Predicting Familial Likelihood of Autism Spectrum Disorder in Infancy Using ECG

Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Technology
in
Computer Science and Engineering

by

Deepa Tilwani
19MCS002
19mcs002@lnmiit.ac.in

Under Guidance of
Dr. Sakthi Balan Muthiah
and
Prof Amit P. Sheth
Artificial Intelligence Institute, University of South Carolina

Department of Computer Science and Engineering
The LNM Institute of Information Technology, Jaipur

July 2022

Predicting Familial Likelihood of Autism Spectrum Disorder in Infancy Using ECG

Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Technology
in
Computer Science and Engineering

by

Deepa Tilwani
19MCS002
19mcs002@lnmiit.ac.in

Under Guidance of
Dr. Sakthi Balan Muthiah
and
Prof Amit P. Sheth
Artificial Intelligence Institute, University of South Carolina

Department of Computer Science and Engineering
The LNM Institute of Information Technology, Jaipur

July 2022

The LNM Institute of Information Technology
Jaipur, India

CERTIFICATE

This is to certify that the thesis entitled *"Predicting Familial Likelihood of Autism Spectrum Disorder in Infancy Using ECG"* submitted by *Deepa Tilwani (19MCS002)* towards partial fulfillment of the requirements for the degree of Master of Technology (M. Tech), is a bonafide record of work carried out by him/her at the Department of Computer Science and Engineering, The LNM Institute of Information Technology, Jaipur, (Rajasthan) India, during the academic session 2019-2022 under my supervision and guidance. The content of the thesis is original and have not been submitted elsewhere for award of any other degree. In my opinion, this thesis is of the standard required to award the said degree

28/07/2022
Date

Adviser: Dr Sakthi Balan Muthiah

Dedicated to my parents, my sister and people diagnosed with Autism Spectrum Disorder.

Acknowledgments

First and foremost, I am grateful to my supervisor, Dr. Sakthi Balan Muthiah, for his guidance, constant support, and persistence throughout my MTech thesis. His understanding and knowledge have motivated me throughout my academic research. I would also want to thank Dr. Amit P. Sheth (Artificial Intelligence Institute, University of South Carolina (AIISC)) for providing the research internship, Dr. Christian O'Reilly (AIISC) for guidance and his expertise in biosignals and Dr. Jessica Bradshaw (University of South Carolina) for collecting and providing data for analysis of this project. I want to express my gratitude to Dr. Manas Gaur and Dr. Ugur Kursuncu for realizing the project on neuro-developmental disorders. I want to express my gratitude to everyone in the Department of Computer Science and Engineering at the LNMIIT for their generous support in my research and making my living so enjoyable. Finally, this would not have been possible without the constant support and encouragement from my parents and siblings, because of whom, I am able to complete and craft my thesis.

Abstract

Autism Spectrum Disorder (ASD) manifests in difficulties in social skills, repeated habits, verbal and non-verbal speech, and individual strengths and differences. According to the World Health Organization, ASD has been on the rise for the past 20 years, affecting 1 in 44. Unfortunately, early diagnosis of ASD in infants is impossible to conduct because electroencephalograph (EEG) signals are difficult to record. However, clinicians can analyze brain development for behavioral data after 2 or 3 years from the time of EEG recordings. Thus, there is a need for a biomarker that can help assess infants early and help health experts design a better treatment plan. One supportive biomarker is Electrocardiograms (ECG), which have not been explored to diagnose ASD. ECGs are cost-effective, easy to record, and can be monitored at the clinic, home, or school. We explore ECG as an alternative to EEG signals and study its potency in predicting the familial likelihood of ASD using a machine learning algorithm. Specifically, we look at subjects who are infants between 3 and 6 months of age for diagnosing ASD using ECG signals. We explored heart rate variability (HRV), and nonlinear measures, such as discrete functional analysis and sample entropy, were computed from ECG and used for classification in the multi-layer perceptron model. We evaluated the model using metrics: accuracy (82.6%), sensitivity (75.0%), specificity (86.7%), F1-score (75.0%), and precision (75.0%). These findings suggest that ECG signals yield information about the likelihood of ASD in infants and could be used as a high potency biomarker for the early diagnosis of ASD.

Acronyms

ADOS The Autism Diagnostic Observation Schedule (ADOS

ADI The Autism Diagnostic Interview

ANS Autonomic Nervous System

ASD Autism Spectrum Disorder

AQ The Autistic Quotient Trait

CDC Center for Disease Control and Prevention

CVI Cardiovagal Index

CVNN Coefficient of variation of normal beat intervals

CSI Cardiac Sympathetic Index

DFA Detrended Fluctuation Analysis

DWT Discrete Wavelet Transform

ECG Electrocardiograms

EDF European Data Format

EMD Empirical Mode Decomposition

EEG Electroencephalograph

fMRI Functional Magnetic Resonance Imaging

HL High Likelihood

HRV Heart Rate Variability

HTI Triangular index of HRV

IMF Intrinsic Mode Function

ML Machine Learning

MLP Multilayer Perceptron

MRI Magnetic Resonance Imaging

NN Normal Beat Intervals

OIX Object Interaction Experiment

pNN Proportion of successive normal beat intervals

PIX Parent Interaction Experiment

LL Low Likelihood

SD Standard Deviation

SQC The Social Connection Questionnaire

SampEn Sample Entropy

TP True Positives

TN True Negatives

Contents

Chapter Page

1 Introduction . . . 1
1.1 The Area of work: Autism Spectrum Disorder . . . 1
1.2 Thesis Motivation . . . 2
1.3 Problem Statement . . . 3
1.4 Thesis Aim . . . 3
1.5 Thesis Contribution . . . 3
1.6 Thesis Organization . . . 4
1.7 Outreach . . . 4

2 Literature Survey . . . 5

3 Data Preparation . . . 8
3.1 Data Collecion . . . 8
3.2 Preprocessing . . . 9
3.2.1 Denoising . . . 10
3.2.1.1 EMD . . . 10
3.2.1.2 DWT . . . 11
3.3 Feature Extraction . . . 13

4 Proposed Work . . . 15
4.1 Pipeline . . . 15
4.2 Method : Multilayer Perceptron . . . 16

5 Results . . . 19
5.1 Results . . . 19
5.2 Discussions . . . 21

6 Conclusions and Future Work . . . 24

Bibliography . . . 25

List of Figures

Figure Page

3.1 Acquisition protocol 9
3.2 PQRST wave in ECG signal 9
3.3 EMD example 11
3.4 Db4 wavelet from Daubechies Family 12
3.5 DWT approximation and details coefficients structure. 12

4.1 Noisy and cleaned QRS complexes 15
4.2 Proposed method for EL vs LL classification. 17
4.3 Structure of a Neuron 18
4.4 Multilayer Perceptron Model 18

5.1 Effect size for extracted features. 22

List of Tables

Table		Page
1.1	Thesis Contribution	4
3.1	Features Extracted from ECG	14
5.1	Features are represented as group mean $\pm$ std. Reported effect sizes are for Cohen's d, which are generally understood as small, medium, and large effects, respectively, for values equal to 0.2, 0.5, and 0.8.	19

Chapter *1*

Introduction

1.1 The Area of work: Autism Spectrum Disorder

Autism spectrum disorder (ASD) is a behavioral syndrome that can be diagnosed as impairments in social contact/interaction or the presence of limited or repetitive behavior, activities, or interests. Children with non-specific symptoms may exhibit abnormal sensory processing abilities, perceptions, physical clumsiness, and anxiety. In ASD, children after three years of age or adults tend to show up symptoms more readily than infants or children up to 1-2 years because behaviors can develop after reaching three years of age. ASD has been linked to genetics and neurologic components; however, it is primarily diagnosed using non-genetic criteria relating to social interaction, imagination, repetitive behaviors, and speech. According to most studies, the actual prevalence of ASD is an average of 1 in every 44 (2.3%), and increasing over the last two decades [30]. To generalize this, there is a need for quicker and more efficient diagnosis solutions, which corresponds to the increased awareness of ASD [41].

There are some behavioral tests and electrophysiological observations for diagnosing ASD. But these tests are time-consuming [34]. Recently, there has been an increased demand for automating procedures to reduce the expense and time associated with the test. The current diagnostic time frame is a significant barrier to treatment. Due to the lengthy process, it will take up to six months to identify a child with ASD. A child must see various doctors, beginning with behavioral pediatricians, neurologists, physicians, or psychologists. The period required to complete an ASD diagnosis is relatively lengthy and time-consuming. An essential step in helping and reducing the symptoms of ASD and improving the quality of life for people who have it is to diagnose it early in life.

On the other hand, there is no known medical test for early diagnosis of ASD. To ascertain whether an individual has a problem with social interaction, doctors must observe that in later years of life rather than in newborns. Additionally, parents and teachers should be mindful of warning symptoms of ASD in adolescent students who go to prevent it from worsening in the future. As a result, machine learning (ML) techniques will significantly improve the operation. It is well established that early intervention is

critical for improving treatment for ASD. Significant technology and ML systems have the potential to make enormous strides in predicting and speeding up complicated and time-consuming diagnostic and care procedures.

