Predicting Adverse Events in Developmental Disabilities Population

James P. McGlothlin[1], Micah Price-Offerman[1], Robbie Beyer[1], George Casey[1] and John P. Barile[2]
[1]*RSM US LLP, Chicago, IL, U.S.A.*
[2]*University of Hawai'i at Mānoa, Havaii, U.S.A.*

Keywords: Artificial Intelligence, Development Disabilities, Population Health, Analytics, Business Intelligence, Predictive Analytics, Machine Learning, Supervised Learning.

Abstract: Individuals with development disabilities can experience a variety of adverse events. We have found that these events are often unreported. In this project, we work with a large government program which assists such individuals. The goal of the project is to use artificial intelligence (AI) and other modern technologies to predict adverse events. This will allow case managers to better avoid adverse events, prepare for them and help the program participants. Our initial results show very good accuracy and precision in identifying risk and predicting participant adverse events.

1 INTRODUCTION

The Developmental Disabilities Division (DDD) supports ~3,600 active participants. DDD is the operating state agency for the Medicaid 1915(c) Home and Community Based Services (HCBS) Waiver for Individuals with Intellectual and Developmental Disabilities. The department's mission is to "Foster partnerships and provide quality person-centered and family-focused services and supports that promote self-determination.". The department's vision is "Individuals with intellectual and developmental disabilities have healthy, safe, meaningful and self-determined lives."

The goal of this product is to determine which participants are most at risk for an adverse event. Our study looks at 10 types of adverse events:

1. Suspected abuse, neglect or financial exploitation
2. Behavior Change
3. Change in Health Requiring Medical Treatment
4. Any Use of Restraints
5. Injury from a Known/Unknown Cause Requiring Medical Treatment
6. Medication Errors and/or Unexpected Reaction to Medications or Treatment
7. Participant's Whereabouts Unknown
8. Use of Prohibited Intervention
9. Use of Seclusion
10. Death

Our research has three goals. The first goal is to identify the majority of adverse events. DDD uses a custom participant records system called Inspire. Anecdotally, it is expected that less than half of adverse events are correctly documented in Inspire. Therefore, we leverage Medicaid claims data to augment the participant records. As the vast majority of the participants have coverage through the government Medicaid program, this method should allow us to identify any hospital visits, clinical visits or medication dispenses for the participants.

The second goal is to predict adverse events before they happen. This is the central exercise of this paper and experiment.

For the purpose of our data and models, the grain is one row per participant per month. So if the participant has been in the program for a year they will have 12 rows. For each participant and month, we predict yes there will be an adverse event or no there will not be. Therefore, while our system does not attempt to predict exactly when an adverse event will occur, it does predict it within the calendar month.

The third and final goal is to utilize these predictions to take prescriptive action and prevent adverse events. This will require making suggestions for interventions and tracking the input of these interventions on participant outcomes.

McGlothlin, J. P., Price-Offerman, M., Beyer, R., Casey, G. and Barile, J. P.
Predicting Adverse Events in Developmental Disabilities Population.
DOI: 10.5220/0013318200003911

In *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2025) - Volume 2: HEALTHINF*, pages 844-849
ISBN: 978-989-758-731-3; ISSN: 2184-4305

2 ARTIFICIAL INTELLIGENCE

The goal of this product is to predict adverse events before they occur. To do this, we leverage machine learning and predictive analytics.

Machine learning (ML) has become a transformative tool in various sectors, and public health is no exception. At its core, machine learning involves algorithms that can learn from data, identify patterns, and make decisions or predictions without explicit programming for each task. In public health, ML is applied to analyze large volumes of data such as electronic health records, genomic information, and social determinants of health. This enables the identification of trends and patterns that may not be immediately obvious to human researchers. With the power of ML, public health systems can improve outcomes through early disease detection, predictive modeling, and more efficient resource allocation (Jordan, 2015) (Bi, 2019).

