Automated Feature Engineering for Single-Trial EEG and Eye-Tracking Classification in Predictive Text Interfaces

Ard Kastrati[1]
ETH Zurich
akastrati@ethz.ch

R. Michael Winters
Pinnacle Group, Inc

Nemanja Djuric
Microsoft Research

Ivan J. Tashev
Microsoft Research

Yu-Te Wang
Academia Sinica
yutewang@citi.sinica.edu.tw

***Abstract*—Brain-Computer Interfaces (BCIs) offer a direct connection between the human brain and digital systems, enabling innovative applications. However, realizing the full potential of BCIs remains challenging due to issues like noise, artifacts, and limited data availability. In this study, we develop a multimodal classifier that integrates electroencephalogram (EEG) and eye-tracking (ET) data to decode user responses to predictive text suggestions. Utilizing an automated feature engineering approach, our pipeline efficiently generates and selects relevant features without extensive manual intervention or deep theoretical insights. Applied to a recent BCI case study involving predictive text input, our method achieved higher classification accuracies compared to traditional approaches. Additionally, it revealed novel insights, such as behavioral patterns where participants did not fully read incorrect predictions and the enhanced performance of multimodal classifiers when combining ICA-preprocessed EEG data with ET data. While automated feature engineering is standard in other domains, it is seldom applied in BCI research. Our findings demonstrate that this approach is a valuable tool for data-driven exploration and the development of competitive single-trial classifiers in novel multimodal BCI paradigms, particularly during the initial stages of research with limited data.**

***Index Terms*—EEG, Eye Tracking, automated feature engineering, predictive interaction systems**

I. Introduction

Brain-computer interfaces (BCIs) enable users to interact directly with digital systems through their neurophysiological signals, offering transformative potential for human-computer interaction and assistive technologies [1]. While significant progress has been made, fully realizing the capabilities of BCIs necessitates the exploration of new experimental setups and paradigms. Several recent studies have proposed novel BCI paradigms, highlighting the potential of using Brain-Computer Interfaces (BCIs) in innovative applications [2]–[4]. One of the recent interesting case studies includes *predictive text inputs* [5], where systems suggest the next word or phrase as users type, as commonly seen in smartphones and modern text editors.

[1] This work was done as a research intern at Microsoft Research, Redmond, WA 98052, USA.

One critical component in developing novel BCI paradigms is the task of building effective *single-trial* classifiers. Developing these classifiers in new paradigms is particularly challenging because persistent issues such as noise, artifacts, and inter-user variability [6] make it difficult to discern whether meaningful neurophysiological signals are present in the data. This challenge is amplified in multi-modal setups [7], [8], where integrating different types of physiological data introduces additional layers of complexity. Traditionally, single-trial classification in the BCI domain involves either deep learning approaches [9] or methods based on feature extraction motivated by theoretical insights [1]. However, both of these approaches exhibit limitations and challenges in the BCI context. On the one hand, deep learning methods, despite their success in various domains, struggle with limited data availability in the BCI domain. This is especially more relevant in the context of early exploration of new paradigms and small datasets. On the other hand, feature extraction, often grounded in existing literature and theories, typically has limitations in terms of classification accuracy. In novel experimental paradigms, existing theoretical frameworks may not fully capture all pertinent aspects of the data, potentially leading to suboptimal feature selection and overlooking important information.

In this study, we show that *automated feature engineering* — a common technique applied in many other domains — can be also successfully applied in BCI, especially in the case of novel paradigms, where the theoretical insights are not fully developed and there is a lack of large scale collection of data to build effective deep learning models. While automated feature engineering is not novel, it is not typically explored in BCI research. This is mainly because, once experiments are collected on a large scale or the paradigm and data are fully studied, deep learning methods and expert-crafted features usually perform best. However, in the initial stages of new paradigms, this is challenging to achieve. We argue that the simplicity and efficiency of automated feature engineering allow for a quick initial exploration and can be a valuable pipeline to be used in the early stages of researching BCI applications.

