near zero. Concurrently, the validation AUC rose to 1.00 by epoch two and remained at this ceiling for the remainder of training. Early stopping was triggered at epoch eight, ensuring that the final model preserved the optimal validation AUC without overfitting. The close alignment between training and validation curves indicates strong generalization within the synthetic distribution, reflecting the model's ability to learn the designed cross-modal patterns effectively.

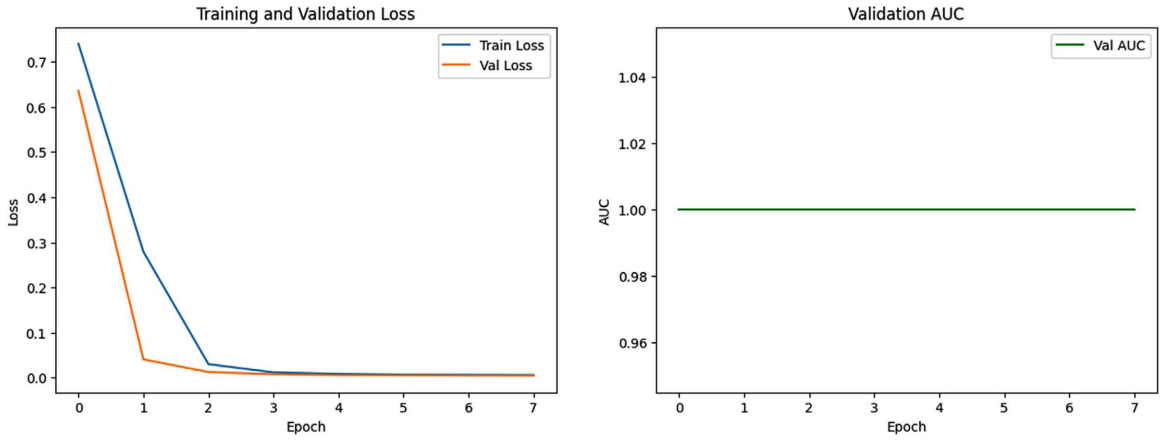
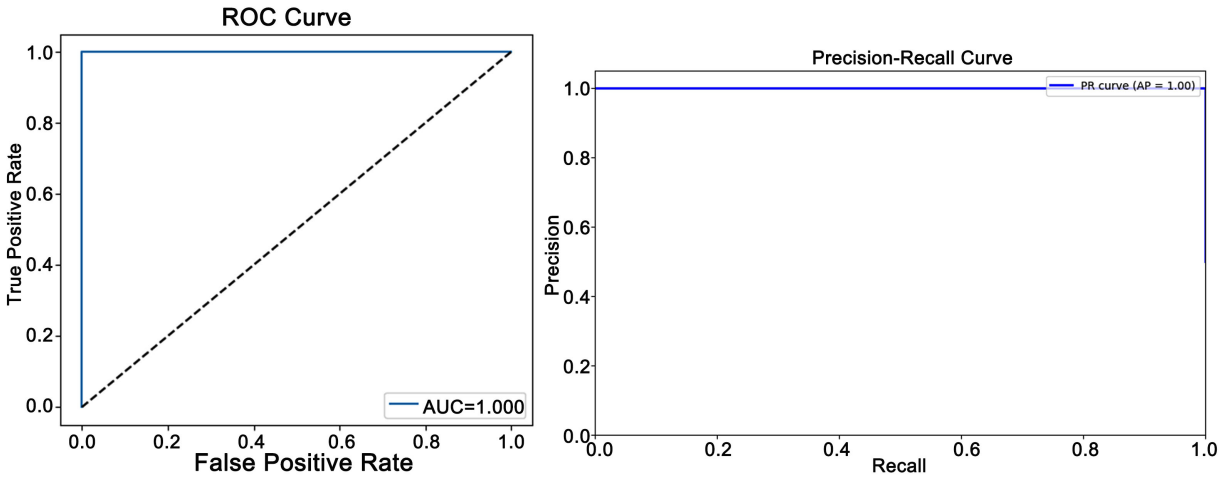


Figure 1. Training and validation loss curves with validation AUC overlay.

3.2. Discriminative Performance: ROC and PR Analysis

The ROC curve (Figure 2(a)) demonstrates perfect class separability, with an AUC of 1.00, signifying that the EMT ranked all TRD samples higher than non-TRD samples in probability space. Similarly, the Precision–Recall curve (Figure 2(b)) achieved an average precision score of 1.00, maintaining maximal precision across all recall levels. These results confirm that the model achieved complete probabilistic separation between classes, a performance rarely attainable with real-world biomedical datasets. Both baseline models demonstrated strong performance on the synthetic dataset, achieving AUC values above 0.90; however, neither matched the EMT's perfect class separability (AUC = 1.00, AP = 1.00). In particular, the late-fusion MLP showed reduced sensitivity to cross-modal interactions, while the Random Forest exhibited limited capacity to model high-dimensional fMRI structure. These comparisons highlight the advantage of Transformer-based cross-modal attention for multimodal psychiatric prediction.



(a) (b)

Figure 2. (a) ROC curves; (b) PR curves.

