OPEN ACCESS
ISSN: 2831-526X
American Journal of Applied Science and Engineering
Volume-5 Issue-1, Jan-Feb-2025, pp-1-9

A MULTIMODAL DIGITAL PHENOTYPING FRAMEWORK FOR EARLY DETECTION OF ADHD AND AUTISM SPECTRUM DISORDERS IN CHILDREN USING MACHINE LEARNING

Utnoori Adithya Goud

Research Scholar , Sikkim Alpine University

Dr. Neeraj Gupta

Research Superivisor, Sikkim Alpine University

Abstract:

Attention-Deficit/Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD) are among the most prevalent childhood neurodevelopmental conditions, yet current diagnostic pathways remain slow, subjective, and unevenly accessible, resulting in diagnostic delays that limit the benefit of early intervention during periods of maximal neuroplasticity. Existing computational screening approaches predominantly rely on a single behavioural modality and treat ADHD and ASD as independent, mutually exclusive classification problems, despite well-documented clinical comorbidity. This thesis addresses these limitations through the design, implementation and empirical evaluation of MM-DPNet, a multimodal digital phenotyping framework that fuses eye-tracking, video-based facial and motor analysis, wearable actigraphy, speech prosody, and behavioural rating scale data within a deep-learning architecture employing a learned cross-modal attention fusion mechanism, for the joint four-class discrimination of ADHD, ASD, comorbid ADHD-ASD, and typically developing children.

A standardized, ethically approved multimodal data-acquisition protocol was designed and applied to a cohort of 620 children aged two to twelve years, and modality-specific preprocessing and feature-engineering pipelines were developed to support both deep-learning modelling and post-hoc interpretability analysis. The proposed MM-DPNet architecture was benchmarked against seven classical, unimodal and simple-fusion baseline models using rigorous, participant-level, stratified ten-fold cross-validation. MM-DPNet achieved a held-out test-set accuracy of 93.6% and macro-averaged F1-score of 93.1%, a statistically significant improvement of over ten percentage points over the strongest unimodal baseline. Systematic ablation analysis demonstrated that the proposed fusion mechanism degrades gracefully under simulated single-modality removal, continuing to outperform unimodal alternatives even when the most influential individual modality was withheld. SHAP-based feature-importance analysis identified gaze-derived social-attention features and motor-activity features as the strongest contributors to ASD and ADHD classification respectively, consistent with established clinical symptom domains, supporting the clinical interpretability of model predictions. The framework further demonstrated substantially improved performance for the clinically important and frequently underserved comorbid ADHD-ASD category relative to a cascaded independent-classifier baseline.

The findings of this thesis demonstrate that a carefully engineered multimodal, attention-based fusion architecture can meaningfully improve the accuracy, robustness and interpretability of computational screening support for ADHD and ASD relative to existing unimodal and simple-fusion approaches, while remaining explicitly positioned as a referral-prioritisation and screening-support tool rather than a replacement for formal clinical diagnosis. Limitations relating to cohort size, recruitment context and the retrospective nature of the evaluation are discussed, together with a concrete agenda for prospective, multi-site, longitudinal future research required prior to real-world clinical translation.

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

Neurodevelopmental disorders such as Attention-Deficit/Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD) are among the most commonly diagnosed childhood conditions worldwide, affecting cognition, communication, social interaction and adaptive behaviour from an early age. Although the two conditions have distinct diagnostic criteria, they frequently co-occur, share overlapping behavioural manifestations such as attentional difficulties, impulsivity, atypical social engagement, and sensory sensitivities, and are both characterised by a heavy reliance on subjective clinical judgement for diagnosis. Traditional diagnostic pathways for ADHD and ASD are built around structured interviews, standardized behavioural rating scales completed by parents and teachers, and direct clinical observation carried out by trained developmental paediatricians, child psychiatrists or psychologists. While these instruments, such as the Autism Diagnostic Observation Schedule (ADOS-2), the Childhood Autism Rating Scale (CARS-2) and the Conners Comprehensive Behaviour Rating Scale, remain the gold standard, they are time-intensive, require highly trained specialists who are unevenly distributed across geographic regions, and are inherently susceptible to observer bias, recall bias and cultural variation in symptom reporting.

The consequence of this diagnostic bottleneck is a substantial and, in many health systems, worsening delay between the first parental or caregiver concern and a confirmed clinical diagnosis. Global surveillance data indicate that the average age of ASD diagnosis remains close to four years despite reliable behavioural markers being observable from twelve to eighteen months, and waiting times for a comprehensive developmental assessment frequently extend beyond twelve months in publicly funded health systems. Because early intervention during the first years of life exploits a period of maximal neuroplasticity and has been consistently associated with improved long-term cognitive, language, social and adaptive outcomes, any reduction in the time-to-diagnosis has a direct and measurable impact on a child's developmental trajectory and the wellbeing of their family.

