IOP Publishing Physiol. Meas. 45 (2024) 12TR01 https://doi.org/10.1088/1361-6579/ad9af4

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

OPEN ACCESS

RECEIVED
15 August 2024

REVISED
28 October 2024

ACCEPTED FOR PUBLICATION
5 December 2024

PUBLISHED
17 December 2024

TOPICAL REVIEW

Artifact identification and removal methodologies for intracranial pressure signals: a systematic scoping review

Tobias Bergmann[1,*], Nuray Vakitbilir[1], Alwyn Gomez[2,3], Abrar Islam[1], Kevin Y Stein[1,4], Amanjyot Singh Sainbhi[1], Noah Silvaggio[3], Izzy Marquez[5], Logan Froese[6] and Frederick A Zeiler[1,2,6,7,8,9,10,*]

[1] Department of Biomedical Engineering, Price Faculty of Engineering, University of Manitoba, Winnipeg, Canada
[2] Section of Neurosurgery, Department of Surgery, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada
[3] Department of Human Anatomy and Cell Science, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada
[4] Undergraduate Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada
[5] Undergraduate Engineering, Price Faculty of Engineering, University of Manitoba, Winnipeg, Canada
[6] Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
[7] Centre on Aging, University of Manitoba, Winnipeg, Canada
[8] Division of Anaesthesia, Department of Medicine, Addenbrooke's Hospital, University of Cambridge, Cambridge, United Kingdom
[9] Pan Am Clinic Foundation, Winnipeg, MB, Canada
[10] Current address: GF231 Health Sciences Centre, 820 Sherbrook Street, Winnipeg, MB R3A 1R9, Canada.
[*] Authors to whom any correspondence should be addressed.

E-mail: bergmant@myumanitoba.ca, Frederick.Zeiler@umanitoba.ca, vakitbir@myumanitoba.ca, gomeza35@myumanitoba.ca, islama9@myumanitoba.ca, steink34@myumanitoba.ca, amanjyot.s.sainbhi@gmail.com, silvaggn@myumanitoba.ca, marquezi@myumanitoba.ca and log.froese@gmail.com

Keywords: artifact management, intracranial pressure, cerebral hemodynamic monitoring, bio-signal analysis

Abstract

Objective. Intracranial pressure measurement (ICP) is an essential component of deriving of multivariate data metrics foundational to improving understanding of high temporal relationships in cerebral physiology. A significant barrier to this work is artifact ridden data. As such, the objective of this review was to examine the existing literature pertinent to ICP artifact management. *Methods.* A search of five databases (BIOSIS, SCOPUS, EMBASE, PubMed, and Cochrane Library) was conducted based on the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines with the PRISMA Extension for Scoping Review. The search question examined the methods for artifact management for ICP signals measured in human/animals. *Results.* The search yielded 5875 unique results. There were 19 articles included in this review based on inclusion/exclusion criteria and article references. Each method presented was categorized as: (1) valid ICP pulse detection algorithms and (2) ICP artifact identification and removal methods. Machine learning-based and filter-based methods indicated the best results for artifact management; however, it was not possible to elucidate a single most robust method. *Conclusion.* There is a significant lack of standardization in the metrics of effectiveness in artifact removal which makes comparison difficult across studies. Differences in artifacts observed on patient neuropathological health and recording methodologies have not been thoroughly examined and introduce additional uncertainty regarding effectiveness. *Significance*. This work provides critical insights into existing literature pertaining to ICP artifact management as it highlights holes in the literature that need to be adequately addressed in the establishment of robust artifact management methodologies.

1. Introduction

The Monro–Kellie doctrine indicates that the absolute volume of the cerebral components contained within the intracranial space should remain constant (Mokri 2001). Under healthy circumstances, these components include: cerebral tissue, cerebral blood, and cerebrospinal fluid (CSF) (Mokri 2001). In some pathological cases, a mass-lesion, such as a tumor or hematoma, contributes as a fourth component of intracranial

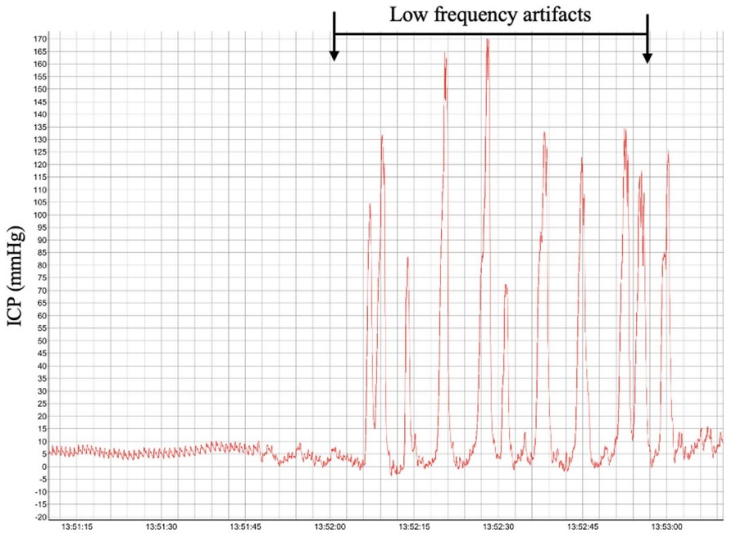

Figure 1. Low frequency artifacts to be replaced with interpolated values after removal. Where ICP = intracranial pressure, mmHg = millimeters of mercury (pressure). Archived data that was used for this figure is from the prospectively maintained signals database at the Multi-omic Analytics and Integrative Neuroinformatics in the HUman Brain (MAIN-HUB) Lab at the University of Manitoba, with ethical approval in place for its access and use (H2017:181, H2020:118, H2024:266).

volume (Munakomi and Das 2024). The components of the intracranial space exert a summative pressure on the skull is known as intracranial pressure (ICP) (Czosnyka 2004, Kawoos *et al* 2015). ICP is a complex waveform as it consists of slow waves from autoregulatory processes and oscillations corresponding to subject respiration (Shiogai *et al* 2010) as well as a pulse waveform consisting of three peaks originating from the cardiac cycle and changes in CSF volume (Zakrzewska *et al* 2021). ICP measurement alongside mean arterial blood pressure (ABP) are used to determine cerebral perfusion pressure (CPP) (ABP-ICP = CPP) (Czosnyka 2004, Kawoos *et al* 2015). CPP is a critical metric for evaluating cerebral autoregulatory (CA) health, which is the ability of the brain vasculature to maintain a pressure gradient across the blood brain barrier conducive to a constant cerebral blood flow, despite changes in ABP or ICP (Miller *et al* 1972, Czosnyka 2004). ICP and CPP monitoring have been widely conducted during the management of traumatic brain injury (TBI) as both these signals are useful in guiding treatment (Carney *et al* 2017, Hawryluk *et al* 2019). Following TBI, maintaining specific thresholds in ICP and CPP have been associated with a decrease in the likelihood of negative outcomes occurring (Carney *et al* 2017). There has been a recent shift towards patient-specific therapeutics in instances of compromised CA to further promote positive outcomes in TBI management (Zeiler *et al* 2021). This has been achieved by targeting specific secondary injury pathways as opposed to relying on global ICP and CPP levels (Sorrentino *et al* 2012, Nourallah *et al* 2018, Zeiler *et al* 2019).

However, there still remains gaps in understanding regarding the high temporal relationships between aspects of cerebral physiology including pressure-flow dynamics, nutrient delivery, cerebrovascular compliance, and autonomic function. As such, there is ongoing investigation into the use of multivariate relationships within cerebral physiology to derive new multi-modal time-series metrics. This aims to improve understanding of subject-specific physiological activity to improve treatment. A major barrier to this development are the abundant artifacts that occur during the clinical measurement of ICP signals (Kim *et al* 2018, Lee *et al* 2020). Erroneous segments of data and increase the incidence of false alarms triggered by erroneous spikes in ICP magnitude (Scalzo *et al* 2013). Additionally, recent work has shown that these segments diminish accuracy and real-time utility of continuously derived metrics such as pressure-reactivity index (Rozanek *et al* 2022); however, future work is needed to elucidate the impacts of artifacts on other ICP derived indices.

Artifacts in ICP signals can occur due to a variety of sources across measurement devices including erroneous QRS (when the waveform does not follow the expected QRS shape) (Asgari *et al* 2009), as well as low frequency artifacts due to coughing, subject motion, device disconnection, and sensor detachment (Dai *et al* 2020). However, there are also device specific errors like those resulting from external ventricular drain (EVD) becoming disconnected (Megjhani *et al* 2023). An exemplar low frequency artifact is included in figure 1. These datasets were recorded using an intraparenchymal strain gauge sampled at 100 Hz.

There are instances in which artifacts are more egregious which makes it difficult to salvage or interpolate any underlying components of the signal due to the duration and severity of the artifact. An example of a

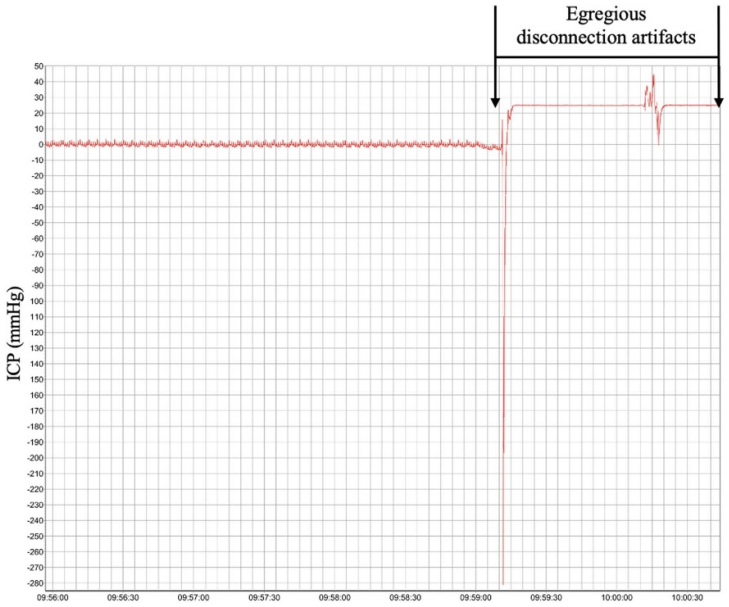

Figure 2. Egregious ICP signal artifacts. Where ICP = intracranial pressure, mmHg = millimeters of mercury (pressure). Archived data that was used for this figure is from the prospectively maintained signals database at the Multi-omic Analytics and Integrative Neuroinformatics in the HUman Brain (MAIN-HUB) Lab at the University of Manitoba, with ethical approval in place for its access and use (H2017:181, H2020:118, H2024:266).

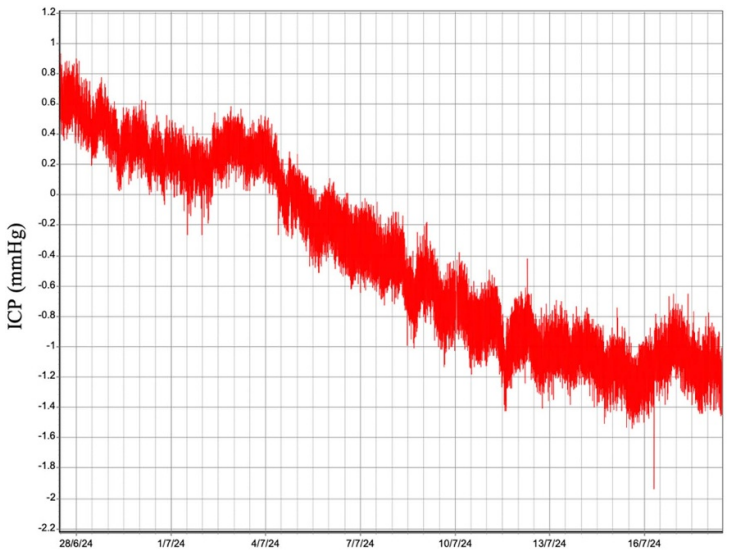

Figure 3. Zero-drift artifact in ICP signal over several days of recording. Where ICP = intracranial pressure, mmHg = millimeters of mercury (pressure). Data was recorded by placing a strain gauge probe connected to an ICP Express box with a sampling rate of 100 Hz in a beaker of water.

more egregious artifacts is depicted in figure 2, this data was sampled at 100 Hz using an intraparenchymal strain gauge. These errors may have resulted from device disconnections or patient motion.

