Classification using deep learning in neurological disorders: a literature review

Paul van der Kleij[1]

[1]Graduate student MSc Technical Medicine, Erasmus MC, TU Delft, The Netherlands

May 2024

1 Introduction

Neurological disorders are the biggest cause of disability worldwide, with potentially far-reaching consequences for an individual's quality of life.[1] These can be burdensome conditions such as Epilepsy, that affect over 50 million people worldwide [2], and Alzheimer's disease, that affect over 7 million people worldwide.[3] An important diagnostic tool for assessing these disorders is the ElectroEncephaloGram (EEG), that was invented in 1929 by Hans Berger.[4, 5] Accurately assessing EEG data requires expert readers, but even then remains susceptible to inter-reader variability.[6] New advances in the field of Artificial Intelligence (AI) have attempted to automate the assessment of EEGs using deep learning algorithms, with promising results.[7] Deep learning is able to extract and process complex EEG features to assess EEG data, but its use in clinical practice is still far from standard. This is often attributed to the lack of transparency and interpretability of deep learning algorithms, calling them 'black boxes'.[8] It has been reported that lack of trust in deep learning, and doubt about false positive results are barriers for the implementation of AI.[9]

This literature review aims to give an overview of the different applications and the performance of deep learning for the classification of patients with (suspected) neurological disorders. Additionally, solutions to augment interpretability and transparency will be highlighted, to evaluate their possible role in the implementation of deep learning in clinical practice.

2 Methods

A literature search was performed on the 18th of March 2024 through Web of Science and PubMed. The search string that was used can be found in *Appendix A*. Since AI has recently undergone major developments [10] , articles before 2010 were not taken into account. From a technical perspective, the search string did not require deep learning to be specifically mentioned as such, but also looked for overarching terms like 'Machine Learning' or 'Artificial Intelligence', since the use of these terms is sometimes interchanged. From a clinical perspective, the subject had to be of neurological nature.

Articles were excluded that 1) did not have a neurological disorder as main subject 2) did not use EEG, 3) did not use deep learning, 4) did not have full-text availability through either the Erasmus MC Medical Library, or TU Delft, 5) did not use deep learning to classify on a patient level, but only employed in-EEG outcome labels, 6) was not available in English, 7) were animal studies.

After screening on full-text for the aforementioned in- and exclusion criteria, data was extracted systematically on the following topics: aim of study, used dataset, input data type and length, classification classes, general methods, feature extraction, deep learning architecture, validation, and performance accuracy. For clinical data pooling, articles will be pooled according to overlapping neurological disorders. Contrarily, technical details on deep learning will be pooled in general, since all included articles use EEG data.

3 Results

3.1 Search Results

1104 articles were initially identified through Web of Science and PubMed, of which 198 were duplicates. Hereafter, screening on title and abstract excluded 810 articles. Full-text assessment excluded 28 articles, and reasons of exclusion were itemized in *Figure 1* of *Appendix B*. A flowchart with an overview of article screening and the inclusion process can be found in *Figure 1* of *Appendix B*. A full overview of included articles is provided in the supplementary excel-file.

The following neurological categories were identified: neurodegenerative disorders, epileptic disorders, (semi-)acute brain injuries, and sleep disorders. Articles that did not fit these categories were assigned to the category 'other'.

3.2 Deep learning

3.2.1 Input data and feature extraction

EEG data was transformed to the spectral or wavelet domain in most instances, using the Fast Fourier Transform (FFT)[11–15] or Wavelet Transform (WT)[16–22], respectively. If a patient cohort was too small to train a deep learning algorithm, EEG data was split into segments/epochs by (non-)overlapping windowing as means for data augmentation.[23, 24] These segments directly represented patients in most cases, but could also be used in a voting system before giving a definitive label to the patient.[25] If spectral data was used, either the PSD was used as input directly [14, 26] , or the PSD was split in underlying frequency bands (e.g. Delta, Alpha, ..) that were then used for feature extraction.[13, 18, 27–34]

Principal or Independent Component Analysis (PCA/ICA) was a main technique used for feature extraction and artifact filtering. [16, 18, 33, 35–39] When eye blink artifacts were frequently observed, the resulting artifact corresponds to the most evident component following the PCA/ICA, enabling quick and computationally efficient filtering. [33] Additionally, because of its inherent filtering of outliers, research that used the PCA/ICA yielded good results in signals with high noise corruption rates. [36–38]

Articles that used raw EEG directly as input data or did not perform extensive manual feature engineering, mostly used a CNN-architecture in their deep learning architecture for feature extraction. [24, 40–43]

3.2.2 Deep learning architectures

The Multi-Layer Perceptron (MLP) was the first deep learning architecture used for the classification of neurological disorders. [11–14, 18, 21, 27–31, 34, 44–51] It was frequently used as comparison standard for other machine learning or deep learning models, but often yielded inferior results. [11, 30, 46, 47] Extensive feature engineering was performed to provide input for the MLP. [44, 45] The need for (manually) engineering features in advance resulted in a less complex and extensive feature space compared to a Convolutional Neural Network or Graph Neural Network. [30, 46]

The most frequently used deep learning architecture was a Convolutional Neural Network (CNN). [13, 16, 17, 19, 20, 22–26, 30, 33, 35, 36, 38, 39, 41–43, 46, 52–74] CNNs were able to learn high-level features of the EEG by using one or more convolutional layers, which provided greater robustness to intra-class variability. [38, 56] Some articles used a CNN only for feature extraction, after which the remaining deep learning architecture consisted of Recurrent Neural Network modules.[38, 61, 74]

A Recurrent Neural Network (RNN) was the second most employed type of deep learning. [29, 39, 74]. Specifically RNNs that included (bi-directional) Long-Short Term Memory (LSTM) modules in their architecture. [15, 32, 37, 40, 41, 52, 54, 61, 67, 75, 76]. The vanishing gradient problem that can occur in RNNs was addressed by using an LSTM module. [41] Gated Recurrent Units (GRU) were mentioned as alternative for solving the vanishing gradient problem. [39, 74] The bi-directional LSTMs were well suited for use in EEG data, because of their capability to learn from past and future timepoints, even in (near) real-time scenarios. [40, 41, 47]

