ChatGPT

Literature Review: Toward an AI-Orchestrated Visual EEG Artifact Detection System

Limitations of Existing EEG Artifact Detection

Electroencephalogram (EEG) recordings are notoriously prone to artifacts – unwanted signals from eye movements, muscle activity, cardiac pulses, line noise, and other non-cerebral sources. Traditional artifact detection/removal techniques include simple thresholding (e.g. rejecting segments exceeding a voltage limit), spectral filters (e.g. removing high-frequency muscle noise), regression using reference channels (e.g. subtracting EOG signals), and blind source separation (BSS) methods like independent component analysis (ICA). While these methods can mitigate artifacts, each has significant limitations. Human experts often must visually inspect and manually reject contaminated segments, which **is time-consuming and subjective** [1] [2]. Manual rejection indiscriminately discards data containing artifacts, at the cost of losing potentially meaningful brain signals [3]. The quality of cleaned EEG depends heavily on expert decisions and vigilance, which can vary between raters and may not scale to large datasets [1] [2]. No universally accepted "gold standard" for artifact identification exists – what is labeled artifact versus brain signal can vary by study and annotator [4] [5].

Many automated algorithms focus on *specific* artifact types or make strong assumptions, reducing their generalizability [6]. For example, dedicated blink detectors (using EOG channels or blink-specific features) work for ocular artifacts but won't catch muscle noise. Methods that rely on auxiliary sensors (EOG for eye movements, EMG for muscle) **depend on reference channels** that are not always available [6]. More sophisticated approaches like ICA can separate mixed sources, but **require expert intervention** to identify which components are artifactual [7]. This manual selection of ICA components is subjective and error-prone, leading to variability in results [7]. Even "automatic" ICA-based pipelines (e.g. ICLabel) still cannot avoid some human tuning or confirmation. Another issue is **computational burden**: techniques like wavelet decomposition or repeated ICA are computationally intensive, making them impractical for real-time or high-volume processing [8]. Simpler heuristics (e.g. fixed thresholds) are fast but too simplistic to handle overlapping or subtle artifacts. Consequently, no single method reliably addresses all artifact types under all conditions [9]. Researchers increasingly combine multiple methods (hybrid approaches) to improve coverage, yet this adds complexity [10]. In summary, current EEG artifact removal is often a patchwork of semi-automated steps and **manual review**, struggling with multi-factorial artifacts, subjectivity in what to remove, and inefficiency for growing EEG datasets [11]. These limitations underscore the need for more robust, generalizable, and efficient artifact detection solutions.

AI/ML in EEG Analysis

Given these challenges, the EEG field has embraced artificial intelligence (AI) and machine learning (ML) to enhance signal analysis. Broadly, ML techniques have achieved state-of-the-art performance in diverse EEG applications – from classifying mental states and pathologies to powering brain–computer interfaces (BCIs). For instance, ML models are used to detect epileptic seizures in EEG with high accuracy, decode motor imagery for BCI control, monitor depth of anesthesia, classify cognitive workload, and even recognize emotions [12] [13]. Early EEG ML approaches typically involved handcrafted feature extraction (e.g. power spectral features, wavelet coefficients, signal statistics) fed into classifiers like support vector machines or decision trees. Such techniques showed promise in tasks like sleep stage scoring and seizure detection, but their performance hinged on the quality of manually engineered features.

In recent years, deep learning has revolutionized EEG analysis by automatically learning features from raw data. **Convolutional neural networks (CNNs)** and recurrent networks (RNNs) have been successfully applied to EEG for event detection and classification [14]. In the context of artifact detection specifically, a variety of AI/ML strategies have emerged. Traditional ML approaches use statistical features (e.g. variance, kurtosis, frequency band ratios) as inputs to classifiers that label segments or channels as "artifact" vs "clean." These can outperform simple threshold rules, but still may miss complex patterns and usually require annotated training data. More recently, **deep learning approaches** learn artifact patterns directly. For example, *Nejedly et al.* (2018) trained a CNN to automatically identify artifacts in intracranial EEG recordings, using a fully automated image-based processing of the multi-channel signals [15]. Their network effectively learned to recognize artifact "signatures" (such as the stereotyped shape of an eye blink or burst of muscle noise) from raw data transformed into 2D representations. Similarly, researchers have applied 1D CNNs or LSTM networks on raw EEG time series to detect artifacts based on temporal patterns [16] [17]. Autoencoder networks have also been proposed to **flag artifacts via reconstruction error** – the idea being that a neural network trained to reconstruct clean EEG will produce larger errors on segments containing artifacts, thereby detecting them [17]. One such model (LSTEEG, an LSTM-based autoencoder) demonstrated superior artifact detection and correction performance compared to prior CNN autoencoders [18].

