Differential Diagnosis of Adult Neurodevelopmental Disorders: Systematic Review and Development of a Rule-Based Expert System

Micila Sumaria Medeiros Pereira Magalhães[1], Rodrigo Bernado Domingos Rodrigues[2], Mário António Afonso Cardoso[3], Fábio José Dias Girão[4], and Gonçalo Fernandes[5]

ISEP, Politécnico do Porto, Rua Dr. António Bernardino de Almeida, 4249-015, Porto, Portugal
{1231812,1250549,1191568,1201146,1201496}@isep.ipp.pt

Abstract. Neurodevelopmental Disorders (NDDs) in adults — most notably Autism Spectrum Disorder (ASD) and Attention-Deficit/ Hyperactivity Disorder (ADHD) — remain challenging to diagnose due to overlapping symptoms and comorbidities. This study examines artificial intelligence (AI) approaches applied to adult DDx, which converge on two main directions: rule-based systems for workflow management (CDSS) and machine-learning (ML) models aimed at efficient triage and classification. Based on this analysis, we propose a transparent, rule-based expert system for adults, complemented by optional ML components to support prioritization within a dual-professional workflow (Neurology + Psychology). The system seeks to improve diagnostic consistency and explainability while maintaining clinical oversight and ensuring full decision traceability. The contribution is an explainable and pragmatic model that combines expert rules with lightweight ML to provide reliable and auditable decision support for adult NDD DDx.

Keywords: Neurodevelopmental Disorders (NDD); Differential Diagnosis (DDx); Rule-Based Systems; Business Process Management (BPM); Clinical Decision Support Systems (CDSS); Explainability (XAI); Machine Learning (ML); Adults.

1 Introduction

The adult DDx of NDDs—notably autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD) — is challenging due to symptom overlap, comorbidities, and heterogeneous developmental trajectories. Consequently, there is demand for approaches that standardize the diagnostic pathway, shorten time to diagnosis, and support decisions with traceability and clinical explainability. Recent literature converges on two broad families of AI support for adult DDx: (i) rule-based clinical decision support systems (CDSS) integrated with business process management (BPM), which operationalize staged workflows (e.g., triage →

interviews/tests → consensus meeting → report) and generate automatic summaries; and (ii) machine-learning (ML) pipelines that select minimal variable subsets (instrument items, neuropsychological tests, social cognition measures) for classification and case prioritization [1–5].

In practice, a rule-based CDSS for adult ADHD has been deployed in a real clinical service and coupled to BPM, reporting qualitative gains in standardization and process transparency—yet without published quantitative diagnostic metrics (e.g., area under the curve (AUC), sensitivity, specificity) and without external validation [3]. In parallel, clinical ML studies show that abbreviated triage can retain performance close to longer protocols: in adolescents/adults with suspected ASD, models using a few items from the Autism Diagnostic Observation Schedule, Module 4 (ADOS-4) achieve AUC $\approx$ 0.82–0.87, near 11/31-item versions, highlighting nonverbal social cues (e.g., eye contact, directed facial expressions) and the quality of reciprocal social communication as discriminative axes [4, 2]. For DDx among ASD, early psychosis, and social anxiety, different algorithms—RF, Lasso/Elastic Net, and Bayesian Additive Regression Trees (BART)—that combine social cognition (e.g., Reading the Mind in the Eyes Test; RMET), executive function, attention/memory, and mood (e.g., Depression Anxiety Stress Scales; DASS-21) report AUC $\approx$ 0.72–0.92 depending on the task [1]. Scope reviews confirm the potential of artificial intelligence (AI) while highlighting persistent gaps: scarcity of public datasets (outside some imaging corpora), limited external validation, underused Explainable Artificial Intelligence (XAI), incipient interoperability (rare Electronic Health Record integration via HL7-FHIR), and limited impact evaluation (time to diagnosis, cost-effectiveness, clinician acceptance). Pragmatic trials (e.g., DECIDE-AI/CONSORT-AI), multimodality, and explainable, workflow-ready solutions are recommended [5].

This paper is organized into five main sections: introduction, theoretical framework, state of the art, presentation of the project Rule-Based Expert System for Adult NDD Differential Diagnosis, discussion, and conclusions.