ICP signals are also often contaminated with high frequency noise which can be filtered out using low-pass filtering (LPF). Additionally, there are also instances of drift in the signal from that occur due to failures in the instrumentation. Depicted in figure 3 is an example of a zero-drift artifact. This gradual drift in ICP magnitude is due to device error. The artifact depicted in figure 4 was measured by placing a strain gauge probe attached to an ICP Express box in a water sample to record instrument zero-drift over several days of recording, which is applicable in long-term recordings of ICP. The sampling rate of the instrument was 100 Hz.

The current practice of manual removal of artifacts is required prior to the use in any diagnostic or prognostic capacity, which requires trained personnel, is resource consuming, and has the potential for human error (Raj *et al* 2022, Rozanek *et al* 2022). Removing artifacts that occur in ICP signals cannot be

simply accomplished using basic thresholding due to the complex morphology and non-stationarity of ICP data. The development of a broadly accepted 'gold standard' method for a fully or semi-automated artifact removal is a topic of present research (Raj *et al* 2022, Rozanek *et al* 2022). This systematic review will help illuminate which fully and semi-autonomous methods have been developed for artifact management in ICP signals as well as their respective efficacies. The findings of this review will provide foundational knowledge for scientists and clinicians to build towards a broadly accepted method in the future. This article serves as the first systematic review that has been conducted to examine the existing artifact identification and removal methods specifically designed for ICP signals. It expands upon the artifact removal methods briefly mentioned in a narrative review conducted by Dai *et al* to discuss the effectiveness of all methods that have been developed for this application (2020).

The objective of this systematically conducted scoping review was to describe the functionality and effectiveness of the artifact management methods that have been developed for ICP signals. Several different categories of artifact management methods have been developed to date; however, in this review, they were placed in two major categories: (1) valid ICP pulse detection algorithms and (2) ICP signal artifact identification or removal methods.

2. Methodology

This systematic scoping review was conducted following the Cochrane Handbook for Systematic Reviews (Page *et al* 2021). The reporting of the results conforms to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis guidelines with the (PRISMA) Extension for Scoping Review (Tricco *et al* 2018, Page *et al* 2021). The review objectives for the search strategy were developed collaboratively by TB and FAZ. Help in the article filtering process was provided by by NV (PhD Student) and LF. The completed PRISMA checklist can be found in appendix A.

2.1. Ethical considerations

All articles included in this review were from previously published journals. As a result, they are expected to have been vetted by their respective journals. Specific ethical approval for this systematically conducted scoping review was therefore not required. Archived data that was used for figures 2 and 3 is from the prospectively maintained signals database at the Multi-omic Analytics and Integrative Neuroinformatics in the HUman Brain (MAIN-HUB) Lab at the University of Manitoba, with ethical approval in place for its access and use (H2017:181, H2020:118, H2024:266).

2.2. Search question and inclusion/exclusion criteria

The main search question examined within this review was, 'What artifact identification and removal techniques exist for full waveform ICP signals?' Secondary questions that were examined looked at the efficacy of the proposed techniques, the types of algorithms that have been used to identify and remove the artifacts, and the ICP recording hardware for which the methods were developed. To be included in this review, the article needed to detail some form of artifact identification or removal method for full waveform ICP signals for human or animal studies. The minimum sampling rate was selected as 50 Hz, as existing literature indicates that this threshold is required to properly encapsulate the full waveform (Holm and Eide 2009). Articles that were excluded were non-English, not full length, used simulated or non-ICP data, or did not include evidence to support the effectiveness of the proposed method(s).

2.3. Search strategy

Searches were conducted across the following five databases: BIOSIS, SCOPUS, EMBASE, PubMed, and Cochrane Library, covering the entire period from the conception of each database up to 19 June, 2024. Dedicated search strings were constructed for each database. These search strings consisted of terms/synonyms for ICP and artifact identification or removal methods. The detailed search string for each database is provided in appendix B. The results of the search from each database were compiled and deduplication was conducted.

2.4. Study selections

A manual review of all unique articles from the initial search was conducted using a two-stage, two-reviewer approach. The first stage involved two reviewers (TB and NV) independently screening the title and abstract of each article for inclusion/exclusion criteria. The second stage involved both reviewers screening the included studies from the first stage to evaluate the entire content of each article for inclusion and exclusion criteria. Any disagreements between the two reviewers were resolved by a third-party (FAZ).

2.5. Data collection

Data collection was conducted by the primary author (TB). Key characteristics for each study were recorded. These characteristics included patient/subject information, data information, and details of the artifact identification/removal methods. Patient/subject information included the sample size, sex, age, and presence of neurological injury. Data information included the sampling rate of the ICP signals, anatomical location, and type of ICP sensor used for measurement. Additionally, any bio-signals being simultaneously recorded were noted. Information regarding the details of ICP signal artifact identification or removal methods described in each article were also extracted. Analysis of each method involved the evaluating the effectiveness of the proposed method (or effectiveness relative to other presented methods) as well as the main results presented in the article. Limitations for each method were also extracted, including those identified by the authors and those identified during the critical analysis of each work.

3. Results

The results of the search strategy that was conducted across five databases (BIOSIS, SCPOUS, EMBASE, PubMed, and Cochrane Library), as well as the reference sections of each article identified as meeting inclusion criteria, are depicted in a PRISMA flow diagram in figure 4. A total of 7652 articles were identified from the initial search. Following deduplication, 5875 articles were screened by title and abstract against inclusion/exclusion criteria. This led to 39 articles that were screened by examining full-length documents. There were 15 articles of the initial search that were included in this review. In addition, there were 4 articles also included in the review that were identified through examination of the reference sections of the included articles from the initial search.

All 19 studies that were included were human-based studies. No animal-based studies met the inclusion criteria. Those studies that indicated patient age demographics included pediatric intensive care unit (ICU) patients (age range not disclosed) (Aboy *et al* 2001, 2005), juvenile (patients between 14 and 18 year of age) (Asgari *et al* 2009, Hu *et al* 2009), and adult (age 18 and older) (Asgari *et al* 2009, Hu *et al* 2009, Scalzo *et al* 2009, Mataczynski *et al* 2022, Megjhani *et al* 2023, Huo *et al* 2024); however, not all articles included patient age demographics (Scalzo *et al* 2010, Feng *et al* 2011, 2012, Scalzo *et al* 2012, 2013, Yang *et al* 2012, Rajagopal *et al* 2016, Megjhani *et al* 2019, Lee *et al* 2020, Martinez-Tejada *et al* 2021, Taco *et al* 2022). The individuals included in these studies were receiving treatment for a variety of pathophysiological issues including TBI (Aboy *et al* 2001, Feng *et al* 2011, 2012, Yang *et al* 2012, Taco *et al* 2022), aneurysmal sub-arachnoid hemorrhage (SAH) (Megjhani *et al* 2023), hydrocephalus (Asgari *et al* 2009, Yang *et al* 2012), intracerebral hemorrhage (ICH) (Scalzo *et al* 2013), intracranial hypertension (IH) (Rajagopal *et al* 2016), and a variety ICP-related issues (Aboy *et al* 2005, Asgari *et al* 2009, Hu *et al* 2009, Scalzo *et al* 2009, 2010, 2012, Rajagopal *et al* 2016, Megjhani *et al* 2019). There were several types of ICP measurement devices that were used to record the ICP signal. These included intraparenchymal strain gauges (Asgari *et al* 2009, Hu *et al* 2009, Scalzo *et al* 2009, 2010, 2012, Lee *et al* 2020, Mataczynski *et al* 2022), intraparenchymal fiber-optic pressure transducers (Aboy *et al* 2001, Yang *et al* 2012), and EVD devices (ventricular catheters) (Aboy *et al* 2001, Megjhani *et al* 2019, 2023, Huo *et al* 2024). There were two studies that indicated ICP measurement and artifact management with ICP was estimated using an intraspinal strain gauge (Calisto *et al* 2010, 2013), which did not satisfy the inclusion criteria for this review.

The sampling rates that were used to collect the ICP data ranged from 50 to 400 Hz; however, some did not list a sampling rate (Feng *et al* 2011, 2012, Lee *et al* 2020, Martinez-Tejada *et al* 2021). Of those that did not include sampling rates, each article included some type of image that indicated that full waveform ICP was included (Feng *et al* 2011, 2012, Lee *et al* 2020, Martinez-Tejada *et al* 2021). There was seldom a formal definition of an 'artifact' defined in any of the articles, the majority compared the results of an algorithm to that of a qualified professional manually cleaning the data. Not all methods that were included made mention to the direct detection and/or removal of artifacts from ICP signals; however, those methods designed to detect valid ICP signals could be leveraged for artifact management, therefore, they were included.

As the scope of this project was to provide a detailed overview of existing methodologies to attain a clean ICP signal, these articles were included. A detailed analysis of the results of this systematic scoping review was split into two main sections: (1) valid ICP pulse detection algorithms and (2) ICP signal artifact identification or removal methods. Each of these sections contained sub-sections detailing specific method types that had been developed for ICP signals.

3.1. Valid ICP pulse detection algorithms

There has been work done to develop algorithms capable of locating specific peaks and relevant morphological features present in valid ICP signals to move away from a reliance on mean ICP. Beat detection algorithms are intended to locate specific peaks that are typical to valid ICP signals (Aboy *et al*

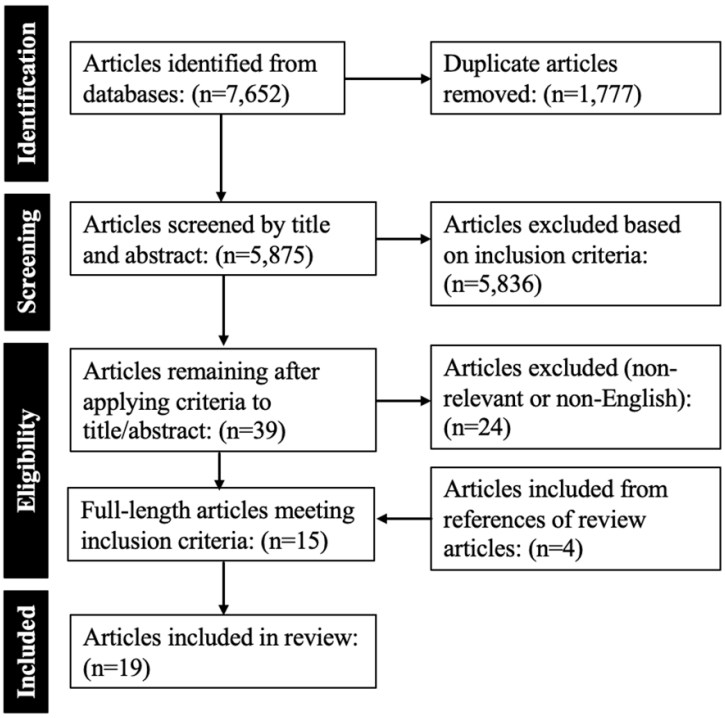

Figure 4. PRISMA flowchart of results for systematically conducted scoping review.

