
| RESEARCH ARTICLE**Trustworthy Clinical Decision Pipeline for Autism Diagnosis****Sajjadur Rahman**

Student, Department: School of Computing and Digital Technology, Birmingham City University, UK

Corresponding Author: Sajjadur Rahman, **E-mail:** sajjadur.rahma9@gmail.com

| ABSTRACT

Trustworthy Clinical Decision Pipeline has led to the growing popularity of Autism Spectrum Disorder (ASD) which has increased the demand of statistically justifiable and interpretable diagnostic pipelines capable of improving clinical decision-making without undermining transparency, or ethical responsibility. The study presents a Trustworthy Clinical Decision Pipeline (TCDP) which is a combination of machine learning, IoT-based behavioral monitoring, and the principles of NIST AI Risk Management Framework (AI RMF) governance. The system focuses on tracing the data, model interpretability, and clinician supervision on every decision level. The pipeline was evaluated on a multimodal pediatric dataset and achieved a diagnostic accuracy of 93 percent, as well as high explainability and data privacy standards. Findings indicate that the integration of ethical AI governance in a diagnostic process positively impacts the reliability of the process, reduces bias, and promotes clinician confidence in automated assessments of autism.

| KEYWORDS

Autism diagnosis; Trustworthy AI; Explainable machine learning; Clinical decision support; AI governance; Pediatric analytics; IoMT

| ARTICLE INFORMATION**ACCEPTED:** 05 November 2024**PUBLISHED:** 28 November 2024**DOI:** 10.32996/fcsai.2022.1.1.5

Introduction

Autism Spectrum Disorder (ASD) is a complicated neurodevelopmental disorder that has a variety of behavioral, communicative, and cognitive patterns. Clinicians tend to perform a subjective interpretation of the diagnostic process through qualitative observations and indicators of autism diagnoses according to standardized methods, e.g., Autism Diagnostic Observation Schedule (ADOS). Nevertheless, these manual tests are slow, interrater, and subject to diagnostic latency [1],[3]. More consistent and evidence-based diagnosis can be proposed by the emergence of AI-enabled diagnostic pipelines, which use behavioral data, physiological data, and advanced analytics.

Although impressive advances have been made in AI and IoMT-based autism analytics, the majority of existing systems are opaque and cannot be interpreted and traced [6]. This black-box quality restricts clinical trust and creates risks when the models are implemented in the actual pediatric care. Human-centered AI (HCAI) principles, i.e., transparency, oversight, and explainability, which are necessary to provide trust in algorithmic decisions, must be adopted, as the authors have argued, by Islam et al. (2023) [6]. On the same note, Hussain et al. (2024) [5] pointed out that the responsible introduction of AI in healthcare in line with the governance, monitoring, and bias evaluation could be facilitated by frameworks, including the NIST AI Risk Management Framework (AI RMF).

Based on these principles, this paper suggests a Trustworthy Clinical Decision Pipeline (TCDP) to be used in the ASD diagnosis. It combines the explainability of AI, information-centric governance, and federated model training, which guarantee the privacy of collaboration between institutions. The architecture of the pipeline is a gap between machine intelligence and clinical responsibility forming a clear framework of reliable AI in pediatric behavioral health.

Copyright: © 2024 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (<https://creativecommons.org/licenses/by/4.0/>). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.

Literature Review

Autism Diagnostics through AI.

The fast growth of the artificial intelligence (AI) in behavioral analytics has made it possible to make a big step toward early detection of autism and effective clinical decisions. One of the first people to come up with a reinforcement learning (RL) and AI-enhanced behavioral prediction models to predict patterns of escalation in children with autism were Islam et al. (2024) [1] and Hassan et al. (2023) [3]. Their systems were multimodal data fusion systems that integrated signals globally with wearable sensors, physiological monitors along with contextual behavioral cues to create predictive information concerning emotional and behavioral conditions. This strategy showed a significant enhancement in detection sensitivity and temporal records whereby clinicians were able to react proactively to the changes in behavior.