Based on these assumptions, a feasible biomarker that can help identify ASD in infancy rather than focusing on behavioral characteristics later in life is needed. In particular, ECG is a well-known tool for assessing the autonomic nervous system (ANS). The ANS is one of the body's major systems for cell functioning. When someone is placed under stress, they become sympathetic, whereas parasympathetic activity increases in an individual during rest. The incidence of increased sympathetic activity has been reported in the literature addressing ANS and ASD, particularly during behavioral trials [4]. Aside from physiological conditions, interactions with the environment and cognitive factors all impact the ANS balance. Heart rate variability (HRV) is the most widely used and flexible tool for measuring ANS extracted from ECG. Because heart rate is influenced by both sympathetic and parasympathetic, its variability can reflect ANS, particularly parasympathetic activity. Therefore, we focused on extracting these features from ECG in individuals and investigating them to study if ECG can be considered a biomarker in ASD.

1.2 Thesis Motivation

Early detection of autism is critical and early intervention is well established in the literature [40, 42]. Newborns are unlikely to exhibit ASD signs; thus, recognizing early symptoms might be challenging for healthcare practitioners, according to American Psychiatric Association, 2000 [2]. Pediatricians have a limited time at each session to discover autistic symptoms. However, each trial for a single patient might take up to 120-180 minutes, which is a significant constraint in observing and determining outcomes. The most critical component in recognizing and treating ASD is to follow newborn development daily. The best person to know their child for developmental observations is a parent. Their monitoring for identifying abnormal changes in the early stages is vital for health experts. According to recent studies [18], parents are frequently right about their child's growth. Pediatricians must consider every scenario offered by parents. So, there is a need for biomarkers that can assist parents in analyzing their children as they develop and keep track of behavioral, social interactions, and learning skills. For this purpose, we analyze ECG as a biomarker, which can be quickly recorded by a small device anywhere and help keep track of heart rate variability.

1.3 Problem Statement

ASD is a syndrome that refers to a set of neurodevelopmental conditions. It encompasses various symptoms due to its "spectrum" nature, which refers to varying degrees of impairment, abilities, and features. Due to this, children with ASD have difficulties in speaking and engaging with peers and others. Anyone who is mildly affected by these signs is considered disabled. The Center for Disease Control and Prevention (CDC) states approximately one in every 44 children is diagnosed with ASD [31]. Early intervention and services will significantly decrease the effect and help cope at an early age. If parents or guardians notice these symptoms, they can help the child to get an ASD evaluation. The primary goal of this work is to speed up ASD diagnosis by developing an ML and time series analysis techniques pipeline that can predict the emergence of ASD using ECG data.

1.4 Thesis Aim

According to available research [27], ASD develops in childhood and continues into adolescence and adulthood. There is no proper medical test available to determine the condition and diagnosis of ASD more quickly. Instead, an examination is conducted based on a medical professional or specialist's observations of a child's actions and behavior. The most critical steps toward mitigating the effects of ASD and improving the well-being for individuals with ASD are early diagnosis and treatment. ECG is included in the broad category of medical biometrics due to its widespread use in clinical diagnostics. With the increased use of ML algorithms [37] in medical diagnostic science, this work examines the likelihood of detecting and evaluating ASD problems in infants using raw ECG. This study compiled a thorough list of distinguishing features and analyses based on activities recorded for each infant using quantitative measurements taken from raw ECG data after signal processing and noise reduction. We considered infants between the ages of three to six months for this study.

The goal was to determine whether an infant had a high likelihood (HL) or a low likelihood (LL) for ASD.

Using an ECG to predict a child's risk of developing ASD will significantly *speed up the current time-consuming diagnosis process.*

1.5 Thesis Contribution

This study highlights the present state of research in the field of ASD research using ECG. ASD-related characteristics were extracted and we compared HL to LL children. Our contribution to the current results is summarized in Table 1.1. In Table 3.1, a full explanation of these features is provided. Our findings revealed a difference between children identified with HL and those diagnosed with LL. This data suggests that ECG could be used as a biomarker for ASD. In Table 1.1 *, authors have used

	Raw ECG	Mean	Median	CVNN	pNN20	SD1SD2	CSI	CVI	SampEn	DFA	HTI
Time Domain Analysis of Heart Rate Variability Signals in Valence Recognition for Children with ASD* [3]	Yes	Yes	Yes	-	-	-	-	-	-	-	-
Can a composite heart rate variability biomarker shed new insights about autism spectrum disorder in school-age children?* [21]	-	Yes	Yes	-	-	-	-	-	Yes	-	-
Heart Rate Variability During a Joint Attention Task in Toddlers With ASD* [7]	-	-	-	Yes	Yes	-	-	-	-	-	-
Our Approach	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 1.1 Thesis Contribution

other features which are not mentioned in this table, but we didn't find it relevant while ablation study and decided not to used them in our study.

1.6 Thesis Organization

The thesis is organized as follows: Chapter 2 discusses the works that are relevant to this study. Chapter 3, discuss about pipeline data collection, pre-processing, and feature extraction. The model design and method are discussed in Chapter 4. In Chapter 5, the study results and discussion are described in detail. Finally, the conclusion and future work are discussed in Chapter 6.

1.7 Outreach

This work highlights a critical problem in the early diagnosis of ASD. For visibility and acceptability of this research we will submit it to the ACM Transactions on Computing for Healthcare (HEALTH) Journal.

Chapter *2*

Literature Survey

In recent decades, children with ASD and diagnosing technologies have been the topic of an increasing number of research investigations. ASD research has been interdisciplinary, utilizing various tools and techniques for multiple goals in educational, psychiatric, and engineering sciences. ASD is a developmental brain impairment that affects the natural development of language and social behaviors [11]. In particular, genetic and neurological variables have been linked to the origin of ASD. In addition to its genetic foundations, ASD is often diagnosed based on behavioral features like social interaction, creative abilities, repetitive behaviors, and communication [13]. These behavioral and communication difficulties can range from challenges with senses (speaking, tasting, and feeling) to lagging language learning and making it difficult to express themselves. ASD is clinically diagnosed utilizing key indications based on behavioral or repetitive gestures.

Children with ASD are difficult to differentiate from healthy infants throughout their first year of life because their brains are still developing. Atypical language trajectories have also been reported, with minor delays at 12 months rising to more substantial delays at 24 months [46]. Such crucial signs as lack of social connection and restricted/repetitive behaviors and desires are revealed by age three. Kids with ASD may be easily distinguished from other behavioral changes during the late preschool and early primary school years. Some experts believe that around 50% of diagnosed children after three years of age have behavioral problems such as aberrant behavior, poor eye contact, and a lack of understanding of parents' voices or interactions. For diagnosing ASD, now tests have been developed using combinations of interviews from a parent, carer, or child. In non-clinical self-administered or parent-based interventions, the autistic quotient trait (AQ) [5] and the social connection questionnaire (SCQ) [6] have been utilized. But AQ can be used for screening people aged 16 years and older. In screening for ASD in high-risk children, the SCQ repeatedly demonstrated a poor balance of sensitivity and specificity [32]. Two other widely used assessment techniques that aid the diagnosis process are the Autism Diagnostic Observation Schedule (ADOS) [28] and the Autism Diagnostic Interview (ADI) [29]. ADOS provides four separate evaluation phases, each applicable to a specific developmental age range and applies across the spectrum from verbally proficient to nonverbal. Patients are evaluated based on their performance with a set of age-appropriate tasks linked with each module, as observed and

reported. ADI is divided into five phases based on communication, social development, and behaviors. Each phase is scored from 0 to 9. When scores in all three categories surpass the prescribed minimal cutoff levels, the patient is diagnosed with ASD. These tests need responses to many questions, consume too much time, and are inefficient. Furthermore, the most challenging aspects of utilizing the ADI and ADOS scores to identify children with ASD is that their findings are not always consistent, and there is no systematic procedure for dealing with discrepancies if one reveals the existence of an ASD while the other does not [15].

For diagnosing ASD, brain image techniques are often considered. A whole-brain magnetic resonance imaging (MRI) scan of high-functioning ASD children and those with developmental language disorder, for example, found reduced functional connectivity between the frontal and parietal-occipital areas, as well as the anterior and posterior insula, striatal subregions, and limbic cortex. The scientists employed functional magnetic resonance imaging (fMRI) to explore the socioemotive circuits of autistic children, especially the amygdala cortical pathways and observed aberrant connectivity of their sensory input [24]. Another study on functional magnetic resonance imaging (fMRI) research, consists 59 infants of age 6-month, revealed substantial changes in the brains of children who would be diagnosed with ASD at 24 months [20]. While these findings using MRI and fMRI are scientifically intriguing, a major barrier must be overcome before biomarker measuring techniques are created for clinical usage. Any brain measuring method should be low-cost and straightforward to apply in the context of an infant to be feasible in a primary care environment.

Electrophysiological observations have the potential to be a reliable biomarker for ASD. EEG is one such modality that is increasingly being examined as a valuable diagnostic tool for detecting aberrant neurodevelopment [10]. Nonlinear approaches and statistical learning for the early identification of ASD using EEG data have been explored widely. The installation of EEG sensors, on the other contrary, can be time-consuming and inconvenient for children with sensory and behavioral issues. EEG is also sensitive to various noise sources, including body movements, heart rates, and eye movements. Furthermore, even professionals trained to analyze EEG often have difficulties identifying some types of noise [38].