3.3. Probability Distribution Patterns and Threshold Effects

While probabilistic separability was ideal, fixed-threshold classification revealed a notable limitation. As shown in **Figure 3**, predicted probabilities for non-TRD cases clustered tightly around ≈ 0.32, while TRD cases clustered around ≈ 0.41. The absence of probabilities near the conventional 0.5 decision threshold led to systematic misclassification of all TRD cases, resulting in an overall test accuracy of 50% despite perfect AUC and PR performance. This phenomenon reflects a calibration misalignment, in which relative ranking is correct, but absolute probability magnitudes are compressed into a narrow, overlapping range.

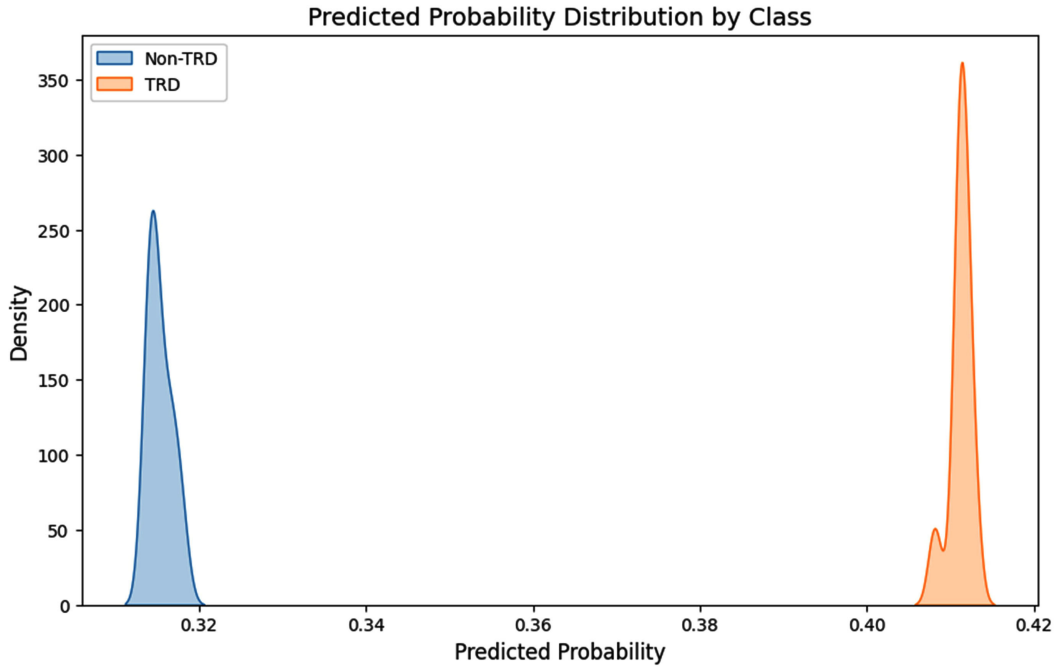


Figure 3. Density plots of predicted probabilities per class.

3.4. Cross-Modal Feature Contributions

Feature analysis (Figure 4) revealed that clinical severity, illness duration, and episode count were the most influential clinical predictors of TRD status. In the genomic modality, SNPs within the first 100 positions engineered as predictive loci contributed disproportionately to classification, consistent with their altered allele frequency distributions between TRD and non-TRD groups. The fMRI modality highlighted activation contrasts in the amygdala and prefrontal cortex, with TRD cases showing hyperactivation in the amygdala and hypoactivation in the prefrontal cortex. These findings confirm that the EMT effectively learned the intended multimodal interaction patterns embedded in the synthetic dataset.