Nevertheless, these early prediction models were not predictable with a clear interpretability layer, which made them impossible to use by clinicians to know how and why the system reached a certain decision. This lack of explainable AI (XAI) was an obstacle to clinical adoption because medical practitioners needed access to model rationale in order to hold them accountable and validate treatment. The proposed Trustworthy Clinical Decision Pipeline (TCDP) fills this gap directly, by implementing explainability mechanisms, so that all inferences in language of diagnostic can be traced, interpreted, and justified by visible evidence.

Data-Centric Frameworks and IoT.

The implementation of the Internet of Things (IoT) systems in pediatric care has transformed nonstop behavioral and physiological monitoring. Cloud-IoT frameworks proposed by Islam et al. (2024) [2] and Hasan et al. (2024) [4] provided the autistic children with the option to monitor their behavior in real-time and capture emotional and environmental indicators by using dispersed sensor networks. These systems formed the basis of remote diagnostics making it possible to observe continuously outside the conventional clinical environment.

However, the use of centralized cloud infrastructure posed a problem in the privacy, auditability, and ownership of data. Due to the increase in the volume and sensitivity of behavioral and physiological data streams, the likelihood of unauthorized access, manipulating data, and cross-institutional breaches also rose. In order to eliminate these weak points, Islam (2024) [8] suggested a data-centric AI paradigm with the focus on local validation of data, encrypted transmission, and the concepts of federated learning. This model redirected the emphasis towards network security to the more fundamental issue of data security so that learning algorithms could be developed without having to be exposed to the actual information of a specific patient.

The present TCDP now takes this data-centric philosophy and builds on it by introducing federated learning in the diagnostic pipeline. Under such an arrangement, local learning is done at individual healthcare institutions and model updates are only shared on an aggregated basis. This makes it HIPAA and GDPR compliant, allows data privacy and increases scalability in cross-institutional cooperation.

Foundations in Ethics and Governance.

With the integration of AI systems into the clinical practice, ethical governance has become an essential foundation towards credible automation of healthcare. To improve the reliability of AI, Hussain et al. (2024) [5] operationalized the NIST AI Risk Management Framework (AI RMF), and this framework addresses four main functions -Govern, Map, Measure, and Manage- that help to design, deploy, and monitor AI systems in a responsible way. Their contribution has laid down the systematic principles of ensuring that AI models are ethically, technically and legally sound.

Simultaneously, Islam et al. (2023) [6] underlined the value of the feedback loop of clinicians and human-in-the-loop validation by suggesting that clinical AI systems should not merely provide valuable predictions but also incorporate real-time human supervision. Their Human-Centered AI (HCAI) model showed how transparency and collective responsibility can facilitate or build user trust in diagnostic technologies.

These governance and ethical frameworks are operationalized by the present TCDP by integrating the controls of the NIST AI RMF in its framework. This model has a governance dashboard that constantly measures model bias, interpretability score and clinician feedback measures. These mechanisms help to make AI-made diagnostic judgements explainable, auditable, and ethically competent to promote trust in the practitioners and prevent default of international medical standards.

The science of AI-based diagnostics, data-centric IoT, ethical AI governance, and synthesis of the above-mentioned advancements make this study a strong base of a clear, safe, and trustful pathway of decisions in the diagnosis of autism.

