COMPARING FINGERS AND GESTURES FOR BCI CONTROL USING AN OPTIMIZED CLASSICAL MACHINE LEARNING DECODER

D. Keller[1], M. J. Vansteensel[1], S. Mehrkanoon[2], M. P. Branco[1]

[1]UMC Utrecht Brain Center, Department of Neurology and Neurosurgery, Utrecht, The Netherlands
[2]Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands

E-mail: m.pedrosobranco@umcutrecht.nl

ABSTRACT: Severe impairment of the central motor network can result in loss of motor function, clinically recognized as Locked-in Syndrome. Advances in Brain-Computer Interfaces offer a promising avenue for partially restoring compromised communicative abilities by decoding different types of hand movements from the sensorimotor cortex. In this study, we collected ECoG recordings from 8 epilepsy patients and compared the decodability of individual finger flexion and hand gestures with the resting state, as a proxy for a one-dimensional brain-click. The results show that all individual finger flexion and hand gestures are equally decodable across multiple models and subjects (>98.0%). In particular, hand movements, involving index finger flexion, emerged as promising candidates for brain-clicks. When decoding among multiple hand movements, finger flexion appears to outperform hand gestures (96.2% and 92.5% respectively) and exhibit greater robustness against misclassification errors when all hand movements are included. These findings highlight that optimized classical machine learning models with feature engineering are viable decoder designs for communication-assistive systems.

INTRODUCTION

Dysfunction of the neuromotor system may precipitate transient or, in severe cases, enduring global loss of motor control. Global dysfunction may be referred to as Locked-In Syndrome (LIS) [1], often characterized by quadriplegia and aphonia. In recent decades, efforts to replace dysfunctional motor control have seen pioneering developments in Brain-Computer Interfaces (BCIs) [2]. BCIs extract information directly from cortical activity to control mechanical or digital effectors without relying on neuromuscular activation, essentially bypassing the muscular output. Restoration of effector control can serve several purposes, ranging from object manipulation [3–5], locomotion and mobility [6] and speech production [7–9]. However, for individuals with severe impairment, the restoration of the communicative agency has been identified as one of the most urgent needs [10].
A simple approach towards communication BCI is automatic letter selection on a digital keyboard [11]. An attractive signal recording modality for communication BCIs is electrocorticography (ECoG) due to its high spa-

tiotemporal precision, good signal-to-noise ratio, and reliable signal stability over extended periods [11, 12]. Several studies have demonstrated that hand movement recognition from ECoG recordings can be performed with high accuracy. Consequently, a variety of hand movements have been explored for this purpose, including but not limited to finger flexion [13–17], reaching and grasping [3–5, 18, 19], and wrist flexion and extension [15], more complex hand gestures [15, 20–22] and handwriting [23]. To identify a reliable motor signature for a unidimensional BCI control signal (i.e., 'brain-click') many studies have examined different types of hand movements in isolation [13, 14, 20–22], but few have compared different hand movements against each other within a unified framework.
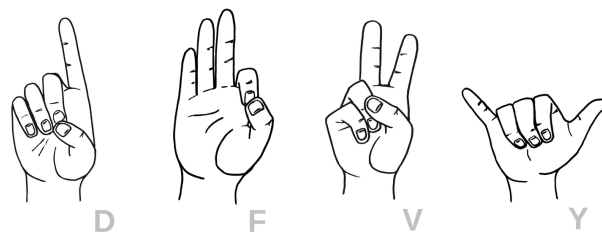

Figure 1: The four gestures executed by subject S1 - S5.

The goal of this work is to contribute to a deeper understanding of the decodability of individual finger flexion and hand gestures against the resting state. Specifically, we explored which hand strategy is the most promising for a reliable brain-click and which is more transferable across subjects. In addition, we aim to extend our analysis to the prospect of multidimensional control with four and eight degrees of freedom (DoF) to investigate which type of hand movement intrinsically yields a better within and across-category discriminability. To overcome the notorious data sparsity in this domain, we employed an optimized feature selection decoder with different classification models and assessed which (offline) machine learning approach yields the best performance on individual hand movements (2-DoF), within hand movement types (4-DoF), and within all hand movements (8-DoF).

