

Early Electrophysiological Aberrations in the Hippocampus of the TgF344-AD Rat Model as a Potential Biomarker for Alzheimer's Disease Prognosis

Faraz Moradi¹⁺, Monica van den Berg^{2,3+}, Morteza Mirjebreili⁴, Lauren Kosten^{2,3}, Marleen Verhoye^{2,3}, Mahmood Amiri^{2,3,5*}, Georgios A. Keliris^{2,3,6*}

1 Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, Canada

2 Bio-Imaging Lab, University of Antwerp, Belgium

3 μ NEURO Research Centre of Excellence, University of Antwerp, Antwerp, Belgium

4 Institute for cognitive science studies, Tehran, Iran

5 Medical Technology Research Center, Kermanshah University of Medical Sciences, Kermanshah, Iran

6 Institute of Computer Science, Foundation for Research & Technology, Hellas, Heraklion, Crete, Greece

+: equal contribution

* Corresponding authors

Mahmood Amiri

Bio-Imaging Lab, University of Antwerp, Belgium

Email: Ma_amiri_bme@yahoo.com

and

Georgios A. Keliris

Institute of Computer Science, Foundation for Research & Technology, Hellas, Heraklion, Crete, Greece

Email: georgios.keliris@outlook.com

Early Electrophysiological Aberrations in the Hippocampus of the TgF344-AD Rat Model as a Potential Biomarker for Alzheimer's Disease Prognosis

Abstract

The hippocampus is thought to guide navigation and has an essential contribution to learning and memory. Hippocampus is one of the brain regions impaired in Alzheimer's disease (AD), a neurodegenerative disease with progressive memory impairments and cognitive decline. Although successful treatments for AD are still not available, developing new strategies to detect AD at early stages before clinical manifestation is crucial for timely interventions. Here, we investigated in the TgF344-AD rat model the classification of AD-transgenic rats versus Wild-type littermates (WT) from electrophysiological activity recorded in the hippocampus of freely moving subjects at an early, pre-symptomatic stage of the disease (6 months old). To this end, recorded signals were filtered in two separate frequency regimes namely low frequency LFP signals and high frequency spiking activity and passed to machine learning (ML) classifiers to identify the genotype of the rats (TG vs. WT). For the low frequency analysis, we first filtered the signals and extracted the power spectra in different frequency bands known to carry differential information in the hippocampus (delta, theta, slow- and fast-gamma) while for the high frequency analysis, we extracted spike-trains of neurons and calculated different distance metrics between them, including Van Rossum (VR), Inter Spike Interval (ISI), and Event Synchronization (ES). These measures were then used as features for classification with different ML classifiers. We found that both low and high frequency signals were able to classify the rat genotype with a high accuracy with specific signals such as the gamma band power, providing an important fraction of information. In addition, when we combined information from both low and high frequency the classification was boosted indicating that independent information is present across the two bands. The results of this study offer a better insight into how different regions of the hippocampus are affected in earlier stages of AD.

Keywords: Alzheimer's Disease, Spike Sorting, Hippocampus, Local Field Potential, Machine Learning

Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder affecting millions of people each year and has become one of the biggest socioeconomic burdens to societies. In the United States alone, over 5.7 million Americans were living with AD in 2018, and estimates indicate that this amount will increase to nearly 14 million by 2050¹. The hallmarks of AD include the excessive aggregation of amyloid- β (A β) and tau proteins in brain areas including the hippocampus and adjacent mediotemporal cortex leading to progressive cognitive decline and, ultimately, dementia². One of the first symptoms in AD is spatial memory impairment and wandering behavior with these symptoms occurring in over 60% of the patients³. Regardless of the many efforts and significant progress scientists and physicians have made in understanding AD, its leading cause and prognosis remain uncertain². Current treatment methods can only decelerate AD progression especially if used in relatively early stages of AD. However, cognitive symptoms that are the leading trigger for AD diagnosis appear only at later stages of the disease timeline. Therefore, it has become essential to develop new strategies for the detection and prognosis of AD at early stages before the clinical manifestation of the disease⁴.

Hippocampus is a vital brain area that is involved in numerous cognitive functions including spatial navigation, learning, and memory⁵. Over several decades, investigations of hippocampal function have focused heavily on neurophysiological network activity, including the gamma and theta rhythms⁶. It has been demonstrated that theta oscillations, ranging in frequency from 4 to 12 Hz, are linked to a variety of cognitive processes and have an important contribution to learning and memory as observed in a wide range of experiments in both humans and animals⁶⁻⁹. Gamma oscillations, on the other hand, are characterized by activity in the frequencies ranging from 30 to 100 Hz¹⁰ and have been linked to functions such as sensory binding¹¹, selective attention¹²⁻¹⁴, transient neuronal assembly formation¹⁵, and information transmission and storage¹⁶⁻¹⁸. Moreover, both types of oscillations, that are part of the low-frequency component of electrophysiological recordings called local field potentials (LFP; \sim 0 - 150 Hz), have been identified as critical components in the consolidation of memories and the performance of executive functions^{6,19-24}. Also, another type of oscillation is high-voltage spindles (HVSs). They range from 5 to 13 Hz and have a characteristic spike-and-wave shape which has been studied mainly in somatosensory systems^{25,26}, and only a few recent studies have seriously investigated their occurrence in the basal ganglia (BG)^{27,28}. Theta and gamma oscillations do not exist in isolation; instead, they interact and co-occur in several areas of the brain²⁹. A specific type of cross-frequency coupling (CFC), named phase-amplitude coupling (PAC), is characterized by the coupling between the phase of the theta and the amplitude of the gamma rhythms²⁶. Moreover, research indicates that PAC in the hippocampus may play an essential role in learning and memory²⁹. For humans, the magnitude of the coupling is found to be positively associated with cognitive processes²⁹⁻³¹. Furthermore, PAC has been proposed to play a key functional role in learning task performance³², facilitation of inter-area communication^{33,34} and can serve as a memory buffer³¹. In relation to AD, there have been numerous articles suggesting that the oscillatory activity including power and/or coupling in theta³⁵⁻³⁷ and gamma^{35,38} rhythms is altered in AD patients, and consistent changes were also found in animal experiments using AD models^{39,40}.

Beyond disturbances in the theta and gamma rhythms in the hippocampus and medial entorhinal cortex (MEC), many research papers also identified AD related disruptions in the activity of single neurons such as place cells and grid cells³⁹⁻⁴¹. Place cells are neurons that selectively respond to specific location in the environment (Place field) first discovered by O'Keefe in CA1 region⁵². The activity of single neurons is characterized by discrete events called action potentials or spikes which are considered as the fundamental element of neuronal communication and can be extracted from the high-frequency component of electrophysiological recordings (> 300 Hz). One or more spikes in a sequence are generated whenever a single neuron is stimulated or even at rest (spontaneous activity), and a sequence of spikes forms what is referred to as a spike train⁴². It is generally accepted that the spike responses of each neuron are governed from two main information processing paradigms: rate coding and/or temporal coding⁴³. Measuring the synchrony of two sets of spike trains has been demonstrated to carry a lot of information in several different contexts. For example, spike train distance metrics were used to estimate the reliability of responses across repeated presentations of the same stimulus and to evaluate the information flow among coupled neurons^{44,45}. Furthermore, some recent studies reported the application of measuring spike train distance metrics for the classification of tactile afferents using artificial spike sequences⁴⁶, raising the possibility that such metrics could potentially be used for the classification of AD in animal models.

With the rise of machine learning (ML) and deep learning (DL), numerous applications have been implemented towards the detection and classification of brain diseases. Many models used high dimensional features from various medical imaging techniques such as magnetic resonance imaging (MRI) and positron emission tomography (PET) for the prediction and diagnosis of AD. To this end, they employed classical ML methods like support vector machine (SVM) and k-nearest neighbor (k-NN), but also DL models such as convolutional neural networks (CNN) and recurrent neural networks (RNN). For example, Lee *et al.*⁴⁷ has proposed a new model to predict AD in which RNN was used considering multimodal features extracted from MRI and longitudinal cerebrospinal fluid (CSF) data. El-Sappagh *et al.*⁴⁸ used ensemble ML classifiers based on Random Forest (RF) to develop an AD diagnosis and progression detection model on 11 MRI modalities. Venugopalan *et al.*⁴⁹ used different models of ML (SVM, RF, and k-NN) and DL models using a fusion of multiple data modalities like MRI and genetic single nucleotide polymorphisms (SNPs) to classify patients into AD, mild cognitive impairment (MCI), and control groups. However, even though the above-mentioned methods demonstrated considerable success, in their great majority, they do not provide insights regarding the underlying disturbances at the local neuronal network and single neuron level.

In this study, we used a very promising rat model of AD that demonstrates all the pathophysiological hallmarks present in AD patients (Tg-F344-AD) and compared it with littermate wild-type (WT) controls. We put forward the hypothesis that local neuronal network features measured with electrophysiological recordings in the hippocampus of freely moving rats at an early time-point during disease progression, when amyloid plaque accumulation is only at the onset (6 months old), can be used in combination ML-based methods to classify the rats into the two respective groups, namely AD and WT. Further, we were interested to identify which of these signals better classify the rats and if classification with a combination of signals can be advantageous. Specifically, we considered both low-frequency LFP signals as well as high-frequency spiking activity. To see the effect of low-frequency information in classification, we performed time-frequency analysis with focus on four specific LFP frequency bands (delta, theta, slow gamma, fast

gamma) as well as CFC measures such as PAC. On the other hand, for the high-frequency analysis, we first used spike sorting algorithms to extract the spike trains of neurons and then computed different spike-train distance metrics such as Van Rossum (VR), Inter Spike Interval (ISI), and Event Synchronization (ES). We then used those measures as classification features to investigate the performance of different ML-classifiers across the two groups. We found that specific low and high frequency signals used with chosen ML methods could provide classification accuracies over 80% and that combining information from both could boost the accuracy to 100%. Our findings provide evidence that the disruption in local neuronal signaling in the hippocampus of an AD rat model could be used at a very early time-point of disease progression to classify between AD and WT rats and thus provide promise for the development of novel disease biomarkers for AD.

MATERIALS AND METHODS

Animals and ethical statement

All procedures were in accordance with the guidelines approved by the European Ethics Committee (decree 2010/63/EU) and were approved by the Committee on Animal Care and Use at the University of Antwerp, Belgium (approval number: 2019-06). Electrophysiological experiments were performed in 6-month-old TgF344-AD rats (N=5) and wild-type littermates (N=4). Rats were group-housed prior to head-stage implantation but housed separately afterward. All animals were kept on a reversed, 12h light/dark cycle, with controlled temperature (20 – 24°C) and humidity (40-60%) conditions. Standard food was provided ad libitum. Animals were water-deprived 24h prior to the habituation and acquisition of the LFP data. After the animal performed the task in the linear track, water was provided for 30 minutes.

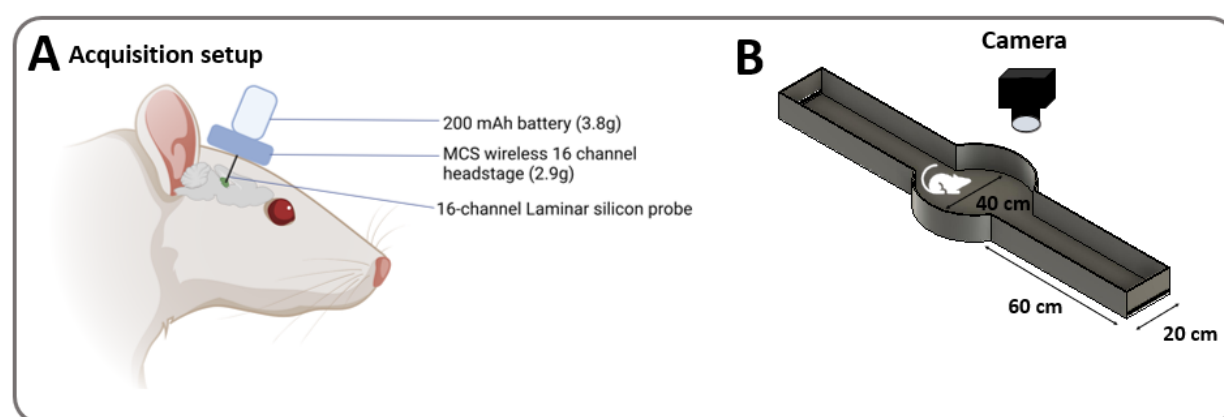


Fig 1. A. Graphical representation of the wireless acquisition system used to record LFP's and neuronal activity. A head-stage was attached to the laminar silicon probe which was placed in the right dorsal hippocampus. The head-stage converts the analog signals recorded from the brain to digital signals, which were then send to the receiver hardware. The head-stage was powered by a small battery mounted on top of the head-stage to allow maximum freedom of movement. B. Graphical representation of the track where the animals freely moved during the recordings. It consisted of a circular area in the middle with a diameter of 40 cm and two rectangular arms of 20x60 cm. At the end of each arm, rewards were periodically provided to motivate the animals to move back and forth.

Chronic hippocampal electrophysiological measurements

Surgical procedure

For the implantation of the electrodes, animals were anesthetized using isoflurane (at 1.0 L/min, induction 5% and maintenance 2-3% isoflurane). Animals were placed in a stereotaxic frame and a craniotomy was made above the right dorsal hippocampus (AP -3.00, ML 2.50). A 16-channel laminar electrode (E16+R-100-S1-L6 NT, Atlas Neuro-engineering, Belgium) with internal reference was placed into the dorsal hippocampus by penetrating the dura (Fig. 1A). The depth of the recording sites was identified by the layer-specific local field potentials (LFP) of the hippocampus (DV 2.5-3.5 mm). The craniotomy was sealed with a sterile silicone gel (Kwik-Cast, WPI). Stainless steel screws were drilled into the skull overlaying the olfactory bulb, left hippocampus and cerebellum, of which the latter served as a ground electrode. Two EMG wires were stitched in the neck muscle in order to record EMG activity. The implant was covered in several layers of dental cement (Stoelting, Co, Dublin, Productnumber 50000,) and the wound was closed. Rats were allowed to recover for at least 7 days.