We apply a simple, and fully automated data-driven approach to a recent BCI case study that investigated users' neurophysiological responses during interactions with commonplace systems that simulated a natural typing environment. While the original analysis focused on grand average responses across all trials and provided the initial insights into this novel application, the potential for real-time, single-trial classification between a correct and incorrect word prediction remained unexplored. Despite the limited data in this dataset and the inherent challenges of early-stage research in novel paradigms, we show that automated feature engineering can help in discovering new insights and building data-driven single-trial classifiers. In an entirely data-driven approach, we automatically uncover novel insights into this BCI case study and additionally develop a single-trial classifier that outperforms traditional approaches. More concretely, through automated feature engineering, we achieve the following results:

1) *Single Modality Classifiers*: Our best single-modality classifier achieved an accuracy of 69% within participants and 64% across participants for EEG, as well as 78% within participants and 74% across participants for Eye Tracking (ET). Notably, our pipeline automatically identified that higher ET accuracy primarily resulted from a distinctive behavioral pattern: participants tend to read the entire suggested word when predictions are correct but not when predictions are incorrect, resulting in varying reaction times.
2) *EEG and ET Hybrid Classifier*: We developed an EEG-ET hybrid classifier that outperformed single-modality classifiers, achieving an accuracy of 81% within and 77% across participants. This underscores the synergy between EEG and ET modalities. Notably, in contrast to the single-modality case, our automatic pipeline showed that the highest accuracy was achieved when Independent Component Analysis (ICA) was applied to EEG data, as this enabled the extraction of true brain activity signals that complemented the eye-tracking data.

II. Related Work

This section discusses relevant research that highlights various methods for feature extraction from eye tracking (ET) data and EEG signals, as well as different approaches of single-trial multimodal classifiers in comparable experimental setups.

a) ET: Zagermann et al. [10] sheds light on the relationship between eye movements and cognitive load. Their comprehensive analysis delves into various eye-tracking features, including saccades, fixations, blinks, and pupil size, revealing their relevance in capturing cognitive processes. In particular, they state that the higher the load the longer the fixations and lower the fixation rates. Additionally, they also conclude that pupil dilates with increasing cognitive load. Finally, they also find that the higher the load, the lower the blinking rate. Salminen et al. [11] focuses on correlating confusion with fixation-level features. Their findings indicate a link between eye-tracking metrics and user confusion. In particular, they use a Random Forest and achieve 70% accuracy to predict user confusion using only fixation features such as fixation duration, fixation position, etc. Another work explores changes in pupil size during auditory tasks in response to mistakes [12]. In particular, when subjects listened to somebody counting from 1 to 19, they observed a sharp spike in pupil diameter when a mistake was made by saying a number out of sequence. This research not only illustrates the potential of pupil size as an indicator of cognitive processes, especially in error-related contexts, but also suggests that pupil size could be a valuable feature in our classification task.

b) EEG: Ferrez and Millan [1] investigate Error-Related Potentials (ErrP) in the context of BCIs. They demonstrate the presence of ErrP when a BCI system makes mistakes and propose building single-trial classifiers to distinguish such cases. Chavarriaga and Millan [13] extend the application of ErrP to passive BCIs, where users monitor the system without providing intentional feedback. They illustrate how ErrP can be used to train an external agent to perform better within the system, showcasing its role in enhancing user interactions. Salazar-Gomez et al. [14] employ ErrP to rectify real-time mistakes made by robots. That application demonstrates the practicality of EEG-based error detection for immediate error correction. [9] showcases the utilization of ErrP to assess the accuracy of a classifier designed for a different BCI setup (SSVEP-based). They use ErrP as a means to estimate when the classification is correct and when it is not, effectively calibrating the model online. This research highlights the adaptability of EEG signals in various BCI scenarios. In addition, there is a recent study on importance of EEG signals related to word recognition with applications in medical domain [15].

c) EEG-ET: There is a range of studies that employ both EEG and ET modalities in different applications. [7] investigated multimodal feature fusion approaches for attention classification in hybrid BCIs using deep learning methods. They compared unimodal EEG and eye-tracking classification to multimodal approaches, including early, middle, and late fusion strategies. On the other hand, [8] presents a method that fuses EEG and ET for emotion recognition by using manual feature engineering. These papers highlight that both deep learning and manual feature engineering methods are used and remain competitive in different applications.

III. Methods

A. Dataset

We use the dataset of a recent study at the intersection of BCIs and commonplace interactive systems within digital technologies. For a comprehensive understanding of the dataset, the details of the study can be found in [5].