One significant application of machine learning in public health is disease prediction and prevention. ML algorithms are capable of processing complex datasets to predict the likelihood of diseases based on various risk factors. For example, ML models have been used to predict the onset of chronic diseases like diabetes and cardiovascular conditions (Siontis, 2012) (Collins, 2012). By analyzing factors such as age, lifestyle, genetics, and environmental influences, ML can forecast the potential for disease in individuals or populations, allowing for early interventions. This predictive power is particularly valuable in resource-limited settings where preventive measures can save lives and reduce healthcare costs. Machine learning is making strides is in epidemiology, especially in tracking and controlling infectious diseases (Ghosh, 2024) (Adegoke, 2024). ML algorithms are being used to analyze patterns in disease spread and to create models for forecasting outbreaks. During the COVID-19 pandemic, ML models were widely employed to predict the spread of the virus, assess healthcare system burdens, and identify effective intervention strategies (Van der Schaar, 2021) (Malik, 2021) (Heidari, 2022). These models relied on a combination of epidemiological data, mobility data, and demographic information. In addition, ML has been applied to track the emergence of antibiotic-resistant bacteria, thereby enhancing surveillance efforts and informing public health responses (Brenda, 2024) .

In addition to modeling, ML is enhancing personalized medicine and treatment in public health (Srinivasaiah, 2024). By analyzing vast datasets, machine learning can tailor healthcare interventions to individuals based on their unique characteristics. This is particularly important in managing chronic diseases, where treatment regimens can vary significantly from one person to another. For instance, ML algorithms can help determine the most effective treatment plans for cancer patients by analyzing genetic data and patient responses to previous treatments (Rafique, 2021) (Quazi, 2022).. This precision medicine approach not only improves individual outcomes but also reduces the inefficiencies of one-size-fits-all healthcare strategies.

Despite its potential, machine learning in public health comes with challenges. These include data privacy concerns, ethical issues regarding algorithmic biases, and the need for sufficient training of healthcare professionals in data science. Moreover, the success of ML models in public health is heavily dependent on the quality of the data used for training these models. Inaccurate, incomplete, or biased data can lead to misleading predictions and decisions. As such, there is an ongoing need for collaboration between data scientists, healthcare professionals, and policymakers to ensure that ML applications are designed, tested, and implemented responsibly.

Supervised learning is a type of machine learning where the model is trained on labeled data, meaning each input is paired with the correct output. The goal is to learn a mapping from inputs to outputs so that, when presented with new, unseen data, the model can predict the correct result (Osisanwo, 2017). The process involves using a dataset with known labels to train the algorithm, which then fine-tunes itself by adjusting its internal parameters to minimize errors between predicted and actual outcomes. This form of learning is widely used in tasks such as classification and regression, where the model learns to categorize data or predict continuous values based on historical examples (Kotsiantis, 2006).

In healthcare, supervised learning has shown significant potential in improving diagnostic accuracy, personalized treatment plans, and predicting patient outcomes. For instance, machine learning models can be trained on medical images like MRIs or X-rays, where the labels correspond to specific diagnoses, enabling the algorithm to assist radiologists in detecting diseases such as cancer or tuberculosis with high accuracy (Sharma, 2025). Supervised learning is also used in predicting patient risk factors, such as the likelihood of developing chronic diseases like diabetes or heart disease, based on historical health data, lifestyle choices, and genetic factors (Islam, 2024). This application helps healthcare professionals provide more tailored treatments and preventative measures, thereby improving patient care and reducing overall healthcare costs (Razzak, 2018).

Predictive analytics models use statistical algorithms and machine learning techniques to analyze historical data and predict future outcomes. These models are built upon data mining and pattern recognition principles, helping organizations forecast trends, behaviors, and events. For example, regression analysis is often used in predictive analytics to identify relationships between variables, while classification algorithms like decision trees or random forests help categorize data points into predefined classes. These predictive models are widely used in various sectors, including finance, healthcare, marketing, and supply chain management, offering insights that guide decision-making and improve operational efficiency (Berrar, 2019).

One of the key advantages of predictive analytics is its ability to enhance decision-making by offering actionable insights based on historical data. Machine learning models, particularly deep learning and neural networks, allow for complex, nonlinear relationships within data to be understood, improving the accuracy of predictions over traditional statistical methods. For instance, in the healthcare sector, predictive models can help identify patients at risk of developing chronic conditions, thereby enabling early intervention and personalized care plans (Chung et al., 2018). These capabilities empower businesses to proactively address issues, reduce costs, and increase customer satisfaction by anticipating needs and actions.