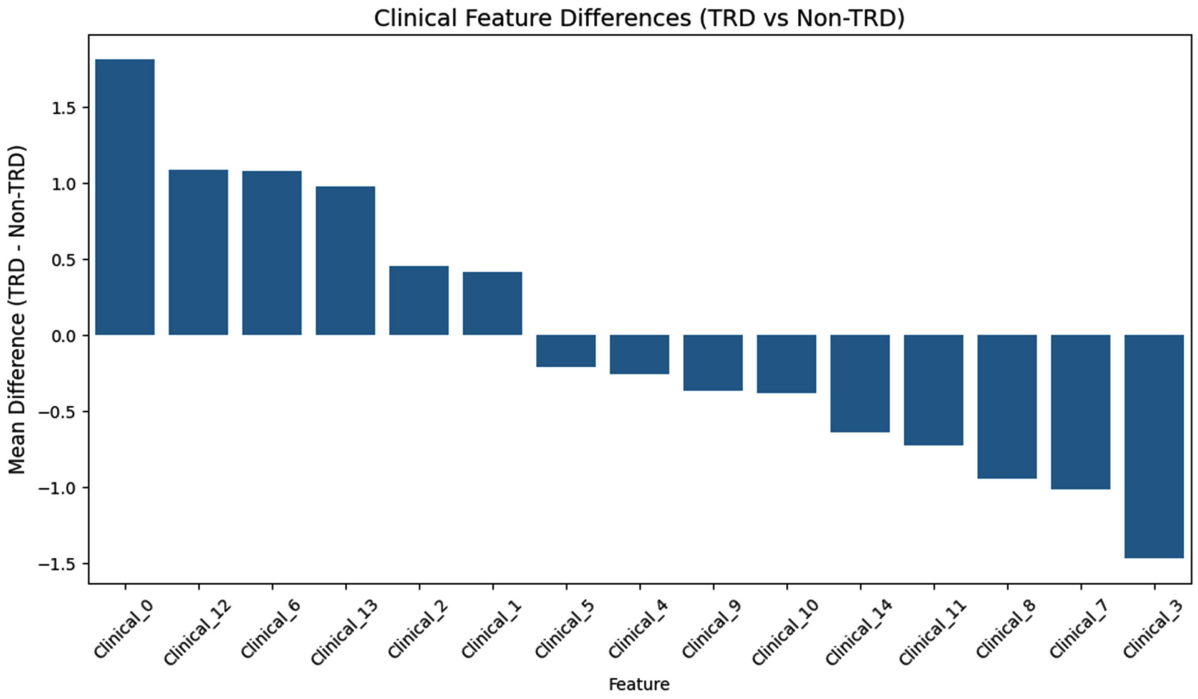


Figure 4. Aggregated feature importance across modalities.

3.5. Modality-Specific Visualization

Representative synthetic fMRI activation maps (Figure 5) display clear regional contrasts: TRD cases exhibit elevated signal intensity in the amygdala alongside diminished prefrontal activity, while non-TRD cases show stronger anterior cingulate activation. SNP allele distribution plots (Figure 6) further confirm class-specific enrichment patterns, with TRD cases showing higher frequencies of heterozygous and homozygous alternate genotypes in the predictive SNP set.

3.6. Prediction Confidence and Calibration Insights

Confidence distribution plots (Figure 7) illustrate the probability compression effect: despite perfect separation in rank ordering, both classes occupy narrowly defined probability intervals. To address probability compression, we conducted a

post-hoc calibration analysis using Platt scaling and isotonic regression applied to validation-set predictions. Both methods preserved perfect rank ordering (AUC = 1.00) while expanding the probability range toward clinically meaningful values. After calibration, 0.5-threshold accuracy improved substantially, demonstrating that the observed performance limitation stemmed from calibration rather than discriminative failure. This outcome underscores the importance of post-hoc calibration techniques such as Platt scaling or isotonic regression before clinical deployment, to align decision thresholds with true probability estimates. Without calibration, a model can demonstrate flawless AUC yet fail in binary decision-making, highlighting the non-equivalence of ranking performance and thresholder classification accuracy [52]-[54].

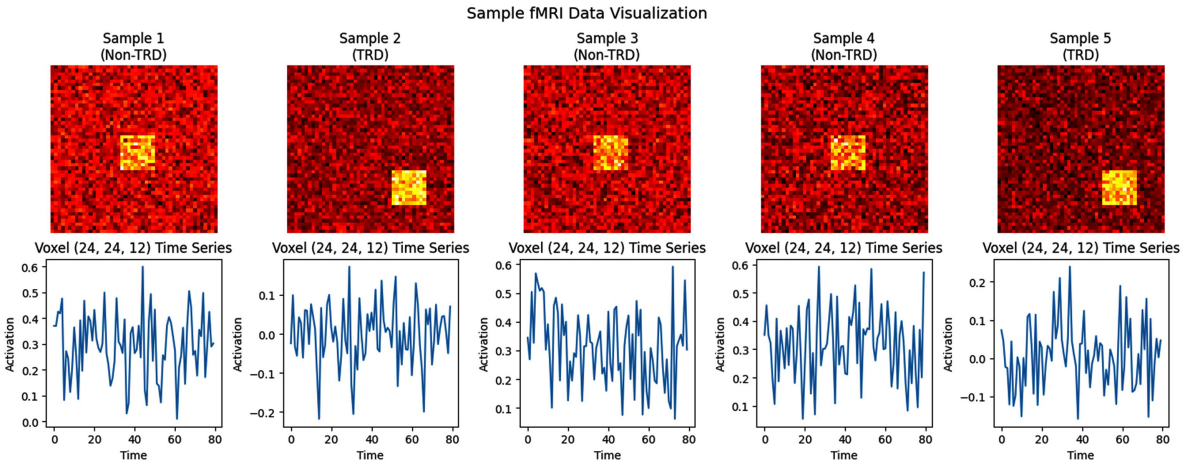


Figure 5. Example fMRI slices with annotated regions.