2001, 2005). The morphological clustering and analysis of ICP pulse (MOCAIP) algorithm/template-matching methods use segmented ICP pulses in comparison to reference library pulses to identify valid signals (Asgari *et al* 2009, Hu *et al* 2009, Scalzo *et al* 2009, 2010, 2012, 2013, Yang *et al* 2012, Megjhani *et al* 2019). Of the methods that were presented, there were five that required auxiliary signals, all of which were electrocardiogram (ECG) signals, to determine the onset of the ICP pulse (Asgari *et al* 2009, Hu *et al* 2009, Scalzo *et al* 2009, 2010, 2012, Megjhani *et al* 2019). Each of these methods were developed using data from different numbers of patients (median = 66.5, range = 3–210) with varying cerebral health. The goal of these algorithms was to accurately differentiate artifactual from non-artifactual signals; however, only a single article that described the morphology of artifactual signals. This article by Asgari *et al* presented a method that was designed with consideration to low frequency artifacts (from erroneous QRS detection, head movements like coughing, patient motion, and sensor detachment) and contamination with other high frequency noise (2009). Four methods were validated using cross-validation (Asgari *et al* 2009, Scalzo *et al* 2009, 2010, 2013), one method split data between template development, training, and testing sets (Yang *et al* 2012), and one method used external validation which involved data from five patients external to the 34 that had been used to develop the model (Megjhani *et al* 2019). Other methods were not trained using data, they were simply validated using a collected dataset, this was referred to as 'direct' validation (Aboy *et al* 2001, 2005, Hu *et al* 2009). Live processing was presented as 'possible' for several of the algorithms; however, it was only actually conducted by Scalzo *et al* (2013). It was indicated that real-time processing would not be possible by Hu *et al* due to its reliance on template matching and computational time (CT) (2009). This information is summarized in table 1.

A more detailed description of the methods and their respective efficacies are included in appendix C (tables C1 and C2). A summary of these appendix tables is displayed in table 2.

3.1.1. Beat detection algorithms

Beat detection algorithms for pulse identification have been developed to move away from the use of gross metrics like mean, standard deviation, and spectral power for ICP signal analysis (Aboy *et al* 2005). Aboy *et al* developed two pressure beat detection algorithms to identify the peaks present in valid ICP signals and focus analysis on the beat-to-beat level (2001, 2005). There are four components to the ICP waveform, these include the percussive peak, dicrotic peak, dicrotic notch, and tidal peak. The algorithm developed by Aboy *et al* utilized LPF, coarse beat detection by incremental sampling, relative quantization error filtering, and local peak detection to determine respective peak locations (2001). The effectiveness of this method was compared to two experts identifying the peaks on an ICP dataset (125 Hz) collected from three severe TBI

Table 1. Algorithm information for valid beat detection and MOCAIP algorithms.

Article	Patients	Specific pathological state	Artifact type	Validation	Live processing
Aboy *et al* (2001)	3	TBI	Not specified	Direct	Possible
Aboy *et al* (2005)	210	TBI, sepsis, cardiac conditions	Not specified	Direct	Possible
Asgari *et al* (2009)	67	Hydrocephalus + other	High and low frequency	Cross-validation (67-fold)	Possible
Hu *et al* (2009)	66	Various ICP conditions	Not specified	Direct	No
Scalzo *et al* (2009)	64	Various ICP conditions	Not specified	Cross-validation (64-fold)	Possible
Megjhani *et al* (2019)	34	Not indicated	Not specified	External validation	Possible
Scalzo *et al* (2012)	128	Various ICP conditions	Not specified	Validated (not specified)	Possible
Scalzo *et al* (2013)	108	TBI, SAH, ICH	Not specified	Cross-validation (10-fold)	Yes
Scalzo *et al* (2010)	128	Various ICP conditions	Not specified	Cross-validation (5-fold)	Possible
Yang *et al* (2012)	48	TBI + hydrocephalus	Not specified	Template + train-test split	Possible, too slow at present

ICP = intracranial pressure, SAH = aneurysmal sub-arachnoid hemorrhage, TBI = traumatic brain injury.

patients. The percent accuracy in locating the peaks indicated that there was a minimum agreement of 99.30% (within an acceptance interval of 8 ms) between the algorithm and the reviewers (Aboy *et al* 2001). This was very similar to the agreement between the reviewers (Aboy *et al* 2001). The second presented algorithm was designed to identify valid ICP pulses (Aboy *et al* 2005). The initial pre-processing used three bandpass elliptic filters with varying cut-off frequencies and additional cut-off frequencies that had been determined using the estimated heart rate. A nearest neighbor feature methodology was then used to identify peaks using relative amplitude and slope information. This information was used to classify the signal components as normal or artifactual. This algorithm attained a specificity of 99.36% and a positive predictivity of 99.43% compared to manual reviewers for a dataset of ICP (125 Hz) pulses randomly selected from a population of 210 patients (Aboy *et al* 2005).

3.1.2. MOCAIP/template-based algorithms

The MOCAIP algorithm is the most rigorously studied algorithm that has been developed for ICP pulse detection (Asgari *et al* 2009, Hu *et al* 2009, Scalzo *et al* 2009, 2010, 2013, Megjhani *et al* 2019). This methodology was largely applied to datasets that had been recorded using intraparenchymal strain gauges (Asgari *et al* 2009, Hu *et al* 2009, Scalzo *et al* 2009, 2010, 2012) and parenchymal fiber optic pressure transducers (Aboy *et al* 2001, Yang *et al* 2012) mainly placed in the right frontal lobe. The MOCAIP algorithm was initially proposed by Hu *et al* (2009). This method uses ECG data simultaneously recorded with the ICP signal to detect the onset of ICP pulses, it then conducts averaging of pulses with similar morphology to remove noise from ICP, determining the correlation of pulses to those in a reference library to identify valid pulses, and finally identify peaks within the ICP pulses (Hu *et al* 2009). The components of the algorithm are:

1. Beat-by-beat detection of the ICP pulse which is estimated using the locations of the QRS wave in the ECG signal. The QRS wave is a particular waveform measured in ECG signals that corresponds to ventricular depolarization (Levick 1991).
2. Clustering of ICP pulses by morphological distance and averaging the pulses corresponding to the largest cluster (dominant pulse) which removes random perturbations of the signal.
3. Non-artifactual pulses are identified using thresholds in regression with a reference library of expert validated ICP pulses.
4. Detection of peaks within dominant pulse using the second derivative of the ICP pulse.
5. Assignment of the detected peaks within the ICP pulse using an assumed Gaussian distribution via probability density functions (PDFs) estimated from ICP pulses from the reference library.

Using the determined optimal parameters, the original MOCAIP algorithm was able to obtain an accuracy of 97.84% in identifying non-artifactual ICP pulses. This method was heavily reliant on having a

Table 2. Summary of valid ICP pulse algorithms.

Artifact removal methods	Number of studies included	Number of subjects	Sensor(s)	Effectiveness
Valid beat detection algorithm	2	4	Ventricular catheter, parenchymal fiber-optic pressure transducer	• For the first method presented by Aboy *et al*, there was an agreement between expert and detection algorithm were 99.40% and 99.30% at an acceptance interval of 8.0 ms for identification of the location of the ICP peaks (2001). Agreement between experts was 99.70%. • For the second method presented by Aboy *et al*, the average sensitivity and positive predictivity of locating the peaks within a 16 ms interval algorithm on 42 539 beats was 99.30% and 99.60% (2005).
MOCAIP algorithms	7	641	Intra-parenchymal strain gauge, EVD	• The MOCAIP algorithm was initially presented by Hu *et al* had an accuracy of 90.17%, 87.56%, and 86.53% for each respective peak (2009). • There was work conducted to improve the MOCAIP algorithm. The most successful of which was an active learning framework to enhance the MOCAIP was able to achieve an AUC of 0.96 ± 0.02, CCR of 0.94 ± 0.02 (Megjhani *et al* 2019). • The extension of MOCAIP developed by Asgari *et al* used SVD instead of correlation for peak detection performed much better achieving a TPR of 96.61%, FPR of 6.61% and a much better computational time of 170 s (2009). • For more challenging datasets, the MOCAIP++ which uses KSR was able to improve the accuracy of prediction for the three ICP peaks from 65.83% to 88.78% (Scalzo *et al* 2010). • The results for the Bayesian tracking algorithm that was developed was able to achieve the average estimation error in each of the three peaks for latency were 8.09 ± 2.0 ms, 6.90 ± 1.7 ms, and 7.46 ± 2.1 ms and for magnitude were 0.04 ± 0.02 millimeters of mercury (mmHg), 0.09 ± 0.06 mmHg, and 0.08 ± 0.04 mmHg. Both of these results were better than the prediction using MOCAIP (Hu *et al* 2009) as the average estimation error for the three peaks were 11.88 ± 7.1 ms, 11.80 ± 7.0 ms, and 11.76 ± 7.0 ms and for magnitude were 0.23 ± 0.18 mmHg, 0.32 ± 0.27 mmHg, 0.33 ± 0.23 mmHg (Scalzo *et al* 2012). • The MOCAIP algorithm was able to be applied for ICP related alarm reduction which was able to improve the AUC-ROC from 55.9 using a threshold-based method to 69.2 ± 1.2–85.9 ± 1.1 (average deviation) using a SR-KDA (Scalzo *et al* 2013). • A method presented by Yang *et al* was similar to the MOCAIP algorithm as it uses hierarchal clustering to select representative segments based on a template library to determine valid ICP signals where it achieved a sensitivity and positive predictive value of 0.9723 and 0.9475, respectively 2012).

AUC = area under the curve, AUC-ROC = area under the receiver operating characteristic curve, CCR = correct classification rate, EVD = extra-ventricular drain, FPR = false positive rate, ICP = intracranial pressure, KSR = kernel spectral regression, MOCAIP = morphological clustering and analysis of ICP pulse, SR-KDA = spectral regression kernel discriminant analysis, SVD = singular value decomposition, TPR = true positive rate.

sound and detailed reference library (Hu *et al* 2009). Because this method requires that a correlation threshold be met between the dominant pulse and that of a pulse in the reference library many iterations had to be conducted (Asgari *et al* 2009). There have been several attempts to develop extensions of the MOCAIP algorithm, specifically targeting the noted weaknesses of the method. These weaknesses include the algorithms accuracy due to the complexity of the ICP signal (Asgari *et al* 2009, Scalzo *et al* 2009), computational requirements, and diversity of the library without redundancy (Asgari *et al* 2009).

The MOCAIP algorithm uses correlation between a pulse and a reference pulse to determine the validity of an ICP pulse. Asgari *et al* proposed an extension of MOCAIP using singular value decomposition (SVD) projections of noise and signal subspaces to identify non-artifactual pulses while it shrinks the required computational requirements and improves accuracy (2009). This method was compared to two variations of the MOCAIP algorithm, one that used a correlation coefficient (r) of 0.97 and another with $r = 1$ to identify the pulses. The true positives (TPs) between these three algorithms did not vary significantly but the false positive rate (FPR) was 47.48% when $r = 0.97$, 28.92% when $r = 1$, and 6.61% using SVD. The CTs were 1927 s when $r = 0.97$, 16 370 s when $r = 1$, and 170 s when SVD was used (Asgari *et al* 2009)

In response to the weaknesses of the MOCAIP algorithm, Scalzo *et al* developed an extension of the MOCAIP using a correlation-based model in place of PDFs for estimating peak locations (2009). This work was furthered by this author through the proposed MOCAIP++ algorithm (2010) to integrate more complex signal morphological analysis using spectral regression (SR), as it can accurately detect the location of the three peaks within the ICP signal without compromising computational efficiency. The most robust comparison of methods was presented by Scalzo *et al* (2010) as it compared the use of three peak regression identification techniques including the Gaussian-based MOCAIP model (Hu *et al* 2009), a Gaussian mixtures model, a SR model, and a kernel spectral regression based model (KSR). The presented peak identification method based off the MOCAIP algorithm was validated using data from 128 patients with varying ICP conditions with a high sampling rate (240–400 Hz) (Scalzo *et al* 2010). The dataset was specified as challenging by manual reviewers; and the best method, which was KSR, was only able to achieve an average accuracy of prediction of the three peaks of 88.78% (Scalzo *et al* 2010).