Experimental Paradigm: The dataset comprises 150 typical English sentences, ranging from 4 to 8 words in length, selected from an existing in-lab dataset. Only one word per sentence is selected for prediction. Each selected word carries a 50% chance of being predicted correctly by the system, which we refer to as a “match.” Conversely, there is an equal 50% chance of an alternative word from the same grammatical category being presented (“mismatch”). Given a suggested word (called also ghost word), the user are asked to press TAB key in case of a match, while in case of mismatches users are required to simply ignore the word and continue typing.

Recording Setup: 32-channel EEG system, utilizing the Brain Products LiveAmp system with active gel electrodes. EEG data is sampled at a rate of 500 Hz, while eye-tracking data is recorded using a Tobii Pro Nano system, sampling eye movements at a rate of 60 Hz. This data was synchronized with the experiment data and computer keyboard using Lab Streaming Layer [16].

Participants: The study includes ten healthy adults with proficiency in English and normal or corrected-to-normal vision. Prior to the study, each participant read and signed a informed consent approved by the an IRB and compliance team at Microsoft Research.

Trials: To gather robust data, each of the ten participants undertakes three runs. During these runs, they engage with 150 different sentences, comprising 50 sentences per run. Consequently, the dataset encompasses a total of 1,500 trials. To ensure data quality, we exclude trials where participants did not press TAB for matches or pressed TAB for mismatches. Also trials with a system lag > 100 msec in displaying the visual stimulus are removed from analysis.

B. Temporal and Spatial Areas of Interest

1) Eye-Tracking Temporal Masking: We mask the eye-tracking data, where specific time points in the data timeseries are replaced with default values based on specific criteria. Here, we explore between sentence masking where only data about participants looking at the sentence on the screen is retained, with any data outside this region being masked with a default value. Similarly, we also explore stricter temporal ghost masking, where data is preserved only when participants gaze at the suggested word (the "ghost") within the sentence. Any data outside this region is masked with a default value. These masking techniques allow us to analyse the feasibility of classifying correct and incorrect word prediction based on the eye movements only while the persons are reading the sentence or the ghost word. An illustration is shown in Fig 1.

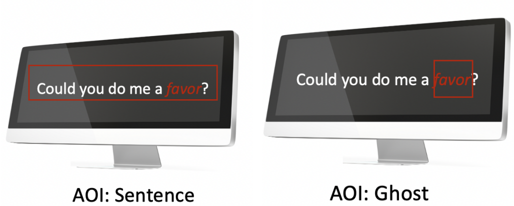

Fig. 1: Sentence and ghost Areas of Interest (AOI) from ET data. When sentence masking is used, only eye movements within the sentence box are kept. Similarly, for ghost masking, only eye movements while the user is looking at the ghost are kept.

a) EEG Spatial Clustering: EEG data acquisition often involves a trade-off due to practical constraints, making it challenging to utilize all 32 electrodes simultaneously. To address this limitation, different electrode clustering configurations are explored. The various clustering options we try are no hair, hair-covered regions, and a combination of both regions. These different clusters allow us to analyse the feasibility of classifying correct and incorrect word predictions based on only specific regions of brain activity collected from a subset of electrodes. See Figure 2 for detailed information of the channels chosen on each cluster.

C. Preprocessing

Eye Tracking Preprocessing: We transform eye tracking data (XY-coordinates on the screen) to relative coordinates. Transforming the data to relative coordinates (X', Y') is achieved by subtracting the coordinates from the position of the text of interest (ghost) as follows: Given the position of the ghost word (X_G, Y_G), then $(X', Y') = (X - X_G, Y - Y_G)$.

EEG Preprocessing: We used EEGLAB [17] and applied several standard preprocessing methods to the EEG data such as filtering, bad channel detection and interpolation, independent component analysis (ICA), and re-referencing to common average. In particular, we performed bandpass filtering between 1–40 Hz and applied a notch filter to eliminate powerline noise at 50 Hz. For bad channel detections, we utilized the CleanRawData function to detect noisy or faulty electrodes and interpolate them based on surrounding electrodes [17]. We experimented with various combinations of these preprocessing methods to assess their impact on the subsequent classification performance.

