Neurotechnology and AI Approach for Early Dementia Onset Biomarker from EEG in Emotional Stimulus Evaluation Task

Tomasz M. Rutkowski1,2,3, Masato S. Abe1, and Mihoko Otake-Matsuura1

Abstract—We present an efficient utilization of a machine learning (ML) method concentrating on the ‘AI for social good’ application. We develop a digital dementia biomarker for early-onset dementia forecast. The paper demonstrates encouraging preliminary results of EEG-wearable-based signal analysis and a subsequent classification adopting a signal complexity test of a multifractal detrended fluctuation analysis (MF DFA) in emotional faces working memory training and evaluation tasks. For the digital biomarker of dementia onset detection, we examine shallow- and deep-learning machine learning models. We report the best median accuracies in a range of 90% for random forest and fully connected neural network classifier models in both emotional faces learning and evaluation experimental tasks. In addition, the classifiers are trained in a ten-fold cross-validation regime to discriminate normal versus mild cognitive impairment (MCI) cognition stages using MF DFA patterns from four-channel EEG recordings. Thirty-five volunteer elderly subjects participate in the current study concentrating on simple wearable EEG-based objective dementia biomarker progression. The reported outcomes showcase an essential social benefit of artificial intelligence (AI) employment for early dementia prediction. Furthermore, we improve ML employment for the succeeding application in this point in time complicated and applied EEG-wearable examination.

Clinical relevance—This project proposes an EEG-wearable-based objective cognitive biomarker candidate for a mild cognitive impairment (MCI) evaluation to substitute conventional idiosyncratic paper and pencil tests.

I. INTRODUCTION

The growth of dementia cases worldwide significantly affects healthcare expenses. Nearly 50 million older adults experience a dementia spectrum of neurocognitive dysfunctions, according to the World Health Organization (WHO) [1]. Moreover, this number is expected to triple within the next three decades [2]. This growing emergency requests feasible employment of AI to improve quick diagnostics for consequent cognitive well-being monitoring and protection with so-called “digital pharma” or “beyond a pill” non-pharmacological-therapeutical (NPT) strategies [3]. A final dementia assessment so far is only possible by postmortem autopsy [2]. A cognitive status exam, such as the Montreal Cognitive Assessment (MoCA) [4, 5], is regularly appointed for the severity of dementia quantification. Objective brain imaging approaches, such as functional magnetic resonance imaging (fMRI) [6] or EEG monitoring [7, 8, 9] concurrently with behavioral tests [10, 11], are recently in continuous increase to implement an early onset prediction of a mild cognitive impairment (MCI) and subsequent monitoring [2].

We propose to examine a wearable EEG device in this project to develop a viable home-based biomarker. We choose to use a commercial wearable MUSE EEG system (InterAxon Inc., Toronto, Canada) to collect EEG data. The MUSE headband affords an acceptable event-related-potential (ERP) and broadband EEG acquisition as confirmed in [12], [13]. However, dry-electrode-based EEG systems result in increased noise in EEG signals relating to clinical-grade amplifiers. To manage the more noisy time series, we recommend employing a signal complexity measure of a multifractal detrended fluctuation analysis (MF DFA), which is more robust to noise [14, 15].

II. METHODS

We conduct EEG recording trials with older adults in the RIKEN Center for Advanced Intelligence Project (AIP). The study adheres to guidelines of human subject involvement and ethical evaluation from the RIKEN Ethical Committee for Experiments with Human Subjects and The Declaration of Helsinki. In the study, 35 elderly take part; a number of females = 22; mean age = 73.5 ± 4.85 years old; recruited from Silver Human Resources Center and Honobono Laboratory, Japan. All participants receive monetary gratification, and they give written informed consent.

We apply a four-channel portable MUSE 2016 headband by InterAxon Inc., Canada. It has been shown already that the MUSE device allows for a reliable EEG capture from preset AF7, AF8, TP9, and TP10 dry electrode locations with a careful setting [12]. The ground reference electrodes are set at the forehead. We prepare our in-house EEG capture environment in Python using muse-lib [13] library, which communicates with MAX [16] visual programming environment short video stimulus presentation. We disinfect the MUSE headband with alcohol before a placement subject’s head with a careful check for EEG average amplitudes not to exceed thresholds indicating EMG contamination. We apply a 50Hz notch filter to remove power line interference, and we further bandpass the signal in a spectrum of 1 ~ 30Hz to minimize EEG out-of-band noise.

The experimental procedure is the same as in the previously published behavioral paradigm by our group [11]. Each participant sits in a chair in front of a computer display. We use a touchscreen tablet in the procedure. The...
computer screen displays video stimuli from a Mind Reading database [17] and response targets in a training mode. In the first training (encoding) part of the experiment, the participant’s task is to learn the procedure by copying a reference emotion judgment on the computer screen in an implicit working memory visuospatial skill. In the final testing (decoding) part of the experiment, no suggestions appear on the screen, and the subjects use newly acquired learning implicit emotion evaluation skills. Our experimental hypothesis is that EEG complexity modulated by cognitive skills shall be a good dementia biomarker candidate. Furthermore, since both training (easy and procedural) and testing (more complex and requiring application of newly learned skills), experimental procedures require from subjects sustained focused attention and short-term implicit learning skills, both paradigms shall elucidate normal versus MCI cognition-related modulation.