Scalzo *et al* proposed a method of tracking signal morphology of ICP signals and the location of peaks using a Bayesian inference algorithm (2012). Using a graphical inference model, a random variable is assigned to each peak based on their location within the pulse and its magnitude. Non-parametric belief propagation is used to determine the relationships between peaks using kernel density estimation. This was calibrated using inputted manually annotated ICP pulse data (400 Hz) from 128 patients with various ICP conditions. This algorithm was able to predict the latency of the ICP signal (using onset of ECG QRS wave) as well as ICP elevation (peaks) in an artificially contaminated ICP dataset. Analysis was conducted on randomly selected 30 min segments of ICP data with an induced sinusoidal temporal dynamic or phase multiplication dynamic (induced noise). The average estimation error in each of the three peaks for latency were 8.09 ± 2.0 ms, 6.90 ± 1.7 ms, and 7.46 ± 2.1 ms and for magnitude were 0.04 ± 0.02 millimeters of mercury (mmHg), 0.09 ± 0.06 mmHg, and 0.08 ± 0.04 mmHg. Both of these results were better than the predicted using MOCAIP (Hu *et al* 2009) as the average estimation error for the three peaks were 11.88 ± 7.1 ms, 11.80 ± 7.0 ms, and 11.76 ± 7.0 ms and for magnitude were 0.23 ± 0.18 mmHg, 0.32 ± 0.27 mmHg, 0.33 ± 0.23 mmHg (Scalzo *et al* 2012).

The applicability of an extension of the MOCAIP in EVD recordings was examined as extracted morphological waveform features were used to differentiate false alarms occurring during ICP signal recording (Scalzo *et al* 2013). ICP waveforms were segmented using the MOCAIP algorithm (Hu *et al* 2009) and the initial tracking of the latency and magnitude of the three peaks was presented by Scalzo *et al* (2012) and was extended by the same author to 24 morphological features including curvature, slope, and decay (Scalzo *et al* 2013). This article also investigated the use of a supervised directionality reduction algorithm presented by Tsai *et al* which decomposes the ICP signal into 24 subspaces (2008). This methodology was described as conditional discretization of morphological features (CDFs) (Scalzo *et al* 2013). It was used as an alternative to using the time-series MOCAIP vectors as inputs to the algorithm. Three regression models were compared for their abilities to detect alarm incidence based on input features from ICP signals (240 Hz) extracted from 108 neurosurgical patients (cases of TBI, SAH, and ICH). The regression models included multiple linear regression (MLR) (Chatterjee and Hadi 1986), SR-discriminant analysis (DA) (Cai *et al* 2007), kernel spectral regression discriminant analysis (SR-KDA) (Cai *et al* 2007), support vector machine (SVM) (Chang and Lin 2001), and extremely randomized trees (ETs) (Geurts *et al* 2006). The average area under the receiver operator characteristic curve (AUC-ROC), which is a performance metric for sensitivity and specificity, indicated that use of the SR-KDA method with CDF demonstrated the best results for accurate alarm detection based on morphology (Scalzo *et al* 2013). This methodology was solely tested on its ability to improve the accuracy of ICP alarms, it is likely that this methodology could be extended to artifact detection with slight alterations.

Megjhani *et al* developed an algorithm to detect valid ICP waveforms from artifactual in data that was recorded using EVDs (2019). It demonstrated encouraging results as it achieved an average AUC-ROC of 0.96 ± 0.02 and average correct classification rate (CCR) of 0.94 ± 0.02. These results were dramatically better than the template matching, ICP stability algorithm, and threshold-based single wave criteria-based methods to which it was compared (Megjhani *et al* 2019).

While it did not directly use the MOCAIP algorithm, Yang *et al* developed the waveform decomposition (WD) methodology of detecting ICP pulse onsets and valid ICP pulses using waveform templates and clustering techniques (2012). A template was constructed based on features extracted from representative ICP segments. Representative segments were automatically selected using hierarchal clustering, the pulse onset of the test signal was selected based on local minima, and a comparison between features was conducted using the χ^2 test statistic. Thresholding was used to determine whether the selected point was that of the ICP pulse onset. Using a tolerance of 30 ms, the WD method was able to achieve a positive predictive value (PPV) and sensitivity of 0.9475 and 0.9723, respectively at selecting the locations of the pulse onsets (Yang *et al* 2012). This methodology was not used to detect artifacts directly; however, it differentiated between normal pulses and artifacts/noise. It should be noted that, at present, this is a very computationally demanding method which could present a barrier for real-time implementation.

3.2. ICP signal artifact identification or removal methods

The methods in this section either identified the locations of artifacts or conducted artifact removal in addition to identification. These methods fall into distinct subcategories: empirical mode decomposition (EMD)-based methods (Feng *et al* 2011, 2012, Martinez-Tejada *et al* 2021), machine learning-based methods (Rajagopal *et al* 2016, Lee *et al* 2020, Mataczynski *et al* 2022, Taco *et al* 2022, Huo *et al* 2024), filter-based methods (Feng *et al* 2012), and wavelet-based methods (Feng *et al* 2012, Megjhani *et al* 2023). There were two methods developed for signals recorded using EVDs (Megjhani *et al* 2023, Huo *et al* 2024) and two developed for intraparenchymal strain gauges (Lee *et al* 2020, Mataczynski *et al* 2022). Four methods used auxiliary signals to either segment ICP pulses (Rajagopal *et al* 2016, Lee *et al* 2020) or jointly for artifact identification (Mataczynski *et al* 2022, Taco *et al* 2022), the other methods used only the recorded ICP signal. There were three methods that were developed specifically for low frequency (spike) artifacts (Feng *et al* 2011, 2012, Martinez-Tejada *et al* 2021). The method developed by Rajagopal *et al* indicated that it was able to remove both high and low frequency artifacts; however, it was only tested based on high frequency (additive Gaussian/Poisson noise) removal (2016). Each of these methods were developed using data from patients of varying cerebral health, with each article using datasets of different numbers of patients (median = 59, range = 5–309). Algorithms not trained using manually annotated datasets were validated by comparing the performance of algorithms compared to manual experts (Feng *et al* 2011, 2012, Martinez-Tejada *et al* 2021, Megjhani *et al* 2023) or compared to an established 'ground truth' signal with added noise (Rajagopal *et al* 2016). This was referred to as 'direct' validation. Other methodologies underwent validation using cross-validation (Lee *et al* 2020, Taco *et al* 2022, Huo *et al* 2024), external validation using data from an alternative center (Huo *et al* 2024), or in performing a train-test split of the data based on the neuropathological health of the patient (Mataczynski *et al* 2022). Live processing was indicated as possible in three of the developed methods (Rajagopal *et al* 2016, Mataczynski *et al* 2022, Huo *et al* 2024); for the remainder of the developed methods, this had not yet been explored. This information is summarized in table 3.

Additional details of the methods presented and their respective efficacy in identifying or removing artifacts are included in appendix C (tables C3–C6). A summary of these appendix tables is displayed in table 4.

3.2.1. EMD-based methods

A significant challenge to artifact identification in ICP data is the non-stationary nature of the ICP signal. As such, it is difficult to simply apply time- or frequency-series filters for artifact identification (Feng *et al* 2011). Three articles (two presenting an identical method) presented methods that apply EMD (Feng *et al* 2011, 2012, Martinez-Tejada *et al* 2021). EMD was used to transform non-stationary bio-signals into a finite number of intrinsic mode functions (IMFs) (Chaudhari *et al* 2016). These are functions with the same number of extrema and zero-crosses initially proposed by Huang *et al* (1998). A methodology was put forward by Feng *et al* that initially decomposed the ICP signals into 16 IMFs with each component corresponding to a different oscillatory frequency (2011, 2012). After which, a robust 3σ (where σ indicates standard deviation) filter was applied to the first IMF such that the locations and durations of large oscillations in the IMF were identified and removed as they corresponded to artifactual episodes. This same filter was iteratively applied to the subsequent IMFs to accurately estimate the width of the artifact more accurately within the time-series data (Feng *et al* 2011, 2012). A typical estimation of σ was deemed

Table 3. Algorithm information for ICP signal identification and removal methods.

Article	Patients	Specific pathological state	Artifact type	Validation	Live processing
Lee *et al* (2020)	309	TBI	High and low frequency	Cross-validation (10-fold)	Not indicated
Mataczynski *et al* (2022)	50	TBI + SAH	Not specified	Train + validate split (TBI), test (SAH)	Yes
Huo *et al* (2024)	60	TBI, SAH, meningitis, ICH, tumor + other	Not specified (can detect artifacts + drainages)	Cross-validation (5-fold) + external validation	Possible
Megjhani *et al* (2023)	229	SAH	EVD system unclamped	Direct	Not indicated
Feng *et al* (2011)	59	Not indicated	Low frequency	Direct	Not indicated
Feng *et al* (2012)	82	TBI	Low frequency	Direct	Not indicated
Rajagopal *et al* (2016)	25	IH	High and low frequency	Direct	Yes
Martinez-Tejada *et al* (2021)	5	Not indicated	Low frequency	Direct	Not indicated
Taco *et al* (2022)	39	TBI	Not specified	Cross-validation (5-fold stratified)	Not indicated

EVD = extra-ventricular drain, IH = intracranial hypertension, SAH = aneurysmal sub-arachnoid hemorrhage, TBI = traumatic brain injury.

inadequate to accurately model the signal in presence of artifacts, as such, this proposed method used 'mean (Calisto *et al* 2010, 2013) absolute deviation' (MAD) to estimate σ. The equation for MAD is defined in equation (1),

$$\begin{aligned} \mathrm{MAD} &= \mathrm{median}_i\left(\left|x_i - \mathrm{median}_i\left(x_j\right)\right|\right) \\ \sigma_{\mathrm{MAD}} &= K * \mathrm{MAD} \end{aligned} \tag{1}$$

where $\mathrm{median}_i\left(x_j\right)$ is the data median, x_i is a time-series data point, K is approximated as 1.4826 based on the assumption that the data is Gaussian, and σ_{MAD} is a robust approximation of σ due to its resistance to artifacts (Feng *et al* 2011, 2012). These two articles both detail validation that was conducted with ICP recordings from 59 patients (Feng *et al* 2011) and 82 TBI patients (Feng *et al* 2012).

The third EMD-based artifact identification algorithm targeted the removal of magnitude spikes in ICP. Similarly, this method identified peaks in successive IMFs from EMD decomposition as locations using thresholding based on the noise levels in summed IMFs (Martinez-Tejada *et al* 2021). This method was validated on 59 human ICP recordings of unknown sampling rate and patient population information (Martinez-Tejada *et al* 2021).

Two of the EMD methods were validated using identical metrics for artifact identification: precision (TP/TP + FPs) and recall (TP/(TP + false negative (FN))) (Feng *et al* 2011, Martinez-Tejada *et al* 2021). However, in the Martinez-Tejada *et al* article, these metrics were based on the ability of the EMD-based article of removing spike artifacts that had been identified by experts (2021) and the Feng *et al* work calculated them based on determination of location and duration of artifactual episodes (2011). The precision and recall of the Feng *et al* method were 100% and 73.6%, respectively (2011), while these metrics were 84% and 77%, respectively, for the Martinez-Tejada method (2021). The results presented by Feng *et al* indicate that an EMD-based algorithm had no FP (no artifacts that are misclassified); however, had significant occurrences of missed artifacts (2011). The EMD-based artifact management method presented by Martinez-Tejada *et al* seemingly also had some difficulty in missing and misclassifying the spike artifacts in the ICP signals (2021). It was recommended by the authors that the adaptive thresholding used on IMFs be improved (Martinez-Tejada *et al* 2021). The other Feng *et al* article proposed the results for this effectiveness in comparison to a wavelet- and filter-based method as the performance gain in metrics of mean-squared error (G_{MSE}), relative absolute error (G_{RAE}), forecast error (G_{FER}), and CT (2012). The results of which were: $G_{\mathrm{MSE}} = 24.7\%$, $G_{\mathrm{RAE}} = 9.07\%$, $G_{\mathrm{FER}} = 13.6\%$, and CT = 312 s.

