A Deep Learning Approach to EEG Subcortical Source Localization

C. Buda[1], B. Gambosi[1], N. Toschi[2,3,4,*] and L. Astolfi[1,*]
[1] *Department of Computer, Control and Management Engineering, University of Rome "La Sapienza", Rome, Italy*
[2] *Department of Biomedicine and Prevention, University of Rome "Tor Vergata", Rome, Italy*
[3] *A.A. Martinos Center for Biomedical Imaging, Harvard Medical School, Boston, MA, USA*
[4] *Department of Psychology, School of Biological Sciences, University of Cambridge, Cambridge, UK*

***Abstract*—Electroencephalography (EEG) offers high temporal resolution but struggles to accurately localize subcortical activity, partly due to the ill-posed nature of the inverse problem and the weak signals from deep structures. Traditional regularized inverse methods are computationally efficient yet often miss deep sources. Here, we introduce a deep learning pipeline specifically designed for subcortical EEG source localization. We generate realistic training data through a custom simulator that combines spatially structured dipole activity, autoregressive time series, controlled synchronization, and distinct forward operators to reduce the inverse crime. Our network maps raw EEG segments directly to subcortical activation, bypassing explicit dipole reconstructions. Compared against nine classical solvers (including MNE, dSPM, sLORETA) across seven different metrics, our approach demonstrates superior localization accuracy and spatial specificity in both cortical and subcortical tests. This mitigates the surface bias typical of standard solutions and highlights the potential of end-to-end deep learning for EEG-based subcortical neuroimaging. Future work will refine simulation realism, explore multi-subject adaptability, and address transfer to real EEG.**

***Keywords*—EEG subcortical source localization, deep learning, simulation pipelines, ill-posed inverse problem**

I. Introduction

Neuroimaging techniques have become central to studying brain activity and dysfunction, with modalities such as functional MRI (fMRI), MEG, and EEG each offering distinct advantages. EEG stands out for its millisecond-level temporal resolution, making it highly sensitive to fast neuronal events. However, EEG source localization is notoriously ill-posed: many different source configurations in the brain can produce virtually indistinguishable scalp potentials, complicating subcortical localization in particular. Deep structures, such as the basal ganglia or hippocampus, contribute smaller signals with more rapid attenuation, further increasing the difficulty of precise localization [1].

Classical approaches to source localization often solve a linear inverse model that incorporates a forward operator describing how dipolar currents in the brain project onto the scalp [2, 3]. Simple methods based on the minimum-norm estimate (MNE), along with subsequent refinements such as dynamic statistical parametric mapping (dSPM) [4], LOw Resolution Electromagnetic TomogrAphy (LORETA), smooth LORETA (sLORETA) [5], exact LORETA (eLORETA) [6], or depth-weighted and orientation-constrained variants [7], have shown excellent performance in cortical reconstructions. Yet, localizing deeper subcortical structures often suffers from smeared or diffuse estimates, as the constraints favor superficial solutions or rely on global regularization priors that may not be optimal for deep dipoles.

Recently, deep learning methods have emerged as promising alternatives, capable of learning complex inverse mappings from EEG measurements to underlying sources without explicitly solving the linear inverse problem [8, 9]. By training on large-scale synthetic data, neural networks can implicitly capture patterns in spatial and temporal structure that standard linear solvers overlook. Although encouraging progress has been made, few studies have tackled the subcortical localization challenge directly, which demands increased realism in training simulations (e.g., appropriate modeling of deeper sources and partial correlations in the source space) and specialized network architectures that can exploit subtle features in short EEG epochs.

In this paper, we introduce a novel deep learning approach tailored for subcortical source localization.

II. Materials and Methods

A. Head Models

We used a publicly available multimodal dataset for both anatomical MRI segmentation and realistic EEG sensor layouts [10]. A single subject's T1-weighted MRI was segmented with FreeSurfer [11] to derive cortical and subcortical surfaces. We employed two pipelines for forward model computation:

- *OpenMEEG* [12, 13]: used for constructing the forward operator incorporated into the classical inverse solvers.
- *DUNEuro* [14]: used for generating synthetic scalp potentials in the training simulations, using slightly different conductivity values to mitigate overfitting to a single forward model (the so-called "inverse crime").