Graph Neural Networks (GNN) were used to closely resemble physiological connectivity in the brain.[30, 32, 77] Using measures such as the Granger Causality or Spectral Coherence, this type of deep learning architecture aimed to classify patient based on connectivity measures in distinctive regions of the brain.[30, 77] The inherent structure of a GNN ensured that there was no spatial bias between electrodes. [30]

For the eventual assigning of classes to patient data most articles employed 1 or 2 Fully Connected (FC) layers, followed by a SoftMax function.[24, 25, 38, 39, 77]

3.2.3 Validation of results

K-fold cross validation was utilized for ensuring robust and generalizable results; folds of k=5 or k=10 were most common.[15, 20, 22, 23, 33, 38, 39, 41, 42, 46, 48, 52, 53, 59, 66, 68, 73, 76, 77] Leave-one(-subject)-out cross validation was also used, especially as means for increasing the data size in the train set.[21, 25, 27, 29, 31, 34, 64, 70]

3.2.4 Visualization and interpretation of deep learning

Feature importance was assessed using SHapley Additive exPlanations (SHAP) in 3 instances.[59, 70, 73] This enabled researchers to interpret the individual contribution of features to the deep learning algorithm performance. Importance of specific data segments was analyzed using Gradient-weighted Class Activation Mapping (Grad-CAM), highlighting class-discriminative regions in either time-varying data or input feature images.[22, 58] Finally, the extent to which specific frequency bands influenced the deep learning classification performance was evaluated using critical band noise masking.[53]

3.3 Epileptic disorders

A total of 22 articles were included in the epilepsy category. [12, 13, 16–19, 23, 24, 33, 39, 47–50, 63–68, 72, 76] The detection and initial diagnosis of epilepsy was the foremost subject (n=12, note: articles can employ multiple subjects), essentially classifying epilepsy patients versus healthy controls.[12, 13, 16–18, 24, 33, 39, 49, 50, 72, 76] The distinction between epilepsy and psychological disorders was also subject to research (n=4), aiming to prevent the general misdiagnosis of epilepsy.[19, 23, 47, 76] The remaining included articles focused on personalizing the diagnosis to fit each patient specifically (n=8).[17, 48, 63–68]

3.3.1 Diagnosing epilepsy using deep learning

In 6 studies, the epilepsy dataset of Bonn, containing single channel EEG data of epilepsy subjects and healthy controls, was sourced for input data. These studies reported classifying epilepsy versus healthy controls with an accuracy of 0.92 to 1.00 .[12, 16, 17, 39, 49, 76] Research that used in-hospital acquired EEG data using the 10-20 system for electrode placement, reached accuracies of 0.80 to 0.96 when performing a similar classification task. [13, 18, 24, 50] Not all deep learning algorithms required the EEG data of an epilepsy patient to be recorded during a seizure, some could also differentiate interictal (epileptic) EEGs versus healthy controls.[24, 33, 49, 72] *Lin et al 2020* even suggested an alternative way of diagnosing epilepsy in children without the presence interictal epileptiform discharges (IED) in a seizure-free EEG.[24]

3.3.2 Specifying epilepsy diagnosis

Psychological disorders that were assessed alongside epilepsy were Schizophrenia (n=2) and Psychogenic Non-Epileptic Seizures (PNES)(n=2).[19, 23, 47, 76] A Support Vector Machine (SVM) was shown to be superior to an Artificial Neural Network (ANN) for distinguishing Shizophrenia versus Epilepsy[47] , but PNES could be distinghuised from epilepsy using deep learning, with an accuracy of 0.86 to 0.94.[19, 23] Specific epilepsy types such as Self-Limiting Epilepsy with Centro-Temporal Spikes (SeLECTS) and Rolandic Epilepsy could be diagnosed with an accuracy of 0.90 and 0.80, respectively.[64, 66] Additionally, evaluating IED spike shapes in these types of epilepsy confirmed the hypothesis that these spikes contain discriminative diagnostic information.[64] Research that focused on classifying epileptic seizures was able to recognise focal seizures

versus generalized ones. Specifying focal seizures even further was possible with Intracranial EEG, localizing the Seizure Onset Zone (SOZ) with a mean deviation of 7mm.[63]

3.4 Neurodegenerative diseases

Neurodegenerative disease was subject to research in 25 articles. The majority was aimed at Alzheimer's Disease (AD) and Mild Cognitive Impairment (MCI)(n=16).[14, 15, 20, 26, 29–31, 34, 35, 37, 42, 60, 62, 69, 71, 77] Diagnosing Parkinson's Disease (PD) was evaluated in 6 instances.[22, 32, 36, 61, 74, 77] 5 articles included other types of neurodegenerative disease.[11, 21, 43, 60, 73]

3.4.1 Differentiating Alzheimer's Disease and Mild Cognitive Impairment

There were high accuracies for diagnosing AD vs. healthy controls (acc. 0.89-0.98)[14, 30, 37], but also promising accuracies for differentiating AD from its prodromal stage MCI.[71] Using locally interpretable AI, higher delta and theta powers were found in patients progressing from healthy to MCI and AD.[34, 35] Most of the research in Alzheimer's classification focuses on epoch-based (or segment-based) classification as a proxy for the patient outcome. *Kim et al 2023* describe a substantial performance gap (Acc. 0.97 to 0.68, resp.) between their epoch-based and patient-based evaluation protocols. They conjectured this gap was mainly attributed to higher variance in EEG acquisition parameters in real-world data. [60]

3.4.2 Parkinson's Disease

There were very high reported accuracy for diagnosing Parkinson's Disease versus healthy controls (0.87-0.99).[22, 32, 36, 74] Because cognitive dysfunction can emerge over time in PD patients there is a clinical need for exploring differences in AD and PD, which was evaluated with an accuracy of 0.93.[77] Using Grad-CAM, *Chang et al 2023* showed that the frontal and temporal lobe are possible regions of interest for diagnosing PD patients. Additionally, regions of high intensity across scale heights in the wavelet domain were absent in patients with PD, but are yet to be linked directly to neurological activity at specific EEG frequencies. It does, however, imply the possible use of Grad-CAM for clinical interpretation of deep learning results.[32]

3.4.3 Multi-class / Other

Other types of neurodegenerative disease were mainly assessed in multi-class deep learning algorithms. Besides classifying neurodegenerative diseases such as Huntington's Disease, Lewy Body Dementia and Creutzfeldt-Jacob's Disease versus healthy controls, they also attempted to differentiate multiple neurodegenerative diseases amongst one another. Where the 2-class algorithms yielded good results (neurodegenerative disease vs. healthy controls, acc. 0.80 to 0.92)[21, 43, 73] , the multiple-class algorithms performed less reliably (acc. 0.67 to 0.68).[43, 60]

