**Beyond Accuracy: Assessing the Impact of EEG Denoising on the Diagnostic Utility of a Pre-Hospital Stroke Triage Model**

Rosmary Blanco1, Jan K. Argasiński1, Maritta N. van Stigt2, Eva E. Groenendijk2, Jonathan M. Coutinho2, Aneta Lisowska3, Henk A. Marquering 2

1Sano Centre for Computational Medicine, Czarnowiejska 36, 30-054 Kraków, Poland

2 Department of Neurology, Department of Radiology and Nuclear Medicine / Biomedical Engineering and Physics, Amsterdam University Medical Center, location University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands

3 Vrije Universiteit / Amsterdam University Medical Center, location University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands
{*r.blanco,* *j.argasinski*}@*sanoscience.org*, {*m.n.vanstigt,* *e.a.groenendijk, j.coutinho, a.j.lisowska, h.a.marquering*}@*amsterdamumc.nl***Keywords**: EEG, AutoML, Artifact Removal, Stroke.

**1. Introduction**

Rapid and accurate diagnosis of acute stroke is essential for improving patient outcomes; however, distinguishing stroke from stroke mimics in pre-hospital settings remains a major clinical challenge [1]. Machine learning (ML) techniques have shown promise in early, data-driven stroke prediction [2]. Electroencephalography (EEG) offers a portable, non-invasive tool for assessing brain function, but its diagnostic utility depends critically on robust signal processing to mitigate noise and artifacts [3]. This study assesses how automated EEG artifact removal methods affect ML model performance, reliability, and classification errors in differentiating ischemic stroke, hemorrhagic stroke, and stroke mimics.

**2. Description of the problem**

Low-density EEG systems with dry electrodes are well-suited for real-world clinical applications because of their portability and rapid setup. However, they are more prone to large, non-stereotypical artifacts, resulting in a low signal-to-noise ratio. Traditional artifact removal methods require expert intervention, making them impractical for real-time or time-sensitive applications [4]. Evaluating EEG denoising is challenging due to the lack of a clean ground-truth signal. Most studies assess efficacy via downstream task performance (e.g., classification accuracy) [5], but this may overlook effects on diagnostic reliability and clinical safety. Automated pipelines are needed to preserve diagnostically relevant neural features under real-world conditions.

**3. Related work**

Prior research has established the value of quantitative EEG features as biomarkers of cerebral ischemia and has increasingly leveraged ML to automate stroke detection [6]. However, most pioneering studies rely on a single, pre-defined denoising pipeline [7]. A systematic comparison of how alternative preprocessing strategies influence the clinical utility and error profile of diagnostic models, particularly for low-density EEG, remains a significant gap in the literature [8].

**4. Solution of the problem**

We analyzed data from 719 patients from the prospective, multicenter ELECTRA-STROKE and AI-STROKE studies, including 389 ischemic strokes and 330 non-ischemic cases (stroke mimics and hemorrhagic strokes). For each patient, 2–3 minutes of resting-state EEG were recorded using a portable 8-channel dry-electrode system. We compared five artifact removal methods: Wavelet Transform (W), Empirical Mode Decomposition (EMD), Artifact Subspace Reconstruction (ASR), and two hybrid approaches, ASR-W and EMD\_W [9], against a minimally processed (filtered) baseline. Each dataset was independently processed with AutoML (AutoPrognosis) [10] to identify the optimal ML pipeline, which optimized the AUROC in a 5-fold cross-validation scheme for unbiased comparison. Final performance was assessed on a held-out test set using AUROC, the DeLong test for statistical comparison, and clinical error analysis, including diagnosis-specific misclassification patterns and clustering of error profiles. Comparative analysis revealed that denoising choice influenced diagnostic value, beyond standard accuracy metrics. The EMD-W pipeline achieved the highest AUROC (0.938), not significantly different from the minimally processed baseline (0.930; p = 0.678). ASR-W performed significantly worse (0.908; p = 0.049). Crucially, error profile analysis showed that EMD-W reduced missed strokes to 15, compared with 20 in the baseline and 31 with ASR. In contrast, ASR\_W accumulated the most errors (56), misclassifying largely stroke mimics such as TIA and “Other” cases. These findings highlight a trade-off: advanced denoising does not always improve overall predictive accuracy but can critically impact diagnostic safety.

