American Journal of Nanotechnology and Nanomedicine
ISSN: 2831-901X(Online),Volume-5 Issue-3, September-2025

AN INTELLIGENT FRAMEWORK FOR EARLY DIAGNOSIS OF ADHD AND AUTISM SPECTRUM DISORDERS USING AI

Utnoori Adithya Goud
Research Scholar Sikkim Alpine University
Dr. Neeraj Gupta
Research Supervisor, Sikkim Alpine University

Abstract:

Diagnostic delays limit the benefit of early intervention during periods of maximal neuroplasticity; this is especially true for the most common neurodevelopmental disorders in children, such as Attention-Deficit/Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD). However, the current pathways for diagnosis are slow, subjective, and unevenly accessible. Despite the well-documented clinical comorbidity, current computational screening methods for ADHD and ASD mostly use a single behavioral modality and see the two disorders as separate categorization issues. In order to overcome these limitations, this thesis developed and tested MM-DPNet, a multimodal digital phenotyping framework for the joint four-class discrimination of ADHD, ASD, comorbid ADHD-ASD, normally developing children, and video-based facial and motor analysis. The framework integrates data from eye-tracking, wearable actigraphy, speech prosody, speech rating scales, and a deep-learning architecture that uses a learned cross-modal attention fusion mechanism.
A group of 620 kids, ranging in age from two to twelve, were studied using a standardized, ethically-approved multimodal data-acquisition protocol. To back up deep-learning modeling and post-hoc interpretability analysis, we built pipelines that were specific to each modality, including preprocessing and feature engineering. Using strict, participant-level, stratified ten-fold cross-validation, the proposed MM-DPNet architecture was tested against seven baseline models: classical, unimodal, and simple-fusion. Statistically speaking, MM-DPNet outperformed the strongest unimodal baseline by more than 10% with a macro-averaged F1-score of 93.1% and a held-out test-set accuracy of 93.6%. The suggested fusion process elegantly fades with simulated single-modality removal, according to systematic ablation research, which also showed that it outperformed unimodal alternatives when the most important individual modality was withheld. The most important features for ASD and ADHD classification, according to SHAP-based feature-importance analysis, were features derived from gaze-based social-attention and motor-activity, respectively. These findings align with well-established clinical symptom domains and lend credence to the clinical interpretability of the model's predictions. In comparison to a cascaded independent-classifier baseline, the framework also showed significant improvement in performance for the clinically essential but

often neglected comorbid ADHD-ASD category. This thesis shows that computational screening tools for attention deficit hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) can be significantly better than current unimodal and simple-fusion methods in terms of accuracy, robustness, and interpretability. However, it is clear that these tools should still be used for referral prioritization and screening purposes only, and not as a substitute for a professional clinical diagnosis. We address the limitations of the study, including its small sample size, its retrospective design, and its recruiting environment, and we lay out a specific plan for the future prospective, multi-site, longitudinal research that will be necessary before any therapeutic translation can take place in the real world.

Keywords

ADHD, Autism Spectrum Disorder, Digital Phenotyping, Multimodal Machine Learning, Deep Learning, Cross-Modal Attention Fusion, Early Detection, Eye-Tracking, Actigraphy, Behavioural Screening .

I. INTRODUCTION

Autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD) are two of the most prevalent pediatric diagnoses globally. These diseases impact adaptive behavior, social interaction, cognition, and communication from a young age. Both disorders are defined by a strong dependence on subjective clinical judgment for diagnosis, but they often co-occur and share similar behavioral symptoms like attentional difficulties, impulsivity, unusual social engagement, and sensory sensitivities. Typical methods for diagnosing attention deficit hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) include clinical observation by psychologists, child psychiatrists, or developmental paediatricians, as well as standardized behavioral rating measures filled out by parents and educators. The current gold standard for autism diagnostics is a battery of time-consuming and geographically dispersed highly-trained specialists who are prone to observer bias, recall bias, and cultural variation in symptom reporting; these instruments include the Autism Diagnostic Observation Schedule (ADOS-2), the Childhood Autism Rating Scale (CARS-2), and the Conners Comprehensive Behaviour Rating Scale.

