Artificial Intelligence in Digital Medicine Across Biomedical Signal Processing and Medical Imaging

Simone Bove[1,†], Rosanna Ferrara[1,†], Martino Giaquinto[1,†], Gennaro Percannella[1,†], Alessia Saggese[1,*,†], Mattia Sarno[1,†], Francesco Tortorella[1,†] and Mario Vento[1,†]

[1]*Department of Information Engineering, Electrical Engineering, and Applied Mathematics (DIEM), University of Salerno*

Abstract

Artificial Intelligence is increasingly offering innovative solutions in the field of digital medicine. From wearable devices that enable continuous physiological tracking to intelligent algorithms that interpret complex medical images, artificial intelligence is introducing novel methods that enhance current clinical approaches. In this work, we propose new solutions for efficient neurological monitoring, brain computer interface systems and robust medical image analysis, addressing key challenges in real-world applicability, personalization and cross-domain generalization.

Keywords

EEG Monitoring, Seizure Detection, Motor Imagery Classification, Medical Image Analysis

1. Introduction

In recent years, the integration of Artificial Intelligence (AI) into digital health has opened new frontiers in both diagnostics and patient care. AI technologies are enabling more precise, data-driven approaches to medicine, with applications ranging from personalized treatment strategies to the use of wearable sensors for continuous health monitoring. Among the most promising developments is the use of AI in analyzing complex biomedical signals and medical images, enhancing the ability to detect subtle patterns and make informed clinical decisions. Within this evolving landscape, new methods are being explored to improve efficiency, usability, and performance in real-world healthcare settings. In this framework, wearable electroencephalography (EEG) based systems are emerging for continuous, real-time monitoring of neurological conditions. However, conventional setups with 21 or more electrodes, following the 10–20 system, often hinder practicality due to their intrusiveness, discomfort, and increased data complexity. To address these limitations, we recently proposed a compact deep learning framework that integrates personalized electrode selection with efficient signal processing for epileptic patients monitoring, specifically optimized for deployment on resource-constrained hardware [1]. Beyond diagnostics, EEG is becoming increasingly attractive in the field of Brain-Computer Interfaces (BCIs), especially with the motor imagery (MI) paradigm [2], based on classification of imagined movement tasks from EEG signals. This enables users to control assistive devices without physical input, with impactful applications in neurorehabilitation, prosthetic control, and intuitive human-machine interaction. While most studies focus on classifying between MI types, the more fundamental task of detecting MI onset against background brain activity remains relatively underexplored. With a look toward real-world applications, we investigated two alternative configurations to improve the accuracy and robustness of MI event detection in the presence of background neural activity [3]. Medical imaging plays a crucial role in modern medicine, enabling non-invasive diagnosis, prognosis, and treatment planning. However, accurately processing and interpreting medical images to identify clinically meaningful

Ital-IA 2025: 5th National Conference on Artificial Intelligence, organized by CINI, June 23-24, 2025, Trieste, Italy

*Corresponding author.

†These authors contributed equally.

sbove@unisa.it (S. Bove); roferrara@unisa.it (R. Ferrara); mgiaquinto@unisa.it (M. Giaquinto); pergen@unisa.it (G. Percannella); asaggese@unisa.it (A. Saggese); masarno@unisa.it (M. Sarno); ftortorella@unisa.it (F. Tortorella); mvento@unisa.it (M. Vento)

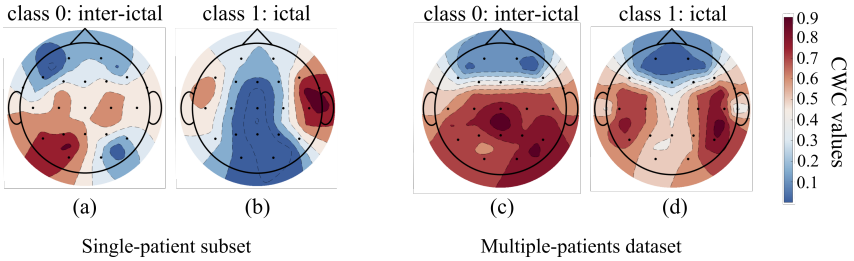

Figure 1: Average CWC maps highlighting neural activity localization during inter-ictal (a,c) and ictal (b,d) phases. Maps in (a,b) are computed from all records of patient chb01 (example case), while (c,d) show the average across all patients in the dataset.