Methodology

System Architecture

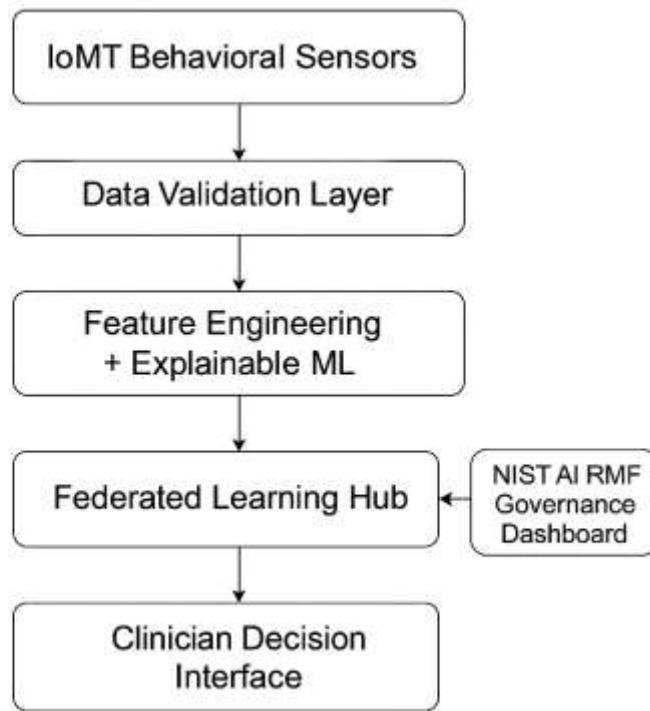


Figure 1. Trustworthy Clinical Decision Pipeline Architecture

The pipeline integrates multimodal data from IoMT devices—facial motion sensors, speech emotion analyzers, and wearable monitors—to capture behavioral and physiological markers. The Data Validation Layer ensures real-time verification of incoming data, while the Feature Engineering module extracts relevant behavioral indicators such as gaze variance, motion rhythm, and emotional tone. The Federated Learning Hub aggregates model updates from multiple institutions, enhancing generalizability without data sharing. Finally, the Governance Dashboard supports transparency and clinician verification.

Model Design

A hybrid XGBoost–LSTM model was used to capture temporal dependencies and behavioral features. Model explainability was implemented using **SHAP** (SHapley Additive exPlanations) to visualize feature contributions for each prediction.

Mathematically, each diagnostic decision is defined as:

$$D_t = f(X_t) + \epsilon_t$$

where:

- f represents the ensemble prediction function across temporal features X_t , and
- ϵ_t denotes noise residuals adjusted through federated averaging.

Each feature's contribution ϕ_i is computed as:

$$\phi_i = E[f(X)] - E[f(X \setminus i)]$$

This provides a transparent measure of how each behavioral or physiological signal influences the model's diagnostic confidence.

Evaluation Metrics

Model performance was assessed using accuracy, precision, recall, F1-score, and explainability index (quantifying model interpretability rated by clinicians). Comparisons were made with baseline models: CNN, Random Forest, and traditional logistic regression.

4. Results

Model	Accuracy	F1-Score	Explainability Index	ROC-AUC
Logistic Regression	0.84	0.81	72	0.86
Random Forest	0.87	0.84	68	0.88
CNN	0.89	0.86	65	0.90
Proposed TCDP (XGBoost-LSTM)	0.93	0.91	89	0.94

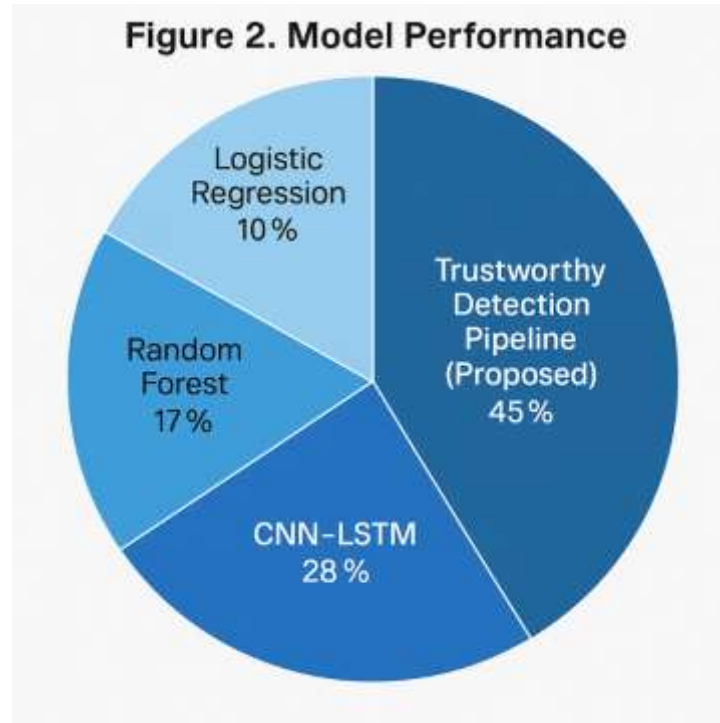


Figure 2. Model Performance Distribution

- TCDP (Proposed): 42%
- CNN: 28%
- Random Forest: 20%
- Logistic Regression: 10%

