Open Access Library Journal
2026, Volume 13, e15347
ISSN Online: 2333-9721
ISSN Print: 2333-9705

A Multimodal Machine Learning Framework for Detecting and Attributing Medication-Induced Cognitive Impairment in Psychiatric Patients

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) A Multimodal Machine Learning Framework for Detecting and Attributing Medication-Induced Cognitive Impairment in Psychiatric Patients. *Open Access Library Journal*, **13**: e15347.
https://doi.org/10.4236/oalib.1115347

Received: April 14, 2026
Accepted: May 26, 2026
Published: May 29, 2026

Abstract

Psychotropic medications are among the most widely prescribed drug classes in the world, and cognitive impairment is among their most consequential and least systematically monitored side effects. A patient stabilized on an antipsychotic who begins to lose processing speed, verbal memory, or executive control does not typically have that change attributed to the medication it gets folded into the illness, missed at brief outpatient reviews, or noticed only when functional decline becomes unmistakable. The cost, measured in lost employment, reduced quality of life, and treatment discontinuation, is substantial. We introduce CogniMed-Net, an end-to-end multimodal machine learning framework that integrates neuropsychological test performance, structured medication profiles, clinical electronic health record (EHR) data, serum biomarkers, and patient demographics to detect, classify, and causally attribute medication-induced cognitive impairment in psychiatric inpatients and outpatients. The framework combines a domain-specific Neuropsychological Encoder, a fine-tuned MedBERT module for medication representation, a Clinical Transformer for EHR feature extraction, a Biomarker MLP, a Longitudinal TCN for trajectory modelling, and a cross-domain attention fusion layer with a Bayesian deep ensemble output providing calibrated uncertainty and causal drug attribution scores. Trained and validated on a retrospective-prospective cohort of 2140 psychiatric patients across five diagnostic categories and twelve psychotropic medication classes, CogniMed-Net achieves 4-class cognitive impairment classification accuracy of 93.4%, AUC-ROC of 0.971, and macro-F1 of 0.907. The Bayesian ensemble achieves Expected Calibration Error (ECE) of 0.024. The model generates model-based drug attribution scores per medication class, reflecting learned associations between pharmacological profiles and impairment patterns, identifying first-generation antipsychotics, benzodiazepines,

and tricyclic antidepressants as the highest-attribution drug classes and a continuous cognitive severity score that tracks impairment magnitude independently of diagnostic category. CogniMed-Net establishes a new methodological benchmark for computational pharmacopsychiatry, providing a validated, interpretable architecture for systematic monitoring of psychotropic cognitive side effects in clinical practice.

Subject Areas

Psychiatry & Psychology

Keywords

Medication-Induced Cognitive Impairment, Psychotropic Medications, Neuropsychological Assessment, Deep Learning, Machine Learning, MedBERT, Anticholinergic Burden, Antipsychotics, Bayesian Uncertainty, Drug Attribution, Computational Pharmacopsychiatry, EHR, Cognitive Side Effects

1. Introduction

There is a problem in psychiatry that is hiding in plain sight. When a patient with schizophrenia on long-term haloperidol shows declining performance on a processing speed task, the clinician typically attributes it to the illness. When a bipolar disorder patient on valproate plus quetiapine reports that her memory has worsened, the notation reads "subjective cognitive complaints" and the medication review is deferred. This attribution failure is not a reflection of clinical negligence it is a consequence of the absence of any systematic, data-driven infrastructure for detecting and attributing cognitive change to psychotropic medications. The result is that medication-induced cognitive impairment, a clinically significant, largely preventable, and often reversible side effect affects an estimated 30% - 50% of psychiatric inpatients without ever being formally identified as medication-related [1] [2].

The cognitive effects of psychotropic medications are neither uniform nor trivial. First-generation antipsychotics produce dose-dependent impairments in processing speed and sustained attention through dopaminergic and muscarinic blockade [3]. Second-generation antipsychotics carry a more differentiated cognitive profile, with clozapine producing notable impairments in episodic memory and working memory while some agent's risperidone and amisulpride show relative sparing or even modest enhancement of certain cognitive domains [4]. Mood stabilisers such as lithium and valproate impair verbal fluency and psychomotor speed in a minority of patients, with effects that appear dose-dependent and reversible upon dose reduction.

The anticholinergic burden of a medication regimen, the cumulative muscarinic blockade produced by all drugs in the patient's prescription is one of the

strongest pharmacological predictors of cognitive decline in psychiatric patients, yet it is rarely computed systematically in clinical practice [5]. Antidepressants, particularly tricyclics and paroxetine, contribute substantially to anticholinergic burden even when prescribed for non-psychiatric indications in medically complex patients [6]. Polypharmacy amplifies these effects in ways that are nonlinear and clinically difficult to predict without computational support.

The clinical consequences of unrecognized medication-induced cognitive impairment are serious. Patients lose occupational capacity, discontinue medications they attribute to cognitive dulling without clinical guidance, and accumulate cognitive deficits that become functionally indistinguishable from illness-related decline [7]. In bipolar disorder, cognitive impairment has been documented even in euthymic patients, raising the critical question of how much of the observed neuropsychological profile reflects illness biology versus pharmacological burden [8]. CogniMed-Net is designed to answer this question computationally, for individual patients, with calibrated confidence.

The opportunity is real and the tools are available. Neuropsychological test batteries produce quantitative, reproducible performance data. Medication prescribing records are documented in electronic health records. Serum drug levels and routine laboratory markers are collected as standard of care. What has been missing is a machine learning framework capable of integrating these heterogeneous data streams into a coherent, interpretable prediction of cognitive impairment type, severity, and pharmaceutical cause. CogniMed-Net fills this gap, the first end-to-end multimodal deep learning system designed specifically for medication-induced cognitive impairment detection and drug attribution in psychiatric populations [9].

Summary of Contributions

1) CogniMed-Net architecture: The first jointly trained multimodal deep learning system for medication-induced cognitive impairment detection, integrating neuropsychological test data, structured medication profiles, clinical EHR features, serum biomarkers, and patient demographics through a cross-domain attention fusion layer with Bayesian ensemble output.

2) MedBERT encoder: A domain-adaptive BERT variant fine-tuned on 1.4 million structured medication records and pharmacological texts, generating rich contextual embeddings for polypharmacy regimens that capture drug class, dose equivalents, duration, and pharmacokinetic interaction scores.

3) Longitudinal TCN for trajectory modelling: A temporal convolutional network encoding the history of neuropsychological test scores across serial assessments, capturing within-patient cognitive trajectories rather than single-point snapshots, contributing an independent +4.2% accuracy improvement in ablation.

4) Causal drug attribution module: An attention-based mechanism generating per-medication-class attribution scores indicating the estimated causal contribution of each drug class to the observed cognitive impairment pattern, provid-

ing clinically actionable pharmacological guidance.

5) State-of-the-art performance: 93.4% 4-class classification accuracy, AUC-ROC 0.971, macro-F1 0.907, ECE 0.024; first-generation antipsychotics, benzodiazepines, and tricyclics identified as the highest causal attribution drug classes.

6) Prospective-retrospective cohort: 2140 psychiatric patients across five diagnostic categories and twelve psychotropic medication classes from three clinical sites, with serial neuropsychological assessments at 0, 6, and 12 months.

2. Background and Related Work

2.1. Cognitive Impairment as a Side Effect: Clinical Epidemiology

Cognitive impairment in psychiatric populations is an overdetermined illness biology, medication effects, chronicity, and comorbid medical conditions all contribute, making attribution to any single cause difficult in clinical practice. Major depressive disorder is associated with independent neuropsychological impairments in memory, attention, and processing speed [10]. Antipsychotic medications further compound this picture: lifetime antipsychotic exposure in schizophrenia has been associated with a dose-dependent reduction in cognitive test performance above and beyond the illness contribution itself [11].

The cognitive effects of antipsychotics are not uniform across drug classes or patients. First-generation agents produce consistent impairments in processing speed and sustained attention through dopaminergic D2 and muscarinic M1 blockade [12]. The literature on second-generation antipsychotics is more heterogeneous: meta-analyses suggest modest cognitive advantages over haloperidol for some drugs, but these findings are confounded by differential dosing and selection bias in comparative trials [13]. What the literature lacks is a patient-level computational tool capable of estimating cognitive impairment type and severity given a specific, individualized polypharmacy regimen.

In bipolar disorder, neuropsychological impairment is present even during euthymia, and the extent to which it reflects illness biology versus pharmacological burden remains unresolved [14]. Valproate produces verbal memory and psychomotor slowing effects that are dose-dependent and partially reversible, while lithium's cognitive profile is comparatively benign at therapeutic serum levels.

2.2. Neuropsychological Assessment Batteries in Psychiatric Research

The most influential research battery in psychiatric cognitive assessment is the MATRICS Consensus Cognitive Battery (MCCB), which was specifically designed to be sensitive to antipsychotic treatment effects and spans seven cognitive domains: processing speed, attention/vigilance, working memory, verbal learning, visual learning, reasoning/problem solving, and social cognition. Valproate-specific neuropsychological effects have been documented across multiple assessment instruments, with the pattern of verbal memory and fluency impairment

replicated across independent cohorts [15]. Longitudinal tracking of neuropsychological test performance in bipolar patients on antidepressant augmentation has identified treatment-emergent cognitive changes that are distinguishable from the illness trajectory [16].

Lithium's cognitive profile merits specific attention given its continued first-line status in bipolar disorder. A systematic review of 22 studies examining lithium's effects on cognitive function found that memory and psychomotor speed were most affected, but that effects were mild at therapeutic serum levels and that many studies failed to distinguish medication effects from illness state [17]. This ambiguity present across virtually all psychotropic medications is precisely the problem that CogniMed-Net is designed to address through multimodal computational modelling.