A notable trend is the integration of computer vision concepts into EEG artifact detection. Some approaches **convert EEG data into visual formats** so that CNNs can leverage their strength in image pattern recognition [19]. For instance, converting multi-channel EEG into an "image" matrix of shape (channels × time) allows use of 2D convolution filters across channels and time [19]. Other methods construct time–frequency images (spectrograms) or topographic maps of EEG channels as input to CNNs, enabling the network to detect artifacts via their distinctive spatial–spectral footprints (e.g. muscle artifacts show broadband high-frequency content, eye blinks show frontal dominance in scalp maps). Indeed, the **ICLabel** system – a recent plug-in for the EEGLAB software – uses a deep CNN to classify ICA components by inspecting their power spectra and scalp maps (essentially treating component features as images). It can distinguish multiple artifact types (eyes, muscle, heart, line noise, etc.) versus brain signal with around 83% accuracy [20]. This exemplifies how **hybrid AI approaches** combine traditional signal processing (ICA decomposition) with modern ML (CNN component labeling) to automate artifact handling. Overall, AI/ML methods in EEG analysis have advanced performance in many tasks, **including artifact detection**, by either mining richer statistical features or employing pattern recognition on raw or visually transformed data. However, most of these AI-based solutions operate as black-box classifiers or automated pipelines – they generally aim to replace manual artifact identification entirely, with less focus on collaborating with human experts. Few systems leverage the full visual intuition of human EEG readers or allow interactive guidance once the model is deployed. These gaps point to an opportunity for more *human-in-the-loop* AI designs in EEG preprocessing.

Computer Vision for Time-Series Data

A promising approach to leverage AI pattern recognition for EEG artifacts is to recast the time-series data as images and apply computer vision models. In domains like speech and finance, researchers have shown that transforming time-series into 2D visual representations and analyzing them with CNNs or vision transformers can yield excellent results [21]. Common strategies for "imaging" a time-series include plotting a conventional time-frequency spectrogram, creating recurrence plots (which depict repeated state trajectories of a dynamical system), or using techniques like Gramian Angular Fields (which encode time-series values into polar coordinate matrices). These transformations make implicit temporal patterns explicit as textures, shapes, or trajectories in an image, to which powerful vision models can be applied. **Time–frequency spectrograms** are especially popular – for example, *Zeng et al.* (2024) demonstrate improved time-series forecasting by converting data into spectrogram images and feeding them to a Vision Transformer, outperforming both statistical models and a deep learning baseline that used raw time-series input [22] [23]. Their vision transformer learned complementary representations in the time and frequency domains, highlighting the value of a visual approach for capturing complex temporal patterns. Other studies have explored simpler encodings like "line-plot images" (rendering the raw signal curve as an image) versus more sophisticated encodings. Generally, **richer representations** (incorporating frequency or state-space information) tend to improve learning, as the model can discern features (like periodicities or transients) that are not obvious in raw waveform alone [21].