The source space comprised a total of 9753 dipoles: 8196 confined to the cortical sheet with fixed (normal) orientations, and 1557 distributed across deep gray-matter structures with unconstrained orientation. This ensures coverage of subcortical areas such as the putamen, hippocampus, thalamus, and other basal ganglia components.

B. Simulation Tool

To train the neural network, we required a simulation tool capable of generating physiologically plausible source activations while preserving the flexibility to systematically vary location, size, and synchrony of active regions.

1) Spatial Module: We randomly select an initial dipole in the subcortical volume or cortex (depending on the simulation parameter). Starting from this "center" dipole, additional dipoles are iteratively added to define an active region of variable radius (5–60 mm). Each dipole d_i is assigned a spatial coefficient $c_i \in [0, 1]$, which decreases radially from the center dipole. The final distribution forms a continuous cloud of dipoles, rather than scattered points, approximating the shape of a focal activation.

2) Temporal Module: Every dipole within the source space (whether active or inactive) is assigned a synthetic autoregressive (AR) time series generated independently. The time series are then normalized to have unit variance. These are designed to be largely uncorrelated across the entire source space, except for a synchronization step applied in the next stage.

3) Synchronization Step: To introduce physiologically inspired correlation within the active region, we blend each dipole's AR time series with the center dipole's time series according to its spatial coefficient, e.g.,

$$s_i(t) \leftarrow c_i \cdot s_{\text{center}}(t) + (1 - c_i) \cdot s_i(t),$$

where $s_i(t)$ is the AR time series originally assigned to dipole i. Hence, larger c_i enforces higher synchronization with the center dipole. This step creates a continuum between highly synchronous signals near the center and largely independent signals outside the active region.

4) Scalp Projection: The final source configuration (dipoles plus time series) is projected through the DUNEuro-generated forward operator to produce scalp potentials at 64 electrodes (the same montage as the real EEG in [10]). Additive Gaussian noise is superimposed to control signal-to-noise ratio (SNR) levels in the training set.

By design, this multi-step simulation produces a wide range of synthetic EEG data with subcortical or cortical foci of varying sizes, realistically correlated time courses, and moderate amounts of noise.

C. Neural Network Architecture and Training

1) Task Definition and Loss: We formulate the training objective as predicting each dipole's "activation coefficient" in the range $[0, 1]$. In other words, the target is a sparse 3D activation map indicating how strongly each dipole belongs to the active region. During training, we use binary cross-entropy over the 9753 output units:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^{N} \Big[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \Big],$$

where $N = 9753$, y_i is the ground truth coefficient (1 for active, 0 for inactive), and $\hat{y}_i$ is the network prediction.

2) Model Architecture: The network ingests 0.2 s of EEG data (50 samples at 250 Hz) from 64 electrodes. We use an initial linear transformation to project from 64 channels $\times$ 50 samples to a higher-dimensional space. A monodirectional LSTM module processes this embedding sequentially, capturing temporal dependencies. Finally, three fully connected layers with ReLU activations process the LSTM hidden state, increasing the feature dimension, with a final sigmoid output layer producing the 9753 activation coefficients.

In initial experiments, LSTM outperformed standard feedforward or 1D convolutional designs, likely due to its capacity to capture time correlations in short EEG segments. Optimization is performed with ADAM, including weight decay and a learning rate scheduler. The training set comprises 10,000 distinct simulations with a single, randomly placed, subcortical or cortical foci of variable extent.

D. Comparison With Classical Inverse Solvers

We benchmark our approach against nine well-established solvers: MNE, dSPM, sLORETA, eLORETA, LORETA, and LAURA (all without orientation constraints or depth weighting) [2, 4, 5, 6, 15], as well as MNE, dSPM, and sLORETA with orientation constraints and depth weighting (0.2 and 0.8, respectively) [7].

Each solver outputs a full spatiotemporal map across the 9753 dipoles. We compute the norm of the three components of each dipole at each time point, apply mild spatial smoothing to reduce noise, normalize to [0,1], and average over time to yield a "final activation coefficient" for each dipole.

E. Evaluation Metrics

We employ six metrics that capture different aspects of source localization quality. Mean Localization Error (MLE) is the average distance between the predicted peak dipole and the true source (where the predicted source is defined as the dipole with maximum activation). $\text{AUC}_{\text{close}}$ and AUC_{far} are AUROC-like scores capturing local vs. global specificity by comparing near-boundary to far-out negative samples. DistRank records how far the true source ranks below the predicted peak. Peak Activity Spread (PAS) gauges how spatially concentrated or diffused the top 20% predicted activation is. FractionActive measures the fraction of the top 5% highest-intensity dipoles that lie within the true region, normalized by region size.

