Classifying Mild Cognitive Impairment from Behavioral Responses in Emotional Arousal and Valence Evaluation Task – AI Approach for Early Dementia Biomarker in Aging Societies –

Tomasz M. Rutkowski[1,*], Masato S. Abe[1,*], Marcin Koculak[2], and Mihoko Otake-Matsuura[1]

Abstract— **The presented paper discusses a practical application of machine learning (ML) in the so-called 'AI for social good' domain and in particular concerning the problem of a potential elderly adult dementia onset prediction. An increase in dementia cases is producing a significant medical and economic weight in many countries. Approximately 47 million older adults live with a dementia spectrum of neurocognitive disorders, according to an up-to-date statement of the World Health Organization (WHO), and this amount will triple within the next thirty years. This growing problem calls for possible application of AI-based technologies to support early diagnostics for cognitive interventions and a subsequent mental wellbeing monitoring as well as maintenance with so-called 'digital-pharma' or 'beyond a pill' therapeutical strategies. The paper explains our attempt and encouraging preliminary study results of behavioral responses analysis in a facial emotion implicit-short-term-memory learning and evaluation experiment. We present results of various shallow and deep learning machine learning models for digital biomarkers of dementia progress detection and monitoring. The discussed machine-learning models result in median accuracies right below a 90% benchmark using classical shallow and deep learning approaches for automatic discrimination of normal cognition versus a mild cognitive impairment (MCI). The classifier input features consist of an older adult emotional valence and arousal recognition responses, together with reaction times, as well as with self-reported university-level degree education and age, as obtained from a group of 35 older adults participating voluntarily in the reported dementia biomarker development project. The presented results showcase the inherent social benefits of artificial intelligence (AI) utilization for the elderly and establish a step forward to advance machine learning (ML) approaches for the subsequent employment of simple behavioral examination for MCI and dementia onset diagnostics.**

Clinical relevance— **This manuscript establishes a behavioral and cognitive biomarker candidate potentially substituting a Montreal Cognitive Assessment (MoCA) evaluation without a paper and pencil test.**

I. INTRODUCTION

Dementia, particularly an aging-associated memory decline, is one of the most critical global challenges in the 21^{st} century's social welfare and mental well-being. Increased longevity and progression of dementia cases affect welfare costs [1]. The Cabinet Office in Japan regularly publishes an annual report on an aging society to address the emergency [2]. United Nations Sustainable Development Goal #3 entitled "Good Health and Well-being " [3] also emphasizes a necessity to approach the aging problem with a focus on healthy lives, and it encourages wellbeing for all at all ages. Contemporary strategies to dementia and mainly the most severe case of Alzheimer's disease (AD) suggest a requirement to advance personalized therapies relying not only on conventional pharmacological interventions but also on lifestyle modifications [4] as well as cognitive maintenance approaches [5], [6]. There is also a social pressure for the dementia early-onset prognostication and subsequent prophylactic measures, as broadly addressed in [1]. All the established pharmacological and the recent 'beyond-a-pill,' or the so-called 'digital-pharma,' therapeutical interventions require reliable biomarkers. Usually, a research focus was on advanced applications of brainwave-related techniques [7], [8], [9], [10], [11], which often entail a more clinical-level environment for a successful application.

A contribution of the reported project is twofold. First, we present a novel behavioral data collection experimental approach, which subsequently could be implemented in a tablet or smartphone application for daily use by adults interested in their cognitive wellbeing monitoring. Next, we apply a range of shallow and deep machine-learning approaches with very encouraging results already on the dataset of 35 participants.

We develop a machine-learning (ML) strategy, belonging to a domain of modern AI for the social or common good. A thriving application shall allow for computerized discrimination of mild cognitive impairment (MCI), defined as the Montreal Cognitive Assessment score MoCA$\leqslant$ 25 [12], [13], versus average level cognition in the elderly using only behavioral responses in a working and implicit/procedural memory learning paradigm constituting a newly acquired skill-testing assignment. A self-reported working-memory decline decreeing a subjective cognitive impairment (SCI), is one of the early signs practiced in the medical community [14]. MCI characterizes also emotional contagion [15] and parietal cortex as well as hippocampus atrophy related spatial memory problems [1], [16]. A visuospatial memory focusing biomarkers has been recently proposed as potentially beneficial research targets [17]. Unlike long-term memory or a language, visuospatial functioning is heavily dependent on parietal lobe integrity, where atrophy

*The authors Tomasz M. Rutkowski and Masato S. Abe contributed equally to this project.

[1]Tomasz M. Rutkowski, Masato S. Abe, and Mihoko Otake-Matsuura are with Cognitive Behavioral Assistive Technology Team, RIKEN Center for Advanced Intelligence Project (AIP), Tokyo, Japan, tomasz.rutkowski@riken.jp

[2]Marcin Koculak is with the Consciousness Lab, Institute of Psychology, Jagiellonian University, Krakow, Poland

in function or structure occurs early in dementia [17].

There is no definite confirmation about the working-implicit/procedural-memory impairment [18], on the other hand, and only the long-term recall is known to be unaltered in dementia cases [15]. Therefore we also incorporate the procedural memory component for the novel dementia biomarker paradigm as we plan to create a more straightforward task for the elderly. In the next section, we describe our experiment and data collection procedures in detail.