D. Manual Feature Engineering & Baseline Classifiers

We follow existing literature and include classifiers based on both methods such as manual feature engineering from related literature, as well as deep learning models.

1) Manual Feature Engineering:

- *Fixation & Saccades:* Motivated from [10], for eye tracking, we leverage the pygazeanalyzer [18] to extract fixations and saccades. From these types of eye movements, we derive a set of informative features, including: number of fixations/saccades in the ghost/sentence (4 features), mean/total duration of fixations/saccades in the ghost/sentence (8 features), first/last fixation/saccade in ghost/sentence (8 features).
- *PCPD:* We consider the relative XY-coordinates and the Percentile Change in Pupil Diameter (PCPD) over time, creating a 3-dimensional time series. This time series is then fed into a Random Forest classifier.
- *xDAWN Dimensionalty Reduction:* Following [9], we begin with dimensionality reduction of EEG data using the xDAWN algorithm, resulting in a reduced dataset with only 4 dimensions.

2) Baseline Classifiers:

- *Classical ML methods:* We use as baseline a Random Forest classifier (for eye tracking and EEG) as well as

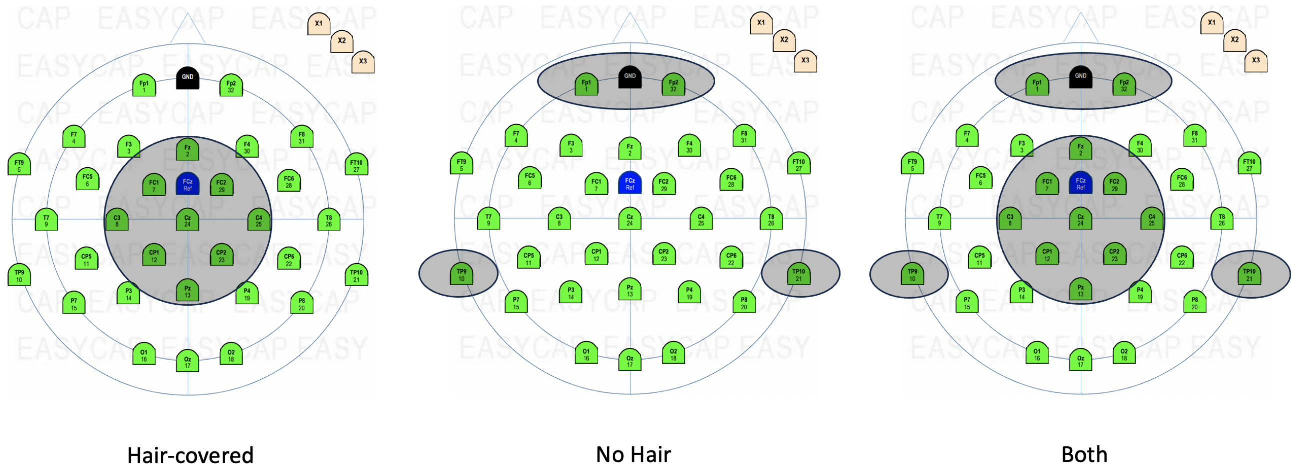

Fig. 2: Spatial Areas of Interest from EEG data: electrodes in the areas covered by hair, no hair, and a combination of both.