We also compare each method's output with a *Random* baseline (random predictions across dipoles) and a *Perfect* reference (the true region mask) to contextualize the performance bounds.

III. Results

We present representative results on four simulated Regions (A–D), each featuring different sizes (5–60 mm) and cortical or subcortical locations. We generate a 2-minute simulated EEG for each Region and then extract 50-sample windows (0.2 s at 250 Hz), producing $\sim$650 samples per Region. The final performance metrics are reported as median values across all samples.

A. Region A (Large Cortical Focus)

Region A is a 60 mm extent patch on the left lateral cortex. Table I summarizes the metrics; Fig. 1 visualizes the reconstructions from the proposed LSTM and sLORETA.

Methods	MLE (mm)	AUC_{close}	AUC_{far}	DistRank	PAS (mm)	FractionActive
Random	80	0.506	0.492	0.913	66	0.060
Perfect	0	1	1	0	45	1
LSTM	9	0.987	1.000	0.002	15	0.987
MNE	15	0.987	1.000	0.006	15	1.000
dSPM	25	0.720	0.995	0.045	23	0.159
sLORETA	5	0.949	1.000	0.000	21	0.871
MNE_{depthw}	20	0.945	1.000	0.022	17	0.867
$dSPM_{depthw}$	43	0.703	0.984	0.140	28	0.043
$sLORETA_{depthw}$	5	0.956	1.000	0.001	20	0.884
eLORETA	16	0.960	1.000	0.002	17	0.901
LORETA	13	0.979	1.000	0.008	15	0.991
LAURA	37	0.674	0.592	0.348	18	0.000

TABLE I: Performance measures on Region A (large cortical activation).

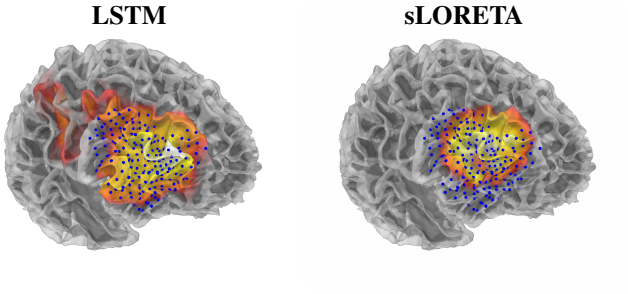

Fig. 1: Reconstruction of Region A using the LSTM (left) and sLORETA (right). Blue dots indicate the ground truth region, while the color indicates predicted activation.

Here, both the LSTM and certain classical methods (e.g., sLORETA, MNE) achieve very accurate localization (MLE ≈5–15 mm). Notably, dSPM shows higher error and a decreased FractionActive, suggesting it tends to overestimate activation outside the target region.

B. Region B (Large Subcortical Focus)

Region B covers the entire left putamen with a 60 mm radial extent. Table II details the performance, and Fig. 2 compares LSTM with depth-weighted sLORETA.

Methods	MLE (mm)	AUC_{close}	AUC_{far}	DistRank	PAS (mm)	FractionActive
Random	67	0.500	0.489	0.632	65	0.053
Perfect	0	1	1	0	16	1
LSTM	3	0.998	1.000	0.001	11	1.000
MNE	48	0.581	0.989	0.152	24	0.000
dSPM	21	0.806	1.000	0.033	21	0.507
sLORETA	16	0.898	1.000	0.024	26	0.766
MNE_{depthw}	37	0.930	1.000	0.023	20	0.952
$dSPM_{depthw}$	19	0.788	1.000	0.049	18	0.301
$sLORETA_{depthw}$	16	0.903	1.000	0.021	23	0.794
eLORETA	27	0.855	1.000	0.036	23	0.627
LORETA	51	0.496	0.757	0.330	14	0.000
LAURA	37	0.656	0.954	0.107	14	0.072

TABLE II: Performance measures on Region B (left putamen).