We also develop a machine-learning-based biomarker, which utilizes behavioral responses in the spatial and working-implicit/procedural-memory testing task. The suggested biomarker shall consequently allow for practical employment in an uncomplicated gaming-style touchpad or smartphone application concerning daily use for older adults enduring cognitive or lifestyle interventions contributing to dementia rise postponement or even possible process reversal as proposed in [4]. The up-to-date methods for dementia diagnostics rely on pencil-and-paper nonobjective psychometric testing, for example, the MoCA [12], the more elaborate physiological or brain imaging analyses [9], or massive multi-sensory datasets [19]. The latter techniques often demand expensive devices, very long testing periods, or advanced clinical environments.

We present an exploratory and following ML/AI behavioral response analysis procedure in which we ask elderly citizens to learn a reasonable new emotional face evaluation skill employing a two-dimensional graph, a so-called emoji-grid [20], of valence and arousal rates, which is an offhand spatial- and implicit-working-memory assignment. Following a short training, the users perform an examination trial in which response time, valence and arousal inputs together with self-reported university-level degree education as well as age create input features to train ML models as outlined in the subsequent methods section. First, we communicate on a pilot study results. We administer the pilot study with a small representation of university graduate students, middle-age, high- and low-scoring on MoCA-scale older adults, as well as reference active-seniors of 80+ years old. The pilot study produces supportive outcomes of behavioral reaction time (RT) and emotional arousal/valence (AV) responses as biomarker nominees. The outcomes of the pilot study encourage the subsequent conclusive project with 35 elderly MoCA-evaluated citizens. We summarize encouraging ML results, allowing for predicting MCI scores in behavioral paradigm and without the requirement for highly subjective paper-and-pencil questionnaires.

The paper is organized as follows. The next section describes data collection, experimental, and machine learning methods used in the study. Results discussion follows, and conclusions with future research directions summarize the manuscript.

II. METHODS

We administered trials with the human participants in the RIKEN Center for Advanced Intelligence Project (AIP) following guidelines and permission of RIKEN Ethical Committee for Experiments with Human Subjects and The Declaration of Helsinki regarding ethical principles for human experimentation.

In the pilot study small groups (less than ten each) of participants we divided into five groups of low-MoCA (MoCA $\leqslant 25$), normal-MoCA (MoCA > 25), ctive 80+ (active professionals with ages above 80 years old), middle-aged (ages $40 \sim 50$ years old), and graduate students (graduate level and $20 \sim 35$ years old). In the final main trial session, 35 elder participants (number of females $= 22$; mean age $= 73.5$ years old; age standard deviation of ± 4.85 years; recruited from Silver Human Resources Center and Honobono Laboratory) took part. All participants received monetary gratification for their participation in the study, and they gave informed written consent.

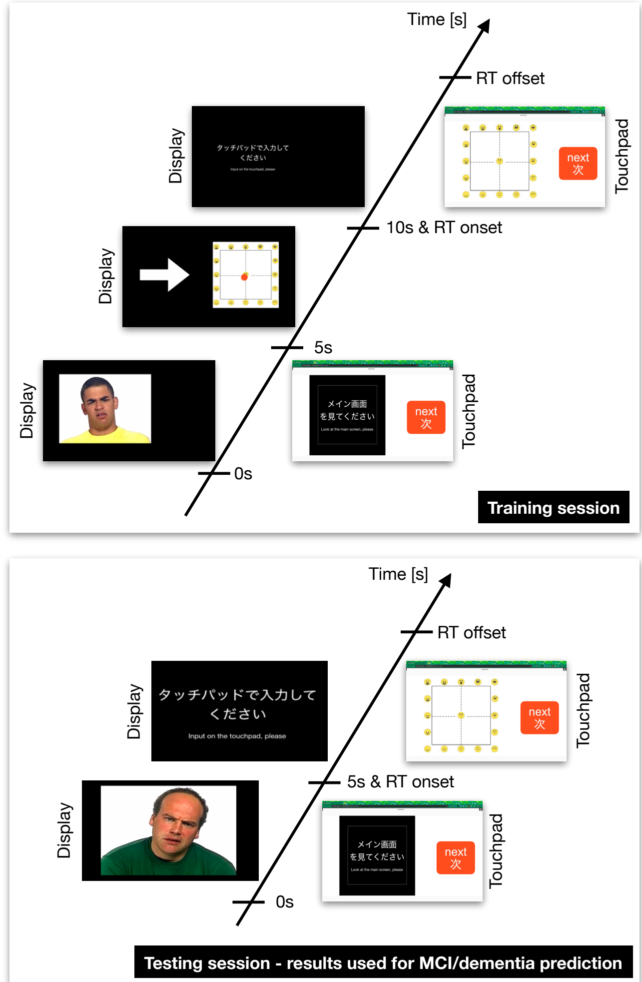

Fig. 1. The experimental procedure timeline. The top panel presents user display utterances on the left and tablet's touchscreen input screen on the right in a single trial example. A user first can see a short facial emotional display randomly chosen from the Mind Reading database [21]. Next, a respective valence and arousal score is shown [22], and the user is requested to memorize it and subsequently to input it on a tablet in the final step of a single trial. The bottom panel presents a timeline of the spatial- and implicit-working-memory testing single trial, in which we omit a suggestion screen.