Linear track acquisition

Animals were habituated to the behavioral room and linear track (Fig. 1 B) for 1 day, after which the rats were trained for 5 consecutive days to walk along the linear track. Cups with sugar water at the end of the arms were used to motivate the animals to walk back and forth. Habituation sessions and training sessions lasted 20-30 min. The combined electrophysiological and behavioral acquisition was performed during two, or three consecutive days, based on the performance (active time) of the rats in the linear track. A wireless head stage (W2100, MultiChannel Systems, Germany) was connected to the laminar electrode, 15 to 30 minutes prior to start of the experiment in the linear track. After the habituation to the acquisition setup, animals were placed in the center of the track and the electrophysiological recordings were started. The movements of the animal were recorded using a camera mounted on the ceiling above the linear track, and the freely available Bonsai acquisition software (Bonsai — Open Ephys (open-ephys.org)) was used to detect the animal position over time. An acquisition session ended when the animal was not moving around for more than 5 minutes.

Data Analysis

Preprocessing of LFP signals

Data was recorded from the 2 EMG channels and the laminar electrode consisting of 14 intracranial channels. The top three intracranial electrodes were located outside the hippocampus and were not used in our analysis. The remaining 11 intracranial channels were used for LFP analysis and spike sorting. To prevent the effect of impedance differences between electrodes on LFP, the LFP power for each electrode was normalized to the power of the 0.5-1 Hz band. The first and last 60 sec of each recording session were excluded to eliminate possible artifacts on the neural activity patterns associated with handling or moving into or out of the linear track. The LFP signals were extracted from the broad-band signal (sampling rate = 10 kHz) using a 3rd-order Butterworth filter (MATLAB function *filtfilt*) with a

low cut-off frequency of 0.1 Hz and a high cut-off frequency of 300 Hz. After filtering the signal was downsampled to 1000 Hz and stored for further analysis. Out of 20 recording sessions from 9 rats (n=4 WT rat, n=5 TG rat), two of the samples (WT3_D1 and TG1_D2) were excluded from our further analysis because of excessive noise.

Power spectrum analysis

The power spectrum density (PSD) was estimated using a 2-second window size (MATLAB function *pwelch*). We computed the power between frequencies 1–100 Hz and normalized it to the power between 0.5-1 Hz as also indicated above. Then, the power at four individual frequency bands namely delta (1-4 Hz), theta (4-12 Hz), slow gamma (30-50 Hz), and fast gamma (50-100 Hz) bands was used for further analysis⁴⁰.

Time-frequency analysis of power and frequency coupling

The Chronux⁵⁰ toolbox in MATLAB (function *mtspecgramc*) was used for the time-frequency analysis. As the duration of recordings for each sample was between 17 to 30 minutes, we have chosen a snippet of 15 minutes (900 seconds) from each recording. The toolbox has parameters like window and step size, tapers which were tuned to obtain proper representations (window and step size = 15, taper = [5 9]). For computing cross-frequency coupling between desired phase and amplitude ranges, the Buzcode toolbox⁵¹ from Buzsaki Lab was used. The modulation index of phase-amplitude between desired ranges was computed for all data from each rat. Each representation is the result of averaging across all other channels.

Detection of sleep spindles

In our analysis, 15 min of recordings were used for all analyses. Automatic detection of high-voltage spindles (HVSs) was identified in the hippocampus when rhythmic negative deflections lower than -0.3 mV occurred while instantaneous power between 6 Hz and 12 Hz was more than double instantaneous delta power for 2 s or longer. The average time-frequency spectrum was calculated within frequencies between 0-20Hz for each animal.

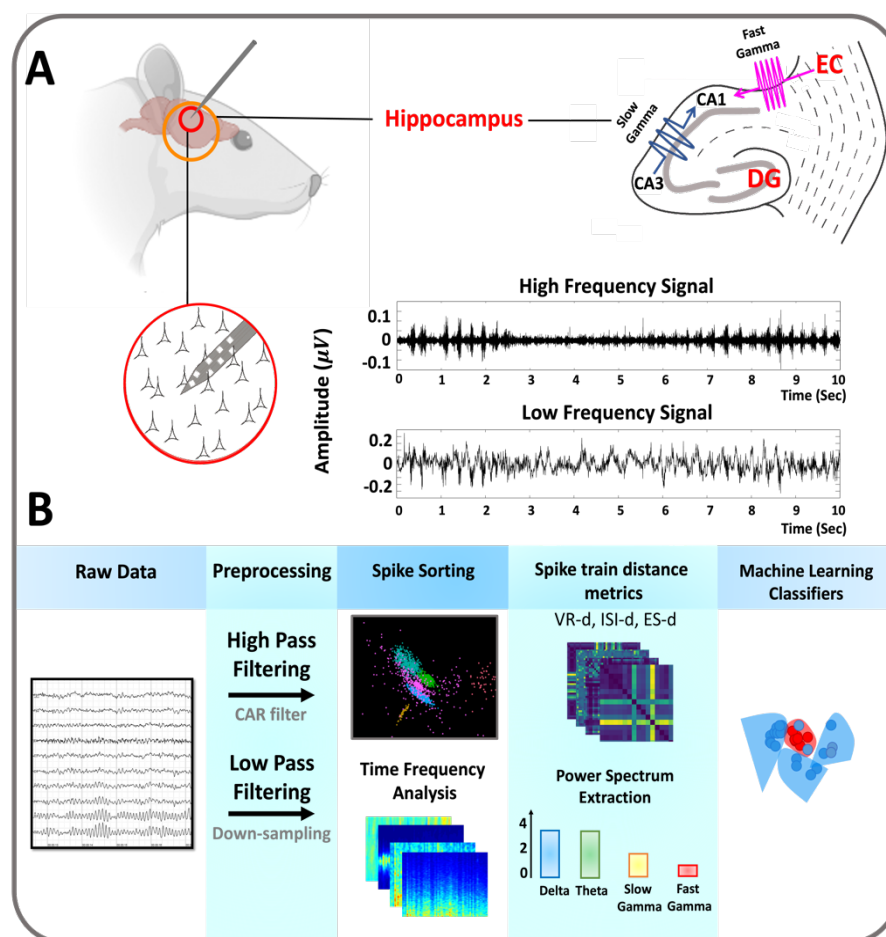


Fig. 2. Representation of the recording and ML models. **A)** The relative positions of Hippocampus, entorhinal cortex (EC) and dentate gyrus (DG) formation are illustrated. Fast and slow gamma oscillations reflect inputs from medial EC and CA3, respectively. Graphs show representative data recorded in the hippocampus of one rat. The spike sorting algorithm runs on the high-frequency signal (band-pass filtered between 300 and 4900 Hz) of the raw data and uses common average reference (CAR) filter to remove mutual noise among channels. The low-frequency signal (band-pass filtered between 0.1 and 300 Hz) of the raw data was analyzed using time-frequency methods to extract desired frequency bands and then perform the classification task. **B)** Graphical representation of the two ML models. In the high-frequency model, spike sorting was performed after data acquisition and preprocessing to obtain the units spike train responses. Next, spike train distance and synchrony metrics were calculated, PCA was applied for dimension reduction and finally this information was fed into the ML classifiers that distinguished TG and WT rats. In the low-frequency model, the primary step was to analyze the time-frequency plots and then estimate the average power for individual frequency bands. By obtaining the most compelling features, the classification of TG and WT rats was then accomplished.

Classification Models

The rat's hippocampus structure is depicted in Fig. 2A. This brain region receives information from various parts of the brain and is believed to support information about the spatial representation of our environment⁵². The most cortical input comes from the entorhinal cortex (EC)⁵³. Encoded information transfers from EC \rightarrow CA1 DG via the perforant path (synapse 1), DG \rightarrow CA3 via mossy fibres (synapse 2), CA3 \rightarrow CA1 via schaffer collaterals (synapse 3)

which together they form the trisynaptic circuit⁵⁴. Therefore, we have decided to do the recording from this brain region as an essential organ in navigation and memory.

In Fig. 2B, the steps undertaken for the classification models are shown. We have put forward two different pipelines for the classification task: 1. Using Low-Frequency Signals (LFS) and 2. Using High-Frequency Signals (HFS). Moreover, we also combined the information from both these models to estimate the total performance of using these local network features for classification. In the LFS model, the average power spectrum density (PSD) is computed for the frequency bands of interest (delta, theta, slow gamma, fast gamma) and then the feature importance method was used by calculating the Mean Decrease in Impurity (MDI) to determine the effectiveness of each frequency band in the classification of TG and WT rats. The LFP data from each rat comprises the model input and is defined as $X (X \in \mathbb{R}^{timepoint \times Channel})$. The HFS model comprises of three main steps: (1) Running spike sorting algorithms using Tridesclous sorter⁵⁵ and SpikeInterface⁵⁶ toolboxes to isolate spike responses of neural activities; (2) Extracting spike train distance and synchrony metrics including Van Rossum (VR-d), Inter Spike Interval (ISI-d), and Event Synchronization (ES-d) for each TG and WT group; and (3) Applying principal component analysis (PCA) to reduce the dimensionality of the feature space.

1. Classification Using Low-Frequency Signals (LFS)

Spectral Power Model

Further analysis on the LFP signal was performed to investigate how TG and WT can be distinguished in low-frequency features using each frequency band's mean power. In this case, the average power of delta (1-4 Hz), theta (4-12 Hz), slow gamma (30-50 Hz), and fast gamma (50-100 Hz) are extracted for every rat as shown in Fig. 3. Therefore, the shape of the extracted feature matrix is $S_{WT} \times 4$ and $S_{TG} \times 4$ where S_{WT} (S_{TG}) is the number of samples for the WT (TG) group. Also, we have analyzed a scenario in which each frequency band is removed one by one from the input features of the ML classifier. Consequently, the shape of each feature matrix is decreased to $S_{WT} \times 1$ and $S_{TG} \times 1$ in the final step. This removal of each frequency band is used to demonstrate how different hippocampal neural oscillations affect the model's performance. Due to the small number of samples, it was not possible to use the commonly used approach that shuffles the data from all samples and then divides them into training, validation, and test sets. Instead, we have used leave-one-subject-out cross-validation (LOSO-CV). In this approach, the number of folds is equal to the number of recording sessions. Thus, the ML classifier is applied once for each sample using all other samples as a training set and the selected sample as a single-item test set. Therefore, we have separated each sample's spike train manually as a test set, and the training was performed using all the remaining spike train features. This process was continued until every recording session was used as a test set.

Feature Importance

The feature importance gives an abstract view about the role of each feature. Although this study aims to build a classifier to discriminate between TG and WT, it is also essential to quantify the importance of each input feature which in turn shows those frequency bands that have been changed in the AD rats. For this purpose, a random forest (RF) classifier was trained with all four frequency bands, and Mean Decrease in Impurity (MDI) has been used as a

measure for evaluating the importance of each feature. The MDI calculates each feature importance as the sum over the number of splits (across all trees) that include the feature, proportionally to the number of samples it splits.

Table. 1 Number of units (spike-trains) extracted after the spike sorting performed on the recorded data. There were 18 recording sessions (samples) from 8 rats (7 samples of WT rats, 11 samples of TG rats).

Number	Subject code	Day	Total Recording Duration (Sec)	Number of units
1	TG1	D1	1624	4
2	TG2	D1	1488	8
3	TG3	D1	1821	2
4	TG4	D1	1563	2
5	TG5	D1	1140	1
6	TG2	D2	1720	7
7	TG3	D2	1801	2
8	TG4	D2	1668	3
9	TG5	D2	1592	2
10	TG2	D3	1801	3
11	TG3	D3	1610	2
12	WT4	D1	1049	2
13	WT5	D1	1506	2
14	WT6	D1	1417	3
15	WT4	D2	1590	2
16	WT5	D2	1528	3
17	WT6	D2	1575	5
18	WT4	D3	1416	2

2. Classification Using High-Frequency Signals (HFS)

Spike Sorting and Multi-unit Extraction

Pre-processing for spike sorting included high-pass filtering with a cutoff at 300Hz and then low-pass filtering with a cutoff at 4900Hz. Also, common reference removal (CAR) filter was used to increase the signal-to-noise ratio.

Then, we used the python-based package Tridesclous Sorter, an automatic spike-sorting toolbox that groups spike responses into separate units based on their similarity in shape and features. We have called these spiking activities “multi-unit,” as neuron isolation from single electrodes on a laminar probe is difficult and thus these may contain

spike responses from a few neighboring neurons. Another python-based package, and namely the Spike Interface toolbox has been used for manual inspection and curation of multi-units. We have defined three operators for each of the steps: Band-pass filter (\odot), Common average reference filter (\oplus), Tridesclous Sorter (\odot). In the preprocessing step, band-pass and CAR filters have been applied to raw data using the Tridesclous toolbox. During the spike sorting step, spikes were detected based on the threshold of 4 standard deviations (SD) of the background noise. All the multi-unit extraction was done automatically. So, the output of each recording session (X) from the spike sorting section would be N multi-units. Each multi-unit has N_s number of spikes. The details of the algorithm are illustrated in Appendix 1.

Further, to remove artifacts such as high frequency noise present in some channels, we have also considered two thresholds on the number of spikes from each multi-unit: the number of spikes in each multi-unit should be higher than 1600 and lower than 360000 for 900 seconds of recordings. After considering this criterion, the number of samples was reduced to 17 and TG3_D3 was removed. The number of multi-units extracted from each recording session is reported in Table. 1. In total, 53 multi-units (19 from WT, 34 from TG) were obtained.