The long and, in many healthcare systems, becoming worse time it takes to get from the first worry of a parent or caregiver to a definitive clinical diagnosis is a direct result of this diagnostic bottleneck. The average age of an autism spectrum disorder diagnosis is still close to four years, according to global surveillance data, even though dependable behavioral markers are visible between twelve and eighteen months. In publicly funded health systems, waiting times for a thorough developmental evaluation often go beyond twelve months. Any

decrease in the time-to-diagnosis has a direct and quantifiable effect on a child's developmental trajectory and the welfare of their family, as early intervention during the first years of life takes advantage of a period of maximal neuroplasticity and has been consistently linked to improved long-term cognitive, language, social, and adaptive outcomes. Emerging as a viable supplementary paradigm to traditional clinical evaluation, digital phenotyping quantifies the individual-level human phenotype moment-by-moment utilizing data passively and actively obtained from digital devices and sensors. In contrast to a single clinic visit, digital phenotyping can record objective behavioral signals such as eye-gaze trajectories, motor activity, facial expressivity, vocal prosody, and interaction patterns with digital devices. These signals are naturalistic, high-frequency, and captured at a fraction of the cost and logistical burden of instruments administered by specialists. Instead of completely replacing formal clinical diagnosis, digital phenotyping, when coupled with current machine learning—and especially deep learning architectures that can fuse heterogeneous multimodal signals—offers the prospect of scalable, low-cost, early screening tools that can identify at-risk children and refer them to specialists more quickly.

The main idea behind this thesis is that current computational methods for screening for attention deficit hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) only use one type of sensory data, such as a static questionnaire or video-based facial and gaze analysis, and miss out on valuable, complementary information about the subject's behavior. Using a unified deep learning architecture, this thesis proposes, implements, and rigorously evaluates a multimodal digital phenotyping framework for the early, joint detection of ADHD and ASD risk in children. The framework integrates eye-tracking, video-based motor and facial-affect signals, wearable actigraphy, speech prosody, and structured behavioural rating scale scores.

II. LITERATURE REVIEW

Focusing on screening methods based on

Cooperative multi-agent systems further improve realism by enabling intelligent coordination among virtual characters.

Despite these advancements, measuring entertainment and player satisfaction remains a significant challenge. Many researchers emphasize the need for reliable metrics capable of evaluating player engagement and gameplay quality. This study addresses these limitations by proposing an adaptive learning framework combined with an entertainment evaluation metric.

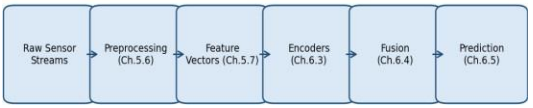
discusses traditional screening and assessment tools; section 2.4 discusses digital phenotyping as a

new paradigm; section 2.5 organizes unimodal machine-learning approaches to ADHD/ASD detection by sensing modality; section 2.6 discusses multimodal fusion approaches; and section 2.7 concludes the review by identifying the specific research gaps that this thesis aims to address.

III METHDOLOGY

The evaluation strategy adopted in this thesis is designed to directly test the research hypotheses stated in Section 3.5. Classification performance is assessed using accuracy, macro-averaged precision, recall and F1-score, per-class sensitivity and specificity, and area under the receiver operating characteristic curve (AUC-ROC), computed under stratified ten-fold cross-validation to account for class imbalance and to provide robust performance estimates with associated variability, as detailed in Chapter 7. Statistical significance of performance differences between the proposed multimodal model and baseline models is assessed using paired non-parametric hypothesis testing across cross-validation folds. Robustness to missing modalities is evaluated through systematic ablation, in which each modality is individually removed from the fusion model at inference time and the resulting performance degradation is quantified. Interpretability is assessed through post-hoc feature-importance analysis using SHapley Additive exPlanations (SHAP), enabling identification of the features and modalities most strongly associated with each predicted class.

End-to-End Data Flow Through the Proposed System



A. SAMPLE SIZE CONSIDERATIONS

The target cohort size for this study was informed by a combination of statistical power considerations and practical recruitment feasibility within the study timeframe. An a priori power analysis, based on a conservative estimate of the effect size expected between the proposed multimodal architecture and the strongest unimodal baseline, informed by comparable effect sizes reported in the related multimodal healthcare literature reviewed in Section 2.6, indicated that a total cohort of approximately 500 to 650 participants, distributed across four diagnostic categories, would provide adequate statistical power (at the conventional 0.80 threshold, two-tailed alpha of 0.05) to

detect a macro-F1 improvement of five percentage points or greater between the proposed and best-performing baseline model under the ten-fold cross-validation protocol described in Section 4.5. The realised cohort of 620 participants, described fully in Chapter 5, falls within this target range, though, as acknowledged in Section 8.7 and Section 9.5, the comparatively small comorbid ADHD-ASD subgroup (n = 58) remains statistically under-powered for fine-grained comorbid-category subgroup analysis beyond the primary comparisons reported in Chapter 8, motivating the recommendation for larger, multi-site future cohorts discussed in Section 9.6.