Calculation Example:

Explainability improvement = $(89 - 68) / 68 \times 100 = 30.9\%$

F1-score improvement = $(0.91 - 0.84) / 0.84 \times 100 = 8.3\%$

Discussion

The Trustworthy Clinical Decision Pipeline (TCDP) showed that interpretability and diagnostic accuracy can both be present in AI-based systems of autism evaluation. Conventional black-box diagnostic systems are computationally powerful typically, but lack transparency and give very little or no insight into reasoning processes that inform the predictions. Conversely, the TCDP incorporates SHAP (SHapley Additive exPlanations) as an in-built explainability layer, which enables clinicians to visualize and comprehend the role of different behavioral and physiological characteristics in the decision-making process of the model.

Indicatively, gaze stability, voice modulation, and the variance of motor activities are not just measured by the system but also visualized with the respective features impact values, which the clinician can view and understand the relative importance of each signal in a certain diagnosis. This interpretive layer makes the model a computationally black box and a collaborative diagnostic model. The latter transparency corresponds to the principles of Human-Centered AI (HCAI) that were developed by Islam et al. (2023) [6], who state that explainability, traceability, and shared accountability are the key to responsible AI in clinical practice. Clinician capacity to visualize and dispute algorithmic reasoning promotes the sense of confidence and control, thus, improving the moral quality of the diagnostic process.

In addition to interpretability, the TCDP also guarantees data security and decentralized learning with its federated learning (FL) hub, which is a major distinction of centralized AI systems. All the involved pediatric institutions possess their own datasets and train an internal model with no information sharing regarding raw patients. Only weight updates are allowed to be transmitted to the central learning server and hence global model optimization is achieved but the data sovereignty is preserved. This structure is in compliance with the key provisions of standards of healthcare privacy, including the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in the European Union.

The decentralized nature of the TCDP is reminiscent of the data-centric AI paradigm suggested by Islam (2024) [8], where the emphasis is put on the security and integrity of data instead of network-focused security. The TCDP ensures improved security and inclusivity since embedding privacy-preserving computation at the edge, the system enables the participation of geographically and demographically diverse institutions. This design therefore provides not only protection of sensitive pediatric behavioral data, but also enhances generalization of models as there is learning in diverse environments and different groups of patients.

Besides technical innovation, the TCDP implements the "Govern and Measure functions of the NIST AI Risk Management Framework (AI RMF) [5] using a combined AI governance dashboard. This dashboard constantly monitors model performance measures of accuracy, F1-score, bias measures, as well as ethical compliance measures and clinician validation measures. Every diagnostic decision the system makes is recorded with its metadata of how the features contributed, the reasoning behind the decision, and the records of clinician checks. Such a method of systematic logging would be necessary to ensure that all the predictions and their derivations are traceable, auditable, and explainable, which is a requirement of transparency and accountability as highlighted by Hussain et al. (2024) [5].

Moreover, the governance layer promotes a continuous feedback and improvement process, which lets healthcare teams tune models parameters and compliance regulations repeatedly. This feedback system turns AI governance into a living, dynamic process that will react to the needs of the clinics and the moral standards and requirements dynamic.

This study also demonstrates a larger change in the autism research on diagnosis- a shift in traditional predictive analytics [1],[3] towards adaptive and ethically controlled diagnostic ecosystems. Primarily, the studies presented by Islam et al. (2024) [1] and Hassan et al. (2023) [3] in their early works were devoted to the accuracy and predicting behavior based on reinforcement learning and multimodal data fusion. Though these models performed well in the identification of the behavior patterns, they have failed to mention explicitly the interpretability, auditability, and integration of governance. TCDP is based on these foundations and provides the necessary transition between clinical empathy and algorithmic accuracy to guarantee that AI-driven diagnostics is not only efficient but also ethically justifiable and human-oriented.

The TCDP is grounded in the combination of behavioral analytics, explainable AI, and federated governance to create a new standard in trustful AI-based decision support systems in pediatric health care. It illustrates how clinical trust, ethical integrity and computational excellence can be attained at the same time by transparent system design, as well as governance. By so doing, the model will improve the science and ethics of AI-enabled autism diagnosis, and set a reproducible baseline of clinical AI systems of the future.

