Evidential Neural Networks for Uncertainty-Aware Alzheimer's Disease Screening from Resting-State EEG

Siddhant Ujjain, Vaibhav Kagathara, Pooja Singh, Abhishek Kumar Singh, Sandeep Kumar, and Tapan Kumar Gandhi

Indian Institute of Technology Delhi, New Delhi, India

Abstract. Electroencephalography (EEG) is a low-cost and non-invasive modality that may support accessible dementia screening, but clinical deployment requires both high discrimination performance and reliable uncertainty estimates. We propose an evidential neural network (ENN) for subject-level screening of Alzheimer's disease and related dementia from resting-state EEG, using a CN versus AD+FTD patient-control formulation. Building on an EEGNet backbone, the model predicts non-negative evidence that induces a Dirichlet distribution over class probabilities, enabling computation of *vacuity* as an epistemic uncertainty measure in a single forward pass. We further introduce a temperature scaled classification objective with an uncertainty-aware regularizer (UMSE) to encourage low vacuity on correctly learned samples while maintaining high uncertainty for ambiguous cases. Experiments on OpenNeuro ds004504 (88 subjects; CN vs AD+FTD) use leakage-safe 5-fold stratified *group* cross-validation with subject-level aggregation. The proposed ENN achieves strong discrimination (AUROC 0.95) while improving calibration and uncertainty quality (lower ECE, Brier, and NLL) compared to softmax EEGNet, MC Dropout, and Deep Ensemble baselines, without increasing inference cost. These results suggest evidential modeling can improve reliability and enable selective prediction for safer EEG-based dementia screening.

Keywords: Alzheimer's disease · EEG · evidential deep learning · uncertainty quantification · calibration · selective prediction

1 Introduction

Alzheimer's disease (AD) and related dementias are a growing global health burden [20], and early identification can support timely intervention, care planning, and clinical trial enrollment. However, standard diagnostic workups often rely on neuroimaging or fluid biomarkers that may be expensive, invasive, or difficult to access at scale. Electroencephalography (EEG) offers an appealing complementary modality: it is non-invasive, low-cost, portable, and captures functional brain dynamics [10] that change with neurodegeneration [3], making it suitable for scalable screening and longitudinal monitoring.

A wide range of machine learning approaches have been explored for EEG-based dementia detection [2]. Classical pipelines typically rely on handcrafted spectral/entropy/connectivity features coupled with standard classifiers, while end-to-end deep models (e.g., CNN/TCN/RNN [17] and EEGNet [13] variants) learn task-specific representations directly from EEG epochs and often achieve strong discrimination. Yet accuracy or AUROC alone is insufficient for clinical decision support: modern neural networks can be miscalibrated and *overconfident when wrong* [8, **?**], causing harmful consequences such as unnecessary follow-up or missed opportunities for early intervention. Reliable uncertainty estimation is therefore essential, particularly in low-data regimes and in the presence of noisy or atypical EEG recordings. Bayesian approximations such as MC Dropout [4] and Deep Ensembles [12] can provide uncertainty estimates, but they increase inference cost, requiring multiple stochastic forward passes or models. Evidential deep learning [18, 19] offers an efficient alternative by predicting non-negative *evidence* parameters that define a Dirichlet distribution [14] over class probabilities, providing both predictive probabilities and an epistemic uncertainty [11] signal in a *single forward pass*. In this framework, *vacuity* (inverse Dirichlet strength) increases when total evidence is low, indicating that the model "does not know." This is especially useful for *selective prediction* [5], where uncertain cases can be deferred for clinician review or additional testing, improving safety at a chosen coverage level.

We present an EEGNet-based evidential model for uncertainty-aware dementia screening from resting-state EEG under a leakage-safe subject-level protocol. We adopt EEGNet as backbone, replacing the softmax head with a Dirichlet evidential layer to estimate class probabilities and vacuity. To improve reliability, we combine a temperature-scaled classification objective with an uncertainty-aware regularizer (UMSE) that encourages low vacuity on correctly learned examples while permitting higher uncertainty on ambiguous cases. We report discrimination, calibration, and selective prediction behavior.