3.5 (Semi-)acute brain injuries

The 12 articles in (semi-)acute brain injury group consists of 2 main groups; Hypoxia/Anoxia[25, 28, 41, 54–56, 58, 70], and Traumatic Brain Injury (TBI).[40, 57, 59, 75]

3.5.1 Hypoxia/Anoxia

The severity of hypoxia was defined in terms of Hypoxic-Ischemic Encephalopathy (HIE) severity grading. With the most recent deep learning studies achieving an accuracy of up to 0.89 for predicting HIE grades. This accuracy was the result of using a voting system on multiple classified segments of patient data, rather than having segments directly represent patient outcome.[25, 56] In the case of anoxia, all research was aimed at classifying good versus poor outcome at 12 or 24-hours post-anoxia. The prognosis that was given by deep learning had an AUC of 0.885-0.89 compared to actual clinical outcome.[54, 55, 58] This method of prognostication also extended to patient suffering from stroke, predicting their proportional recovery with an 0.88-0.92 accuracy compared to poor recovery.[70]

3.5.2 Traumatic brain injury

Detecting TBI through deep learning was done with an accuracy of 0.78-1.00.[59, 75] Moreover, moderate cases of TBI were also evaluated, resulting in comparable accuracies of 0.72-0.93. This could represent a first step towards addressing a fundamental challenge in the diagnosis of concussions, since it is currently dependent on self-reported symptoms.[40, 57]

3.6 Sleep disorders

There were 3 articles that performed classification of a Sleep disorder on a patient level.[44, 52, 53] An accuracy of 0.86-0.92 was achieved in diagnosing Insomnia using data from only 1 EEG channel. This indicates simplified sleep monitoring hardware could be sufficient for a reliable diagnosis, possibly improving in-home ambulatory monitoring options.[44]

Rapid Eye Movement (REM) Sleep Behavior Disorder (RBD) could be classified with an accuracy of 0.89 compared to healthy controls. *Brink-Kjaer et al 2022* showed this performance accuracy while removing extensive pre-processing steps of the EEG from a polysomnography, proving the feasibility of an end-to-end sleep classification model.[52]

3.7 Other

5 articles discussed topics that did not fit the aforementioned groups.[27, 38, 45, 46, 51] Similar to the 'Post-anoxic' group, some focused on classifying good versus bad prognoses. *Vuckovic et al. 2018* used EEG data to predict the onset of Central Neuropathic Pain in patients with Spinal Cord Injury with an accuracy of 0.90.

4 Discussion

The generalizability of the included deep learning algorithms remains up for debate. Even though most research used a form of (cross-) validation, this validation data was initially part of the same (single) data set. This makes an interpretation of reported accuracies difficult, since the robustness and inherent bias of a dataset could vary between studies. Other medical fields have called upon their scientific community to develop a general test dataset that could function as a litmus test of deep learning performance.[78] Even though in the epileptic disorder category a large portion of articles used the same dataset, making comparison amongst them possible, their performance could not directly be translated to clinical practice. Since the Bonn and Bern-Barcelona datasets only consisted of 23s single-channel EEG data for each patient, there is a distinctive gap to the multi electrode recordings of clinical EEGs.

This use of segmented EEG data also became apparent in studies that did not use the Bonn or Bern-Barcelona datasets, since most of the studies used segment-based classification as a proxy for patient outcome. *Kim et al. 2023* observed a large performance gap between outcomes on single segments versus outcomes on patient level.[60] *Raurale et al. 2020* also demonstrated that including more segments in patient classification, using a voting structure, improved classification results.[25] This underlines the impact of segmental labelling, since EEG segments of a diseased individual could also (sporadically) appear healthy. The use of a tailored voting system for each specific neurological disorder, could resolve this issue. For example, an individual that is being assessed for PD, with 1% of their segments being categorized as diseased, could be categorised as healthy. Whereas an individual that is being assessed for epilepsy, with a similar percentage of diseased segments, could be categorized as epileptic.

Even though a quantitative comparison of deep learning types was not issued in this review, the included articles did give a view of some benefits that result from their architecture. CNNs were used for automatic feature extraction, even in cases that used raw (but windowed) EEG data. This could enable potential backpropagating explanatory modules to more specifically pinpoint regions of interest in the original EEG.[52, 54] GNNs architecture closely resembled the placement of electrodes, enabling functional connectivity measures to be used for evaluation.[30, 77] Future research could consider using GNNs to classify for the presence of known physiological brain connectivity, such as the Default Mode Network (DMN). Decreases in the DMN have been suggested to be indicative of AD.[79]

The evaluation of individual feature importance using SHAP, GradCAM and critical band noise masking assesses 3 domains: spatial, time, spectral. In an ideal case, these 3 domains are combined for clinical interpretation of deep learning results. In reality, most techniques only give a transparent view of 1 or 2 domains. It is essential that explanatory AI bridges the gap by mimicking its results to the way a neurologist would interpret an EEG. That is, highlighting segments of a specific electrode in the original EEG, with an annotation that tells the neurologist what (spectral) feature influenced the classification outcome. *Raab et al. 2023* developed an eXplainable AI for EEG (XAI4EEG) module, based on two SHAP explainers, that encompasses the aforementioned criteria. Following a user study, a substantially lower time for validating predictions and an increase in interpretability, trust and confidence were reported.[80]

As the medical field progresses to using more deep learning in clinical practice [81] , the framework of scientific evidence needs to shift as well. The contents of this review demonstrate that scientific research on deep learning in neurological disorders is mainly formatted to investigate and improve technical feasibility and performance. To make routine use possible in a hospital, the end-users of such algorithms should also be taken into account. Research could focus on user verification to integrate deep learning results in the current structure of clinical decision making. A possible future step is to then set up prospective studies, with outcome labels that are not only the result of deep learning classification, but outcomes that reflect the final clinical decision making of neurologists based on the deep learning results.