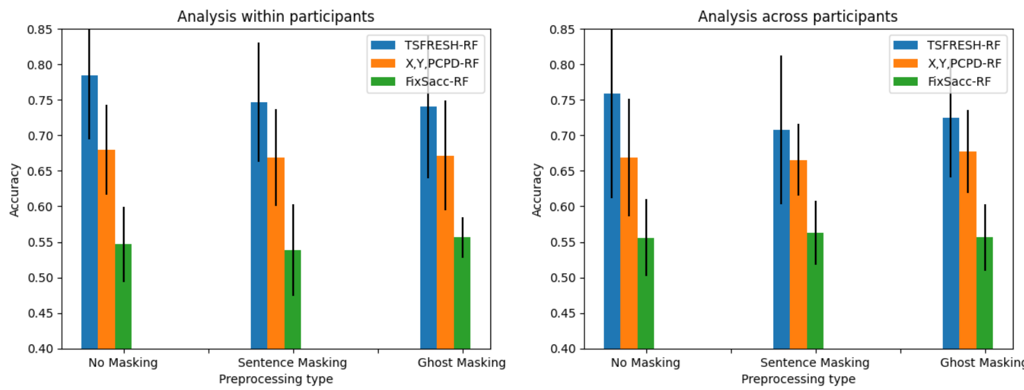

Fig. 3: Eye Tracking experiments (within and across participants). All models make use of Random Forest (RF) classifier, with the main difference in (1) TSFresh-RF we use automated feature engineering pipeline from TSFresh library to extract the features, (2) X,Y,PCPD-RF we manually extract the percentile change of pupil diameter and feed the raw X,Y eye movements to Random Forest Classifier and (3) FixSacc-RF we manually extract 20 features related to fixations and saccades.

Linear Discriminant Analysis (LDA) classifier (for EEG motivated from [19]). These are two typical classical ML methods used in eye tracking and EEG where features are manually extracted.

- *CNN:* We also include a simple Convolutional Neural Network consisting of only two layers, followed by a fully connected layer in our baselines. We train the model using binary cross-entropy loss.
- *LSTM:* Similar to the previous model (CNN), we also insert an LSTM building block between the last convolution layer and the subsequent fully connected layer to capture time dependencies in the signal.
- *Transformer:* In our baseline models, we also include a model that is attention-based. We perform two types of attentions — through electrodes and through time. We use a transformer encoder building block for both attention building blocks and then concatenate the embeddings before feeding them to a fully connected neural network.

E. Automated Feature Engineering & Multi-modal Classifier

Instead of extracting specific features or using the raw timeseries data, we utilize the TSFresh library [20] to extract tens of thousands of general-purpose but interpretable features. In addition, we run a permutation feature importance analysis, which allows us to determine which of these thousands of features are important in our setup. Finally, for classification, we use a Random Forest classifier with the most important features. We also test a multi-modal model, where we leverage the features extracted using TSFresh from both ET and EEG data and simply concatenate them before feeding them to the permutation feature importance (PFI) pipeline or Random Forest classifier. The whole pipeline can be run automatically and does not require any manual feature engineering.

IV. Results

A. Eye tracker analysis

Fig. 3 shows the key findings from our eye-tracking experiments conducted under three distinct masking scenarios. Con-

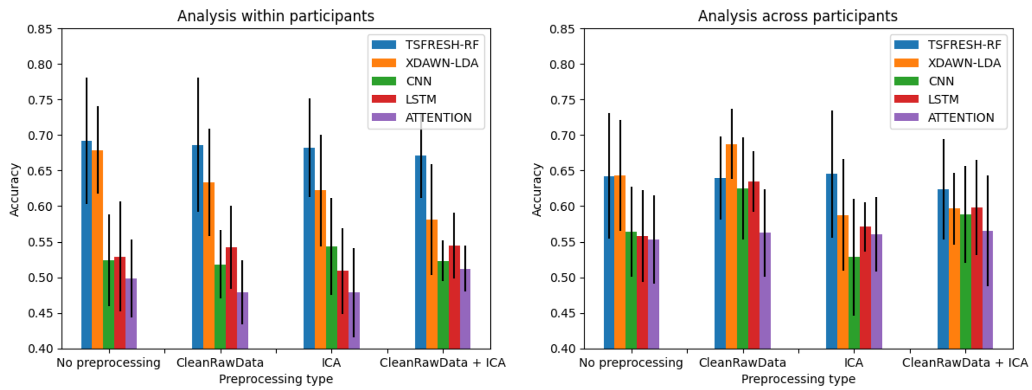

Fig. 4: EEG Experiments (within and across participants) with different classifiers across different preprocessing methods. TSFresh-RF represents the model based on automated feature engineering and a random forest (RF) classifier. In the xDAWN-LDA baseline, we perform dimensionality reduction using xDAWN algorithm and use the extracted features in an Linear Discriminant Analysis (LDA) classifier. The CNN, LSTM, and Attention models are deep learning based models.

ducted inter and across participants, the results consistently indicate that the TSFresh-RF (automated feature engineering with a random forest classifier) model outperforms other models across these scenarios. Particularly noteworthy is that the highest accuracy is consistently achieved when no masking is applied to the eye-tracking data.