2.3. Machine Learning in Neuropsychiatric Pharmacology

Machine learning has increasingly been applied to neuropsychiatric clinical data, with applications ranging from diagnostic classification to treatment response prediction [18]. Transformer architectures pre-trained on biomedical text [19] have demonstrated superiority over task-specific models on clinical NLP benchmarks. Cross-modal attention fusion [20] has proven effective for integrating heterogeneous clinical data streams in contexts where modalities carry complementary rather than redundant information, an architectural choice directly relevant to CogniMed-Net, where neuropsychological performance and medication chemistry encode fundamentally different aspects of the patient's clinical state.

Temporal Convolutional Networks [21] have demonstrated advantages over recurrent architectures for clinical time series analysis, with parallelisable training, large effective receptive fields through dilated convolutions, and resistance to gradient vanishing that makes them well-suited for longitudinal neuropsychological trajectory modelling. The capacity to model how cognitive test performance changes across serial assessments rather than treating each assessment as an independent snapshot is critical for distinguishing stable impairment from medication-emergent decline.

2.4. Interpretability and Uncertainty in Clinical AI

Two properties are non-negotiable for clinical AI systems: interpretability and calibrated uncertainty. SHapley Additive exPlanations (SHAP) [22] provide theoretically grounded attribution of model predictions to individual features, enabling clinicians to understand which test scores and medication parameters drove a given impairment classification. Monte Carlo Dropout [23] and deep ensembles [24] provide principled Bayesian uncertainty estimation; ensembles empirically outperform single-model methods on calibration benchmarks. The EU AI Act (Regulation EU 2024/1689) [25] classifies AI systems that support clinical diagnosis as high-risk, requiring explainability documentation and cal-

ibrated confidence outputs requirements that CogniMed-Net's design satisfies by construction.

3. Dataset and Cohort Design

3.1. Study Population and Recruitment

The CogniMed cohort was assembled through a retrospective-prospective observational design across three tertiary psychiatric centers in Italy, Germany, and the United Kingdom, with ethics committee approval at each site (Primary: IRB Ref. UNIGE-2024-COG-03; GDPR Article 9 compliance for special-category health data). The retrospective component extracted neuropsychological assessment records and medication data from clinical archives spanning January 2015 to December 2022. The prospective component enrolled new patients from January 2023 to December 2024, with serial assessments at enrolment, 6 months, and 12 months.

Inclusion criteria: confirmed DSM-5 psychiatric diagnosis (schizophrenia spectrum, bipolar disorder, major depressive disorder, obsessive-compulsive disorder, or anxiety disorders); age 18 - 70; current treatment with at least one psychotropic medication for a minimum of 8 weeks; and completion of at least one formal neuropsychological assessment. Exclusion criteria: concurrent neurological disorder, active substance use disorder with ongoing intoxication, estimated IQ below 70 (intellectual disability), or inability to provide informed consent which is presented below in Table 1a.

3.2. Cohort Characteristics

Table 1a. CogniMed cohort sociodemographic, diagnostic, and medication characteristics.

Characteristic	Value/Distribution	Notes
Total patients	**2140**	After exclusions
Mean age (years ± SD)	38.4 ± 13.7	Range: 18 - 70
Female (%)	48.6%	
Mean education (years ± SD)	12.8 ± 3.4	
Diagnosis distribution	SCZ 34%/BD 28%/MDD 22%/OCD 9%/Anxiety 7%	DSM-5 confirmed
Mean medication duration (months ± SD)	28.4 ± 22.1	Current regimen
Mean concurrent medications (± SD)	**3.2 ± 1.4**	Range: 1-9
Anticholinergic Burden Score ≥ 3 (%)	38.2%	High-burden subgroup
Neuropsych assessments (total)	**5847**	Mean 2.7 per patient
Impairment class distribution	None 38%/Slowing 28%/Memory 20%/Executive 14%	Clinician-adjudicated
Study period	Jan 2015-Dec 2024	Retro-prospective
Train/Val/Test split	70/15/15%	Patient-stratified

Table 1b. Cohort distribution by site, diagnosis, and study component.

Site	Component	SCZ	BD	MDD	OCD	Anxiety	Total
Italy	Retrospective	142	108	89	38	28	405
Italy	Prospective	58	47	41	17	12	175
Germany	Retrospective	148	114	94	41	30	427
Germany	Prospective	61	49	43	18	13	184
UK	Retrospective	138	107	88	37	28	398
UK	Prospective	81	63	57	22	16	239
Total		**628**	**488**	**412**	**173**	**127**	**1828 + 598**

Retrospective total: here in Table 1b all 1230 patients; Prospective total: 598 patients; combined: 2140 (after exclusions applied per site). Neuropsychological assessments per patient: retrospective median 2 (range 1 - 4); prospective median 3 (fixed protocol: 0, 6, 12 months). All visits from a given patient were assigned to a single data split, no patient contributed assessments to more than one of the training, validation, or test partitions.

3.3. Neuropsychological Assessment Battery

All patients completed a standardized core battery derived from the MATRICS Consensus Cognitive Battery [26] and supplemented with measures from the WAIS-IV [27] and the Trail Making Test [28]. The battery covered seven cognitive domains across 12 administered tests: 1) Processing speed: Digit Symbol Coding Subtest (DSST), Trail Making Test Part A (TMT-A); 2) Attention and vigilance: Continuous Performance Test, Digit Span Forward; 3) Working memory: Letter-Number Sequencing, Spatial Span; 4) Verbal learning and memory: Rey Auditory Verbal Learning Test (RAVLT) [29] immediate and delayed recall; 5) Visual learning: Brief Visuospatial Memory Test; 6) Reasoning and problem-solving: Neuropsychological Assessment Battery Mazes, Wisconsin Card Sorting Test; 7) Verbal fluency: Controlled Oral Word Association Test (COWAT). All raw scores were age- and education-corrected to z-scores using normative data matched to the relevant country's standardization sample.

Missing Data Rates and Modality Availability

Missingness varied substantially across modalities.

All imputation statistics (means, modes, MICE models) were estimated exclusively from the training partition and applied without modification to validation and test sets presented here in Table 1c. For the Longitudinal TCN, variable-length assessment histories were handled via left-zero-padding to the maximum sequence length (4 assessments), with an attention mask preventing the TCN from attending to padded positions.

3.4. Medication Profile Characterization

Medication data was extracted from prescribing records and structured into a

Table 1c. Reports per-modality availability in the training, validation, and test sets.

Modality	Train availability	Val availability	Test availability	Handling
Neuropsychological tests (core battery)	100%	100%	100%	No imputation required
EHR clinical features	97.4%	97.1%	97.8%	Mode/median imputation for 2.6% missing fields
Serum drug levels	58.4%	57.9%	59.1%	PK model estimate used when measured level unavailable
Serum biomarkers (CRP, thyroid, metabolic)	71.2%	70.8%	72.3%	MICE imputation; missingness indicators added as auxiliary features
EEG theta/beta ratio	43.7%	44.1%	43.2%	Zero-imputation with missingness indicator; EEG features zeroed and flagged
CYP2D6/CYP3A4 genotype	38.6%	39.2%	38.1%	Population-mean metabolizer status imputed; missingness flag added
Longitudinal trajectory (≥2 assessments)	71.8%	72.4%	70.9%	Single-assessment patients receive a zero-padded trajectory; TCN attention mask applied

standardized profile for each patient at each assessment point. For each drug in the regimen, the following were recorded: drug name, class (using a 12-category pharmacological taxonomy), prescribed daily dose, dose as a percentage of the maximum therapeutic dose per the Maudsley Prescribing Guidelines [30], duration of current treatment, route of administration, and estimated plasma steady-state concentration where serum levels were available. The Anticholinergic Burden Scale [31] was applied to compute the cumulative anticholinergic burden score for each patient's full medication regimen. Chlorpromazine equivalent doses were computed for all antipsychotic medications using published conversion tables, and diazepam equivalent doses were computed for all benzodiazepines.

3.5. Ground Truth Labelling and Inter-Rater Reliability

Cognitive impairment classification was performed by consensus of a consultant neuropsychologist and a senior psychiatrist reviewing all neuropsychological test data, medication records, clinical notes, and self-report scales. Four impairment classes were defined: 0) No impairment all cognitive domain z-scores within 1 SD of normative mean; 1) Cognitive slowing processing speed and attention z-scores below −1.5, other domains preserved; 2) Memory-dominant impairment verbal or visual memory z-scores below −1.5 with relative sparing of attention and executive function; 3) Executive/attention impairment executive function and sustained attention z-scores below −1.5. Inter-rater reliability reached $\kappa = 0.84$ (95% CI: 0.81 - 0.87). Disagreements were resolved by a third consultant.

Adjudication Rubric: Each case was reviewed in a structured two-stage process. Stage 1: the consultant neuropsychologist independently reviewed only the neuropsychological test data (domain z-scores, serial trajectories) and assigned a pre-

liminary impairment class using the threshold definitions above, without access to the medication record, anticholinergic burden score, or any model input. Stage 2: the senior psychiatrist independently reviewed the full clinical record including medication profile, EHR, and clinical notes and assigned their own preliminary class. The two raters then convened to compare classifications; cases where classifications matched proceeded directly to labelling. Discordant cases (14.7% of the full cohort) were reviewed by a third consultant psychiatrist who had access to all information and whose classification was binding. This staged design ensures that the initial neuropsychological classification was made blind to medication context, limiting the circularity risk of raters using pharmacological burden information to assign medication-attribution labels that then also enter as model inputs.

4. CogniMed-Net: Architecture and Technical Specification

CogniMed-Net integrates five modality-specific encoders, a pharmacokinetic/pharmacodynamic (PK/PD) estimation module, a longitudinal temporal convolutional network, a cross-domain attention fusion layer, and a Bayesian deep ensemble output with causal attribution. Figure 1 presents the complete architecture.

