JOURNAL OF
EMERGING INVESTIGATORS

Article

Prediction of preclinical Aβ deposit in Alzheimer's disease mice using EEG and machine learning

Mari Igarashi[1], Heechul Jun[2]
[1]University High School, Irvine, California
[2]Department of Anatomy and Neurobiology, University of California, Irvine, California

SUMMARY

Alzheimer's disease (AD) is a common disease affecting 6 million people in the U.S., but no cure exists. To create therapy for AD, it is critical to detect amyloid-β protein in the brain at the early stage of AD because the accumulation of amyloid-β over 20 years is believed to cause memory impairment. However, it is difficult to examine amyloid-β in patients' brains. The development of a simple method to predict the presence of amyloid-β without harming patients would be helpful for clinical applications. Electroencephalography (EEG) is a testing method that records brain activity using electrodes attached to the scalp without harming the brain. In this study, we hypothesized that we could accurately predict the presence of amyloid-β using EEG data and machine learning. To test this, we analyzed published EEG data recorded from healthy mice and AD mice. Machine classifiers were trained using EEGs of old mice with and without amyloid-β. We then tested if the best trained machine classifier (Gaussian Naïve Bayes (GNB) classifier) could predict amyloid-β in young mice from their EEG data. Results showed that GNB classifier can predict if a given mouse has amyloid-β or not at 81% accuracy when theta waves were used for classifier training. The accuracy of the classifier was at chance level when beta or gamma waves were used. These results indicate that amyloid-β presence can be predicted at high accuracy using EEG and machine learning and suggest that this method may be useful for early diagnosis of AD patients.

INTRODUCTION

Today in the U.S., one in nine people 65 years and older have Alzheimer's disease (AD) (1). However, there is no cure for AD currently available. One of the leading theories for the cause of AD is that amyloid precursor protein creates amyloid-β protein that accumulates in the brain and impairs neuronal functions (2). Another protein, tau, forms neurofibrillary tangles, which eventually kill neurons (3). In AD patients, the deposition of amyloid-β starts around age 60 and gradually disrupts brain function until memory impairment emerges around age 80 (4). One of the major problems in creating a cure or therapy for AD is the fact that it is extremely difficult to diagnose patients in early stages. Brain sections are needed to examine the existence of amyloid-β, but biopsies cannot be performed on brains of living people. Older people who are having amyloid-β accumulation but no memory difficulties do not visit clinics for memory impairments. This period with amyloid-β accumulation but without memory impairments is called the "preclinical stage" of AD (5). At the time when people have memory impairment and visit clinics for their memory problems, patients are in the "clinical stage" of AD. However, neurons and brain circuits are severely damaged already by the clinical stage, and treatment will be difficult because neurons do not regenerate (5). Thus, for addressing the clinical burden of AD, it is critical to diagnose amyloid-β accumulation at the preclinical stage before neuronal death occurs.

Electroencephalography (EEG) (6) is an effective way to record brain activity without damaging brain cells by placing electrodes on the scalp. This test detects electrical charges from brain activities. Using frequencies, the waves of EEG can be categorized into different bands. These bands include theta (5–7 Hz), beta (16–30 Hz), and gamma (41–80 Hz) waves. It was previously reported that old AD mice (10 months of age) show distinct patterns of EEG powers compared to healthy old mice (6). These findings suggest that EEG data could be useful for identifying AD in early stages, if we can predict the existence of amyloid-β from EEG data. For this prediction, we hypothesized if machine learning can provide a high predictive power. Machine learning technology has advanced dramatically over the past few decades, and it has become a common tool in scientific research. Machine learning is an artificial intelligence algorithm that learns through training. It can detect subtle differences, and after training with existing data with known answers, machine learning classifiers can predict answers for incoming data. Many different classifiers are available, such as the support vector machine and the Naïve Bayes classifier, and each has an optimal type of data for usage (7). Machine learning can provide better prediction performance than traditional statistical methods. We thus hypothesized that amyloid-β accumulation in young mice could be predicted using EEG data and machine learning. We tested this hypothesis using published EEG data obtained from healthy mice and AD mice (8). Our results showed that Gaussian Naïve Bayes classifier can predict the existence of amyloid-β at 81% accuracy from EEG data, suggesting that this method may be useful for early diagnosis of AD patients.