MATERIALS AND METHODS

Data description: This study is based on two ECoG data sets, consisting of finger flexion, including the index finger, little finger, and thumb, and hand gestures associated with the American Sign Language letters D, F, V, and Y [21, 24] (see Figure 1).

The data were collected from 8 subjects ($N_G = 5$, $N_F = 6$) in an epilepsy monitoring unit of the University Medical Center Utrecht (see Table 1). These subjects had 32, 64 or 128 high-density subdural ECoG electrodes with an inter-electrode distance of 3 or 4mm and an exposed diameter of 1 to 1.3mm (AdTech, Racine, USA; or PMT Corporation, Chanhassen, MN, USA) implemented over the hand-knob region of the sensorimotor cortex. The ECoG data were recorded using a 128-channel Micromed LTM system (subjects 1 - 5; Treviso, Italy; 22 bits, hardware bandpass filter 0.15–134.4Hz; sampling frequency 512Hz) and a Blackrock system (subjects 6 - 8; Microsystems LLC, Salt Lake City, USA, digital bandpass filter 0.3 - 500Hz; sampling frequency 2000Hz). Data were converted to the BIDS standard format [25].

Table 1: Subject Details

Subject	Task	Trials (C / T)	Age	Sex	Hand	Handedness	Hemisphere	Grid (incl.)
S1	G F	37 / 74 90 / 181	19	F	Right	Right	Left	4x8 (32)
S2	G	68 / 138	45	F	Left	Left	Right	8x8 (59)
S3	G	34 / 69	29	M	Right	Right	Left	4x8 (29)
S4	G F	32 / 67 90 / 181	19	M	Right	Right	Left	4x8 (31)
S5	G F	34 / 69 88 / 177	42	M	Right	Right	Left	4x8 (32)
S6	F	89 / 179	30	F	Left	Right	Right	16x8 (123)
S7	F	85 / 171	20	F	Right	Right	Left	8x8 (64)
S8	F	84 / 169	36	F	Right	Right	Left	16x8 (128)

Note. Trials are presented as the ratio of hand movement trials per condition (C) out of all trials (T; including the rest trials). In Grid, (incl.) indicates the number of channels included. Abbreviations: Gesture, G; Finger, F, Male, M; Female, F.

Experimental Design: Subjects were instructed to initiate movements based on visual cues that were presented in a randomized, event-driven design. For the gestures, the subjects imitated the depicted gesture after stimulus onset and maintained the posture until the end of the trial before returning to a resting position. Each subject performed 10 trials with an intertrial interval of 4.4s and a run duration of 6.7m. Rest trials were implicitly calculated from a small time interval before the onset of the next movement. For finger flexion, the subjects performed two finger flexions immediately after cue onset and then returned to a resting position afterward. In contrast to the gestures, each movement was interleaved with an explicit resting trial. The design consisted of 30 trials with an intertrial interval of 7s and a run duration of 8.2m. In both experiments, each subject performed the tasks with the hand contralateral to the grid location, and subject 2 performed the task twice. In addition, a data glove (5DT, Irvine CA, USA, 20 ms sampling time) was used during both experiments to record motor activity.

Preprocessing: Data preprocessing included the removal of bad trials and channels (identified by [21, 24] based on data glove data and raw signal inspection), followed by common average referencing, notch, and bandpass filtering (56 Hz - 130 Hz) to remove artifacts. Finger flexion data, sampled at 2000 Hz, were downsampled to 512 Hz for consistency across subjects. The data were then subsequently aligned with movement onset markers obtained from data glove recordings and segmented accordingly. For decoding individual hand movements and within hand movement types a segmentation window of $W_F = [-0.5, 1.5s]$ and $W_G = [-0.5s, 2.5s]$ was used. For decoding all hand movements, the two 4-DoF settings for subject 5 were combined with a common segmentation window of $W_{FG} = [-0.5s, 2s]$. In all three settings $t = 0$ represents the motion-aligned stimulus presentation.