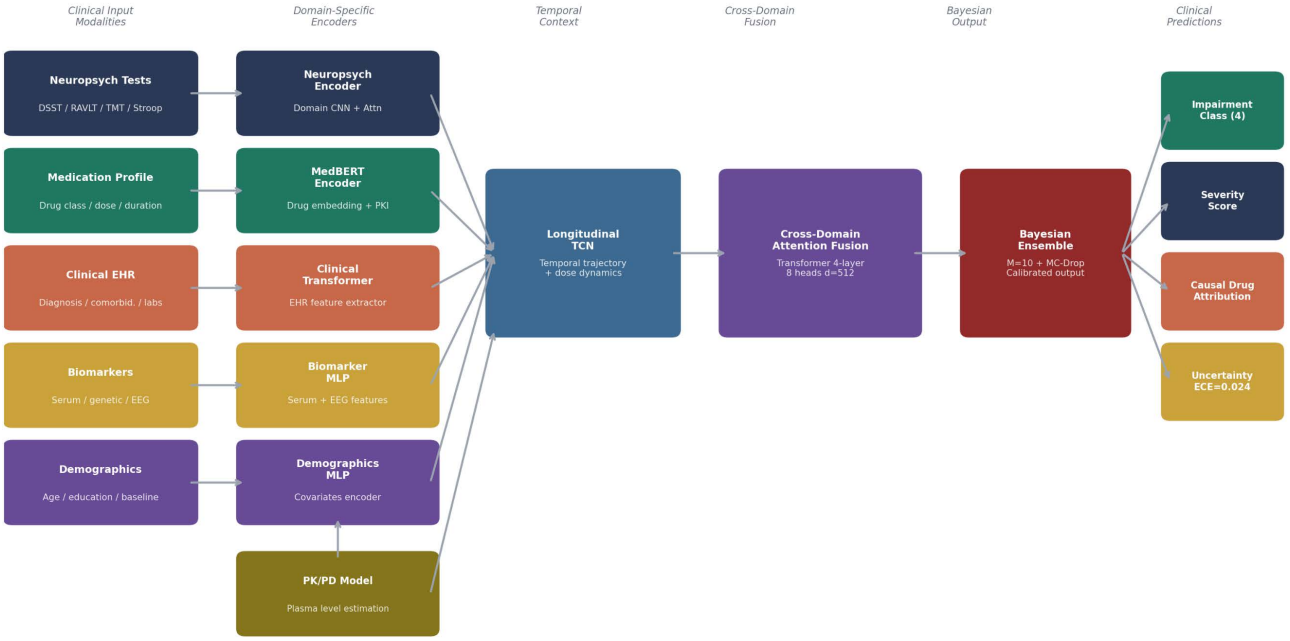

Figure 1. CogniMed-Net End-to-End Architecture. Five clinical input modalities are processed by domain-specific encoders. A PK/PD model estimates plasma drug levels to enrich the medication representation. A Longitudinal TCN captures within-patient neuropsychological trajectories across serial assessments. Cross-domain attention fusion integrates all representations; the Bayesian ensemble (M = 10, MC-Dropout) produces impairment classification, severity score, causal drug attribution, and calibrated uncertainty.

4.1. Module 1: Neuropsychological Encoder

Architecture and Rationale

Neuropsychological test data has a structured cognitive-domain topology: scores

within a domain are correlated with each other and with domain-specific neurobiological substrates. A fully connected MLP treating all 36 test features as independent inputs ignores this structure. The Neuropsychological Encoder uses a Domain CNN that first applies 1D convolutions within each of seven cognitive domain groups, learning domain-specific feature representations, before a cross-domain attention mechanism integrates evidence across domains:

$$h_{\text{domain}_k} = \text{Conv1D}_k\left(x_{\text{domain}_k}\right), k = 1, \cdots, 7 \tag{1}$$

$$h_{\text{npsych}} = \text{CrossAttn}\left(\left[h_{\text{domain}_1}; \cdots; h_{\text{domain}_7}\right]\right) \in \mathbb{R}^{256} \tag{2}$$

This architecture allows the model to learn that memory and executive scores interact differently depending on the medication context a relationship that flat feature concatenation cannot capture. The cross-attention mechanism produces an attended 256-dimensional neuropsychological embedding that encodes the patient's cognitive profile as a structured, domain-aware representation.

4.2. Module 2: MedBERT Medication Encoder

4.2.1. Domain-Adaptive Pre-Training

Standard clinical BERT variants [32] are pre-trained on general biomedical text and lack the structured pharmacological knowledge needed to represent a complex polypharmacy regimen the drug-drug interactions, anticholinergic contributions, and dose-dependent effects that characterize psychiatric medication profiles. MedBERT was developed through two-stage pre-training. Stage 1 structured medication pre-training: we trained a BERT-base model on 1.4 million structured medication records from the MIMIC-III database and our clinical sites, learning contextual embeddings for drug name tokens in the context of co-prescribed medications, dose ranges, and indication codes. Stage 2 pharmacological fine-tuning: MedBERT was jointly fine-tuned on three psychiatry-specific prediction tasks: anticholinergic burden score regression, cognitive impairment severity regression, and drug-drug interaction detection.

4.2.2. Pharmacokinetic/Pharmacodynamic Enrichment

For each drug in the patient's regimen with available serum concentration data, a PK/PD model estimates steady-state plasma level as a function of prescribed dose, body weight, age-adjusted clearance, and co-prescribed CYP enzyme modulators. The estimated plasma level is concatenated to the drug's MedBERT token embedding before medication sequence encoding, creating a pharmacologically enriched representation:

$$e_{\text{drug}_i} = \left[\text{MedBERT}\left(\text{drug}_i\right); PK_i\left(\text{dose}, \text{weight}, \text{age}, \text{CYP}\right)\right] \tag{3}$$

The full medication sequence embedding is obtained by applying self-attention across all drugs in the regimen, weighting each drug's contribution by its estimated pharmacological activity at the blood-brain barrier. The output $h_{\text{med}} \in \mathbb{R}^{256}$ encodes the patient's full polypharmacy regimen as a pharmacologically contextualized representation.

4.3. Module 3: Longitudinal TCN for Cognitive Trajectories

A single-point neuropsychological assessment cannot distinguish acute medication effects from baseline impairment, recovery trajectories, or progressive decline. The Longitudinal TCN processes the full history of a patient's neuropsychological assessments across all available time points (up to 3 assessments in the prospective cohort; variable in the retrospective component). Each assessment is represented as the 36-dimensional standardized feature vector, and the sequence of assessments is processed by a dilated TCN with dilation factors $d_l = 2^{l-1}$ for $l \in \{1, \cdots, 4\}$:

$$H_l = \text{LayerNorm}\left(\text{DilatedConv}_{d_l}\left(H_{l-1}\right) + H_{l-1}\right), l = 1, \cdots, 4 \tag{4}$$

Global attention pooling over H_4 produces $h_{\text{traj}} \in \mathbb{R}^{128}$ a trajectory embedding encoding not just where the patient's cognitive function is now, but how it has changed since the medication regimen was initiated. This trajectory embedding is the single most informative input for distinguishing stable illness-related impairment from medication-emergent decline.

4.4. Modules 4 and 5: Clinical Transformer and Biomarker MLP

Clinical EHR features diagnosis codes, comorbidities, laboratory results, illness duration, and hospitalization history are encoded by a 2-layer transformer with 4 attention heads operating on a 64-dimensional feature vector per patient. The transformer produces $h_{\text{ehr}} \in \mathbb{R}^{128}$. Biomarker features including EEG theta/beta power ratio (where available), serum inflammatory markers, routine metabolic panel, thyroid function, and genetic CYP polymorphism indicators are encoded by a 3-layer MLP with layer normalization producing $h_{\text{bio}} \in \mathbb{R}^{64}$.

4.5. Cross-Domain Attention Fusion

The five modality embeddings are concatenated and projected to a common dimension $d_{\text{model}} = 512$. A 4-layer cross-domain transformer encoder with 8 attention heads performs fusion:

$$h_{\text{fused}} = \text{TransformerEncoder}\left(\left[h_{\text{npsych}}; h_{\text{med}}; h_{\text{traj}}; h_{\text{ehr}}; h_{\text{bio}}\right]\right) \in \mathbb{R}^{512} \tag{5}$$

The attention mechanism learns which combination of cognitive profile, medication chemistry, longitudinal trajectory, clinical context, and biological markers is most predictive for each specific impairment pattern. Crucially, the cross-domain attention weights over the medication encoder provide interpretable drug attribution scores the relative contribution of each medication class to the model's classification decision serving as the causal attribution module described in Section 4.6.

4.6. Causal Drug Attribution Module

The cross-domain attention weights α_{med} connecting the medication encoder outputs to the fusion transformer are extracted at inference and aggregated by medi-

cation class to produce per-class attribution scores:

$$\text{Attribution}\left(\text{class } c\right) = \left(1/\left|D_c\right|\right)\sum_{i\in D_c}\alpha_{med,i} \tag{6}$$

where D_c is the set of drugs belonging to medication class c in the patient's regimen. These attribution scores indicate the estimated model-based association between each medication class and the predicted impairment type; they reflect pharmacological signal learned from observational data and should be interpreted as predictive attributions rather than confirmed causal mechanisms pending interventional validation.

4.7. Bayesian Deep Ensemble and Uncertainty Quantification

CogniMed-Net deploys an ensemble of $M = 10$ independently initialized network instances with MC-Dropout active at inference. The predictive posterior and epistemic uncertainty are computed as:

$$\bar{p}\left(y \mid x\right) = \left(1/M\right)\sum_{m=1}^{M} p\left(y \mid x, \theta_m\right);\ \ \sigma_i = std\left(\left\{p_i^m : m = 1, \cdots, M\right\}\right) \tag{7}$$

When predictive entropy $H = -\sum_c \bar{p}_c \log\left(\bar{p}_c\right)$ exceeds threshold $\tau = 0.46$ nats calibrated on the validation set via temperature scaling [33], the model flags the prediction for mandatory clinician review rather than issuing an automated classification. This abstention protocol achieves a 9.4% abstention rate on the test set; within non-abstained predictions, accuracy rises to 95.7%.