ECG is another physiological observation being investigated for ASD. ECG is captured by wireless or wired sensors frequently utilized in heart monitoring and cardiovascular analysis. It analyzes the electric field between the electrodes on the body to record the electrical signal. It is a non-stationary, time-varying signal in which the intervals between consecutive heartbeats change with time. The heart rate is a commonly utilized characteristic generated from ECG readings. A typical ECG pulse period indicates a steady heartbeat that frequently includes peaks and ECG waves. ECG signals can depict our hearts' complex mechanical and electrical processes, hence symbolizing the heart's most helpful information. Because of its stability and depiction of the heart, ECG is currently widely employed in various devices such as cardiac diagnosis of disorders, sleep apnea prevention, and heart rate tracking. Another ECG-derived metric is HRV. In recent years, attention has focused on HRV for diagnosis, which has emerged as a sensitive and noninvasive predictor of ASD extracted from ECG [17]. HRV is a method

for determining the difference in beat-to-beat cycles of heart rhythms, and it influences parasympathetic and sympathetic autonomic function behavior.

HRV in individuals with ASD is lesser than in non-autistic populations [43]. HRV changes have been linked to many behavioral changes and physical health issues caused by ANS dysfunction. They compared resting-state HRV in persons with ASD to those who were not autistic. The outcomes of their investigation revealed that those with ASD had lower HRV than adults without ASD [43]. These findings suggest that ANS dysfunction might indicate a high risk for ASD. Researchers investigated that ANS function can be evaluated by HRV in people with ASD [7]. Also, adults with ASD were anticipated to have lower HRV than normal subjects. This thesis looks at the ANS functioning in infants with ASD from both the LL and HL groups. We hypothesized that HL infants would have lower HRV than LL infants. We extracted the HRV features and used them further for classification.

Chapter *3*

Data Preparation

3.1 Data Collecion

A sample size of 75 ECG signals were obtained from *N=23* infants aged 3-6 months. Subjects were either at a high familial probability for ASD (HL), since they had an older sibling with ASD *(n=7)* or at a low familial probability for ASD (LL) because no first or second-degree relatives had ASD *(n=16)*. On the infant's chest, a tiny device containing an ECG sensor (Active Cardio; CamNTech) with two electrodes was attached, and raw ECG data were captured in the laboratory. These electrodes are connected to a galvanometer to calculate the potential difference between the peaks. Infants took part in two tests in which they engaged with either items (toys) or their parents. During the object interaction experiment (OIX), the infants were put on their parents' laps, and five toys (objects) were shown on a table in front of them for 1-2 minutes. During the first trial, the parent did not engage with their newborn. The newborns were put on a table during the parent interaction experiment (PIX), and parents engaged with them without toys for a total of nine minutes in the following sequence: five minutes of interaction, two minutes of ignoring the infants, and two minutes of extra interaction. Furthermore, all of the sessions were video taped using a camera attached to the computer to record stimulus (OIX/PIX) time logs and saved in mp4 format. Figure 3.1 shows the data acquisition protocol.

Both experiments were carried out on the same infant. Some patients have only OIX or PIX recordings, either due to technological difficulties or because newborn fussing probably caused an early conclusion of the trial. The European Data Format (EDF) was used to save the ECG recordings. To facilitate exact annotation of interaction time in OIX/PIX, ECG and videos are synced. ECG segments were manually retrieved for the OIX and PIX trials using EDFbrowser [1]. Each segment continued for 8-12 minutes, for a maximum of about 20 minutes per child. Each ECG recording is segmented in 35 to 40 contiguous 30-second slice using EDFbrowser. OIX and PIX segments were combined for the analysis. Data collection is done at the Early Social Development Lab, University of South Carolina [2].

[1] https://www.teuniz.net/edfbrowser/

[2] https://www.esdilab.com/

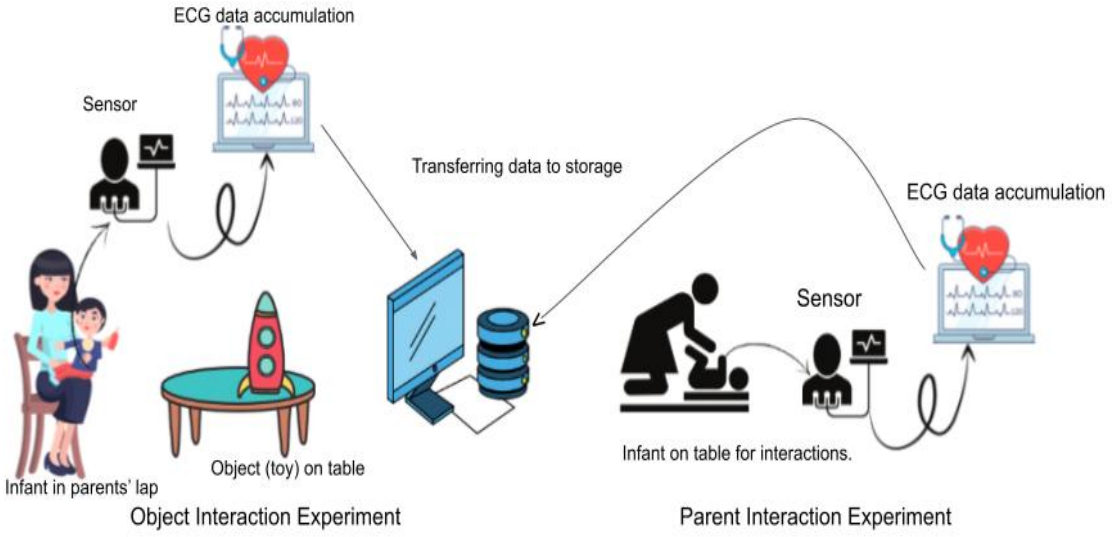

Figure 3.1 Acquisition protocol

Experiment with an object interaction in the left, with a toy on the table and an infant on the parent's lap. Experiment on the right side with the infant on the surface and the parent interacting with the infant. The ECG sensor collects all data and stores it in the same location

3.2 Preprocessing

The initial data preparation stage for the machine learning model is ECG signal processing, which is necessary to reduce noise from input data. ECG consists of morphological and temporal features, such as the P, QT, QRS and T waves.

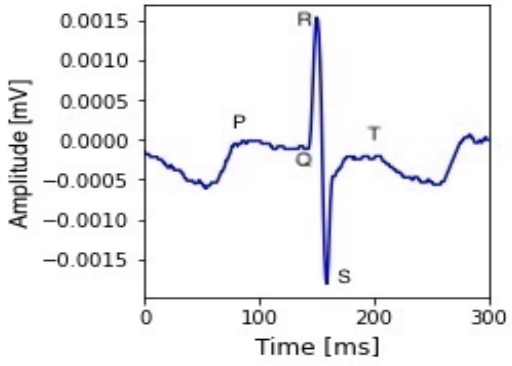

Figure 3.2 PQRST wave in ECG signal

Combining all these waves, a single wave is formed. It is known as the PQRST wave, as shown in Figure 3.2, which is often contaminated by the noise. There is a need for preprocessing step of the ECG data before feature extraction, which helps to improve retrieved features and increase system

performance. Several types of noise can contaminate ECG. During the acquisition of ECG signals, various types of noise are introduced, including baseline drift caused by the child's inhalation and exhalation, electromyogram noise generated by the contraction and movement of human muscles, artifacts caused by electrode movement (due to sudden body movements), channel noise caused by white Gaussian noise introduced during channel transmission, power line interference, and other types of excessive noise. These forms of noise can produce distortion in the ECG waveform, obscuring the small details that are crucial for diagnosis. As a result, many denoising approaches have been developed to optimize the extraction of relevant information from the ECG.

3.2.1 Denoising

Artifacts have a negative influence on automated classification. Because of this, we cleaned the ECG signals using two parallel techniques available in the literature, which are based on signal reconstructions utilizing empirical mode decomposition (EMD) [9] and discrete wavelet transform (DWT) [33].

3.2.1.1 EMD

Artifacts that affect ECG most are baseline wander. This artifact appears from respiration and lies between 0.15 and 0.3 Hz. This artifact's denoising signal is needed to diminish the irregularities in beat morphology. EMD is a powerful, valuable method for eliminating baseline wander by decomposing non-stationary and nonlinear signals. A signal x(t) is decomposed into oscillatory components for which the instantaneous frequency can be defined, and these components are called intrinsic mode function (IMF). These IMF can satisfy two conditions: (1) the number of local maxima and local minima in the entire data must either equal or differ by one; and (2) the mean m(t) of the envelope of local maxima at any point and the envelope defined by the local minima is zero. The envelopes calculated for each IMFs are symmetric, and the mean value is zero. Due to these qualities, IMF is known as being monocomponent. In addition, each mode function carries information on how the frequency of the original signal varies over time when it decomposes a time-domain signal into IMFs. As a result, the EMD doesn't need to be linear or time-invariant.

First IMF is calculated as the difference d(t) between x(t) and m(t), relation for d(t) can be showed as d(t):

$$d(t) = x(t) - m(t) \tag{3.1}$$

Now, d(t) is treated as a new signal and using this new signal, its upper and lower envelopes and their mean $m_2(t)$ is calculated, then a new diffrence as $d_2(t)$:

$$d_2(t) = d(t) - m_2(t) \tag{3.2}$$

This process is continued and checks the satisfying condition for IMF each time until the last residue has no value variation. The original signal can be reconstructed as the sum of IMF's and the final residue.

$$x(t) = \sum_{i=1}^{n} IMF + h_n(t) \tag{3.3}$$

We chose *n=3* because early evaluations revealed that the first three IMFs had the best signal-to-noise ratio for QRS complexes. An output for EMD on raw signal is shown in Figure 3.3.