We segment EEG into eight-second long intervals with MNE version 0.23.0 package in Python [18]. We start each segment at video stimulus onset time. Each segment in each electrode channel separately is next processed for the signal complexity evaluation as follows. In chaos theory, stochastic processes analysis, and time-series processing, a detrended fluctuation analysis (DFA) [19], [20] method is often employed for determining a statistical self-affinity. The DFA helps analyze time series that appear to be the so-called long-memory processes, which could be characterized by diverging correlation resulting in a time power-law decaying autocorrelation function or 1/f noise frequency spectrum profile (EEG is such a signal as shown in [21], [22]). The generalized MFDFA calculation procedure consists of five steps, of which the first three are essentially identical to the conventional DFA procedure as in [19], [20]. The two additional steps, as proposed in [14] and efficiently implemented in [15], require averaging of overall segments from the standard DFA procedure to obtain the \( q^{th} \) order fluctuation function and subsequently a determination of a scaling behavior of the fluctuation functions by analyzing log-log plots versus the scale for each value of order \( q \) (details in [14]). For stationary time series, the MFDFA at order 2 is equal to a well-known Hurst exponent; thus, the MDFA is also regarded as a generalized Hurst exponent [14]. We obtain MFDFA features, using a Python implantation by [15], for each channel separately from eight-second-long post-stimulus intervals (short videos with emotional facial expressions from [17]) of the recoded and segmented EEG time series. The four-channel derived feature vectors resulting from each trial are used together as input for machine learning model training and evaluation using a ten-fold cross-validation procedure with careful separation of training and testing subsets to avoid overfitting.

We examine machine learning models, as available in the scikit-learn library version 0.24.2 [23], for binary classification of MCI versus normal cognition of the 35 participants (22 MCI and 23 normal) in our emotional facial expression short video working memory study using input MDFA feature vectors (62 on average from each subject; 1106 normal and 1088 feature vectors) obtained from four EEG channels. We apply a ten-fold cross-validation procedure due to a limited number of available classifier training examples, with a chance level of 50%. Furthermore, we apply the machine learning models to classify the participants into two groups (MCI and normal).
learning training and evaluation procedures to EEG epoch-derived features, not on a leave-one-one-subject level due to the abovementioned limited number of training samples. Similarly, as in our previously published study [11], we assess the following machine learning models using input feature standard scaling by deducting a mean and dividing by a signal variance: a logistic regression (LR) with a liblinear solver and a maximum iteration number of 1000; a linear discriminant analysis (LDA) with a least-squares solver; a linear support vector machine (linearSVM) using a 1/2-penalty linear kernel and a squared hinge loss function; a radial basis function support vector machine (rbfSVM) using a kernel coefficient $\gamma = 1/4$; a polynomial support vector machine (polySVM) with a second-degree polynomial kernel and $\gamma = 1/4$ and an independent term in kernel function $\text{cof}_0 = 1.0$; a sigmoid support vector machine (sigmoidSVM) with $\gamma = 1/4$; a random forest classifier (RFC) using a number of trees in the forest equal to 50, split criterion by a mean squared error, without a maximum tree depth limitation, and a minimum number of samples required for a split set to 2; a fully connected deep neural network (FNN) with rectified linear units (ReLU), using a configuration of a single input and three hidden layers (512, 256, 128, 32, and 16 units), a two-unit output softmax layer, and 50000 training epochs, with an ADAM optimizer, a learning rate set to 0.001, and a log-loss function. We use 10% of training data for evaluation in a ten-fold-cross-validation run for each above machine learning model.

III. RESULTS

Results of MFDFA distributors we summarized in Figures 1 and 2 for subject training and testing sessions respectively. We observed statistically significantly higher, as evaluated with Wilcoxon rank sums tests for $p_k \ll 0.01$, MFDFA values for MCI (MoCA $\leq 25$) evaluated subjects comparing to normal cognition (MoCA $> 25$) participants in our study group.

Results of EEG binary classification of normal (MoCA $> 25$) versus MCI (MoCA $\leq 25$) we summarized in the form of bar-plots with error bars depicting 95% confidence intervals of classification results using the evaluated classifiers for training (encoding) and testing (decoding) tasks in Figures 3 and 4, respectively. The RFC and FNN classifiers resulted in the best median accuracy exceeding 90% in training and testing tasks. Both best classifiers resulted in slightly better, but not statistically significantly different, results in the subject training (encoding) phase comparing to testing (decoding).

IV. CONCLUSIONS

We reported a project showcasing novel results in EEG response distributions in working memory implicit training (encoding) and testing (decoding) tasks. We showed that elderly subject MF DFA patterns obtained from four-channel wearable EEG recordings resulted in significantly differing temporal and frontal electrode locations in implicit working memory training and testing tasks using short videos with emotional facial expressions. The machine learning median classification accuracies resulted in a range of 90% for the best methods using RFC and FNN models as summarized in Figures 3 and 4 for subject training and testing tasks, respectively.

The well-off employment of such an AI/ML-based dementia onset forecast shall serve well the aging societies globally.

We also acknowledge the intrinsic limitations of the developed approach as we only replicate human-error-prone
subjective cognitive evaluation criteria rendered to binary MCI thresholds at a level MoCA ⩽ 25, which are only for the picom intervention program: a preliminary report.” BMC geriatrics, vol. 20, no. 1, pp. 1–10, 2020.