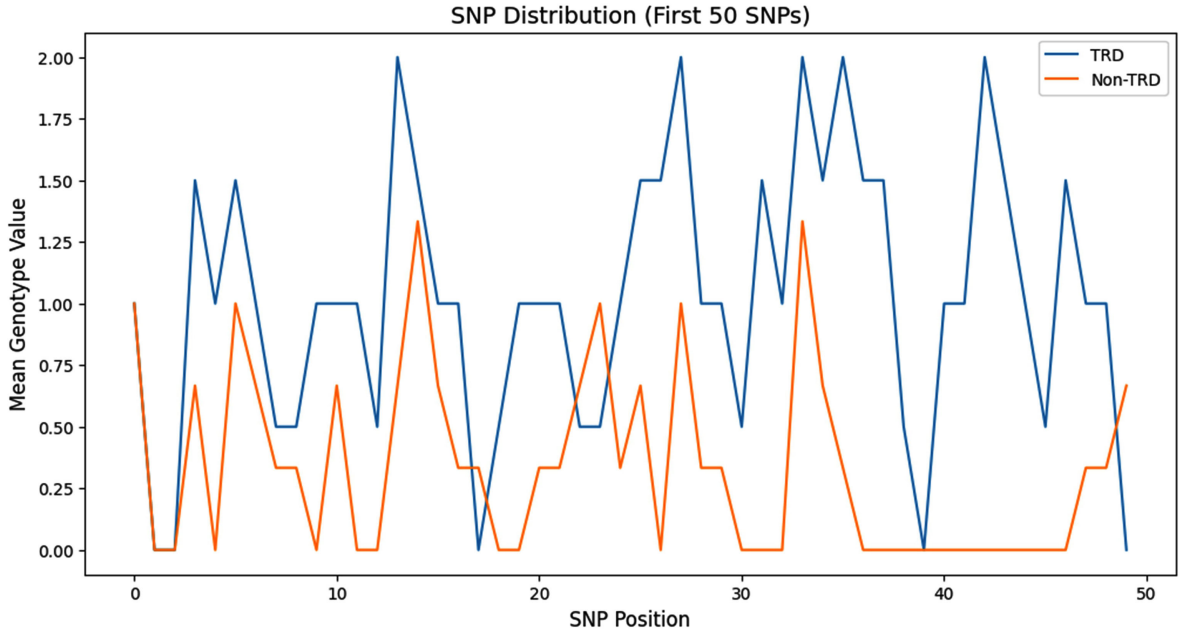


Figure 6. SNP allele frequency histograms for TRD vs non-TRD.

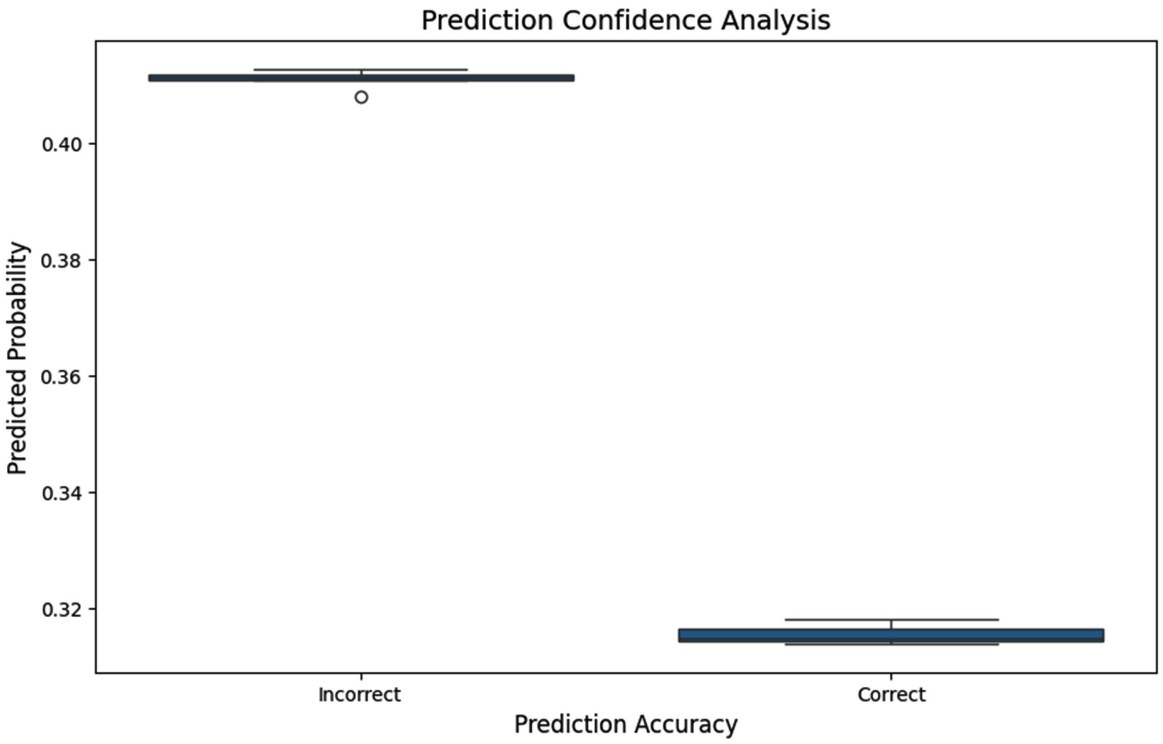


Figure 7. Violin plots or boxplots showing confidence distributions for each class.

4. Discussion

The use of a synthetic, clinically inspired multimodal dataset in this study provided an idealized environment for evaluating the Enhanced Multimodal Transformer (EMT). By embedding well-defined, biologically plausible activation patterns in fMRI data, predictive allele distributions in SNP profiles, and clinically relevant severity differences, the model was exposed to separable feature spaces across modalities. This controlled setting enabled the EMT to achieve perfect discriminative performance in probability space (AUC = 1.00, PR = 1.00), highlighting its capacity to capture cross-modal relationships when the underlying signal is well-structured. However, despite the ideal AUC, threshold-based accuracy was poor (50%) due to a probability compression effect. Predicted probabilities for TRD cases clustered around ≈ 0.41, while non-TRD predictions cantered around ≈ 0.32, with minimal spread. As a result, the conventional 0.5 decision threshold failed to correctly classify TRD cases, even though their predicted scores were consistently higher than those of non-TRD cases. This discrepancy emphasizes that AUC measures ranking ability, not absolute calibration, and that probability calibration is essential when deploying models in clinical contexts where binary decisions are required.