5 Limitations

This literature review aimed to give an overview of the application of deep learning in classifying neurological disorders. By setting the specific scope of this review, some closely affiliated articles were excluded that might (indirectly) have clinical or technical implications on this topic. This becomes apparent when highlighting the articles in the 'Sleep disorder' group. Only 3 articles were included in this review, which contrasts their relative abundance in the used online medical libraries. The large number of exclusions is contributed to the fact that most of the 'Sleep Disorder' articles focus on using deep learning for the automatic characterization of polysomnographic data segments. Rather than classifying on the patient level, deep learning was used as a tool to lighten the analytical load, by automatically allocating PSG/EEG-data segments to a certain sleep stage. Effectively being an in-EEG outcome label, this has no consequence for the clinical decision making on a patient level. Additionally, the search string only includes 'Sleep disorder'-articles that have been self-reported as being of neurological nature. The search string did not include a separate search term for this specific sub-category, as was the case for most of the other assessed categories. On the whole, this might yield a misconstrued view on the current or future role of deep learning in these neurological fields.

The use of other types of machine learning for the classification of neurological disorders were neglected in this review. For some neurological diseases, classifiers that are less complex and easier to interpret might be sufficient, such as the k-Nearest Neighbour (kNN) or Random Forest (RF). This eludes the notorious 'black box' of deep learning, and promotes interpretability. Because even though new explanative analysis techniques such as the SHAP values and Grad-CAM methods have been developed, they are yet to find their way into clinical practice.

Classification on a patient level was a prerequisite for including articles in this review, since our future research is focused on classifying on a patient level. This excluded articles that only employed an in-EEG outcome label. However, with most included articles classifying patients based on segmental data, the in- or exclusion of an article could be equivocal. In order to make uniform in- and exclusions across neurological disorders, the final labeling of outcomes was definitive. If the labeling was symptomatic or solely descriptive of EEG features, the article was excluded, if the labeling metioned patient outcome, the article was included. A future literature review should reduce a possible selection bias by using multiple reviewers for article inclusion, according to the PRISMA 2020 statement.[82]

This literature review focused primarily on the accuracy of deep learning algorithms. Other performance metrics (e.g. sensitivity, specificity, false positive rate,...) were not reported in the majority of research. Because these other performance metrics were reported only occasionally, drawing parallels or conclusions based on these outcomes was difficult. In order to prevent misinterpretation of results, most additional performance metrics were not used in this review. Accuracy can depict the general performance of an algorithm quite well, but the clinical implementation of deep learning in specific disease classification calls for a more elaborate view of model performance. For instance, in post-anoxia patients it is crucial that no patient is unjustly labelled as 'poor outcome', corresponding to a perfect specificity of 1.0. Consequently, when opting for high specificity in a model, assuming a constant AUC, this is at the expense of the sensitivity and accuracy. Taking this into account, it is evident that using only 1 performance metric gives an incomplete view of model dynamics, rendering it unviable for clinical use in some cases.

6 Conclusion

The development of deep learning algorithms for the classification of neurological diseases gives promising results. Most literature is still focused on improving classification accuracies, in order to make eventual clinical use viable. Additionally, there is an increasing number of studies that report some form of feature importance analysis, giving insight in the decision making of deep learning algorithms, partially unclouding the 'black box'.

Acknowledgements

This literature review was made as part of the education programme MSc Technical Medicine (Erasmus MC, TU Delft). The author would like to thank Wichor Bramer from the Erasmus MC medical library for his assistance in setting up the literature search string

References

1. Steinmetz, J. D. *et al.* Global, regional, and national burden of disorders affecting the nervous system, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. English. *The Lancet Neurology* **23.** Publisher: Elsevier, 344–381. ISSN: 1474-4422, 1474-4465. https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(24)00038-3/fulltext (2024) (Apr. 2024).
2. *Epilepsy* en. https://www.who.int/news-room/fact-sheets/detail/epilepsy (2024).
3. 2023 Alzheimer's disease facts and figures. eng. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **19,** 1598–1695. ISSN: 1552-5279 (Apr. 2023).
4. Rayi, A. & Murr, N. I. eng. in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2024). http://www.ncbi.nlm.nih.gov/books/NBK563295/ (2024).
5. Tudor, M., Tudor, L. & Tudor, K. I. [Hans Berger (1873-1941)–the history of electroencephalography]. hrv. *Acta Medica Croatica: Casopis Hravatske Akademije Medicinskih Znanosti* **59,** 307–313. ISSN: 1330-0164 (2005).
6. Grant, A. C. *et al.* EEG interpretation reliability and interpreter confidence: a large single-center study. eng. *Epilepsy & Behavior: E&B* **32,** 102–107. ISSN: 1525-5069 (Mar. 2014).
7. Tveit, J. *et al.* Automated Interpretation of Clinical Electroencephalograms Using Artificial Intelligence. *JAMA Neurology* **80,** 805–812. ISSN: 2168-6149. https://doi.org/10.1001/jamaneurol.2023.1645 (2024) (Aug. 2023).
8. Ahmed, M. I. *et al.* A Systematic Review of the Barriers to the Implementation of Artificial Intelligence in Healthcare. *Cureus* **15,** e46454. ISSN: 2168-8184. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10623210/ (2024).
9. Reus, E. E. M., Visser, G. H., Sommers-Spijkerman, M. P. J., van Dijk, J. G. & Cox, F. M. E. Automated spike and seizure detection: Are we ready for implementation? eng. *Seizure* **108,** 66–71. ISSN: 1532-2688 (May 2023).
10. Kaul, V., Enslin, S. & Gross, S. A. History of artificial intelligence in medicine. eng. *Gastrointestinal Endoscopy* **92,** 807–812. ISSN: 1097-6779 (Oct. 2020).
11. Rodrigues, P. M. & Teixeira, J. P. Classification of Electroencephalogram Signals Using Artificial Neural Networks. English. *2010 3rd International Conference on Biomedical Engineering and Informatics,* 808–812. ISSN: 1948-2914 (2010).
12. Koçer, S. & Canal, M. R. Classifying epilepsy diseases using artificial neural networks and genetic algorithm. *J Med Syst* **35,** 489–498. ISSN: 0148-5598 (Print) 0148-5598 (Linking). (0008) (2011).
13. Guerrero, M. C., Parada, J. S. & Espitia, H. E. EEG signal analysis using classification techniques: Logistic regression, artificial neural networks, support vector machines, and convolutional neural networks. *Heliyon* **7,** e07258. ISSN: 2405-8440 (Print) 2405-8440 (Electronic) 2405-8440 (Linking). (0006) (2021).