patterns exceeds human capabilities. Hence, researchers began developing systems for automatically analyze such images. AI, particularly deep learning (DL), has proven highly effective for tasks such as disease classification, localization, and segmentation on medical images, by automatically learning relevant features for such task from raw images. However, DL models require large, annotated datasets, which is a challenge in the medical domain, where the available datasets are highly heterogeneous, varying in terms of tasks they are designed for, as well as in terms of domains from which they are sampled, e.g., different datasets may contain images acquired through different scanners or protocols. Such heterogeneity poses two main challenges: (1) developing models that leverage related tasks to improve performance, and (2) ensuring models trained on limited domains generalize to unseen ones without significant performance loss. The first challenge has been addressed in studies on Hep-2 cell analysis, while ongoing research focuses on the second in the context of histopathology image analysis.

2. Seizure Detection based on EEG

The proposed system for seizure detection is based on a lightweight convolutional neural network (CNN) optimized for real-time operation and tailored to individual patients through patient-specific electrode configuration and model training. A data-driven, classifier-independent channel selection method reduces the standard EEG setup to just two personalized channels, using eigen-decomposition-derived weights to ensure robustness and interpretability. This approach maintains diagnostic accuracy while minimizing computational load, enabling efficient deployment on wearable, non-invasive EEG devices.

2.1. Experimental Results

The approach was validated on the CHB-MIT EEG dataset [4], which includes 686 records from 24 pediatric patients, sampled at 256 Hz. The dataset contains over 950 hours of EEG data with annotated ictal and inter-ictal periods. The signals were denoised using a 60 Hz notch filter and standardized with z-score normalization. To identify the most informative EEG channels, we used an alternative representation that highlights scalp regions with the highest neural activity over time, following the three-step procedure described in [1]. First, the signals were filtered using a second-order Butterworth bandpass filter to isolate standard clinical frequency bands: δ [0.5–4 Hz), θ [4–8 Hz), α [8–13 Hz), β [13–35 Hz), and γ [35–80 Hz). The filtered data were segmented into 1-second windows, each of them decomposed using principal component analysis (PCA) into 21 principal components (PCs). For each window, channel contributions (loadings) were weighted by the corresponding PC's explained variance, and summed to compute a single Channel Weight Coefficient (CWC) per channel. In the second step, CWCs were averaged across time and frequency bands, with min-max normalization applied per band to ensure equal weighting. This process was repeated for all records of a given patient, obtaining normalized average CWC maps for inter-ictal and ictal phases. The most informative channels were selected based on the highest CWCs in each phase. Applying this method across all patients revealed personalized informative channel pairs, as illustrated for patient chb01 in Fig. 1a,b, for inter-ictal and ictal phases respectively. Data from these selected channels were segmented into non-overlapping

Table 1
System performance of the seizure detection system across the different configurations analyzed

Metrics	Patient-specific system			Hybrid system
	ictal channel	inter-ictal channel	dual-channel	dual-channel
Balanced accuracy	0.77 ± 0.17	0.69 ± 0.18	0.83 ± 0.15	0.76 ± 0.18
Detection delay [s]	13.6 ± 11.1	17.7 ± 15.1	11.5 ± 9.75	13.1 ± 9.6
False positive per hour	0.21 ± 0.27	0.09 ± 0.13	0.15 ± 0.18	0.17 ± 0.21
Detected seizures number	136 / 173	98 / 173	152 / 173	130 / 173

4-second windows, which were used to train a lightweight, patient-specific, classification model to distinguish between inter-ictal (class 0) and ictal (class 1) segments. The detection model is based on a CNN optimized for low-power, microcontroller-based platforms [1] and adapted to support either two or one input channel [5], enabling evaluation of two system setups: the Dual-Channel and Single-Channel configurations. Table 1 summarizes the results, reported as mean and standard deviation across all patients, for both configurations. The Dual-Channel configuration provides the best overall performance, achieving a balanced accuracy of 0.83. In more details, it correctly detected 152 out of 173 ictal events, with the shortest average detection delay of 11.5 seconds and a moderate false positive (FP) rate of 0.15 FP/h. Among the Single-Channel configurations, the ictal channel performs better in accuracy detection (0.77) but has a higher false positive rate (FP/h = 0.21). In contrast, the inter-ictal channel is less accurate (0.69) but far more robust against FP, reaching a FP/h of 0.09. Overall, Dual-Channel setup strikes the best balance between invasiveness and performance, which are comparable, and in some cases even superior, to those of recently proposed systems that rely on a large number of channels [1]. Interestingly, applying the selection algorithm across all subjects allows for a generalizable channel selection strategy [6]. Following this approach, as illustrated in Fig. 1c,d, the inter-ictal phase exhibits widespread activity with a peak at the central area, which shifts toward the right temporal regions, during the ictal phase. Therefore, to reduce the dependency on patient-specific electrode setups, we also evaluated a hybrid configuration, combining a globally optimized channel pair with a patient-specific model [6]. As shown in the last column of Table 1, this approach allowed to reach a balanced accuracy of 0.76, detecting 130 ictal events with an average delay of 13 seconds, while maintaining a relatively low false positive rate of 0.17 FP/h. These results make the hybrid system as a promising compromise between detection performance and practical implementation.

