**Literature Review**

**Introduction**

Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by deficits in social communication and restricted, repetitive behaviors, with a prevalence of approximately 1 in 36 children in the United States (Maenner et al., 2020). Early diagnosis, ideally before age five, is critical for initiating interventions that improve developmental outcomes (Zwaigenbaum et al., 2015). Socioeconomic disparities, such as differences in household income and parental education, contribute to diagnostic delays, particularly in underserved communities (Durkin et al., 2017). Machine learning offers a promising approach to integrate behavioral and socioeconomic data for predictive modeling, potentially reducing these disparities. This literature review synthesizes recent research from 2012 to 2024 on socioeconomic factors, behavioral data, and machine learning applications in early ASD diagnosis, supporting the capstone project’s aim to develop a predictive model using socioeconomic and behavioral data.

**Socioeconomic Factors in ASD Diagnosis**

Socioeconomic status (SES), encompassing income, parental education, and healthcare access, significantly influences ASD diagnosis timing and prevalence. Durkin et al. (2017) analyzed data from 1.3 million 8-year-olds across multiple U.S. sites, finding that higher SES, measured by area-level income and education, was associated with increased ASD prevalence, likely due to better access to diagnostic services (Durkin et al., 2017). Specifically, ASD prevalence was 17.2 per 1,000 in high-income areas (> $90,000 median household income) compared to 7.1 per 1,000 in low-income areas (≤ $30,000), indicating barriers to early diagnosis for lower-SES families. Similarly, Yan et al. (2023) studied 165 Chinese parents of children with ASD, finding that higher SES, indicated by parental income and education, was linked to greater parental involvement in early interventions, mediated by reduced parenting stress (Yan et al., 2023). These findings highlight the need for predictive tools to prioritize high-risk cases in low-SES populations.

In contrast, Rai et al. (2012) conducted a Swedish population-based study of 4,709 children with ASD and found no significant association between parental education or income and ASD risk, suggesting that universal healthcare systems may mitigate SES-related disparities (Rai et al., 2012). This variability across healthcare systems underscores the importance of context-specific models. For the capstone project, variables such as medication use and informant type (e.g., parent, teacher) can serve as SES proxies to analyze healthcare access and family engagement.

**Behavioral Data and Early ASD Prediction**

Behavioral data, including developmental screenings and parent-reported measures, are essential for early ASD identification. Zwaigenbaum et al. (2015) reviewed screening tools like the Modified Checklist for Autism in Toddlers (M-CHAT), finding that parent-reported behaviors (e.g., limited eye contact, delayed gesturing) at 18–24 months predicted ASD with 70–80% sensitivity (Zwaigenbaum et al., 2015). Measures of social responsiveness and developmental milestones are critical for predictive modeling. Wiggins et al. (2020) analyzed data from 2,210 children and found that those evaluated by 36 months had improved developmental outcomes, emphasizing the value of early behavioral screening (Wiggins et al., 2020).

Informant variability impacts behavioral data accuracy. Yu et al. (2024) examined parent and teacher reports, finding that mothers in low-SES families reported higher ASD symptom severity than other informants, possibly due to cultural or educational biases (Yu et al., 2024). Variables capturing informant identity can help account for these biases, improving predictive model reliability.

**Machine Learning for ASD Prediction**

Machine learning has shown strong potential for early ASD diagnosis using behavioral and socioeconomic data. Abbas et al. (2018) developed a random forest model using behavioral data from children aged 2–10, achieving 88.4% accuracy in ASD classification (Abbas et al., 2018). Key features included diagnostic scores and parent-reported developmental milestones, such as age of first words. Duda et al. (2016) used a support vector machine with parent questionnaire data from 1,500 children, achieving 89% accuracy in distinguishing ASD from ADHD (Duda et al., 2016), supporting the feasibility of high-accuracy models for young children.

Incorporating socioeconomic variables enhances model performance. Delobel-Ayoub et al. (2015) combined SES indicators (e.g., parental education, neighborhood income) with behavioral data in a logistic regression model, achieving 75% accuracy in predicting ASD diagnosis likelihood across 3,200 children (Delobel-Ayoub et al., 2015). SES variables improved predictions by capturing disparities in diagnostic access. For the capstone project, combining demographic variables (e.g., age, sex, medication status) with behavioral measures (e.g., social affect, developmental milestones) can support the goal of achieving at least 80% accuracy.

**Gaps and Opportunities**

Current research reveals gaps in early ASD prediction. Many studies focus on children aged 4–8, limiting applicability to those under five (Abbas et al., 2018; Wiggins et al., 2020). Retrospective developmental data, such as age of symptom onset, can address this gap. Additionally, while SES is a known predictor of diagnostic delays, its integration into machine learning models is limited due to data availability (Delobel-Ayoub et al., 2015). The capstone project can leverage SES proxies like medication status and informant type to fill this gap. Finally, cultural variability in SES effects, as noted by Rai et al. (2012), suggests the need for U.S.-specific models, which the project can explore.