14. Ferri, R. *et al.* Stacked autoencoders as new models for an accurate Alzheimer's disease classification support using resting-state EEG and MRI measurements. *Clin Neurophysiol* **132,** 232–245. ISSN: 1872-8952 (Electronic) 1388-2457 (Linking). (0001) (2021).

15. Gkenios, G., Latsiou, K., Diamantaras, K., Chouvarda, I. & Tsolaki, M. Diagnosis of Alzheimer's disease and Mild Cognitive Impairment using EEG and Recurrent Neural Networks. *Annu Int Conf IEEE Eng Med Biol Soc* **2022,** 3179–3182. ISSN: 2694-0604 (Electronic) 2375-7477 (Linking). (0007) (2022).

16. Shams, M. & Sagheer, A. A natural evolution optimization based deep learning algorithm for neurological disorder classification. *Biomed Mater Eng* **31,** 73–94. ISSN: 1878-3619 (Electronic) 0959-2989 (Linking) (2020).

17. Xin, Q., Hu, S., Liu, S., Zhao, L. & Zhang, Y. D. An Attention-Based Wavelet Convolution Neural Network for Epilepsy EEG Classification. *IEEE Trans Neural Syst Rehabil Eng* **30,** 957–966. ISSN: 1558-0210 (Electronic) 1534-4320 (Linking) (2022).

18. Işik, H. & Sezer, E. Diagnosis of epilepsy from electroencephalography signals using multilayer perceptron and Elman Artificial Neural Networks and Wavelet Transform. *J Med Syst* **36,** 1–13. ISSN: 0148-5598 (Print) 0148-5598 (Linking). (0002) (2012).

19. Lo Giudice, M. *et al.* Permutation Entropy-Based Interpretability of Convolutional Neural Network Models for Interictal EEG Discrimination of Subjects with Epileptic Seizures vs. Psychogenic Non-Epileptic Seizures. *Entropy (Basel)* **24.** ISSN: 1099-4300 (Electronic) 1099-4300 (Linking). (0001) (2022).

20. Huggins, C. J. *et al.* Deep learning of resting-state electroencephalogram signals for three-class classification of Alzheimer's disease, mild cognitive impairment and healthy ageing. *J Neural Eng* **18.** ISSN: 1741-2552 (Electronic) 1741-2552 (Linking). (0006) (2021).

21. Morabito, F. C. *et al.* Deep Learning Representation from Electroencephalography of Early-Stage Creutzfeldt-Jakob Disease and Features for Differentiation from Rapidly Progressive Dementia. *Int J Neural Syst* **27,** 1650039. ISSN: 1793-6462 (Electronic) 0129-0657 (Linking). (0003) (2017).

22. Shaban, M. & Amara, A. W. Resting-state electroencephalography based deep-learning for the detection of Parkinson's disease. *PLoS One* **17,** e0263159. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking) (2022).

23. Lo Giudice, M. *et al.* Convolutional Neural Network Classification of Rest EEG Signals among People with Epilepsy, Psychogenic Non Epileptic Seizures and Control Subjects. *Int J Environ Res Public Health* **19.** ISSN: 1660-4601 (Electronic) 1661-7827 (Print) 1660-4601 (Linking). (0011) (2022).

24. Lin, L. C., Ouyang, C. S., Wu, R. C., Yang, R. C. & Chiang, C. T. Alternative Diagnosis of Epilepsy in Children Without Epileptiform Discharges Using Deep Convolutional Neural Networks. *Int J Neural Syst* **30,** 1850060. ISSN: 1793-6462 (Electronic) 0129-0657 (Linking). (0005) (2020).

25. Raurale, S. A., Boylan, G. B., Lightbody, G. & O'Toole, J. M. Grading the severity of hypoxic-ischemic encephalopathy in newborn EEG using a convolutional neural network. *Annu Int Conf IEEE Eng Med Biol Soc* **2020,** 6103–6106. ISSN: 2694-0604 (Electronic) 2375-7477 (Print) 2375-7477 (Linking). (0007) (2020).

26. Bi, X. & Wang, H. Early Alzheimer's disease diagnosis based on EEG spectral images using deep learning. *Neural Netw* **114,** 119–135. ISSN: 1879-2782 (Electronic) 0893-6080 (Linking). (0006) (2019).

27. Vuckovic, A., Gallardo, V. J. F., Jarjees, M., Fraser, M. & Purcell, M. Prediction of central neuropathic pain in spinal cord injury based on EEG classifier. *Clin Neurophysiol* **129,** 1605–1617. ISSN: 1872-8952 (Electronic) 1388-2457 (Linking). (0008) (2018).

28. Erani, F. *et al.* Electroencephalography Might Improve Diagnosis of Acute Stroke and Large Vessel Occlusion. *Stroke* **51,** 3361–3365. ISSN: 1524-4628 (Electronic) 0039-2499 (Print) 0039-2499 (Linking). (0011) (2020).

29. Seo, J., Laine, T. H., Oh, G. & Sohn, K. A. EEG-Based Emotion Classification for Alzheimer's Disease Patients Using Conventional Machine Learning and Recurrent Neural Network Models. *Sensors (Basel)* **20.** ISSN: 1424-8220 (Electronic) 1424-8220 (Linking). (0012) (2020).

30. Klepl, D., He, F., Wu, M., Blackburn, D. J. & Sarrigiannis, P. EEG-Based Graph Neural Network Classification of Alzheimer's Disease: An Empirical Evaluation of Functional Connectivity Methods. *IEEE Trans Neural Syst Rehabil Eng* **30,** 2651–2660. ISSN: 1558-0210 (Electronic) 1534-4320 (Linking) (2022).

31. Kim, D. & Kim, K. Detection of Early Stage Alzheimer's Disease using EEG Relative Power with Deep Neural Network. *Annu Int Conf IEEE Eng Med Biol Soc* **2018,** 352–355. ISSN: 2694-0604 (Electronic) 2375-7477 (Linking). (0007) (2018).

32. Chang, H., Liu, B., Zong, Y., Lu, C. & Wang, X. EEG-Based Parkinson's Disease Recognition via Attention-Based Sparse Graph Convolutional Neural Network. *IEEE J Biomed Health Inform* **27,** 5216–5224. ISSN: 2168-2208 (Electronic) 2168-2194 (Linking). (0011) (2023).