A. Experimental Procedure

During the experimental session, we tested the participants in pairs, but they were seated in such a way that they could not see each other stimulus display and response touchpad. Each participant was sitting in a chair in front of a computer display. A touchscreen tablet was also used in the procedure. The computer screen displayed stimuli and response targets in a training mode. The touchscreen tablet was recording participant's responses on a two-dimensional emoji-grid [20] together with response times after each video clip ended. The experimental procedure was coded in a visual programming environment MAX by *Cycling '74, USA*. The experiment was divided into two similarly structured parts, as is depicted in Figure 1.

In the first part of both pilot and final experiments, the participants' task was to learn the procedure by copying a reference emotion judgment shown on the computer screen. At the beginning of each trial, on the left side of the screen, a short video clip with facial emotional expression from a Mind Reading database [21] developed initially as a teaching emotion-recognition aid for people with autism spectrum disorders. Next, a two-dimensional grid was displayed on the right side with the horizontal dimension describing the valence of emotion and the vertical arousal [22], [20], respectively. On this grid, an associated emotion score was displayed in the form of an orange dot for five seconds. To help participants understand the meaning of those dimensions, we placed a set of emoji depicting emotions [20], as shown in the top panel within Figure 1. The participant had to respond by touching an identical position on a grid displayed on the tablet's touchscreen in a place where the reference judgment was shown. Emojis were placed around both grids on the screen and the tablet. After doing it, the participant had to touch a next button displayed on the tablet, which triggered the subsequent trial start (see Figure 1). Tablet was unresponsive during the presentation of the video and the emotion grid, so participants had to watch the whole material before giving an answer and progressing to the next trial.

For every participant, an experimental session consisted of 72 video display trials ($5 \sim 7$ seconds each) with 24 distinct emotion classes [21]. Three distinct videos reproduced every emotion with actors varying in age, gender, and skin color. The sequence of the videos was randomized before the trial but was the same for every participant. Reference judgments were taken from a published article [22]. After a short break, participants started the second part, which contained the same amount of stimuli and the same emotions as the first part. However, new videos were chosen as the stimuli, so participants have not seen them before [21]. Another significant difference was the absence of suggestions on the screen after a video was played (refer to a bottom panel in Figure 1). Therefore participant's task was to reproduce a previously learned spatial-evaluation judgment of emotions seen in each video concerning the valence and arousal dimensions. Thus the spatial- and implicit/procedural-working-memory was tested together with recorded reaction (thinking intervals) times estimating a cognitive load [10], [23].

B. Behavioral Data Recording for the Subsequent AI Application

During the data recording operations, arousal and valence score responses, as well as the reaction times were registered together by the stimulus display program developed in a visual programming ecosystem MAX by *Cycling '74, USA*. At the end of each testing session, all results were saved automatically in a text file and next preprocessed for feature extraction. While in the pilot study we analyzed the actual valence and arousal values (see two lower panels in Figure 2), than in the final experiment we obtained absolute response score errors to unify response spreads and reaction times as,

$$v_e(i,s) = |v_d(i) - v_t(i,s)|, \tag{1}$$
$$a_e(i,s) = |a_d(i) - a_t(i,s)|, \tag{2}$$
$$r_t(i,s) = t_t(i,s) - t_o(i), \tag{3}$$

where $s = 1,\ldots,20$ identified the participant in our final study; $i = 1,\ldots,72$ represented the i^{th} video clip presented in the experiment; $v_e(i,s)$ and $a_e(i,s)$ were the valence and arousal errors related to emotional stimulus i and participant s, respectively; $v_d(i)$ and $a_d(i)$ were the video clip assigned ground truth emotional scores from [21], [22]; $v_t(i,s)$ and $a_t(i,s)$ the actual response inputs by a user number s on a touchpad after the video clip number i, which reflected the learned emotion evaluation in the spatial- and working-memory task; $r_t(i)$ was a reaction time obtained as an interval between user response $t_t(i,s)$ and the i^{th} video clip end at a timestamp $t_o(i)$. A single video clip evaluation feature vector $\mathbf{F}_{i,s}$ related to video clip i and participant s for each evaluated next classifier in training and subsequent leave-one-participant-out cross-validation procedure has been built as follows,

$$\mathbf{F}_{i,s} = \begin{bmatrix} v_d(i) \\ v_e(i,s) \\ a_d(i) \\ a_e(i,s) \\ r_t(i,s) \\ e(s) \\ g(s) \end{bmatrix}, \tag{4}$$

where the quantities $v_e(i,s), a_e(i,s)$ and $r_t(i,s)$ were obtained from equations (1) to (3) for participant and video number s and i, respectively; $e(s) \in \{0,1\}$ denoted a self-reported education level; and $g(s)$ age. A pairwise comparison scatter plots together with linear regression fits are summarized together with classifier output MCI levels in Figure 3.