3.2.2. Machine-learning-based methods

Another method of circumventing difficulties with artifact identification and removal of non-stationarity of ICP signals is in the extraction of certain features and analyzing them using neural networks (NNs) (Lee *et al* 2020). Three novel techniques have been developed for ICP artifact identification leveraging different NN-based techniques (Lee *et al* 2020, Mataczynski *et al* 2022, Taco *et al* 2022). There were two additional

Table 4. Summary of ICP signal artifact identification or removal methods.

Artifact removal methods	Number of studies included	Number of subjects	Sensor(s)	Effectiveness
Empirical mode decomposition-based methods	3	141	Not indicated	• An EMD-based method that decomposes ICP signal into 16 IMFs, filters the signal using a 3σ filter, and removes artifacts using an ARMA model was developed by Feng *et al* (2011). It had a precision of 100%, a recall of 73.6%, and an artifact width accuracy of 82%. This method was developed mainly to address short high-frequency artifacts. • A similar method was developed by Feng *et al* using an ARIMA model in place of ARMA. It obtained a MSE of 24.7%, RAE of 9.07%, a FE of 13.6%, and a computational time of 312 s (2012). It performed worse than the compared median filtering and wavelet-based methods. • The algorithm presented by Martinez-Tejada *et al* built upon the previous two EMD-based methods to specifically remove spike artifacts. It achieved precision of 84% and recall of 77% (2021).
Machine learning-based methods	5	483	Intra-parenchymal strain gauge, EVD	• The SCAE combined with CNN method for artifact identification proposed by Lee *et al* had the highest net prediction rate and sensitivity compared to the linear support vector machine (LSVM) and Radial basis function kernel support vector machine (KSVM). The net prediction rate for artifact removal was 94.1% and the sensitivity was 96.2% (2020). • Taco *et al* presented a comparison of several data learning algorithms for artifact identification in time, frequency, and time–frequency domains (2022). The LSTM method had the best accuracy in the time and frequency domain with 94.81% and 92.96% (tied with BiLSTM in frequency domain). The BiLSTM method had the best accuracy in the time–frequency domain, The CNN had consistently the best computational efficiency. • There was a ResNet-based method that was compared to several neural network methods by Mataczynski *et al* (2022). Some methods had solely an ICP signal as input, some used ABP. The single channel (ICP) ResNet had the highest 'best accuracy' and 'strict accuracy'. It significantly outperformed the CNN-SCAE method presented by Lee *et al* (2020). • Two processing pipelines, one using PCA-based hyperparameter optimization and one using sequential feature selection, were combined with two estimators which were histogram-based gradient boosting and extremely randomized trees. OvR AUCROC and AUPRC were highest in PCA & HGB (0.893 (0.871–0.918)) and PCA & HGB (0.610 (0.518–0.683)) (Huo *et al* 2024). • A subspace-based tracking framework was developed by Rajagopal *et al* (2016). By projecting ICP pulses into the define subspace, this algorithm could predict certain variables expected in successive pulses. The algorithm was able to use this to detect artifacts and replace them. The RS01 filter performed the best compared to generic kernels when the additive white Gaussian noise variance was 15%, 25%, and 35% (SNR = 25.8013, 25.5826, 25.8380, respectively) and when the Poisson noise variance was 25% and 35% (SNR = 25.8184, 25.9031, respectively) (Rajagopal *et al* 2016).

(Continued.)

Table 4. (Continued.)

Filter-based methods	1	82	Intra-parenchymal strain gauge	• A methodology was proposed by Feng *et al* that used a median filter and MAD filter to identify artifacts and an ARIMA model to remove them achieved an MSE of 24.6%, RAE of 10.2%, FE of 8%, and a computational time of 24 s (2012). The accuracy of this method seemed was very statistically similar to the wavelet and EMD-based methods to which it was compared; however, it had much better computational efficiency (Feng *et al* 2012). 9/30/24 2:34:00 PM.
Wavelet-based methods	2	311	Intra-parenchymal strain gauge	• A wavelet-based method that transformed the ICP signal using Haar wavelets and used its impulses to detect artifacts was presented by Feng *et al* (2012) the artifacts were removed using an ARIMA model. This method performed better by metrics of MSE (21.8%) and RAE (5.4%) compared to the presented EMD-based and median filter-based methods in this article. It was worse compared to the median filter-based method in metrics of FE and computational efficiency. • A method using a Morlet wave transformation using synchrosqueezing transform (SST) to calculate instantaneous frequencies to determine good ICP signals and remove artifacts was presented by Megjhani *et al* (2023). This method was intended to be used to label ICP waveforms as either normal or artifactual. Its performance was manually evaluated by experts. 86% of the labels were deemed 'Good', 2.0% were 'Ok with Missing' indicating partial correctness, 6.2% were 'Missing' which indicated that there was no label on the waveform, and 5.4% were 'False Positive' (Megjhani *et al* 2023).

ABP = arterial blood pressure, ARMA = auto regressive moving average, ARIMA = auto regressive integrated moving average, AUCROC = area under the receiver operating characteristic curve, AUPRC = area under the precision–recall curve, BiLSTM = bidirectional long short-term memory, CNN = convolutional neural network, EMD = empirical mode decomposition, FIR = finite impulse response, FE = forecast error, HGB = histogram-based gradient boosting, ICP = intracranial pressure, IMF = intrinsic mode function, KSVM = kernel support vector machine, LSTM = long short-term memory, LSVM = linear support vector machine, MAD = median absolute deviation, MSE = mean squared error, OvR = one-vs-rest, PCA = principal component analysis, RAE = relative absolute error, ResNet = residual network, SCAE = sparse convolutional autoencoder, SST = synchrosqueezing transform.

methodologies presented that used different machine learning techniques to identify and remove ICP signal artifacts (Rajagopal *et al* 2016, Huo *et al* 2024).

Lee *et al* proposed a method leveraging ABP signals recorded simultaneously with ICP (2020). This algorithm initially used the ABP signal to detect the pulse onset of the ICP signal, using the incidence and ending of the wave pulse as the basis for segmentation. After this, these pulses were normalized using cubic spline interpolation such that they are a constant length while maintaining the integrity of the signal morphology. The normalized pulses were then converted into a representative image using a stacked convolutional autoencoder (SCAE) and are classified as normal or artifactual using convolutional NN (CNN) filters (Lee *et al* 2020). The proposed CNN was compared to the use of linear SVM learning (LSVM) and radial basis function kernel SVM learning (KSVM). The models were developed, trained, and tested using data from 309 TBI patients. The efficacy of artifact removal in ABP and ICP was validated by comparing the results of the NN-based methods to manually cleaned data for 10 datasets from randomly selected patients. CNN was the most successful in artifact identification as it had highest net prediction rate for ABP and ICP artifact removal rates of 97.0% and 94.1%, respectively as well as the highest sensitivity for ABP and ICP at 97.3% and 96.2%, respectively (indicating it had the lowest number of artifacts misclassified).

Taco *et al* presented a methodology performing multi-modal data cleaning of ICP, ABP and ECG signals (ICP/ABP sampled at 125 Hz, ECG down sampled from 500 to 125 Hz) (2022). The proposed method used information from each bio-signal to conduct multivariate data cleaning based on different patterns identified. The NN-based methodologies that were discussed for this application included a CNN-model, long short-term memory (LSTM) model, bidirectional LSTM (both of which are recurrent NN models), and a transformer model (self-attention-based NN). Each of these NN models were used to try to identify artifacts in multivariate data in time domain, frequency domain (using Fourier transformation (FT)) and time–frequency domain (also through FT). Datasets were collected from 39 TBI patients and were stratified for cross-validation (Taco *et al* 2022). It was indicated that the CNN and LSTM methods were the most robust. The CNN model had a 92.22% accuracy in the frequency domain and the LSTM model had a 94.81% accuracy in the time domain. CNN was by far the most computationally efficient (GPU time of 7 min 24 s and 7 min 3 s in the time and frequency domains, respectively, while all others were over 1 h); however, LSTM was slightly more accurate. As such, the most advantageous method is dependent upon the application (Taco *et al* 2022).

The model proposed by Mataczynski *et al* used a one-dimensional vector of the normalized ICP signal samples as inputs to a NN, as such, this method involved first using a pulse detection algorithm to segment the data (Bishop and Ercole 2018, Mataczynski *et al* 2022) followed by cubic resampling and pulse scaling (Mataczynski *et al* 2022). The raw data was sampled from 50 neuro-ICU patients (39 TBI and 11 SAH). This article presented the performance of several NN-based models that were used for signal pulse classification. The classification of the pulses was in five categories: 'normal', 'possibly pathological', 'likely pathological', 'pathological', and 'artifact/error'. Some of the methods presented used ICP signals exclusively, while some had inputs of both ICP and ABP. The models compared included three 1-D residual neural networks (ResNet) (dual channel ABP and ICP, single channel ICP, and Siamese feature extractors), LSTM fully CNN (LSTM-FCN), fully connected NN (used as baseline), and an adapted version of that presented by Lee *et al* (2020), Mataczynski *et al* (2022). The presented advantage of ResNets was the use of residual connections between its layers for better error propagation (Mataczynski *et al* 2022). The efficacy of these NN-based methods was measured using 'strict' and 'best' accuracy based on the classes used. 'Best accuracy' accounted for a segment of data possibly being included in one classification based on uncertainty in the manual classification. The ResNet based method and the SCAE-CNN proposed by Lee *et al* (2020) performed similarly in training and validation; however, the ResNet based methods performed decidedly better on the test set of randomly selected 650 pulses. The single channel ResNet, performed the best in metrics of 'best accuracy' and 'strict accuracy' with results of 86.00% and 81.85%, respectively, compared to the adapted Lee *et al* method with 68.00% and 64.62%, respectively. The results of this study indicate that the method presented by Lee *et al* (2020) performs significantly worse than the Mataczynski *et al* (2022) methods; however, this is a poor comparison as the Mataczynski *et al* (2022) methods were those intended to classify five types of signal pulses whereas the Lee *et al* (2020) method was designed to solely classify pulses as artifactual or valid.

The results of each of these articles identify a variation of a CNN as an effective and efficient method for NN-based artifact removal due to its reduced CT and ability to easily extract morphological features from the ICP signals (Lee *et al* 2020, Mataczynski *et al* 2022, Taco *et al* 2022).

Huo *et al* presented two processing pipelines that leveraged machine learning techniques for implementation (2024). The methods that were presented used features extracted from the ICP signals to determine the onset of artifacts. The features extracted included summary statistics, entropy, root mean sum of squared distance, frequency domain features, and power spectra amongst others for a total of 94 features.

There were two preprocessing methods employed. The first of these methods used quantile transformation, principal component analysis (PCA), and performed hyperparameter optimization using a grid search approach. The second used a combined algorithm of sequential feature selection (SFS) and hyperparameter optimization. These two preprocessing methods were each combined with two estimators, histogram-based gradient boosting (HGB) and ETs. The cross-validation and validation of these models used data sampled from EVD in multi-center (40 Zurich and 20 Berlin) patients with varying Glasgow Coma Scales. This method was intended to label data segments as normal, artifactual, or corresponding to a drainage. It discarded the portions of signals that were identified as invalid. The performance of these models was measured using one-versus-rest (OvR) AUC-ROC and the area under the precision–recall curve (AUPRC) using a nested 5-fold cross-validation as well as for a labeled external dataset. The dataset was manually annotated by an expert for validation. Cross-validation indicated that the PCA preprocessing method outperformed the SFS, as the PCA & HGB method yielded the highest average OvR AUC-ROC and AUPRC values of 0.945 (95% confidence interval (CI): 0.92–0.969) and 0.692 (95% CI: 0.453–0.93), respectively (Huo *et al* 2024). A limitation of this method is that it classifies artifacts only after the single pulse waveform has been detected. Single pulses are not always present for every artifact (Huo *et al* 2024).