3. Motor Imagery based on EEG

MI decoding from EEG is traditionally addressed as a classification problem between different MI classes (e.g. left vs right hans movement). However, real-world applications require distinguishing intentional MI from background brain activity. We addressed this challenge by comparing two strategies and including a background class across those pertaining to MI tasks: a hierarchical two-stage model and a compact end-to-end approach. Both show a performance drop when detecting background activity, but the end-to-end model offers a better balance of accuracy and complexity, making it more suitable for portable use.

3.1. Experimental Results

We analyzed EEG data from the BCI Competition 2008 Graz dataset B [7], with nine subjects performing left- and right-hand motor imagery tasks. The signals were recorded at 250 Hz from electrodes C3, C4, and Cz (referenced to Fz), including background activity during fixation and breaks, and 4-second MI segments registered after a cue. Preprocessing involved bandpass filtering (in the range [4–49] Hz), outlier removal, segmentation into 4-second windows labeled as left, right, or background (taken from 4 seconds before cues), and exclusion of artifact-contaminated segments, resulting in a background class twice the size of each event class. Data were z-score normalized per channel using training statistics. Two configurations were tested: (i) a hierarchical multistage approach separating movement detection

Table 2
Balanced accuracy and model complexity for the Multistage (MS) Classification approach (Stage 1, Stage 2, Combined) and the Single-Stage Classification approach.

	Balanced accuracy										Model Complexity	
Classifier	**B01**	**B02**	**B03**	**B04**	**B05**	**B06**	**B07**	**B08**	**B09**	**Average**	**# Params (k)**	**FLOPs (M)**
MS-Stage 1	0.93±0.01	0.79±0.01	0.78±0.01	0.93±0.01	0.85±0.02	0.94±0.01	0.84±0.01	0.88±0.01	0.86±0.02	0.87±0.06	2.17	1.84
MS-Stage 2	0.69±0.02	0.58±0.03	0.64±0.03	0.97±0.01	0.89±0.01	0.81±0.02	0.86±0.02	0.91±0.01	0.84±0.02	0.80±0.13	2.17	1.84
MS-Combined	0.75±0.02	0.57±0.02	0.58±0.01	0.92±0.01	0.82±0.01	0.80±0.02	0.77±0.02	0.85±0.01	0.75±0.02	0.76±0.12	4.34	3.68
Single stage	0.74±0.02	0.56±0.02	0.59±0.02	0.91±0.01	0.79±0.02	0.79±0.02	0.77±0.01	0.83±0.02	0.72±0.03	0.74±0.11	2.67	1.84

and MI classification, and (ii) a single-stage multi-class classifier distinguishing background, left, and right MI. Both used EEGNet [2], a compact CNN effective in capturing EEG spatial-temporal patterns with few parameters. Models were patient-specific, trained via 5-fold cross-validation, and evaluated on the dataset's separate test set.