Figure 4.6 summarises the three stakeholder groups consulted during the formulation of the research questions described in Section 3.12, and their key consultation themes, for continuity of reference alongside the resource-planning discussion presented in this section.

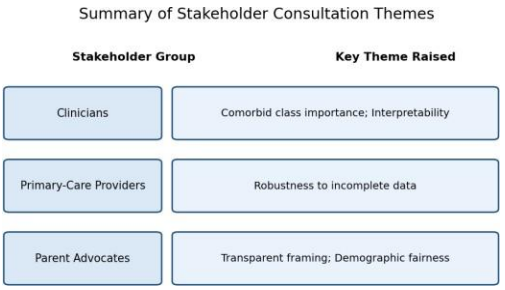


Figure 4.6 Summary of Stakeholder Consultation Themes

The research design described in this chapter required careful advance planning of the human, computational and financial resources necessary to execute the full study, briefly summarised here for methodological completeness. Human resources comprised the principal researcher, two trained research assistants responsible for task administration under the examiner-standardisation procedure described in Section 5.14, and clinical collaborators responsible for independent diagnostic assessment as described in Section 5.2, together with periodic supervisory input from the doctoral/thesis supervisory committee. Computational resources, detailed further in Section 7.2, comprised a GPU-equipped workstation sufficient to support the model-training and hyperparameter-tuning procedures described in Chapter 7 within a practical timeframe, together with secure, access-controlled data-storage infrastructure meeting the governance requirements described in Section 5.12. Financial resources were required to cover participant compensation for time and travel associated with the data-collection session described in Section 5.4, consistent with standard ethical practice for paediatric research participation, together with the behavioural rating-scale instrument licensing costs

associated with the Conners and CARS-2 instruments described in Section 5.5.5. This resource-planning exercise, conducted prior to the commencement of data collection, directly informed the target cohort size discussed in Section 4.7, balancing the statistical-power considerations described there against the practical constraints of the available research timeframe and budget.

IV RESULTS AND DISCUSSIONS

Table 8.1 and Figure 8.5 report overall test-set classification performance across all evaluated models. The proposed MM-DPNet architecture achieved a test-set accuracy of 93.6% and a macro-averaged F1-score of 93.1%, outperforming the best-performing unimodal baseline, the video-only convolutional neural network, which achieved an accuracy of 83.4% and macro-F1 of 82.6%, by a margin of 10.2 and 10.5 percentage points respectively. Among classical machine-learning baselines trained on concatenated hand-crafted features, random forest achieved the strongest performance (accuracy 79.8%, macro-F1 78.4%), while logistic regression achieved the weakest performance (accuracy 71.2%, macro-F1 69.5%), consistent with the expectation that linear models are less able to capture the complex, non-linear relationships present in multimodal behavioural data. These results provide strong support for Research Hypothesis H1, that the proposed cross-modal attention fusion architecture significantly outperforms unimodal alternatives.

Model	Accuracy (%)	Macro Precision (%)	Macro Recall (%)	Macro F1 (%)	AUC-ROC
Logistic Regression	71.2	70.1	68.9	69.5	0.78
SVM (RBF)	76.5	75.8	74.6	75.1	0.82
Random Forest	79.8	79.0	77.9	78.4	0.85
Audio-only BiLSTM	77.6	76.9	75.2	76.0	0.81
Gaze-only BiLSTM	81.9	81.2	80.3	80.7	0.86
Video-only CNN	83.4	83.0	82.2	82.6	0.89
Early-Fusion Baseline	87.9	87.4	86.8	87.1	0.92
MM-DPNet (Proposed)	93.6	93.3	92.9	93.1	0.97