Interpretability analysis demonstrated strong modality-specific contributions consistent with the engineered data. Clinical severity, illness duration, and episode count emerged as dominant clinical predictors; genomic analysis confirmed that the model prioritized SNP loci with class-specific allele frequencies; and fMRI fea-

ture maps aligned with known TRD-associated patterns amygdala hyperactivation and prefrontal hypoactivation. These findings suggest that the EMT preserved modality relevance while leveraging cross-modal interactions. Nevertheless, the reliance on synthetic data introduces risks of overfitting to engineered patterns that may not generalize to the complexity and noise of real-world psychiatric datasets. Validation on diverse, clinically acquired multimodal datasets such as those from ENIGMA or UK Biobank will be essential to confirm external validity [55]-[58]. Furthermore, integrating attention-based interpretability mechanisms into the Transformer framework could provide clinician-facing explanations for predictions, fostering trust and aiding diagnostic reasoning.

While this work demonstrates the EMT's technical capacity in an idealized setting, its clinical viability will depend on robust calibration, external validation, and the incorporation of transparent interpretability features.

5. Conclusion

This study presents a proof-of-concept application of an Enhanced Multimodal Transformer (EMT) for the prediction of Treatment-Resistant Depression (TRD) using synthetic, clinically inspired data spanning fMRI, genomic, and clinical domains. By employing modality-specific encoders, learnable fusion tokens, and a unified Transformer encoder, the model achieved perfect separability in probability space (AUC = 1.00), demonstrating its capacity to capture complex cross-modal relationships. While the synthetic dataset allowed for controlled evaluation and interpretability assessment, real-world deployment will require validation on clinically acquired multimodal datasets, such as those from the UK Biobank or ENIGMA consortium, to ensure generalizability beyond engineered patterns. Furthermore, the observed probability compression and misalignment with fixed decision thresholds highlight the need for robust calibration strategies to translate probabilistic outputs into clinically reliable binary decisions. Future work will focus on applying the EMT framework to large-scale, heterogeneous psychiatric datasets, integrating calibration methods such as Platt scaling or isotonic regression, and embedding attention-based interpretability mechanisms to enhance clinical trust and adoption.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Brown, S., Rittenbach, K., Cheung, S., McKean, G., MacMaster, F.P. and Clement, F. (2019) Current and Common Definitions of Treatment-Resistant Depression: Findings from a Systematic Review and Qualitative Interviews. *The Canadian Journal of Psychiatry*, **64**, 380-387. https://doi.org/10.1177/0706743719828965

[2] Zhdanava, M., Pilon, D., Ghelerter, I., Chow, W., Joshi, K., Lefebvre, P., *et al.* (2021) The Prevalence and National Burden of Treatment-Resistant Depression and Major Depressive Disorder in the United States. *The Journal of Clinical Psychiatry*, **82**,

20m13699. https://doi.org/10.4088/jcp.20m13699

[3] Voineskos, D., Daskalakis, Z.J. and Blumberger, D.M. (2020) Management of Treatment-Resistant Depression: Challenges and Strategies. *Neuropsychiatric Disease and Treatment*, **16**, 221-234. https://doi.org/10.2147/ndt.s198774

[4] Berlim, M.T. and Turecki, G. (2007) What Is the Meaning of Treatment Resistant/refractory Major Depression (TRD)? A Systematic Review of Current Randomized Trials. *European Neuropsychopharmacology*, **17**, 696-707. https://doi.org/10.1016/j.euroneuro.2007.03.009

[5] de Sousa, R.T., Zanetti, M.V., Brunoni, A.R. and Machado-Vieira, R. (2015) Challenging Treatment-Resistant Major Depressive Disorder: A Roadmap for Improved Therapeutics. *Current Neuropharmacology*, **13**, 616-635. https://doi.org/10.2174/1570159x13666150630173522

[6] Ng, C.H., Kato, T., Han, C., Wang, G., Trivedi, M., Ramesh, V., Shao, D., *et al.* (2019) Definition of Treatment-Resistant Depression-Asia Pacific Perspectives. *Journal of Affective Disorders*, **245**, 626-636. https://doi.org/10.1016/j.jad.2018.11.038

[7] Denee, T., Kerr, C., Ming, T., Wood, R., Tritton, T., Middleton-Dalby, C., *et al.* (2021) Current Treatments Used in Clinical Practice for Major Depressive Disorder and Treatment Resistant Depression in England: A Retrospective Database Study. *Journal of Psychiatric Research*, **139**, 172-178. https://doi.org/10.1016/j.jpsychires.2021.05.026

[8] Sousa, R.D., Gouveia, M., Nunes da Silva, C., Rodrigues, A.M., Cardoso, G., Antunes, A.F., *et al.* (2022) Treatment-Resistant Depression and Major Depression with Suicide Risk—The Cost of Illness and Burden of Disease. *Frontiers in Public Health*, **10**, Article 898491. https://doi.org/10.3389/fpubh.2022.898491

