Chapter 3: Predictive Models in ASD: An Overview

Early identification and therapeutic intervention in Autism Spectrum Disorder (ASD) is paramount. The consensus among clinicians and researchers is that identifying and supporting children with ASD as early in life as possible, ideally before the age of three years, results in much better long-term outcomes. Interventions that are designed to target the early inclinations of brain development (including treatment with Applied Behavior Analysis (ABA), play-based therapy, and speech therapy) generally work best when implemented during periods of critical brain development. Indeed, research demonstrates that children who receive intensive high-quality early interventions show greater improvements in cognitive skills, language abilities, and adaptive functioning compared to children who receive later diagnosis and intervention (Dawson et al., 2010; Zwaigenbaum et al., 2015).

A late diagnosis not only delays the availability of these essential services but also places a substantial emotional burden on families and an economic one on society. For example, because children may miss the critical windows for learning social and communication skills that they will need if their diagnosis is delayed, families may be left to grapple for years in a state of limbo and engage in unguided efforts to help their child. However, traditional diagnostic methods are lengthy, expensive, and require intensive training for such assessments; many clinicians still resort to in-person diagnosis, using the Autism Diagnostic Observation Schedule (ADOS-2) and the Autism Diagnostic Interview-Revised (ADI-R). Youth commonly wait for a trained diagnostician, particularly in rural or low-resource areas, resulting in long waiting lists and an average age of diagnosis that is still higher than ideal (Wiggins et al., 2015). This diagnostic bottleneck highlights the need for a paradigm shift, from reactive to predictive identification.

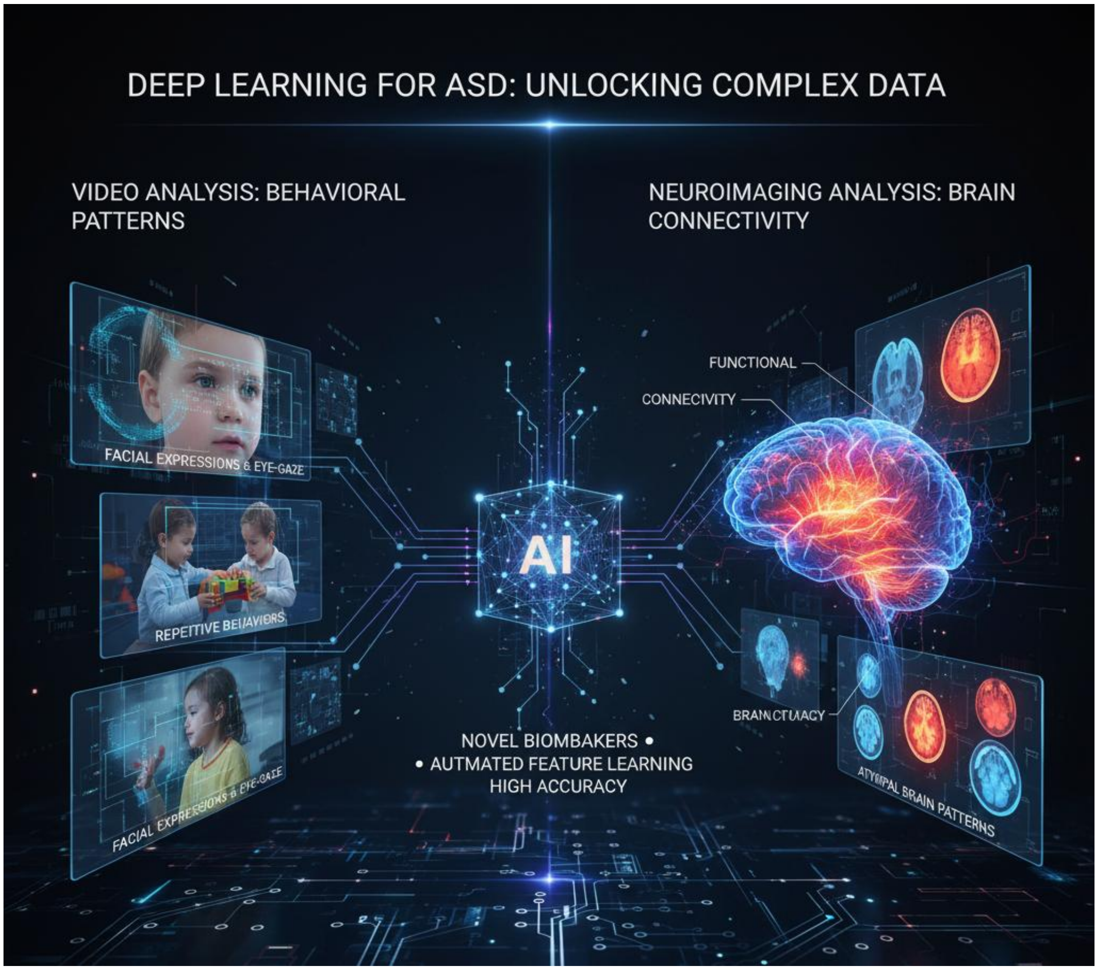
Predictive modeling using machine learning can provide a scalable, objective method for identifying those at highest risk of ASD. These models use data to surface patterns, correlations, and anomalies that are often too subtle or complex for human beings to observe. By crunching through extensive data, including early behavioral milestones a child accomplishes, which are documented in EHRs, to high-resolution digital data from video and audio recordings, machine learning (ML) models can identify potential cases with high precision at much earlier ages. For example, one might train a model on data from parent-completed questionnaires that identify certain features (e.g., atypical eye contact, vocalizations, or motor movements) that, along with them, predict a high likelihood of having ASD. One strength of these models is their capacity to incorporate data from multiple modalities. For example, they could integrate genetic risk scores with early-life environmental exposures and a child's developmental trajectory to generate better and more accurate predictions of risk. Given that a biosignature such as this could be further developed into an early diagnostic aid in pediatric clinics, it would prompt healthcare workers to the necessity of comprehensive testing and cut intervention times by orders of magnitude. Instead of waiting for a developmentally delayed child to show more apparent signs or for a parent to raise concern, predictive models could put the diagnostic wheels in motion early on, leaving no child behind. This data-driven discovery model may revolutionize the diagnostic pathway, rendering it a less biased and more efficient process that is better equipped to promote effective outcomes for individuals with ASD.