Calculation of metrics reflecting Temporal Coding

In neuroscience, analyzing neural coding realizes the relationship between stimulus and neural response produced by the neuronal network. In temporal coding, the spike times of an individual neuron are considered to carry essential information. Here, to take into consideration the temporal complexity of spike sequences, we extracted different synchronization and distance metrics. These measures can compare two spike patterns and provide information about their relationship (similarity or dissimilarity). We have used three spike distance metrics to compute the spike train relationships among the units of each genotype group: 1-Van Rossum distance (VR-d), 2-Inter-Spike Interval distance (ISI-d), and 3-Event- Synchronization Distance (ES-d). VR-d is a metric for the dissimilarity of spike trains and measures the distance between two sets of spike trains by transforming them into continuous functions by convolving them with an exponential kernel⁵⁷. ISI-d assesses the dissimilarity between spike trains based on instantaneous rate synchrony calculated by the time intervals between spikes. ES-d also measures spike synchrony albeit using a different approach. In contrast to ISI-d, it measures similarity instead of dissimilarity⁵⁸. To calculate these metrics, isolated multi-units from each group of TG and WT have been concatenated to make a unified list of spike trains. The total number of spike trains used for each group is defined by N_{TG} , N_{WT} . In summary, all three spike distance metrics were calculated for each pair of spike trains. For more information on the mathematical description of each metric we refer to Appendix 2.

Machine Learning Classifier

Currently, the DL and ML models have a great potential to be used in hospitals and health care systems for clinical decision support in a variety of diseases including cancers^{59,60} and AD (image analysis)^{61,62}. The number of dimensions in the feature space is one of the important parameters in an ML model that should be controlled to achieve reasonable accuracy while avoiding overfitting. In the final step of our model, we use unsupervised dimension reduction methods to reduce the feature dimensions. For this purpose, we applied PCA and the components carrying the highest explained variance were chosen. We have employed this approach for different numbers of components (1-9), and the best

performance was achieved by considering three components ($P=3$). The training and validation steps were the same as the method discussed in the spectral analysis section.

Statistics

Mean \pm SEM is provided for comparing different frequency bands. Statistical testing for frequency band power, PAC, and spike-train metrics, assumed non-parametric distributions and thus Mann Whitney U tests were used to evaluate differences between experimental groups.

Results

Low-Frequency Signal (LFS) Analysis

Several studies have shown that gamma oscillations are commonly disrupted in AD³⁹⁻⁴¹. More specifically, fast gamma oscillations were shown to be diminished in the hippocampal CA1 region of a knock-in mouse AD model as well as the TgF344-AD rat model^{39,40,63-65}. To evaluate if gamma and potentially other rhythms are impaired at an early stage in the hippocampus of TgF344-AD rats relative to WT controls, we computed the average PSD for different rhythms and specifically delta (1-4 Hz), theta (4-12 Hz), slow gamma (30-50 Hz) and fast gamma (50-100 Hz) (see Materials and Methods). In Figure 3, the average PSD for each pair of these bands is illustrated per recording session (WT: blue circles, TG: red circles). In addition, the diagonal panels show a univariate distribution that represents the marginal distribution of the data in each band (Fig.3). This illustration allows the comparison across the TG and WT groups for each pair of the four frequency bands to identify the ones that can reasonably discriminate between them. The results showed no apparent difference among TG and WT rats when the delta and theta frequency bands were considered with the samples of the two groups close to each other and a high degree of overlap between the distributions. On the contrary, when the slow and fast gamma oscillations were considered, it was immediately apparent that the separation between the distributions of PSD values across the two groups was substantial, promising higher discriminability in feature space and a better classification performance.

To better understand in which frequency bands differences occurred between the two groups, we plotted the power spectra, time-frequency plots, and normalized-power bar plots along with statistical analysis (Fig. 4). We found no significant differences in the average normalized power for delta and theta oscillations that were very similar across both groups (Delta; WT: 1.29 ± 0.03 ; TG: 1.09 ± 0.08 ; $p > 0.01$, Fig. 4 Theta; WT: 1.39 ± 0.19 ; TG: 1.2 ± 1.2 ; $p > 0.01$, Mann-Whitney U test, $n = 7$ WT and $n = 11$ TG recording sessions). Inversely, for both slow (30–50 Hz) and fast (50–100 Hz) gamma oscillations a significant reduction in average PSD observed (Slow gamma; WT: 0.13 ± 0.008 ; TG: 0.075 ± 0.008 ; $p > 0.01$, Fig. 4 Fast gamma; WT: 0.048 ± 0.0024 ; TG: 0.023 ± 0.0026 ; $p > 0.01$, Mann-Whitney U test, $n = 7$ WT and $n = 11$ TG recording sessions).

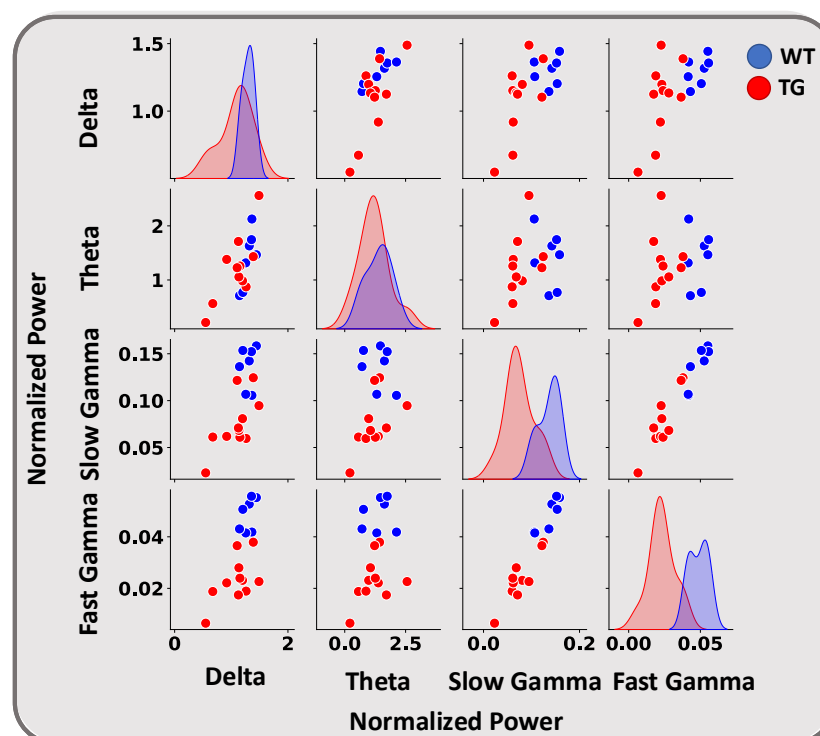


Fig. 3. Pairwise comparison of the normalized power in delta (1-4 Hz), theta (4-12 Hz), slow gamma (30-50 Hz), and fast gamma (50-100 Hz) for TgF344-AD (TG) and Wildtype (WT) groups. Each point represents the average PSD of a single sample in the specified band. The diagonal plots are the univariate distributions that show the amount of overlap between the distribution of two groups in individual frequency bands. The other plots compare WT and TG for specified frequency bands. Blue/Red points are recording session (samples) of WT (n = 7) /TG (n = 11), respectively.

Previous studies used electrophysiology recording in the same rat model have indicated an attenuation in the theta phase – gamma amplitude coupling (PAC) in the hippocampus of 8-9 months old animals^{65,66}. To investigate if PAC, which plays an essential role in memory formation²⁹, shows attenuation already at an earlier time point during disease progression, we assessed PAC across the WT and AD groups at 6 months of age. To this end, we calculated the modulation index (MI) capable of detecting cross-frequency coupling between two different frequency ranges of interest in a single signal as previously described by Tort et al.³². Example results from two TG and two WT rats are shown in Figure 5. The PAC spectrograms (Fig. 5B) revealed that large amplitude slow and fast gamma oscillations occur in the peak to falling phase of theta oscillations in WT; however, slow gamma oscillations seem to be attenuated in TG. Therefore, we have decided to compute the MI for fast and slow gamma among all samples (Fig. 5C). The analysis revealed no significant MI difference for fast gamma oscillation (Fig. 5C fast gamma; WT: $0.000342 \pm 2.8 \times 10^{-5}$; TG: $0.00037 \pm 7 \times 10^{-5}$; Mann-Whitney U test, n = 7 WT and n = 11 TG recording sessions); however, slow gamma coupling was significantly attenuated in TG (Fig. 5C slow gamma; WT: $0.00026 \pm 4 \times 10^{-5}$; TG: $0.000155 \pm 1.7 \times 10^{-5}$; Mann-Whitney U test p > 0.01, n = 7 WT and n = 11 TG recording sessions).

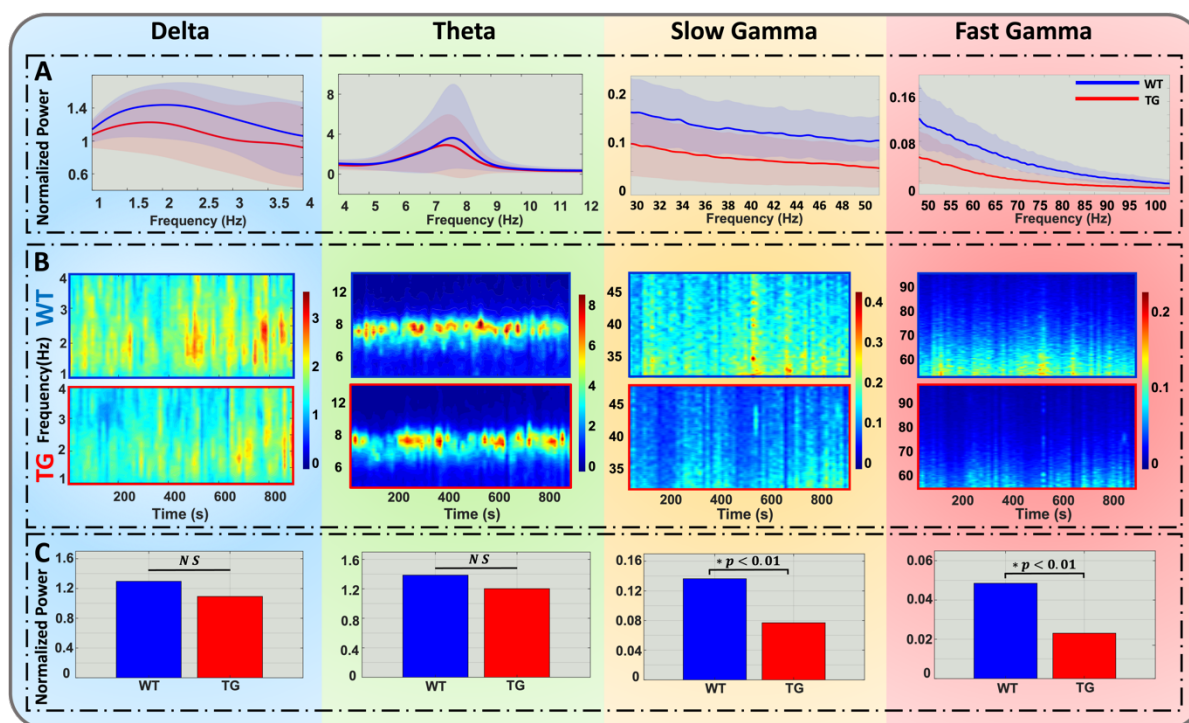


Fig. 4. Comparison of the power spectral density (PSD) (A), time-frequency plots (B), and average PSD of 4 different frequency bands (Delta: 1–4 Hz; Theta: 4–12 Hz; Slow Gamma: 30–50 Hz; Fast Gamma: 50–100 Hz) (C). (A) From left to right average mean PSD of the delta, theta, slow gamma, and fast gamma in WT rats ($n = 7$ recording session) and TG ($n = 11$ recording session). (B) The time-frequency spectra for WT (upper panels) and TG rats (lower panels) for different frequencies of interest. (C) Normalized power for WT and TG rat in the frequency bands of interest. WT = wildtype, TG = TgF344-AD

Synchronous oscillations in various frequency ranges have been recorded in several nuclei of the basal BG and are thought to be an information processing mechanism. HVSs are 5–13 Hz spike-and-wave oscillations, which are commonly recorded in rats, and which have been reported in some recent studies where their occurrence in the BG has been investigated⁶⁷. This distinct oscillatory pattern was present in the cortical area of TgF344-AD in previous studies and very rare in WT rats^{68,69}. However, their occurrence rate was not investigated in the hippocampus of the AD rat model in the earlier stage of AD. HVSs are commonly characterized by an increase in their instantaneous power spectral density values; this must be incorporated into any HVS detection algorithm. The continuous wavelet transforms (CWT), and discrete wavelet transform (DWT) has been effectively applied for the detection of HVS-like nonstationary features such as sleep spindles and spike-wave discharges. We have investigated the alternations of HVSs in both time and frequency domains. The average time-frequency analysis of the two groups indicates no significant difference (Fig. 6A). We have further our analysis to make statistical inferences by calculating the HVS incidence metric (number of events/duration). Our results indicate there was no significant difference between the two groups in the occurrence rate of HVSs (Fig. 6B; WT: 2.7 ± 0.8 ; TG: 2.5 ± 0.6 ; Mann-Whitney U test, $n = 7$ WT and $n = 11$ TG recording sessions).