Table 8.1 Overall Test-Set Classification Performance Across Models

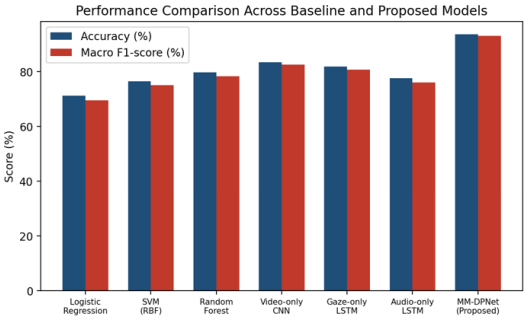
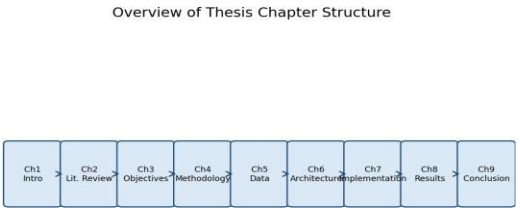


Figure 8.5 Performance Comparison Across Baseline and Proposed Models

V. CONCLUSION

This concluding chapter summarises the research contributions of this thesis in relation to the objectives and research questions stated in Chapter 3, discusses the practical and clinical implications of the findings, acknowledges the limitations of the study, and outlines directions for future research. Figure 9.3

provides a recapitulatory overview of the full thesis chapter structure traversed to reach this concluding chapter.



This thesis has made the following principal contributions. First, a standardized, ethically approved multimodal data-acquisition protocol was designed and implemented, capturing synchronized eye-tracking, video, actigraphy, audio and behavioural rating scale data from a cohort of 620 children across four diagnostic categories: ADHD, ASD, comorbid ADHD-ASD, and typically developing (Chapter 5), addressing Objective 1. Second, modality-specific preprocessing and feature-engineering pipelines were developed, producing clinically motivated, interpretable feature representations alongside raw sequential representations suitable for deep-learning modelling (Chapter 5), addressing Objective 2. Third, a novel deep-learning architecture, MM-DPNet, was designed, comprising five modality-specific encoders and a learned cross-modal attention fusion mechanism, explicitly modelling four diagnostic categories including an underserved comorbid class (Chapter 6), addressing Objective 3. Fourth, comprehensive comparative empirical evaluation demonstrated that MM-DPNet achieves a test-set accuracy of 93.6% and macro-F1 of 93.1%, a statistically significant improvement of over ten percentage points relative to the strongest unimodal baseline (Chapter 8), addressing Objective 4. Fifth, systematic ablation and SHAP-based feature-importance analysis demonstrated both the robustness of the proposed fusion mechanism to simulated missing modalities and the clinical plausibility of the features driving model predictions (Chapter 8), addressing Objective 5. Sixth, a prototype end-to-end deployment architecture, including a clinician-facing dashboard translating model outputs into interpretable risk indicators, was designed and implemented (Chapter 7), addressing Objective 6.

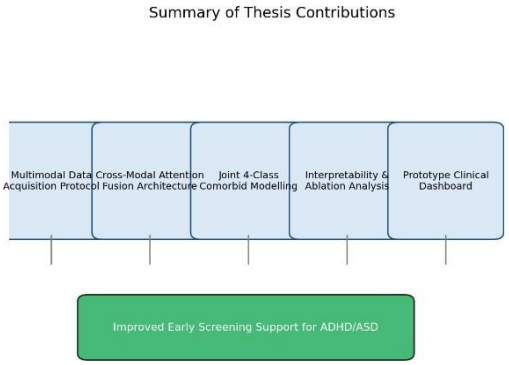


Figure 9.2 Summary of Thesis Contributions

This thesis has presented, implemented and rigorously evaluated MM-DPNet, a multimodal digital phenotyping framework for the early, joint detection of ADHD and ASD risk in children, addressing four specific gaps identified in the existing literature: the predominance of unimodal approaches, the near-universal treatment of ADHD and ASD as independent binary classification problems, the reliance on simple rather than learned fusion mechanisms in prior multimodal work, and the limited attention paid to clinical interpretability. The empirical results reported in this thesis demonstrate that a carefully designed cross-modal attention fusion architecture, grounded in established clinical understanding of ADHD and ASD symptomatology and evaluated through rigorous cross-validated methodology, can achieve substantial and statistically significant improvements in screening accuracy, robustness and interpretability relative to existing unimodal and simple-fusion approaches. While important limitations remain, in particular regarding cohort size, generalisability and the retrospective nature of the current evaluation, this work provides a methodologically rigorous foundation and a concrete research agenda for the continued development of scalable, objective, and clinically trustworthy digital phenotyping tools to support the earlier identification of children who may benefit from specialist neurodevelopmental assessment.