Features were extracted using a continuous Morlet wavelet transformation, which produced spectral power features for the high-frequency band (60 Hz - 126 Hz) in 2 Hz frequency bins. To reduce the feature space the power was averaged and the time dimension was decimated to $T_G = 154$ and $T_F = 102$ time points per channel for the fingers and gestures, respectively. The resulting feature vectors were used for subsequent model training. Preprocessing was conducted in Python (v3.9) using the MNE library (v1.16).

Decoder: The architecture of the decoder, depicted in Figure 2, revolves around an optimized data-driven feature engineering approach for conventional classical machine learning classifiers. The decoder encompasses four modules: Normalization, Incremental Feature Selection (IFS), Feature Reduction (FR), and Classification.

The initial step of the decoder normalises the spectral power of the spatio-temporal feature vector (N) via a Box-Cox transformation [26], followed by mean centering and unit variance scaling to ensure data normality and variance stabilisation. Feature selection employs Variance Thresholding (VT), Mutual Information Criterion (MIC) [27, 28], and Recursive Feature Elimination (RFE). RFE iteratively removes a set of features (S = 1e-3 * N) corresponding to the least important coefficients, and akin to MIC, retains a subset of the best K temporal features across all channels. Each selection method generates a binary mask indicating the retained features. The individual modules are applied incrementally, with the specific combination and its parameter configuration being delegated to a Bayesian optimization algorithm, which avoids manual tuning and efficiently navigates through the high-dimensional parameter space. In particular, this approach aims to balance the advantages of filter and wrapper methods [29] to remove noisy and redundant features while prioritising discriminative ones. The Feature Reduction (FR) method can be applied in isolation or in conjunction with IFS processing. In the classification phase, four algorithms were selected to compete with each other: Logistic Regression (LR), Linear Discrim-
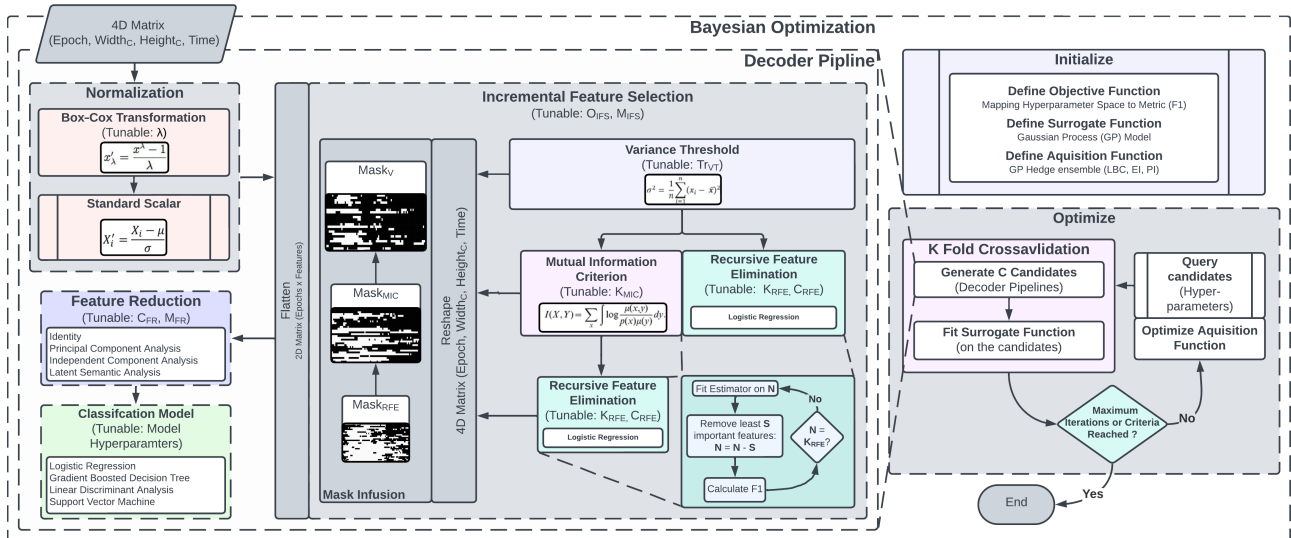
Figure 2: General decoder architecture. The pipeline is composed of four modules: Normalization, Incremental Feature Selection (IFS), Feature Reduction (FR), and a Classification Model. A Box-Cox (BC) transformation and a Standard Scalar normalize the signal. The incremental selection procedure can recruit different combinations of Variance Thresholding (VT), Mutual Information Criterion (MIC), and Recursive Feature Elimination (RFE) to select the relevant features from the spatio-temporal representation. RFE iteratively removes the set of S least important features from the pool N until it reaches K_{RFE} features. After feature selection, the feature space can be further compressed with various Feature Reduction (FR) methods. The resulting vector is forwarded to one of four classifiers. A Bayesian optimization algorithm orchestrates the order and method of feature selection (O_{IFS}, M_{IFS}), other pipeline configuration (λ_{BC}, Tr_{VT}, K_{MIC}, K_{RFE}, C_{RFE}, C_{FR}), and various model-specific hyperparameters. Dotted lines represent meta-routing processes; single and double-lined boxes represent dynamic and predefined processes, respectively. Abbreviations: LBC, Lower Confidence Bound; EL, Negative Expected Improvement; PI Negative Probability Improvement