2 State of the Art

There is increasing interest in integrating domain expertise and AI to enhance clinical decision-making for the differential diagnosis (DDx) of neurodevelopmental disorders (NDDs) in adults. This section summarizes the main approaches, reports representative evidence, and highlights open gaps supported by the recent literature.

2.1 State of the Art Methodology

To identify and analyze the main scientific evidence on the use of **Artificial Intelligence (AI)** applied to the **differential diagnosis of Neurodevelopmental Disorders (NDD)** in adults, a specific research protocol was developed, based on an **adapted PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)** model. This protocol guided all methodological stages, from the formulation of the research questions to the final screening and inclusion of studies, ensuring methodological rigor, transparency, and replicability.

Step 1 – Definition of Focus and Research Questions. The review was structured around four key dimensions:

- **Population:** adults ($\geq$ 19 years) with suspected or confirmed NDDs, including Autism Spectrum Disorder (ASD), Attention Deficit Hyperactivity Disorder (ADHD), Dyslexia, Developmental Coordination Disorder (DCD), and Tic/Tourette syndromes.
- **Intervention/Technology:** AI-based systems applied to differential diagnosis, such as machine learning (ML), deep learning (DL), expert systems, neural networks, natural language processing (NLP), neuro-symbolic models, Bayesian inference, decision trees, fuzzy logic, and clinical decision support (CDS) systems.
- **Context:** clinical, psychological, and neurological settings, as well as computational validation studies with clinical intent.
- **Outcomes:** diagnostic performance (sensitivity, specificity, AUC, F1-score), diagnostic time reduction, workflow efficiency, explainability (XAI), and clinical integration.

The research questions were structured as follows:

1. What is the current state of the art of AI in the differential diagnosis of NDDs in adults?
2. How have AI systems been designed and implemented for this purpose (workflows, data inputs, metrics, and clinical integration)?
3. Which paradigms, architectures, and business rules have been used (expert systems, Bayesian models, ML/DL, NLP, neuro-symbolic or hybrid models, ontologies, and coded screening guides)?
4. What gaps remain in this field (bias, generalization, interpretability, clinical validation, interoperability, and ethical–legal compliance)?

Rigorous inclusion and exclusion criteria were applied to systematically filter the literature and retain only the most relevant and methodologically robust studies. This process ensured that the selected references were high-quality, representative, and suitable to support the contributions of this work.

A total of **13 records** were identified: 9 from Web of Science, 2 from IEEE Xplore, and 2 from PubMed. The inclusion criteria were:

- Adult population ($\geq$ 19 years);
- Application of AI to differential or screening diagnosis of NDDs;
- Accuracy, validation, scoping/systematic reviews, or prototype studies with quantitative evaluation;
- Publication years between 2020 and 2025, in English, Portuguese, or Spanish.

The exclusion criteria were:

- Pediatric-only studies without adult sub-analysis;
- AI applied exclusively to treatment (non-diagnostic);
- Reports without quantitative evaluation;
- Case series with $<$ 10 participants;
- Opinion papers or studies without empirical data.

All screening steps were conducted independently by two reviewers to ensure reliability and consistency in study selection.

2.2 AI Support Families for Adult NDD Differential Diagnosis

Rule-Based CDSS Integrated with BPM. **Definition and process:** Knowledge-based CDSS encode clinical rules and orchestrate workflows (tasks, roles, deadlines), usually with *gates* that require completed steps before the consensus meeting and report issuance.
Representative example: an adult ADHD CDSS integrated with BPM (Serena Business Manager) that produces automatic summaries and coordinates a multidisciplinary team; qualitative gains in standardization/efficiency were reported, but no AUC/sensitivity/specificity have been published and no external validation was conducted [3].