Contributions. (1) We propose an EEGNet-based evidential model that produces Dirichlet evidence and vacuity, enabling single-pass epistemic uncertainty estimation for EEG-based dementia screening. (2) We perform leakage-safe subject-level evaluation using 5-fold stratified group cross-validation and compare against a broad suite of baselines, including classical ML, deterministic deep networks, and uncertainty baselines (MC Dropout and Deep Ensembles). (3) We provide a comprehensive reliability study that includes calibration metrics (ECE, Brier, NLL), ablations of temperature scaling and uncertainty regularization, and selective prediction analysis using vacuity.

2 Methods

2.1 Dataset and Preprocessing

We use the OpenNeuro ds004504 [15] resting-state EEG dataset with 88 subjects in three diagnostic groups: cognitively normal controls (CN), Alzheimer's disease (AD), and frontotemporal dementia (FTD) (Table 1). We formulate a

binary screening task by merging dementia subtypes into a single patient class (Patient = AD+FTD) versus controls (CN). This follows a screening formulation where the first decision separates cognitively normal subjects from neurodegenerative patient groups; subtype-specific AD-vs-FTD diagnosis is outside this work's scope and requires larger cohorts. All evaluation is at the *subject level* using leakage-safe splits (no subject appears in more than one fold).

Table 1. Dataset characteristics (OpenNeuro ds004504 [15]).

Item	Value
Subjects (total)	88
CN / AD / FTD	29 (33%) / 36 (41%) / 23 (26%)
Task	Resting-state
Channels	19 (10–20)
Sampling (orig → resamp)	500 Hz → 128 Hz
Epoch length	4.0 s (512 samples)
Epochs (post-preproc)	15,496
Binary task	Patient = AD+FTD vs Control = CN

Raw EEG $\mathbf{X}_{\text{raw}} \in \mathbb{R}^{C\times T}$ is band-pass filtered (0.5–45 Hz), notch-filtered (50/60 Hz when applicable), re-referenced to the common average, resampled (500→128 Hz), and segmented [7] into 4 s non-overlapping epochs. We reject epochs with peak amplitude exceeding $\theta_{\text{amp}} = 200\,\mu$V. Each epoch is represented as $\mathbf{X} \in \mathbb{R}^{1\times C\times T}$ with C=19 and T=512.

2.2 Model Architecture

Our model comprises (i) an EEGNet feature extractor and (ii) an evidential head [1] that induces a Dirichlet distribution over class probabilities. The full pipeline is illustrated in Fig. 1.

EEGNet Backbone Each epoch $\mathbf{X} \in \mathbb{R}^{1\times C\times T}$ is mapped to a feature vector

$$\mathbf{z} = f_\theta(\mathbf{X}), \tag{1}$$

where f_θ follows the EEGNet design (factorized temporal and spatial filtering with separable convolutions) to obtain compact spatiotemporal features under limited data.

Evidential Output Head (Dirichlet Evidence) We predict non-negative evidence $\mathbf{e} \in \mathbb{R}^K_{\geq 0}$,

$$\mathbf{e} = \text{Softplus}(\mathbf{W}\mathbf{z} + \mathbf{b}), \qquad e_k \geq 0, \tag{2}$$

and form Dirichlet parameters

$$\boldsymbol{\alpha} = \mathbf{e} + \mathbf{1}, \qquad \alpha_k > 0, \tag{3}$$

yielding $\boldsymbol{\pi} \sim \mathrm{Dir}(\boldsymbol{\alpha})$. The predictive mean is

$$p_k = \mathbb{E}[\pi_k] = \frac{\alpha_k}{S}, \qquad S = \sum_{j=1}^{K} \alpha_j. \tag{4}$$