inant Analysis (LDA), Boosted Decision Trees (BDT), and Support Vector Machines (SVM). A Majority Class Predictor, which predicts the most frequent hand movement, was evaluated on the data to establish a 'chance' baseline. Performance is assessed using the F1 scoring metric, adjusted for label imbalance with the inversely weighted class distribution.

For this 'black-box' optimization problem, the Bayesian algorithm [30] approximates an expensive non-smooth objective function by inference, essentially guiding the search process based on prior results. To find an optimal decoder candidate in the large parameter space, a Gaussian Process model [31] with a hedging portfolio strategy [32] is used, where hedging probabilistically choose the best acquisition function from three candidates: Lower Confidence Bound, Negative Expected Improvement or Negative Probability Improvement. The search is restricted to a maximum of 256 candidates, exploring a hyperparameter space, ranging from 9 (LDA) to 20 (BDT) configurations for different algorithms, of which up to 60% are conditional hyperparameters; the number-of-components hyperperameter for the FR step was shared among all three methods. Model performance is evaluated using stratified 10-fold crossvalidation, with the best candidate further assessed through leave-one-out crossvalidation. The decoder pipeline adheres to the scikit-learn architecture, ensuring compatibility with the scikit-learn library and its derivatives. The implementation is in Python 3.9, using scikit-learn (v1.4.0) and xgboost (v2.0.3).

Statistical Analysis: The analysis relies on a Friedman ANOVA to identify a general effect and Dunn's test with Benjamini-Hochberg's false discoveries rate correction the post hoc analysis and pairwise comparison. The statistical analysis was performed in Python 3.9, using scipy (v1.13.0) and scikit-posthocs (v0.9.0).

RESULTS

Individual Hand Movements: In the context of individual hand movements, all classification models exhibited a high F1 performance (averaged across subjects), exceeding 98.0% for each finger flexion and gesture (details summarised in Table 2 and Figure 3 A). Notably, the Index finger (99.59%), Gesture V (99.11%), and Gesture F (98.09%) were the most promising candidates for brain-click BCI control. Interestingly, they share a commonality in index finger flexion. Within the hand gestures, no gesture significantly outperformed the others (Friedman ANOVA and Dunn's test; ns). Similarly, within the finger flexions, the overall effect was significant, $\chi^2(3) = 15.32$, p = $4.71e^{-4}$), but no finger flexion was significantly different from the others (Dunn's test). Moreover, decoding performance remained remarkably stable for each subject across all hand movements and models, with consistent trends in variability observed for each subject (i.e., subject 1 consistently had the lowest and highest scores for the fingers and the gestures, respectively).

Hand Movements Types: For the hand movement types, a different trend emerged. For 4-DoF classifica-

Table 2: Mean performance across all subjects of the Optimize Feature Selection Decoder for Individual Movements (2-DoF), Within Types (4 DoF) and Within all Hand Movements (8-DoF).