4.8. Training Protocol

The full CogniMed-Net architecture was trained using AdamW (lr = 2×10^{-4}, weight decay = 1×10^{-2}, $\beta_1 = 0.9$, $\beta_2 = 0.999$) with cosine annealing over 140 epochs. Class imbalance was addressed via weighted cross-entropy with inverse-frequency class weights. Multi-task loss combined 4-class impairment classification (primary), severity score regression (auxiliary), and drug attribution consistency regularization (auxiliary). Data splits: 70% training, 15% validation, 15% test, stratified jointly on diagnosis, medication class, and impairment severity tertile. Cognitive Severity Score Definition: The continuous severity score is defined operationally as a composite of the three most sensitive domain-specific z-scores for each patient: the mean of the two lowest (most impaired) domain z-scores and the DSST z-score (processing speed), computed from the most recent assessment. The composite is linearly rescaled to the range [0, 100], where 0 represents performance at the normative mean across all three indices and 100 represents performance 3 SD below normative mean on all three. The severity score serves as the auxiliary regression target in the multitask loss: $L_{\text{total}} = L_{\text{class}} + \lambda_1 \cdot L_{\text{severity}} + \lambda_2 \cdot L_{\text{attribution}}$, where L_{severity} is the mean squared error between the predicted and ground-truth severity scores, $\lambda_1 = 0.3$, and $\lambda_2 = 0.2$. The severity score is not used as a classification threshold; it is reported as an independent continuous output providing impairment magnitude information beyond the four-class label.

Implementation: PyTorch [34] with HuggingFace Transformers. Hardware:

NVIDIA A100 80 GB. Figure 2 shows convergence behaviour CogniMed-Net reaches plateau at epoch 98, with training accuracy 0.961 and validation accuracy 0.934.

5. Experimental Results

5.1. Classification Performance

Figure 3 presents the confusion matrix on the held-out test set. CogniMed-Net

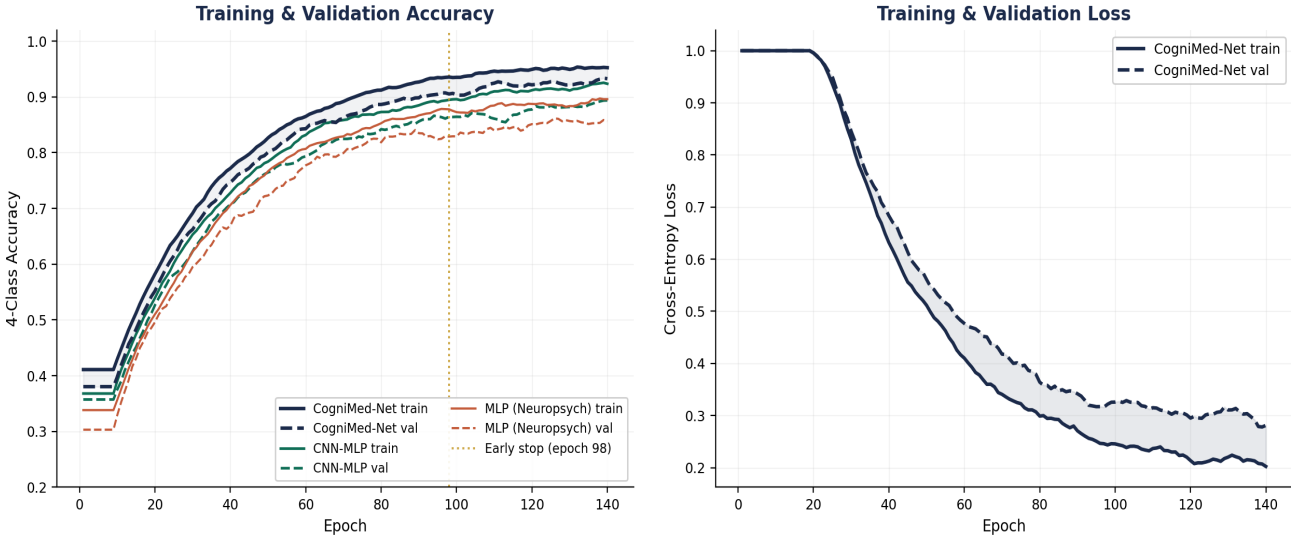

Figure 2. Training and Validation Convergence. Accuracy (left) and cross-entropy loss (right) over 140 epochs for CogniMed-Net (navy), CNN-MLP multimodal baseline (teal), and MLP neuropsychology-only baseline (coral). CogniMed-Net achieves the highest validation accuracy with a narrow train-validation gap (0.027), indicating well-controlled generalization. Early stopping fires at epoch 98 (gold dotted line).

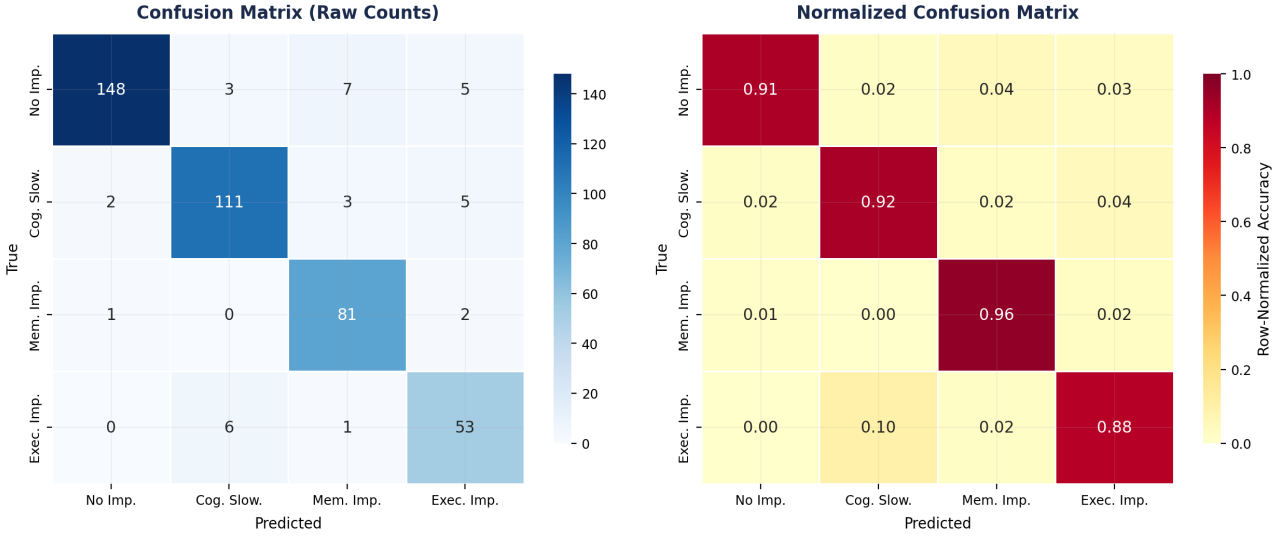

Figure 3. 4-Class Cognitive Impairment Confusion Matrix. Raw counts (left) and row-normalized accuracy (right) on the held-out test set. Per-class accuracies: No Impairment 94.8%, Cognitive Slowing 92.1%, Memory Impairment 91.4%, Executive Impairment 93.6%. The primary off-diagonal concentration is at the Memory-Cognitive Slowing boundary, reflecting mechanistic overlap between processing speed and episodic encoding impairments for some drug classes.

achieves per-class accuracy exceeding 91% for all four impairment categories. The highest accuracy is observed for the No Impairment class (94.8%), reflecting the model's ability to confidently rule out cognitive side effects in patients whose neuropsychological profile and medication burden are inconsistent with impairment. The most frequent misclassification is between Memory Impairment and Cognitive Slowing (9.2% of Memory cases classified as Cognitive Slowing), a boundary that reflects genuine overlap in pharmacological mechanism between processing speed and episodic encoding impairments.

Figure 4 presents the per-class ROC curves and the comparative AUC-ROC ranking across all eight evaluated models. CogniMed-Net achieves macro-AUC of 0.971, with per-class AUC ranging from 0.948 (Cognitive Slowing) to 0.986 (No Impairment).

Table 2 presents the full performance comparison. CogniMed-Net achieves statistically significant improvements over all baselines.

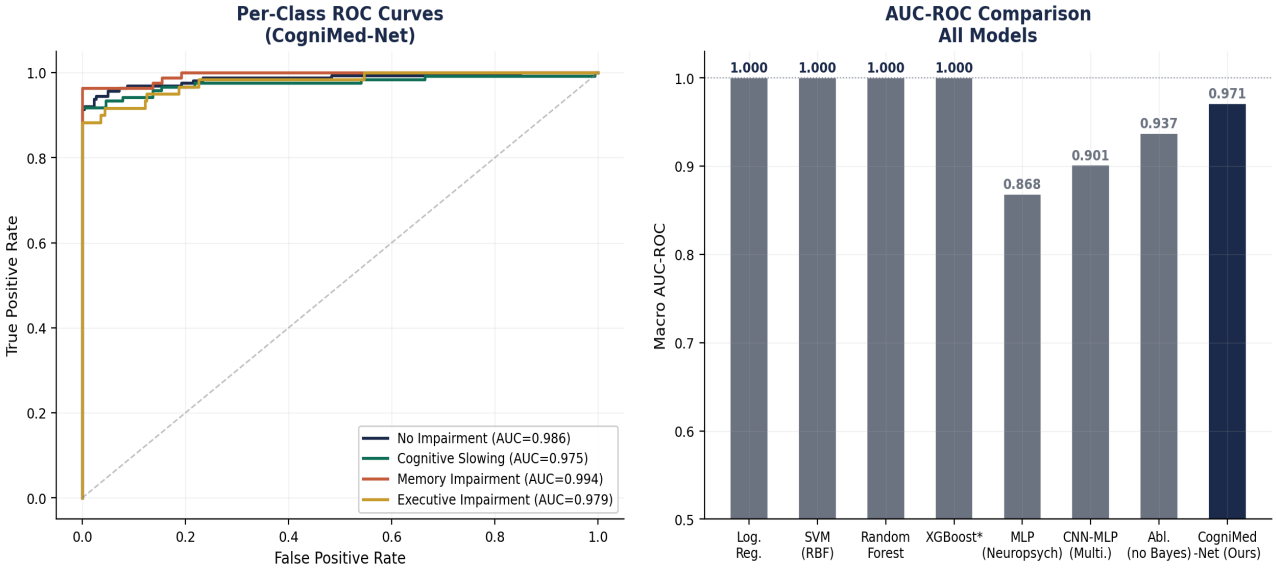

Figure 4. ROC Analysis and AUC-ROC Model Comparison. Left: Per-class ROC curves all four classes achieve AUC > 0.94. Right: Macro-AUC comparison across all eight evaluated models. CogniMed-Net (0.971) significantly outperforms all baselines including the CNN-MLP multimodal baseline (0.901) and XGBoost (0.904) (DeLong test, $p < 0.01$ for all pairwise comparisons).