Further, we performed a permutation feature importance (PFI) analysis on the TSFresh-RF model. The analysis revealed that the most critical feature for classification was the maximum value of the X-coordinate of eye movement. Since the XY-coordinates of eye movements is transformed relative to the ghost word, the maximum value of the X-coordinate represents how far the participant looked on the right of the start position of the ghost word. This observation suggests that the extent to which participants scan the ghost word directly correlates with its classification as a match or mismatch. We provide a visual representation of the relationship between word length (in characters) and scan length (in seconds) for ghost words in Figure 5. Notably, for mismatches, the scan length does not increase with word length, confirming our previous observation. This indicates that participants tend to simply stop from reading incorrect ghost words until the end. The TSFresh-RF model successfully utilizes this information, achieving an accuracy of approximately 78% within participants and approximately 76% across participants.

B. EEG analysis

In Figure 4, we present the results for all EEG models under different EEG preprocessing methods, both within and across participants. The experiments for within participants demonstrate that the TSFresh-RF model consistently delivers the best performance. Similar patterns are observed across participants, except for the case of CleanRawData, where the XDAWN-LDA model performs better. Surprisingly, preprocessing the EEG data did not result in performance improvements. In fact, combining CleanRawData and ICA led to a decrease in model accuracy. To this end, we conducted a permutation

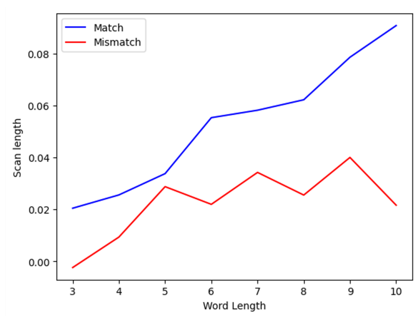

Fig. 5: The relation between word length and scan length. This insight was automatically discovered from the features extracted from TSFresh library.

feature importance (PFI) analysis on the EEG model. The analysis revealed that the most important electrode was FP2. This observation suggests that the model might be utilizing information similar to that of the eye tracker, as FP2 is closely positioned to the eye and captures eye movements. Notably, preprocessing, especially ICA, removes these artifacts, leading to a degradation in performance.

C. Eye tracker and EEG analysis

In Figure 6, we present the results for the hybrid model (TSFresh-RF) under different preprocessing types, electrode clusters, and eye-tracking data masking scenarios. Within participants, the model achieves 81% accuracy when both CleanRawData and ICA are applied to EEG data, and the combined cluster of EEG electrodes is used without any masking on the eye-tracking data. This differs from the EEG-only setup, where preprocessing did not improve performance due to the removal of eye movement information by ICA. However, in the hybrid model, preprocessing enhances performance, as the eye movement information is already present in the eye-tracking data, making the cleaned EEG data complementary. Similar patterns are observed for across participants, although