Research comparing different time-series-to-image mappings found that methods like recurrence plots (RP) often outperform alternative encodings when coupled with deep CNN classifiers. For instance, one study transformed 1D student learning behavior time-series into four image types (pixel intensity plot, sine-wave transformation, RP, and Gramian field) and fed them into CNN-based classifiers. The **recurrence plot encoding yielded the highest accuracy (~95%)**, surpassing other image forms and far outperforming conventional machine learning on the original time-series [24] [25]. This suggests that capturing the inherent structure of a time-series in a 2D pattern can significantly boost classification performance, effectively leveraging CNNs' ability to recognize shapes and textures. Similarly, in a seismic signal classification task, converting earthquake time-series into images enabled CNNs to achieve accuracy comparable to traditional expert-designed features. A comparison of encoding techniques in that domain noted that there is **no consensus yet on the optimal image representation or resolution** – for example, wavelet-based spectrogram images reached about 79.5% accuracy, slightly higher than a novel time-segmentation image method at 76.8%, but at a cost of much larger image size [26]. Notably, the CNN approach performed on par with the manual feature-based method, indicating that with the right encoding, automated visual feature learning can match expert knowledge [26]. However, the study also underscores practical challenges: preparing a huge number of images for training is computationally intensive, and one must carefully choose image parameters (width, height, color channels) to balance information content and efficiency [27]. **In summary, applying computer vision to time-series data opens powerful avenues** – it allows complex temporal dynamics to be learned as visual patterns, and advanced architectures like CNNs and vision transformers can exploit this to improve accuracy in classification and forecasting tasks. The strengths reported include higher classification accuracy and feature learning flexibility [25], while challenges involve deciding on the best transformation, handling the added computational load of image data, and ensuring that important temporal context isn't lost in the conversion. For EEG artifact detection, this approach is enticing: artifacts often have telltale shapes in time or frequency (e.g. a blink looks like a large slow wave in frontal channels; muscle noise looks like a high-frequency burst). Encoding EEG epochs as images (such as multi-channel multi-time snapshots or frequency plots) could enable a vision model to spot these patterns more reliably than a purely temporal algorithm.

AI as Human Augmentation

Across many fields, a consensus is emerging that AI works best **as an augmentation to human experts** rather than an outright replacement – especially for visual analysis tasks where context and judgment are key. In medical imaging, for example, dozens of AI tools have been developed to assist radiologists in detecting abnormalities in X-rays, MRIs, and CT scans. Rather than diagnosing patients autonomously, these tools often serve as a *second reader* or triage system: the AI can flag suspicious lesions or prioritize scans for review, and the radiologist makes the final decisions [28]. Such human-AI collaboration has been shown to improve workflow efficiency. A recent meta-analysis found that when AI assistance is integrated, radiologists' image reading time decreased by ~27% on average, without loss of diagnostic accuracy [29]. In routine screening scenarios (like mammography), having an AI perform an initial pass to mark likely-normal cases and obvious anomalies can drastically reduce the human workload (in some studies, AI pre-screening cut the number of images requiring full manual review by over 60% [29]). Crucially, the **combination of human + AI outperforms either alone** in many settings. Purely automatic AI systems raise ethical and legal concerns and may falter on unusual cases, while human experts alone may be slower or prone to oversight on tedious high-volume tasks [30] [28]. The optimal solution is a synergy: use AI's speed and consistency to support the human's expertise and contextual understanding [28].

Designing collaborative and trustworthy human-AI systems requires careful attention to user experience and transparency. Experts need to feel that the AI is a helpful colleague, not a mysterious "black box" undermining their autonomy. Studies have found that clinicians are more accepting of AI tools that *minimally disrupt their workflow and respect their decision-making authority* [31] [32]. For instance, radiologists reported high satisfaction with an AI system for chest X-ray reading when it provided actionable suggestions without demanding extra clicks or overriding their judgments [31]. This was true even though the AI's internal logic was not fully explainable – indicating that **usability and alignment with human workflows can matter more for trust than full explainability** [31] [32]. Another principle is to allow the human some control over the AI's operation. In EEG artifact annotation, one approach used an active learning loop where a CNN initially learned from experts' labels, then suggested segments for the experts to relabel, gradually improving the model [33] [34]. The system let users adjust the confidence threshold for the AI's decisions, effectively tuning how aggressive the automated labeling was [35]. By raising the threshold for certainty, the user could have the AI only auto-annotate very obvious artifacts and defer ambiguous cases to human judgment, ensuring high precision; or lower the threshold to have the AI handle more cases to save time, accepting some false positives that the human would later check [35]. This kind of user-guided AI operation builds trust and makes the collaboration flexible. In general, **human-AI teams benefit from clear role definitions** – for example, the AI flags potential problems and the human confirms or corrects them. In quality control for manufacturing, this translates to AI-based vision systems spotting microscopic defects or inconsistencies, then human inspectors verifying and handling edge cases. AI significantly reduces human fatigue and error by doing the initial screening, addressing the fact that manual inspectors can miss defects due to tiredness or inconsistent attention [36]. Meanwhile, human oversight ensures that unusual or high-stakes decisions get the benefit of human reasoning. One caveat from recent human factors research is that experts must be trained to *appropriately trust* the AI. If they **underweight the AI's alerts** or ignore recommendations due to bias or poor calibration, the partnership falters [37]. On the other hand, over-reliance without double-checking can let errors slip through. Thus, the most effective systems are those that provide **accurate, interpretable alerts and foster calibrated trust**, so that humans and AI correct each other's mistakes and combine their strengths. In summary, the lesson for an EEG artifact context is that an AI system should assist the neurophysiologist by consistently catching likely artifacts and suggesting actions, but the human should guide the process and make final decisions, aided by an interface that allows easy review and override of the AI's suggestions.