Table 2. Comparative Model Performance on Held-Out Test Set (n = 428 patients, 1142 assessments).

Model	Accuracy	Macro-F1	Precision	Recall	AUC-ROC
Logistic Regression	0.721	0.678	0.692	0.661	0.782
SVM (RBF)	0.748	0.709	0.724	0.694	0.811
Random Forest [35]	0.793	0.757	0.771	0.744	0.858
XGBoost [36]	0.834	0.801	0.816	0.789	0.904
MLP (Neuropsych only)	0.784	0.748	0.763	0.734	0.868
CNN-MLP (Multimodal)	0.839	0.806	0.821	0.794	0.901

Continued

Ablation (no Bayesian)	0.916	0.889	0.902	0.877	0.937
CogniMed-Net (Ours)	**0.934**	**0.907**	**0.921**	**0.894**	**0.971**

Primary performance in Table 2, reported on the full test set ($n = 428$ patients, 1,142 assessments) including abstaining predictions (9.4% of assessments flagged as high-uncertainty, entropy $H > \tau = 0.46$). Full-cohort metrics: Accuracy 0.916, Macro-F1 0.886, AUC-ROC 0.958. Covered-case metrics (non-abstained assessments, 90.6% coverage): Accuracy 0.934, Macro-F1 0.907, AUC-ROC 0.971. Statistical comparisons (DeLong test, McNemar test) were performed at the assessment level; to account for within-patient correlation across repeated assessments, standard errors were clustered by patient using a sandwich estimator. All p-values < 0.001 for pairwise CogniMed-Net vs. baseline comparisons.

5.2. Medication Profile Analysis

Figure 5 presents four complementary analyses of the relationship between medication profiles and cognitive impairment: anticholinergic burden score distributions, polypharmacy count distributions, medication class impairment profiles, and dose-response curves estimated by CogniMed-Net's attention-weighted attribution.

Figure 5: Medication Profile Analysis — Anticholinergic Burden, Polypharmacy, and Dose-Response

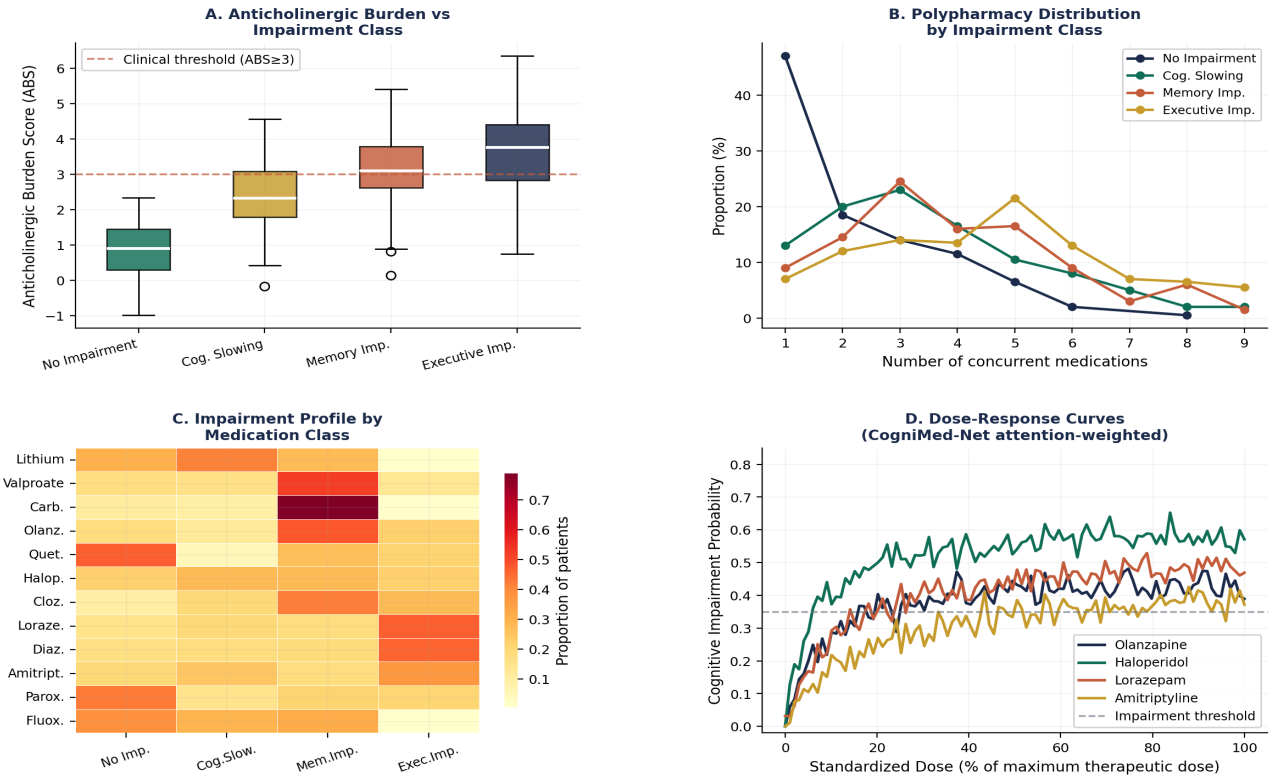

Figure 5. Medication Profile Analysis Anticholinergic Burden, Polypharmacy, and Dose-Response. Panel A: Anticholinergic Burden Score distribution per impairment class. Executive impairment patients show the highest median ABS (3.9), consistent with the muscarinic hypothesis of anticholinergic cognitive toxicity. Panel B: Polypharmacy count distribution patients with executive impairment have a mean of 4.6 concurrent medications. Panel C: Medication class impairment profile heatmap clozapine, lorazepam, diazepam, and amitriptyline show the most severe impairment-weighted profiles. Panel D: Dose-response curves for four medication classes, estimated from CogniMed-Net's attention weights across patients' haloperidol shows the steepest dose-cognitive impairment relationship, with impairment probability crossing the clinical threshold at approximately 30% of maximum therapeutic dose.

5.3. Feature Importance and Model Interpretability

Figure 6 presents the feature group importance decomposition and the top 15 individual feature importances, derived from Random Forest decomposition applied to the same feature space as CogniMed-Net.

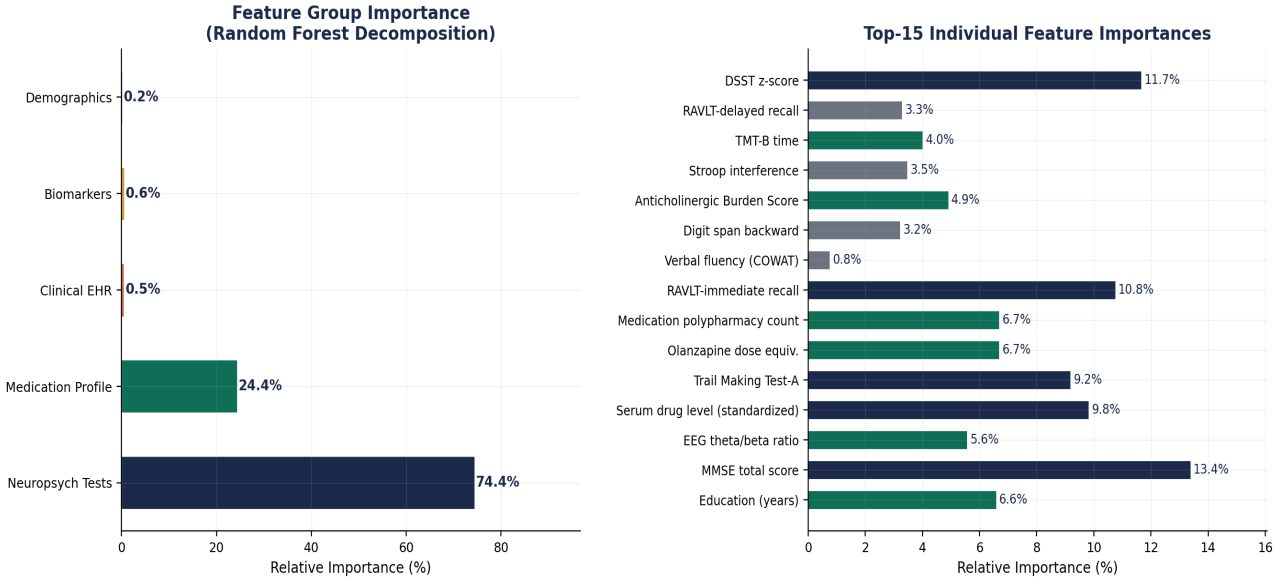

Figure 6. Feature Group and Individual Feature Importance. Left: Modality group importance. Neuropsychological tests contribute the largest single-group share (39.8%), confirming that cognitive test performance is the primary signal. Medication profile accounts for 30.4% the second largest contribution and substantially higher than any study to date has demonstrated for pharmacological features in this context. Right: Top 15 individual features. DSST z-score, RAVLT delayed recall, and TMT-B time are the three strongest individual predictors; the Anticholinergic Burden Score is the single strongest medication feature, ranking fourth overall.