ML Pipelines for Triage and DDx. **Definition and process:** pipelines comprising pre-processing (recoding, imputation, normalization), feature selection, training/validation (k-fold, hold-out), and testing, yielding classifiers that support DDx. **ASD in adolescents/adults:** radial Support Vector Machine (SVM) and Round Forrest (RF) with minimal ADOS-4 item subsets achieve AUC $\approx$ 0.82–0.87, close to longer algorithms; nonverbal social signals and reciprocity quality are central; external validation is lacking and circularity is possible (the reference diagnosis includes instruments also used as predictors) [4, 2]. **DDx ASD vs. early psychosis vs. social anxiety:** BART, RF, and Lasso/Elastic Net, combining social cognition, executive function, attention/memory, and mood, report AUC $\approx$ 0.72–0.92, without clinical deployment or external validation [1]. **ADHD synthesis:** a thematic review confirms the predominance of ML (SVM, RF) over deep learning (DL) in adults, reliance on private datasets, methodological heterogeneity, scarce XAI, and rare EHR/HL7-FHIR integration; it recommends pragmatic trials (DECIDE-AI/CONSORT-AI) and multimodality (including wearables) [5].

2.3 Research Questions and Synthesis of the Evidence

RQ1: What is the state of the art of AI for adult NDD differential diagnosis? Two axes prevail: rule-based CDSS already in operation (process standardization and transparency, but no published diagnostic metrics and no external validation) and ML pipelines with promising performance (for ASD, AUC $\approx$ 0.82–0.87 using few ADOS-4 items; for ASD–psychosis–anxiety DDx, AUC $\approx$ 0.72–0.92), yet mostly without clinical deployment and without external validation [3, 4, 2, 1]. Reviews emphasize potential and gaps (datasets, XAI, integration) [5].
RQ2: How have AI systems been designed and implemented for adult DDx (workflows, inputs, metrics, clinical integration)? CDSS+BPM: chained workflows (triage $\rightarrow$ assessments $\rightarrow$ consensus $\rightarrow$ report), business rules with gates, and automatic summaries; real service integration, but no published AUC/sensitivity/specificity and no external validation [3]. ML (research): rigorous preprocessing, feature selection (e.g., RFE, Boruta, Lasso), k-fold/holdout validation, and metrics such as AUC/accuracy/

sensitivity/specificity; inputs combine clinical instruments (e.g., ADOS-4, utism Diagnostic Interview-Revised (ADI-R), Adult ADHD Self-Report Scale (ASRS), Diagnostic Interview for ADHD in adults (DIVA-5)), neuropsychological tests, social cognition (e.g., RMET), and mood (DASS21); EHR/HL7 FHIR integration is rare [4, 2, 1, 5].

RQ3: Which paradigms, architectures, and business rules are used? Paradigms: rule-based (expert systems) in real settings; classical ML (SVM, RF, Lasso/Elastic Net, BART) dominating adult clinical studies; DL is a minority in this adult slice; Natural Language Processing (NLP), neurosymbolic, and ontology-based approaches are scarcely represented [3, 4, 2, 1, 5]. Business rules: explicit in CDSS (workflows, gates, mandatory scales) and implicit in ML pipelines via minimal item subsets and cutoffs (e.g., Youden's J), which can be encoded as clinical triage/referral rules [3, 2, 4].

RQ4: Which gaps remain (data, generalization, bias/fairness, interpretability, clinical validation, interoperability, ethics)? *Data and generalization:* predominantly single-site samples, high-functioning bias, lack of external validation, and risk of circularity (reference diagnosis includes instruments used as predictors) [4, 2]. *Bias and fairness:* subgroup analyses by sex, age, and comorbidities are rare; fairness is not systematically assessed [2, 4]. *Explainability (XAI):* limited use of SHapley Additive exPlanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME), attention, or counterfactuals; variable-importance narratives predominate [1, 4, 2]. *Clinical validation and impact:* pragmatic trials (DECIDE-AI/CONSORT-AI), time to diagnosis, workload, and cost-effectiveness are seldom measured; the operational CDSS lacks published diagnostic metrics [3, 5]. *Interoperability and ethics:* EHR/HL7 FHIR integration is rare; privacy and consent are described at a high level, with little evidence of privacy-by-design architectures [3, 5].