Models	Fingers (vs. rest)						Gestures (vs. rest)								Multi-DoF (incl. rest)					
	Index		Little		Thumb		Gesture D		Gesture F		Gesture V		Gesture Y		4-Finger		4-Gesture*		8-Hand**	
	A	F1	A	F1	A	F1	A	F1	A	F1	A	F1	A	F1	A	F1	A	F1	A	F1
Chance	75.3	64.6	75.5	64.9	74.9	64.2	83.3	75.7	79.8	70.8	79.4	70.3	79.8	70.9	50.3	33.7	56.6	40.9	50.4	33.8
LR	98.2	98.3	98.3	98.4	95.9	96.0	97.4	97.5	98.2	98.2	98.7	98.7	96.6	96.8	95.7	95.7	91.3	91.5	87.4	87.7
BDT	99.6	**99.6**	98.1	98.1	98.1	**98.0**	97.6	97.7	98.7	98.7	98.4	98.5	98.2	98.2	95.8	95.8	92.0	92.0	88.6	88.7
LDA	99.3	99.3	98.7	**98.8**	97.5	97.6	98.1	**98.1**	98.7	98.7	98.0	98.0	96.4	96.6	96.1	96.1	91.1	91.1	93.1	**92.9**
SVM	99.5	99.4	98.8	**98.8**	97.7	97.6	97.9	98.0	98.8	**98.89**	99.1	**99.1**	98.4	**98.4**	96.7	**96.6**	92.5	**92.5**	92.7	92.6

Note. Values are in %. *The 4-DoF gesture decoding includes Gesture F, Y, and V. ** The 8-DoF decoding of all hand movements was only obtained from subject 5. Abbreviations: Accuracy, A; Versus, vs; Inclusive, incl.
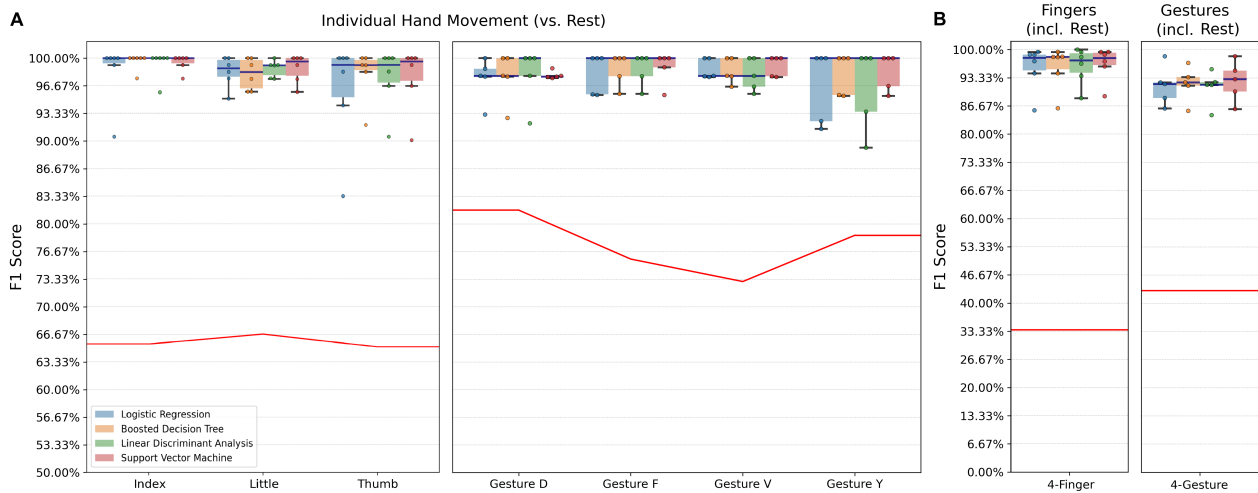

Figure 3: Box plots for different machine learning models of the decoder for (A) all individual fingers and (B) the three multi-DoF comparisons. Each point represents one subject, with the upper and lower error bars representing an interquartile range of 25 and 75, respectively, and where omitted when performance for one or more subjects exceeded this range. The 4-DoF gesture decoding includes Gesture F, Y, and V. The red line represents the highest chance level among all subjects.

tion, finger flexion (95.7% - 96.6%) outperformed the hand gestures (91.1% - 92.5%; average across subjects), which could be statistically verified (Dunn's test), p = 0.031; we excluded the worst decodable gesture (Gesture D) to ensure a similar task complexity. An extension to 8-DoF classification preserves a high F1 score for Subject 5 (88.7% - 92.9%), with remarkably minimal confusion between gestures and fingers. On visual inspection, fingers exhibit more confusion with the resting state, while gestures are more often confused among themselves (as depicted in Figure 4).