33. Rijnders, B., Korkmaz, E. E. & Yildirim, F. Hybrid machine learning method for a connectivity-based epilepsy diagnosis with resting-state EEG. *Med Biol Eng Comput* **60,** 1675–1689. ISSN: 1741-0444 (Electronic) 0140-0118 (Linking). (0006) (2022).

34. Sibilano, E. *et al.* An attention-based deep learning approach for the classification of subjective cognitive decline and mild cognitive impairment using resting-state EEG. *J Neural Eng* **20.** ISSN: 1741-2552 (Electronic) 1741-2552 (Linking). (0002) (2023).

35. Jeong, T., Park, U. & Kang, S. W. Novel quantitative electroencephalogram feature image adapted for deep learning: Verification through classification of Alzheimer's disease dementia. *Front Neurosci* **16,** 1033379. ISSN: 1662-4548 (Print) 1662-453X (Electronic) 1662-453X (Linking) (2022).

36. Chu, C. *et al.* Deep learning reveals personalized spatial spectral abnormalities of high delta and low alpha bands in EEG of patients with early Parkinson's disease. *J Neural Eng* **18.** ISSN: 1741-2552 (Electronic) 1741-2552 (Linking). (0012) (2021).

37. Alessandrini, M. *et al.* EEG-Based Alzheimer's Disease Recognition Using Robust-PCA and LSTM Recurrent Neural Network. *Sensors (Basel)* **22.** ISSN: 1424-8220 (Electronic) 1424-8220 (Linking). (0005) (2022).

38. Li, S. *et al.* Aided diagnosis of cervical spondylotic myelopathy using deep learning methods based on electroencephalography. *Med Eng Phys* **121,** 104069. ISSN: 1873-4030 (Electronic) 1350-4533 (Linking). (0011) (2023).

39. Bouallegue, G., Djemal, R., Alshebeili, S. A. & Aldhalaan, H. A Dynamic Filtering DF-RNN Deep-Learning-Based Approach for EEG-Based Neurological Disorders Diagnosis. English. *IEEE Access* **8,** 206992–207007. ISSN: 2169-3536 (2020).

40. Thanjavur, K. *et al.* Recurrent neural network-based acute concussion classifier using raw resting state EEG data. *Sci Rep* **11,** 12353. ISSN: 2045-2322 (Electronic) 2045-2322 (Linking). (0006) (2021).

41. Choi, Y. A. *et al.* Deep Learning-Based Stroke Disease Prediction System Using Real-Time Bio Signals. *Sensors (Basel)* **21.** ISSN: 1424-8220 (Electronic) 1424-8220 (Linking). (0006) (2021).

42. Şeker, M. & Özerdem, M. S. Deep insights into MCI diagnosis: A comparative deep learning analysis of EEG time series. *J Neurosci Methods* **403,** 110057. ISSN: 1872-678X (Electronic) 0165-0270 (Linking). (0003) (2024).

43. Hata, M. *et al.* Precise Discrimination for Multiple Etiologies of Dementia Cases Based on Deep Learning with Electroencephalography. *Neuropsychobiology* **82,** 81–90. ISSN: 1423-0224 (Electronic) 0302-282X (Linking) (2023).

44. Shahin, M. *et al.* Deep Learning and Insomnia: Assisting Clinicians With Their Diagnosis. *IEEE J Biomed Health Inform* **21,** 1546–1553. ISSN: 2168-2208 (Electronic) 2168-2194 (Linking). (0011) (2017).

45. Mohammadi, M. R. *et al.* EEG Classification of ADHD and Normal Children Using Non-linear Features and Neural Network. English. *Biomed. Eng. Lett.* **6,** 66–73. ISSN: 2093-9868. (0005) (2016).

46. Doborjeh, M. *et al.* Prediction of Tinnitus Treatment Outcomes Based on EEG Sensors and TFI Score Using Deep Learning. *Sensors (Basel)* **23.** ISSN: 1424-8220 (Electronic) 1424-8220 (Linking). (0001) (2023).

47. Prabhakar, S. K. & Rajaguru, H. Performance Analysis of ApEn as a Feature Extraction Technique and Time Delay Neural Networks, Multi Layer Perceptron as Post Classifiers for the Classification of Epilepsy Risk Levels from EEG Signals. English. *Computational Intelligence, Cyber Security and Computational Models, Icc3 2015* **412,** 89–97. ISSN: 2194-5357 (2016).

48. J, P. *et al.* Detection of Focal and Non-Focal Electroencephalogram Signals Using Fast Walsh-Hadamard Transform and Artificial Neural Network. *Sensors (Basel)* **20.** ISSN: 1424-8220 (Electronic) 1424-8220 (Linking). (0009) (2020).

49. Lin, Q. *et al.* Classification of Epileptic EEG Signals with Stacked Sparse Autoencoder Based on Deep Learning. English. *Intelligent Computing Methodologies, Icic 2016, Pt Iii* **9773,** 802–810. ISSN: 0302-9743 (2016).

50. Cabrerizo, M., Ayala, M., Jayakar, P. & Adjouadi, M. Classification and Medical Diagnosis of Scalp Eeg Using Artificial Neural Networks. English. *Int. J. Innov. Comp. Inf. Control* **7,** 6905–6918. ISSN: 1349-4198. (0012) (2011).

51. Akben, S. B., Subasi, A. & Tuncel, D. Analysis of repetitive flash stimulation frequencies and record periods to detect migraine using artificial neural network. *J Med Syst* **36,** 925–931. ISSN: 0148-5598 (Print) 0148-5598 (Linking). (0004) (2012).

52. Brink-Kjaer, A. *et al.* End-to-end Deep Learning of Polysomnograms for Classification of REM Sleep Behavior Disorder. *Annu Int Conf IEEE Eng Med Biol Soc* **2022,** 2941–2944. ISSN: 2694-0604 (Electronic) 2375-7477 (Linking). (0007) (2022).

53. Barnes, L. D., Lee, K., Kempa-Liehr, A. W. & Hallum, L. E. Detection of sleep apnea from single-channel electroencephalogram (EEG) using an explainable convolutional neural network (CNN). *PLoS One* **17,** e0272167. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking) (2022).

54. Zheng, W. L. *et al.* Predicting neurological outcome in comatose patients after cardiac arrest with multiscale deep neural networks. *Resuscitation* **169,** 86–94. ISSN: 1873-1570 (Electronic) 0300-9572 (Print) 0300-9572 (Linking). (0012) (2021).