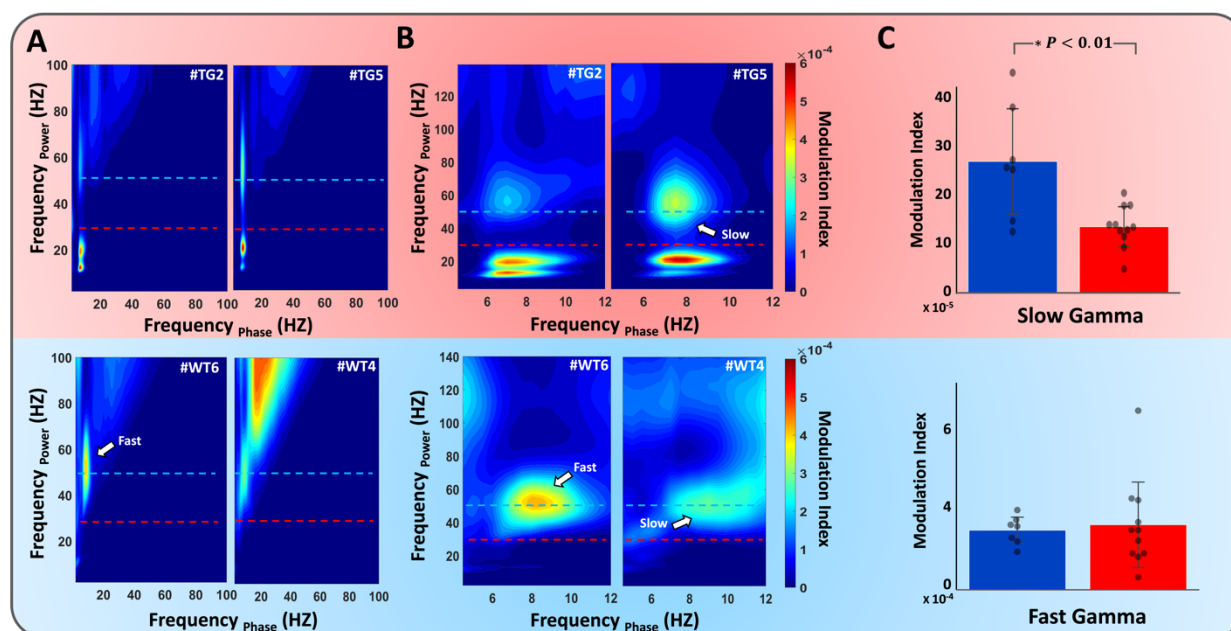


Fig. 5. Slow gamma coupling was diminished in the hippocampal CA1 of 6 months old TgF344-AD rats. spectrograms of phase-amplitude coupling in the hippocampus were represented for two TG and WT rats by computing the modulation index to measure the coupling between desired frequency ranges. **A)** Comodulograms showing the phase and power of 0-100Hz frequency for two representative TgF344-AD (TG) rats (top row) two representative wildtype (WT) rats (bottom row). Modulation index values were coded by color and plotted according to theta and gamma frequency. The dotted lines denote 30 and 50 Hz borders between theta, slow gamma, and fast gamma oscillations. **B)** Comodulogram showing the phase of the theta band (5-12 Hz) in regard to the power in the 0-140Hz frequency band **C)** Modulation Index was computed for coupling of slow (30-50 Hz) and fast gamma (50-100 Hz) oscillations with theta rhythms on all recording sessions (WT: n = 7, TG: n = 11). Each black point represents each recording session of both groups. TG rats exhibited a significant decrease in slow gamma modulation index compared to WT (* $p < 0.01$).

Subsequently, we proceeded to use ML classifiers (see Materials and Methods) to estimate their performance and ability to identify the rats from the WT and TG groups. To investigate how each frequency band affects the model's accuracy, we first trained the ML classifiers with all frequency bands included. In Fig. 7A, the total accuracy among all samples is shown while Fig. 7B-C presents the performance separately in each group. To estimate the importance of each band, the fast gamma, slow gamma, theta, and delta oscillations have been removed one by one and their effects on the classification accuracy were calculated (Fig. 7D). When all frequency bands were used as input of the classifiers (RF, k-NN and SVM) the model performance was as expected maximum. However, the discrimination performance dropped significantly after eliminating fast and slow gamma oscillations indicating their importance in the classification of WT and AD rats.

Given that the number of samples from each rat was slightly different, the above-mentioned classification results could be susceptible to bias against rats with a low number of samples. Therefore, as a control, the results of the LFS classification task were reproduced with a selection of an unbiased sample (one sample from each subject) to ensure that the same results could be achieved. These results are available in the supplementary information (Fig S1, S2, S3).

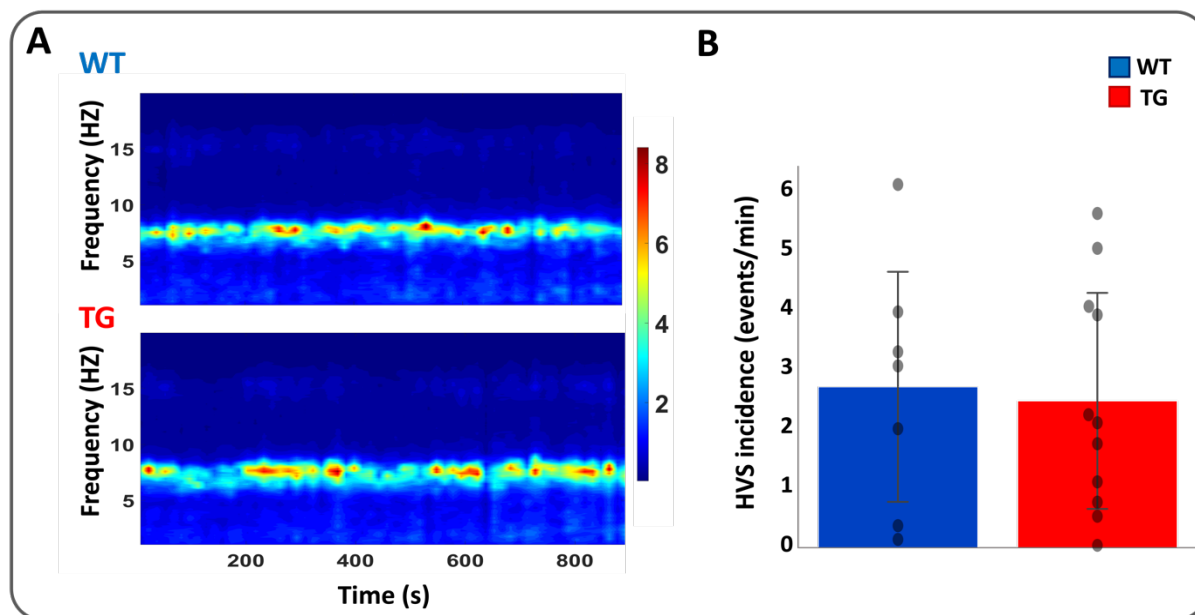


Fig. 6. High-voltage spindles (HVSs) in the cortex of aged wild-type (WT) and TgF344-AD rats. **A)** Average time-frequency plots with a duration of 15 min of the hippocampus from WT and TG rats (WT: $n = 7$, TG: $n = 11$). Large negative deflections in the LFP and prominent bouts of high-power activity between 6 and 9 Hz in the spectrogram are HVS events. **B)** Comparison of HVS incidence rate between TG and WT rat models.

High-Frequency Signal (HFS) Analysis

Signals such as the spectral power in different frequency bands and phase-amplitude coupling, which we used in the LFS analysis, provide straightforward, measurable quantities that are also associated with specific cognitive mechanisms that are relatively well understood. In the case of HFS, the average firing rate of neurons is one of the quantities that is similarly straightforward to calculate and interpret. However, although the firing rate is a quantity that has been extensively used in neuroscience, its use is in most cases coupled with a stimulus (or task) and it is usually the stimulus response function that provides information about the underlying mechanisms and/or aberrations due to pathology. On the other hand, in the case of activity during relatively free behavior such as the one the rats in our study performed, it is very difficult to draw conclusions based on the firing rate alone. The huge variability in the base firing rates, which depend on cell types, brain areas or specific layers within them, as well as variability caused by the behavior itself, make the evaluation of spike trains based on the firing rate problematic and difficult to understand. To this end, spike train distance metrics have been developed that provide a powerful alternative approach to evaluate spiking activity in more general conditions. Moreover, depending on the spike-trains distance metric used, we can deduce information whether changes depend on underlying principles of neural coding such as rate or temporal codes. Here, we used three different spike-train distance metrics: VR-d, a parametric distance based on the temporal structure of a pair of spike trains, ISI-d, a parameter-free spike distance metric that extracts information from inter-spike intervals by evaluating the ratio of instantaneous firing rates, and ES-d, a complementary measure that is sensitive to spike coincidences. Given that these metrics depend on different characteristics extracted from the spike trains, the performance of each metric in the classification task can provide information about the underlying mechanisms being disrupted in AD.

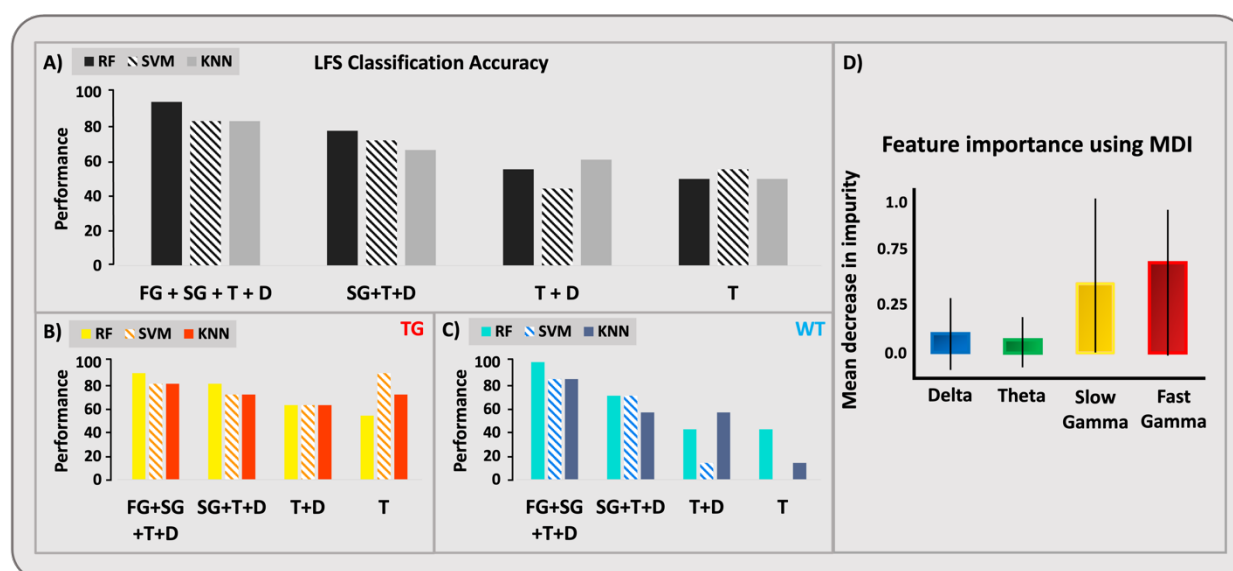


Fig. 7. Classification accuracy of LFS model with three ML classifiers: k-nearest neighbors (k-NN), Support vector machine (SVM), and Random Forest (RF). Frequency bands: Fast Gamma (FG), Slow Gamma (SG), Theta (T), and Delta (D). (A) the overall accuracy. (B) the classification accuracy only for TG rats. (C) the classification accuracy only for WT rats. From left to right in each section, the frequency bands are removed from the classifier's input one by one to see their effects on the accuracy. (D) the results of feature importance analysis. The MDI for each frequency band is calculated which shows the importance of slow and fast gamma oscillations.

In Fig. 8 A and B, the spike distances among all spike patterns calculated in the TG and WT groups are shown. It is immediately obvious that the patterns for each metric show differences and similarities across the two groups. VR-d is a parametric spike train distance which is based on the temporal structure of two pairs of spike trains. In Fig 8 (left panels), we can observe that the VR distance between WT multi-units is generally small. However, the distance between TG multi-units is not consistent, and some spike trains have considerable distance from the others. ISI-d is a parameter-free spike distance metric that extracts information from inter-spike intervals by evaluating the ratio of instantaneous firing rates. In Fig. 8 (middle panels), the ISI distances among both TG and WT spike trains demonstrate similar values and variability and thus it is not immediately obvious if this metric can provide information for classification. A major weakness of ISI-d is the fact that it is not well suited to track synchrony changes based on spike coincidences. On the contrary, ES-d is a complementary metric that is sensitive to spike coincidences while sharing the fundamental advantages of ISI-d. In Fig 8 (right panels), like ISI-d, the values and variability across the two groups are similar and thus it is difficult to judge any difference between the two groups.

As indicated in the Materials and Methods, dimensionality reduction via PCA was subsequently performed and the three components explaining the highest variability were used as features for further analysis. In Figure 9, the results of the PCA are presented in 3D feature space (Fig. 9). Using this representation, it becomes easier to observe that in the case of VR-d the data points of the WT group are very close to each other forming a cluster that is separated from the datapoints of the AD group. Thus, VR-d is a promising metric for classification. For the other two metrics, ISI-d and ES-d, the data points of the two groups are intermixed and thus it is more difficult to predict the results of classification.