Classification Models: When training classification models within an optimised feature engineering framework, no model emerges as significantly superior to the others (Friedman ANOVA;), $\chi^2(4) = 60.48$, p = $2.3e^{-12}$, (Dunn's test; ns), although, all perform significantly above chance level (Dunn's Test), p < $1.28e^{-8}$. In general, Boosted Decision Trees (BDT) and Support Vector Machines (SVM) demonstrate the highest classification performance across all conditions, except for Little Finger and Gesture D.

DISCUSSION

The current work demonstrates that optimised spatio-temporal feature engineering of finger flexion and hand gestures, recorded from high-density ECoG, enables reliable decoding for one-dimensional brain-click, 4- and 8-DoF decoding tasks, even with very small data volumes. Notably, within each category, no single hand movement emerged as superior decodable. However, upon qualitative inspection, the index finger and Gestures V and F, all sharing index finger flexion, appeared as the most promising candidates. Moreover, the fingers exhibited a better performance in more complex 4- and 8-DoF decoding; finger flexion may possess more discriminative properties for multi-DoF tasks. Finally, no classical machine learning model outperformed the others, but BDT and SVM may have a small advantage.

Hand gestures entail a more complex interplay of motor components than simple finger flexion, including wrist and finger flexion, and lateral extension, among others. However, our results revealed that individual finger flexion alone yields a near perfect neuroelectrical signature.
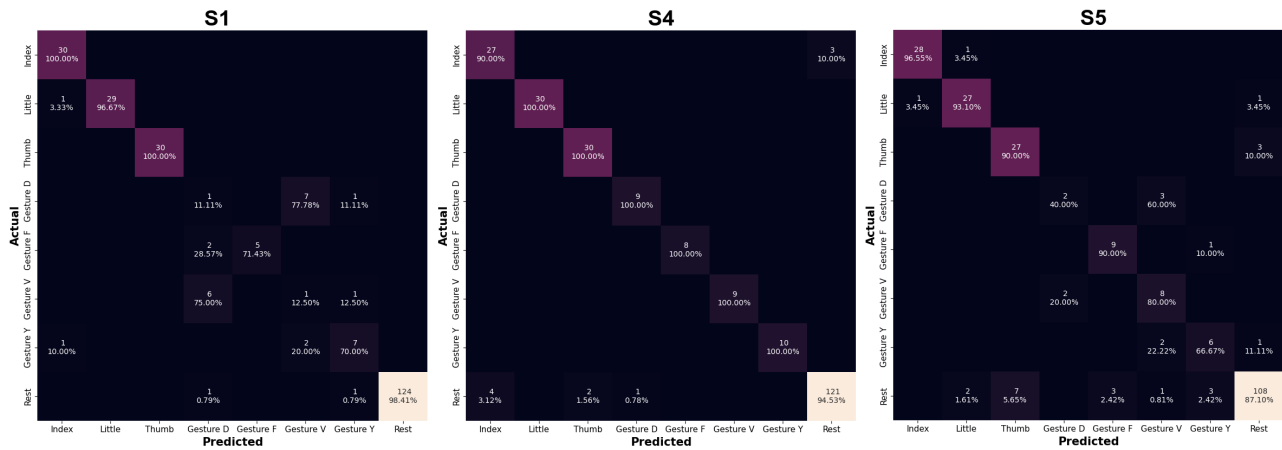
Figure 4: Confusion matrices of the LR model for subjects 1, 4, and 5. In each confusion matrix, the horizontal axis represents the predicted hand movement (or rest), and the vertical axis represents the ground truth hand movement (or rest). Henceforth, the diagonal elements represent the total correct values predicted per hand movement. The colour coding represents the proportion of absolute counts for each hand movement ranging from black (0%) to white (100%).