C. Machine Learning Methods

We tested classifiers available in *the scikit-learn* library version 0.22.1 [24] and Tensorflow 2.1.0 [25] for binary classification of MCI versus normal cognition of the 35 participants in our final study using input features $\mathbf{F}_{i,s}$ from equation (4). We used a leave-one-participant-out procedure with carefully balanced labels from both MCI and normal cognition classes in each cross-validation run. Thus, a chance

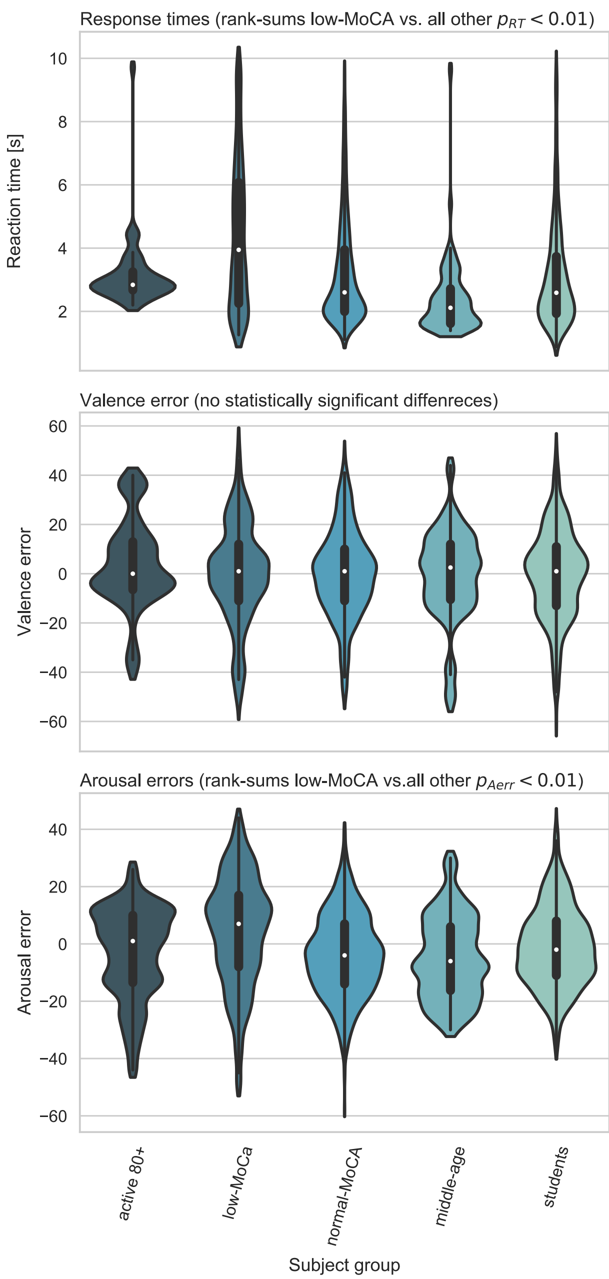

Fig. 2. Three panels present the participant reaction-time (RT), valence and arousal error experimental results from a behavioral pilot study in which short-video-clips with emotional expressions presented to the five user groups. The white dots within the distribution plots depict median results, and the thick-black-lines the error-bars the 25-percentile-intervals, respectively. Response times and arousal errors for the low-MoCA group resulted in statistically significant differing quantities (larger values) comparing to all the other participants with $p_{RT} < 0.01$ and $p_A < 0.01$, respectively, as tested with the Wilcoxon rank-sum tests. Emotional valence response error distributions did not differ significantly among all the evaluated groups in the presented study.

level in every classification trial was guaranteed to be 50% as marked in results Figure 6. The following methods and appropriate steps were implemented:

- **Logistic regression (LR):** standard scaling of input features $\mathbf{F}_{i,s}$ by removing the mean and dividing by a variance; application of a *liblinear* solver; setting a maximum iteration number to 1000.
- **Linear discriminant analysis (LDA):** application of a least-squares solver without shrinkage.
- **Linear support vector machine (linearSVM):** standard scaling of input features $\mathbf{F}_{i,s}$ by removing the mean and dividing by a variance; application of linear kernel with $l2-$penalty; loss set to squared hinge.
- **Radial basis function support vector machine (rbfSVM):** application of a radial basis function kernel; with a kernel coefficient *gamma* set to $1/7$ representing an inverse of feature vector $\mathbf{F}_{i,s}$ length.
- **Polynomial support vector machine (polySVM):** standard scaling of input features $\mathbf{F}_{i,s}$ by removing the mean and dividing by a variance; application of a second degree polynomial kernel; with a kernel coefficient *gamma* set to $1/7$ representing an inverse of feature vector $\mathbf{F}_{i,s}$ length; with an independent term in kernel function $coef0 = 1.0$.
- **Sigmoid support vector machine (sigmoidSVM):** standard scaling of input features $\mathbf{F}_{i,s}$ by removing the mean and dividing by a variance; application of a sigmoid kernel; with a kernel coefficient *gamma* set to $1/7$ representing an inverse of feature vector $\mathbf{F}_{i,s}$ length.
- **Random forest classifier (RFC):** with a number of trees in the forest set to 200; mean squared error used as a split criterion; no maximum tree depth limitation; and a minimum number of samples required for a split set to 2.
- **Fully connected deep neural network (FNN):** with densely connected rectified linear units (ReLU), configured in one input and five hidden layers with $32, 64, 256, 512, 256$, and 32 units, respectively; the above middle layer with 512 units followed by dropout activation set to 50%; an output softmax layer with two units; the whole network consisted of $290,146$ trainable parameters; the training conducted using batches of 128 and maximum 500 epochs with early stopping set for no validation loss improvement for more than 7 epochs using an ADAM optimizer with a learning rate set to 0.001, and a loss function of a binary-cross-entropy; 10% of training data was used for validation in each leave-one-participant-out run, respectively.