Uncertainty (vacuity). We use Dirichlet strength S to define epistemic uncertainty (vacuity)

$$u = \frac{K}{S}. \tag{5}$$

Training Objective (Temp-CE + UMSE) For an epoch-label pair $(\mathbf{X}, y)$, we optimize

$$\mathcal{L} = \mathcal{L}_{\text{Temp-CE}} + \lambda\, \mathcal{L}_{\text{UMSE}}. \tag{6}$$

Temp-CE. Let $\mathbf{g} = \mathbf{W}\mathbf{z} + \mathbf{b} \in \mathbb{R}^K$ be the pre-activation logits (before the Softplus used for evidence). Temperature scaling [8] with $\tau > 0$ gives

$$\tilde{\mathbf{p}} = \mathrm{Softmax}\Big(\frac{\mathbf{g}}{\tau}\Big), \qquad \mathcal{L}_{\text{Temp-CE}} = -\log \tilde{p}_y, \tag{7}$$

with τ annealed toward 1 during training.

UMSE. Let $c = \mathbb{I}[\hat{y} = y]$ where $\hat{y} = \arg\max_k p_k$. We set a target vacuity

$$u^\star = (1 - c)\, u_{\text{hi}} + c\, u_{\text{lo}}, \qquad 0 \le u_{\text{lo}} < u_{\text{hi}} \le 1, \tag{8}$$

and minimize

$$\mathcal{L}_{\text{UMSE}} = (u - u^\star)^2. \tag{9}$$

Selective prediction. At inference time, we can abstain on uncertain subjects using subject-level uncertainty (e.g., mean vacuity). For coverage γ, we accept the least-uncertain γ fraction and compute risk on the accepted set, producing risk–coverage curves that quantify the safety–coverage trade-off.

2.3 Training and Evaluation Protocol

We evaluate using 5-fold stratified **group** cross-validation (group = subject) to prevent leakage. Models are trained on epochs and pooled to subjects by mean aggregation:

$$\mathbf{p}^{(s)} = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbf{p}_i^{(s)}. \tag{10}$$

We report Accuracy, area under the receiver operating characteristic curve (AUROC), area under the precision–recall curve (AUPRC), Sensitivity, Specificity, and F1. Uncertainty is assessed with expected calibration error (ECE) [16], Brier score [**?**], and negative log-likelihood (NLL); selective prediction is evaluated via risk–coverage curves [6], which show error rate as a function of accepted coverage.

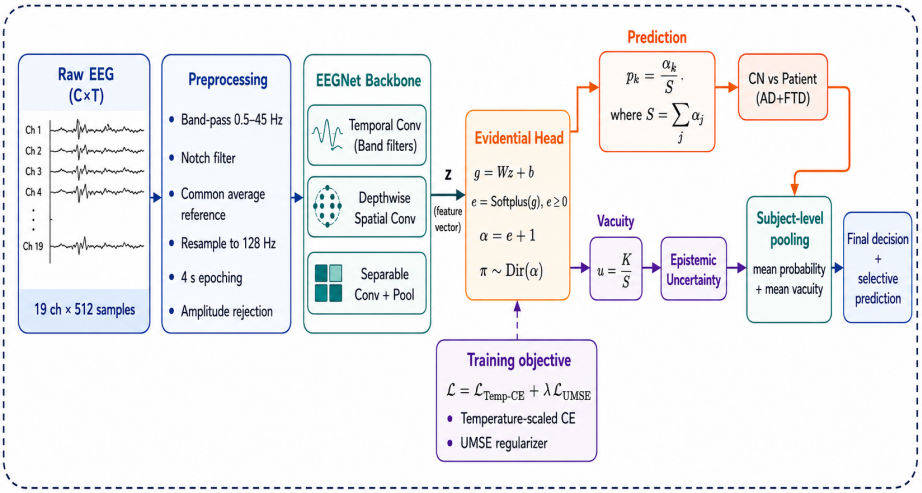

Fig. 1. Evidential EEG classification pipeline. Preprocessed multi-channel EEG is encoded by an EEGNet backbone, followed by an evidential layer that models class probabilities and epistemic uncertainty via a Dirichlet distribution, enabling uncertainty-aware subject-level prediction between cognitively normal (CN) and patient groups.