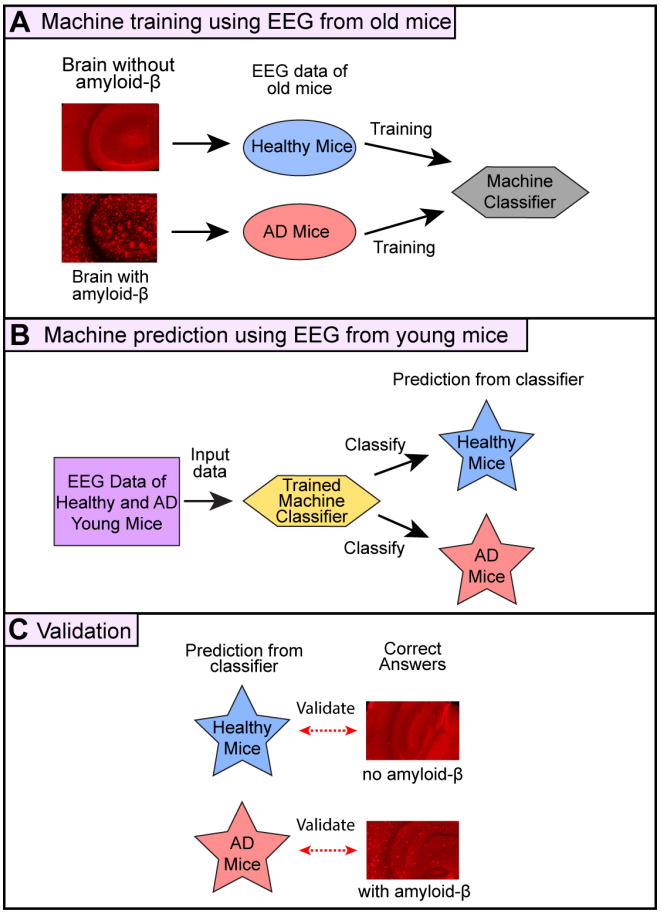

Figure 1: Analytical procedure and workflow used in this study. **A:** Machine classifiers were trained using EEG data from healthy old and AD old mice groups. **B:** EEG data from healthy young and AD young were used as inputs to the classifier for prediction, without telling which group each data was derived from. The classifier returned predicted output (healthy or AD) for each mouse. **C:** The output from the classifier was tested against the truth. The output was validated with the correct answer, which is the existence of amyloid-β in the brain.