III. RESULTS

The current project first resulted with an encouraging pilot study using a small and nonuniform sample of participants, and next the carefully balanced sample of 35 older adults confirmed a possibility of binary classification of MCI (MoCA $\leqslant 25$) versus normal (MoCA > 25) cognition using only participant behavioral responses in the spatial- and implicit-working-memory task. A detailed discussion of results follows.

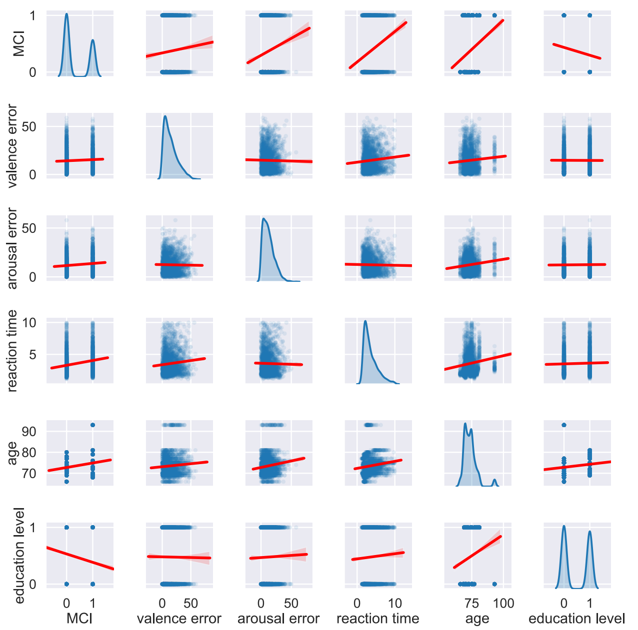

Fig. 3. A collection of scatter plots showing pairwise relationships of all output/target (MCI; where 1 represents those with MoCA ≤ 25 and 0 those with normal cognition, respectively) and input (with following abbreviations: valence_err = user valance input error; arousal_err = user arousal input error; RT = user response time; edu = user self-reported education level; age = user self-reported age) features used in the subsequent machine-learning analysis. Red lines depict linear regression fits with shaded confidence intervals of the pairwise data distributions.

A. Pilot Study Statistical Data Analysis of Behavioral Responses

The pilot and preliminary study results are summarized in Figure 2 in the form of distribution plots of reaction-times, valence, and arousal response errors. Out of five participant groups, only the low-MoCA group (MoCA $\leqslant 25$) resulted in significantly differing median responses as tested with Wilcoxon rank-sums test. The response times resulted in $p_{RT} < 0.01$, while similarly, the arousal response errors were at the same statistical significance level of $p_{Aerr} < 0.01$,, respectively. The valence errors did not differ significantly among the participant groups.

TABLE I

COMPARISON OF CLASSIFIERS IN LEAVE-ONE-PARTICIPANT-OUT CROSS-VALIDATION FOR THE TARGET CLASS MEDIAN F_1-SCORES, WHERE $F_1 = (precision \cdot recall)/(precision + recall)$

Classifier	Median F_1-score
LR	**0.945**
LDA	0.921
shrinkage LDA	0.899
linear SVM	0.933
rbf SVM	0.783
polynomial SVM	0.738
sigmoid SVM	0.898
RFC	0.790
FNN	0.941