The overall performance of the Multistage Classification system is summarized in Table 2 (third row), showing an average balanced accuracy of 0.76 (±0.12) across nine subjects. Subject B04 achieved the highest accuracy at 0.92, while B02 had the lowest at 0.57. To better understand these results, it is useful to examine the performance of each stage individually. Stage 1, which distinguished combined left/right movements (event) from background, demonstrated strong discriminative capability, reaching an average balanced accuracy of 0.87 (±0.06), as reported in the first row of Table 2. Several subjects, including B01, B04, and B06, exceeded 0.90, with the lowest accuracy being 0.77 (B03). In Stage 2, EEGNet was implemented in a configuration widely adopted in the literature to classify the two event sub-classes: left and right movements, achieving an average balanced accuracy of 0.80, in line with previously reported findings [2]. Stage 2 showed greater inter-subject variability (0.13), impacting overall system performance. In the second approach, EEGNet was configured to directly classify among the three target classes, i.e. background, left, and right, within a single-stage architecture. This streamlined setup yielded an average balanced accuracy of 0.74 across all subjects, with a standard deviation of 0.11, indicating moderate inter-subject variability. Comparing this to the multistage method (Table 2, last vs. third row), the single stage model's performance is comparable, with only about a 2% lower accuracy, indicating that the added complexity of the multistage system offers limited gains. Subject-specific trends were consistent across both methods, highlighting the impact of inter-subject variability, independently on the method. Notably, the single stage approach halves both the number of parameters (from 4.34k to 2.67k) and the computational cost (FLOPs from 3.68M to 1.84M), making it better suited for resource-limited or portable applications.

4. AI in the context of medical image analysis

As anticipated in Section 1, the use of deep neural networks in the medical imaging context poses two main challenges, i.e., the development of models capable of exploiting the inherent knowledge related to distinct but related task, and the development of robust models against shifts in the data domain.

4.1. Experimental results

To address the first challenge, we proposed a multi-task learning (MTL) approach for developing a model capable of simultaneously carrying out multiple medical imaging tasks, related to image segmentation and classification. As pointed out in [8], indeed, there exist inherent or complementary relationships among different medical imaging tasks. Based on such observations, we propose a MTL approach for carrying out multiple diagnostic tasks, related to image segmentation and image classification. Specifically, we propose an architecture, built on the well-known U-Net, which is designed to learn multiple tasks at once, by learning those latent representations from input data, that are useful for all

the tasks. The proposed approach was evaluated in the context of Hep-2 image analysis. Specifically, in a preliminary work [9], the proposed approach was evaluated in the tasks of intensity classification and specimen segmentation, showing promising outcomes; then, the work was extended to the task of staining pattern classification as third task [10]. To jointly handle classification and segmentation tasks on medical images, a modification of the U-Net architecture was proposed.

Compared to the standard U-Net, the proposed architecture, reported in Figure 2, shows additional

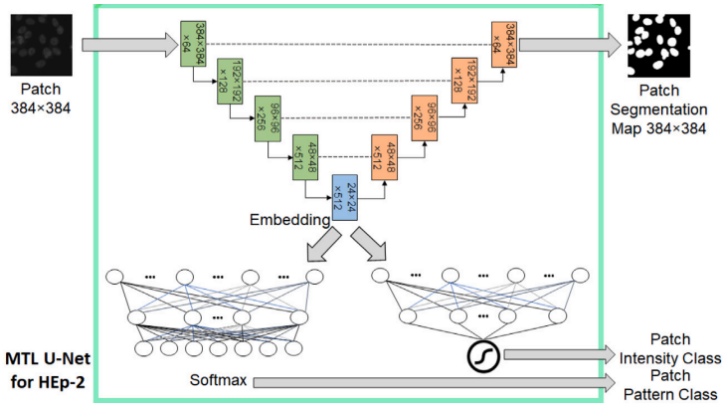

Figure 2: Proposed MTL architecture for Hep-2 specimen segmentation, intensity classification and staining pattern classification.

task-specific heads, connected at the end of the U-Net encoding path. Each head was designed to carry out a certain task by exploiting the latent representations extracted by the encoder, which are therefore shared across multiple tasks.

The proposed approach was evaluated on the dataset I3A [11], and, specifically, on the portion named I3A Task-2, devoted to the specimen-level Hep-2 image analysis. This portion was obtained by 252 distinct wells, each obtained by a distinct patient. In particular, each well was sampled in four distinct and not overlapped locations, providing four specimen images per well. Overall the I3A Task-2 dataset consists of 1008 specimen images.

Table 3
Performance of our MTL U-Net compared to single task models. The values are expressed as mean ± standard deviation across the 100 repetitions (in stratified random partitioning 80% train and 20% test).