55. Tjepkema-Cloostermans, M. C. *et al.* Outcome Prediction in Postanoxic Coma With Deep Learning. *Crit Care Med* **47,** 1424–1432. ISSN: 1530-0293 (Electronic) 0090-3493 (Linking). (0010) (2019).

56. Raurale, S. A. *et al.* Grading hypoxic-ischemic encephalopathy in neonatal EEG with convolutional neural networks and quadratic time-frequency distributions. *J Neural Eng* **18.** ISSN: 1741-2552 (Electronic) 1741-2560 (Print) 1741-2552 (Linking). (0003) (2021).

57. Lai, C. Q. *et al.* Detection of Moderate Traumatic Brain Injury from Resting-State Eye-Closed Electroencephalography. *Comput Intell Neurosci* **2020,** 8923906. ISSN: 1687-5273 (Electronic) 1687-5265 (Print) (2020).

58. Jonas, S. *et al.* EEG-based outcome prediction after cardiac arrest with convolutional neural networks: Performance and visualization of discriminative features. *Hum Brain Mapp* **40,** 4606–4617. ISSN: 1097-0193 (Electronic) 1065-9471 (Print) 1065-9471 (Linking). (0011) (2019).

59. Boshra, R., Ruiter, K. I., DeMatteo, C., Reilly, J. P. & Connolly, J. F. Neurophysiological Correlates of Concussion: Deep Learning for Clinical Assessment. *Sci Rep* **9,** 17341. ISSN: 2045-2322 (Electronic) 2045-2322 (Linking). (0011) (2019).

60. Kim, M. J., Youn, Y. C. & Paik, J. Deep learning-based EEG analysis to classify normal, mild cognitive impairment, and dementia: Algorithms and dataset. *Neuroimage* **272,** 120054. ISSN: 1095-9572 (Electronic) 1053-8119 (Linking). (0005) (2023).

61. Dar, M. N., Akram, M. U., Yuvaraj, R., Gul Khawaja, S. & Murugappan, M. EEG-based emotion charting for Parkinson's disease patients using Convolutional Recurrent Neural Networks and cross dataset learning. *Comput Biol Med* **144,** 105327. ISSN: 1879-0534 (Electronic) 0010-4825 (Linking). (0005) (2022).

62. Amini, M., Pedram, M., Moradi, A. & Ouchani, M. Diagnosis of Alzheimer's Disease by Time-Dependent Power Spectrum Descriptors and Convolutional Neural Network Using EEG Signal. *Comput Math Methods Med* **2021,** 5511922. ISSN: 1748-6718 (Electronic) 1748-670X (Print) 1748-670X (Linking) (2021).

63. Makaram, N. *et al.* Deep Learning-Based Visual Complexity Analysis of Electroencephalography Time-Frequency Images: Can It Localize the Epileptogenic Zone in the Brain? English. *Algorithms* **16.** (0012) (2023).

64. Misiunas, A. V. M., Rapsevicius, V., Samaitiene, R. & Meskauskas, T. Electroencephalogram spike detection and classification by diagnosis with convolutional neural network. English. *Nonlinear Anal.-Model Control* **25,** 692–704. ISSN: 1392-5113 (2020).

65. Yang, S. *et al.* Early prediction of drug-resistant epilepsy using clinical and EEG features based on convolutional neural network. *Seizure* **114,** 98–104. ISSN: 1532-2688 (Electronic) 1059-1311 (Linking). (0001) (2024).

66. Liu, X. *et al.* Classification of self-limited epilepsy with centrotemporal spikes by classical machine learning and deep learning based on electroencephalogram data. *Brain Res* **1830,** 148813. ISSN: 1872-6240 (Electronic) 0006-8993 (Linking). (0002) (2024).

67. Alshaya, H. & Hussain, M. Classification of Epileptic Seizure Types Using Multiscale Convolutional Neural Network and Long Short-Term Memory. English. *Mathematics* **11.** (0009) (2023).

68. San-Segundo, R., Gil-Martín, M., D'Haro-Enríquez, L. F. & Pardo, J. M. Classification of epileptic EEG recordings using signal transforms and convolutional neural networks. *Comput Biol Med* **109,** 148–158. ISSN: 1879-0534 (Electronic) 0010-4825 (Linking). (0006) (2019).

69. You, Z. *et al.* Alzheimer's Disease Classification With a Cascade Neural Network. *Front Public Health* **8,** 584387. ISSN: 2296-2565 (Print) 2296-2565 (Electronic) 2296-2565 (Linking) (2020).

70. Lin, P. J. *et al.* A Transferable Deep Learning Prognosis Model for Predicting Stroke Patients' Recovery in Different Rehabilitation Trainings. *IEEE J Biomed Health Inform* **26,** 6003–6011. ISSN: 2168-2208 (Electronic) 2168-2194 (Linking). (0012) (2022).

71. Drage, R. *et al.* A novel deep learning approach using AlexNet for the classification of electroencephalograms in Alzheimer's Disease and Mild Cognitive Impairment. *Annu Int Conf IEEE Eng Med Biol Soc* **2022,** 3175–3178. ISSN: 2694-0604 (Electronic) 2375-7477 (Linking). (0007) (2022).

72. Chirasani, S. K. R. & Manikandan, S. A deep neural network for the classification of epileptic seizures using hierarchical attention mechanism. *Soft comput* **26,** 5389–5397. ISSN: 1432-7643 (Print) 1433-7479 (Electronic) 1432-7643 (Linking) (2022).

73. Watanabe, Y. *et al.* A deep learning model for the detection of various dementia and MCI pathologies based on resting-state electroencephalography data: A retrospective multicentre study. *Neural Netw* **171,** 242–250. ISSN: 1879-2782 (Electronic) 0893-6080 (Linking). (0003) (2024).

74. Lee, S., Hussein, R., Ward, R., Jane Wang, Z. & McKeown, M. J. A convolutional-recurrent neural network approach to resting-state EEG classification in Parkinson's disease. *J Neurosci Methods* **361,** 109282. ISSN: 1872-678X (Electronic) 0165-0270 (Print) 0165-0270 (Linking). (0009) (2021).

