Diagnostics and Prognostics with High Dimensional Spatial-Temporal Data: From Structures to Human Brains

Yan Xue[1], Yuxiang Zhou[2], Yongming Liu[3]

[1,3] Arizona State University, Tempe, AZ, 85287, USA

yxue37@asu.edu

Yongming.Liu@asu.edu

[2] Mayo Clinic Hospital, Phoenix, AZ 85054, USA

zhou.yuxiang@mayo.edu

ABSTRACT

Diagnostics and prognostics with high-dimensional spatial-temporal data require innovative methodologies due to the inherent complexity of such datasets. This thesis explores the challenges of diagnostics and prognostics in high-dimensional spatial-temporal data, extending from physical structures to complex human brain analyses through resting-state functional magnetic resonance imaging (rs-fMRI). Drawing an analogy to engineering structural health monitoring using spatial-temporal vibration data, the approach leverages techniques from engineering diagnostics and prognostics data analytics to handle clinical problems with similar characteristics. A pioneering approach is developed to analyze multimodal datasets that not only include advanced rs-fMRI features—Amplitude of Low-frequency Fluctuations (ALFF), Regional Homogeneity (ReHo), Euler Characteristics (EC), and Fractal Analysis—but also encompass a wide array of clinical data. This integration includes infant developmental metrics such as birth weight and gestational age, maternal health factors like BMI and fat mass, and environmental influences including dietary intake and mental health during pregnancy. The study establishes a robust computational framework that uses advanced machine learning algorithms to analyze the interplay of these diverse data types, enhancing the precision and predictive power of our models for early childhood development. Initial validations have demonstrated the effectiveness of this comprehensive approach in identifying ADHD, with ongoing efforts aimed at expanding the methodology to address a broader range of developmental disorders. This work not only advances the diagnostic and prognostic capabilities in medical imaging but also significantly contributes to the field of Prognostics and Health Management (PHM). By providing a solid foundation for managing and understanding high-dimensional and multimodal spatial-temporal data across various disciplines, it bridges the gap between engineering and clinical diagnostics, demonstrating the potential for cross-disciplinary innovation.

1. PROBLEM STATEMENT:

The application of high-dimensional spatial-temporal data analysis, specifically through technologies such as resting-state functional magnetic resonance imaging (rs-fMRI), presents significant engineering challenges in the diagnostic and prognostic domains, ranging from structural analysis to neurological disorders. While rs-fMRI offers unparalleled insights into brain function, leveraging this data for the early diagnosis of developmental disorders such as ADHD involves navigating vast datasets and complex interactions within the brain (Canario et al., 2021; Zhang et al., 2019). Current methodologies in engineering and data science often struggle to efficiently process and extract actionable insights from such multidimensional data, limiting their practical utility in clinical settings (Gupta et al., 2022).

Drawing an analogy from engineering prognostics, where spatial-temporal vibration data is used to monitor structural health, similar challenges are faced in clinical prognostics with rs-fMRI data. Both fields require the development of sophisticated computational frameworks to handle the complexity inherent in high-dimensional datasets. In clinical settings, this involves integrating multimodal data, including genetic, environmental, and neurophysiological information, with high-dimensional imaging data. These frameworks must not only manage the complexity of rs-fMRI data but also align it with clinical diagnostic criteria, a task beyond the capabilities of standard analytical tools and approaches.

There is a critical need for advanced engineering solutions that can synthesize these diverse datasets to improve diagnostic accuracy and offer predictive insights into the progression of neurodevelopmental disorders. Addressing this challenge requires innovative computational approaches that can parse, integrate, and analyze rs-fMRI alongside other relevant modalities, enhancing the capability to detect subtle biomarkers early in development. This thesis aims to bridge the gap between complex data analysis and clinical application by developing methods that effectively apply engineering principles to medical diagnostics. In doing so, it enhances the field of Prognostics and Health Management (PHM), demonstrating the potential for cross-disciplinary innovation to improve the early detection and management of neurodevelopmental disorders.

2. EXPECTED CONTRIBUTIONS:

The expected contributions for the project are as follows:

Contribution 1- Innovative and Scalable Diagnostic Algorithms for High-Dimensional Spatial-Temporal Data:

This thesis will introduce scalable, state-of-the-art algorithms designed specifically to handle and interpret high-dimensional spatial-temporal data obtained through rs-fMRI. These algorithms will leverage advanced machine learning techniques, including deep learning and ensemble methods, to efficiently process the data, identify intricate patterns, and detect anomalies that are indicative of developmental disorders such as ADHD. Central to these algorithms is their ability to perform sophisticated dimension reduction, essential for managing the high-dimensional nature of rs-fMRI data. This process not only enhances the interpretability of complex brain activities but also maintains the integrity and predictive quality of the data, facilitating earlier and more accurate diagnostic outcomes. The scalability of these algorithms ensures they are well-equipped to handle increasing volumes of data and growing complexity, making them indispensable for future diagnostic and prognostic applications across various domains of high-dimensional spatial-temporal analysis.