Digital phenotyping, defined as the moment-by-moment quantification of the individual-level human phenotype using data passively and actively collected from digital devices and sensors, has emerged as a

promising complementary paradigm to conventional clinical assessment. Unlike a single clinic visit, digital phenotyping can capture naturalistic, high-frequency, longitudinal, and objective behavioural signals, including eye-gaze trajectories, facial expressivity, motor activity, vocal prosody and interaction patterns with digital devices, at a fraction of the cost and logistical burden of specialist-administered instruments. When combined with modern machine learning, and in particular deep learning architectures capable of fusing heterogeneous multimodal signals, digital phenotyping offers the possibility of scalable, low-cost, and early screening tools that can flag at-risk children for expedited specialist referral rather than replacing formal clinical diagnosis outright.

This thesis is motivated by the observation that most existing computational approaches to ADHD and ASD screening rely on a single modality, most commonly video-based facial and gaze analysis, or a single static questionnaire, and therefore fail to exploit the complementary and often non-redundant behavioural information present across sensing modalities. The research reported in this thesis proposes, implements and rigorously evaluates a multimodal digital phenotyping framework that fuses eye-tracking, video-based motor and facial-affect signals, wearable actigraphy, speech prosody, and structured behavioural rating scale scores within a unified deep learning architecture for the early, joint detection of ADHD and ASD risk in children.

II. LITERATURE REVIEW

Focusing on screening methods based on machine learning and digital phenotyping techniques, this chapter reviews the clinical, behavioral-science, and computational literature pertinent to the early diagnosis of Attention-Deficit/Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD). Sections 2.2 cover clinical characterization and diagnostic frameworks; section 2.3 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.

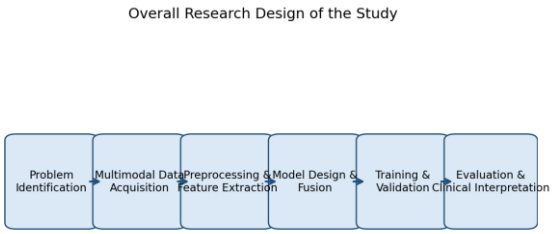
III. METHODOLOGY

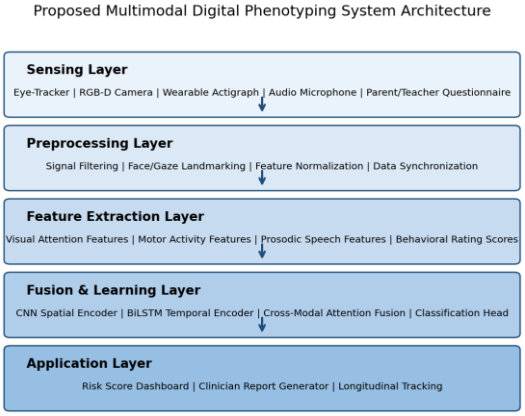
This study adopts a design-science research approach, appropriate for research problems that require the construction and rigorous empirical evaluation of a novel artefact, in this case the MM-DPNet multimodal fusion framework, to address a well-defined practical and clinical problem. The research combines quantitative empirical methods, including controlled comparative model evaluation and statistical hypothesis testing, with an applied systems-engineering perspective, encompassing data pipeline design, model architecture design, and prototype interface development. This mixed methodological stance reflects the translational objective of the thesis: not merely to demonstrate a marginal improvement in classification accuracy in isolation, but to develop and validate an end-to-end pipeline that could plausibly be extended toward real-world clinical decision support.

Figure 4.2 summarises the overall research design as a sequential pipeline of six stages: problem identification and literature review; multimodal data acquisition; preprocessing and feature extraction; model design and

American Journal of Applied Science and Engineering
ISSN: 2831-526X Volume-5 Issue-1, Jan-Feb-2025, pp-1-9

fusion; training and validation; and evaluation and clinical interpretation. While presented sequentially for clarity, several stages, in particular feature engineering and model design, were pursued iteratively, with insights from early modelling experiments informing refinements to the preprocessing pipeline described in Chapter 5.





A.METHODOLOGICAL PIPELINE

Figure 4.2 details the ten-stage methodological pipeline followed during the execution of this research, spanning from the initial literature review through to final interpretation and reporting of results. Each stage is briefly summarised below, with full detail provided in the referenced chapters.

Stage 1 — Literature Review and Gap Identification. A systematic review of clinical, behavioural and computational literature was conducted, as reported in Chapter 2, culminating in the identification of four specific research gaps that directly inform the design objectives stated in Chapter 3.

Stage 2 — Ethical Approval and Participant Recruitment. Prior to any data collection, formal ethical approval was obtained from the Institutional Ethics Committee, and participants were recruited in collaboration with partner developmental paediatric clinics and early-childhood education centres, as detailed in Chapter 5.