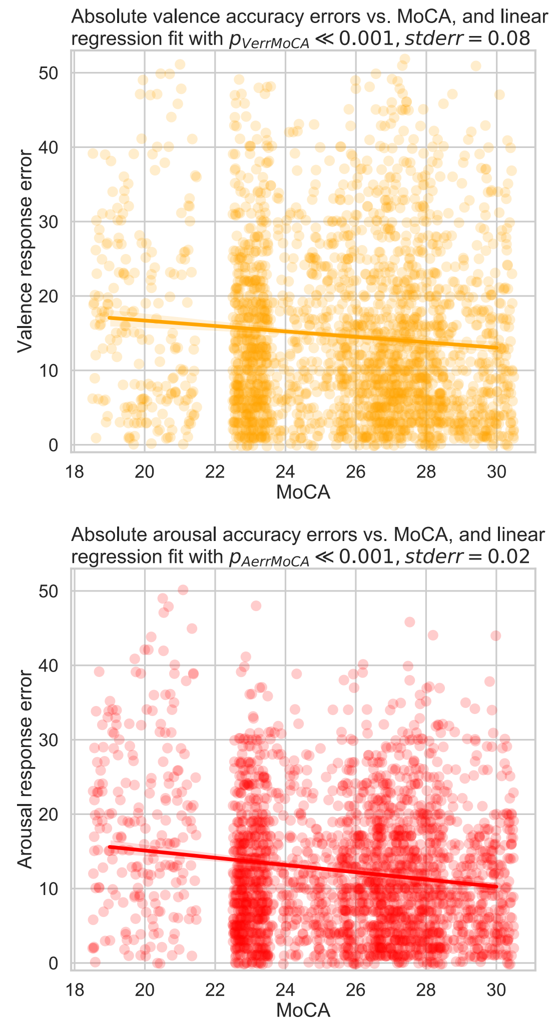

Fig. 4. The elderly participant results in the form of valence (top panel), and arousal (bottom, respectively) response absolute errors in the spatial- and implicit-working-memory task plotted in function MoCA cognitive evaluations. We also depict linear regression fits with confidence intervals of 95-percentile levels, which show an apparent linear increase of both absolute response errors for lower-MoCA-scoring participants. We summarize statistics of the linear regression fits above the graphs in the form of p-values and standard errors ($stderr$).

B. Multimodal Experiment for MCI Inference from Behavioral Response Classification Results

The final study results are summarized in form behavioral feature distributions in Figures 4 and 5, for arousal and valence response errors, as well as response times, respectively. We also performed linear regression fit analysis of the detailed results (p-values and standard errors) depicted in Figures 4 and 5. All the behavioral results did linearly increase for lower MoCA scores. We present the pairwise comparison of the classification input features in Figure 3. The feature sets were very noisy, thus a simple statistical analysis was not sufficient for a final application.

The results of shallow and deep learning classifiers (at the current stage a small sample of participants did not allow for a successful application of more complex deep

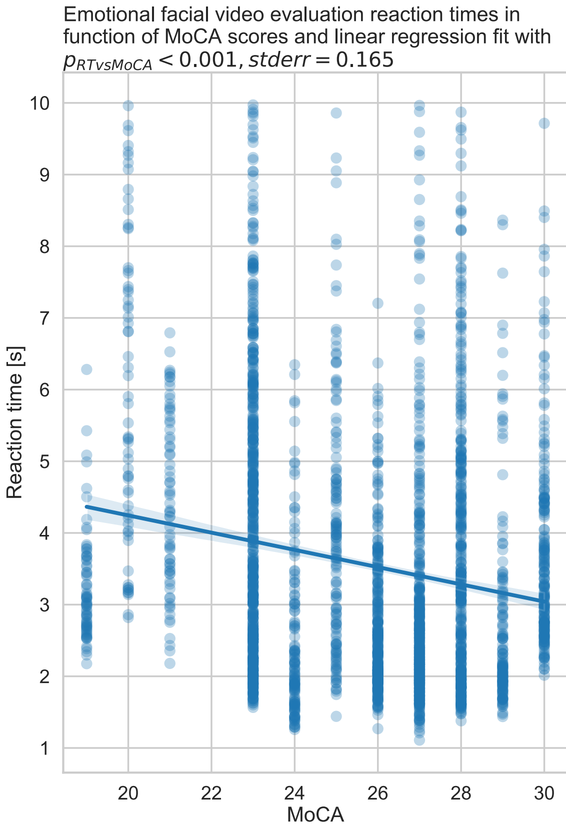

Fig. 5. The elderly participant results in the form of reaction times in seconds in the spatial- and implicit-working-memory task plotted in function MoCA cognitive evaluations. We also depict the linear regression fit with confidence intervals of 95-percentile levels, which also confirms an apparent linear increase of both absolute response errors for lower-MoCA-scoring participants.