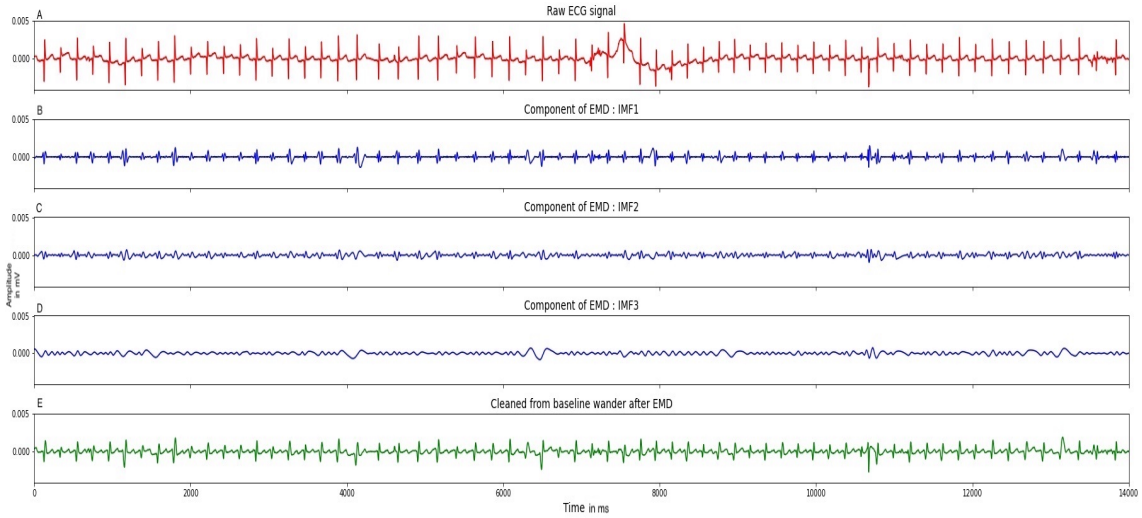

Figure 3.3 EMD example

(A) A 30 segment raw ECG signal with baseline wander. (B-D) EMD components (IMF1, IMF2, IMF3) calculated on raw ECG. (E) Clean signal obtained from EMD after adding components (IMF1+IMF2+IMF3).

3.2.1.2 DWT

The DWT is a valuable technique for analyzing non-stationary data and a commonly utilized for ECG noise removal. DWT offers to represent a signal in both frequency and time domain to explore quasi-periodic properties of ECG signal. The DWT signal is represented as a linear combination of the wavelet coefficients and the mother wavelet. It is defined as:

$$DWT(x[t], y_{high}[t], y_{low}[t]) = 2^{\frac{-y_{high}}{2}} \sum x[t]\psi^*(2^{-y_{high}}t - y_{low}) \tag{3.4}$$

where, x[t] is the ECG signal, ψ^* is the complex conjugate of the analyzing wavelet function, $y_{high}[t]$ and $y_{low}[t]$ are the high pass and low pass filter parameters respectively [1]. A filter bank is used to create a multi-resolution decomposition of signals (described in Equation 3.4). If N is the length of the signal, then it is decomposed up to $log_2 N$ levels by down-sampling the signal based on desired cut-off frequency. High pass and low pass filters are used to downsample the signal iteratively. Equations 3.5 and 3.6 shows the output of the filters: high $(h[n])$ and low $(g[n])$. Wavelet decomposition is performed up to level 6 as shown in Figure 3.5, resulting in a set of approximation coefficients (C6) and six sets of detail coefficients (d1,...,d6).

$$y_{low}[t] = \sum_{t=-\infty}^{\infty} x[t]g[2n-t] \tag{3.5}$$

$$y_{high}[t] = \sum_{t=-\infty}^{\infty} x[t]h[2n-t] \tag{3.6}$$

Wavelet families include Daubechies, Symlets, Biorthogonal, Haar, Coiflets, Morlet, Meyer, and Mexican Hat. However, the Daubechies (Db4) wavelet has been demonstrated to provide more precise information for ECG signals than others. Significantly, this wavelet resembles QRS complexes, and the power spectrum is focused around low frequencies. As a result, we selected the Db4 wavelet to denoise in our system. Db4 wavelet is shown in Figure 3.4

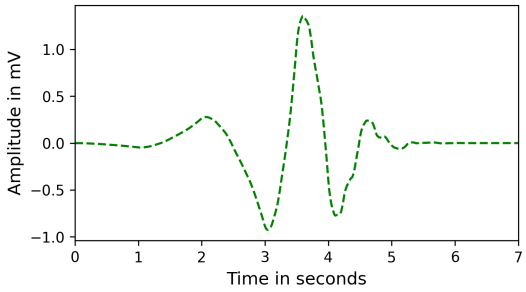

Figure 3.4 Db4 wavelet from Daubechies Family

$$[C_n, d_n] = DWT(s, \tau) \tag{3.7}$$

here, C_n and d_n are approximation and details coefficients vectors.

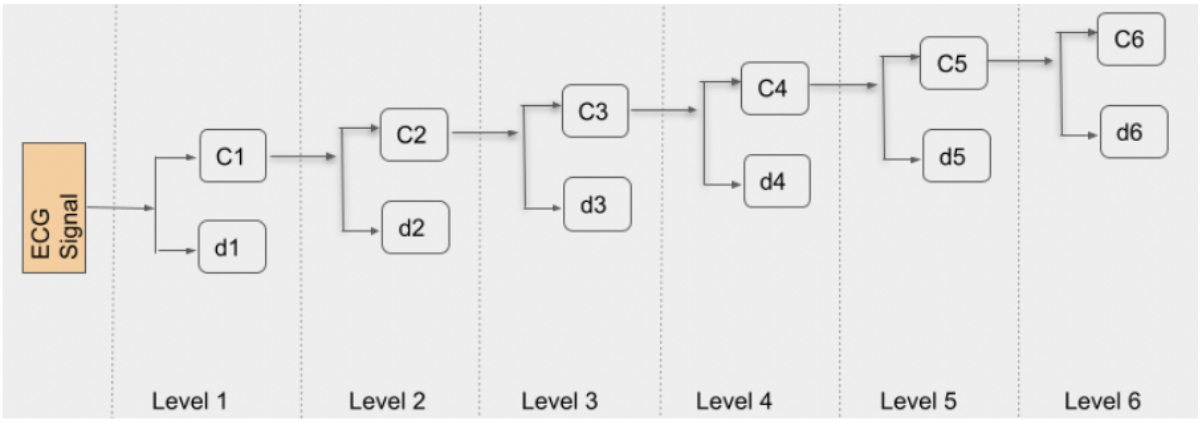

Figure 3.5 DWT approximation and details coefficients structure.

As described in [22], the threshold of 0.15 is applied to eliminate the noise captured in detail coefficients. After applying threshold, signal is reconstructed using inverse discrete wavelet transform on coefficients.

$$x_{rec}[t] = idwt(th(DWT(x[t], y_{high}[t], y_{low}[t]), 0.15), "db4") \tag{3.8}$$

here, th stands for threshold parameter, x_{rec} shows the reconstructed and clean signal after applying idwt. This clean signal can be used for the feature extraction.

3.3 Feature Extraction

The initial stage in the feature extraction process is to identify the peaks in the ECG data. We extracted features parallelly from the EMD and DWT processed signals. EMD processed signals are first used to determine the QRS complex. The biosignal processing in Python (biosspy) toolbox [12] is used to identify QRS complex, which is based on the Hamilton peak segmenting approach [23]. To assess variations from QRS complexes, we extracted: time-domain, geometrical methods, and non-linear features. Time-domain measurements are perhaps the easiest to consider. These approaches determine the heart rate at any moment in time or the intervals between successive normal complexes. After identifying the QRS complex, HRV features were calculated using normal beat intervals (NN), defined as the time interval between two consecutive R peaks of the QRS complex. These NN intervals' mean and median are time-domain measures and widely explored for ASD. The time-domain measure also includes the NN intervals higher than 20ms (NN20) and the percentage derived by dividing NN20 by the total number of NN intervals (pNN20), which helps to assess parasympathetic activity. The coefficient of variation of the R-R interval (CVNN) is also a time domain measure, and it is a measure to evaluate parasympathetic function in these infants.

Another way to analyze the NN interval is to convert them into a geometric pattern, and these are called geometrical methods for HRV analysis. We calculated the HRV triangular index (HTI). HTI is a geometrical method that calculates the integral density distribution. It can be defined as

$$\text{HTI} = \frac{\text{total number of NN}}{\text{number of NN intervals in the modal bin}} \tag{3.9}$$

Here, the modal bin depends on the bin's length when the NN interval is converted to a geometric pattern.

We explored non-linear domain measures also, which include Cardiac Sympathetic Index (CSI), Cardiovagal Index (CVI), and Poincaré plot (PP) analysis. CSI and CVI can be calculated from the Lorenz plot. For calculating these, the length of the longitudinal (L) and transverse (T) axes were calculated within each elliptic distribution. Then CVI was calculated by taking log of product of L and T *($log10(L * T)$)* transformation and CSI as a L/T transformation [44]. PP indexes are also used to assess ASD. It is a

measure that shows each R-R interval of a tachogram projected against the preceding R-R interval. The center point in PP shows the average R-R interval length for the tachogram. For quantitative analysis of a PP, an ellipse is fitted at the center, matching with the center point of the markings and comparing points from two longitudinal and transverse lines traveling through the center point of the data. Here, the length of the longitudinal line is called SD2, and the length of the transverse line is defined as SD1 of the PP data in the perpendicular direction. We considered the ratio of SD1 and SD2, which we called SD1SD2 for analyzing HRV.