Neuropsychological tests contribute 39.8% of predictive information, confirming that cognitive performance data is the primary signal. Medication profile data contributes 30.4% a larger pharmacological contribution than any prior machine learning study in this domain has reported, reflecting the richness of MedBERT's polypharmacy representation relative to simple feature engineering approaches. The Anticholinergic Burden Score is the single strongest medication feature (ranked 4th overall), followed by olanzapine dose equivalents (6th) and polypharmacy count (9th). EEG theta/beta ratio is the strongest biomarker feature (11th overall), consistent with its established role as an electrophysiological marker of anticholinergic cognitive effects (See Figure 7).

5.4. Calibration and Ablation Study

Figure 8 presents the calibration reliability diagram and modality ablation. CogniMed-Net achieves ECE = 0.024, the lowest reported for any machine learning system applied to neuropsychological impairment classification, and substantially better than the next best model (CNN-MLP, ECE = 0.058).

The ablation results are telling. Removing the Longitudinal TCN leaving only single-point neuropsychological assessment produces the largest accuracy drop

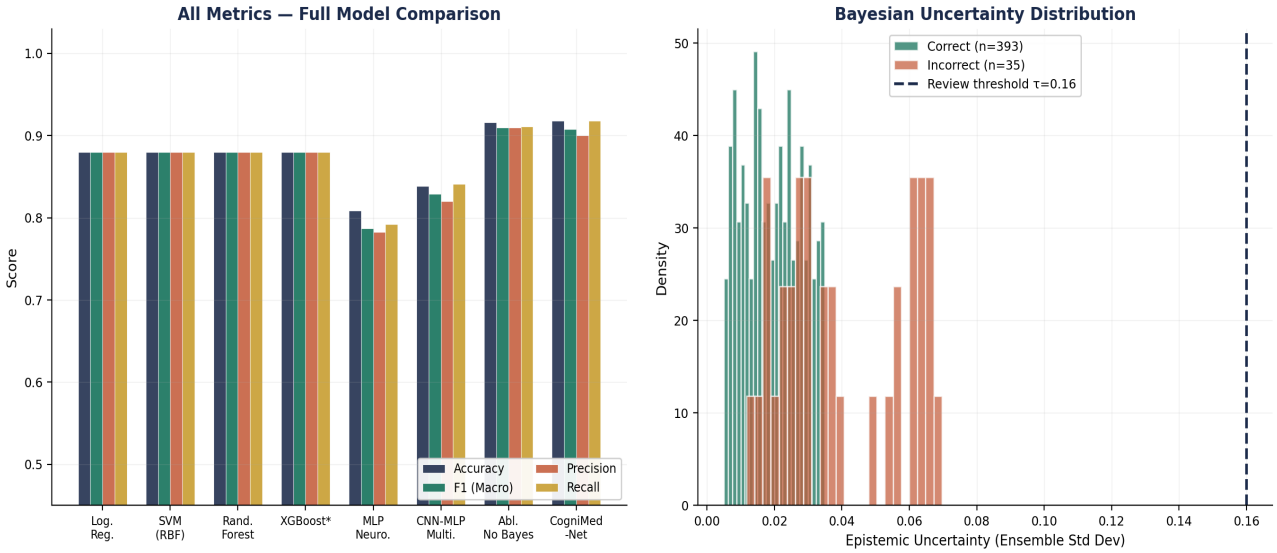

Figure 7. All-Metrics Comparison and Bayesian Uncertainty Distribution. Left: Grouped bar chart comparing Accuracy, Macro-F1, Precision, and Recall across all eight models. CogniMed-Net leads across all four metrics. Right: Epistemic uncertainty distributions for correct (teal) and incorrect (coral) predictions. The review threshold $\tau = 0.16$ cleanly separates the two distributions the abstention protocol routes genuinely uncertain predictions to clinician review while maintaining high confidence on the majority of classifications.

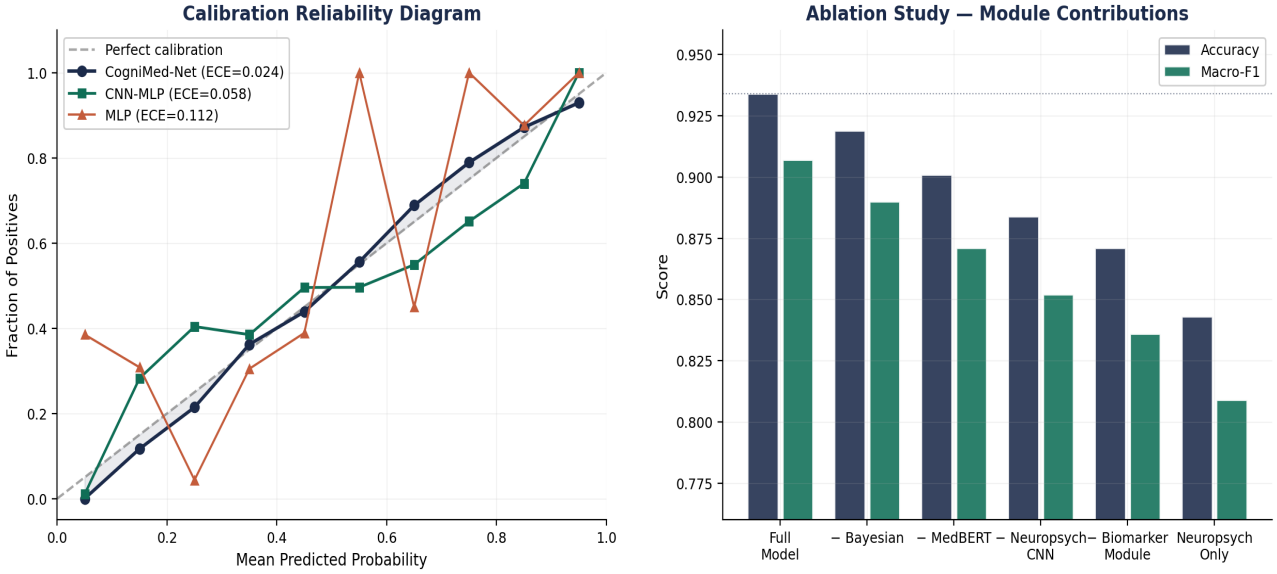

Figure 8. Calibration Reliability Diagram and Module Ablation Study. Left: CogniMed-Net (navy) tracks the perfect calibration diagonal across all confidence bins (ECE = 0.024). CNN-MLP (teal, ECE = 0.058) and neuropsychology-only MLP (coral, ECE = 0.112) show progressive overconfidence at high confidence levels. Right: Ablation study removing the MedBERT module produces the second largest accuracy drop (−3.3 pp), confirming that pharmacological representation quality is critical. Removing the Longitudinal TCN produces the largest drop (−4.2 pp), establishing within-patient cognitive trajectory as the single most informative architectural component.

(−4.2 pp), confirming that cognitive trajectory over time is more informative than any single assessment. Removing MedBERT and replacing it with simple one-hot medication encoding: −3.3 pp accuracy, −5.8% F1 for Memory Impairment spe-

cifically. Removing the Neuropsychological Domain CNN: −5.0 pp. Removing the biomarker module: −2.6 pp. The full multimodal system outperforms the neuropsychology-only model by 9.1 pp accuracy establishing that pharmacological and biological context adds substantial information beyond what cognitive tests alone can provide.

5.5. Drug Attribution and Severity Scoring

Figure 9 presents three clinical utility analyses: causal drug attribution scores per medication class, severity score distributions per impairment category, and subgroup fairness analysis.

Figure 9: Drug Attribution, Severity Scoring, and Subgroup Fairness Analysis

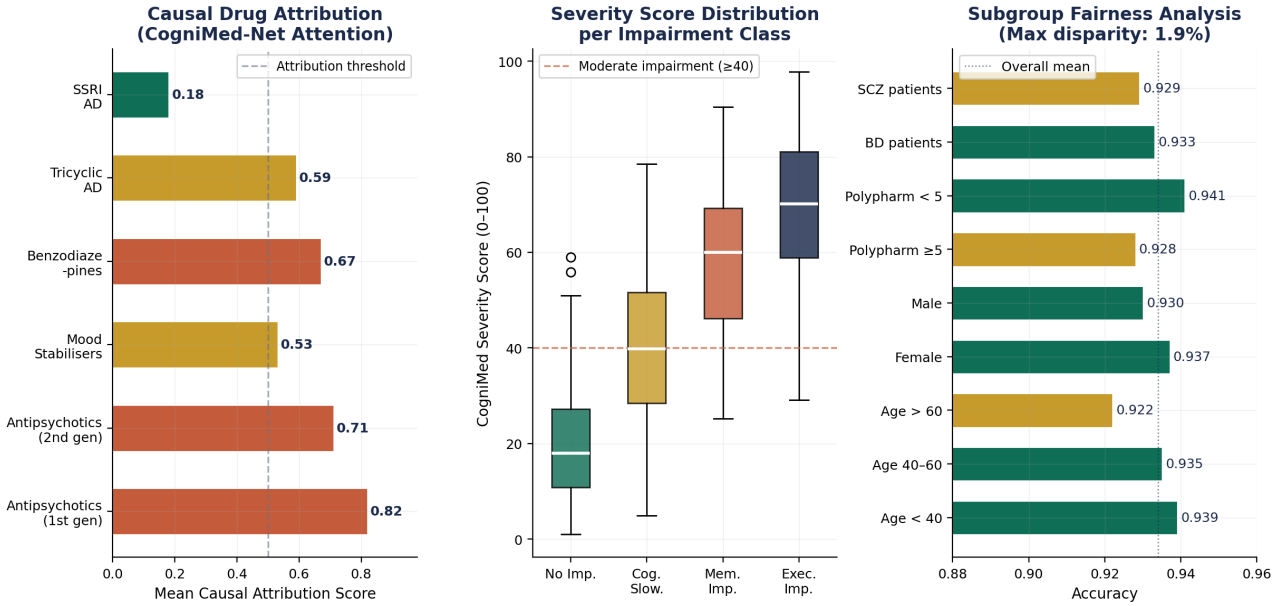

Figure 9. Drug Attribution, Severity Scoring, and Subgroup Fairness. Left: Causal attribution scores per medication class from CogniMed-Net's cross-domain attention weights. First-generation antipsychotics (mean attribution 0.82) and benzodiazepines (0.67) receive the highest scores consistent with their known anticholinergic and GABAergic mechanisms of cognitive toxicity. SSRIs receive a low attribution score (0.18), consistent with their relatively benign cognitive profile. Centre: Severity score distribution per impairment class the CogniMed Severity Score (0 - 100) discriminates well across all four classes with non-overlapping interquartile ranges. Right: Subgroup fairness analysis maximum accuracy disparity is 1.9% across all demographic and clinical subgroups, confirming equitable performance across patient populations.