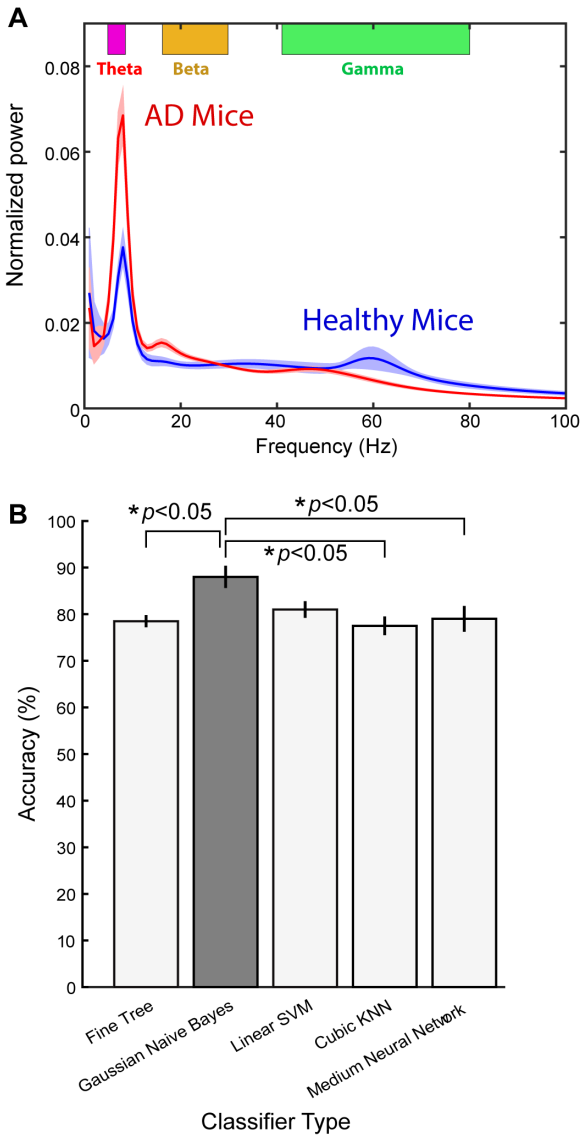

Figure 2: Training of machine classifiers using EEG from old mice. **A:** Power spectrum of EEG data recorded from healthy old mice (blue, N=10 mice) and AD old mice (red, N=10 mice). Raw EEG powers were summed across 1–100 Hz bins, and powers at each bin were divided by the summed power for normalization. Powers are shown in mean (bold lines) ± standard error (transparent lines). **B:** Prediction accuracy of different machine classifiers, cross-validated using EEG of old mice. Gaussian Naïve Bayes classifier provided the highest accuracy. *=p<0.05, ANOVA followed post-hoc Bonferroni test. N=10 cross-validation samples.

RESULTS

Training of machine classifier using EEG from old mice

We used data from a previously published paper was to perform this experiment (8). The dataset contains EEG data obtained from four groups of mice: healthy young mice (N=5 mice), healthy old mice (N=10 mice), AD young mice (N=6 mice), and AD old mice (N=10 mice). We defined young mice as those between 3 and 5 months of age, and old mice as those between 7 and 12 months of age. We used amyloid precursor protein knock-in (APP-KI) mice as AD mice (4). APP-KI are genetically-modified mice having mutated amyloid precursor protein (APP) that produces amyloid-β accumulation in the brain. On the other hand, healthy mice do not have mutated APP and thus do not accumulate amyloid-β. A previous study showed that young APP-KI mice were at the preclinical AD stage where brains have amyloid-β accumulation, but these animals did not exhibit signs of memory impairment (6). Old APP-KI mice are at the clinical AD stage with both amyloid-β accumulation and memory impairment (6).

To leverage the EEG data from both preclinical and clinical stages, we designed our own analytical strategy (**Figure 1**). First, we trained machine classifiers using EEG data from old healthy and old AD mice at the clinical stage (**Figure 1A**). Next, we used these trained machine classifiers to predict amyloid-β deposition from EEG data of young healthy mice and young AD mice at the preclinical stage (**Figure 1B**). Finally, we validated the predictions by comparing them to amyloid-β accumulation results in individual mice that were

examined histologically and reported in the previous paper (**Figure 1C**) (6). We expected that differences in EEG data would be larger at the clinical stage than the preclinical stage, and thus we used EEG from the clinical stages for training machine classifiers. Prediction at the preclinical stage has high clinical relevance.

We pre-processed EEG data using power spectrum analysis, and calculated power spectrum of the EEG data from each mouse at 1–100 Hz bins (**Figure 2A**). The mean powers of three representative wave bands were calculated: theta (5¬–7 Hz), beta (16–30 Hz), and gamma (41–80 Hz) bands. This procedure provided three measures (theta, beta, and gamma powers) from each mouse. Theta and beta bands are thought to be involved in the communication between distant brain regions, whereas gamma bands are shown to have synchronizing neurons in each brain region (6). We tested each of these bands in our analyses.