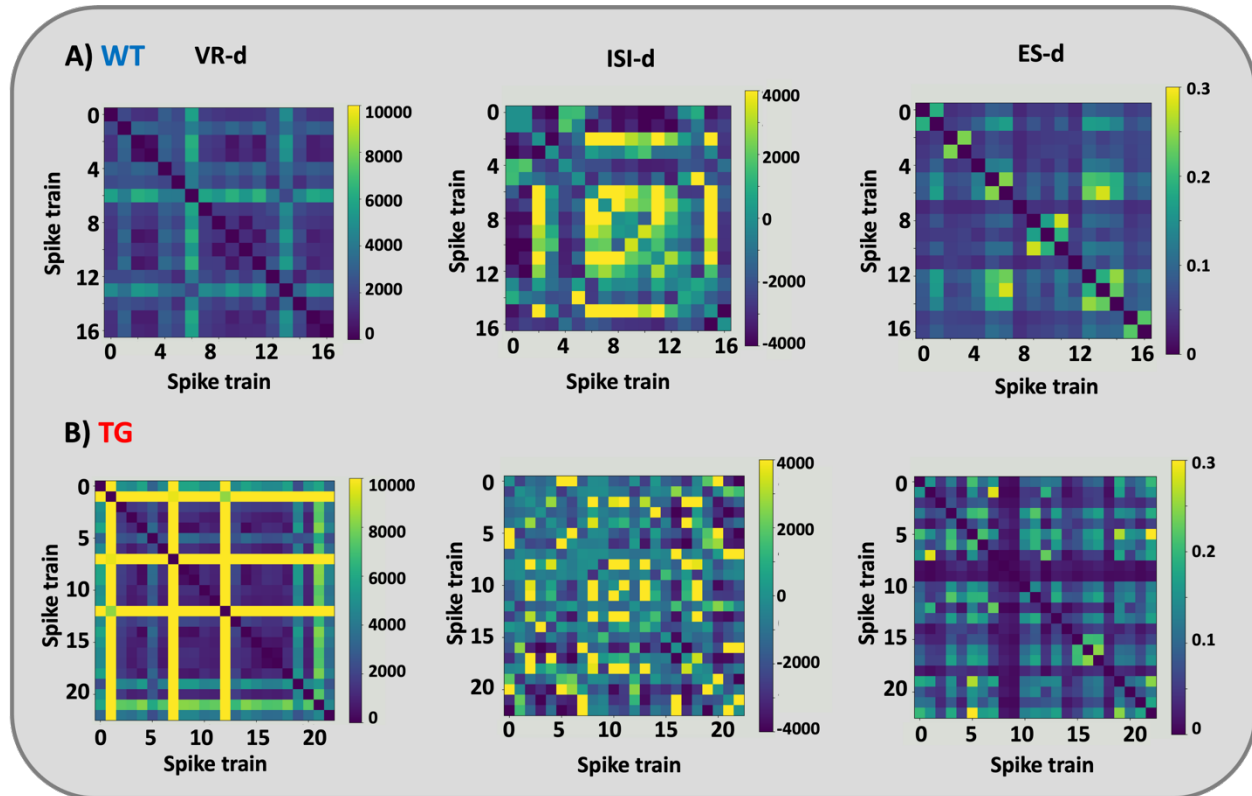


Fig. 8. The spike distance metrics for individual spike responses in (A) TG and (B) WT groups. From left to right, van Rossum distance (VR-d), inter-spike interval distance (ISI-d) and event synchronization distance (ES-d). The blue color shows more similarity, and the yellow color indicates how far the two spike trains are.

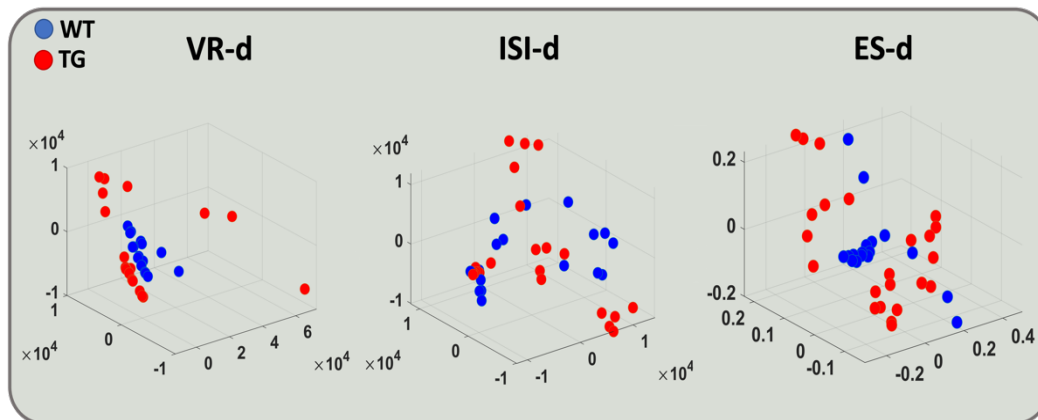


Fig. 9. The feature space of each distance metric in 3D space. The red and blue dots represent TG and WT rats respectively. Each figure is the result of the application of PCA algorithm on the matrix of spike train distances (Fig. 7). Three principal components from each spike distance matrix were considered as the inputs of the classifiers.

Using the features extracted from PCA analysis we then proceeded to using classifiers and namely SVM, k-NN, and RF to classify the two groups of rats based on HFS (Fig. 10). Note that for k-NN, different values of k from 3 to 13 have been tested to evaluate the model. However, we have only reported the result of $k = 5$ as the best result. The

classification results were compatible with the observed separation of groups in the feature space (Fig. 9) with the features based on the VR-d providing the highest total classification accuracy for all classifiers while the ISI-d had the lowest total performance (Fig. 10A). Notably, when the performance for each group was plotted separately (Fig. 10 B-C) a slightly different behavior across groups was observed. For the WT group VR-d showed clearly the highest accuracy reaching 100% for SVM and k-NN classifiers while ISI-d performed very poorly; ES-d had intermediate performance. For the TG group, on the other hand, the maximum classification accuracy was in general less but comparable for each metric.

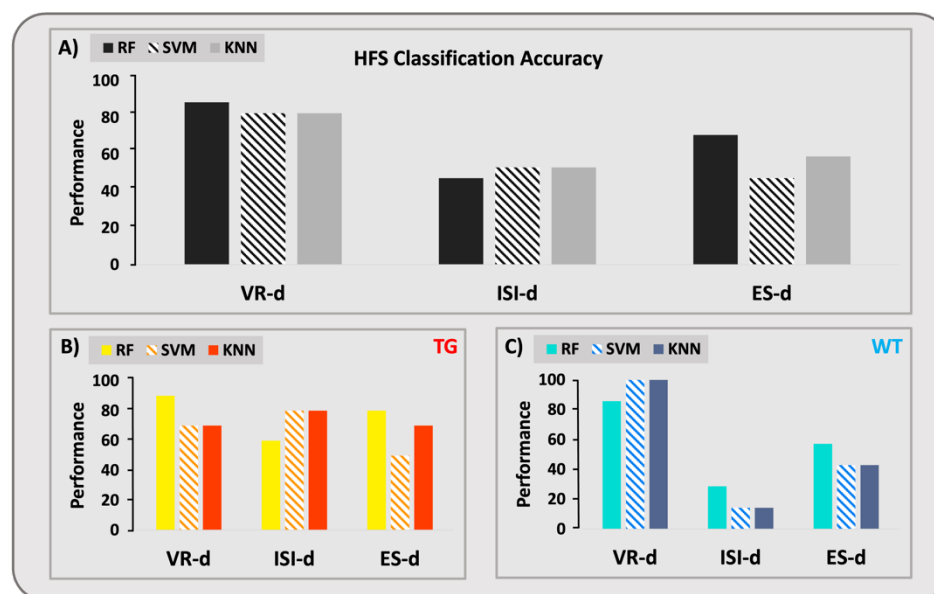


Fig. 10. Classification accuracy for three different classifiers: RF, SVM, k-NN. LOSO-CV was used for training and testing individual classifiers. (A) the total performance for the classification of TG and WT samples is presented. The classification performance for (B) TG rats and (C) for WT rats separately.

Concatenation of HFS and LFS Analysis

As we reported in the previous sections, we have used two types of analysis to extract features from low and high-frequency signals respectively. Because each analysis extracts features with distinct characteristics, the LFS and HFS models do not share information. Therefore, it is possible to combine LFS and HFS information by stacking these models. The simplest form of stacking can be described as an ensemble learning technique where the predictions of multiple classifiers (referred as level-one classifiers) are used as new features to train a meta-classifier⁷⁰. Fig. 11A shows how LFS and HSF models have been stacked as level-one classifiers and logistic regression as meta-classifier. The predictions of level-one models (P_1 , P_2) are stacked and used as input to train the final prediction's meta-classifier. To stack models, the best model of HFS (RF model using VR-d features) and the best model of LFS (RF model) were chosen. The meta-model is often a simple model, providing a smooth interpretation of the predictions made by the base models. As a result, linear models are often used as the meta-model, such as linear regression for regression tasks (predicting a numeric value) and logistic regression for classification tasks (predicting a class label). Although this is common, it is not required. Logistic regression is a statistical technique used to estimate the probability of a discrete

outcome given an input variable. Also, it is best thought of as linear regression modified for classification problems⁷¹. The major distinction between linear and logistic regression is that the range of logistic regression is limited to values between 0 and 1. Table 1 provides the classification performance of the proposed ML method. According to the Classification results, we have achieved 80.23% and 94.44%, respectively, using high and low-frequency features. However, using the combination of high and low-frequency features using stacking models, the accuracy was 100%, and the feature space of two classes was easily separable with a single line (Fig. 11B).

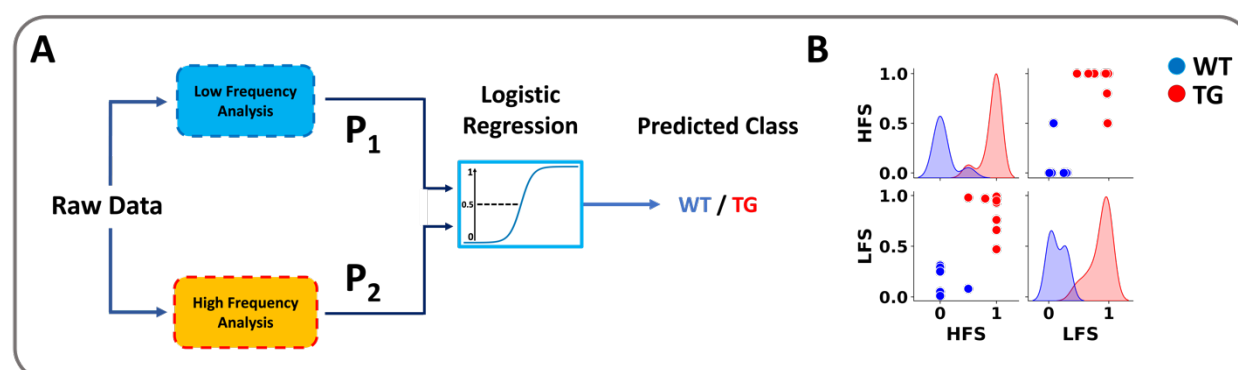


Fig. 11. A) Stacking model pipeline is represented. P_1 is the probabilistic prediction of the LFS model and P_2 is the average prediction of each mouse's spike train using the HFS model. Next, P_1 and P_2 are used as input of meta-model (logistic regression). B) The feature space of meta-model for classification. The red and blue dots represent TG and WT rats respectively.

Table 1. Comparison of the classification models

Model	TG	WT	Total
LFS (RF)	%90.9	%100	%94.44
HFS (RF - VR-d)	%90	%85.71	%80.23
Stacked	%100	%100	%100

Discussion

Alzheimer's disease is characterized by the progressive accumulation of amyloid- β plaques and neurofibrillary tau tangles in the brain and is gradually leading to cognitive decline through multiple stages starting with mild memory loss, increasing deficits in daily life activities, loss of executive functions, and eventually dementia^{72,73}. To date, no therapies exist able to stop or decelerate this devastating condition, while diagnosis is possible only years after the onset of pathophysiological events, presenting a major challenge towards developing successful therapeutic approaches. Thus, current efforts are focusing in identifying early disease biomarkers able to provide prognostic features that will allow patient identification at early time points.

In this study, we recorded electrophysiological signals (LFPs and neuronal spiking activity) from the hippocampus of freely moving TgF344-AD rats and WT littermates, at an early phase of disease progression. Previous studies that characterized in detail the pathophysiology of this animal model, have indicated its great potential over other murine

models, because it gradually expresses all hallmarks of AD in humans including the rarely observed (in murine models) spontaneous accumulation of endogenous hyperphosphorylated tau (pTau) and neurodegeneration^{74,75}. Very importantly, the relatively slow disease progression dynamics in this model that simulate different stages of AD in humans, offer remarkable opportunities for identifying novel AD biomarkers, understanding the underlying mechanisms of pathology and cognitive decline, and testing new therapeutic approaches. At the period we performed our recordings (6 months old rats), only mild A β accumulation is observed in the cortex and hippocampus, cognitive deficits are by and large absent, and thus this time-point is accepted to simulate the early pre-symptomatic phase of AD in humans. We hypothesized, that subtle changes in electrophysiological activity in the hippocampus, could be potentially used to classify individual rats across the two genotypes, AD and WT.

The hippocampus, an area amongst the first to be affected in Alzheimer's patients and strongly implicated in learning and memory processes, has been consistently shown to exhibit abnormal morphology and function in AD patients, with a remarkable reduction in volume as well as aberrant activation during cognitive tasks^{68,69,76,77}. Consistent with these clinical observations, pre-clinical and translational research in transgenic animal models of AD pathology have repeatedly shown hippocampal dysfunctions and in particular abnormalities in oscillatory brain rhythms. Accumulating evidence suggests that disturbances in synaptic and neuronal function are leading to impairments of coordinated activity at early pre-symptomatic stages of the disease with major targets the brain networks that support memory and cognition^{73,78}. Aberrations in the activity of these brain circuits range from subtle changes in oscillatory rhythms to more profound alterations that are reflected in EEG signals at later stages of the disease. Such impairments in brain rhythms, considered to underlie compromised hippocampal-dependent cognitive function, were also shown in pre-clinical animal models of AD before the onset of behavioral deficits⁷⁹⁻⁸¹.