Rajagopal *et al* presented a single subspace decomposition tracking-based method designed to remove noise. This author defines 'egregious noise' as signal segments that can be introduced by electronic equipment, electrode transients, and patient motion as well as sensor displacement (defined in this paper as 'low frequency artifacts'). It was indicated that this algorithm was capable of removing spikes in signal deviation; however, it was validated on signals contaminated with additive white Gaussian noise and Poisson noise (2016)

An algorithm was to used extract ICP pulses from each signal using the ECG QRS wave to segment the pulse onsets as described by Choi *et al* (2010), Rajagopal *et al* (2016). The pulses were decomposed into three variables to describe normalized pulsatile information (length and amplitude), mean amplitude, and width of the pulse from onset to the elevation plateau (referred to as 'starting time-index') (Rajagopal *et al* 2016). An idealized signal ('ground truth') was created using clustering based on relative time-index information while preserving morpho-temporal locality to be used to evaluate the reconstruction accuracy of the algorithm. The initial signal was stored as reference data.

A subspace that could consistently project morphologically similar input samples was defined using SR-KDA. SR-KDA has integrated optimal graph embedding such that subspace similarities are also graphical similarities between input pulses. During learning, the subspace considers the mean amplitude and starting time-index information. The tracking framework is reliant upon the assumption that consecutive ICP pulses should have relatively similar morphologies. This is referred to as the RajagopalScalzo01 (RS01) algorithm and it functions by projecting pulses onto the subspace where sequential tracking is used to predict the locations of the successive samples in the subspace using k-nearest neighbor regression. Using the estimation of the successive signal morphology, the algorithm can detect artifacts in the signal. In the presence of artifacts, this algorithm will use graph searching to find similar pulses to estimate its morphology with the artifacts removed using reference data.

The RS01 algorithm was compared to three LPF generic kernels: LPF1, LPF2, and LPF3 (corresponding to variances of 5%, 10%, and 20% of normalized ICP amplitude). This algorithm was tested using the ground truth ICP data sampled from 25 TBI patients simultaneously with ECG (both at 400 Hz) that had additive white Gaussian noise and Poisson noise. The signal-to-noise ratio (SNR) was calculated between the noisy signal and the cleaned signal. The RS01 filter performed the best when the additive white Gaussian noise variances were 15%, 25%, and 35% (SNR = 25.8013, 25.5826, 25.8380, respectively) and when the Poisson noise variance was 25% and 35% (SNR = 25.8184, 25.9031, respectively). LPF1 performed between the other algorithms when the white Gaussian noise variance was 5% (SNR = 38.3530) and Poisson noise was 5% and 15% (SNR = 41.0770, and 26.2788, respectively). The RS01 seems to function well when noise is of high variance, and not as well when the noise is smaller (Rajagopal *et al* 2016).

3.2.3. Filter-based methods

Average filtering is a tool that can be used to reduce noise and artifacts while remaining simple, as such, it is particularly advantageous for online applications due to its minimal computational requirements (Feng *et al* 2012). There was a single method that leveraged this methodology for ICP artifact identification. Feng *et al* presented a median filter-based method to extract the trend from the non-stationary ICP signal and use it to determine the artifacts within the signal. The trend component ($T(n)$) of the k-order median filter was the median value based on a window of k time-series data points ($X(n)$). The optimal number for k was determined to be 50. The residual component ($R(n)$) was calculated $R(n) = X(n) - T(n)$ as it was deemed an indicator for artifacts in the ICP signal. The MAD filter previously discussed in section 3.1 was used to determine whether a segment was artifactual or normal. The results for this particular method were

$G_{MSE} = 24.6\%$, $G_{RAE} = 10.2\%$, $G_{FER} = 6.6\%$, and CT = 24 s (Feng *et al* 2012). This indicates similar performance in G_{MSE}, G_{RAE}, and G_{FER} to the EMD-based method while exhibiting dramatically better computational efficiency. An auto-regressive integrated moving average (ARIMA) model was used to remove the identified artifacts (Feng *et al* 2012). This filter-based method identifies artifacts with sound efficiency and high accuracy.

3.2.4. Wavelet-based methods

There are two articles that present methods using wavelet transformations for artifact management, they each employ a different type of wavelet transformation (Feng *et al* 2012, Megjhani *et al* 2023). The method proposed by Megjhani *et al* utilizes a continuous wavelet transformation (CWT) (equation (2)) to decompose a time-series signal into wavelet coefficients in the time–frequency domain (2023),

$$W(\tau, s) = \int_{-\infty}^{\infty} x(t) * \psi^* \left(\frac{t-\tau}{s} \right) \mathrm{d}t \tag{2}$$

where $W(\tau, s)$ denotes the wavelet coefficients, t denotes time, $x(t)$ denotes the input ICP signal, ψ^* denotes the complex conjugate of the mother wavelet function, τ denotes translation of the wavelet, and s is the dilation of the mother wavelet (Megjhani *et al* 2023). The mother wavelet used in the Megjhani *et al* method is the Morlet wavelet is displayed in equation (3) (2023). It is a sine wave multiplied by a Gaussian,

$$\psi(t) = \frac{1}{\sqrt{\pi * f_b}} * e^{j2\pi f_c t} * e^{\left(-\frac{t^2}{f_b}\right)}. \tag{3}$$

The parameter f_c denotes the center frequency of the wavelet, f_b denotes the bandwidth of the wavelet, and were both chosen as 1 for this application (Megjhani *et al* 2023). Additionally, the synchrosqueeze transform (SST) was used to sharpen the time–frequency domain signal. Artifact removal was conducted on EVD-based ICP data (240 Hz) sampled from 229 patients with SAH (125 female/104 male, 62 developed delayed cerebral ischemia). A continuous Morlet wavelet transformation was performed on the recorded ICP signals. The resulting CWT coefficients were squared to obtain signal power. The magnitude of the signal power between 0.33 and 4.16 Hz was used to determine when the signal was valid. Further, Otsu thresholding was used to determine whether the signal was true ICP data by using binary classification for the foreground and background (noise/artifacts) (Xu *et al* 2011, Megjhani *et al* 2023). As this method was developed for EVD ICP data, it was not noted how it might perform when data was recorded using intraparenchymal microsensors. This method was validated using one-hour segments that were randomly selected to be manually annotated by two experts. The algorithm had a success rate of 86.0% (8.2% poor segmentation or missed segmentation of artifacts, 5.4% FPs).

The second wavelet-based method presented by Feng *et al* used the Haar wavelet basis function to detect any rapid changes in ICP signal magnitude (2012). The absolute values of the wavelet coefficients (energy) across the frequencies at a particular time instance are aggregated as artifacts have a detectable impulse in this energy. The first derivative of the aggregated energy signal is filtered using the MAD filter to disseminate the normal signal from artifacts. The results for this method were $G_{MSE} = 21.8\%$, $G_{RAE} = 5.4\%$, $G_{FER} = 8.0\%$, and CT = 57 s. As such, when it was compared to the EMD-based and median filter-based methods, it performed worse than both in G_{MSE}, G_{RAE}, and G_{FER} and was only better than the EMD-based method in computational efficiency. A major limitation of this method that was noted is the selection of the basis function (Haar wavelet parameters), which highly impactful on its performance; however, there was no indication of how to select them when the method is applied.

3.2.5. Artifact removal methods

The aforementioned sections detail the methods that have been developed to identify artifactual segments of ICP data. Artifactual segments can be removed by directly removing the data points from the data stream; however, a more robust methodology would retain information regarding the underlying signal despite the initial presence of artifacts (Feng *et al* 2011). Feng *et al* developed two EMD-based methods for artifact detection that also used an auto-regressive moving average (ARMA) method (2011) and an ARIMA (Feng *et al* 2012) method to maintain the integrity of the ICP signal despite removing artifacts as opposed to removing the data altogether. The subspace tracking-based model developed by Rajagopal *et al* uses inverse mapping to produce a denoised signal after it has been projected onto the subspace and the coordinates of the successive samples have been estimated (2016).

Several articles included in this review either did not make direct mention to how artifacts would be removed following identification or suggested that the artifactual segments simply be discarded. An example

of this is the wavelet-based methodology presented by Megjhani *et al* (2023). This article does present an artifact removal method, however, in the presence of noise or artifacts, the data segment is simply removed. While simply identifying and removing artifactual segments remains valuable, in instances where the underlying signal can be maintained or when missing data points can be interpolated using temporal forecasting tools, it is extremely valuable for the utility of the dataset. Discarding segments of signals is inevitable in cases of some egregious artifacts where it is not possible to maintain the underlying signal. As seen in the example provided in figure 2, the ICP signal measured is unusable and will need to be removed entirely.

4. Discussion

This systematic review provides a detailed overview of the current landscape of artifact identification and removal for ICP full-waveform signals. In this review evaluating the performance and utility of 20 artifact identification/removal methods, a meta-analytical approach to the comparison of the results of this systematic review proved difficult due to the heterogeneity of several aspects of each study included.

There were nine methods presented in section 3.1 that were designed to identify certain features of valid ICP waveforms. While not directly identifying artifacts, these methodologies could be potentially applicable in this domain. The MOCAIP algorithm presented by Hu *et al* was the basis of many of these algorithms as they aimed to improve upon identified weaknesses in its performance (2009). As such, the effectiveness in the performance of some presented algorithms was internally compared to that of the MOCAIP algorithm (Asgari *et al* 2009, Scalzo *et al* 2010). The comparison of the effectiveness of each method between articles was difficult due to the varying metrics used, data sampling rates, pathological states of the patient populations examined, and ICP monitoring devices used. It remained difficult to identify a leading method for valid signal identification. The method presented by Megjhani *et al* indicated promising results as described in section 3.1.2; however, it had only been developed using a small dataset ($n = 34$) on ICP signals that had been recorded using an EVD system (2019).

There were five articles presenting NN-based and machine learning-based methods that reported some of the most robust analysis and encouraging artifact removal results in this review. Three articles presented in-depth comparisons of several NN-based methodologies and indicated that the use of CNN methods (or CNN variations) had the best performance in artifact identification and removal (Lee *et al* 2020, Mataczynski *et al* 2022, Taco *et al* 2022). The most recently published of these articles by Mataczynski *et al* (2022) included the CNN method presented by Lee *et al* (2020) in its comparison of how well different methods could classify ICP signal segments as 'normal', 'possibly pathological', 'pathological' or 'artifact/error'. These results indicated that the ResNet (CNN variation) had superior performance to the method proposed by Lee *et al* (2020), indicating that the use of residual connections to minimize error propagation in the CNN framework is beneficial in improving classification accuracy. However, it should be noted that, in this work, the method presented by Lee *et al* was extended as it was initially designed to only label signal segments as artifactual or valid. Additionally, the machine learning-based method presented by Rajagopal *et al* (2016) indicated strong results evaluated using SNR in the presence of white Gaussian and Poisson noise. The other machine learning-based method, the most successful of which was the PCA & HGB method presented by Huo *et al* (2024), also performed well in differentiating artifacts from valid signals as indicated by metrics of AUC-ROC and AUPRC.

Another notable method presented was a filter-based method. This methodology was presented by Feng *et al* (2012). It was a median filter coupled with a MAD filter, respectively. A comparison was conducted by Feng *et al* between a median filter-based method, EMD-based method, and a wavelet-based method (Feng *et al* 2012). The median filter-based method performed equal or better than the other methods by metrics of MSE, RAE, and FER but had the best computational efficiency.