Unlike cortical activations, subcortical sources impose stricter requirements on forward modeling and regularization strategies. While depth-weighted or oriented variants (e.g., MNE_{depthw}) partially mitigate surface bias, only the LSTM consistently achieves MLE under 10 mm. Classical methods

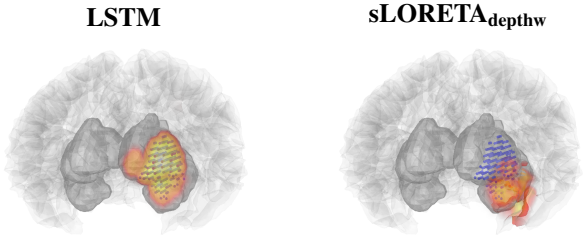

Fig. 2: Reconstruction of Region B using the LSTM (left) and $sLORETA_{depthw}$ (right). Blue dots indicate the ground truth region, while the color indicates predicted activation.

tend to extend activity toward superficial regions, resulting in increased PAS and reduced FractionActive.

C. Regions C and D (Medium Extent Cortical and Subcortical)

We also tested two medium-sized (≈30 mm) activations. Region C is cortical, located on the left parietal area. Region D is subcortical, covering the left hippocampus. Tables III and IV summarize the performance.

Methods	MLE (mm)	AUC_{close}	AUC_{far}	DistRank	PAS (mm)	FractionActive
Random	82	0.465	0.510	0.552	66	0.048
Perfect	0	1	1	0	30	1
LSTM	10	0.995	1.000	0.001	12	1.000
MNE	14	0.920	1.000	0.012	12	0.919
dSPM	25	0.488	0.986	0.168	19	0.065
sLORETA	22	0.940	1.000	0.008	25	0.984
MNE_{depthw}	45	0.671	1.000	0.084	30	0.387
$dSPM_{depthw}$	44	0.330	0.967	0.303	21	0.000
$sLORETA_{depthw}$	22	0.944	1.000	0.010	27	0.968
eLORETA	24	0.923	1.000	0.014	24	0.968
LORETA	14	0.895	0.992	0.017	14	0.758
LAURA	41	0.386	0.516	0.674	20	0.000

TABLE III: Performance measures on Region C (medium cortical source).

Methods	MLE (mm)	AUC_{close}	AUC_{far}	DistRank	PAS (mm)	FractionActive
Random	48	0.467	0.488	0.391	65	0.038
Perfect	0	1	1	0	7	1
LSTM	10	0.978	1.000	0.005	10	0.955
MNE	38	0.567	0.984	0.121	22	0.000
dSPM	21	0.678	1.000	0.089	17	0.218
sLORETA	12	0.874	1.000	0.015	24	0.797
MNE_{depthw}	36	0.926	1.000	0.013	20	0.910
$dSPM_{depthw}$	19	0.790	1.000	0.056	16	0.286
$sLORETA_{depthw}$	13	0.872	1.000	0.015	23	0.782
eLORETA	17	0.835	1.000	0.023	23	0.707
LORETA	39	0.433	0.683	0.393	13	0.000
LAURA	56	0.855	0.996	0.018	20	0.654

TABLE IV: Performance measures on Region D (medium subcortical source in left hippocampus).

In Region C, both LSTM and MNE variants localize well (MLE ≈ 10–14 mm), whereas dSPM has a larger DistRank and lower FractionActive. Region D, a hippocampal focus, proves more challenging for classical solvers: MNE has MLE = 38 mm, while the LSTM only has 10 mm. Depth-weighted MNE and sLORETA partially improve subcortical performance, but still underperform the LSTM.

In all cases, the deep model tends to preserve focal energy at or near the ground-truth source, mitigating superficial spread.

This is particularly evident in subcortical configurations (Regions B and D), where classical linear solvers frequently distribute activation toward higher conductivity cortical surfaces.

IV. Discussion

We introduced a deep learning framework that learns directly from brief scalp recordings, using an LSTM to decode subcortical activity. A custom simulation pipeline generates diverse training examples with realistic spatial distributions, correlated time courses, and distinct forward operators. Results on four representative regions show that our approach consistently surpasses classical solvers in localizing deep sources, maintaining tighter spatial focus and lower errors. Even for larger cortical regions, the deep model achieves comparable or better performance than state-of-the-art inverse methods.