Consequently, the addition of supplementary motor components may not increase decodability but rather confront the decoder with a motor signature that inherently has a higher variability in its signal. As we evaluated the decoder on small data volumes, an effect of additional components on decoder performance may emerge with larger sample sizes. Extending this rationale, gestures might possess a more intricate spatial and temporal pattern, recruiting various neuronal populations from a larger effector field in the sensorimotor cortex. The absence of advanced feature transformation techniques in our classical machine learning approach might hinder the decoder from exploiting the full potential of the gesture's electrical signature. In addition, differences in sample size, with fewer than 30 trials for fingers and only 10 trials for gestures, may impact the comparison, while the overall small sample size may not provide sufficient power to detect a potential statistical difference.

In line with the evident discernible difference between the 4-DoF types, misclassification errors were more pronounced for gestures, with frequent confusion among different fingers and the resting state. Notably, substantial confusion between hand movements and the resting state suggests a potential contamination of the rest periods with unintended movement. This may be attributed to the design of the experimental design of the gestures, which lacked separate explicit rest trials. Incorporating a threshold based on data glove recordings could be beneficial. However, defining true rest may not be practically feasible for real BCI applications, as it necessitates extensive subject training to suppress such activity [11, 12, 19], and a more naturalistic approach would be to build a decoder that can successfully discriminate between meaningful and non-task-related sporadic motor activity. Furthermore, some features along the temporal dimension may not reflect actual motor activity, but 'resting' activity, especially before the movement onset and towards the end of the segmentation window.

The proposed decoder design surpassed prior approaches evaluated on gesture data for four out of five subjects [20, 21], albeit a direct comparison is difficult due to differences in task complexity. For communication assistance systems, the proposed decoder design can offer a viable alternative to deep learning approaches for one-dimensional brain-click tasks [19] and even larger DoF applications [14, 17], where data acquisition is challenging. Although trained offline, the decoder can process individual segments of preprocessed data in as little as 2 - 10ms. Consequently, although theoretically deployable in an online setting, regular offline retraining on new data is necessary to address concept drifts for ensuring long-term stability, in particular for individuals with neurodegenerative diseases. Importantly, the experiments involved movement execution by epileptic individuals, further validation in attempted movement is imperative to extend applicability to online BCI settings for individuals with LIS.

CONCLUSION

Electrocorticography data provides a high-resolution spatiotemporal feature representation, forming a suitable foundation to tailor an optimised classical machine learning decoder with automatic feature engineering to the large feature space. We demonstrate that within this framework, both finger flexion and hand gestures enable reliable decoding across multiple subjects, and when extended to a multiple degrees of freedom, maintain high discriminability between hand movements. For click-based letter selection in communication-assistive BCI systems, the index finger flexion emerges as an optimal candidate. Moreover, all tested models consistently exhibit high classification performance across multiple subjects - a comparable performance to deep learning approaches.

REFERENCES

[1] Plum F, Posner JB. The diagnosis of stupor and coma. Oxford University Press, USA (2000).

[2] Thakor NV. Translating the brain-machine interface. Science translational medicine. 2013;5(210):210ps17–210ps17.

[3] Hochberg LR *et al.* Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. Nature. 2012;485(7398):372–375.

[4] Collinger JL *et al.* High-performance neuroprosthetic control by an individual with tetraplegia. The Lancet. 2013;381(9866):557–564.

[5] Bouton CE *et al.* Restoring cortical control of functional movement in a human with quadriplegia. Nature. 2016;533(7602):247–250.

[6] Lorach H *et al.* Walking naturally after spinal cord injury using a brain–spine interface. Nature. 2023:1–8.

[7] Moses DA *et al.* Neuroprosthesis for decoding speech in a paralyzed person with anarthria. New England Journal of Medicine. 2021;385(3):217–227.

[8] Willett FR *et al.* A high-performance speech neuroprosthesis. Nature. 2023;620(7976):1031–1036.

[9] Metzger SL *et al.* A high-performance neuroprosthesis for speech decoding and avatar control. Nature. 2023;620(7976):1037–1046.