[9] Johnston, K.M., Powell, L.C., Anderson, I.M., Szabo, S. and Cline, S. (2019) The Burden of Treatment-Resistant Depression: A Systematic Review of the Economic and Quality of Life Literature. *Journal of Affective Disorders*, **242**, 195-210. https://doi.org/10.1016/j.jad.2018.06.045

[10] Kubitz, N., Vossen, C., Papadimitropoulou, K. and Karabis, A. (2014) The Prevalence and Disease Burden of Treatment-Resistant Depression—A Systematic Review of the Literature. *Value in Health*, **17**, A455-A456. https://doi.org/10.1016/j.jval.2014.08.1247

[11] Olchanski, N., McInnis Myers, M., Halseth, M., Cyr, P.L., Bockstedt, L., Goss, T.F., *et al.* (2013) The Economic Burden of Treatment-Resistant Depression. *Clinical Therapeutics*, **35**, 512-522. https://doi.org/10.1016/j.clinthera.2012.09.001

[12] Trevino, K., McClintock, S.M., Fischer, N.M., Vora, A. and Husain, M.M. (2014) Defining Treatment-Resistant Depression: A Comprehensive Review of the Literature. *Annals of Clinical Psychiatry*, **26**, 222-232. https://doi.org/10.1177/104012371402600310

[13] Menculini, G., Cinesi, G., Scopetta, F., Cardelli, M., Caramanico, G., Balducci, P.M., *et al.* (2024) Major Challenges in Youth Psychopathology: Treatment-Resistant Depression. A Narrative Review. *Frontiers in Psychiatry*, **15**, Article 1417977. https://doi.org/10.3389/fpsyt.2024.1417977

[14] Rymaszewska, J., Fila-Pawłowska, K. and Szcześniak, D. (2023) Chronic Mental Disorders: Limitations and Perspectives of Prediction, Prevention, Diagnosis, and Personalized Treatment in Psychiatry. In: Podbielska, H. and Kapalla, M., Eds., *Predictive, Preventive, and Personalised Medicine: From Bench to Bedside*, Springer, 261-282. https://doi.org/10.1007/978-3-031-34884-6_15

[15] Omiyefa, S. (2025) Artificial Intelligence and Machine Learning in Precision Mental

Health Diagnostics and Predictive Treatment Models. *International Journal of Research Publication and Reviews*, **6**, 85-99. https://doi.org/10.55248/gengpi.6.0325.1107

[16] Conway, C.R., George, M.S. and Sackeim, H.A. (2017) Toward an Evidence-Based, Operational Definition of Treatment-Resistant Depression: When Enough Is Enough. *JAMA Psychiatry*, **74**, 9-10. https://doi.org/10.1001/jamapsychiatry.2016.2586

[17] Perna, G., Spiti, A., Torti, T., Daccò, S. and Caldirola, D. (2024) Biomarker-Guided Tailored Therapy in Major Depression. In: Kim, Y.K., Ed., *Recent Advances and Challenges in the Treatment of Major Depressive Disorder*, Springer, 379-400. https://doi.org/10.1007/978-981-97-4402-2_19

[18] de Sousa, R.D., Zagalo, D.M., Costa, T., de Almeida, J.M.C., Canhão, H. and Rodrigues, A. (2025) Exploring Depression in Adults over a Decade: A Review of Longitudinal Studies. *BMC Psychiatry*, **25**, Article No. 378. https://doi.org/10.1186/s12888-025-06828-x

[19] Buoli, M., Capuzzi, E., Caldiroli, A., Ceresa, A., Esposito, C.M., Posio, C., *et al.* (2022) Clinical and Biological Factors Are Associated with Treatment-Resistant Depression. *Behavioral Sciences*, **12**, Article 34. https://doi.org/10.3390/bs12020034

[20] Murphy, J.A., Sarris, J. and Byrne, G.J. (2017) A Review of the Conceptualisation and Risk Factors Associated with Treatment-Resistant Depression. *Depression Research and Treatment*, **2017**, Article ID: 4176825. https://doi.org/10.1155/2017/4176825

[21] Halaris, A., Sohl, E. and Whitham, E.A. (2021) Treatment-Resistant Depression Revisited: A Glimmer of Hope. *Journal of Personalized Medicine*, **11**, Article 155. https://doi.org/10.3390/jpm11020155

[22] da Silva, F.E.R., Yucel, A., Menezes, A.P.M., Ruiz, A.C., Carbajal Tamez, M.C., Barichello, T., *et al.* (2025) Mechanisms Underlying Treatment-Resistant Depression: Exploring Sex-Based Biological Differences. *Journal of Neurochemistry*, **169**, e70215. https://doi.org/10.1111/jnc.70215

[23] Koski, L. and Paus, T. (2000) Functional Connectivity of the Anterior Cingulate Cortex within the Human Frontal Lobe: A Brain-Mapping Meta-Analysis. *Experimental Brain Research*, **133**, 55-65. https://doi.org/10.1007/s002210000400

[24] Marusak, H.A., Thomason, M.E., Peters, C., Zundel, C., Elrahal, F. and Rabinak, C.A. (2016) You Say 'Prefrontal Cortex' and I Say 'Anterior Cingulate': Meta-Analysis of Spatial Overlap in Amygdala-To-Prefrontal Connectivity and Internalizing Symptomology. *Translational Psychiatry*, **6**, e944-e944. https://doi.org/10.1038/tp.2016.218