A very well-known method for time series signals to evaluate the complexity of a system is called sample entropy [39]. It is a non-linear domain, and we hypothesized it can indicate the self-similarity in the time series and helps to classify the HL/LL groups more accurately. Detrended fluctuation analysis (DFA) is another non-linear method for non-stationary signals characterization [35]. It provides information about the correlations of heartbeat intervals. It can help analyze when the infant is getting attention towards the stimulus. A summary of all the features is shown in Table 3.1.

Acronym	**Extracted From**	**Feature Description**
Mean	EMD	Mean of the NN .
Median	EMD	Median of the NN.
pNN20	EMD	Stands for the proportion of successive NN with a difference larger than 20 ms .
HTI	EMD	HRV Triangular Index is the integral of the NN density histrogram divided by its height .
SD1SD2	EMD	Stands for SD1/SD2. It describes the ratio of short to long-term variations in NN fluctuations and reflects sympathovagal balance .
CSI	EMD	The Cardiac Sympathetic Index (CSI) quantifies the sympathetic nervous system activity indicating increased arousal
CVI	EMD	The Cardiovagal Index (CVI) quantifies the parasympathetic nervous system activity.
CVNN	EMD	Coefficient of variation of NN, calculated as standard deviation of NN divided by mean of NN.
SampEn	DWT	Sample entropy is a statistical method used to assess the complexity of physiological time-series.
DFA	DWT	Detrended Fluctuation Analysis is used to analyse the scaling behavior of non-stationary time-series.

Table 3.1 Features Extracted from ECG

Chapter *4*

Proposed Work

4.1 Pipeline

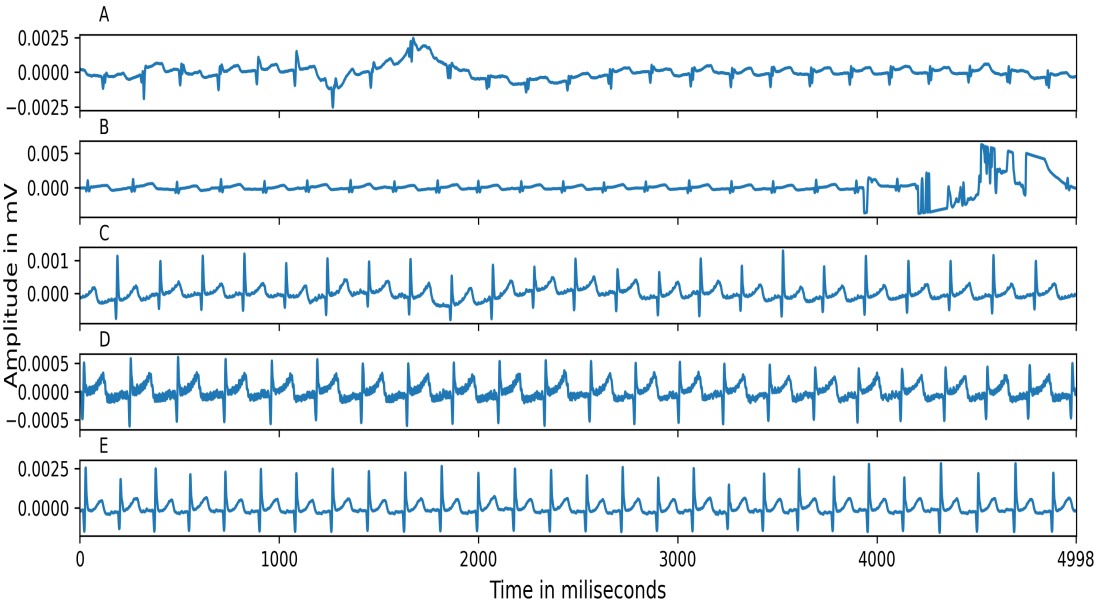

Figure 4.1 Noisy and cleaned QRS complexes

(A-B) Type of segments which we excluded from the dataset. (C) Example of baseline wandering. (D) Example of power line interference.(E) Clean signal, after preprocessing.

Our approach focuses on raw ECG to predict the familial outcome for ASD as HL vs. LL summarized in Figure 4.2. It involves the following steps: synchronization of the edf and video time logs, extraction of the OIX/PIX experiments, segmenting the experiments into non-overlapping 30 seconds slices, parallel preprocessing of each slice using EMD and DWT to eliminate artifacts, feature extraction parallel from EMD and DWT output, and classification of signals using multi-layer perceptron. These signals are collected at two different frequencies, i.e., 512 and 1024 Hz. We down-sampled the frequency signal to 512 Hz to ensure uniform sampling across each slice. After downsampling, the

Hamilton technique uses slope, amplitude, and width information to reliably detect QRS complexes. The QRS complex of the ECG helps us to calculate the features which we can use further for classification. Slices which was not able to detect the QRS complex are rejected (e.g., 4.1.A-B) were rejected. A total of 1483 slices of 30 seconds were considered after rejecting 123 (7.65%) slices from the total. 4.1.C-D show baseline wandering and power line interference. These two types of artifacts were corrected so that these segments could be used for our analyses. 4.1.E gives an example of a clean signal following preprocessing. Each slice of 30 seconds has a valid number of QRS complexes to calculate features from parallel preprocessing defined in Section 3.3. We calculated the mean across each extracted feature for allowing subject-wise classification. We used the leave one out cross-validation [45] technique to ensure that training and testing have both classes and it is adequately evaluated.

4.2 Method : Multilayer Perceptron

The Multilayer Perceptron (MLP) is a feed-forward neural network with three layers: input, hidden, and output. The input layer receives the extracted feature vector to be processed. The output layer performs the required task to classify HL vs. LL. The model's design consists of the hidden layers between the input and output layers, which are considered the computational engine of the MLP. Data flows from the input to the output layer, and the neurons in the model are trained with the backpropagation learning algorithm. Our model uses the MLP implementation from Sci-kit-Learn [1] and a maximum of 1000 iterations for convergence. We used an input layer with ten neurons (i.e., one per feature listed in 3.1), three hidden layers with 400, 200, and 100 neurons, respectively, and an output layer with two neurons associated with labels EL or LL, as shown in Figure 4.4. Subsequent layers are fully connected. Neuron weights and bias parameters are adjusted during training using classic backpropagation. The structure of a neuron is shown in Figure 4.3.

The neuron output is filtered through a logistic function:

$$f(x) = 1/(1 + exp(-x)) \tag{4.1}$$

The output of the neuron j at the layer l, before being filtered by $f(x)$, is given by h_j^l and it computes the biased and weighted sum of the inputs x_i^l:

$$h_j^l = \sum_{i=1}^{I} w_{ji}^l x_i^l + b_j^l \tag{4.2}$$

where w_{ji}^l are the weights and b_j^l is the bias. The output of the neuron is passed to the logistic activation function to produce the output y_j^l for this perceptron, which will also becomes the input for the units of the next layer:

$$y_j^l = f(h_j^l) = x_j^{l+1} \tag{4.3}$$

[1] https://scikit-learn.org/

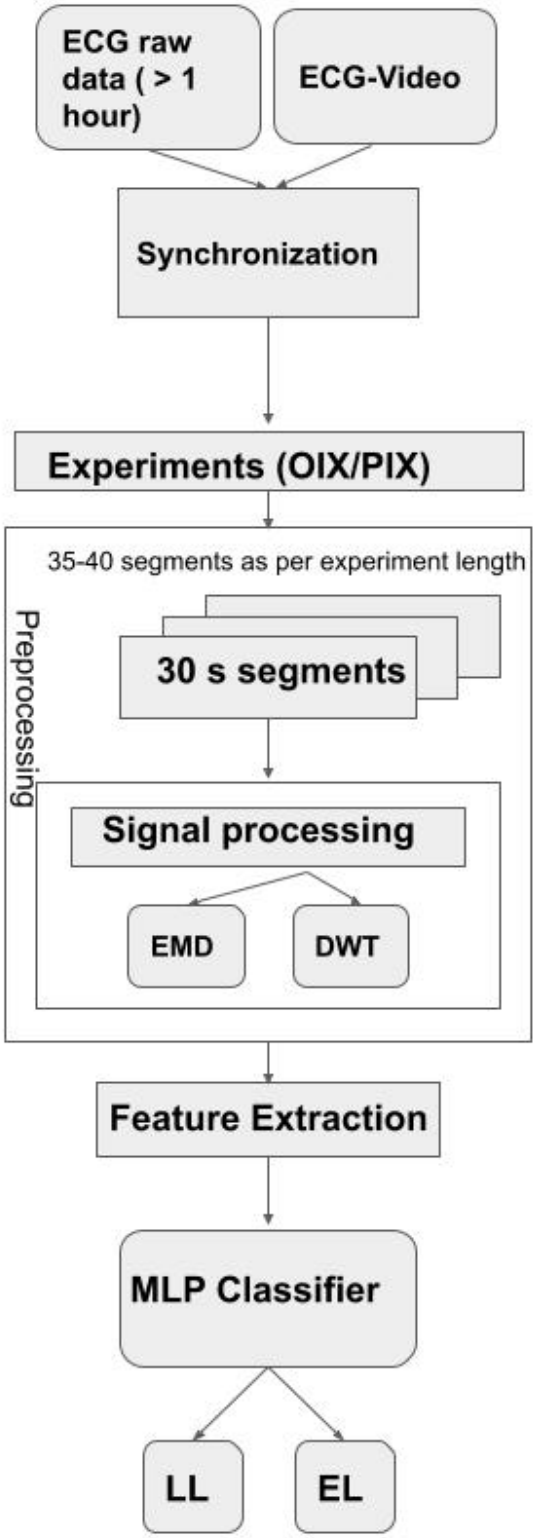

Figure 4.2 Proposed method for EL vs LL classification.