3.1 Types of Predictive Models Applied to ASD

The use of predictive models in ASD research falls within a broad category of machine learning methodologies, each of which is appropriate for specific analytical objectives. Most of the models are classification models because we are responding to the binary question of whether an individual has been diagnosed with Autism Spectrum Disorder or not. Several 'traditional' ML algorithms have been successfully used for this purpose. One type of ML model, decision trees, does so literally by building a model of decisions (in the diagnostic case, about whether or not to classify an individual as having ASD) based on simple rules that can be generated from data, such as: "if child has delayed speech AND prefers sameness, then classify as autism." The simplicity and interpretability of decision trees render them a valuable tool for clinicians. Even more advanced (and hence usually a better classification model) are methods such as Random Forests and SVM. In particular, SVMs are well-suited for ASD research, particularly task-related neuroimaging studies, because they can find a hyperplane to discriminate between two classes of data points in multidimensional space (e.g., distinguish brain scans of people with AS from those with TD) (Solek et al., 2024; Thabtah & Peebles, 2020). These models, based on data obtained from clinical evaluations, behavioral observations, or genetics, have yielded promising results in identifying the association between individuals with and without ASD (Thabtah et al., 2020).

In addition to mere classification, the regression models predict a continuous measure (e.g., the severity of ASD symptomatology, ranging from 0.3 to 1, or the social responsiveness score of an individual). Based on patients' data, a regression model can estimate the severity score of their symptoms, providing clinicians with a numerical measure to track a patient's progression over time or evaluate the efficacy of an intervention. For instance, a model might be trained on behavioral data from a child at different ages to predict their future social communication scores, allowing therapists and parents to anticipate when the child may struggle and tailor interventions accordingly. This shift goes from a "one-shot" diagnosis towards dynamic monitoring of the process. Although conventional regression models are helpful, the complexity of ASD necessitates more advanced nonlinear techniques for modeling to fully describe the complex relationships among different features.

The most significant improvements over the past few years have been driven by deep learning models, a family of neural networks that excel at making sense of unstructured, high-dimensional data, such as video, audio, and raw neuroimaging scans.



Convolutional Neural Networks (CNNs), for example, are employed to analyze facial features extracted from video recordings to identify subtle variations in expressions or eye-gaze patterns that may be indicative of ASD (Tariq et al., 2023). However, another compelling use case is for neuroimaging, where deep learning can be used to process fMRI data in the search for abnormal functional connectivities in the brains of individuals with ASD, and to diagnose them with a high degree of accuracy compared to control groups (Heinsfeld et al., 2018). Such models work so well because they can automatically learn intricate features from the raw data without requiring us to perform elaborate feature engineering. For example, a CNN does not need to be programmed with information about what a "social smile" is; the system learns the pattern for itself from thousands of examples. Although they are computationally demanding and data-intensive, these deep learning methods could mark the next era for ASD prediction models, with the promise of uncovering previously inaccessible biological and behavioral features that were overlooked by classical approaches.