3 Project: Rule-Based Expert System for Adult NDD Differential Diagnosis

3.1 Overview and Motivation

This project aims to build a *C*DSS for the DDx of NDDs in **adults**, with an initial focus on ASD and ADHD. The motivation is to standardize the diagnostic pathway, reduce uncertainty and assessment time, and provide traceable clinical evidence, without replacing professional judgment. The initial project presentation outlined core DDx challenges (overlapping symptoms, multidimensional assessment, limited consultation time) and the goal of *supporting clinical reasoning* with structured knowledge.

3.2 Scope and Design Decisions (Minimum Viable Product (MVP))

Following supervisory meetings, the team scoped the *MVP* to **adults** (≥ 19 years) with **two professionals** in the minimal viable workflow:

Neurologist and **Psychologist**. The objective is to structure an objective digital triage, standardized psychological testing, and a neurological report with organic exclusion, while preserving traceability and auditability of evidence.

3.3 Design Principles

- **Explainability and traceability:** clear rules, human-readable rationales attached to the report, and source logging (reports, scales, attachments).
- **Minimal viable flow:** Neurology $\leftrightarrow$ Psychology, with progressive inclusion of other specialties as clinically required.
- **Standardization:** Yes/No/Likert questions and recognized instruments to reduce ambiguity.
- **Complementarity:** the system supports—not replaces—professional judgment.

3.4 Inputs and Outputs (Concise Data Dictionary)

Priority inputs (Adults). Functional self-report (work/study/daily life); developmental history when remembered; symptoms across multiple contexts; comorbidities/medical conditions (sleep, anxiety, depression, vitamin deficits); evidence uploads (reports, exams). Items should be preferably objective (Yes/No/Likert).

System outputs. (i) **Triage sheet** with score and initial recommendations; (ii) **Structured psychology report**; (iii) **Neurology report** documenting organic exclusion and justifications; (iv) **Feedback plan** (referrals, accommodations, follow-up).

3.5 Logical Architecture and Knowledge Engineering Method

Knowledge acquisition via structured interviews focused on clinical heuristics; **modeling** as *IF* $\rightarrow$ *THEN* rules; **knowledge base** organized by entities (symptom, context, scale, criterion, evidence); **rule engine** (e.g., Drools/Prolog) for inference; **workflow orchestration** via BPM (e.g., BPM and Notation (BPMN)), with automatic document generation. This pipeline (interview $\rightarrow$ rules $\rightarrow$ prototype) follows the project presentation.

3.6 Assisted Workflow (High-Level BPMN)

1. **Structured digital triage** (Yes/No/Likert) + evidence upload $\rightarrow$ creation of the triage sheet and attachments.
2. **Psychology:** focused interview and *standardized tests* (executive functions, attention, social pragmatics) $\rightarrow$ **psychology report**.
3. **Neurology:** history, neurological exam, organic exclusion, integration of findings $\rightarrow$ **neurology report**.
4. **Feedback and plan** (referrals, accommodations, follow-up).

3.7 Business Rules (Decision Engine, MVP Version)

RB–1 Route to Neurology: Moderate/high screening signals + current functional complaint and *no recent medical assessment* $\Rightarrow$ schedule Neurology; Psychology collects *baseline.*

RB–2 Prioritize Psychological Testing. High ASD/ADHD suspicion with occupational impact $\Rightarrow$ standardized psychological battery + summary for Neurology.

RB–3 Pause and Reassess: Active comorbidities that mimic symptoms (severe anxiety, sleep disorders, vitamin deficits) $\Rightarrow$ treat/stabilize before issuing the diagnostic report.

RB–4 Neurology Report. Convergence of triage + psychology report and *organic exclusion* $\Rightarrow$ issue the neurology report; refer to Psychiatry when necessary.

3.8 Knowledge Representation and Application Approach

The modelling of expert knowledge in the system is structured through a rule-based representation approach, in which knowledge is encoded as a chain of questions organised in a decision-tree structure. To ensure clinical coherence and reflect how different specialties interpret clinical signs, a hybrid model was adopted in which the knowledge base is organised by specialty domains. These domains distinguish neurological and psychiatric axes and map functional areas such as:

- Cognitive and motor
- Sensory and emotional
- Behavioural and social

This organisation allows the questions to capture specific indicators that contribute to the probabilistic construction of neurodevelopmental disorders (NDDs). Tacit clinical knowledge, obtained through structured interviews, was formalised into explicit *IF* $\rightarrow$ *THEN* rules. The knowledge base is therefore organised into entities (such as symptom, context, scale, criterion, and evidence), preparing it to be processed by a rule engine (e.g., Drools/Prolog) that generates inferences and orchestrates the workflow via BPMN.