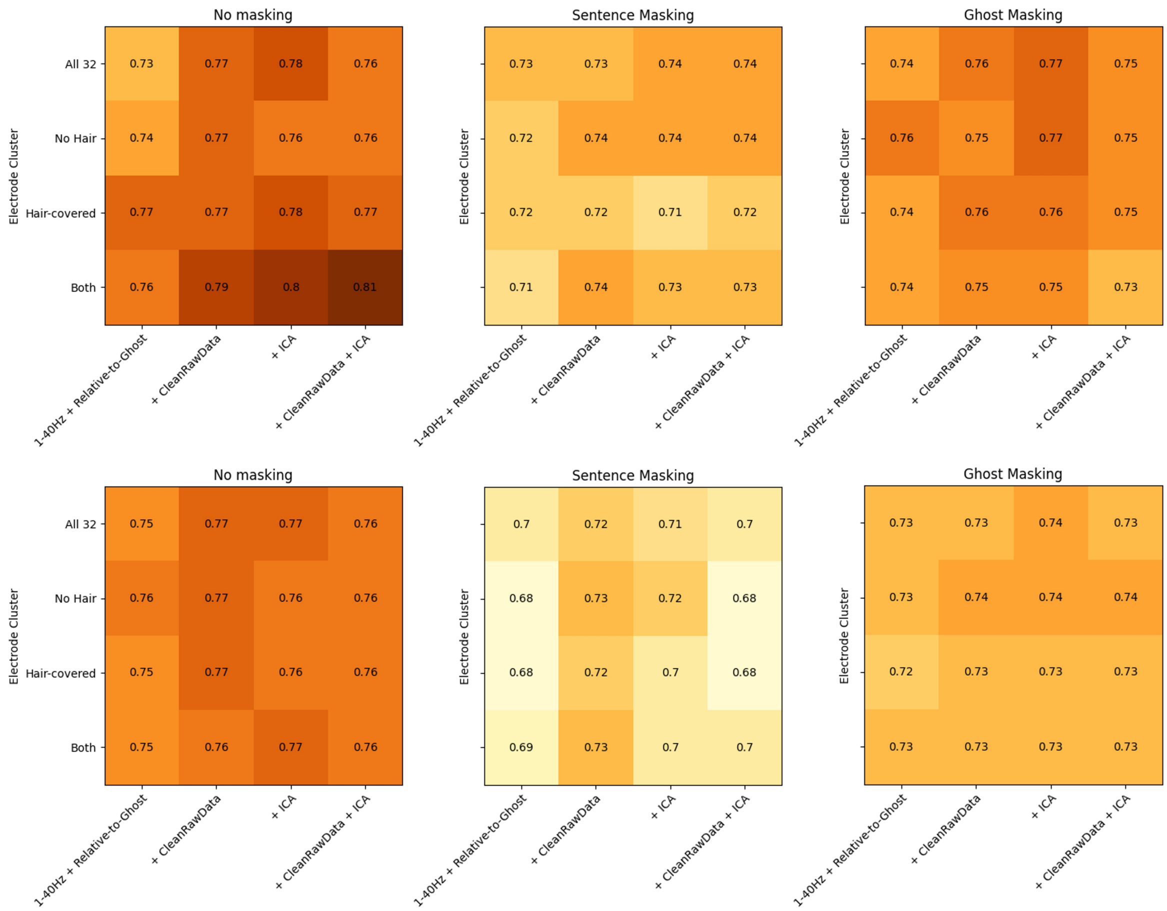

Fig. 6: Hybrid Model (EEG-ET) Experiments for each preprocessing type, for each cluster type, for each masking type. First row = within participants, second row = across participants.

here, CleanRawData or ICA alone perform slightly better than their combination. The best accuracy achieved across participants with existing methods is 77%.

V. Conclusion

In this study, we developed a multimodal classifier combining EEG and eye-tracking (ET) data for a BCI application involving predictive text systems. We demonstrated the feasibility and utility of automated feature engineering in BCI setups, particularly during the initial stages of data exploration when theoretical frameworks are still evolving and data availability is limited. We showed that pipelines with automated feature engineering can be effectively utilized in BCI contexts, offering a practical and efficient tool for generating and selecting relevant features without the need for extensive manual effort or deep theoretical insights.

This approach yielded valuable insights, including the automatic detection of behavioral patterns such as participants not fully reading incorrectly predicted words. Furthermore, we discovered that integrating ICA-preprocessed EEG data with ET data enhanced classification performance by allowing brain features to complement eye-tracking information. In contrast, in the single-modality EEG-only case, ICA inadvertently removed informative eye movement artifacts, reducing classification accuracy. The inclusion of ET data mitigated this issue, enabling cleaner brain activity signals to synergize with eye-tracking data for improved predictions.

Although our findings highlight the benefits of automated feature engineering in early-stage BCI research, we acknowledge its limitations. Automated feature engineering is not intended to replace existing methods or achieve superior performance in all scenarios. We anticipate that deep learning models, supported by larger datasets and well-developed theoretical frameworks, will outperform automated feature engineering approaches in more mature research contexts. However, the ease of implementation and the ability to derive meaningful insights make automated feature engineering a valuable addition to the BCI research toolkit, particularly for novel paradigms and initial explorations. Automated feature engineering is a standard approach in many domains related to timeseries classification, but, surprisingly, it is not commonly applied in the BCI context. Our study demonstrates that it is effective for data-driven exploration and competitive single-trial classifier development in multimodal setups.

References

[1] P. W. Ferrez and J. del R. Millan, "Error-related eeg potentials generated during simulated brain–computer interaction," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 3, pp. 923–929, 2008.