We then trained machine classifiers by providing theta, beta and gamma powers and animal type ("healthy old" or "AD old"). When provided with these data, the machine classifiers determined optimized parameters for their implemented functions (i.e. trained) to provide the best classification for the provided data. The same data from healthy old mice and AD old mice was run through the trained machine classifier using a 10-fold cross-validation method. We initially tested five representative machine classifiers available in MATLAB software. Out of the five different machine classifiers that were trained, the Gaussian Naïve Bayes classifier proved to be the most accurate with 88.0 ± 2.4% accuracy. The other classifiers, Fine Tree, Linear SVM, Cubic KNN, Medium Neural Network had accuracies at 78.5 ± 1.3%, 81.0 ± 1.8%, 77.5 ± 2.0%, and 79.0 ± 2.8%, respectively. The Gaussian Naïve Bayes Classifier had a significantly higher accuracy than the three other classifiers ($p<0.05$, post-hoc Bonferroni test after ANOVA).

Gaussian Naïve Bayes classifier predicted amyloid-β in young mice with high accuracy

We then used the Gaussian Naïve Bayes Classifier, trained with EEG data from healthy old mice and AD old mice, to examine the prediction accuracy from EEG of healthy young mice and AD young mice as inputs, without telling the classifier which mouse type each EEG was recorded from. We calculated prediction accuracies by validating with the correct answers, which were the mouse type each EEG was recorded from (healthy young or AD young). We repeated this procedure by using three distinct frequency bands from the EEG data (theta, beta, and gamma) to obtain three separate prediction accuracies (**Figure 3**). Results show that the classifier was able to predict AD in young mice with an accuracy of 80.9 ± 0.03% using theta wave data. However, beta and gamma wave data provided prediction accuracies at the chance level of 48.1 ± 0.02% and 55.2 ± 0.03%, respectively. Theta wave EEG data provided the higher prediction accuracy compared to other waves ($p<0.001$, post-hoc Bonferroni test after ANOVA). These results indicate that, when theta waves were used, the Gaussian Naïve Bayes classifier predicted the existence of amyloid-β accumulation in preclinical stage brains with an accuracy as high as 81%.