Synthesis and Opportunity for a Visual EEG Artifact Detection System

The literature highlights a clear gap: EEG artifact handling remains labor-intensive and imperfect, yet emerging techniques in AI and computer vision offer new possibilities to close this gap. Traditional EEG cleaning methods are limited by their **narrow focus, reliance on manual labor, and inability to adapt to all artifact types** [6] [9]. Pure automation with machine learning, while promising in research, still faces practical hurdles such as scarce labeled data, black-box behavior, and limited integration into neuroscientists' workflows. We also see that human experts, by visually examining EEG plots, can often intuitively detect artifacts – leveraging patterns like a muscle spike's high-frequency fuzz or an eye blink's large frontal deflection. However, doing this by eye for thousands of epochs is not feasible, and consistency between reviewers is a concern. This is precisely where an **AI-orchestrated, computer vision-based artifact detection system** can make a difference. The idea is to harness the pattern-recognition power of modern vision models (CNNs, transformers) by presenting EEG data in a visual format they can understand, *while keeping a human in the loop to guide and validate the process*.

Such a system would take raw EEG recordings and convert them into informative images – for example, per-epoch multi-channel plots, time-frequency maps, or topographic voltage maps – creating a canvas on which artifacts become visually salient. An ensemble of AI agents (or an orchestrated model) would then analyze these images. One can imagine specialized detectors for different artifact morphologies: an agent tuned to spot eye blink shapes, another to catch muscle noise texture in the spectrogram, etc., all coordinated by a higher-level AI that integrates their outputs. Crucially, **prompt-guided AI** can be employed to allow dynamic interaction and guidance. For instance, an EEG technician could input a prompt or parameter (through a UI) indicating "Focus on muscle artifacts in the beta band" or "Ignore ECG artifacts for this dataset," and the AI orchestrator would adjust the detection strategy accordingly. This prompt-guidance might be implemented via a large language model (LLM) that interprets the expert's instructions and configures the vision models (a novel approach that merges human instruction with automated analysis). The system would flag segments or independent components that likely contain artifacts, accompanied by visual highlights (e.g. marking the portion of the EEG image that appears artifactual) to make review easy. The human expert then reviews these flags in a streamlined interface, confirming true artifacts and rejecting any false alarms. Over time, the AI can learn from the expert's confirmations/corrections (active learning), continually improving its accuracy and aligning with the specific criteria of that lab or clinician – much like the iterative model refinement demonstrated by active learning in artifact annotation [33] [34].

The novelty of this approach lies in **integrating three domains that have traditionally been separate**: (1) advanced EEG signal processing (to generate rich visual representations), (2) state-of-the-art computer vision AI (to detect complex patterns in those representations), and (3) human-in-the-loop design (to ensure the system acts as a cognitive aid rather than an autonomous black box). By doing so, the system addresses the current limitations identified in the literature. It can handle multifactorial artifacts because the vision models can be trained on a wide variety of artifact examples (eye, muscle, electrode pop, etc.) and recognize each by its signature in the image domain – something a simple threshold or single-method approach cannot do [6]. It drastically reduces dependence on tedious manual cleaning: the AI performs the heavy initial screening, and the expert's role shifts to one of oversight and targeted intervention, which is far more efficient for large datasets [2] [36]. Yet, because the expert remains in control – setting prompts, reviewing flags – the process retains transparency and flexibility. The human reviewer can inject contextual knowledge (e.g. "the patient has a pacemaker, so what looks like an artifact heartbeat is expected") via prompts or by simply disregarding certain AI flags, and the system can adapt accordingly. This **augmented intelligence** approach ensures that the final cleaned EEG is both consistent (thanks to the AI's systematic pattern recognition) and reliable (thanks to human validation for edge cases).