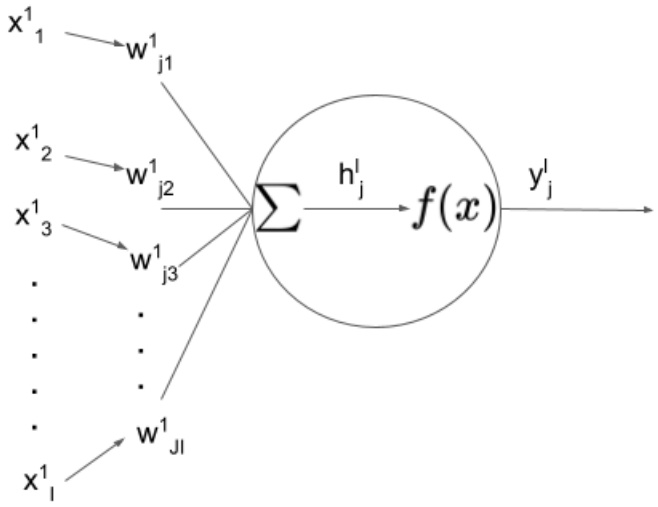

Figure 4.3 Structure of a Neuron

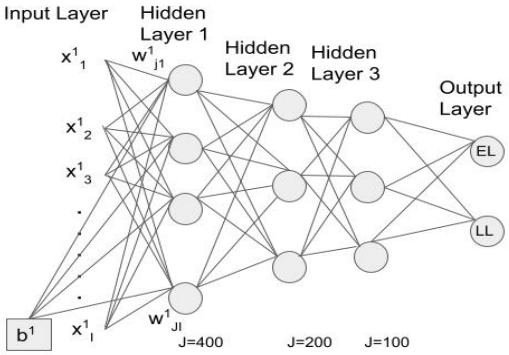

Figure 4.4 Multilayer Perceptron Model

Weights are omitted for layers > 1 only to avoid crowding the schema.

Chapter *5*

Results

5.1 Results

Feature	LL Group	HL Group	t-test	Effect Size
Mean	453.5 $\pm$ 77.96	494.0 $\pm$ 147.1	(t=0.870, p=0.393)	0.381
Median	420.8 $\pm$ 44.19	437.9 $\pm$ 58.2	(t=0.789, p=0.438)	0.346
CVNN	0.233 $\pm$ 0.148	0.286 $\pm$ 0.227	(t=0.676, p=0.506)	0.296
pNN20	29.41 $\pm$ 20.96	41.14 $\pm$ 19.27	(t=1.312, p=0.203)	0.575
HTI	5.928 $\pm$2.822	6.279 $\pm$ 2.092	(t=0.308, p=0.760)	0.135
SD1SD2	0.865 $\pm$ 0.143	0.971 $\pm$ 0.233	(t=1.351, p=0.190)	0.592
CSI	1.395 $\pm$ 0.305	1.262 $\pm$ 0.468	(t=0.826, p=0.417)	0.362
CVI	4.831 $\pm$ 0.650	5.013 $\pm$ 0.769	(t=0.598, p=0.555)	0.262
SampEn	0.147 $\pm$ 0.024	0.119 $\pm$ 0.029	(t=2.389, p=0.0263)	1.046
DFA	1.215 $\pm$ 0.095	1.226 $\pm$ 0.168	(t=0.187, p=0.853)	0.082

Table 5.1 Features are represented as group mean $\pm$ std. Reported effect sizes are for Cohen's d, which are generally understood as small, medium, and large effects, respectively, for values equal to 0.2, 0.5, and 0.8.

All infants tolerated the ECG sensors well and data loss was minimal. Features extracted for every 30-second segment are averaged within recordings and then across recording sessions for subjects with repeated measurements, producing a matrix of 10 features per 23 subjects. Mean and standard deviation (std) were calculated for all features at a group level and presented as mean $\pm$ std in 5.1. The HL group shows higher values for extracted HRV features mean, median, pNN20, HTI, SD1SD2, CVNN than the LL group. Also, CVI (parasympathetic index) and DFA (extracted from DWT) of ECG signals were found higher in the HL group. We used an MLP classifier for classifying extracted features for HL/LL groups.

Metrics used to assess these results are accuracy, precision, sensitivity, specificity, and F1, with HL and LL representing the positive and negative cases, respectively. Sensitivity is defined as the true positive rate as shown in Equation 5.2, while specificity is defined as the true negative rate shown in Equation 5.5. Precision is defined as one minus the false positive rate shown in Equation 5.3 and accuracy is the proportion of correctly predicted subjects shown in Equation 5.1. F1 score provides a harmonic mean score to capture both precision and sensitivity shown in Equation 5.4.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{5.2}$$

$$Precision = \frac{TP}{TP + FP} \tag{5.3}$$

$$F1 = \frac{2 * Precision * Sensitivity}{Precision + Sensitivity} \tag{5.4}$$

$$Specificity = \frac{TN}{FP + TN} \tag{5.5}$$

Here TP, TN, FN, and FP represent true positives, true negatives, false negatives, and false positives.

Performances achieved for subject classification are listed in **??**. For comparison, we also report the results from another study using an ensemble classifier [3]. Since the authors only reported their accuracy, only a partial comparison is possible. To our knowledge, this is the only other study reporting classification results using raw ECG with machine learning for diagnostic applications in ASD. Note, however, that whereas these authors classified ASD vs neurotypical children, we classified HL vs LL infants.

Classification Algorithm	Accuracy	Precision	Sensitivity	F1 Score	Specificity
XGBoost	0.59 (± 0.22)	0.57 (± 0.22)	0.59 (± 0.14)		
Extra trees	0.56 (± 0.15)	0.53 (± 0.17)	0.55 (± 0.19)		
K nearest neighbors (k = 5)	0.47 (± 0.22)	0.49 (± 0.21)	0.50 (± 0.25)		
Random forest	0.50 (± 0.08)	0.48 (± 0.14)	0.50 (± 0.15)		
Artificial neural network	0.44 (± 0.33)	0.40 (± 0.39)	0.47 (± 0.38)		
Logistic regression	0.54 (± 0.14)	0.53 (± 0.20)	0.55 (± 0.19)	-	-
SVM	0.47 (± 0.09)	0.44 (± 0.09)	0.49 (± 0.08)		
SGD	0.44 (± 0.26)	0.47 (± 0.36)	0.48 (± 0.28)		
LASSO-LARS	0.47 (± 0.26)	0.44 (± 0.35)	0.47 (± 0.30)		
Gradient boosted trees	0.50 (± 0.19)	0.48 (± 0.27)	0.49 (± 0.26)		
Decision tree [21]	0.52 (± 0.33)	0.53 (± 0.3	0.54 (± 0.35)		
Ensemble Model [3]	75.3%	-	-	-	-
Proposed Method (MLP)	0.82(± 0.39)	0.75 (± 0.30)	0.75 (± 0.30)	0.75 (± 0.30)	0.86 (± 0.49)

5.2 Discussions

The primary goal of this study was to classify infants in terms of their familial likelihood for ASD (HL vs LL). HL infants show elevated rates of ASD diagnosis and elevated ASD symptomology, making this preliminary study a first step in ASD biomarker development. Our approach has high ecological validity considering the naturalistic, experimental context (interaction with objects and parents) and the wearable, wireless and non-invasive sensors. Such an approach can be easily scaled up and reach more participants by collecting data at home or remotely (i.e., without an experimenter on site). Considering the difficulty of classifying HL vs LL infants based only on ECG, our results using an MLP classifier showed high classification accuracy (82.6%), sensitivity (75.0%), specificity (86.7%), F1-score (75.0%), and precision (75.0%). These results demonstrate promise for using wearable, wireless, and user-friendly ECG devices in health settings to assess ANS activation and screen for ASD likelihood. This is in contrast to EEG devices, which generally rely on multiple electrodes that are more sensitive to artifacts, making data collection with infants more difficult. Other neuroimaging techniques like functional magnetic resonance imaging are even less practical due to higher costs and lower accessibility. While ECG signal disruption can occur, our denoising and preprocessing methods have corrected artifacts without significant data loss.

Research suggests a link between lower HRV and atypical brain functioning, and our results are consistent with multiple studies showing atypical HRV in ASD [14]. Here, we hypothesized that ASD-specific differences in autonomic regulation would result in between-group differences in HRV. To evaluate this relationship, we calculated CSI (sympathetic index) and CVI (parasympathetic index). These variables have previously been associated with a variety of psychological and physical health outcomes, including social behavior, cognition, emotional functioning, and physical health [26, 19]. Although these features are not statistically significant in our small sample size as shown in Table 5.1, CVI and

CSI values were higher and lower, respectively, in HL compared to LL infants. These findings suggest that HL infants demonstrate lower-paced breathing and a lower heart rate. In addition, HL infants demonstrated increased mean, median, pNN20, HTI, and CVNN compared to LL infants (5.1). While these trends are consistent with previous studies [8, 16], the lack of statistical significance and small effect size warrant additional research with larger sample sizes.