Contribution 2- Development of an Advanced Computational Framework for Multimodal Data Fusion:

We will create a sophisticated framework capable of fusing rs-fMRI data with other clinical and environmental data. This innovative framework will leverage state-of-the-art data fusion techniques to effectively manage and synthesize complex datasets from diverse modalities. By integrating genetic, environmental, and neuroimaging data, the framework is designed to deliver a comprehensive analysis of the factors influencing neurodevelopmental disorders, thus enhancing diagnostic accuracy and the capacity for effective prognostic assessments.

Contribution 3- Advanced Predictive Models for Early Diagnosis:

The research will advance predictive modeling techniques to forecast developmental disorders from an early age. By leveraging machine learning methods, these models will provide early diagnostic insights, enabling proactive clinical interventions.

Contribution 4- Impact on Clinical Practices and Public Health Strategies:

The integrated diagnostic framework will transform clinical practices by offering more accurate and timely diagnostics. It will also inform public health strategies, advocating for early screening and personalized interventions based on comprehensive neurobiological and environmental risk profiles. This could lead to policy changes incorporating advanced imaging and detailed assessments into routine pediatric healthcare.

3. RESEARCH PLAN:

The research project is projected to be completed within the scheduled three years starting August 2023. To effectively develop and validate an integrated diagnostic framework for early detection of developmental disorders, the research plan is elaborated as follows:

<u>Year One:</u>

Completion of ADHD Study: Finalize the analysis and documentation of the ADHD diagnostic study using rs-fMRI data.

Establishment of Collaboration with Medical Institution: Formalize the partnership with a medical institution providing infant data and outline data-sharing protocols. Begin preprocessing the initial set of 40 infant cases provided by the institution, ensuring data quality and consistency. Integrate initial clinical data (e.g., birth metrics, maternal health indicators) with rs-fMRI data.

<u>Year Two:</u>

Expansion of Infant Data Analysis: Preprocess and analyze more infant cases from the collaborating medical institution. Integrate the expanded dataset with existing clinical and environmental data. Develop initial predictive models for identifying developmental disorders in infants using machine learning techniques.

Development of Integrated Diagnostic Framework: Create a computational framework that combines rs-fMRI data with clinical and environmental factors. Employ advanced machine learning algorithms (e.g., deep learning, ensemble methods) to develop predictive models. Validate the models using cross-validation techniques and compare their performance against existing diagnostic methods.

Year Three:

Comprehensive Model Validation and Refinement: Perform extensive validation of the models using diverse datasets to test their generalizability. Refine models to enhance accuracy and reliability based on validation feedback.

Implementation of Diagnostic Tools: Develop user-friendly software tools or applications for clinicians to use in early diagnosis. Conduct pilot studies in collaboration with medical institutions to test the tools in real-world settings.

3.1. Work Performed:

Our research aimed at developing integrated diagnostic and prognostic models for childhood and infantile developmental disorders using resting-state functional magnetic resonance imaging (rs-fMRI). The initial phase focused on enhancing the diagnosis of ADHD, employing advanced neuroimaging techniques and data analysis methods. We utilized the DPABI Matlab toolbox(Yan et al., 2016) for preprocessing rs-fMRI data, ensuring rigorous quality control through steps including slice timing correction, realignment, coregistration, spatial normalization using DARTEL, and spatial smoothing.

To address the extensive voxel dataset, we applied two distinct region selection methods: a hypothesis-driven approach using two-sample T-tests to identify brain regions where Amplitude of Low-Frequency Fluctuations (ALFF) and Regional Homogeneity (ReHo) features significantly differed between ADHD and control groups(Yu-Feng et al., 2007), and a data-driven approach using the Smith Atlas to select predefined brain regions for further analysis. We extracted ALFF, ReHo, Euler Characteristics (EC), and Fractal features from the selected brain regions to capture various aspects of brain activity and structure.

To streamline the feature space into a more manageable format, we applied Uniform Manifold Approximation and Projection (UMAP)(McInnes et al., 2018) for dimensionality reduction. Given the often-limited sample sizes in neuroimaging studies, we employed Data Augmentation techniques to enhance the robustness of the dataset. For classification, we used a Support Vector Machine (SVM) algorithm, ensuring the model's predictive validity through cross-validation(Kim et al., 2021).

Table 1. Comparison of accuracy in ADHD-200 classification models.