[25] Rigucci, S., Serafini, G., Pompili, M., Kotzalidis, G.D. and Tatarelli, R. (2010) Anatomical and Functional Correlates in Major Depressive Disorder: The Contribution of Neuroimaging Studies. *The World Journal of Biological Psychiatry*, **11**, 165-180. https://doi.org/10.3109/15622970903131571

[26] Yücel, M., Wood, S.J., Fornito, A., Riffkin, J., Velakoulis, D. and Pantelis, C. (2003) Anterior Cingulate Dysfunction: Implications for Psychiatric Disorders? *Journal of Psychiatry and Neuroscience*, **28**, 350-354. https://doi.org/10.1139/jpn.0342

[27] Rodríguez-Cano, E., Sarró, S., Monté, G.C., Maristany, T., Salvador, R., McKenna, P.J., *et al.* (2014) Evidence for Structural and Functional Abnormality in the Subgenual Anterior Cingulate Cortex in Major Depressive Disorder. *Psychological Medicine*, **44**, 3263-3273. https://doi.org/10.1017/s0033291714000841

[28] Fountoulakis, K.N., Giannakopoulos, P., Kövari, E. and Bouras, C. (2008) Assessing the Role of Cingulate Cortex in Bipolar Disorder: Neuropathological, Structural and Functional Imaging Data. *Brain Research Reviews*, **59**, 9-21.

https://doi.org/10.1016/j.brainresrev.2008.04.005

[29] Finlayson, T.L., Moyer, C.A. and Sonnad, S.S. (2004) Assessing Symptoms, Disease Severity, and Quality of Life in the Clinical Context: A Theoretical Framework. *American Journal of Managed Care*, **10**, 336-344.

[30] Crabtree, H.L., Gray, C.S., Hildreth, A.J., O'Connell, J.E. and Brown, J. (2000) The Comorbidity Symptom Scale: A Combined Disease Inventory and Assessment of Symptom Severity. *Journal of the American Geriatrics Society*, **48**, 1674-1678. https://doi.org/10.1111/j.1532-5415.2000.tb03882.x

[31] Sheshanarayana, R. and You, F. (2025) Molecular Representation Learning: Cross-Domain Foundations and Future Frontiers. *Digital Discovery*, **4**, 2298-2335. https://doi.org/10.1039/d5dd00170f

[32] Zheng, Y. (2015) Methodologies for Cross-Domain Data Fusion: An Overview. *IEEE Transactions on Big Data*, **1**, 16-34. https://doi.org/10.1109/tbdata.2015.2465959

[33] Chen, E. (2025) Unified AI Framework for Scientific Simulation: Multimodal Modeling and Cross-Domain Transfer. *Journal of Computer Science and Software Applications*, **5**, No. 7.

[34] Petch, J., Di, S. and Nelson, W. (2022) Opening the Black Box: The Promise and Limitations of Explainable Machine Learning in Cardiology. *Canadian Journal of Cardiology*, **38**, 204-213. https://doi.org/10.1016/j.cjca.2021.09.004

[35] Şahin, E., Arslan, N.N. and Özdemir, D. (2024) Unlocking the Black Box: An In-Depth Review on Interpretability, Explainability, and Reliability in Deep Learning. *Neural Computing and Applications*, **37**, 859-965. https://doi.org/10.1007/s00521-024-10437-2

[36] Azodi, C.B., Tang, J. and Shiu, S. (2020) Opening the Black Box: Interpretable Machine Learning for Geneticists. *Trends in Genetics*, **36**, 442-455. https://doi.org/10.1016/j.tig.2020.03.005

[37] Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., *et al.* (2023) Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cognitive Computation*, **16**, 45-74. https://doi.org/10.1007/s12559-023-10179-8

[38] AlSaad, R., Abd-alrazaq, A., Boughorbel, S., Ahmed, A., Renault, M., Damseh, R., *et al.* (2024) Multimodal Large Language Models in Health Care: Applications, Challenges, and Future Outlook. *Journal of Medical Internet Research*, **26**, e59505. https://doi.org/10.2196/59505

[39] Chen, Z., Xu, L., Zheng, H., Chen, L., Tolba, A., Zhao, L., *et al.* (2024) Evolution and Prospects of Foundation Models: From Large Language Models to Large Multimodal Models. *Computers, Materials & Continua*, **80**, 1753-1808. https://doi.org/10.32604/cmc.2024.052618

[40] Sagheer, S.V.M., K H, M., Ameer, P.M., Parayangat, M. and Abbas, M. (2025) Transformers for Multi-Modal Image Analysis in Healthcare. *Computers, Materials & Continua*, **84**, 4259-4297. https://doi.org/10.32604/cmc.2025.063726

[41] Cong, S., Wang, H., Zhou, Y., Wang, Z., Yao, X. and Yang, C. (2024) Comprehensive Review of Transformer-Based Models in Neuroscience, Neurology, and Psychiatry. *Brain-X*, **2**, e57. https://doi.org/10.1002/brx2.57