75. Lai, C. Q., Ibrahim, H., Abd Hamid, A. I. & Abdullah, J. M. Classification of Non-Severe Traumatic Brain Injury from Resting-State EEG Signal Using LSTM Network with ECOC-SVM. *Sensors (Basel)* **20.** ISSN: 1424-8220 (Electronic) 1424-8220 (Linking). (0009) (2020).

76. Prabhakar, S. K. & Lee, S. W. SASDL and RBATQ: Sparse Autoencoder With Swarm Based Deep Learning and Reinforcement Based Q-Learning for EEG Classification. *IEEE Open J Eng Med Biol* **3,** 58–68. ISSN: 2644-1276 (Electronic) 2644-1276 (Linking) (2022).

77. Cao, J., Yang, L., Sarrigiannis, P. G., Blackburn, D. & Zhao, Y. Dementia classification using a graph neural network on imaging of effective brain connectivity. *Comput Biol Med* **168,** 107701. ISSN: 1879-0534 (Electronic) 0010-4825 (Linking). (0001) (2024).

78. Isensee, F. *et al. nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation* arXiv:2404.09556 [cs]. Apr. 2024. `http://arxiv.org/abs/2404.09556` (2024).

79. Yıldırım, E. & Soncu Büyükişcan, E. Default Mode Network Connectivity in Alzheimers Disease. eng, tur. *Turk Psikiyatri Dergisi = Turkish Journal of Psychiatry* **30,** 279–286. ISSN: 2651-3463 (2019).

80. Raab, D., Theissler, A. & Spiliopoulou, M. XAI4EEG: spectral and spatio-temporal explanation of deep learning-based seizure detection in EEG time series. en. *Neural Computing and Applications* **35,** 10051–10068. ISSN: 1433-3058. `https://doi.org/10.1007/s00521-022-07809-x` (2024) (May 2023).

81. Huang, T., Ma, L., Zhang, B. & Liao, H. Advances in deep learning: From diagnosis to treatment. eng. *Bioscience Trends* **17,** 190–192. ISSN: 1881-7823 (July 2023).

82. Page, M. J. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. eng. *BMJ (Clinical research ed.)* **372,** n71. ISSN: 1756-1833 (Mar. 2021).

Appendix

A. Search string

PubMed

("Artificial Intelligence"[mj] OR Artificial-Intelligen*[ti] OR Deep-Learning*[ti] OR Deep-transfer-Learning*[ti] OR Machine-Learning*[ti] OR neural-network*[ti]) AND (Electroencephalography[mj] OR Electroencephalogra*[ti] OR eeg[ti] OR eegs[ti] OR ieeg[ti]) AND ("Nervous System Diseases"[mh] OR Nervous-System-Dis*[tiab] OR Neurological-Dis*[tiab] OR Neurologic-Dis*[tiab] OR Epilep*[tiab] OR Voice-Dis*[tiab] OR Amyotrophic-Lateral-Sclero*[tiab]) AND ("Classification"[mh] OR Classification[sh] OR Classification*[tiab] OR Prognosis[mh] OR Prognos*[tiab] OR predict*[tiab] OR Phenotype[mh] OR Phenotyp*[tiab] OR subphenotyp*[tiab]) AND 2010:2024[dp] NOT (animals[mh] NOT humans[mh]) AND English[la]

Web of Science

(TI=Artificial-Intelligen* OR TI=Deep-Learning* OR TI=Deep-transfer-Learning* OR TI=Machine-Learning* OR TI=neural-network*) AND (TI=Electroencephalogra* OR TI=eeg OR TI=eegs OR TI=ieeg) AND ((TI=Nervous-System-Dis* OR AB=Nervous-System-Dis*) OR (TI=Neurological-Dis* OR AB=Neurological-Dis*) OR (TI=Neurologic-Dis* OR AB=Neurologic-Dis*) OR (TI=Epilep* OR AB=Epilep*) OR (TI=Voice-Dis* OR AB=Voice-Dis*) OR (TI=Amyotrophic-Lateral-Sclero* OR AB=Amyotrophic-Lateral-Sclero*)) AND ((TI=Classification* OR AB=Classification*) OR (TI=Prognos* OR AB=Prognos*) OR (TI=predict* OR AB=predict*) OR (TI=Phenotyp* OR AB=Phenotyp*) OR (TI=subphenotyp* OR AB=subphenotyp*)) AND PY=(2010-2024) AND LA=(english)

B. Flowchart

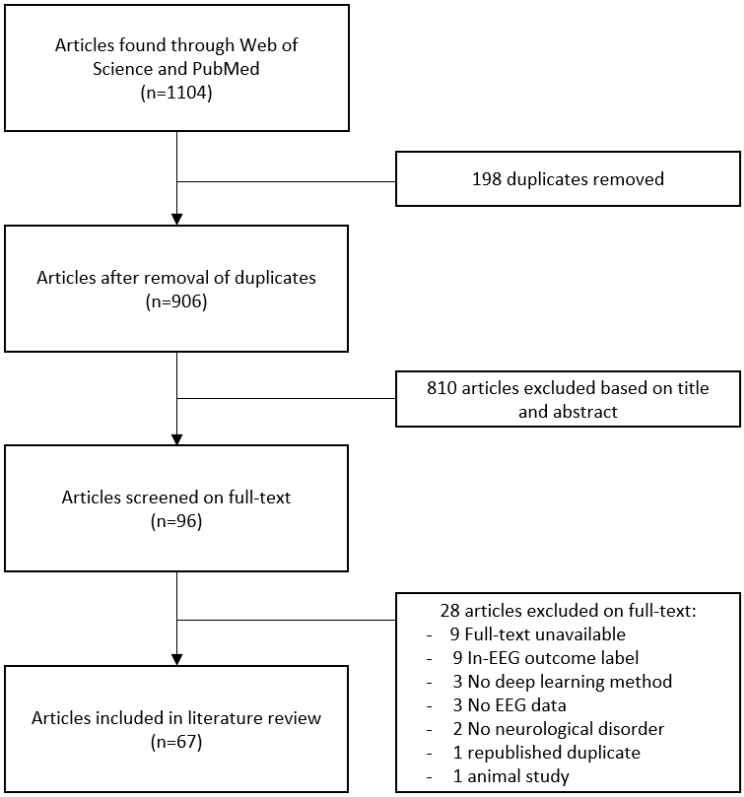

Figure 1: A flowchart describing the different steps of article inclusion