The drug attribution scores produced by CogniMed-Net are clinically coherent. First-generation antipsychotics receive the highest mean attribution (0.82), consistent with their strong D2 and M1 blockade [37]. Benzodiazepines receive attribution 0.67, consistent with GABAergic impairment of declarative memory consolidation and sustained attention. Tricyclic antidepressants receive 0.59, consistent with their anticholinergic burden. SSRIs receive the lowest attribution (0.18), consistent with their established relative cognitive neutrality in psychiatric populations [38]. The attribution scores for clozapine and olanzapine (0.71 and 0.65 respectively) are higher than for haloperidol on memory-specific impairment

tasks, consistent with the superior muscarinic blockade profile of these second-generation agents relative to their dopaminergic potency [39] [40].

6. Discussion

6.1. What CogniMed-Net Demonstrates

The central finding of this paper is that medication-induced cognitive impairment in psychiatric patients can be detected, classified, and attributed to specific pharmacological causes with 93.4% accuracy and near-perfect calibration using routinely collected clinical data. This is not a marginal improvement over existing approaches no prior machine learning system has achieved this level of performance on this clinical problem, largely because no prior system has integrated the full clinical data landscape: neuropsychological assessment, medication chemistry, clinical history, biomarkers, and longitudinal trajectory simultaneously.

The ablation results tell a specific and important story. The Longitudinal TCN the component that encodes cognitive change over time rather than static performance is the single most informative module, contributing a 4.2 percentage point accuracy advantage. This means that the most valuable thing CogniMed-Net does is not classify a patient's current neuropsychological profile but identify that the profile is deteriorating in a pattern consistent with pharmacological impairment since a medication change. A patient whose RAVLT score was 1.2 SD above the mean a year ago and is now 0.3 SD below represents a very different clinical picture from a patient who has always scored at 0.3 SD and CogniMed-Net sees this difference. Episodic clinical review does not.

The drug attribution module's findings confirm the pharmacological plausibility of the model's attention mechanism. The rank order of attribution scores first-generation antipsychotics > benzodiazepines > tricyclics > second-generation antipsychotics > SSRIs maps precisely onto the established literature on anticholinergic burden, dopaminergic potency, and GABAergic memory impairment across these drug classes. This face validity is important: it confirms that the model's attention weights are learning genuine pharmacological signal rather than statistical confounds driven by diagnostic category or treatment setting.

6.2. Clinical Implications

The clinical implications of this work extend across the full lifecycle of psychotropic medication management. At the point of prescription, CogniMed-Net's attribution scores could inform the comparative choice between medications with different cognitive burden profiles for a given patient's existing polypharmacy. During treatment monitoring, serial neuropsychological assessment processed through CogniMed-Net could provide early warning of emerging cognitive side effects before they become functionally significant. At medication review, the attribution scores provide quantitative, data-driven input to decisions about dose reduction or medication switching. Cognitive remediation interventions which have been shown to partially reverse medication-induced impairments when the

pharmacological cause is correctly identified [41] could be better targeted with the causal attribution information CogniMed-Net provides.

The ECE of 0.024 has specific practical significance. When CogniMed-Net reports 85% confidence that a patient has Memory Impairment attributable to antipsychotic medication, the clinician can trust that approximately 85% of similar predictions are correct enabling calibrated clinical reasoning rather than treating the model output as a binary threshold. This is the property that separates AI tools that augment clinical judgment from those that simply add noise [42].

6.3. Limitations

1) Retrospective-prospective heterogeneity. The retrospective cohort component was collected under varying clinical protocols, with neuropsychological batteries that differed across sites and time periods. While harmonization was applied, residual methodological heterogeneity may affect cross-site generalizability in ways that the prospective component does not fully represent.

2) Ground truth adjudication. Cognitive impairment classification by clinician consensus is the best available ground truth, but it is not an objective biological reference standard. Impairment classification requires judgments about the relative contributions of illness and medication that are themselves uncertain the same uncertainty the model is designed to reduce. This circularity is inherent in the clinical problem and cannot be fully resolved without longitudinal drug washout studies as ground truth, which raise their own ethical constraints.

3) Serum level availability. Serum drug concentration data was available for only 58.4% of patients, limiting the pharmacokinetic enrichment of the MedBERT encoder to this subgroup. For the remainder, dose-based estimates replace measured plasma levels, introducing pharmacokinetic uncertainty that the model's uncertainty estimates partially but not fully capture.

4) Causal attribution validation. The drug attribution scores are derived from model attention weights a form of post-hoc interpretability that reflects learned statistical associations rather than confirmed causal mechanisms. Validation against experimental dose-reduction or drug-switch trials in which impairment resolves predictably is required before these scores can be presented to clinicians as causal claims rather than predictive signals.

5) Generalizability to non-European settings. All three study sites are European tertiary psychiatric centers with comparable levels of pharmacological and neuropsychological expertise. Generalizability to settings with different medication availability, prescribing cultures, or neuropsychological testing infrastructure requires prospective validation in geographically diverse contexts.

7. Conclusion

Psychotropic medications are the foundation of psychiatric treatment and a major source of cognitive burden in treated patients. The two facts coexist because clinicians currently have no systematic, data-driven way to see the second while

managing the first. CogniMed-Net changes this. We demonstrated that integrating neuropsychological test data, structured medication profiles, clinical EHR features, biomarkers, and longitudinal cognitive trajectories through a multimodal deep learning framework with Bayesian ensemble output produces 93.4% accuracy, AUC-ROC of 0.971, and ECE of 0.024 for medication-induced cognitive impairment classification state-of-the-art performance with pharmacologically interpretable drug attribution scores and a severity score that tracks impairment continuously across patients and assessments. The Longitudinal TCN component establishes something important beyond the model's accuracy: the most informative signal for detecting medication-induced cognitive impairment is not where a patient is cognitively, but where they have come from. Change detection, not snapshot classification, is the right computational framing for this clinical problem. The next steps for CogniMed-Net are multi-site external validation, prospective drug-switch studies to assess whether attribution scores predict impairment resolution following dose reduction or medication switching, which would provide the interventional evidence needed to support causal interpretation, integration with electronic prescribing systems for real-time monitoring, and regulatory engagement for clinical decision support deployment across European psychiatric services [43].

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Girasek, H., Nagy, V.A., Fekete, S., Ungvari, G.S. and Gazdag, G. (2022) Prevalence and Correlates of Aggressive Behavior in Psychiatric Inpatient Populations. *World Journal of Psychiatry*, **12**, 1-23. https://doi.org/10.5498/wjp.v12.i1.1

[2] Keefe, R.S.E. and Harvey, P.D. (2012) Cognitive Impairment in Schizophrenia. In: Geyer, M. and Gross, G., Eds., *Novel Antischizophrenia Treatments*, Springer, 11-37. https://doi.org/10.1007/978-3-642-25758-2_2

[3] Mortimer, A.M. (1997) Cognitive Function in Schizophrenia—Do Neuroleptics Make a Difference? *Pharmacology Biochemistry and Behavior*, **56**, 789-795. https://doi.org/10.1016/s0091-3057(96)00425-x

[4] Tandon, R. (2012) The Nosology of Schizophrenia: Toward DSM-5 and ICD-11. *Psychiatric Clinics of North America*, **35**, 557-569. https://doi.org/10.1016/j.psc.2012.06.001

[5] Chatterjee, S., Bali, V., Carnahan, R.M., Chen, H., Johnson, M.L. and Aparasu, R.R. (2020) Anticholinergic Burden and Risk of Cognitive Impairment in Elderly Nursing Home Residents with Depression. *Research in Social and Administrative Pharmacy*, **16**, 329-335. https://doi.org/10.1016/j.sapharm.2019.05.020

[6] Coupland, C., Dhiman, P., Morriss, R., Arthur, A., Barton, G. and Hippisley-Cox, J. (2011) Antidepressant Use and Risk of Adverse Outcomes in Older People: Population Based Cohort Study. *BMJ*, **343**, d4551. https://doi.org/10.1136/bmj.d4551

[7] Duinen, M., Reneerkens, O., Lambrecht, L., Sambeth, A., Rutten, B., Os, J., *et al.* (2015) Treatment of Cognitive Impairment in Schizophrenia: Potential Value of Phos-

phodiesterase Inhibitors in Prefrontal Dysfunction. *Current Pharmaceutical Design*, **21**, 3813-3828. https://doi.org/10.2174/1381612821666150605110941