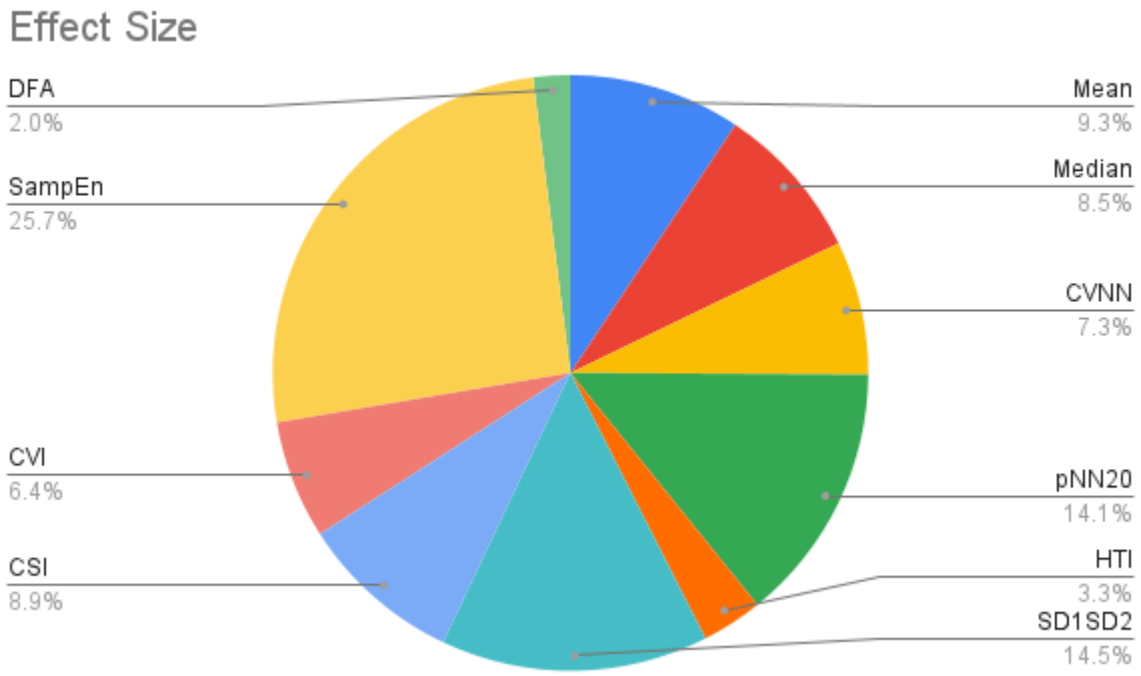

Figure 5.1 Effect size for extracted features.

Effect sizes for the features evaluated in this study ranged from small to large. A pie chart for effect size for seeing how significant each feature is which helps in further classifying HL vs LL, is shown in Figure 5.1, with two features demonstrating a medium effect (effect size > 0.5; pNN20, SD1SD2) and one feature demonstrating a large effect (effect size > 1.0; SampEn). The SampEn feature was the only one to reach statistical significance. Given the small sample size of this study, statistical significance would be expected only for features with large effect sizes. Thus, although not statistically significant with the current sample size, many of the features showing small to medium effect sizes may still contribute helpful information to our classifier (i.e., a lack of statistical significance is not proof of an absence of relationship). We further note that the sample entropy (the only the statistically significant feature) is lower in the HL group. This is similar to prior studies of reduced heart variability in people with ASD [43], given that entropy is a measure of disorder or randomness, a concept very close to variability. Finally, we note that DFA, which captures short-term fluctuations, showed only a very

small effect size and may not be very sensitive in classifying ASD likelihood. In contrast, the SD1SD2 features, which reflect the sympathovagal balance [25], showed the second-largest effect size (0.592) and should, therefore, be considered for future study.

Comparing the accuracy of our system with previously reported results on a very similar problem [3] (i.e., ASD diagnosis vs ASD likelihood classification) and [21] (i.e ASD, typical development (TD), depression and conduct problems (CP), a multi-class classification approach with HRV features) suggests good performance of our approach. Unfortunately, in [3] previous authors did not report specificity, sensitivity, precision, or F1 score, which would have allowed for a more meaningful comparison. But in [21], authors report precision and sensitivity, which suggests that our approach is outperforming that study. They reported highest accuracy 0.59 (± 0.22) with XGBoost classifier, which also shows that proposed method is significant. Moreover, accuracy is known to be biased in the case of unbalanced dataset (i.e., high accuracy can be obtained on an unbalanced dataset even if the classifier performs very poorly for the smaller class) and is generally insufficient to assess the performance of a classifier properly. Therefore, we supplemented accuracy with precision, sensitivity, specificity, and F1 score to allow future studies to compare their results against our own.

While interpreting the results of this study, we have some limitations. First, our study only looked at HRV parameters in an unbalanced dataset with a small sample size. Therefore, the conclusions of this study should be corroborated in the future by a larger sample. In addition, this study only included infants with an elevated familial likelihood for ASD rather than infants who were later confirmed to have a diagnosis of ASD. We were not able to reproduce the results from [3] and [21] because of unavailability of data.

Chapter *6*

Conclusions and Future Work

Analysis of ECG and heart rate variability have become an important avenue of biomarker research [36]. Because it is a non-invasive method and non-complex implementation. In addition, HRV is a useful measure for studying sympathetic and parasympathetic functions, particularly in infancy. This study demonstrated how HRV, sample entropy, DFA, CSI, and CVI extracted from ECG can predict the familial likelihood of ASD in 3-6 month infants. In the future, we plan to build upon this preliminary study to explore further HRV and sample entropy as potential biomarkers of ASD in infancy. Plans are to replicate these results in a larger sample of infants and children with confirmed ASD diagnoses and non-ASD developmental disorders (e.g., hyperactivity disorder, language disorder, attention deficiency). We will develop a system to help parents monitor the infant anytime using ECG and have an easy GUI and visualization to track performance.

Bibliography

[1] P. S. Addison. Wavelet transforms and the : a review. *Physiological measurement*, 26 5:R155–99, 2005.

[2] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR*. Autor, Washington, DC, 4th ed., text rev. edition, 2000.

[3] B. Anandhi, J. Selvaraj, I. Anusuya, and H. Das. Time domain analysis of heart rate variability signals in valence recognition for children with autism spectrum disorder (asd). *IRBM*, 06 2021.

[4] E. Bal, E. Harden, D. Lamb, A. V. Van Hecke, J. W. Denver, and S. W. Porges. Emotion recognition in children with autism spectrum disorders: Relations to eye gaze and autonomic state. *Journal of autism and developmental disorders*, 40(3):358–370, 2010.

[5] S. Baron-Cohen, S. Wheelwright, R. Skinner, J. Martin, and E. Clubley. The autism-spectrum quotient (aq): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31:5–17, 02 2001.

[6] S. Berument, M. Rutter, C. Lord, A. Pickles, and A. Bailey. Autism screening questionnaire: Diagnostic validity. *The British journal of psychiatry : the journal of mental science*, 175:444–51, 12 1999.

[7] L. Billeci, A. Tonacci, A. Narzisi, Z. Manigrasso, M. Varanini, F. Fulceri, C. Lattarulo, S. Calderoni, and F. Muratori. Heart rate variability during a joint attention task in toddlers with autism spectrum disorders. *Frontiers in Physiology*, 9, 2018.

[8] L. Billeci, A. Tonacci, A. Narzisi, Z. Manigrasso, M. Varanini, F. Fulceri, C. Lattarulo, S. Calderoni, and F. Muratori. Heart rate variability during a joint attention task in toddlers with autism spectrum disorders. *Frontiers in Physiology*, 9:467, 05 2018.

[9] M. Blanco-Velasco, B. Weng, and K. Barner. ECG signal denoising and baseline wander correction based on the empirical mode decomposition. *Computers in biology and medicine*, 38:1–13, 02 2008.

[10] G. Brihadiswaran, D. Haputhanthri, S. Gunathilaka, D. Meedeniya, and S. Jayarathna. EEG-based processing and classification methodologies for autism spectrum disorder: A review. *Journal of Computer Science*, 15:1161.1183, 09 2019.

[11] E. by Edmund Sonuga-Barke. Journal of child psychology and psychiatry. *Journal of Child Psychology and Psychiatry*, 35(5):vii–viii, 1994.

[12] C. Carreiras, A. P. Alves, A. Lourenço, F. Canento, H. Silva, A. Fred, et al. *BioSPPy: Biosignal Processing in Python*. 2015.

[13] B. Chakrabarti, F. Dudbridge, L. Kent, S. Wheelwright, G. Hill-Cawthorne, C. Allison, S. Banerjee-Basu, and S. Baron-Cohen. Genes related to sex steroids, neural growth, and social-emotional behavior are associated with autistic traits, empathy, and asperger syndrome. *Autism Research*, 2, 06 2009.