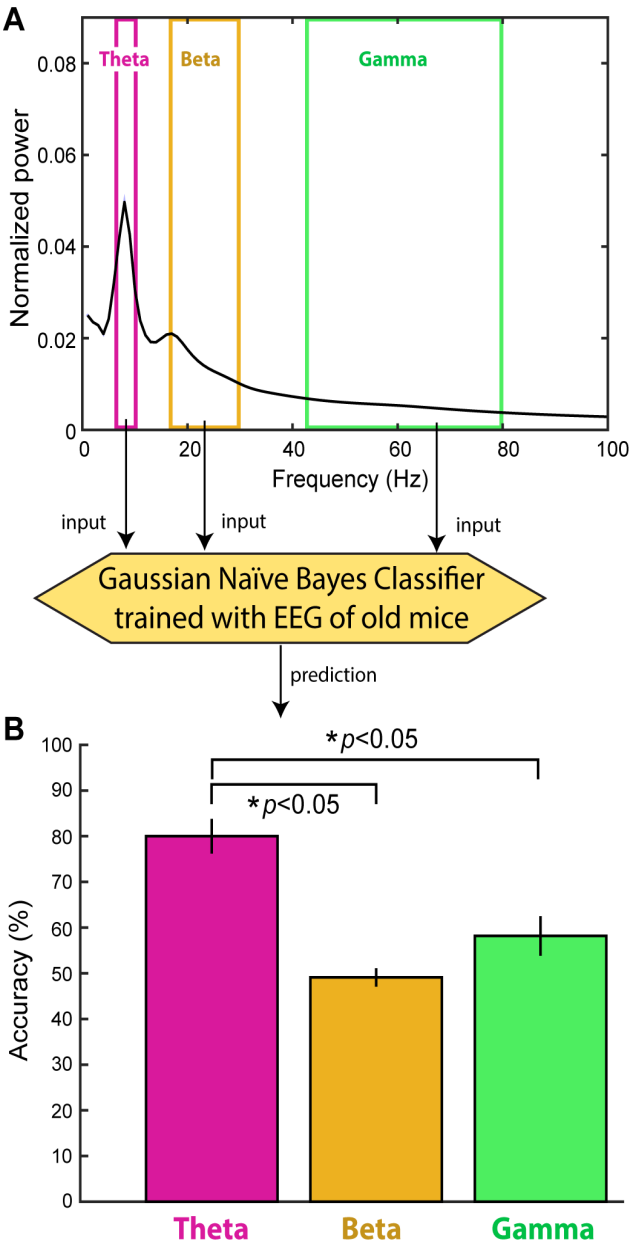

Figure 3: Gaussian Naïve Bayes classifier predicted amyloid-β in young mice with high accuracy. A: An example EEG power spectrum recorded from young mice. Theta, beta, and gamma powers were independently used for inputs to the Gaussian Naïve Bayes classifier trained with EEG of old mice. **B:** Prediction accuracy of Gaussian Naïve Bayes classifier using EEG of young mice. Theta wave data provided the highest accuracy. *=p<0.05, ANOVA followed by post-hoc Bonferroni test. N=10 bootstrap repetitions.

DISCUSSION

In this study, we showed that the existence of amyloid-β in AD mice can be predicted in preclinical stage brains at

high accuracy using EEG and machine learning. A similar study using machine learning and EEG from old AD patients recently reported 85% accuracy, but preclinical stage patients were not tested (9). By contrast, our results show that machine classifiers can provide up to 81% accuracy even for preclinical AD mice, when machine classifiers were trained with data from clinical stage AD mice. Early diagnosis of amyloid-β is critical; therefore our results may show promise for preclinical AD patients.

The selection of classifiers and EEG wave bands used in this study may not be the best for human EEG data and should be tested in future studies. In this study, the Gaussian Naïve Bayes classifier performed the best out of the classifiers tested, although other classifiers may also provide good prediction for human EEG data. As for the wave bands, the theta band (5–7 Hz) yielded 81% accuracy, whereas the beta band (16–30 Hz) and the gamma band (41–80Hz) yielded 48% and 55% accuracies, respectively. This may be because theta waves show the earliest change in power at a young age. The increase of theta oscillation power observed in AD mice in our study is consistent with a study showing increased theta oscillation power in AD patients with mild cognitive impairment (10). This suggests an importance of theta band for early detection. Theta oscillations are thought to be a brain mechanism that coordinates activity of various brain regions, and the loss of theta oscillations in AD may result in cognitive impairment (6). However, this should be carefully tested in human data in future studies.

The next step would be to perform the same experiment on AD patients during both preclinical and clinical stages. If this method could predict amyloid-β deposition at a high accuracy in the future, it could be used for diagnosis well before the progression of AD within patients. In the future, people could receive the option to record their EEG and monitor potential risk of carrying amyloid-β deposition, along with other medical checks. Together with a future cure or form of therapy, this approach provides a potential therapeutic benefit for enhancing management and care toward AD patients.

MATERIALS AND METHODS

Pre-processing of EEG data

EEG data, recorded from Healthy mice and AD mice for approximately one hour, has been published previously (6). All procedures were performed using MATLAB software (MathWorks). EEG data was separated into four groups according to the age of mice (young [3–5 months old] or old [10 months old]) and existence of amyloid-β (healthy or AD) examined after EEG recording (Jun et al., 2020): healthy old (N=10 mice), AD old (N=10 mice), healthy young (N=5 mice), and AD young mice (N=6 mice). Power spectrum analysis was performed for EEG data from each mouse to obtain power spectrum (dB) for each frequency (1–100 Hz, in a 1 Hz bin) using MATLAB software. Mean powers of theta (5¬–7 Hz), beta (16–30 Hz), and gamma (41–80 Hz) bands were calculated. This procedure provided three measures (power spectrum data for theta, beta, and gamma bands) for each mouse.