However, despite their power, predictive analytics models also come with challenges. The effectiveness of these models is highly dependent on the quality and volume of the data being analyzed. Inaccurate, incomplete, or biased data can lead to incorrect predictions, which could be detrimental in areas like finance or healthcare. Moreover, predictive models can be resource-intensive, requiring significant computational power and expertise to develop and maintain. It is also crucial to continually update the models with new data to ensure they remain relevant and accurate. Addressing these challenges requires robust data governance practices and collaboration between data scientists and domain experts (Aguirre, 2019).

3 IMPLEMENTATION APPROACH

This project is intended to be used in a commercial setting by hospital providers, so that they can comply with the requirements of patient registries with less burden to hospital staff. Therefore, we wanted to only use commercially available and respected software products which have been approved to handle protected health information (PHI) under the United States's HIPAA (Health Insurance Portability and Accountability Act of 1996) (Moore, 2019).

Therefore, we chose to implement our project using software available from Microsoft including Azure, Azure Machine Learning (AML) (Barga, 2015) (Barnes, 2015) and OpenAI.

Azure Machine Learning is a cloud service from Microsoft designed to streamline the machine learning process. It provides a variety of tools for building, training, and deploying machine learning models, catering to data scientists, developers, and organizations. The platform supports integration with popular frameworks and offers features such as automated machine learning (AutoML), model versioning, and deployment in a secure, scalable environment. Notable capabilities include automated hyperparameter tuning, experiment tracking, and seamless integration with Azure's cloud infrastructure for efficient model management. Azure Machine Learning also supports collaborative development through integrated notebooks and offers monitoring and management tools post-deployment. It accommodates both code-based and low-code development, making it accessible for users with different skill levels. This flexibility enables businesses to advance their AI projects while ensuring governance, security, and scalability in production environments. (Barnes, 2015).

The department uses two source systems to manage participant records: Inspire and MedQuest. Inspire is a custom solution built in Microsoft Dynamics. Med-Quest is a custom solution built to manage Medicaid claims.

Our evaluation determined that adverse events were overwhelmingly under-reported in Inspire. Augmenting the participate record with corresponding claims data allows us to identify any adverse event that resulted in a medical claim. Figure 1 shows the data flow diagram for both the data abstraction and the machine learning.

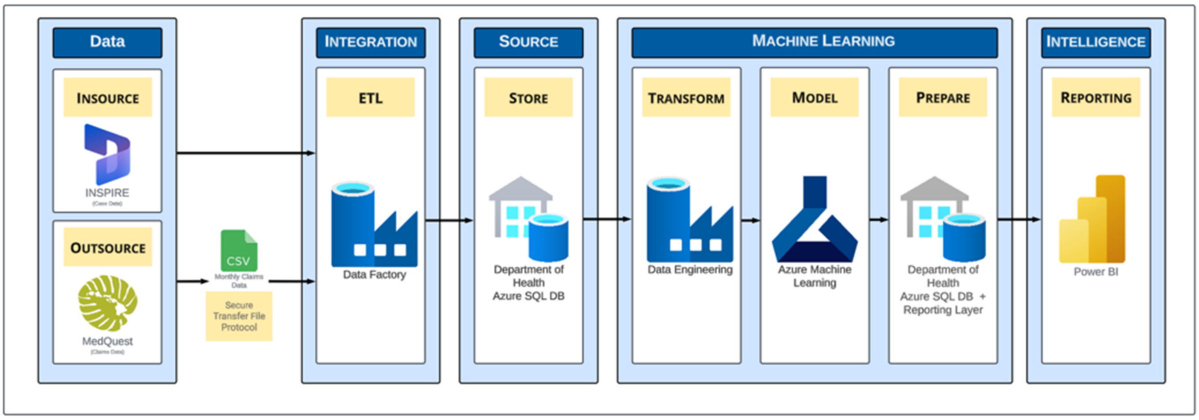

Figure 1: Architectural data flow diagram.