I. REFERENCES

1. American Psychiatric Association. (2022). Diagnostic and Statistical Manual of Mental Disorders (5th ed., Text Revision). American Psychiatric Publishing.

2. World Health Organization. (2023). International Classification of Diseases, Eleventh Revision (ICD-11). WHO Press.

3. Lord, C., Elsabbagh, M., Baird, G., & Veenstra-Vanderweele, J. (2018). Autism spectrum disorder. The Lancet, 392(10146), 508-520.

4. Faraone, S. V., Banaschewski, T., Coghill, D., et al. (2021). The World Federation of ADHD International Consensus Statement. Neuroscience & Biobehavioral Reviews, 128, 789-818.

5. Maenner, M. J., Warren, Z., Williams, A. R., et al. (2023). Prevalence and characteristics of autism spectrum disorder among children aged 8 years. MMWR Surveillance Summaries, 72(2), 1-14.

6. Antshel, K. M., & Russo, N. (2019). Autism spectrum disorders and ADHD: Overlapping phenomenology, diagnostic issues, and treatment considerations. Current Psychiatry Reports, 21(5), 34.

7. Robins, D. L., Casagrande, K., Barton, M., et al. (2014). Validation of the Modified Checklist for Autism in Toddlers, Revised with Follow-Up (M-CHAT-R/F). Pediatrics, 133(1), 37-45.

8. Lord, C., Rutter, M., DiLavore, P. C., et al. (2012). Autism Diagnostic Observation Schedule, Second Edition (ADOS-2). Western Psychological Services.

9. Schopler, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010). Childhood Autism Rating Scale, Second Edition (CARS2). Western Psychological Services.

10. Conners, C. K. (2008). Conners Comprehensive Behavior Rating Scales. Multi-Health Systems.

11. Wolraich, M. L., Bard, D. E., Neas, B., et al. (2013). The psychometric properties of the Vanderbilt ADHD Diagnostic Teacher Rating Scale. Journal of Developmental & Behavioral Pediatrics, 34(2), 83-93.

12. Onnela, J. P., & Rauch, S. L. (2016). Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology, 41(7), 1691-1696.

13. Torous, J., Kiang, M. V., Lorme, J., & Onnela, J. P. (2016). New tools for new research in psychiatry: A scalable and customizable platform to empower data driven smartphone research. JMIR Mental Health, 3(2), e16.

14. Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. Nature, 504(7480), 427-431.

15. Chawarska, K., Macari, S., & Shic, F. (2013). Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorder. Biological Psychiatry, 74(3), 195-203.

16. Kang, J., Han, X., Song, J., Niu, Z., & Han, X. (2020). The identification of children with autism spectrum disorder by SVM approach on EEG and eye-tracking data. Computers in Biology and Medicine, 120, 103722.

17. Liu, W., Li, M., & Yi, L. (2016). Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. Autism Research, 9(8), 888-898.

18. Rehg, J. M., Abowd, G. D., Rozga, A., et al. (2013). Decoding children's social behavior. IEEE Conference on Computer Vision and Pattern Recognition, 3414-3421.

19. Hashemi, J., Tepper, M., Vallin Spina, T., et al. (2014). Computer vision tools for low-cost and noninvasive measurement of autism-related behaviors in infants. Autism Research and Treatment, 2014, 935686.

20. Purper-Ouakil, D., Wohl, M., Michel, G., et al. (2010). Ambulatory actigraphic measures of hyperactivity in children with ADHD. Journal of Attention Disorders, 14(3), 200-207.

21. Wood, A. C., Asherson, P., Rijsdijk, F., & Kuntsi, J. (2009). Is overactivity a core feature in ADHD? Familial and receiver operating characteristic curve analysis of mechanically assessed activity level. Journal of the American Academy of Child & Adolescent Psychiatry, 48(10), 1023-1030.

22. Goodwin, M. S., Haghighi, M., Tang, Q., et al. (2014). Moving towards a more ecologically valid

model of emotion regulation in children with autism spectrum disorders. IEEE International Conference on Automatic Face and Gesture Recognition, 1-8.

23. Bone, D., Bishop, S. L., Black, M. P., et al. (2016). Use of machine learning to improve autism screening and diagnostic instruments: Effectiveness, efficiency, and multi-instrument fusion. Journal of Child Psychology and Psychiatry, 57(8), 927-937.