[14] Y.-C. Cheng, Y. C. Huang, and W.-L. Huang. Heart rate variability in individuals with autism spectrum disorders: a meta-analysis. *Neuroscience and Biobehavioral Reviews*, 118, 08 2020.

[15] A. Couteur, G. Haden, D. Hammal, and H. McConachie. Diagnosing autism spectrum disorders in preschool children using two standardised assessment instruments: The ADI-R and the ADOS. *Journal of autism and developmental disorders*, 38:362–72, 03 2008.

[16] C. Daluwatte, J. Miles, S. Christ, D. Beversdorf, T. Takahashi, and G. Yao. Atypical pupillary light reflex and heart rate variability in children with autism spectrum disorder. *Journal of autism and developmental disorders*, 43, 12 2012.

[17] J. Denver, S. Reed, and S. Porges. Methodological issues in the quantification of respiratory sinus arrhythmia. *Biological psychology*, 74:286–94, 03 2007.

[18] M. Di Renzo, V. Guerriero, A. Pagnacco, M. Petrillo, L. Racinaro, S. D'Errico, and F. Bianchi di Castelbianco. Attunement and paternal characteristics in care relationships in the presence of children diagnosed with autism. *International Journal of Environmental Research and Public Health*, 18(4), 2021.

[19] N. Dodo and R. Hashimoto. Autonomic nervous system activity during a speech task. *Frontiers in Neuroscience*, 13, 04 2019.

[20] R. Emerson, C. Adams, T. Nishino, H. Cody, J. Wolff, L. Zwaigenbaum, J. Constantino, M. Shen, M. Swanson, J. Elison, S. Kandala, A. Estes, K. Botteron, L. Collins, S. Dager, A. Evans, G. Gerig, H. Gu, R. Mckinstry, and J. Piven. Functional neuroimaging of high-risk 6-month-old infants predicts a diagnosis of autism at 24 months of age. *Science Translational Medicine*, 9, 06 2017.

[21] M. Frasch, C. Shen, H.-T. Wu, A. Müller, E. Neuhaus, R. Bernier, D. Kamara, and T. Beauchaine. Can a composite heart rate variability biomarker shed new insights about autism spectrum disorder in school-age children? *Journal of Autism and Developmental Disorders*, 51, 05 2020.

[22] R. Haddadi, E. Abdelmounim, M. Hanine, and B. Abdelaziz. Discrete wavelet transform based algorithm for recognition of qrs complexes. *International Conference on Multimedia Computing and Systems -Proceedings*, 0:375–379, 04 2014.

[23] P. Hamilton. Open source ECG analysis. In *Computers in Cardiology*, pages 101–104, 2002.

[24] S. Holiga, J. Hipp, C. Chatham, P. Garces, W. Spooren, X. Liogier d'Ardhuy, A. Bertolino, C. Bouquet, J. Buitelaar, C. Bours, A. Rausch, M. Oldehinkel, M. Bouvard, A. Anouck, M. Caralp, S. Gueguen, M. Ly, J. Houenou, C. Beckmann, and J. Dukart. Patients with autism spectrum disorders display reproducible functional connectivity alterations. 04 2018.

[25] P. Kamen, H. Krum, and A. M. Tonkin. Poincaré plot of heart rate variability allows quantitative display of parasympathetic nervous activity in humans. *Clinical science*, 91 2:201–8, 1996.

[26] S. Laborde, E. Mosley, and J. F. Thayer. Heart rate variability and cardiac vagal tone in psychophysiological research – recommendations for experiment planning, data analysis, and data reporting. *Frontiers in Psychology*, 8:213, 2017.

[27] A. Levy and A. Perry. Outcomes in adolescents and adults with autism: A review of the literature. research in autism spectrum disorders, 5(4), 1271-1282. *Research in Autism Spectrum Disorders - RES AUTISM SPECTR DISORD*, 5:1271–1282, 10 2011.

[28] C. Lord, S. Risi, L. Lambrecht, E. H. Cook, B. L. Leventhal, P. C. DiLavore, A. Pickles, and M. L. Rutter. The autism diagnostic observation schedule—generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30:205–223, 2000.

[29] C. Lord, M. Rutter, and A. Couteur. Autism diagnostic interview-revised - a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of autism and developmental disorders*, 24:659–85, 11 1994.

[30] M. Maenner, K. Shaw, J. Baio, A. Washington, M. Patrick, M. DiRienzo, D. Christensen, L. Wiggins, S. Pettygrove, J. Andrews, M. Lopez, A. Hudson, T. Baroud, Y. Schwenk, T. White, C. Rosenberg, L.-C. Lee, R. Harrington, M. Huston, and P. Dietz. Prevalence of autism spectrum disorder among children aged 8 years — autism and developmental disabilities monitoring network, 11 sites, united states, 2016. *MMWR. Surveillance Summaries*, 69:1–12, 03 2020.

[31] M. Maenner, K. Shaw, A. Bakian, D. Bilder, M. Durkin, A. Esler, S. Furnier, L. Hallas, J. Hall-Lande, A. Hudson, M. Hughes, M. Patrick, K. Pierce, J. Poynter, A. Salinas, J. Shenouda, A. Vehorn, Z. Warren, J. Constantino, M. DiRienzo, R. Fitzgerald, A. Grzybowski, M. Spivey, S. Pettygrove, W. Zahorodny, A. Ali, J. Andrews, T. Baroud, J. Gutierrez, A. Hewitt, L. Lee, M. Lopez, K. Mancilla, D. McArthur, Y. Schwenk, A. Washington, S. Williams, and M. Cogswell. Prevalence and characteristics of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, united states, 2018. *MMWR Surveillance Summaries*, 70(11):1–16, Dec. 2021.

[32] I. Oosterling, N. Rommelse, M. de Jonge, R. J. van der Gaag, S. Swinkels, S. Roos, J. Visser, and J. Buitelaar. How useful is the social communication questionnaire in toddlers at risk of autism spectrum disorder? *Journal of Child Psychology and Psychiatry*, 51(11):1260–1268, 2010.

[33] K. Palanisamy, M. M, and S. Yaacob. ECG signal denoising using wavelet thresholding techniques in human stress assessment. *International Journal on Electrical Engineering and Informatics*, 4, 06 2012.

[34] M. Penner, E. Anagnostou, and W. Ungar. Practice patterns and determinants of wait time for autism spectrum disorder diagnosis in canada. *Molecular Autism*, 9, 03 2018.

[35] A. A. Pranata, G. W. Adhane, and D. S. Kim. Detrended fluctuation analysis on ECG device for home environment. *IEEE Annual Consumer Communications and Networking Conference (CCNC)*, pages 126–130, 2017.

[36] D. Quintana, G. Alvares, and J. Heathers. Guidelines for reporting articles on psychiatry and heart rate variability (graph): recommendations to advance research communication. *Translational Psychiatry*, 6, 05 2016.

[37] S. Raj and S. Masood. Analysis and detection of autism spectrum disorder using machine learning techniques. *Procedia Computer Science*, 167:994–1004, 01 2020.

[38] M. M. Rashid, N. bin Sulaiman, A. P. P. A. Majeed, R. M. Musa, A. F. A. Nasir, B. S. Bari, and S. Khatun. Current status, challenges, and possible solutions of EEG-based brain-computer interface: A comprehensive review. *Frontiers in Neurorobotics*, 14, 2020.

[39] J. Richman and J. Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *American journal of physiology. Heart and circulatory physiology*, 278:H2039–49, 07 2000.

[40] D. Rotholz, A. Kinsman, K. Lacy, and J. Charles. Improving early identification and intervention for children at risk for autism spectrum disorder. *Pediatrics*, 139, 01 2017.

[41] A. Russell, C. Murphy, C. E. Wilson, N. Gillan, C. Brown, D. Robertson, M. Craig, Q. Deeley, J. Zinkstok, K. Johnston, G. McAlonan, D. Spain, and D. Murphy. The mental health of individuals referred for assessment of autism spectrum disorder in adulthood: A clinic report. *Autism : the international journal of research and practice*, 20, 10 2015.

[42] A. Swanson, Z. Warren, W. Stone, A. Vehorn, E. Dohrmann, and Q. Humberd. The diagnosis of autism in community pediatric settings: Does advanced training facilitate practice change? *Autism : the international journal of research and practice*, 18, 07 2013.

[43] R. Thapa, G. Alvares, T. Zaidi, E. Thomas, I. Hickie, S. Park, and A. Guastella. Reduced heart rate variability in adults with autism spectrum disorder. *Autism Research*, 12, 04 2019.

[44] M. Toichi, T. Sugiura, T. Murai, and A. Sengoku. A new method of assessing cardiac autonomic function and its comparison with spectral analysis and coefficient of variation of R-R interval. *Journal of the autonomic nervous system*, 62 1-2:79–84, 1997.

[45] G. Webb, C. Sammut, C. Perlich, T. Horváth, S. Wrobel, K. Korb, W. Noble, C. Leslie, M. Lagoudakis, N. Quadrianto, W. Buntine, L. Getoor, G. Namata, J. Jin, J.-A. Ting, S. Vijayakumar, S. Schaal, and L. De Raedt. *Leave-One-Out Cross-Validation*. 01 2010.

[46] L. Zwaigenbaum, S. Bryson, T. Rogers, W. Roberts, J. Brian, and P. Szatmari. Behavioral markers of autism in the first year of life. *International journal of developmental neuroscience : the official journal of the International Society for Developmental Neuroscience*, 23:143–52, 04 2005.