4.1. Limitations of literature

There are several pertinent limitations to the literature examined in this review that made it difficult to elucidate a leading method in artifact management for ICP signals. There was vast heterogeneity in the metrics used to describe the effectiveness of each method, sensor types used for data recording, and patient pathology considered in these articles presenting novel artifact management methods.

In the current literature, there was inadequate consideration of the impact of variance in pathophysiological health on the effectiveness of a particular method in removing artifacts. The work presented by Mataczynski *et al* indicated that there was a significant difference in best accuracy and strict accuracy in the test set and the validation set, which corresponded to different neuropathophysiological

states (2022). In a comparison of several NN-based models, each model had been developed using training data sampled from TBI patients. Testing and validation were conducted using SAH ($n = 650$ pulses) and TBI patients ($n = 6812$ pulses), respectively. The results of the most successful method (single-channel ResNet) indicated a best and strict accuracy in the test set (SAH data) were 86.00% and 81.85%, respectively, and these same metrics in the validation set (TBI data) were 95.17% and 92.95%, respectively. This indicates that there needs to be one of two paths forward: (1) external validation using data from patients with varying pathology for a method extended to a broad range of patients or (2) the scope of patients for which a particular artifact management tool should be used should be clearly defined. There were no other articles that applied methodologies to patient sets with different neurological health than those used for training the methods. This is of specific importance for methods that use specific thresholds or ICP pulse templates to differentiate valid ICP pulses from artifacts. In instances of cerebral pathological processes or surgical interventions like decompressive craniectomy, the magnitude and morphology can change dramatically which can affect the accuracy of the analysis of the signal.

There were a variety of different sensor types that were used for data collection for each article which could have resulted in different artifact morphologies being present. Similarly to instances of specific neuropathology or surgical interventions, the development of an algorithm using a particular sensor type may affect its ability to identify or remove all signal artifacts if measured using a different sensor. An example of this is drainage artifacts present in EVD recordings that are absent from other sensor types. There were no instances in which the change in effectiveness of a method was noted when it was extended to an ICP signal recorded using a different device.

Tables 1 and 3 depict the artifact types for which each algorithm was developed. There were very few articles in which artifact types were specified. Four methods were explicitly stated to have been designed with the capability to consider both high and low frequency artifacts (Asgari *et al* 2009, Lee *et al* 2020), three that were stated to consider high frequency artifacts (Feng *et al* 2011, 2012, Martinez-Tejada *et al* 2021), two methods capable of removing EVD specific artifacts corresponding to being unclamped or drainages occurring (Megjhani *et al* 2023, Huo *et al* 2024), and a single method dedicated to removing signal noise (Rajagopal *et al* 2016). It is essential that an automated method for adequately cleaning ICP signals be capable of removing all artifact types, as such, vague descriptions of different artifact categories are not sufficient. Empirical definitions of artifacts that can be identified or removed by each method are essential for a more robust comparison like those depicted in the Dai *et al* narrative review (2020).

As discussed in section 3.2.5, only a few articles included in this review that provided any information regarding a methodology to remove artifacts while maintaining the structure of the signal. The outright removal of segments of the signal containing artifacts will significantly diminish the utility of the signal.

There has been inadequate external validation conducted on the artifact management methods that have been developed thus far. There were four articles that presented methods that had been externally validated on other datasets; however, only one article used external data from a patient with a different neuropathology. As discussed, there are so many variables that can be overlooked in artifact identification if the datasets on which these algorithms are trained and validated do not adequately encapsulate the scope in which it will eventually be used. Without robust external validation, each artifact management method developed should only be used for data from the patient type, sensor type, and artifact type for which it was initially designed.

There were several articles that mentioned the real-time implementation of artifact identification methodologies; however, there were few articles that either directly stated that real-time application had been conducted (Rajagopal *et al* 2016, Mataczynski *et al* 2022) or provided data indicating it was possible (Huo *et al* 2024). More research in the real-time implementation of artifact management is necessary.

The use of auxiliary signals like ECG or ABP to detect the onset of signals was present in seven methods included in this review. Auxiliary signal use can present a barrier to widespread implementation of a particular method. The use of ECG and ABP simultaneously recorded does not present as significant of a barrier due to the prevalence of these bio-signals in ICU recording.

4.2. Limitations of review

There were limitations associated with the scope of this review. Only English studies were included, as such, there was a language bias present. There was only literature included that had been published prior to 19 June, 2024, which could have led to recently published articles being missed. There were articles that were included that did not directly identify artifacts but identified valid ICP signals and there were no criteria included in the initial search string for these specific article types. As such, there could have been valid ICP detection algorithms that were missed in this search. Additionally, due to the heterogeneity of the study types, there may have been inadequate conclusions made in comparing articles due to a lack of sufficient information for a meta-analysis.

4.3. Future recommendations

There is extensive work that must be done prior to the real-time implementation of artifact management in ICP signals. External validation using more diverse datasets must be conducted to improve understanding of the variability in ICP and their artifacts in the context of differing recording methodologies and neuropathology health. This knowledge base is critical as it determines the scope of implementation of a particular method. A better-defined scope of the capabilities of an artifact management method using better defined dataset criteria, effectiveness metrics, and empirical definitions of artifacts in ICP signals will allow for a far more robust comparison of artifact management tools in this application. Work must also be done to consider the removal of artifacts while maintaining the structure of the signal such that, in the cases of less egregious artifacts, the signal can remain useful despite its raw form having artifacts present.

Several different techniques that have been explored as options for artifact management; however, without a robust comparison of the models using consistent effectiveness metrics and artifact definitions, it remains nebulous as to which method type is the most advantageous. Both ML-based and filter-based methods have provided encouraging results, but more work should be done with more diverse and larger datasets to obtain greater certainty. Ultimately, a sequential approach applying several methodologies may be required to adequately remove the diverse artifacts present in ICP signals.

Future work that will be conducted in this lab will involve the development of a methodology that employs several signal analysis techniques, many of which were discussed in this review. This layered approach will involve the following:

1. Threshold-based methods—will be used to detect extraneous data points as well as signal drift based on the expected magnitude of signals.
2. Time-series autoregression-based methods—will be used to detect large magnitude spikes common in high frequency artifacts that occur during patient motion.
3. Wavelet or FT-based methods—the transformation of the time-series ICP into the time–frequency or frequency domains will allow for the detection of artifacts in the oscillatory behavior of ICP data.
4. Waveform morphology detection-based methods—a catalog will be developed based on high frequency data such that complex morphological artifacts can be identified using their morphological structure like those that have been based off the Hu *et al* morphological clustering and analysis of ICP algorithm (Hu *et al* 2009), which has been the basis for several morphology-based signal detection algorithms for ICP signals (Asgari *et al* 2009, Scalzo *et al* 2010, 2012).

All these techniques will be applied using recursive and CNNs and will be merged into a single autonomous artifact management pipeline using machine learning techniques. In the presence of different artifact types, specifically those that are high frequency, different methodologies will be investigated to maintain the underlying signal. Methodologies such as ARMA models (Feng *et al* 2011) and ARIMA models (Feng *et al* 2012) have been employed for ICP signals specifically; however, work must still be conducted in this domain to explore the applicability of more complex statistical models such as dynamic regression, *k*-nearest neighbors, or other methods of interpolation of adjacent points.

In future works, it would be beneficial to develop a more detailed library of different erroneous pulse types based on their origin (subject motion, device disconnection, coughing, etc); however, this would require some kind of real-time annotation of the ICP signals which was not available during the collection of the data included in the presented examples.

5. Conclusion

A search was conducted across five databases to examine the current literature on artifact management methods for ICP signals. This search yielded methods of two categories: (1) valid ICP pulse detection algorithms, (2) ICP artifact identification and removal methods. This is a thorough review of existing ICP artifact management methods. However, several factors make it premature to identify a single leading method. These factors include different metrics of effectiveness as well as varying artifacts present based on patient neuropathology and sensor types. Future work is necessary, particularly in real-time model applications, to better understand the impacts of these factors on the effectiveness of different leading methods and method types. This work is critical as it is a foundational step to build towards improving understanding of high temporal relationships within cerebral physiology.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary information files).

Acknowledgment

FAZ is supported through the Endowed Manitoba Public Insurance (MPI) Chair in Neuroscience/TBI Research Endowment, NSERC ((DGECR-2022-00260, RGPIN-2022-03621, ALLRP-578524-22, ALLRP-576386-22, I2IPJ 586104–23, and ALLRP 586244-23), Canadian Institutes of Health Research (CIHR), the MPI Neuroscience Research Operating Fund, the Health Sciences Centre Foundation Winnipeg, the Canada Foundation for Innovation (CFI) (Project #: 38583) and Research Manitoba (Grant #: 3906 and 5429).

TB is supported through NSERC (ALLRP-578524-22 and Canada Graduate Scholarship—Masters (CGS-M)).

NV is supported by NSERC (RGPIN-2022-03621, ALLRP-578524-22, ALLRP-576386-22, ALLRP 586244-23), the Graduate Enhancement of Tri-Council Stipends (GETS)—University of Manitoba, and the University of Manitoba Graduate Fellowship (UMGF)—Biomedical Engineering.

ASS is supported through the University of Manitoba Graduate Fellowship (UMGF)—Biomedical Engineering, NSERC (RGPIN-2022-03621), and the Graduate Enhancement of Tri-Council Stipends (GETS)—University of Manitoba.

AI is supported by a University of Manitoba Dept of Surgery GFT Grant, the University of Manitoba International Graduate Student Entrance Scholarship (IGSES), the University of Manitoba Graduate Fellowship (UMGF)—Biomedical Engineering, and by NSERC (ALLRP-578524-22).

AG is supported through a CIHR Fellowship (Grant #: 472286).

KYS is supported through the NSERC CGS-D program (CGS D-579021-2023), University of Manitoba R.G. and E.M. Graduate Fellowship (Doctoral) in Biomedical Engineering, and the University of Manitoba MD/PhD program.

LF is supported through a NSERC Post-Doctoral Fellowship (PDF).

IM is supported through the Undergraduate Research Award (URA) program at the University of Manitoba, and NSERC (ALLRP 586244-23).

NS is supported through NSERC (ALLRP-578524-22).

Funding

This work was directly supported through the Endowed Manitoba Public Insurance (MPI) Chair in Neuroscience and the Natural Sciences and Engineering Research Council of Canada (NSERC; ALLRP-578524-22 and Canada Graduate Scholarship—Masters (CGS-M)).

Conflict of interest

The authors report no conflicting interests.

Author contributions

Conceptualization, T.B. and F.A.Z.; methodology, T.B. and F.A.Z.; validation, N.V., A.G., A.I., K.Y.S., A.S.S., N.S., I.M., L.F., and F.A.Z.; formal analysis, T.B.; investigation, T.B.; resources, F.A.Z.; data curation, T.B., writing—original draft preparation, T.B.; writing—review and editing, N.V., A.G., A.I., K.Y.S., A.S.S., N.S., I.M., L.F., and F.A.Z.; visualization, F.A.Z. and T.B.; supervision, F.A.Z. and L.F.; project administration, F.A.Z.; funding acquisition, F.A.Z. All authors have read and agreed to the published version of the manuscript.

Appendix A

Table A1. PRISMA ScR checklist.

Section and Topic	Item #	Checklist item	Location where item is reported
Title			
Title	1	Identify the report as a systematic review.	Pg 1
Abstract			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Pg 1
Introduction			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Pg 3–4 (section 1)
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Pg 4 (section 1)
Methods			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Pg 4 (section 2.2)
Information sources	6	Specify all databases, registers, websites, organizations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Pg 4 (section 2.3)
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Pg 4 (section 2.3) + Appendix B (Pg 24)
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Pg 4 (section 2.4)
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Pg 4–5 (section 2.5)
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Pg 4–5 (section 2.5)
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	N/A

(Continued.)