To identify potential early aberrations in electrophysiological signals in the hippocampus of our TgF344-AD rat model, we first performed analysis of the low frequency signals (LFPs). Two different frequency ranges of gamma oscillations thought to be generated in the hippocampus were analyzed: slow gamma (25–50 Hz) and fast gamma (50–100 Hz)^{24,82–84}. We found that both the slow as well as the fast gamma oscillations were reduced in the AD rats in comparison to WTs (Fig. 4). This is consistent with findings in which gamma oscillations were found to be disrupted in AD patients³⁸, in mice^{39,40}, but also in some previous experiments in the same transgenic rat model, albeit performed at a slightly later stage of disease progression (8-9 months of age)^{65,66}. Noteworthy, those studies reported significant decreases in the slow gamma power, while in our study – performed at an earlier AD stage - we, in addition, observed strong decreases in the fast gamma oscillations. We speculate that this difference could potentially relate to compensation processes that get activated at these slightly later stages and are resulting in cognitive resilience that is persistent until much later time points^{85,86}. On the other hand, our analysis of HVSs indicates no significant difference between the HVSs occurrence rate between two groups. Earlier studies have shown that HVSs would increase during the AD in the cortical region of TgF344-AD rats^{68,69}. These spindles can reach the hippocampus by the neocortical-entorhinal cortex path⁸⁷. Therefore, our assumption would be although these spindles have become widespread in the cortical region in the previous studies, they still have not reached the hippocampus in the earlier stages of AD.

In the hippocampus, gamma oscillations are nested within the slower theta oscillations with this co-modulation being proposed as a mechanism that recruits cell assemblies throughout the theta cycle⁸⁸. The coupling between the phase of the theta with the amplitude of gamma oscillations has been demonstrated to support cognitive processes and memory formation in humans³¹, monkeys²⁹, and rodents^{32,83,89}. To probe potential changes in these coupling in the TgF344-AD rat model we performed analysis of cross-frequency coupling (Fig. 5). We found a significant reduction of the modulation index in the theta phase with slow gamma but not the fast gamma amplitude in the AD rats. Low-frequency gamma is associated with gamma locking and signal transfer in the hippocampus from CA3 → CA1 (Fig. 2A)²⁴. Previous research in TgF344-AD has found a reduced fidelity in the activity of place cells recorded from the hippocampus CA2/3 while place cells recorded from CA1 showed persistent fidelity until much later stages of the disease⁹⁰. Thus, the reduced theta-phase slow gamma amplitude coupling we observe could stem from such early aberrations in the CA3 circuit. On the other hand, the absence of changes in theta-phase to fast gamma amplitude coupling, which relates to the signal transfer from MEC → CA1, indicates that this part of the circuit is able to function in a relatively normal fashion in the earlier stages of AD, although some synaptic deficits were found *in vitro*⁹¹.

Using the information and insights we obtained from the frequency band power analysis, we performed classification using combinations of the power in different frequency bands in our effort to understand the contribution of each band on classification accuracy. As expected from visualizing the pairwise combinations of the PSD values for different frequency bands (Fig. 3), our results indicated that the best accuracy is achieved when average power in the gamma band is one of the classifier's inputs while if only theta and delta band information is used, classification is close to chance levels. Comparing the performance of different ML classifiers, we observed that their performance was similar on average, with the random forest achieving the highest total classification (Fig. 7). The importance of gamma band activity in classification across the groups is in line with previous findings^{39,40,63–65,81,92} suggesting that gamma oscillations are essential in spatial memory and get disrupted during early stages of AD. Moreover, this finding provides support to suggestions that gamma frequency entrainment or reactivation have positive effects and can be developed towards potential treatments for AD^{93,94}.

Having established that low frequency electrophysiology signals carry important information for classification, we then performed analysis of the high frequency signals by performing spike sorting to extract the spike trains of activity from different cells. Then, we calculated three spike-train distance metrics (VR-d, ISI-d, ES-d) across pairs of units in the TG and WT animals and used this as features for classification of the genotype. We found that VR-d provided the highest classification performance with over 80% total accuracy while the ISI-d and ES-d performance was on average close to chance levels. Interestingly, the ISI-d performance was dissociated between the two genotype groups showing high classification performance in TG animals but low in WT. In our experiments, we have recorded data from the hippocampus CA1 and CA3 regions^{95–97} while the rats were freely moving in a linear track. We thus conjecture that aberrations in spike-trains driving the classification are related at least in part to place cell activity. Many recent studies have shown disruption of place cells activity in AD^{39,40,90}. A study by Galloway and colleagues, explicitly studied place cell activity in the hippocampus of TgF344-AD rats between 12 and 20 months of age⁹⁰. They found age related reductions in the fidelity of place cell firing in CA2/3 but not CA1. These subfields of hippocampus are part of the well-known trisynaptic loop which includes the dentate gyrus, CA3 and CA, with each of these

subserving a different function⁹⁸. The pyramidal neurons of CA3 exhibit burst activity and project to the CA1 pyramidal neurons via the Schaffer Collaterals. Noteworthy, the synapses of these CA3 neurons on CA1 dendrites are highly modifiable and very dynamic⁹⁹. In addition, approximately one third of synaptic connections in CA1 are received directly from the layer III of entorhinal cortex via the perforant pathway in contrast to CA2 and CA3 neurons, which receive direct projections from entorhinal cortex layer II^{100,101}. Given that previous studies have already shown that CA1 place cells are able to demonstrate normal place fields in the absence of CA3 input, albeit not without inputs from the layer III of the entorhinal cortex^{102,103}, it would be perhaps possible to normalize hippocampal activity in AD by stimulating these layer III projections from the entorhinal cortex to CA1.

To further understand if the aberrations in hippocampal spiking activity are providing the same (or additional) information for classification as the low frequency signals, we decided to perform a concatenation of the HFS and LFS analyses into a combined ML model. We found that stacking information from the two analyses, boosted the classification to 100% accuracy. This could indicate that although a lot of information in these frequency ranges may be shared (e.g., due to the phase locking of spiking to the theta and/or gamma rhythms), additional independent information can help the accurate classification. Thus, our analysis is consistent with previous findings indicating that that connectivity between CA3 and CA1 subfields of the hippocampus is disrupted due to aberrations in CA3 activity while CA1 seems resilient.

In conclusion, the results of this paper indicate that with a combination of ML models, feature extraction methods based on similarities and distances in the neural responses (high-frequency components of the neuronal data) and time-frequency analysis of LFP signal (low-frequency components of the neuronal data) recorded from the rat's hippocampus, we can classify healthy and AD rats at an early, pre-symptomatic stage of Alzheimer's disease. These promising findings provide a clear indication that information from a combination of signals and electrophysiological measurements, could accelerate diagnosis of AD and assist scientists in the development of more effective therapies.

Acknowledgments

This study was supported by the Fund for Scientific Research Flanders (FWO) (grant agreement G048917N to GAK) (G045420N to MV) and the Stichting Alzheimer Onderzoek (SAO-FRA-2018003 to MV).

Author Contributions

M.v.B and G.A.K designed the electrophysiology experiments and G.A.K supervised the data acquisition. G.A.K. and M.A designed the simulations and M.A supervised the data analysis. F.M and M.A performed the high-frequency analysis. M.M, F.M, and M.A performed the low-frequency analysis. M.v.B performed the electrophysiology experiments and data acquisition in freely moving rats with the assistance of L.K. The initial draft of the manuscript was written by F.M, M.M, M.A, G.A.K. All authors edited the manuscript and accepted the final version.

References

1. Alzheimer's Association. 2015 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* **11**, 332–384 (2015).
2. Lane, C. A., Hardy, J. & Schott, J. M. Alzheimer's disease. *European Journal of Neurology* **25**, 59–70 (2018).
3. Hope, T. *et al.* The structure of wandering in dementia. *International Journal of Geriatric Psychiatry* **9**, 149–155 (1994).
4. Albert, M. S. *et al.* The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia* **7**, 270–279 (2011).
5. Amaral, D. G. & Witter, M. P. The three-dimensional organization of the hippocampal formation: A review of anatomical data. *Neuroscience* **31**, 571–591 (1989).
6. GyörgyBuzsáki. Theta Oscillations in the Hippocampus. *Neuron* **33**, 325–340 (2002).
7. Vertes, R. P. Hippocampal theta rhythm: A tag for short-term memory. *Hippocampus* vol. 15 (2005).
8. Lisman, J. & Redish, A. D. Prediction, sequences and the hippocampus. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**, (2009).
9. Lega, B. C., Jacobs, J. & Kahana, M. Human hippocampal theta oscillations and the formation of episodic memories. *Hippocampus* **22**, (2012).
10. Buzsáki, G., Anastassiou, C. A. & Koch, C. The origin of extracellular fields and currents-EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience* vol. 13 (2012).
11. Singer, W. & Gray, C. M. Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience* vol. 18 (1995).
12. Fries, P., Reynolds, J. H., Rorie, A. E. & Desimone, R. Modulation of oscillatory neuronal synchronization by selective visual attention. *Science (1979)* **291**, (2001).
13. Börgers, C., Epstein, S. & Kopell, N. J. Background gamma rhythmicity and attention in cortical local circuits: A computational study. *Proc Natl Acad Sci U S A* **102**, (2005).
14. Jensen, O., Kaiser, J. & Lachaux, J.-P. Human gamma-frequency oscillations associated with attention and memory. *Trends in Neurosciences* **30**, 317–324 (2007).
15. Harris, K. D., Csicsvari, J., Hirase, H., Dragoi, G. & Buzsáki, G. Organization of cell assemblies in the hippocampus. *Nature* **424**, (2003).
16. Montgomery, S. M. & Buzsáki, G. Gamma oscillations dynamically couple hippocampal CA3 and CA1 regions during memory task performance. *Proc Natl Acad Sci U S A* **104**, (2007).
17. Driver, J. E. *et al.* Impairment of hippocampal gamma (γ)-frequency oscillations in vitro in mice overexpressing human amyloid precursor protein (APP). *European Journal of Neuroscience* **26**, (2007).
18. Lisman, J. The theta/gamma discrete phase code occurring during the hippocampal phase precession may be a more general brain coding scheme. *Hippocampus* **15**, (2005).
19. Berry, S. D. & Thompson, R. F. Prediction of learning rate from the hippocampal electroencephalogram. *Science (1979)* **200**, (1978).
20. Winson, J. Loss of hippocampal theta rhythm results in spatial memory deficit in the rat. *Science (1979)* **201**, (1978).

21. Sederberg, P. B., Kahana, M. J., Howard, M. W., Donner, E. J. & Madsen, J. R. Theta and Gamma Oscillations during Encoding Predict Subsequent Recall. *Journal of Neuroscience* **23**, (2003).
22. Sederberg, P. B. *et al.* Hippocampal and Neocortical Gamma Oscillations Predict Memory Formation in Humans. *Cerebral Cortex* **17**, 1190–1196 (2006).
23. Lisman, J. & Buzsáki, G. A neural coding scheme formed by the combined function of gamma and theta oscillations. *Schizophrenia Bulletin* vol. 34 (2008).
24. Colgin, L. L. *et al.* Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature* **462**, 353–357 (2009).
25. Robinson, D. L., Goldberg, M. E. & Stanton, G. B. Parietal association cortex in the primate: sensory mechanisms and behavioral modulations. *Journal of Neurophysiology* **41**, 910–932 (1978).
26. Nicolelis, M. A. L., Baccala, L. A., Lin, R. C. S. & Chapin, J. K. Sensorimotor Encoding by Synchronous Neural Ensemble Activity at Multiple Levels of the Somatosensory System. *Science* (1979) **268**, 1353–1358 (1995).
27. Deransart, C. *et al.* Single-unit Analysis of Substantia Nigra Pars Reticulata Neurons in Freely Behaving Rats with Genetic Absence Epilepsy. *Epilepsia* **44**, 1513–1520 (2003).
28. Slaght, S. J. On the Activity of the Corticostriatal Networks during Spike-and-Wave Discharges in a Genetic Model of Absence Epilepsy. *Journal of Neuroscience* **24**, 6816–6825 (2004).
29. Canolty, R. T. & Knight, R. T. The functional role of cross-frequency coupling. *Trends in Cognitive Sciences* vol. 14 (2010).
30. Händel, B. & Haarmeier, T. Cross-frequency coupling of brain oscillations indicates the success in visual motion discrimination. *Neuroimage* **45**, (2009).
31. Axmacher, N. *et al.* Cross-frequency coupling supports multi-item working memory in the human hippocampus. *Proc Natl Acad Sci U S A* **107**, (2010).
32. Tort, A. B. L., Komorowski, R. W., Manns, J. R., Kopell, N. J. & Eichenbaum, H. Theta-gamma coupling increases during the learning of item-context associations. *Proc Natl Acad Sci U S A* **106**, (2009).
33. Gregoriou, G. G., Gotts, S. J., Zhou, H. & Desimone, R. High-Frequency, long-range coupling between prefrontal and visual cortex during attention. *Science* (1979) **324**, (2009).
34. Siegel, M., Warden, M. R. & Miller, E. K. Phase-dependent neuronal coding of objects in short-term memory. *Proc Natl Acad Sci U S A* **106**, (2009).
35. van Deursen, J. A., Vuurman, E. F. P. M., Verhey, F. R. J., van Kranen-Mastenbroek, V. H. J. M. & Riedel, W. J. Increased EEG gamma band activity in Alzheimer's disease and mild cognitive impairment. *Journal of Neural Transmission* **115**, (2008).
36. Czigler, B. *et al.* Quantitative EEG in early Alzheimer's disease patients - Power spectrum and complexity features. *International Journal of Psychophysiology* **68**, (2008).
37. Moretti, V. D. *et al.* EEG markers discriminate among different subgroup of patients with mild cognitive impairment. *American Journal of Alzheimer's Disease and other Dementias* **25**, (2010).
38. Herrmann, C. S. & Demiralp, T. Human EEG gamma oscillations in neuropsychiatric disorders. *Clinical Neurophysiology* vol. 116 (2005).
39. Jun, H. *et al.* Disrupted Place Cell Remapping and Impaired Grid Cells in a Knockin Model of Alzheimer's Disease. *Neuron* **107**, 1095-1112.e6 (2020).