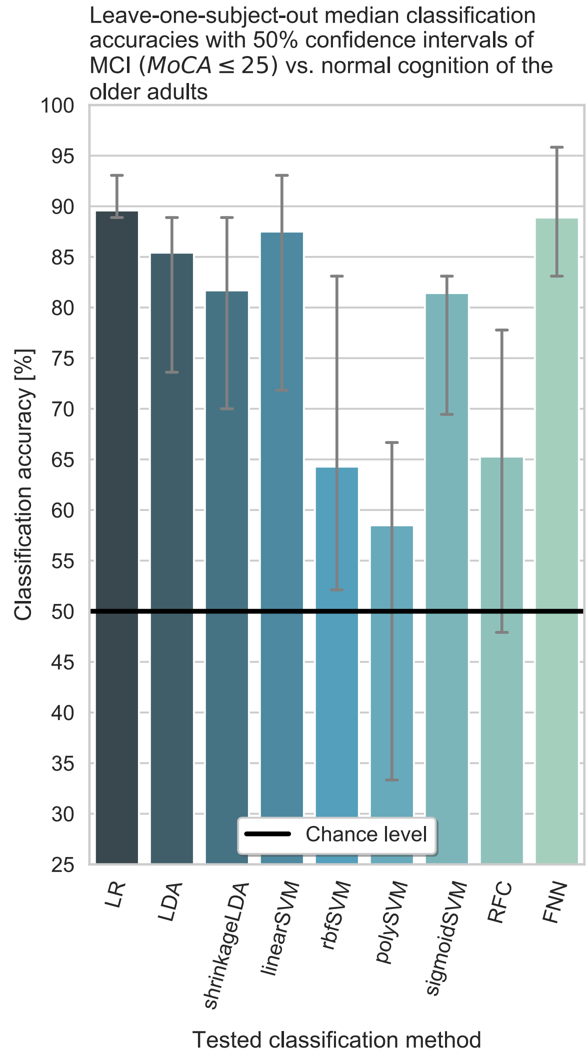

Fig. 6. Median accuracy classification results of binary classifiers discriminating between MCI (MoCA $\leqslant 25$) versus normal cognition (MoCA > 25) cases together with confidence intervals of $50^{th}-$percentile. The successfully evaluated shallow learning classifiers [24] were as follows: logistic regression (LR); linear discriminant analysis (LDA); linear support vector machine (linearSVM); radial basis function support vector machine (rbfSVM); polynomial support vector machine (polySVM); the sigmoid support vector machine (sigmoidSVM); the random forest classifier (RFC); and the fully connected deep neural network (FNN) [25] was also applied, respectively. Detailed settings of the evaluated classifiers are summarized in Section II-C in this paper.

learning methods) are summarized in Figure 6 in the form of median bar plots together with $50-$percentile error-bars. The majority of classifiers scored way above a chance level of 50% with very encouraging median results for the linear regression and the fully connected deep neural network (FNN) just below a 90% benchmark.

The median $F1-$score results [24] in leave-one-participant-out cross-validation scenarios are summarized in Table I with the best mean scores obtained for LR– and FNN–based classifiers. Those two classifiers also resulted in the best median accuracies, although the FNN method had some higher accuracies as depicted with error bars in Figure 6.

We plan that after collecting a comprehensive database in near-future, a further elaborate deep learning model would allow for even more thriving classification accuracies of MCI versus healthy elderly discrimination.

IV. CONCLUSIONS

The reported project produced two significant results. First, in the discussed pilot study, we distinguished three behavioral responses as nominees for MCI pathology biomarking from a gaming-style paradigm involving the spatial- and implicit-working-memory task of emotional arousal and valence scoring together with reaction time records in mere video clips evaluation task. In the closing study concerning older adults with identified MoCA rates, we were able to assess several shallow and deep learning classifiers. The classification in a binary organization of MCI (MoCA $\leqslant 25$) against standard cognition (MoCA > 25) levels in the mere emotional faces evaluation task resulted in reliable accuracy outcomes (median classification accuracies just below 90% for the best methods) as summarized in Figure 6. The discussed innovative strategy for behavioral responses in emotional faces evaluation task concerning spatial- and implicit-working-memory-based original skill acquisition together with classification outcomes contribute to advancement in research and improvement of new dementia-level-estimation behavioral biomarkers for the elderly, for whom possible early determination of cognitive decline, as well as a life improvement, are indispensable.

The successful employment of such an AI/ML-based dementia onset forecast shall lead to healthcare cost-reducing benefiting all the aging societies worldwide.

We also acknowledge the inherent limitations of the discussed approach as we only infer human-error-prone nonobjective cognitive evaluation rates translated to binary MCI thresholds at a level MoCA $\leqslant 25$, which are only proxy predictors of dementia. AI-based dementia predictors, if used without proper evaluation, might also pose a danger of ill-usage or abuse; thus, proper ethical standards will need to be in place too. A possible dementia diagnostic gain for medical doctors using the proposed method would be a possibility to drop paper-and-pencil tests, as well as to administer the proposed methodology inspections more frequently to observe therapy progress.

In the succeeding research project, we intend to assess the developed techniques with a more extensive representation of ordinary versus MCI or even dementia diagnosed citizens. We also plan to merge the proposed behavioral measures with brainwaves, especially EEG and fNIRS, datasets for even more reliable final classification. We also recognize that the future application of AI methods for fully interactive paradigms in closed-loop user behavior and brainwave monitoring shall lead to even more impactful results.

REFERENCES

[1] G. Livingston, A. Sommerlad, V. Orgeta, S. G. Costafreda, J. Huntley, D. Ames, C. Ballard, S. Banerjee, A. Burns, J. Cohen-Mansfield, *et al.*, "Dementia prevention, intervention, and care," *The Lancet*, vol. 390, no. 10113, pp. 2673–2734, 2017.

[2] Japanese Goverment, "Annual report on the aging society," 2018.

[3] United Nations, "United nations sustainable development goals," August 2019.

[4] D. Bredesen, *The End of Alzheimer's: The First Programme to Prevent and Reverse the Cognitive Decline of Dementia*. Vermilion, 2017.

[5] S. J. Czaja, W. R. Boot, N. Charness, W. A. Rogers, and J. Sharit, "Improving Social Support for Older Adults Through Technology: Findings From the PRISM Randomized Controlled Trial," *Gerontologist*, vol. 58, no. 3, pp. 467–477, 2018.

[6] R. M. G. Reinhart and J. A. Nguyen, "Working memory revived in older adults by synchronizing rhythmic brain circuits," *Nature Neuroscience*, vol. 22, no. 5, pp. 820–827, 2019.