One of the most important factors for successful model is to determine which features are valuable to the prediction and have a true causal effect. Data fields which we evaluated and leveraged include participant age, race, use of medical devices, care setting, level of participation in program, history of trauma, medications and history of schizophrenia,

4 RESULTS

This research is in early stages of development and test.

Our model made 24,146 predictions. The algorithm predicted adverse events in 6,181 instances. In 6,156 instances the adverse event occurred. Therefore, when an adverse event is predicted, there is a 99.6% chance the adverse event will occur. There were only 25 false positives, representing only 0.1% of the preductions. There were 246 false negatives, where the model failed to predict a adverse event. This represents 1.0% percent of predictions. The predictions were correct overall 98.9% of the time.

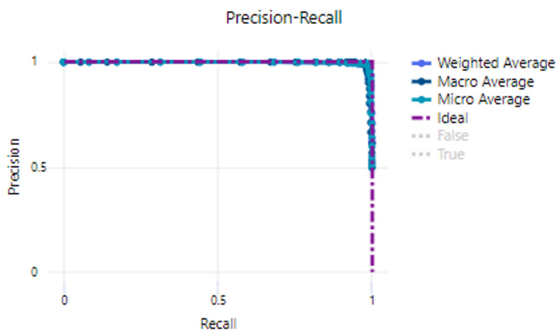

Figure 2: Precision and Recall.

Figure 2 shows the precision and recall for our model. Figure 3 the calibration curve and predicted probability.

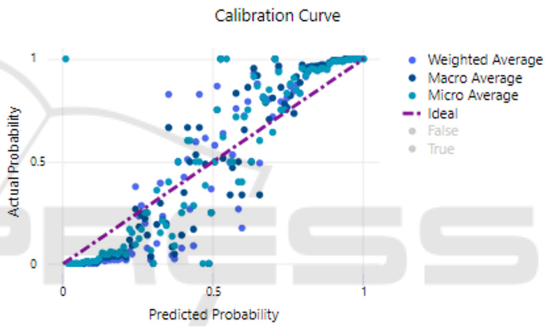

Figure 3: Calibration Curve.

5 NEXT STEPS

Currently, our research predicts the occurrence of any adverse event in a given month. It would be preferable if we could predict exactly which adverse event will occur. For a next step, we will attempt to model each adverse event separately. This will also allow for a distinct set of features per event, instead of forcing a common set of prediction features across all adverse events.

Additionally, we realize that more model tuning, training and rigor is needed to show the statistical significance of our algorithm and process over time.

There is a large amount of free text data collected for the program participants. As a future long term step, we would like to edit the records management software to utilize drop down lists and not allow free text. In order to get value out of the vast amount of text data already collected, we plan to use generative AI to summarize the data.

As stated in our introduction, the third and final goal of the project is to utilize these prediction to take prescriptive action and prevent adverse events. This will require making suggestions for interventions and tracking the input of these interventions on participant outcomes. This is left for future work and has not yet been attempted. To do this properly, we will need to provide integrated data into the case management system enabling our case managers to take appropriate action.

6 CONCLUSIONS

It is vital that both case workers and researchers know when participants have adverse events. By augmenting the participant records with claims data, we were able to almost double the number of known adverse events.

The primary purpose of this initiative is to predict adverse events before they happen. While this is a preliminary evaluation, our early results show an exceptional 98.9% accuracy across all predictions. This shows the promise of AI to help these participants.

REFERENCES

Aguirre, S., Mahr, D., & Rivas, M. (2019). Data-driven predictive analytics in business decision-making. Springer.

Berrar, D. (2019). Predictive modeling applications in business and industry. Wiley-IEEE Press.

Chung, W. Y., Park, J. H., & Kim, H. (2018). Applications of predictive analytics in healthcare. Springer.

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, *9*(1), 381-386.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255-260.

Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is machine learning? A primer for the epidemiologist. *American journal of epidemiology*, *188*(12), 2222-2239.

Panch, T., Szolovits, P., & Atun, R. (2018). Artificial intelligence, machine learning and health systems. *Journal of global health*, *8*(2).