40. Nakazono, T. *et al.* Impaired In Vivo Gamma Oscillations in the Medial Entorhinal Cortex of Knock-in Alzheimer Model. *Frontiers in Systems Neuroscience* **11**, (2017).
41. Mably, A. J., Gereke, B. J., Jones, D. T. & Colgin, L. L. Impairments in spatial representations and rhythmic coordination of place cells in the 3xTg mouse model of Alzheimer's disease. *Hippocampus* **27**, 378–392 (2017).
42. Rieke F & Warland D. Spikes: exploring the neural code. *MIT press*.
43. Bear M, Connors B & Paradiso MA. *Neuroscience: Exploring the Brain, Enhanced Edition: Exploring the Brain*. (2020).
44. Mainen, Z. F. & Sejnowski, T. J. Reliability of Spike Timing in Neocortical Neurons. *Science (1979)* **268**, 1503–1506 (1995).
45. Reyes, A. D. Synchrony-dependent propagation of firing rate in iteratively constructed networks in vitro. *Nature Neuroscience* **6**, 593–599 (2003).
46. Yavari, F., Amiri, M., Rahatabad, F. N., Falotico, E. & Laschi, C. Spike train analysis in a digital neuromorphic system of cutaneous mechanoreceptor. *Neurocomputing* **379**, 343–355 (2020).
47. Lee, G., Nho, K., Kang, B., Sohn, K.-A. & Kim, D. Predicting Alzheimer's disease progression using multi-modal deep learning approach. *Scientific Reports* **9**, 1952 (2019).
48. El-Sappagh, S., Alonso, J. M., Islam, S. M. R., Sultan, A. M. & Kwak, K. S. A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease. *Scientific Reports* **11**, 2660 (2021).
49. Venugopalan, J., Tong, L., Hassanzadeh, H. R. & Wang, M. D. Multimodal deep learning models for early detection of Alzheimer's disease stage. *Scientific Reports* **11**, 3254 (2021).
50. <http://chronux.org>.
51. <https://github.com/buzsakilab/buzcode>.
52. O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research* **34**, 171–175 (1971).
53. INSAUSTI, R. & AMARAL, D. G. Hippocampal Formation. in *The Human Nervous System* 871–914 (Elsevier, 2004). doi:10.1016/B978-012547626-3/50024-7.
54. WITTER, M. P., WOUTERLOOD, F. G., NABER, P. A. & van HAEFTEN, T. Anatomical Organization of the Parahippocampal-Hippocampal Network. *Ann N Y Acad Sci* **911**, 1–24 (2006).
55. <https://github.com/tridesclous/tridesclous>.
56. Buccino, A. P. *et al.* SpikeInterface, a unified framework for spike sorting. *Elife* **9**, (2020).
57. Rossum, M. C. W. van. A Novel Spike Distance. *Neural Computation* **13**, 751–763 (2001).
58. Satuvuori, E. & Kreuz, T. Which spike train distance is most suitable for distinguishing rate and temporal coding? *Journal of Neuroscience Methods* **299**, 22–33 (2018).
59. Esteva, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).
60. Weng, S., Xu, X., Li, J. & Wong, S. T. C. Combining deep learning and coherent anti-Stokes Raman scattering imaging for automated differential diagnosis of lung cancer. *Journal of Biomedical Optics* **22**, 1 (2017).
61. Suk, H.-I., Lee, S.-W. & Shen, D. Deep sparse multi-task learning for feature selection in Alzheimer's disease diagnosis. *Brain Structure and Function* **221**, 2569–2587 (2016).

62. Liu, S. *et al.* Multimodal Neuroimaging Feature Learning for Multiclass Diagnosis of Alzheimer's Disease. *IEEE Transactions on Biomedical Engineering* **62**, 1132–1140 (2015).
63. Morrone, C. D. *et al.* Regional differences in Alzheimer's disease pathology confound behavioural rescue after amyloid- β attenuation. *Brain* **143**, (2020).
64. Goutagny, R. *et al.* Alterations in hippocampal network oscillations and theta-gamma coupling arise before A β overproduction in a mouse model of Alzheimer's disease. *European Journal of Neuroscience* **37**, (2013).
65. Bazzigaluppi, P. *et al.* Early-stage attenuation of phase-amplitude coupling in the hippocampus and medial prefrontal cortex in a transgenic rat model of Alzheimer's disease. *Journal of Neurochemistry* **144**, (2018).
66. Joo, I. L. *et al.* Early neurovascular dysfunction in a transgenic rat model of Alzheimer's disease. *Scientific Reports* **7**, (2017).
67. Goldberg, J. A., Rokni, U., Boraud, T., Vaadia, E. & Bergman, H. Spike synchronization in the cortex-basal ganglia networks of parkinsonian primates reflects global dynamics of the local field potentials. *Journal of Neuroscience* **24**, (2004).
68. Stoiljkovic, M., Kelley, C., Stutz, B., Horvath, T. L. & Hajós, M. Altered Cortical and Hippocampal Excitability in TgF344-AD Rats Modeling Alzheimer's Disease Pathology. *Cerebral Cortex* **29**, (2019).
69. Stoiljkovic, M., Kelley, C., Horvath, T. L. & Hajós, M. Neurophysiological signals as predictive translational biomarkers for Alzheimer's disease treatment: Effects of donepezil on neuronal network oscillations in TgF344-AD rats. *Alzheimer's Research and Therapy* **10**, (2018).
70. Sagi, O. & Rokach, L. Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* vol. 8 (2018).
71. Tolles, J. & Meurer, W. J. Logistic regression: Relating patient characteristics to outcomes. *JAMA - Journal of the American Medical Association* vol. 316 (2016).
72. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica* vol. 82 (1991).
73. Selkoe, D. J. & Hardy, J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Molecular Medicine* **8**, (2016).
74. Drummond, E. & Wisniewski, T. Alzheimer's disease: experimental models and reality. *Acta Neuropathologica* vol. 133 (2017).
75. Webster, S. J., Bachstetter, A. D., Nelson, P. T., Schmitt, F. A. & van Eldik, L. J. Using mice to model Alzheimer's dementia: An overview of the clinical disease and the preclinical behavioral changes in 10 mouse models. *Frontiers in Genetics* **5**, (2014).
76. Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P. & Thompson, P. M. The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology* vol. 6 (2010).
77. Han, S. D. *et al.* Beta amyloid, tau, neuroimaging, and cognition: Sequence modeling of biomarkers for Alzheimer's Disease. *Brain Imaging and Behavior* **6**, (2012).
78. Palop, J. J. & Mucke, L. Amyloid-B-induced neuronal dysfunction in Alzheimer's disease: From synapses toward neural networks. *Nature Neuroscience* vol. 13 (2010).
79. Howlett, D. R. *et al.* Cognitive correlates of A β deposition in male and female mice bearing amyloid precursor protein and presenilin-1 mutant transgenes. *Brain Research* **1017**, (2004).

80. Villette, V. *et al.* Decreased rhythmic GABAergic septal activity and memory-associated θ oscillations after hippocampal amyloid- β pathology in the rat. *Journal of Neuroscience* **30**, (2010).
81. Scott, L. *et al.* Age-dependent disruption in hippocampal theta oscillation in amyloid- β overproducing transgenic mice. *Neurobiology of Aging* **33**, (2012).
82. Canolty, R. T. *et al.* High Gamma Power Is Phase-Locked to Theta Oscillations in Human Neocortex. *Science* (1979) **313**, 1626–1628 (2006).
83. Belluscio, M. A., Mizuseki, K., Schmidt, R., Kempter, R. & Buzsáki, G. Cross-frequency phase-phase coupling between theta and gamma oscillations in the hippocampus. *Journal of Neuroscience* **32**, (2012).
84. Scheffer-Teixeira, R., Belchior, H., Leão, R. N., Ribeiro, S. & Tort, A. B. L. On high-frequency field oscillations (>100 Hz) and the spectral leakage of spiking activity. *Journal of Neuroscience* **33**, (2013).
85. Goodman, A. M., Langner, B. M., Jackson, N., Alex, C. & McMahon, L. L. Heightened hippocampal β -adrenergic receptor function drives synaptic potentiation and supports learning and memory in the TgF344-AD rat model during prodromal Alzheimer's disease. *Journal of Neuroscience* **41**, (2021).
86. Morrone, C. D., Lai, A. Y., Bishay, J., Hill, M. E. & McLaurin, J. Parvalbumin neuroplasticity compensates for somatostatin impairment, maintaining cognitive function in Alzheimer's disease. *Translational Neurodegeneration* **11**, 26 (2022).
87. Rolls, E. T. Hippocampo-cortical and cortico-cortical backprojections. *Hippocampus* **10**, (2000).
88. Jackson, J., Goutagny, R. & Williams, S. Fast and slow gamma rhythms are intrinsically and independently generated in the subiculum. *Journal of Neuroscience* **31**, (2011).
89. Wulff, P. *et al.* Hippocampal theta rhythm and its coupling with gamma oscillations require fast inhibition onto parvalbumin-positive interneurons. *Proc Natl Acad Sci U S A* **106**, (2009).
90. Galloway, C. R. *et al.* Hippocampal place cell dysfunction and the effects of muscarinic M1 receptor agonism in a rat model of Alzheimer's disease. *Hippocampus* **28**, (2018).
91. Smith, L. A. & McMahon, L. L. Deficits in synaptic function occur at medial perforant path-dentate granule cell synapses prior to Schaffer collateral-CA1 pyramidal cell synapses in the novel TgF344-Alzheimer's Disease Rat Model. *Neurobiology of Disease* **110**, 166–179 (2018).
92. Zhang, X. *et al.* Impaired theta-gamma coupling in APP-deficient mice. *Scientific Reports* **6**, (2016).
93. Iaccarino, H. F. *et al.* Gamma frequency entrainment attenuates amyloid load and modifies microglia. *Nature* **540**, 230–235 (2016).
94. Mably, A. J., Gereke, B. J., Jones, D. T. & Colgin, L. L. Impairments in spatial representations and rhythmic coordination of place cells in the 3xTg mouse model of Alzheimer's disease. *Hippocampus* **27**, 378–392 (2017).
95. Shapiro, M. Plasticity, hippocampal place cells, and cognitive maps. *Archives of Neurology* **58**, (2001).
96. O'Keefe, J. An allocentric spatial model for the hippocampal cognitive map. *Hippocampus* **1**, (1991).
97. O'Keefe, J. Chapter 22 A computational theory of the hippocampal cognitive map. *Progress in Brain Research* **83**, (1990).

98. Stepan, J., Dine, J. & Eder, M. Functional optical probing of the hippocampal trisynaptic circuit in vitro: Network dynamics, filter properties, and polysynaptic induction of CA1 LTP. *Frontiers in Neuroscience* vol. 9 (2015).
99. Amaral, D. G., Ishizuka, N. & Claiborne, B. Chapter Neurons, numbers and the hippocampal network. *Progress in Brain Research* **83**, (1990).
100. Llorens-Martín, M. *et al.* Selective alterations of neurons and circuits related to early memory loss in Alzheimer's disease. *Frontiers in Neuroanatomy* **8**, (2014).
101. Steward, O. & Scoville, S. A. Cells of origin of entorhinal cortical afferents to the hippocampus and fascia dentata of the rat. *Journal of Comparative Neurology* **169**, (1976).
102. Brun, V. H. *et al.* Impaired Spatial Representation in CA1 after Lesion of Direct Input from Entorhinal Cortex. *Neuron* **57**, (2008).
103. Brun, V. H. *et al.* Place cells and place recognition maintained by direct entorhinal-hippocampal circuitry. *Science (1979)* **296**, (2002).
104. Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001).