3 Experiments and Results

3.1 Implementation Details

EEG preprocessing used MNE-Python [7], as described in Sec. 2.1. Deep models were trained with Adam and early stopping using a *group-preserving* validation split; class imbalance was handled with class-weighted losses. Leakage is enforced via 5-fold stratified *group* cross-validation, and all metrics are reported at the *subject level* by mean pooling.

3.2 Baselines and Main Results

We compare against (i) classical ML on handcrafted subject-level statistics, (ii) deterministic deep baselines trained end-to-end on raw epochs (EEGNet, ResNet1D [9], TCN, GRU), and (iii) uncertainty baselines (MC Dropout, Deep Ensembles). All methods share the same leakage-safe subject-level protocol (5-fold stratified group CV; subject prediction by mean aggregation).

Table 2 summarizes subject-level performance over folds. Deterministic deep baselines achieve strong discrimination but can be unreliable (e.g., lower specificity), motivating uncertainty-aware modeling. The proposed evidential model achieves the best AUROC and AUPRC while retaining single-pass inference (1x), as also shown in Fig. 2. Compared to Bayesian approximations, it avoids the inference overhead of MC Dropout (20–30x) and Deep Ensembles (3–5x), while remaining competitive or superior across operating characteristics.

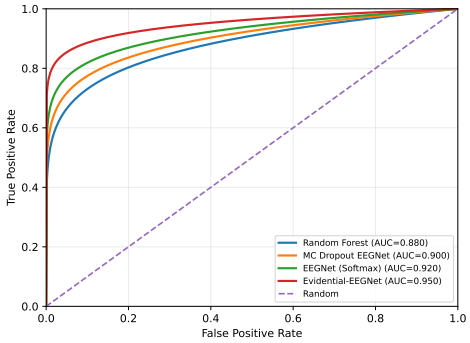

Fig. 2. ROC curves for top baselines and the proposed ENN (subject-level).

Table 2. Subject-level performance (mean±std over 5 folds).

Model	Acc	AUROC	AUPRC	Sens	Spec	F1	Inf. Cost
LogReg	0.75 ± 0.13	0.81 ± 0.12	0.91 ± 0.06	0.74 ± 0.14	0.76 ± 0.28	0.80 ± 0.11	1x
GaussianNB	0.58 ± 0.10	0.76 ± 0.18	0.89 ± 0.08	0.41 ± 0.13	0.93 ± 0.09	0.56 ± 0.13	1x
k-NN	0.80 ± 0.14	0.89 ± 0.09	0.95 ± 0.04	0.79 ± 0.16	0.79 ± 0.14	0.83 ± 0.12	1x
Random Forest	0.81 ± 0.13	0.88 ± 0.18	0.95 ± 0.05	0.88 ± 0.16	0.80 ± 0.26	0.88 ± 0.15	1x
XGBoost	0.62 ± 0.09	0.79 ± 0.13	0.91 ± 0.06	0.53 ± 0.15	0.75 ± 0.14	0.63 ± 0.13	1x
Linear SVM	0.62 ± 0.14	0.78 ± 0.20	0.91 ± 0.07	0.59 ± 0.18	0.67 ± 0.33	0.66 ± 0.20	1x
SVM-RBF	0.76 ± 0.12	0.86 ± 0.16	0.94 ± 0.05	0.79 ± 0.12	0.75 ± 0.23	0.82 ± 0.11	1x
EEGNet (Softmax)	0.74 ± 0.15	0.92 ± 0.08	0.96 ± 0.05	0.83 ± 0.28	0.53 ± 0.21	0.79 ± 0.17	1x
ResNet1D	0.77 ± 0.11	0.91 ± 0.11	0.96 ± 0.04	0.82 ± 0.20	0.67 ± 0.38	0.82 ± 0.10	1x
TCN	0.71 ± 0.15	0.89 ± 0.11	0.95 ± 0.04	0.79 ± 0.25	0.58 ± 0.43	0.78 ± 0.16	1x
GRU	0.66 ± 0.13	0.74 ± 0.27	0.87 ± 0.11	0.78 ± 0.15	0.69 ± 0.12	0.79 ± 0.11	1x
MC Dropout EEGNet	0.72 ± 0.07	0.90 ± 0.10	0.96 ± 0.03	0.84 ± 0.10	0.53 ± 0.41	0.80 ± 0.03	20–30x[a]
Deep Ensemble EEGNet	0.82 ± 0.06	0.92 ± 0.10	0.96 ± 0.03	0.85 ± 0.10	0.67 ± 0.38	0.86 ± 0.05	3–5x[b]
EEGNet-EDL (CE only)	0.81 ± 0.07	0.88 ± 0.13	0.95 ± 0.05	0.87 ± 0.16	0.72 ± 0.32	0.82 ± 0.03	1x
Evidential-EEGNet (Proposed) (Temp-CE+UMSE)	0.83 ± 0.07	0.95 ± 0.04	0.97 ± 0.02	0.88 ± 0.22	0.78 ± 0.25	0.84 ± 0.08	1x