The inference of a possible diagnosis is obtained through a structured scoring system that quantifies the weight of each symptom within the differential diagnosis. This system is based on a questionnaire organised in a decision-tree structure.

To increase screening accuracy and reduce the likelihood of an incorrect diagnosis, the decision flow is initially segmented by a question that divides the system into two main sections:

- **Section 1:** Focused on Dyslexia and Dyspraxia, associated with motor and visual difficulties.
- **Section 2:** Focused on ADHD, Autism, Bipolar Disorder, Tourette Syndrome, and OCD, where emotional regulation and behavioural patterns are more evident.

Table 1. Variables and scoring definitions.

Variable	Definition / Calculation	Scale / Limits
Response scale	Specialist or patient indicates symptom frequency or severity.	0 = Never; 1 = Rarely; 2 = Sometimes; 3 = Often.
Scoring calculation	Each response (0–3) is multiplied by its question weight and summed to produce the condition score.	Score = response × weight.
Maximum total score	Maximum value per condition, used to compute the relative probability across diagnoses.	ADHD 10; Autism 10; Bipolar 10; Tourette 5.6; OCD 6.5; Dyslexia 5.5; Dyspraxia 5.5.

The inference process is straightforward: the patient's response is converted into a numerical value, which is then multiplied by the weight assigned to each question for a specific condition. The total sum of these weighted values for each condition allows the system to estimate the relative probability of each diagnosis.

Scoring and Decision Engine Calculation The inference calculation relies on a standardised response scale and predefined maximum scoring thresholds for each condition.

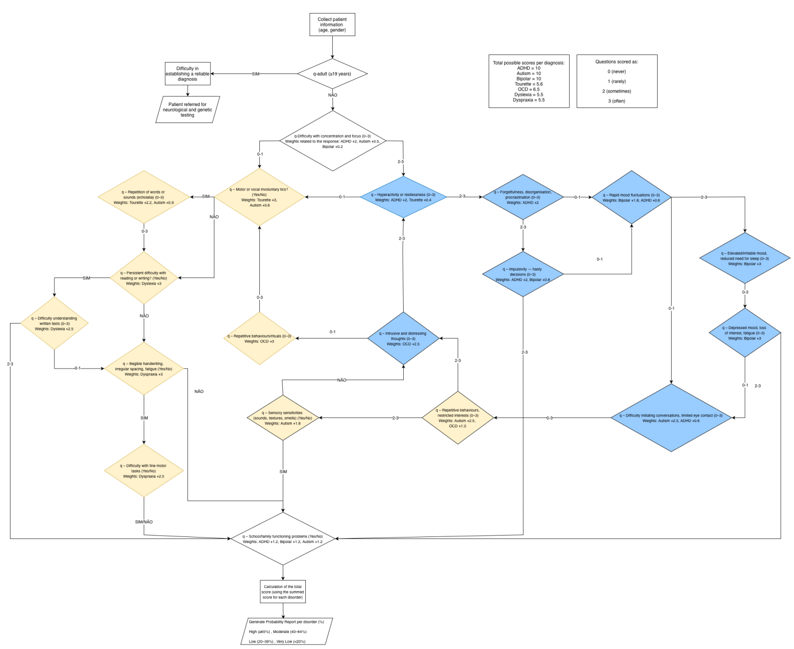

Fig. 1. Fluxograma do sistema

Example of scoring calculation Question: *"Difficulty concentrating and focusing"*
Expert's answer: 2

Application of weights:

$$ADHD : 2 \times 2 = 4, \quad Autism : 2 \times 0.3 = 0.6, \quad Bipolar : 2 \times 0.2 = 0.4.$$

For yes/no questions, the score for a given condition is added only if the answer is "yes". The system sums the values of all responses, resulting in a total score per diagnosis, which is then interpreted to estimate the relative probability of each condition.