Appendix 1

The algorithm for spike sorting and multi-unit classification model

Input: LFP signals from each rat

- **C**: Number of electrode channels
- **Pnts**: Number of time points or the during of signal
- **X_s**: Input data to spike sorter
- **f**: Band pass filter # shape: (**Pnts**, **C**)
- **N**: Number of multi-units found by spike sorters
- **N_s**: Number of spikes in each founded multi-unit
- **P**: Number of PCA components
- **label_{actual}** ∈ {0, 1}: The class label of each multi-unit (TG or WT)
- **label_{pred prob}** ∈ [0 1]: The probability of predicted class for each sample

Output:

- **label_{pred}** ∈ {0, 1}: The predicted class for each multi-unit

Operators:

- \otimes : Band-pass filter operator
- \oplus : Common reference removal
- \ominus : Down sample
- \odot : Spike sorter (Tridesclous)
- \parallel : Concatenation
- **F**: Spike train distance metrics (VR, ISI, and ES distance)
- **A**: Mean PSD of the delta, theta, slow gamma and fast gamma
- **p**: Principle component analysis with **P** components
- **M**: ML classifiers (RF, SVM, and k-NN)

HFS Procedure:

Preprocessing

$$\begin{aligned} X &= f \otimes X && \# \text{ shape: } (Pnts, C) \\ X &= \oplus (X) && \# \text{ shape: } (Pnts, C) \end{aligned}$$

Spike Sorting

$$\begin{aligned} N &= \odot(X) && \text{Number of multi-units extracted from each rat LFP Data} \\ N_{TG}, N_{WT} &&& \text{Number of multi-units for each group of WT and TG} \end{aligned}$$

Feature Extraction

$$\begin{aligned} N_{TG} \times N_{TG}, N_{WT} \times N_{WT} &= F(N_{TG}, N_{WT}) && \text{Spike train feature distance metrics} \\ N_{TG} \times 3, N_{WT} \times 3 &= p(N_{TG}, N_{WT}) && \text{Principle component analysis} \end{aligned}$$

ML Classifiers

$$label_{pred} = M(N_{TG} \times 3, N_{WT} \times 3, label_{actual})$$

Classification using ML models and leave one out cross validation

LFS Procedure:

Preprocessing

$$\mathbf{X} = \mathbf{f} \circledast \mathbf{X} \quad \# \text{ shape: } (\mathbf{Pnts}, \mathbf{C})$$

$$\mathbf{X} = \ominus (\mathbf{X}) \quad \# \text{ shape: } (\mathbf{Pnts}, \mathbf{C})$$

Feature Extraction

$$\mathbf{N}_{TG} \times 4, \mathbf{N}_{WT} \times 4 = \mathbf{A}(\mathbf{N}_{TG}, \mathbf{N}_{WT}) \quad \text{Extraction of average frequency band of each rat}$$

ML Classifiers

$$\mathbf{label}_{pred} = \mathcal{M}(\mathbf{N}_{TG} \times 4, \mathbf{N}_{WT} \times 4, \mathbf{label}_{actual}) \quad \text{Classification using ML models}$$

Stacking Procedure:

Level one Classifiers

$$\mathbf{label}_{pred\ prob\ HFS} = \mathcal{M}_{HFS}(\mathbf{N}_{TG} \times 3, \mathbf{N}_{WT} \times 3, \mathbf{label}_{actual}) \quad \text{Classification using ML models}$$

$$\mathbf{label}_{pred\ prob\ LFS} = \mathcal{M}_{LFS}(\mathbf{N}_{TG} \times 4, \mathbf{N}_{WT} \times 4, \mathbf{label}_{actual}) \quad \text{Classification using ML models}$$

Concatenation

$$\mathbf{X} = \mathbf{label}_{pred\ prob\ HFS} \parallel \mathbf{label}_{pred\ prob\ LFS} \quad \# \text{ shape: } (2, \mathbf{C})$$

Level two Classifier

$$\mathbf{label}_{pred} = \mathcal{M}(\mathbf{X}, \mathbf{label}_{actual}) \quad \text{Classification using ML models}$$

Appendix 2

The mathematical description of individual metrics.

Van Rossum distance (VRd):

The Van Rossum metric estimates the distance of two spike trains and has been proved to be a reliable metric as the distance between multiple responses to the same stimulus is usually smaller than the distance between responses to the other stimuli. The discrete spike trains x and y are transformed into continuous functions through convolving each spike t_k with an exponential kernel $H(t) = \exp(-\frac{t}{\tau_R})$. τ_R is the time constant and H is the Heaviside step function. After proper tradeoff, $\tau_R = 0.05$ was selected. From the resulted waveforms $x(t)$ and $y(t)$, the VRd, D_{VR} can be calculated as: [33]

$$D_{RD}(\tau_R) = \frac{1}{\tau_R} \int [x(t) - y(t)]^2 dt$$

Inter spike interval distance (ISI-distance):

The ISI- distance, DI , measures the immediate rate difference between spikes. It is based on a time-resolved profile that defines a dissimilarity value for each instance. We assign t , the time of the previous spike to each moment to obtain this feature:

$$t_p^{(n)}(t) = \text{Max}\{t_i^{(n)} | t_i^{(n)} \leq t\} \text{ for } t_1^{(n)} \leq t \leq t_{M_n}^{(n)}$$

$$t_F^{(n)}(t) = \text{Max}\{t_i^{(n)} | t_i^{(n)} > t\} \text{ for } t_1^{(n)} \leq t \leq t_{M_n}^{(n)}$$

$$X_{ISI}^{(n)}(t) = t_F^{(n)}(t) - t_p^{(n)}(t)$$

$$I_{m,n}(t) = \frac{|X_{ISI}^{(n)}(t) - X_{ISI}^{(m)}(t)|}{\text{Max}\{X_{ISI}^{(n)}(t) | X_{ISI}^{(m)}(t)\}}$$

$$I(t) = \frac{2}{N(N-1)} \sum_{n=1}^{N-1} \sum_{m=N+1}^N I_{n,m}(t)$$

$$D_1 = \frac{1}{t_e - t_s} \int_{t_s}^{t_e} I(t) dt$$

The Event-Synchronization distance (ESd):

The Event-Synchronization distance is based on the event relative timing. In this article, the event is defined by a spike in the spike patterns. The spikes t_i and t_j in the separate spike trains are synchronous only if their time difference is less than the delay time τ_{ij} , which is obtained by:

$$\tau_{ij} = \frac{\min\{t_{i+1} - t_i, t_i - t_{i-1}, t_{j+1} - t_j, t_j - t_{j-1}\}}{2}$$

$$c(m|n) = \sum_{i=1}^{N_m} \sum_{j=1}^{N_n} J_{ij}$$

$$J_{ij} = \begin{cases} 1 & 0 < t_i^m - t_j^n \leq \tau_{ij} \\ 0.5 & t_i^m = t_j^n \\ 0 & \text{otherwise} \end{cases}$$

$$D_{\text{Event-Syn}} = 1 - \frac{c(m|n) + c(n|m)}{\sqrt{N_m N_n}}$$

Mean Decrease Impurity (MDI):

Gini importance measures the importance of variable X in predicting variable Y in the random forest for all nodes averaged over all N_T trees in the forest¹⁰⁴. This measure is also called Mean Decrease Impurity, here denoted by *IncNodePurity*:

$$\text{IncNodePurity}(X_m) = \frac{1}{N_T} \sum_T \sum_{t \in T: v(s_t) = X_m} p(t) \Delta i(s_t, t, t_l, t_r)$$

where $v(s_t)$ is the variable used in split s_t and $\Delta i(s_t, t, t_l, t_r)$ is the impurity decrease of a binary split s_t dividing node t into a left node t_l and a right node t_r :

$$\Delta i(s_t, t, t_l, t_r) = i(t) - \frac{N_{t_l}}{N_t} \cdot i(t_l) - \frac{N_{t_r}}{N_t} \cdot i(t_r)$$

where N is the sample size, $p(t) = \frac{N_t}{N}$ the proportion of samples reaching t , and $p(t_l) = \frac{N_{t_l}}{N}$ and $p(t_r) = \frac{N_{t_r}}{N}$ are the proportion of samples reaching the left node t_l and the right node t_r respectively.

Supplemental Information

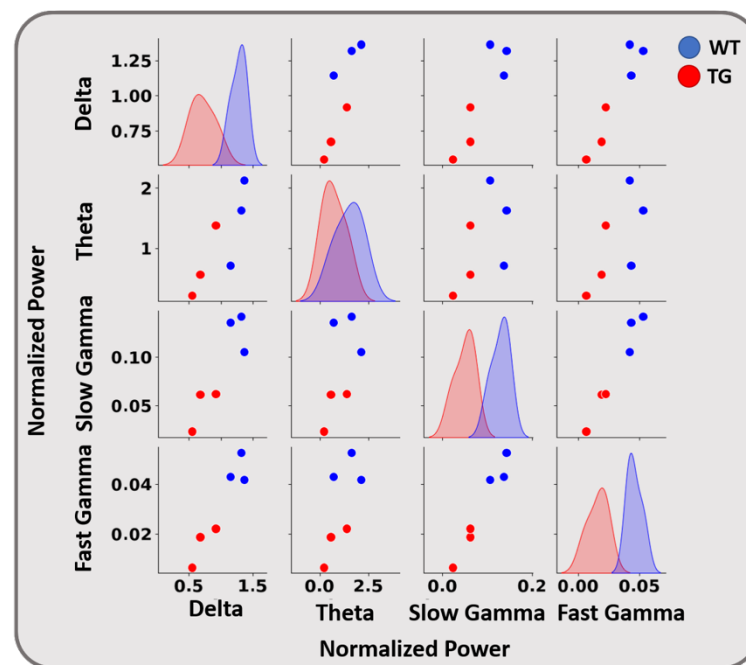


Figure S1. Pairwise comparison of the normalized power in delta (1-4 Hz), theta (4-12 Hz), slow gamma (30-50 Hz), and fast gamma (50-100 Hz) for TgF344-AD (TG) and Wildtype (WT) groups. Each point represents the average PSD of a single sample in the specified band. The diagonal plots are the univariate distributions that show the amount of overlap between the distribution of two groups in individual frequency bands. The other plots compare WT and TG for specified frequency bands. Blue/Red points are recording session (samples) of WT (n = 3) /TG (n = 3), respectively (TG1, TG2, TG3, WT4, WT5, WT6).

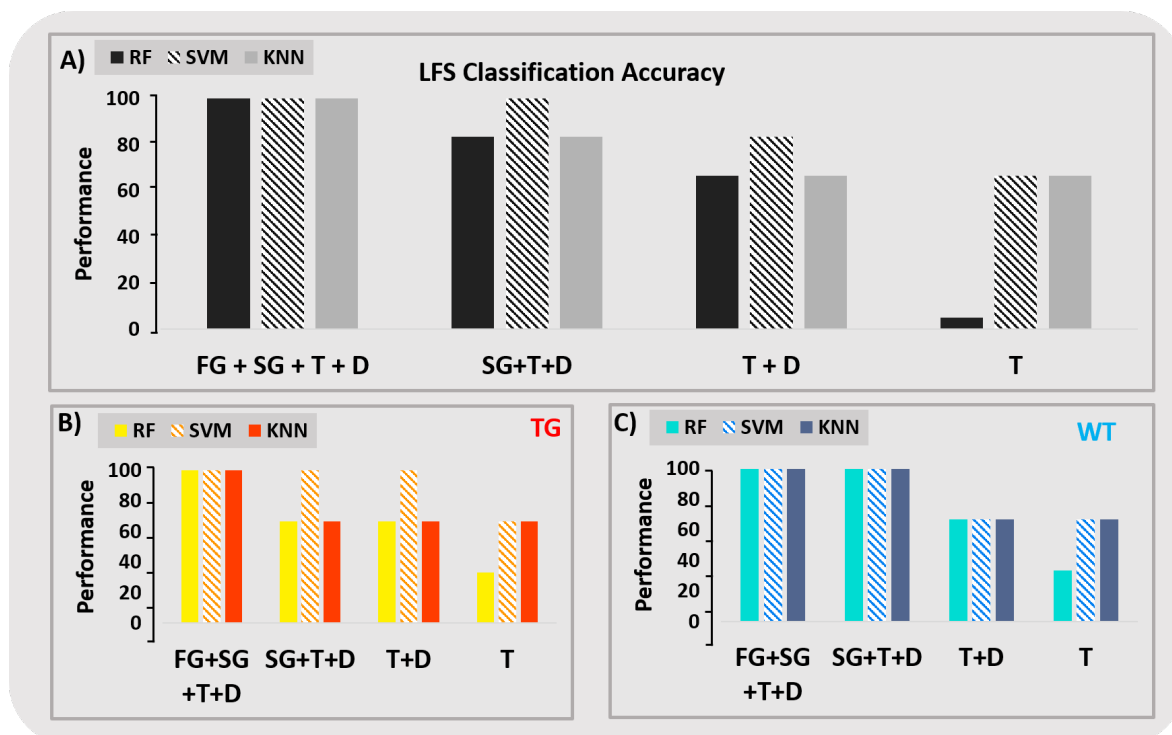


Figure S2. Classification accuracy of LFS model with three ML classifiers: k-nearest neighbors (k-NN), Support vector machine (SVM), and Random Forest (RF) for 6 rats from both groups of TgF344-AD and WT (TG1, TG2, TG3, WT4, WT5, WT6). Frequency bands: Fast Gamma (FG), Slow Gamma (SG), Theta (T), and Delta (D). (A) the overall accuracy. (B) the classification accuracy only for TG rats. (C) the classification accuracy only for WT rats. From left to right in each section, the frequency bands are removed from the classifier's input one by one to see their effects on the accuracy.

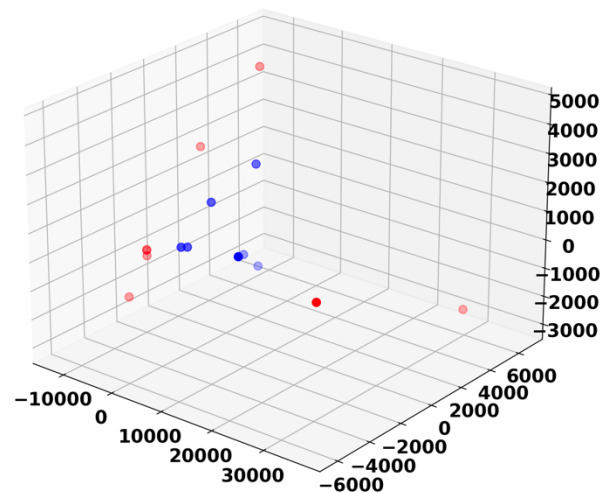


Figure S3. The feature space of VR-d by considering 6 rats from both groups of TgF344-AD and WT (TG1, TG2, TG3, WT4, WT5, WT6) in 3D space. The red and blue dots represent TG and WT rats respectively. Each figure is the result of the application of PCA algorithm on the matrix of spike train distances (Fig. 7). Three principal components from each spike distance matrix were considered as the inputs of the classifiers.