[a] T stochastic forward passes (e.g., T=30).
[b] M independently trained models (e.g., M=5).

To isolate the effect of the output head and objective on an identical backbone, Table 3 reports gains over EEGNet (Softmax), showing consistent improvements in discrimination without increasing inference cost.

3.3 Ablation Study

We ablate (i) the evidential head (EDL), (ii) temperature-scaled training (Temp-CE), and (iii) uncertainty regularization (UMSE); results are reported in Table 4. EDL with CE alone modestly improves calibration over softmax (ECE 0.24 vs. 0.31) but not as much as the full model, and UMSE alone degrades discrimination (AUROC 0.78), indicating that uncertainty regularization is most effective when paired with a confidence-controlled objective. The full model (Temp-CE+UMSE) yields the best trade-off, achieving the highest AUROC and the lowest ECE/NLL among all variants in Table 4. The low accuracy of Temp-CE

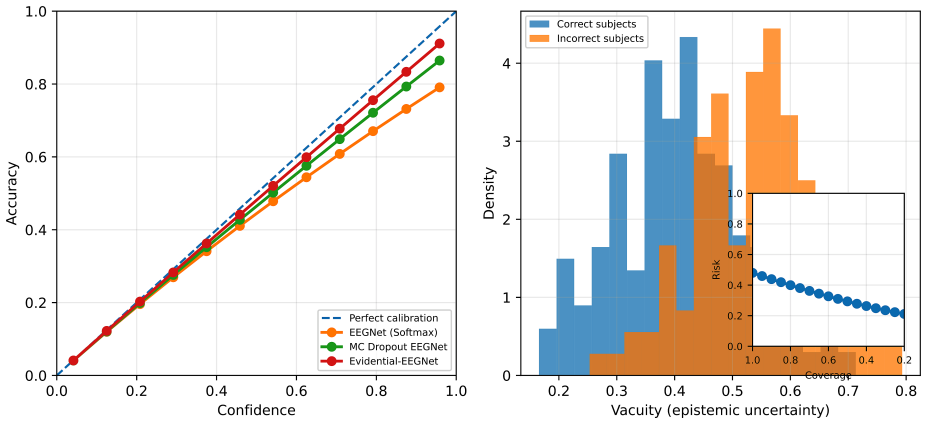

Fig. 3. Calibration and uncertainty analysis. Reliability diagram and vacuity-based selective prediction at the subject level.

Table 3. Improvement over EEGNet (Softmax) baseline (subject-level means over folds). Base. denotes Baseline and Evid.-EEGNet denotes Evidential-EEGNet (Proposed).