Ravì, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G. Z. (2016). Deep learning for health informatics. *IEEE journal of biomedical and health informatics*, *21*(1), 4-21.

Johnson, A. E., Pollard, T. J., & Mark, R. G. (2017, November). Reproducibility in critical care: a mortality prediction case study. In *Machine learning for healthcare conference* (pp. 361-376). PMLR.

Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. *Annual review of biomedical engineering*, *19*(1), 221-248.

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 738). New York: Springer.

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, *160*(1), 3-24.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.

Moore, W., & Frye, S. (2019). Review of HIPAA, part 1: history, protected health information, and privacy and security rules. *Journal of nuclear medicine technology*, *47*(4), 269-272.

Barga, R., Fontama, V., Tok, W. H., & Cabrera-Cordon, L. (2015). *Predictive analytics with Microsoft Azure machine learning* (pp. 221-241). Berkely, CA: Apress.

Barnes, J. (2015). Azure machine learning. Microsoft Azure Essentials. 1st ed, Microsoft.

Siontis, G. C., Tzoulaki, I., Siontis, K. C., & Ioannidis, J. P. (2012). Comparisons of established risk prediction models for cardiovascular disease: systematic review. *Bmj*, *344*.

Collins, G. S., & Moons, K. G. (2012). Comparing risk prediction models. *Bmj*, *344*.

Ghosh, S., & Sharma, V. (2024). Tracking of Disease—A Review of the State of the Art of Technology for Next Generation Healthcare. *Deep Learning in Internet of Things for Next Generation Healthcare*, 242-268.

Adegoke, B. O., Odugbose, T., & Adeyemi, C. (2024). Data analytics for predicting disease outbreaks: A review of models and tools. *International journal of life science research updates [online]*, *2*(2), 1-9.

Van der Schaar, M., Alaa, A. M., Floto, A., Gimson, A., Scholtes, S., Wood, A., ... & Ercole, A. (2021). How artificial intelligence and machine learning can help healthcare systems respond to COVID-19. *Machine Learning*, *110*, 1-14.

Malik, Y. S., Sircar, S., Bhat, S., Ansari, M. I., Pande, T., Kumar, P., ... & Dhama, K. (2021). How artificial intelligence may help the Covid-19 pandemic: Pitfalls and lessons for the future. *Reviews in medical virology*, *31*(5), 1-11.

Heidari, A., Jafari Navimipour, N., Unal, M., & Toumaj, S. (2022). Machine learning applications for COVID-19 outbreak management. *Neural Computing and Applications*, *34*(18), 15313-15348.

Branda, F., & Scarpa, F. (2024). Implications of Artificial Intelligence in Addressing Antimicrobial Resistance: Innovations, Global Challenges, and Healthcare's Future. *Antibiotics*, *13*(6), 502.

Rodríguez-González, A., Zanin, M., & Menasalvas-Ruiz, E. (2019). Public health and epidemiology informatics: can artificial intelligence help future global challenges? An overview of antimicrobial resistance and impact of

climate change in disease epidemiology. *Yearbook of medical informatics*, *28*(01), 224-231.
Srinivasaiah, B. (2024). The Power of Personalized Healthcare: Harnessing the Potential of Machine Learning in Precision Medicine. *International Journal of Science and Research*, 426-429.
Rafique, R., Islam, S. R., & Kazi, J. U. (2021). Machine learning in the prediction of cancer therapy. *Computational and Structural Biotechnology Journal*, *19*, 4003-4017.
Quazi, S. (2022). Artificial intelligence and machine learning in precision and genomic medicine. *Medical Oncology*, *39*(8), 120.
Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, *48*(3), 128-138.
Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, *26*, 159-190.
Sharma, R., Kumar, N., & Sharma, V. (2025). Computer Vision for Disease Detection—An Overview of How Computer Vision Techniques Can Be Used to Detect Diseases in Medical Images, Such as X-Rays and MRIs. *AI in Disease Detection: Advancements and Applications*, 77-98.
Islam, R., Sultana, A., & Islam, M. R. (2024). A comprehensive review for chronic disease prediction using machine learning algorithms. *Journal of Electrical Systems and Information Technology*, *11*(1), 27.