Moreover, employing a **visual AI** on EEG data could reveal subtle artifact patterns that might elude traditional methods. For example, a vision transformer might pick up a faint, repetitive oscillation across channels that indicates electrode noise – a pattern that might not trigger any single-channel threshold but is obvious when viewing the multi-channel EEG as an image. Through such capabilities, the proposed system could achieve a level of artifact detection thoroughness and consistency not currently attainable. Importantly, it does so *collaboratively*: the AI components are orchestrated to assist, not replace, the EEG analyst. This aligns with best practices from medical imaging and quality control, where AI's value is maximized by maintaining human oversight [28] [32]. By adjusting the AI's autonomy (e.g. via confidence thresholds as noted above) the system can be tuned to virtually eliminate missed artifacts (high sensitivity mode, where the human weeds out false positives) or to operate with minimal human input (high specificity mode, where only a small percent of doubtful cases are sent for review) [35]. Such adaptability would make it useful in both research (where one might favor thorough cleaning and examine each rejection) and clinical settings (where time is critical).

In conclusion, the convergence of neuroscience and computer vision literature suggests that an AI-orchestrated visual EEG artifact detection system is a timely innovation. It directly tackles the shortcomings of existing methods—manual burden, single-method rigidity, and scalability issues—by fusing them with AI's pattern recognition prowess and a human-centered design. This approach is novel in treating EEG not just as a signal to be algorithmically filtered, but as a *visual scene* to be interpreted by an AI "eye," guided by expert knowledge. The anticipated advantages include **greater artifact removal accuracy across diverse types, improved consistency and objectivity in cleaning, and major gains in efficiency** for large EEG datasets. By using prompt-guided AI agents that work in concert with human experts, the system embodies a form of "intelligent augmentation" that could set a new standard for how EEG data is preprocessed – making it faster, more reliable, and ultimately more trustable for downstream neuroimaging and BCI applications.

Sources: The arguments and findings above draw upon current literature on EEG artifact methods [6] [8], machine learning in EEG [19] [20], time-series computer vision techniques [21] [25], and human-AI collaboration principles from medicine and engineering domains [28] [32]. The synthesis integrates these insights to articulate the need and rationale for the proposed system.

---

[1] [2] [4] [5] [11] [33] [34] [35] Improved Manual Annotation of EEG Signals through Convolutional Neural Network Guidance - PMC

https://pmc.ncbi.nlm.nih.gov/articles/PMC9532020/

[3] [6] [7] [8] [9] [10] Removal of Artifacts from EEG Signals: A Review - PMC

https://pmc.ncbi.nlm.nih.gov/articles/PMC6427454/

[12] [16] [18] EEG Artifact Detection and Correction with Deep Autoencoders

https://arxiv.org/html/2502.08686v1

[13] [14] Frontiers | Status of deep learning for EEG-based brain–computer interface applications

https://www.frontiersin.org/journals/computational-neuroscience/articles/10.3389/fncom.2022.1006763/full

[15] [17] [19] [20] arxiv.org
https://arxiv.org/pdf/2106.13081

[21] [22] [23] [2403.11047] From Pixels to Predictions: Spectrogram and Vision Transformer for Better Time Series Forecasting
https://arxiv.org/abs/2403.11047

[24] [25] Imaging time series using the Gramian Angular Field. (Left) Time series... | Download Scientific Diagram
https://www.researchgate.net/figure/maging-time-series-using-the-Gramian-Angular-Field-Left-Time-series-Middle-Polar_fig1_375878644

[26] [27] Frontiers | Encoding Time-Series Ground Motions as Images for Convolutional Neural Networks-Based Seismic Damage Evaluation
https://www.frontiersin.org/journals/built-environment/articles/10.3389/fbuil.2021.660103/full

[28] [29] [30] Impact of human and artificial intelligence collaboration on workload reduction in medical image interpretation | npj Digital Medicine
https://www.nature.com/articles/s41746-024-01328-w

[31] [32] Towards human-AI collaboration in radiology: a multidimensional evaluation of the acceptability of AI for chest radiograph analysis in supporting pulmonary tuberculosis diagnosis - PMC
https://pmc.ncbi.nlm.nih.gov/articles/PMC11796096/

[36] Quality Control in Manufacturing with AI Visual Inspection
https://www.loopr.ai/blogs/quality-control-in-manufacturing-with-ai-visual-inspection

[37] Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology
https://alexmoehring.com/assets/pdf/agarwal-et-al-diagnostic-ai.pdf