[10] Kageyama Y *et al.* Nationwide survey of 780 japanese patients with amyotrophic lateral sclerosis: Their status and expectations from brain–machine interfaces. Journal of Neurology. 2020;267:2932–2940.

[11] Vansteensel MJ *et al.* Fully implanted brain–computer interface in a locked-in patient with als. New England Journal of Medicine. 2016;375(21):2060–2066.

[12] Pels EG *et al.* Stability of a chronic implanted brain-computer interface in late-stage amyotrophic lateral sclerosis. Clinical Neurophysiology. 2019;130(10):1798–1803.

[13] Xie T, Zhang D, Wu Z, Chen L, Zhu X. Classifying multiple types of hand motions using electrocorticography during intraoperative awake craniotomy and seizure monitoring processes—case studies. Frontiers in neuroscience. 2015;9:353.

[14] Xie Z, Schwartz O, Prasad A. Decoding of finger trajectory from ecog using deep learning. Journal of neural engineering. 2018;15(3):036009.

[15] Thomas TM *et al.* Simultaneous classification of bilateral hand gestures using bilateral microelectrode recordings in a tetraplegic patient. MedRxiv. 2020:2020–06.

[16] Pradeepkumar J, Anandakumar M, Kugathasan V, Lalitharatne TD, De Silva AC, Kappel SL. Decoding of hand gestures from electrocorticography with lstm based deep neural network. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC). 2021, 420–423.

[17] Yao L, Zhu B, Shoaran M. Fast and accurate decoding of finger movements from ecog through riemannian features and modern machine learning techniques. Journal of Neural Engineering. 2022;19(1):016037.

[18] Jiang T *et al.* Characterization and decoding the spatial patterns of hand extension/flexion using high-density ecog. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2017;25(4):370–379.

[19] Crone N *et al.* A click-based electrocorticographic brain-computer interface enables long-term high-performance switch-scan spelling. Research Square. 2023.

[20] Bleichner MG, Freudenburg ZV, Jansma JM, Aarnoutse EJ, Vansteensel MJ, Ramsey NF. Give me a sign: Decoding four complex hand gestures based on high-density ecog. Brain Structure and Function. 2016;221:203–216.

[21] Branco MP, Freudenburg ZV, Aarnoutse EJ, Bleichner MG, Vansteensel MJ, Ramsey NF. Decoding hand gestures from primary somatosensory cortex using high-density ecog. NeuroImage. 2017;147:130–142.

[22] Li Y *et al.* Gesture decoding using ecog signals from human sensorimotor cortex: A pilot study. Behavioural neurology. 2017;2017.

[23] Xu M, Zhou W, Shen X, Wang Y, Mo L, Qiu J. Swin-tcnet: Swin-based temporal-channel cascade network for motor imagery ieeg signal recognition. Biomedical Signal Processing and Control. 2023;85:104885.

[24] Siero JC, Hermes D, Hoogduin H, Luijten PR, Ramsey NF, Petridou N. Bold matches neuronal activity at the mm scale: A combined 7 t fmri and ecog study in human sensorimotor cortex. Neuroimage. 2014;101:177–184.

[25] Pernet CR *et al.* Eeg-bids, an extension to the brain imaging data structure for electroencephalography. Scientific data. 2019;6(1):103.

[26] Box GE, Cox DR. An analysis of transformations. Journal of the Royal Statistical Society Series B: Statistical Methodology. 1964;26(2):211–243.

[27] Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. Physical review E. 2004;69(6):066138.

[28] Ross BC. Mutual information between discrete and continuous data sets. PloS one. 2014;9(2):e87357.

[29] Tsamardinos I, Aliferis CF. Towards principled feature selection: Relevancy, filters and wrappers. In: International Workshop on Artificial Intelligence and Statistics. 2003, 300–307.

[30] Mockus J, Mockus J. The bayesian approach to local optimization. Springer (1989).

[31] Rasmussen CE, Williams CK, *et al.* Gaussian processes for machine learning. Springer (2006).

[32] Hoffman M, Brochu E, De Freitas N, *et al.* Portfolio allocation for bayesian optimization. In: UAI. 2011, 327–336.