Task	Method	Metric	Performance (%)
Specimen Segmentation	U-Net	Segmentation Accuracy	89.27 ± 3.44
	MTL U-Net		**90.49** ± 1.92
Intensity Classification	DenseNet121	Accuracy	91.42 ± 3.61
	Single Trunk		88.78 ± 1.96
	MTL U-Net		**95.83** ± 2.21
Pattern Classification	DenseNet121	MCA	88.91 ± 3.42
	Single Trunk		88.95 ± 1.45
	MTL U-Net		**96.02** ± 2.19

Experiments have been carried out based on a stratified random train/test partitioning, which we repeated 100 times. Specifically, we maintained the distribution of class labels by adopting a stratified sampling strategy. During each iteration, we allocated 80% of the samples 20% to the test set. Additionally, we ensured that samples from the same patient were exclusively assigned to either the training or test subset. From the experimental results, shown in Table 3, it is possible to observe that the proposed method improves state-of-the art methods in all the tasks of specimen segmentation, staining pattern classification, and intensity classification, in terms of segmentation accuracy (SA), mean class accuracy (MCA) and accuracy respectively.

4.2. Future directions: robust medical imaging algorithms under domain shift

Recently, a research activity concerning the development of AI models robust to domain shifts in the context of medical imaging has been launched. Domain shift arises when test data differ in distribution from training data, due to variations in scanners, acquisition protocols, or patient populations. This shift often leads to a significant drop in the performance of deep neural networks. As a result, there is growing interest in developing methods for improving robustness in the area of domain adaptation and domain generalization. In medical imaging, this problem is particularly relevant, as the available datasets for training an AI model are scarce and are not fully representative of the actual variability in the data. Specifically, our research activity focuses on addressing domain shift in histopathology, a problem that gained growing interest in recent years, in which domain shift across images can be caused by variations in the acquisition scanner or protocols, as well as by variations in depicted tumor types or sample's species.

Acknowledgements

This work was partially supported by MUR (Italian Ministry for University and Research) through the DIEM Department of Excellence Project 2023–2027 (law 232/2016) and the project "MUltimodal, Multitask, unBalanced Learning Environments (MUMBLE).

Generative AI Declaration

During the preparation of this work, the authors used ChatGPT in order to: grammar and spelling check, paraphrase and reword. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] R. Ferrara, M. Giaquinto, G. Percannella, L. Rundo, A. Saggese, Personalizing seizure detection for individual patients by optimal selection of eeg signals, Sensors 25 (2025) 2715.

[2] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, B. J. Lance, Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces, Journal of neural engineering 15 (2018) 056013.

[3] S. Bove, M. Giaquinto, G. Percannella, A. Saggese, M. Vento, The impact of non-task-related neural activity in eeg-based motor imagery classification, 2025 IEEE INTERNATIONAL CONFERENCE ON Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE25) (2025). [under review].

[4] J. Guttag, CHB-MIT Scalp EEG Database, https://doi.org/10.13026/C2K01R, 2010. Version 1.0.0, PhysioNet.

[5] R. Ferrara, M. Giaquinto, G. Percannella, A. Saggese, M. Vento, Wearable eeg systems: toward single-channel personalized configurations for seizure detection, in: Ninth National Congress of Bioengineering-Proceedings., Patron Editore, 2025. [Just accepted].

[6] R. Ferrara, M. Giaquinto, G. Percannella, A. Saggese, M. Vento, Personalized vs unpersonalized medicine in wearable based seizure detection, in: 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2025. [Just accepted].

[7] R. Leeb, C. Brunner, G. Müller-Putz, A. Schlögl, G. Pfurtscheller, Bci competition 2008–graz data set b, Graz University of Technology, Austria 16 (2008) 1–6.

[8] Y. Zhao, X. Wang, T. Che, G. Bao, S. Li, Multi-task deep learning for medical image computing and analysis: A review, Computers in Biology and Medicine 153 (2023) 106496.

[9] G. Percannella, U. Petruzzello, P. Ritrovato, L. Rundo, F. Tortorella, M. Vento, Joint intensity classification and specimen segmentation on hep-2 images: a deep learning approach, in: 2022 26th International Conference on Pattern Recognition (ICPR), 2022, pp. 4343–4349.
[10] G. Percannella, U. Petruzzello, F. Tortorella, M. Vento, A multi-task learning u-net model for end-to-end hep-2 cell image analysis, Artificial Intelligence in Medicine 159 (2025) 103031.
[11] P. Hobson, G. Percannella, M. Vento, A. Wiliem, Competition on cells classification by fluorescent image analysis, in: Proc. 20th IEEE Int. Conf. Image Process.(ICIP), 2013, pp. 2–9.