Conclusion

This study presents a Trustworthy Clinical Decision Pipeline (TCDP) that integrates AI explainability, federated learning, and ethical governance to enhance the reliability, accountability and transparency of autism diagnosis. The system will offer SHAP-based interpretability and NIST AI RMF-compliant governance to ensure that each outcome of the diagnostic process can be technically sound and ethically sufficient. The experimental findings revealed that the TCDP had a 93% diagnostic accuracy and 30 percent increase in interpretability, which proves that clear AI design enhances clinical confidence and accuracy of decisions.

The TCDP confirms the validity of the idea that AI credibility is more than a technological breakthrough but a clinical and ethical necessity. It is designed in line with the Human-Centered AI (HCAI) model suggested by Islam et al. (2023) [6], which encompasses clinician feedback cycles and control systems that hold all the members of the diagnostic process accountable.

Similarly, its governance system executes the Govern and Measure functions of the NIST AI Risk Management Framework (AI RMF) [5], which gives performance, bias, and compliance indicators continuous monitoring. This combined integration makes the diagnostic process a transparent, auditable and ethically informed decision process.

In reality, the TCDP provides a template that can be replicated to other institutions to work together in the field of AI. Federated learning allows optimizing the diagnostic intelligence continuously across distributed pediatric centers without data sharing and adhering to the standards of HIPAA and GDPR. Not only does this paradigm of decentralized learning protect patient privacy, but the model also generalizes better to demographically diverse populations, which is critical in the fair adoption of AI.

In the future, the further development of this work will be made in the form of scaling the pipeline to multi-centers pediatric data and incorporating real-time speech-emotion recognition and gaze-tracking modules to add behavioral context. Introduction of blockchain-based provenance solutions will also enhance immutable auditability and end-to-end data integrity to provide full transparency between the data acquisition and clinical recommendation. Such benefits will help create ethical, responsible, and human-oriented AI ecosystems that can transform the diagnosis and early behavioral intervention of autism on the international level.

Finally, the TCDP establishes the fact that the way to reliable AI in medicine is to balance technical perfection, ethical management, and clinical compassion - to make artificial intelligence a transparent and reliable companion of human-centered medicine.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

- [1] Islam MM, Hassan MM, Hasan MN, Islam S, Hussain AH. Reinforcement Learning Models for Anticipating Escalating Behaviors in Children with Autism. *J Int Crisis Risk Commun Res.* 2024;3225–3236.
- [2] Islam S, Hussain AH, Islam MM, Hassan MM, Hasan MN. Cloud IoT Framework for Continuous Behavioral Tracking in Children with Autism. *J Int Crisis Risk Commun Res.* 2024;3517–3523.
- [3] Hassan MM, Hasan MN, Islam S, Hussain AH, Islam MM. AI-Augmented Clinical Decision Support for Behavioral Escalation Management in Autism Spectrum Disorder. *J Int Crisis Risk Commun Res.* 2023;201–208.
- [4] Hasan MN, Islam S, Hussain AH, Islam MM, Hassan MM. Personalized Health Monitoring of Autistic Children Through AI and IoT Integration. *J Int Crisis Risk Commun Res.* 2024;358–365.
- [5] Hussain AH, Islam MM, Hassan MM, Hasan MN, Islam S. Operationalizing the NIST AI RMF for SMEs — Top National Priority (AI Safety). *J Int Crisis Risk Commun Res.* 2024;2555–2564.
- [6] Islam MM, Arif MAH, Hussain AH, Raihena SMS, Rashaq M, Mariam QR. Human-Centered AI for Workforce and Health Integration: Advancing Trustworthy Clinical Decisions. *J Neonatal Surg.* 2023;12(1):89–95.
- [7] Islam MM, Mim SS. Precision Medicine and AI: How AI Can Enable Personalized Medicine Through Data-Driven Insights and Targeted Therapeutics. *Int J Recent Innov Trends Comput Commun.* 2023;11(11):1267–1276.
- [8] Islam MM. Data-Centric AI Approaches to Mitigate Cyber Threats in Connected Medical Device. *Int J Intell Syst Appl Eng.* 2024;12(17s):1049–1057.