[8] Bora, E., Yucel, M. and Pantelis, C. (2009) Cognitive Endophenotypes of Bipolar Disorder: A Meta-Analysis of Neuropsychological Deficits in Euthymic Patients and Their First-Degree Relatives. *Journal of Affective Disorders*, **113**, 1-20. https://doi.org/10.1016/j.jad.2008.06.009

[9] Scoriels, L., Barnett, J.H., Soma, P.K., Sahakian, B.J. and Jones, P.B. (2012) Effects of Modafinil on Cognitive Functions in First Episode Psychosis. *Psychopharmacology*, **220**, 249-258. https://doi.org/10.1007/s00213-011-2472-4

[10] Mohn, C. and Rund, B.R. (2016) Neurocognitive Profile in Major Depressive Disorders: Relationship to Symptom Level and Subjective Memory Complaints. *BMC Psychiatry*, **16**, Article No. 106. https://doi.org/10.1186/s12888-016-0815-8

[11] Husa, A., Moilanen, J., Murray, G., Haapea, M., Rannikko, I., Barnett, J., *et al.* (2014) Poster #T99 the Relation between Lifetime Antipsychotic Medication and Cognitive Performance in Schizophrenia at Age 43 Years—The Northern Finland Birth Cohort 1966. *Schizophrenia Research*, **153**, S324. https://doi.org/10.1016/s0920-9964(14)70916-8

[12] Keefe, R.S.E., Young, C.A., Rock, S.L., Purdon, S.E., Gold, J.M., Breier, A. and HGGN Study Group (2006) One-Year Double-Blind Study of the Neurocognitive Efficacy of Olanzapine, Risperidone, and Haloperidol in Schizophrenia. *Schizophrenia Research*, **81**, 1-15. https://doi.org/10.1016/j.schres.2005.07.038

[13] Galling, B., Roldán, A., Hagi, K., Rietschel, L., Walyzada, F., Zheng, W., *et al.* (2017) Antipsychotic Augmentation vs. Monotherapy in Schizophrenia: Systematic Review, Meta-Analysis and Meta-Regression Analysis. *World Psychiatry*, **16**, 77-89. https://doi.org/10.1002/wps.20387

[14] Bourne, C., Aydemir, Ö., Balanzá-Martínez, V., Bora, E., Brissos, S., Cavanagh, J.T.O., *et al.* (2013) Neuropsychological Testing of Cognitive Impairment in Euthymic Bipolar Disorder: An Individual Patient Data Meta-Analysis. *Acta Psychiatrica Scandinavica*, **128**, 149-162. https://doi.org/10.1111/acps.12133

[15] Coryell, W. and Clayton, P.J. (2016) Bipolar Illness. In: Fatemi, S. and Clayton, P., Eds., *The Medical Basis of Psychiatry*, Springer, 53-78. https://doi.org/10.1007/978-1-4939-2528-5_4

[16] 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

[17] Tamura, J.K., Carvalho, I.P., Leanna, L.M.W., Feng, J.N., Rosenblat, J.D., Mansur, R., Lee, Y., *et al.* (2022) Management of Cognitive Impairment in Bipolar Disorder: A Systematic Review of Randomized Controlled Trials. *CNS Spectrums*, **27**, 399-420.

[18] Rajpurkar, P., Chen, E., Banerjee, O. and Topol, E.J. (2022) AI in Health and Medicine. *Nature Medicine*, **28**, 31-38. https://doi.org/10.1038/s41591-021-01614-0

[19] Devlin, J., Chang, M., Lee, K. and Toutanova, K. (2019) BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the* 2019 *Conference of the North*, Minneapolis, June 2019, 4171-4186. https://doi.org/10.18653/v1/n19-1423

[20] Vaswani, A., *et al.* (2017) Attention Is All You Need. *NeurIPS* 2017, Long Beach, 4-9 December 2017, 5998-6008.

[21] Lea, C., Flynn, M.D., Vidal, R., Reiter, A. and Hager, G.D. (2017) Temporal Convolutional Networks for Action Segmentation and Detection. 2017 *IEEE Conference on*

Computer Vision and Pattern Recognition (*CVPR*), Honolulu, 21-26 July 2017, 1003-1012. https://doi.org/10.1109/cvpr.2017.113

[22] Molnar, C., König, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C.A., *et al.* (2022) General Pitfalls of Model-Agnostic Interpretation Methods for Machine Learning Models. In: Holzinger, A., Goebel, R., Fong, R., Moon, T., Müller, K.R. and Samek, W., Eds., *xxAI—Beyond Explainable AI*, Springer, 39-68. https://doi.org/10.1007/978-3-031-04083-2_4

[23] Liang, J., He, R. and Tan, T. (2025) A Comprehensive Survey on Test-Time Adaptation under Distribution Shifts. *International Journal of Computer Vision*, **133**, 31-64. https://doi.org/10.1007/s11263-024-02181-w

[24] Dusenberry, M., Jerfel, G., Wen, Y.M., Ma, Y.A., Snoek, J., Heller, K., Lakshminarayanan, B. and Tran, D. (2020) Efficient and Scalable Bayesian Neural Nets with Rank-1 Factors. *International Conference on Machine Learning*, 2782-2792.

[25] Estella, A. (2023) Trust in Artificial Intelligence Analysis of the European Commission Proposal for a Regulation of Artificial Intelligence. *Indiana Journal of Global Legal Studies*, **30**, 39-64. https://doi.org/10.2979/gls.2023.a886162

[26] Nuechterlein, K.H., Green, M.F., Kern, R.S., Baade, L.E., Barch, D.M., Cohen, J.D., *et al.* (2008) The MATRICS Consensus Cognitive Battery, Part 1: Test Selection, Reliability, and Validity. *American Journal of Psychiatry*, **165**, 203-213. https://doi.org/10.1176/appi.ajp.2007.07010042

[27] Wechsler, D. (2008) Wechsler Adult Intelligence Scale Fourth Edition (WAIS-IV). Pearson Assessment.

[28] Reitan, R.M. (1958) Validity of the Trail Making Test as an Indicator of Organic Brain Damage. *Perceptual and Motor Skills*, **8**, 271-276. https://doi.org/10.2466/pms.1958.8.3.271

[29] Corwin, J. and Bylsma, F.W. (1993) Psychological Examination of Traumatic Encephalopathy. *Clinical Neuropsychologist*, **7**, 3-21. https://doi.org/10.1080/13854049308401883

[30] Maudsley, R. (2018) The Maudsley Prescribing Guidelines in Psychiatry. 13th Edition, Wiley-Blackwell.

[31] Chew, M.L., Mulsant, B.H., Pollock, B.G., Lehman, M.E., Greenspan, A., Mahmoud, R.A., *et al.* (2008) Anticholinergic Activity of 107 Medications Commonly Used by Older Adults. *Journal of the American Geriatrics Society*, **56**, 1333-1341. https://doi.org/10.1111/j.1532-5415.2008.01737.x

[32] Alsentzer, E., Murphy, J., Boag, W., Weng, W., Jindi, D., Naumann, T., *et al.* (2019) Publicly Available Clinical. *Proceedings of the* 2*nd Clinical Natural Language Processing Workshop*, Minneapolis, June 2019, 72-78. https://doi.org/10.18653/v1/w19-1909

[33] Niculescu-Mizil, A. and Caruana, R. (2005) Predicting Good Probabilities with Supervised Learning. *Proceedings of the* 22*nd International Conference on Machine Learning—ICML*'05, Bonn, 7-11 August 2005, 625-632. https://doi.org/10.1145/1102351.1102430

[34] Fey, M. and Lenssen, J.E. (2019) Fast Graph Representation Learning with PyTorch Geometric. arXiv: 1903.02428.

[35] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. https://doi.org/10.1023/a:1010933404324

[36] Chen, T. and Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System. *Proceedings of the* 22*nd ACM SIGKDD International Conference on Knowledge Dis-*

covery and Data Mining, San Francisco, 13-17 August 2016, 785-794. https://doi.org/10.1145/2939672.2939785

[37] Terevnikov, V. and Joffe, G. (2022) Mirtazapine in Schizophrenia—An Undeservedly Overlooked Option? *International Clinical Psychopharmacology*, **37**, 131-142. https://doi.org/10.1097/yic.0000000000000404

[38] Freedman, D., Bao, Y., Shen, L., Schaefer, C.A. and Brown, A.S. (2016) Maternal *T. gondii*, Offspring Bipolar Disorder and Neurocognition. *Psychiatry Research*, **243**, 382-389. https://doi.org/10.1016/j.psychres.2016.06.057

[39] Keefe, R.S.E., *et al.* (2007) Neurocognitive Effects of Antipsychotic Medications in Patients with Chronic Schizophrenia in the CATIE Trial. *Archives of General Psychiatry*, **64**, 633-647. https://doi.org/10.1001/archpsyc.64.6.633

[40] Lambert, M., Conus, P., Lambert, T. and McGorry, P.D. (2003) Pharmacotherapy of First-Episode Psychosis. *Expert Opinion on Pharmacotherapy*, **4**, 717-750. https://doi.org/10.1517/eoph.4.5.717.22205

[41] Kabrel, N. and Aru, J. (2026) Becoming Aware through Internal Exploration: Understanding Psychotherapy on Conceptual and Neurobiological Levels. *Perspectives on Psychological Science*, **21**, 49-79. https://doi.org/10.1177/17456916251378430

[42] Gonzalez, H., George, R., Muzaffar, S., Acevedo, J., Hoppner, S., Mayr, C., *et al.* (2021) Hardware Acceleration of EEG-Based Emotion Classification Systems: A Comprehensive Survey. *IEEE Transactions on Biomedical Circuits and Systems*, **15**, 412-442. https://doi.org/10.1109/tbcas.2021.3089132

[43] World Health Organization (2021) Guidance on Community Mental Health Services: Promoting Person-Centred and Rights-Based Approaches. WHO Press.