[7] Z. Chen, A. Cichocki, and T. Rutkowski, "Constrained non-negative matrix factorization method for eeg analysis in early detection of alzheimer disease," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. V, pp. 893–896, 2006.

[8] J. Mcbride, X. Zhao, N. Munro, G. Jicha, C. Smith, and Y. Jiang, "Discrimination of Mild Cognitive Impairment and Alzheimer's Disease Using Transfer Entropy Measures of Scalp EEG HHS Public Access," *J Heal. Eng*, vol. 6, no. 1, pp. 55–70, 2015.

[9] A. Horvath, A. Szucs, G. Csukly, A. Sakovics, G. Stefanics, and A. Kamondi, "EEG and ERP biomarkers of Alzheimer's disease: a critical review," *Frontiers in Bioscience*, vol. 23, pp. 183–220, 2018.

[10] T. M. Rutkowski, Q. Zhao, M. S. Abe, and M. Otake, "AI neurotechnology for aging societies - task-load and dementia EEG digital biomarker development using information geometry machine learning methods," in *AI for Social Good Workshop at the Neural Information Processing Systems (NeurIPS) 2018*, (Montreal, Canada), pp. 1–4, December 8, 2018.

[11] T. M. Rutkowski, M. Koculak, M. S. Abe, and M. Otake-Matsuura, "Brain correlates of task–load and dementia elucidation with tensor machine learning using oddball BCI paradigm," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8578–8582, May 2019.

[12] P. Julayanont, Z. Nasreddine, M. Brousseau, M. Borrie, H. Chertkow, and N. Phillips, "The Montreal cognitive assessment memory index score (MOCA-MIS) and total MOCA score to help predict mci conversion to Alzheimer's disease," *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, vol. 8, no. 4, p. P372, 2012.

[13] M. Nara, M. Sugie, T. Takahashi, T. Koyama, R. Sengoku, Y. Fujiwara, O. Shuichi, K. Harada, S. Kyo, and H. Ito, "Japanese version of the montreal cognitive assessment cut-off score to clarify improvement of mild cognitive impairment after exercise training in community-dwelling older adults: Moca-j score to clarify mci improvement," *Geriatrics & Gerontology International*, vol. 18, 02 2018.

[14] S. Emrani, D. J. Libon, M. Lamar, C. C. Price, A. L. Jefferson, K. A. Gifford, T. J. Hohman, D. A. Nation, L. Delano-Wood, A. Jak, K. J. Bangen, M. W. Bondi, A. M. Brickman, J. Manly, R. Swenson, and R. Au, "Assessing working memory in mild cognitive impairment with serial order recall," *J. Alzheimer's Dis.*, vol. 61, no. 3, pp. 917–928, 2018.

[15] A. Gutchess, *Cognitive and Social Neuroscience of Aging*. Cambridge Fundamentals of Neuroscience in Psychology, Cambridge University Press, 2019.

[16] D. Howett, A. Castegnaro, K. Krzywicka, J. Hagman, D. Marchment, R. Henson, M. Rio, J. A. King, N. Burgess, and D. Chan, "Differentiation of mild cognitive impairment using an entorhinal cortex-based test of virtual reality navigation," *Brain*, vol. 142, pp. 1751–1766, 05 2019.

[17] S. Salimi, M. Irish, D. Foxe, J. R. Hodges, O. Piguet, and J. R. Burrell, "Can visuospatial measures improve the diagnosis of alzheimer's disease?," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 10, pp. 66–74, 2018.

[18] S. D. Slotnick, *Cognitive Neuroscience of Memory*, ch. Implicit Memory, pp. 129–149. Cambridge Fundamentals of Neuroscience in Psychology, Cambridge University Press, 2017.

[19] R. Chen, F. Jankovic, N. Marinsek, L. Foschini, L. Kourtis, A. Signorini, M. Pugh, J. Shen, R. Yaari, V. Maljkovic, *et al.*, "Developing measures of cognitive impairment in the real world from consumer-grade multimodal sensor streams," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2145–2155, ACM, 2019.

[20] A. Toet, D. Kaneko, S. Ushiama, S. Hoving, I. de Kruijf, A.-M. Brouwer, V. Kallen, and J. B. F. van Erp, "EmojiGrid: A 2D Pictorial Scale for the Assessment of Food Elicited Emotions," *Front. Psychol.*, vol. 9, p. 2396, nov 2018.

[21] S. Baron-Cohen, *Mind Reading - The Interactive Guide to Emotions*. London, UK: Jessica Kingsley Publishers, 2004.

[22] A. B. Warriner, V. Kuperman, and M. Brysbaert, "Norms of valence, arousal, and dominance for 13,915 english lemmas," *Behavior research methods*, vol. 45, no. 4, pp. 1191–1207, 2013.

[23] T. M. Rutkowski, M. S. Abe, M. Koculak, and M. Otake-Matsuura, "Cognitive assessment estimation from behavioral responses in emotional faces evaluation task - ai regression approach for dementia onset prediction in aging societies -," in *NeurIPS Joint Workshop on AI for Social Good*, vol. Track 1 - Producing Good Outcomes, (Vancouver, Canada), pp. 1–4, December 14, 2019.

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[25] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.