Development of an Innovative Mouse Model for

Alzheimer’s Disease Prediction

Karthik Rayapati

*Abstract*—In project, goal is to develop a mouse model for Alzheimer’s prediction using machine learning models using mouse experiment data. Since we are using a complex cognitive information-based dataset that has both behavioral and cognitive metrics from mouse models of Alzheimer’s disease, it is expected to predict the progression of disease with better accuracy. In our approaches, we used statistical, neural network based, probabilistic, and baseline models for different kinds of viewpoints in our problem statement. So, this would provide us with different kinds of unique insights from each machine learning model. This would help find a way for early detection or intervention of Alzheimer’s disease diagnosis.

*Index Terms*—Alzheimer’s disease, MRI, machine learning, convolutional neural networks, recurrent neural networks, data preprocessing, Dementia, Classification, Regression, Mouse Mod-

els

# I. INTRODUCTION

A

LZHEIMER’S Disease (AD), is a progressive neurodegenerative disorder seen in many individuals all around the world. And it is hard to be detected early in diagnosis and in treatment as well. But, with advancements and improvements to machine learning (ML) field, possibilities in the detection and analysis of AD have transformed rapidly. Our project showcases a new mouse model using different ML techniques to predict Alzheimer’s disease growth and accuracy. This approach is possible only because of data-driven insights and machine learning, which helps us understand Alzheimer’s disease patterns or insights. This is critical for finding ways to predict and prevent disease in the early stages. Our methodology used different ML models of different approaches so that we have unique perspectives of detection or prediction of AD.

I handled the data preprocessing and analysis of our project. I was responsible for implementing data augmentation and normalization to ensure our mouse model data was standardized for model training. And handled the data visualization aspect of our project.

# III. RELATED WORK

Recent machine learning developments have fueled Alzheimer’s research even more than ever, mainly in the cases of early detection and prediction. Many studies were done on the same topic with the use of neural networks and statistical models, with the goal of analyzing clinical data. These studies were mainly done to demonstrate the potential of AD diagnosis. Our research is based on or inspired by these related works. And we are focusing on an Alzheimer’s mouse model. Although this method is not heavily used or explored, it is impactful in our opinion as we have clinical data to back up our predictions and analysis.

# IV. APPROACH

My approach to using different kinds of machine learning models to analyze the mouse model data is, in our opinion, effective. I built and trained a statistical model for primary analysis. And used a neural network-based model for pattern recognition of the data. A probabilistic model is also used for prediction-based analysis. And since I didn’t find a baseline model for picked dataset, I built a baseline model as well to set benchmark or threshold standards. Model selection was decided based on different evaluation metrics, and this decision was also inspired by other related research works. Our approach had the following things: heavy data preprocessing, proper feature selection, hyper-parameter tuning, and model optimization. This is to make sure that the prediction accuracy is in fact valid and reproducible on test data as well.

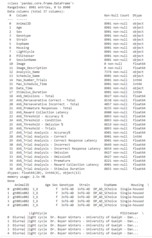


Fig. 1. Dataset - EDA

# V. EXPERIMENTAL RESULTS

## A. Data Preprocessing and Model Training

I started my project by cleaning and handling the data using the well-known data preprocessing techniques of normalization and dimensionality reduction. This enhanced model performance greatly. During the training phase, I adjusted the hyperparameters and also added cross-validation techniques to tune my models and make sure our results weren’t biased.

## B. Model Evaluation and Performance

All models’ performance should be compared at the end to find the best-performing model. The evaluation metrics of any common regression model and neural networks were used for this purpose. ” Mean Squared Error” (MSE),” Mean Absolute Error”(MAE), and the R-squared value were chosen as our selection criterion metrics and are used, showcased. Out of all the machine learning models we trained, the statistical model SVM gave the best results in prediction accuracy. Although the neural network model performed pretty well in identifying patterns in our dataset, the Bayesian regression (probabilistic)



Fig. 2. Testing Data: Predictions

model gave better insights at the end. And as you might expect, the baseline model is only used for comparing the performance of all other models.

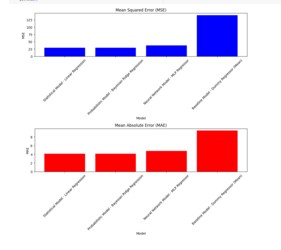


Fig. 3. Evaluation Results.

# VI. CONCLUSION

At the end, my mouse model for Alzheimer’s disease prediction performed well at finding the accuracy of Alzheimer’s disease. Since I used multiple machine learning models, I did not only focus on end metrics but also on Alzheimer’s analytics. I understood that different models would provide unique insights that could help in early diagnosis and intervention techniques for Alzheimer’s. Each model provided a unique view for my project, and at the end, when all of them were put together, they gave me a complete picture of AD and its characteristics. I hope that our work will help future developments in this topic.



Fig. 4. Machine Learning Model Outputs

# VII. REPRODUCIBILITY

To make sure that my work could be easily reproduced, I commented on my code with brief text explaining all the steps I took in the code, including but not limited to data processing, model configuration, model evaluation, and data visualization. My Python code would help all other researchers duplicate or reproduce my project to verify and cross-check my findings. They could also extend my work by adding new things to my project as well.

# VIII. DATASET DESCRIPTION

my project was using a mouse model dataset that had various cognitive behavioral inputs or outputs of Alzheimer’s Mouse Trails. Although the dataset is complex and multidimensional, it needed to be like that, as I was dealing with a unique challenge. My dataset had data from different Alzheimer’s mouse models that were tested in different conditions. So, it is a great resource to test my machine learning models on.

# IX. FUTURE WORK

Looking further, we think that my project could be beneficial for future research and improvements. Future research based on my project could focus on using much more advanced machine learning models run on huge computer machines so that they can provide better accuracy. They could use different ensemble methods and much more complex neural networks to find intricate patterns in the data. Adding to this, using new mouse model data could provide a better understanding of how AD disease has transformed.

# X. ETHICAL ISSUES

As much as it is important to solve this problem, there is always the question of ethics when it comes to clinical trials, particularly animal and human trials. Machine learning in medical research might also pose the question of ethical considerations as to which data to consider. Because, when it comes to human data, topics of privacy and consent and the responsible use of model data are always things to be considered and handled properly. So, techniques like data anonymization and proper data handling for the ethical treatment of animal subjects in research are very much needed.

# XI. LIMITATIONS OF THE STUDY

Although my project presented good findings in the field of Alzheimer’s prediction, it is also important to focus on the limitations as well. I encountered a few issues when analyzing the Mouse Model data, such as the fact that the conditions and places of the trails are predetermined and don’t have many options to choose from. So, this might have introduced a bias to my models unknowingly. Hence, for all future research, it is important to make sure to find all biasing features and make sure they are handled properly before they move on to the modeling phase, which is, in my opinion, needed.

# XII. REFERENCES

1. Miya Bishara Yokoyama, “Mouse Models of Alzheimer’s Disease,” *National Library of Medicine*, Jun. 2021.
2. Scott J. Webster, “Using mice to model Alzheimer’s dementia: an overview of the clinical disease and the preclinical behavioral changes in 10 mouse models,” *National Library of Medicine*, Mar. 2016.