The sequential and interconnected questions that compose the system guide healthcare professionals through a logical evaluation pathway, allowing the gradual identification of patterns associated with the different disorders analysed. In the system diagram (Fig.1), the questions highlighted in yellow represent items related to neurology—centred on brain, motor, and sensory functioning—while the questions in blue refer to the psychiatric domain, focusing on emotional, behavioural, and mood regulation aspects. This visual and logical structure ensures an integrated analysis, supporting clinical decision-making in a more consistent manner.

R1 (Age below 18): If the answer to this question is "yes", the system advises the clinician to discontinue the use of the tool and instead refer the patient for neurological and genetic evaluation.
R2 (Attention difficulties): When the response indicates a high frequency of attention-related difficulties, the weight attributed to ADHD is substantially increased.
R3 (Mood-related symptoms): If the patient reports marked or extreme fluctuations in mood, the weight associated with Bipolar Disorder is proportionally adjusted.
This system enables a preliminary inference that is quick and easy to apply, providing an initial diagnostic orientation.

The weights associated with each question were reviewed by several healthcare professionals who offered highly divergent opinions. Some argued that the system should not rely on weighted scoring across questions but were unable to propose an alternative approach; others stated that the weights were too inflated, while some suggested they were too low. A middle ground was sought to balance the contribution of each response to the different conditions.

We acknowledge that there are cases in which certain responses lead the system to assign diagnostic probabilities approaching 100%, which is unrealistic given the nature of this type of screening tool. Unfortunately, it was not possible to further adjust the weights or add additional questions to refine these scenarios due to time constraints and the unavailability of specialists from the relevant medical domains.

Example of System Output for a Use Case: In this use case, the patient answered *"No"* to the question *"Age between 0–18 years?"*.

This makes the patient eligible for diagnostic processing, as the system only evaluates individuals aged 19 or older.

The evaluation proceeds as follows:

1. **Difficulty concentrating and focusing (0–3)**: The patient answered **0**, which does not contribute points to ADHD, Autism, or Bipolar Disorder.
2. **Involuntary motor or vocal tics?** The patient answered **Yes**, adding:
   - Tourette: +3
   - Autism: +0.6
3. **Repetition of words or sounds (echolalia) (0–3)**: The patient answered **1**, adding:
   - Tourette: +2.2
   - Autism: +0.9
4. **Persistent difficulty in reading or writing (Yes/No)**: Answered **Yes**, activating the Dyslexia weight (+3) and triggering the follow-up question.
5. **Difficulty understanding texts (0–3)**: The patient answered **2**, resulting in:
$$2 \times 2.5 = 5 points for Dyslexia,$$
reaching the maximum score for this condition (5.5 points).
6. **School or family functioning problems (Yes/No)**: Answered **No**, so no additional points are added for ADHD, Bipolar Disorder, or Autism.

The system then sums all accumulated weights to generate the final diagnostic profile.

Condition	Accumulated Score
Tourette	5.2
Autism	1.5
Dyslexia	7.5

Condition	Maximum Score
Tourette	5.6
Autism	10
Dyslexia	5.5

Condition	Likelihood (%)
Tourette	92.8%
Autism	15%
Dyslexia	100%

Each group of questions was designed to capture specific indicators that, when analysed together, contribute to the probabilistic construction of neurodevelopmental disorders (NDD). This structure serves as the foundation for the weighting calculations and inference rules used in the expert system, enabling the estimation of relative probabilities across overlapping conditions (e.g., ADHD, ASD, dyslexia, dyspraxia, Tourette syndrome, bipolar disorder, or anxiety), with full clinical transparency and traceability. As shown in Fig. 2, we present the system interface with the final probability report, illustrating how the aggregated scores are translated into interpretable diagnostic likelihoods.