Proposed paper	Model name	LOSO accuracy	Proposed paper	Model name	LOSO accuracy
Riaz et al.[2017]	FCNet	60.4	Kim et al.[2021]	ASSRNN	70.46
Riaz et al.[2018]	DeepfMRI	67.9		Our model using ALFF	71.2
Zhang et al.[2020]	SCCNN-Attention	68.6		Our model using ReHo	66.7
Zhang et al.[2020]	SCCMM-RNN	63.6		Our model using combined feature	75.5

In our study, the accuracy of ADHD diagnosis varied significantly with different brain regions. The Leave-One-Subject-Out (LOSO) accuracy of our model, which integrated multiple features, reached 75.5%, surpassing the performance of the reference model(Kim et al., 2021).

However, when applying Amplitude of Low-Frequency Fluctuations (ALFF) and Regional Homogeneity (ReHo) features to brain regions well-documented in the literature, and utilizing Uniform Manifold Approximation and Projection (UMAP) for dimensionality reduction, we did not observe a distinct separation between the control and ADHD groups.

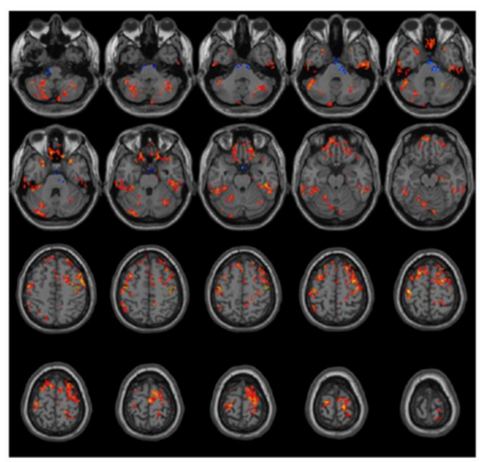

Figure 1. Brain regions with differing ReHo (blue), ALFF (red), and their overlap (green).

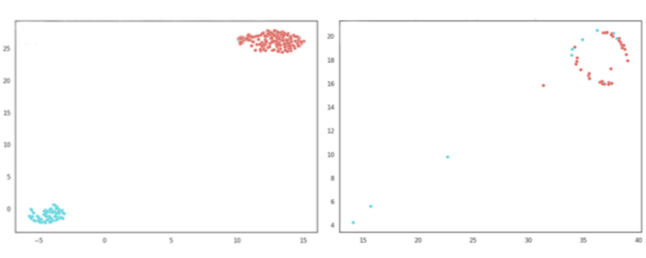

Figure 2. UMAP Projections of ALFF and Reho for training (left) and testing(right) sets, respectively, with ADHD in blue, healthy controls in red.

3.2. Remaining Work:

The next phases of our research will focus on expanding and refining our diagnostic and prognostic models for early detection of developmental disorders in infants. We will continue preprocessing and analyzing the initial 40 infant cases from the collaborating medical institution to ensure data quality and consistency, and incorporate an additional infant cases currently being collected and organized by the collaborating medical institution. Leveraging the recent breakthroughs in Curvature Augmented Manifold Embedding and Learning (CAMEL)(Liu, 2024), we will transition from using Uniform Manifold Approximation and Projection (UMAP) to CAMEL, which offers significant improvements for supervised learning tasks. We will enhance our classification algorithms with advanced machine learning techniques, including updated SVM models and deep learning, to improve accuracy and robustness. Extensive validation will ensure generalizability using diverse datasets. We will develop and pilot user-friendly diagnostic tools, and advocate for their integration into pediatric healthcare through high-impact publications and conferences.

4. CONCLUSION:

This research aims to significantly advance the early diagnosis of developmental disorders in infants by integrating advanced rs-fMRI analysis with comprehensive clinical data. Leveraging recent methodological innovations such as CAMEL and enhanced machine learning techniques, we aim to develop robust, user-friendly diagnostic tools. Our extensive validation process and collaboration with medical institutions will ensure these tools are effective and ready for clinical application, ultimately contributing to improved pediatric healthcare outcomes.

REFERENCES

Kim, B., Park, J., Kim, T., & Kwon, Y. (2021). *Finding essential parts of the brain in rs-fMRI can improve diagnosing ADHD by Deep Learning* (arXiv:2108.10137). arXiv. http://arxiv.org/abs/2108.10137

Liu, Y. (2024). *Curvature Augmented Manifold Embedding and Learning* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2403.14813

McInnes, L., Healy, J., & Melville, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction* (Version 3). arXiv. https://doi.org/10.48550/ARXIV.1802.03426

Yan, C.-G., Wang, X.-D., Zuo, X.-N., & Zang, Y.-F. (2016). DPABI: Data Processing & Analysis for (Resting-State) Brain Imaging. *Neuroinformatics*, *14*(3), 339–351. https://doi.org/10.1007/s12021-016-9299-4

Yu-Feng, Z., Yong, H., Chao-Zhe, Z., Qing-Jiu, C., Man-Qiu, S., Meng, L., Li-Xia, T., Tian-Zi, J., & Yu-Feng, W. (2007). Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain and Development*, *29*(2), 83–91. https://doi.org/10.1016/j.braindev.2006.07.002