Machine Learning Procedure

Step 1: Machine learning classifiers were trained using EEG data from Healthy old and AD old mice groups. Five representative machine classifiers were trained: Fine Tree Classifier, Gaussian Naïve Bayes Classifier, Linear Support Vector Machine, Cubic k-Nearest Neighbor Classifier, and Medium Neural Network Classifier. The performance of classifiers was tested using a 10-fold cross-validation procedure: Data was randomly divided 10 equal-sized groups, and then 9 of these groups are used for training and 1 of these groups are used for testing. This procedure was repeated 10 times.

Step 2: EEG data from healthy young and AD young mice were then tested. Power spectrum data in theta, beta and gamma bands from healthy young and AD young EEG was used as inputs for the previously trained Gaussian Naïve Bayes classifier, without telling which group each data was derived from. The classifier returned predicted output ("healthy" or "AD") for each mouse.

Step 3: The output from the classifier was tested. The output was validated with the correct answer, whether the data came from a healthy old mouse without amyloid-β, or AD mouse with amyloid-β deposition examined histologically in the previous paper (6). Steps 2–3 were repeated for N=10 trials using the bootstrapping method (11 random samples were chosen from N=5 healthy young mice and N=6 AD young mice by allowing multiple sampling from same animal, repeated 10 times). Accuracy of machine prediction was statistically examined using analysis of variance (ANOVA) and Bonferroni post-hoc test.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Li, Ms. Hudash, and Ms. Kwon for their advice and mentorship through the project as well as their science teacher, Ms. Kaak, and their parents for supporting them.

Received: May 22, 2022
Accepted: August 14, 2022
Published: November 29, 2022

REFERENCES

1. National Institute on Aging. "Alzheimer's Disease Fact Sheet." www.nia.nih.gov/health/alzheimers-disease-fact-sheet Date Accessed, Apr 18, 2022.
2. Makin, Simon. "The amyloid hypothesis on trial." *Nature*, vol. 559, no. 7715, 2018, pp. S4-S7. Doi.org/10.1038/d41586-018-05719-4
3. Congdon EE, et al. "Tau-targeting therapies for Alzheimer disease". *Nat Rev Neurol*. vol. 14 no. 7, 2018, pp. 399-415. doi: 10.1038/s41582-018-0013-z

4. Sasaguri, Hiroki, et al. “APP mouse models for Alzheimer's disease preclinical studies.” *EMBO J*, vol. 36, no. 17, 2017, pp. 2473-2487. doi:10.15252/embj.201797397
5. Dubois, Bruno, et al. “Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria.” *Alzheimer’s and Dementia*, vol. 12, no. 3, 2016, pp. 292-323, doi:10.1016/j.jalz.2016.02.002.
6. Jun, Heechul, et al. “Disrupted Place Cell Remapping and Impaired Grid Cells in a Knockin Model of Alzheimer's Disease.” *Neuron*, vol. 107, no. 6, 2020, pp. 1095-1112, doi:10.1016/j.neuron.2020.06.023
7. Mathworks. “MATLAB Machine Learning Onramp.” www.mathworks.com/learn/tutorials/machine-learning-onramp.html Date Accessed, Apr 18, 2022.
8. Jun, Heechul, et al. “Protocol for remapping of place cells in disease mouse models.” *STAR Protocols*, vol. 2, no. 3, 2021, pp.100759, doi:10.1016/j.xpro.2021.100759
9. Kim, Nam Heon, et al. “Machine Learning to Predict Brain Amyloid Pathology in Pre-dementia Alzheimer's Disease Using QEEG Features and Genetic Algorithm Heuristic.” *Frontier in Computational Neuroscience*, vol. 15, 2021, pp. 755499, doi:10.3389/fncom.2021.755499
10. Musaeus, Christian Sandøe, et al. “EEG Theta Power Is an Early Marker of Cognitive Decline in Dementia due to Alzheimer's Disease.” *Journal of Alzheimer's Disease*, vol. 64, no. 4, 2018, pp. 1359-1371, doi:10.3233/JAD-180300