Metric	Base.	Evid.-EEGNet	Δ
Accuracy	0.74	0.83	+0.09
AUROC	0.92	0.95	+0.03
AUPRC	0.96	0.97	+0.01
F1	0.79	0.84	+0.05

alone (0.68) alongside its high AUROC (0.93) reflects threshold displacement under annealed temperature: discrimination is preserved but the decision boundary is not optimally placed without the uncertainty regularizer. Vacuity is higher on incorrect than correct predictions across all EDL variants, supporting its use as an epistemic signal for abstention.

3.4 Uncertainty, Calibration, and Selective Prediction

We assess reliability using ECE, Brier score, and NLL, and evaluate whether uncertainty supports selective prediction (abstention). As shown in Table 5, the proposed ENN improves probabilistic quality over all baselines while preserving single-pass inference. Notably, MC Dropout yields a higher Brier score than the softmax baseline (0.23 vs. 0.21), consistent with prior observations that stochastic averaging without explicit calibration can redistribute probability mass in ways that worsen proper scoring rules [**?**]. Beyond calibrated probabilities, evidential learning provides epistemic uncertainty via vacuity u, enabling a safety mechanism that abstains on high-uncertainty subjects. Fig. 3 (left) confirms the proposed model's confidence curve lies closest to the calibration diagonal.

Table 4. Ablation study (mean over folds). Lower ECE/NLL is better.

Variant	Acc	AUROC	ECE	NLL	Mean Vac. (err)	Mean Vac. (corr)
EEGNet (Softmax)	0.74	0.92	0.31	0.81	–	–
EEGNet-EDL (CE only)	0.81	0.88	0.24	0.58	0.49	0.33
+ Temp-CE only	0.68	0.93	0.30	0.63	0.52	0.47
+ UMSE only	0.71	0.78	0.32	0.73	0.55	0.49
Evidential-EEGNet (Full)	**0.83**	**0.95**	**0.22**	**0.47**	0.54	0.41

Table 5. Uncertainty quality (mean±std over 5 folds). Lower ECE/Brier/NLL is better.

Method	ECE	Brier	NLL
EEGNet (Softmax)	0.31 ± 0.06	0.21 ± 0.13	0.81 ± 0.38
MC Dropout EEGNet	0.28 ± 0.04	0.23 ± 0.02	0.65 ± 0.09
Deep Ensemble EEGNet	0.29 ± 0.07	0.21 ± 0.07	0.68 ± 0.22
Evidential-EEGNet (Proposed) (Temp-CE+UMSE)	0.22 ± 0.06	0.16 ± 0.02	0.47 ± 0.05

We operationalize selective prediction by ranking subjects by vacuity and rejecting the most uncertain, reporting risk (error rate) versus coverage. Informative uncertainty should reduce risk at matched coverage by preferentially abstaining on likely errors; consistent with Table 4, vacuity is higher for incorrect than correct predictions. The resulting risk–coverage behavior is summarized in Fig. 3 (right), where decreasing coverage corresponds to increasing abstention, indicating that the ENN provides actionable uncertainty for safer EEG-based screening.

4 Discussion

We have shown that replacing the softmax head of EEGNet with a Dirichlet evidential layer, trained with temperature-scaled cross-entropy and uncertainty regularization, yields gains in discrimination and calibration (lower ECE/Brier/NLL) without increasing inference cost. Unlike MC Dropout or deep ensembles, the proposed model requires only a single forward pass, an important property for latency-sensitive settings such as outpatient triage or portable EEG deployments. The resulting vacuity score provides an interpretable epistemic signal that enables *selective prediction*: high-vacuity cases can be deferred to clinician review, reducing the risk of overconfident misclassification and aligning model behavior with safety-oriented clinical pathways.

Limitations and Future Work. Our binary formulation (AD+FTD vs. CN) may obscure subtype-specific signatures, and the moderate size of OpenNeuro ds004504 (88 subjects, single site) limits generalizability; multi-class extension and multi-center external validation are natural next steps. Evaluating

robustness under distribution shift (e.g., montage variation, medication effects) and incorporating clinical covariates remain important directions for translation.