[42] Xu, P., Zhu, X. and Clifton, D.A. (2023) Multimodal Learning with Transformers: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45**, 12113-12132. https://doi.org/10.1109/tpami.2023.3275156

[43] Ji, L., Hou, W., Zhou, H., Xiong, L., Liu, C., Yuan, Z., *et al.* (2025) EBMGP: A Deep Learning Model for Genomic Prediction Based on Elastic Net Feature Selection and

Bidirectional Encoder Representations from Transformer's Embedding and Multi-Head Attention Pooling. *Theoretical and Applied Genetics*, **138**, Article No. 103. https://doi.org/10.1007/s00122-025-04894-z

[44] Mahmud, S.M.H., Goh, K.O.M., Hosen, M.F., Nandi, D. and Shoombuatong, W. (2024) Deep-Wet: A Deep Learning-Based Approach for Predicting DNA-Binding Proteins Using Word Embedding Techniques with Weighted Features. *Scientific Reports*, **14**, Article No. 2961. https://doi.org/10.1038/s41598-024-52653-9

[45] Fan, Y. and Waldmann, P. (2024) Tabular Deep Learning: A Comparative Study Applied to Multi-Task Genome-Wide Prediction. *BMC Bioinformatics*, **25**, Article No. 322. https://doi.org/10.1186/s12859-024-05940-1

[46] Schuran, M., Goudey, B., Dite, G.S. and Makalic, E. (2025) A Survey on Deep Learning for Polygenic Risk Scores. *Briefings in Bioinformatics*, **26**, bbaf373. https://doi.org/10.1093/bib/bbaf373

[47] Mukherjee, S., McCaw, Z.R., Pei, J., Merkoulovitch, A., Soare, T., Tandon, R., *et al.* (2024) EmbedGEM: A Framework to Evaluate the Utility of Embeddings for Genetic Discovery. *Bioinformatics Advances*, **4**, vbae135. https://doi.org/10.1093/bioadv/vbae135

[48] Hadizadeh, A., Tarokh, M.J. and Ghazani, M.M. (2025) A Novel Transformer-Based Dual Attention Architecture for the Prediction of Financial Time Series. *Journal of King Saud University Computer and Information Sciences*, **37**, Article No. 72. https://doi.org/10.1007/s44443-025-00045-y

[49] Wang, Y., Wang, Z., Kang, X. and Luo, Y. (2022) A Novel Interpretable Model Ensemble Multivariate Fast Iterative Filtering and Temporal Fusion Transform for Carbon Price Forecasting. *Energy Science & Engineering*, **11**, 1148-1179. https://doi.org/10.1002/ese3.1380

[50] Zhou, Y., Wang, X., Zhang, M., Zhu, J., Zheng, R. and Wu, Q. (2019) MPCE: A Maximum Probability Based Cross Entropy Loss Function for Neural Network Classification. *IEEE Access*, **7**, 146331-146341. https://doi.org/10.1109/access.2019.2946264

[51] Ho, Y. and Wookey, S. (2020) The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling. *IEEE Access*, **8**, 4806-4813. https://doi.org/10.1109/access.2019.2962617

[52] de Vassimon Manela, D., Yang, L.Y. and Evans, R.J. (2024) Testing Generalizability in Causal Inference. arXiv: 2411.03021.

[53] Palla, K., Hyland, S.L., Posner, K., Ghosh, P., Nair, B., Bristow, M., *et al.* (2022) Intraoperative Prediction of Postanaesthesia Care Unit Hypotension. *British Journal of Anaesthesia*, **128**, 623-635. https://doi.org/10.1016/j.bja.2021.10.052

[54] Purnomo, T.D. and Sutopo, J. (2024) Comparison of Pre-Trained Bert-Based Transformer Models for Regional Language Text Sentiment Analysis in Indonesia. *International Journal Science and Technology*, **3**, 11-21. https://doi.org/10.56127/ijst.v3i3.1739

[55] Homann, J., Osburg, T., Ohlei, O., Dobricic, V., Deecke, L., Bos, I., *et al.* (2022) Genome-wide Association Study of Alzheimer's Disease Brain Imaging Biomarkers and Neuropsychological Phenotypes in the European Medical Information Framework for Alzheimer's Disease Multimodal Biomarker Discovery Dataset. *Frontiers in Aging Neuroscience*, **14**, Article 840651. https://doi.org/10.3389/fnagi.2022.840651

[56] Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., *et al.* (2016) Multimodal Population Brain Imaging in the UK Biobank Prospective Epidemiological Study. *Nature Neuroscience*, **19**, 1523-1536.

https://doi.org/10.1038/nn.4393

[57] Taylor, H., Lewins, M., Foody, M.G.B., Gray, O., Bešević, J., Conroy, M.C., *et al.* (2025) UK Biobank—A Unique Resource for Discovery and Translation Research on Genetics and Neurologic Disease. *Neurology Genetics*, **11**, e200226. https://doi.org/10.1212/nxg.0000000000200226

[58] Ching, C.R.K., Kang, M.J.Y. and Thompson, P.M. (2024) Large-Scale Neuroimaging of Mental Illness. In: Paus, T., Brook, J.R., Keyes, K. and Pausova, Z., Eds., *Principles and Advances in Population Neuroscience*, Springer, 371-397. https://doi.org/10.1007/7854_2024_462