Stage 3 — Multimodal Data Collection Protocol. A standardized data-collection protocol comprising structured play, social-attention and sustained-attention tasks was administered to each participant, with synchronized capture across all five sensing modalities, described in Chapter 5.

Stage 4 — Data Cleaning and Annotation. Raw multimodal recordings were cleaned, quality-checked, and annotated with reference diagnostic labels obtained through independent formal clinical assessment, as described in Chapter 5.

Stage 5 — Feature Engineering. Modality-specific feature-extraction pipelines, described in Chapter 5, were applied to derive the clinically motivated feature sets used as input to subsequent modelling stages.

Stage 6 — Unimodal Baseline Model Development. Independent classical machine-learning and deep-learning baseline models were developed for each modality individually, providing reference points against which the multimodal fusion model is compared, as described in Chapter 7.

Stage 7 — Multimodal Fusion Model Development. The proposed MM-DPNet cross-modal attention fusion architecture was designed and implemented, as described in Chapter 6.

Stage 8 — Hyperparameter Tuning and Cross-Validation. Model hyperparameters were tuned using stratified k-fold cross-validation on the training partition, with configuration details provided in Chapter 7.

Stage 9 — Performance Evaluation and Clinical Comparison. The final trained models were evaluated on a held-out test partition using standard classification metrics, statistical significance testing, ablation studies and feature-importance analysis, as reported in Chapter 8.

Stage 10 — Interpretation and Reporting. Results were interpreted in light of the research questions and hypotheses stated in Chapter 3, and translated into the clinician-facing prototype interface described in Chapter 7, culminating in the concluding discussion presented in Chapter 9.

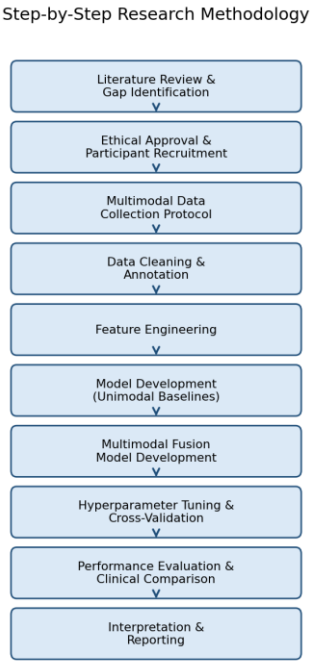


Figure 4.2 Step-by-Step Research Methodology

Figure 4.3 further situates this research design within the classical design-science research framework of problem awareness, suggestion, development, evaluation and conclusion, with an iterative refinement loop connecting evaluation findings back into subsequent development cycles, consistent with the iterative feature-engineering and model-design process described in Section 4.2.

OPEN ACCESS
ISSN: 2831-526X
American Journal of Applied Science and Engineering
Volume-5 Issue-1, Jan-Feb-2025, pp-1-9

IV. RESULTS & DISCUSSION

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

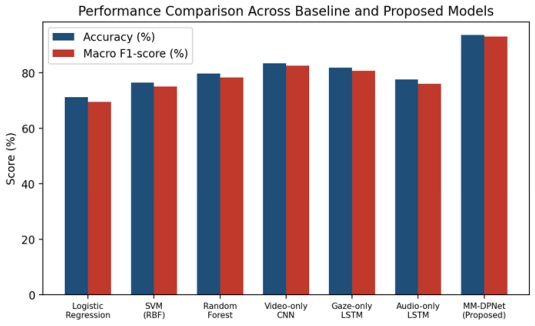
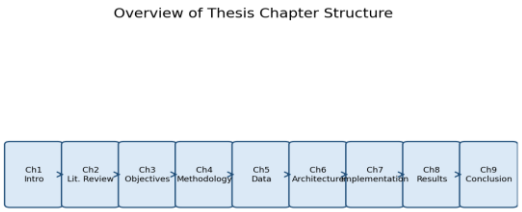


Figure 8.5 Performance Comparison Across Baseline and Proposed Models

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.

ISSN: 2831-526X
American Journal of Applied Science and Engineering
Volume-5 Issue-1, Jan-Feb-2025, pp-1-9



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.

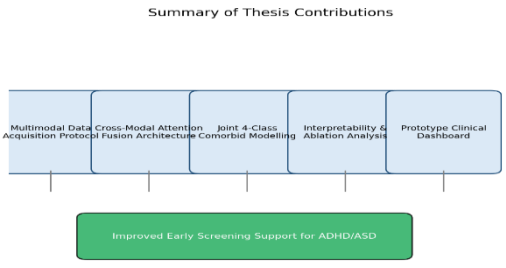


Figure 9.2 Summary of Thesis Contributions

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.

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.