5 Conclusion

We presented an evidential neural network for uncertainty-aware Alzheimer's disease screening from resting-state EEG. By equipping an EEGNet backbone with a Dirichlet evidential head trained via Temp-CE and UMSE, the model achieves the highest subject-level discrimination (AUROC 0.95, AUPRC 0.97) and best calibration (ECE 0.22, Brier 0.16, NLL 0.47) in a single forward pass, avoiding the inference overhead of MC Dropout and deep ensembles. The resulting vacuity signal provides a principled epistemic uncertainty measure that supports selective prediction, enabling safer clinical workflows where uncertain subjects are deferred for expert review.

Acknowledgments

We gratefully acknowledge Prof. Sandeep Kumar's support through the Core Research Grant (CRG, RP04820), which enabled this work.

Disclosure of Interests

The authors have no competing interests in the paper.

References

1. Amini, A., Schwarting, W., Soleimany, A., Rus, D.: Deep evidential regression. Advances in Neural Information Processing Systems **33**, 14927–14937 (2020)
2. Cassani, R., Estarellas, M., San-Martin, R., Fraga, F.J., Falk, T.H.: Systematic review on resting-state EEG for Alzheimer's disease diagnosis and progression assessment. Disease Markers **2018**, 5174815 (2018)
3. Dauwels, J., Vialatte, F., Cichocki, A.: Diagnosis of Alzheimer's disease from EEG signals: Where are we standing? Current Alzheimer Research **7**(6), 487–505 (2010)
4. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: international conference on machine learning. pp. 1050–1059. PMLR (2016)
5. Geifman, Y., El-Yaniv, R.: Selective classification for deep neural networks. Advances in neural information processing systems **30** (2017)
6. Geifman, Y., El-Yaniv, R.: Selectivenet: A deep neural network with an integrated reject option. In: International conference on machine learning. pp. 2151–2159. PMLR (2019)
7. Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hämäläinen, M.S.: Mne software for processing meg and eeg data. neuroimage **86**, 446–460 (2014)

8. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: International conference on machine learning. pp. 1321–1330. PMLR (2017)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
10. Jeong, J.: EEG dynamics in patients with Alzheimer's disease. Clinical Neurophysiology **115**(7), 1490–1505 (2004)
11. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? Advances in neural information processing systems **30** (2017)
12. Lakshminarayanan, B., Pritzel, A., Blundell, C.: Simple and scalable predictive uncertainty estimation using deep ensembles. Advances in neural information processing systems **30** (2017)
13. Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J.: EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces. Journal of Neural Engineering **15**(5), 056013 (2018)
14. Malinin, A., Gales, M.: Reverse kl-divergence training of prior networks: Improved uncertainty and adversarial robustness. Advances in neural information processing systems **32** (2019)
15. Miltiadous, A., Tzimourta, K.D., Giannakeas, N., Tsipouras, M.G., Afrantou, T., Ioannidis, P., Tzallas, A.T.: A dataset of scalp EEG recordings of Alzheimer's disease, frontotemporal dementia and healthy subjects from routine EEG. OpenNeuro (2023), doi:10.18112/openneuro.ds004504.v1.0.5
16. Naeini, M.P., Cooper, G., Hauskrecht, M.: Obtaining well calibrated probabilities using bayesian binning. In: Proceedings of the AAAI conference on artificial intelligence. vol. 29 (2015)
17. Schirrmeister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., Ball, T.: Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapping **38**(11), 5391–5420 (2017)
18. Sensoy, M., Kaplan, L., Kandemir, M.: Evidential deep learning to quantify classification uncertainty. In: Advances in Neural Information Processing Systems. vol. 31, pp. 3179–3189 (2018)
19. Ujjain, S., Soren, V.N., Kumar, S., Gandhi, T.K.: Evidence-based epileptic seizure detection. In: 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 1–6. IEEE (2025)
20. World Health Organization: Dementia: Key facts. https://www.who.int/news-room/fact-sheets/detail/dementia (2023), accessed: 2024-01-15