[2] J. Ha, S. Park, and C.-H. Im, "Novel hybrid brain-computer interface for virtual reality applications using steady-state visual-evoked potential-based brain–computer interface and electrooculogram-based eye tracking for increased information transfer rate," *Frontiers in Neuroinformatics*, vol. 16, 2022. [Online]. Available: https://www.frontiersin.org/journals/neuroinformatics/articles/10.3389/fninf.2022.758537

[3] K. Liu, Y. Yu, and Z. Zhou, "A novel bci paradigm combined with augmented reality," in *2021 China Automation Congress (CAC)*, 2021, pp. 53–56.

[4] A. Nijholt, R. J. K. Jacob, M. Andujar, B. F. Yuksel, and G. Leslie, "Brain-computer interfaces for artistic expression," in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–7. [Online]. Available: https://doi.org/10.1145/3170427.3170618

[5] S. K. Mehdizadeh, E. Cutrell, R. M. Winters, N. Djuric, Y. Cheng, I. J. Tashev, and Y. T. Wang, "Eeg and eye-tracking error-related responses during predictive text interactions: A bci case study," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 2023, pp. 1–4.

[6] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for eeg-based brain-computer interfaces: A 10-year update," *Journal of Neural Engineering*, vol. 15, 02 2018.

[7] L.-M. Vortmann, S. Ceh, and F. Putze, "Multimodal eeg and eye tracking feature fusion approaches for attention classification in hybrid bcis," *Frontiers in Computer Science*, vol. 4, 2022. [Online]. Available: https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2022.780580

[8] W.-L. Zheng, B.-N. Dong, and B.-L. Lu, "Multimodal emotion recognition using eeg and eye tracking data," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 5040–5043.

[9] K.-J. Chiang, D. Emmanouilidou, H. Gamper, D. Johnston, M. Jalobeanu, E. Cutrell, A. D. Wilson, W. An, and I. Tashev, "A closed-loop adaptive brain-computer interface framework: Improving the classifier with the use of error-related potentials," in *International IEEE/EMBS Conference on Neural Engineering*, IEEE. IEEE, May 2021.

[10] J. Zagermann, U. Pfeil, and H. Reiterer, "Measuring cognitive load using eye tracking technology in visual computing," in *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, ser. BELIV '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 78–85. [Online]. Available: https://doi.org/10.1145/2993901.2993908

[11] J. Salminen, M. Nagpal, H. Kwak, J. An, S.-g. Jung, and B. J. Jansen, "Confusion prediction from eye-tracking data: Experiments with machine learning," in *Proceedings of the 9th International Conference on Information Systems and Technologies*, ser. ICIST '19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3361570.3361577

[12] J. Klingner, R. Kumar, and P. Hanrahan, "Measuring the task-evoked pupillary response with a remote eye tracker," in *ETRA '08: Proceedings of the 2008 symposium on Eye tracking research applications*. New York, NY, USA: ACM, 2008, pp. 69–72.

[13] R. Chavarriaga and J. d. R. Millan, "Learning from eeg error-related potentials in noninvasive brain-computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 4, pp. 381–388, 2010.

[14] A. F. Salazar-Gomez, J. DelPreto, S. Gil, F. H. Guenther, and D. Rus, "Correcting robot mistakes in real time using eeg signals," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 6570–6577.

[15] H. Jang, S.-K. Kim, J. Ha, and L. Kim, "Early detection of alzheimer's disease through analysis of eeg responses to word recognition," in *2024 12th International Winter Conference on Brain-Computer Interface (BCI)*, 2024, pp. 1–4.

[16] "Lab streaming layer," https://github.com/sccn/labstreaminglayer, [Online; accessed 2-February-2024].

[17] A. Delorme and S. Makeig, "Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.

[18] E. Dalmaijer, S. Mathôt, and S. Stigchel, "Pygaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments," *Behavior Research Methods*, vol. 46, 11 2013.

[19] B. Blankertz, C. Schäfer, G. Dornhege, and G. Curio, "Single trial detection of eeg error potentials: A tool for increasing bci transmission rates," in *Artificial Neural Networks — ICANN 2002*, J. R. Dorronsoro, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 1137–1143.

[20] "Tsfresh," https://tsfresh.readthedocs.io/en/latest/, [Online; accessed 2-February-2024].