Despite these improvements, our simulator still depends on simplified AR processes and assumes radially symmetric activation patterns. Future developments will include biologically realistic connectivity, address multi-region and multi-subject scenarios, and enhance robustness to noise and artifacts inherent in real EEG data. By integrating deep learning with domain-aware simulation, this work addresses long-standing EEG localization challenges, particularly in deeper brain structures.

V. Conclusion

We presented a neural approach specifically geared toward subcortical EEG source localization. Pairing an LSTM-based architecture with tailored simulations yielded superior accuracy compared to nine inverse solvers. This highlights the potential of end-to-end deep learning to advance EEG-based imaging of deep brain regions. Ongoing work focuses on richer biological constraints, multi-subject generalization, and domain adaptation for real EEG.

Acknowledgments

We thank the open-access data providers and the developers of FreeSurfer, DUNEuro, and OpenMEEG. This work was partially supported by the AEGEUS project funded by the European Union, Horizon Europe Programme (GA 101099210).

References

[1] S. Baillet, J. Mosher, and R. Leahy, "Electromagnetic brain mapping," *IEEE Signal Processing Magazine*, vol. 18, no. 6, pp. 14–30, 2001.

[2] M. Hämäläinen and R. Ilmoniemi, "Interpreting magnetic fields of the brain: minimum-norm estimates," *Med. Biol. Eng. Comput.*, vol. 32, pp. 35–42, 1994.

[3] B. Van Veen, W. Van Drongelen, M. Yuchtman, and A. Suzuki, "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 9, pp. 867–880, 1997.

[4] A. Dale, A. Liu, B. Fischl, R. Buckner, E. Halgren *et al.*, "Dynamic statistical parametric mapping: Combining fmri and meg for high-resolution imaging of cortical activity," *Neuron*, vol. 26, no. 1, pp. 55–67, 2000.

[5] R. Pascual-Marqui, "Standardized low resolution brain electromagnetic tomography (sloreta): Technical details," *Methods Find Exp Clin Pharmacol.*, vol. 24, no. Suppl D, pp. 5–12, 2002.

[6] ——, "Discrete, 3d distributed, linear imaging methods of electric neuronal activity. part 1: Exact, zero error localization," *Math. Physics Biol. Physics Neurons Cogn.*, vol. 0710, 2007, arXiv:0710.3341.

[7] F. Lin, J. Belliveau, A. Dale, and M. Hämäläinen, "Distributed current estimates using cortical orientation constraints," *Hum. Brain Mapp.*, vol. 27, pp. 1–13, 2006.

[8] R. Sun, A. Sohrabpour, G. Worrell, and B. He, "Deep neural networks constrained by neural mass models improve electrophysiological source imaging of spatiotemporal brain dynamics," *Proc. Natl. Acad. Sci. U.S.A.*, 2021.

[9] L. Hecker, R. Rupprecht, L. Tebartz Van Elst, and J. Kornmeier, "Convdip: a convolutional neural network for better eeg source imaging," *Front. Neurosci.*, vol. 15, p. 569918, 2021.

[10] Q. Telesford, E. Gonzalez-Moreira, T. Xu *et al.*, "An open-access dataset of naturalistic viewing using simultaneous eeg-fmri," *Sci Data*, vol. 10, p. 554, 2023.

[11] A. Dale, B. Fischl, and M. Sereno, "Cortical surface-based analysis. i. segmentation and surface reconstruction," *NeuroImage*, vol. 9, pp. 179–194, 1999.

[12] A. Gramfort, T. Papadopoulo, E. Olivi, and M. Clerc, "Openmeeg: Open-source software for quasistatic bioelectromagnetics," *BioMed Eng OnLine*, vol. 9, p. 45, 2010.

[13] A. Gramfort, M. Luessi, E. Larson, D. Engemann, M. Hämäläinen *et al.*, "Meg and eeg data analysis with mne-python," *Front. Neurosci.*, vol. 7, no. 267, pp. 1–13, 2013.

[14] S. Schrader, A. Westhoff, M. Piastra, C. Miinalainen, C. Engwer *et al.*, "Duneuro – a software toolbox for forward modeling in bioelectromagnetism," *PLoS ONE*, vol. 16, no. 6, p. e0252431, 2021.

[15] R. De Peralta-Menendez and S. Gonzalez-Andino, "A critical analysis of linear inverse solutions to the neuroelectromagnetic inverse problem," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 4, pp. 440–448, 1998.