**5. Conclusions and future work**

Evaluation of preprocessing pipelines should go beyond accuracy metrics, and it is advisable to test multiple denoising methods to identify the approach that maximizes diagnostic utility and ensures model safety. This work represents a foundational step toward building reliable, data-driven physiological models for virtual human twin platforms, demonstrating that rigorous, automated processing is essential to translate noisy, real-world EEG data into clinically trustworthy solutions.

**Acknowledgements.** EU Horizon 2020 (857533), FNP (MAB PLUS/2019/13), Polish Ministry of Education and Science (MEiN/2023/DIR/3796).

References

1. Lumley, Hannah A., et al. "A scoping review of pre-hospital technology to assist ambulance personnel with patient diagnosis or stratification during the emergency assessment of suspected stroke." BMC emergency medicine 20.1 (2020): 30,

<https://doi.org/10.1186/s12873-020-00323-0>

1. Amann, Julia. "Machine learning in stroke medicine: Opportunities and challenges for risk prediction and prevention." Artificial Intelligence in Brain and Mental Health: Philosophical, Ethical & Policy Issues (2022): 57-71,

[https://link.springer.com/chapter/10.1007/978-3-030-74188-4\_5](https://link.springer.com/chapter/10.1007/978-3-030-74188-4_5#citeas:~:text=DOI-,https%3A//doi.org/10.1007/978%2D3%2D030%2D74188%2D4_5,-Published)

1. Sharma, Ramnivas, and Hemant Kumar Meena. "Emerging trends in EEG signal processing: A systematic review." SN Computer Science 5.4 (2024): 415,

<https://doi.org/10.1007/s42979-024-02773-w>

1. Mumtaz, Wajid, Suleman Rasheed, and Alina Irfan. "Review of challenges associated with the EEG artifact removal methods." Biomedical Signal Processing and Control 68 (2021): 102741,

<https://doi.org/10.1016/j.bspc.2021.102741>

1. Del Pup, Federico, et al. "The more, the better? Evaluating the role of EEG preprocessing for deep learning applications." IEEE Transactions on Neural Systems and Rehabilitation Engineering (2025),

<https://doi.org/10.1109/TNSRE.2025.3547616>

1. Lanzone, J., et al. "Quantitative measures of the resting EEG in stroke: a systematic review on clinical correlation and prognostic value." Neurological Sciences 44.12 (2023): 4247-4261,

<https://doi.org/10.1007/s10072-023-06981-9>

1. Sebastián-Romagosa, Marc, et al. "EEG biomarkers related with the functional state of stroke patients." Frontiers in neuroscience 14 (2020): 582,

<https://doi.org/10.3389/fnins.2020.00582>

1. Prakash, V., Kumar, D. (2024). Artifact Detection and Removal in EEG: A Review of Methods and Contemporary Usage. In: Santosh, K., Sood, S.K., Pandey, H.M., Virmani, C. (eds) Advances in Artificial-Business Analytics and Quantum Machine Learning. COMITCON 2023. Lecture Notes in Electrical Engineering, vol 1191. Springer, Singapore, <https://doi.org/10.1007/978-981-97-2508-3_19>
2. Behera, Sandhyalati, and Mihir Narayan Mohanty. "EEG artifact detection and removal techniques: A brief review." Computational Techniques in Neuroscience (2023): 143-187,

<https://doi.org/10.1201/9781003398066>

1. Imrie, Fergus, et al. "AutoPrognosis 2.0: Democratizing diagnostic and prognostic modeling in healthcare with automated machine learning." PLOS digital health 2.6 (2023): e0000276,

<https://doi.org/10.1371/journal.pdig.0000276>