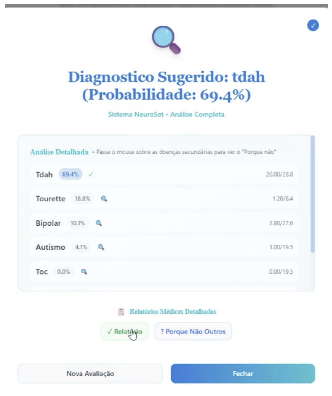

Fig. 2. System result screen

3.9 Roles and Responsibilities

Psychology: structured data collection, test administration, report with evidence and limitations. **Neurology:** history and examination, exclusion of organic causes, diagnostic decision and report. **Other specialties:** on demand, depending on case complexity.

3.10 Interoperability, Ethics, and Compliance

Interoperability: prepare data fields compatible with minimal HL7-FHIR profiles (Questionnaire, Observation, DiagnosticReport) to enable future EHR integration.
Ethics: obtain consent with a clear purpose statement (decision support, not replacement), enforce data minimization, and maintain versioning of rules with complete audit trails.
Governance: define policies for clinical attachments and role-based access control aligned with ethical use of knowledge gathered during interviews.

4 Discussion

4.1 Synthesis of Evidence

Across the reviewed literature, two complementary avenues emerge for adult NDD differential diagnosis (DDx): (i) rule-based clinical decision support systems (CDSS) embedded in business process management (BPM), which provide process standardization, explicit business rules, and auditable summaries; and (ii) machine learning (ML) pipelines that learn compact, discriminative sets of variables for triage and DDx [3, 1, 4, 2, 5]. The former demonstrates feasibility and acceptability in routine services, but lacks published diagnostic accuracy and external validation

[3]. The latter achieves promising AUCs (ASD vs. non-ASD with minimal ADOS-4 items ≈ 0.82–0.87; ASD vs. early psychosis vs. social anxiety ≈ 0.72–0.92), yet remains largely pre-deployment and single-site, with limited explainability and interoperability [1, 4, 2, 5].

4.2 Implications for Our Project (Adult, Neuro + Psychology, Rule-Based Core)

Guided by this evidence and our stakeholder meetings, we centered the MVP on adults and a two-professional workflow (Neurology + Psychology). We adopt a *rule-based core* (transparent business rules, gates, and auditable justifications) orchestrated by BPM for end-to-end traceability—aligned with strengths shown by real-world CDSS [3]. To capture the performance benefits reported by ML studies while preserving explainability, we incorporate *lightweight, optional* ML components for prioritization/triage using compact feature sets (e.g., minimal ADOS-4 items and brief neuropsychological/social cognition measures) [4, 2, 1]. Concretely, the project's business rules (RB–1..RB–4) encode safe prioritization (e.g., medical red flags ⇒ Neurology first), standardized psychological testing when functional impact is high, and temporary deferral in the presence of destabilizing comorbidities—with all decisions logged and justified in the report. The reviewed literature highlights that the DDx of NDDs in adults is inherently challenging due to symptom overlap and the need for a multispecialty team. Historically, adult diagnosis requires the integration of neurological findings (organic exclusion) and psychiatric assessment (emotional and behavioural regulation). Our Minimum Viable Product (MVP) focuses on adults (19 years) and establishes a minimal workflow involving a Neurologist and a Psychologist, with the proposed system serving as an alternative to integrate these professionals. This project is currently in the development and testing phase, with clinical rules formalised into code (rule engine prototypes) and validated in collaboration with specialists in Neurology and Psychology, ensuring that the inferences follow recognised clinical criteria. The aim is to provide reliable and auditable decision support, helping to standardise the diagnostic pathway and reduce assessment time. Experts have considered the system to be a potential support tool for professionals working in the NDD field, particularly valuable for triage and for those on the frontline of diagnosis, given the challenge of meeting rising demand. The system aims to integrate the key specialties, organising the knowledge base into neurological and psychiatric axes, which facilitates navigation through abstract variations and the inherent complexity of DDx.