All in all, recent literature emphasizes the role of socioeconomic and behavioral data in early ASD diagnosis. Higher SES is linked to earlier diagnosis due to better healthcare access (Durkin et al., 2017; Yan et al., 2023), while behavioral measures like parent-reported symptoms are reliable predictors (Zwaigenbaum et al., 2015). Machine learning models achieve high accuracy but often underutilize SES data (Abbas et al., 2018; Duda et al., 2016). This capstone project addresses these gaps by developing a predictive model combining socioeconomic and behavioral data, contributing to equitable early screening and intervention research.

**References**

Abbas, H., Garberson, F., Glover, E., & Wall, D. P. (2018). Machine learning approach for early detection of autism by combining questionnaire and home video screening. *Journal of the American Medical Informatics Association: JAMIA*, *25*(8), 1000–1007. <https://doi.org/10.1093/jamia/ocy039>

Delobel-Ayoub, M., Ehlinger, V., Klapouszczak, D., Maffre, T., Raynaud, J. P., Delpierre, C., & Arnaud, C. (2015). Socioeconomic Disparities and Prevalence of Autism Spectrum Disorders and Intellectual Disability. *PloS one*, *10*(11), e0141964. <https://doi.org/10.1371/journal.pone.0141964>

Duda, M., Ma, R., Haber, N., & Wall, D. P. (2016). Use of machine learning for behavioral distinction of autism and ADHD. *Translational psychiatry*, *6*(2), e732. <https://doi.org/10.1038/tp.2015.221>

Durkin, M. S., Maenner, M. J., Baio, J., Christensen, D., Daniels, J., Fitzgerald, R., Imm, P., Lee, L. C., Schieve, L. A., Van Naarden Braun, K., Wingate, M. S., & Yeargin-Allsopp, M. (2017). Autism Spectrum Disorder Among US Children (2002-2010): Socioeconomic, Racial, and Ethnic Disparities. *American journal of public health*, *107*(11), 1818–1826. <https://doi.org/10.2105/AJPH.2017.304032>

Maenner, M. J., Shaw, K. A., Baio, J., EdS1, Washington, A., Patrick, M., DiRienzo, M., Christensen, D. L., Wiggins, L. D., Pettygrove, S., Andrews, J. G., Lopez, M., Hudson, A., Baroud, T., Schwenk, Y., White, T., Rosenberg, C. R., Lee, L. C., Harrington, R. A., Huston, M., … Dietz, P. M. (2020). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. *Morbidity and mortality weekly report. Surveillance summaries (Washington, D.C. : 2002)*, *69*(4), 1–12. <https://doi.org/10.15585/mmwr.ss6904a1>

Rai, D., Lewis, G., Lundberg, M., Araya, R., Svensson, A., Dalman, C., Carpenter, P., & Magnusson, C. (2012). Parental socioeconomic status and risk of offspring autism spectrum disorders in a Swedish population-based study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*(5), 467–476.e6. <https://doi.org/10.1016/j.jaac.2012.02.012>

Wiggins, L. D., Durkin, M., Esler, A., Lee, L. C., Zahorodny, W., Rice, C., Yeargin-Allsopp, M., Dowling, N. F., Hall-Lande, J., Morrier, M. J., Christensen, D., Shenouda, J., & Baio, J. (2020). Disparities in Documented Diagnoses of Autism Spectrum Disorder Based on Demographic, Individual, and Service Factors. *Autism research : official journal of the International Society for Autism Research*, *13*(3), 464–473. <https://doi.org/10.1002/aur.2255>

Yan, T., Hou, Y., & Liang, L. (2023). Family Socioeconomic Status and Parental Involvement in Chinese Parents of Children with Autism Spectrum Disorder: A Moderated Mediation Model. *Healthcare (Basel, Switzerland)*, *11*(9), 1281. <https://doi.org/10.3390/healthcare11091281>

Yu, Y., Ozonoff, S., & Miller, M. (2024). Assessment of Autism Spectrum Disorder. *Assessment*, *31*(1), 24–41. <https://doi.org/10.1177/10731911231173089>

Zwaigenbaum, L., Bauman, M. L., Stone, W. L., Yirmiya, N., Estes, A., Hansen, R. L., McPartland, J. C., Natowicz, M. R., Choueiri, R., Fein, D., Kasari, C., Pierce, K., Buie, T., Carter, A., Davis, P. A., Granpeesheh, D., Mailloux, Z., Newschaffer, C., Robins, D., Roley, S. S., … Wetherby, A. (2015). Early Identification of Autism Spectrum Disorder: Recommendations for Practice and Research. *Pediatrics*, *136 Suppl 1*(Suppl 1), S10–S40. <https://doi.org/10.1542/peds.2014-3667C>