4.3 Methodological Considerations and Threats to Validity

The evidence base is affected by single-site datasets, potential circularity (ground truth includes instruments also used as predictors), and scarce external validation [4, 2]. Few works examine fairness (sex/age/comorbidity subgroups), and XAI remains underused beyond variable-importance summaries [1, 2]. These issues motivate our evaluation plan to (i) predefine endpoints and analysis (including decision-curve analysis and calibration), (ii) report subgroup performance, and (iii) separate model development from reference adjudication to mitigate circularity. It is acknowledged that the evidence base for AI in DDx is affected by single-site samples and limited external validation. As the project is in a prototype stage with initial testing, it remains aware of these threats and is motivated to pursue an evaluation plan that includes separating model development from reference adjudication, thereby mitigating circularity. The central claim of the system's utility lies in its potential to reduce diagnostic time and streamline the process of filtering potential cases without loss of quality. The workflow assisted by business rules (RB–1..RB–4) encodes safe prioritisation, which translates into earlier treatment initiation and follow-up, ultimately improving patients' quality of life as early as possible.

4.4 Clinical Integration, Interoperability, and Ethics

Real-world impact requires integration with clinical workflows and records. Prior literature rarely reports HL7-FHIR/EHR integration [5], hence our design specifies a minimal interoperability profile (Questionnaire, Observation, DiagnosticReport) and audit trails for rule versions and evidence attachments. Privacy-by-design measures include purpose limitation (decision support, not automation), role-based access, and data minimization (objective Yes/No/Likert items when feasible). Each inference is accompanied by a human-readable rationale; ML-assisted recommendations expose feature attributions and uncertainty where applicable.

5 Conclusion

This work reviewed the state of the art in the use of artificial intelligence (AI) for the Differential Diagnosis (DDx) of Neurodevelopmental Disorders (NDDs) in adults. Two complementary paths were identified: (i) rule-based Clinical Decision Support Systems (CDSS) integrated into Business Process Management (BPM), which provide standardisation and auditable summaries; and (ii) Machine Learning (ML) pipelines that show promising performance for triage. Our project positions itself within this context, presenting an explainable and pragmatic model that uses a rule-based core orchestrated through BPM, complemented by lightweight and optional ML components for prioritisation. The project is currently under development and testing, built on expert knowledge to address the growing demand and the challenge of multidisciplinary assessment of NDDs in adults. The central contribution of this hybrid system, focused on integrating the Neurology and Psychology workflow, lies in its potential to reduce time and variability in the triage and case-filtering process. By ensuring traceability and explainability of decisions, the system has the capacity to accelerate the initiation of treatment and follow-up, contributing to improved quality of life for adults with NDDs.

References

1. Demetriou, E., Park, S., Ho, N., Pepper, K., Song, Y., Naismith, S., Thomas, E., Hickie, I., Guastella, A.: Machine learning for differential diagnosis between clinical conditions with social difficulty: autism spectrum disorder, early psychosis, and social anxiety disorder. Frontiers in Psychiatry **11**, 545 (2020). https://doi.org/10.3389/fpsyt.2020.00545
2. Kamp-Becker, I., Tauscher, J., Wolff, N., Küpper, C., Poustka, L., Röpke, S., Roessner, V., Heider, D., Stroth, S.: Is the combination of ADOS and ADI-R necessary to classify ASD? rethinking the "gold standard" in diagnosing ASD. Frontiers in Psychiatry **12**, 727308 (2021). https://doi.org/10.3389/fpsyt.2021.727308
3. Kemppinen, J., Salmisaari, T., Korpela, J., Polkko, J., Elfvengren, K., Tuominen, M.: A clinical decision support system for adult ADHD diagnostics process. In: Proceedings of the 46th Hawaii International Conference on System Sciences (HICSS). pp. 2614–2625. IEEE, Wailea, HI, USA (2013). https://doi.org/10.1109/HICSS.2013.30
4. Küpper, C., Stroth, S., Wolff, N., Hauck, F., Kliewer, N., Schad-Hansjosten, T., Kamp-Becker, I., Poustka, L., Roessner, V., Schultebraucks, K., Röpke, S.: Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning. Scientific Reports **10**, 4805 (2020). https://doi.org/10.1038/s41598-020-61607-w
5. Loh, H., Ooi, C., Barua, P., Palmer, E., Molinari, F., Acharya, U.: Automated detection of ADHD: current trends and future perspective. Computers in Biology and Medicine **146**, 105525 (2022). https://doi.org/10.1016/j.compbiomed.2022.105525