Table A1. (Continued.)

Section and Topic	Item #	Checklist item	Location where item is reported
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	All articles published in academic journals, as such, biases were assumed to have been screened
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	N/A
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	N/A
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	N/A
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	All data items for each method were tabulated and are included in appendices C (tables C1–C6)
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	N/A
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	N/A
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	N/A
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Pg. 18 (section 4.2)
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	N/A
Results			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Pg 5–6 (section 3)
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Pg 5 (section 3)
Study characteristics	17	Cite each included study and present its characteristics.	Pg 5–17 (sections 3.1 and 3.2)
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	All articles published in academic journals, as such, biases were assumed to have been screened

(Continued.)

Table A1. (Continued.)

Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Pg 5–17 (sections 3.1 and 3.2)
Results of syntheses	20a	For each synthesis, briefly summarize the characteristics and risk of bias among contributing studies.	All articles published in academic journals, as such, biases were assumed to have been screened
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	No statistical synthesis conducted
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Pg 5–17(sections 3.1 and 3.2)
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Pg 5–17 (sections 3.1 and 3.2)
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Pg 18 (section 4.2)
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	N/A
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Pg 17 (section 4)
	23b	Discuss any limitations of the evidence included in the review.	Pg 17–18 (section 4.1)
	23c	Discuss any limitations of the review processes used.	Pg 18 (section 4.2)
	23d	Discuss implications of the results for practice, policy, and future research.	Pg 18–19 (section 4.3)
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Review was not registered
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Protocol not prepared
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	N/A
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Pg 20 (Funding)
Competing interests	26	Declare any competing interests of review authors.	Pg 20 (Conflicts of interest)
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	N/A

Appendix B

The detailed strings of keywords that were used as search parameters in BIOSIS, SCOPUS, EMBASE, PubMed, and Cochrane Database were:

'Intracranial pressure' OR 'ICP' OR 'Intracranial Pressures' OR 'Pressure, Intracranial' 'Pressures, Intracranial' OR 'Intercranial pressure' OR 'Intercranial Pressures' OR 'Pressure, Intercranial' OR 'Pressures, Intercranial' OR 'Cranial pressure' OR 'Cranial Pressures' OR 'Pressure, Cranial' OR 'Pressures, Cranial' OR 'Subarachnoid Pressure' OR 'Pressures, Subarachnoid' OR 'Pressure, Subarachnoid' OR 'Subarachnoid Pressures' OR 'Intracranial hypertension' OR 'IH' OR 'Hypertension, Intracranial' OR 'Intracranial hypotension' OR 'Hypotension, intracranial' OR 'Intercranial hypertension' OR 'Hypertension, intercranial' OR 'Intercranial hypotension' OR 'Hypotension, intercranial' OR 'Cranial hypertension' OR 'Hypertension, cranial' OR 'Cranial hypotension' OR 'Hypotension, cranial' OR 'Subarachnoid hypertension' OR 'Hypertension, cranial' OR 'Subarachnoid hypotension' OR 'Subarachnoid, cranial'

AND

'Artifact management techniques' OR 'Artefact management techniques' OR 'Artifact management' OR 'Artefact management' OR 'Artifact removal method' OR 'Artefact removal method' OR 'Artifact removal' OR 'Artefact removal' OR 'Artifact management system' OR 'Artefact management system' OR 'Artifact removal technique' OR 'Artefact removal technique' OR 'Error management' OR 'Error removal' OR 'Error management method' OR 'Error removal method' OR 'Error management technique' OR 'Error removal technique' OR 'Error management system' OR 'Error removal system' OR 'error' OR 'artifact' OR 'artefact'

Appendix C

Table C1. Valid pulse detection algorithms.

Reference	Subject information	Data type (sampling rate)	Sensor type (and sensor location, if provided)	Artifact removal method(s) used	Methods compared	Effectiveness	Study results and conclusions	Limitations
Aboy *et al* (2001)	3 patients with severe TBI at pediatric ICU, age demographic indicated as 'pediatric'	ICP (125 Hz)	Ventricular catheter or parenchymal fiber-optic pressure transducer	Method is used to detect peaks. There are four phases: 1. Lowpass filter—this algorithm uses a moving average filter to reduce the impact of high-frequency noise on the signal quality. The resulting signal will resemble a sinusoid that contains one cycle per beat. 2. Detection of coarse beats—the resulting signal is assumed to have one peak per heartbeat, this is the basis to detect the temporal location of each beat. 3. Quantization error filter—quantization errors are often present in the sampled signal. A moving average filter with a length dictated by the RQE is used to address the quantization error. 4. Local peak detection—search for certain peaks in heartbeat. These include percussion peak (P), tidal peak (T), dicrotic notch (N), and dicrotic peak (D). Search is conducted in 0.24 s before and after each coarse beat. PTND components are detected in that order.	Two experts identified each percussion component for each patient's recording and were screened for significant artifacts.	Percentage accuracy was calculated using total number of beats detected and combination of false negatives and positives for approximately 1000 beats each. Agreements between expert and detection algorithm were 99.40% and 99.30% at an acceptance interval of 8.0 ms. Agreement between experts was 99.70%.	The results indicate that the algorithm used was as accurate as the experts were between one another. The experts took 1 h to identify the beats in 8 min of recording, the algorithm tool 1.5 s Algorithm was approximately as accurate at identifying a significant artifacts.	No indication as to how to deal with artifacts if located. Small cohort used (3 patients), the algorithm was only tested on 1000 beats.

(Continued.)

Table C1. (Continued.)

Reference	Subject information	Data type (sampling rate)	Sensor type (and sensor location, if provided)	Artifact removal method(s) used	Methods compared	Effectiveness	Study results and conclusions	Limitations
Aboy *et al* (2005)	210 patients admitted to pediatric ICU (60 TBI, 60 with sepsis, 90 with cardiac issue), 2 entire recordings used, age demographic indicated as 'pediatric'	ICP (125 Hz) ABP SpO_2	Sensor type and location not indicated	1. Preprocessing—uses three bandpass filters. Filter #1: removes trend and high frequency noise. Resulting signal then used to determine cut off frequencies of next filters. Filter #2: removes oscillations above 2.5 times the heart rate such that the resulting signal has reduced high-frequency noise and only one cycle per heart contraction. Filter #3: detrends signal using high pass filter. It removes frequencies that are below 0.5 times signal and smooths the signal using a filter with cut off at 10 times the heart rate. 2. Spectral heart rate estimation—the signal is partitioned and a harmonic PSD technique is employed to robustly detect the heart rate encapsulated by the ICP signal. 3. Peak detection and decision logic—non-linear rank-based filters are used to identify the location of peaks within the previously cleaned signals. These rank filters are applied both to magnitude and slope spikes of the signal such that the location of the peaks can be accurately determined. The interval beat intervals (IBI) are used to measure the temporal difference between successive signals such that peaks are not missed. Median filtering is used to remove impulsive noise. Thresholding can be conducted to the residual signals to locate detection errors. 4. Nearest neighbor decision logic—a combination of slope and beat amplitude information is used to determine whether certain peak in the smoothed signal is valid or not using a nearest neighbor algorithm. 5. IBI Classification Logic—based on IBI of data containing peaks, looks for cases in which artifact removal method has 'over detected' or misdetected. This process is repeated until the IBIs of the dataset have a value within an acceptable range.	Two experts manually annotated sections of the data, total of 42 539 beats by one, and 7128 beats by another.	Sensitivity and positive predictivity for these algorithms were calculated for 2300 beats. The average sensitivity and positive predictivity between the algorithm the experts were 99.40% and 99.40%, respectively, with an acceptance interval of 16 ms. The average sensitivity and positive predictivity between experts were 99.57% and 100%, respectively.	Algorithm is nearly as accurate as the experts are with each other. Similar or superior performance is expected with higher sampling rate.	IBI element of this algorithm is not computationally efficient as it must make several passes through.

ABP = arterial blood pressure, Hz = Hertz, IBI = inter-beat interval, ICP = intracranial pressure, ICU = intensive care unit, PSD = power spectral density, QRE = relative quantization error, SpO_2 = oxygen saturation.

Table C2. Morphological information extraction algorithm.

Reference	Subject information	Data type (sampling rate)	Sensor location (and sensor location, if provided)	Artifact removal method(s) used	Methods compared	Effectiveness	Study results and conclusions	Limitations
Asgari *et al* (2009)	67 patients from Adult Hydrocephalus center (33 M/34 F), ages 14–94 years	ICP (400 Hz or 240 Hz)	Intra-parenchymal strain gauge placed in right frontal lobe.	Extension of MOCAIP algorithm using singular value decomposition (SVD) that defines a signal and noise subspace (designed with the intension to reduce computational cost, reduce the need to manage and maintain the ICP pulse library). New pulses are validated based on the ratio of energies of the projected signal into the subspace. This is used to create a more diverse library. MOCAIP phases include: 1. Pulse detection 2. Pulse clustering 3. Pulse validation 4. Detection and designation of three subcomponent peaks. Using the MOCAIP pulse validation, thresholds of the correlation coefficient are reached to validate the pulse (which is time and computationally consuming). The proposed SVD method uses stretching/shrinking the timescale (using spline interpolation) and normalization to match the segmented pulse to reference library pulses. A new 'effective reference library' was defined as a set of bases for the signal subspace. The experimental segment is projected onto the defined noise and signal subspaces. A threshold of the ratio of the energies in these subspaces determines the validity of the pulse. Diversification of the library conducted through selective extension using different base values for the signal subspace can be treated as a new reference library of valid ICP pulses.	Evaluation of performance of method for valid ICP identification through 67-fold cross validation comparing SVD-based method and correlation-based method.	Validity of dominant ICP pulses assessed using energy ratio test calculating true positive rate (TPR) and false positive rate (FPR) were systematically calculated; results for each fold (TPR/FPR) used to calculate ROC. Wanted to show that TPR is the same as correlation-based method but FPR is lower. Computational time (CT) also evaluated. Lowest FPR for correlation-based method was when correlation coefficient threshold (r) set as 1: TPR = 97.00% FPR = 28.92% CT = 16 370 s When $r = 0.97$ (value set by Hu *et al* (2009)) TPR = <99% FPR = 47.48% CT = 1927.2 s For SVD-based, $C = 3$ dB TPR = 96.61% FPR = 6.61% CT = 170 s	SVD method has much better computational efficiency and FPR without compromising dramatically on TPR compared to the Hu *et al* MOCAIP correlation-based algorithm (2009).	SVD-based method measures validity based on shape, not by absolute value of features. ICP wave is typically tri-phasic but morphology is not always consistent between patients.

(Continued.)

Table C2. (Continued.)

Reference	Subject information	Data type (sampling rate)	Sensor location (and sensor location, if provided)	Artifact removal method(s) used	Methods compared	Effectiveness	Study results and conclusions	Limitations
Hu *et al* (2009)	66 patients with various ICP-related conditions (32 M/34 F), patient age demographic was 61 years (14–94)	ICP (400 Hz or 240 Hz) ECG (400 Hz or 240 Hz)	Intra-parenchymal strain gauge placed in right frontal lobe.	Morphological clustering and analysis of ICP pulse (MOCAIP) is presented in this article. The steps of this algorithm are as follows: 1. Uses previously developed algorithm for identifying ICP pulse using ECG QRS wave (Hu *et al* 2008b) with a backstop in case of ECG errors using ICP constraints. 2. Hierarchal clustering was used to extract a representative ICP pulse by determining the dominant ICP wave from the average ICP pulse of the cluster. 3. Reference library of ICP pulses were used to determine whether dominant ICP pulses were valid or artifactual. The pulse cluster corresponding to an ICP pulse must correlate with those from reference library. Changes were made to correlation coefficient thresholds to attain differing levels of effectiveness. 4. Detection of the peaks in the ICP signal involves four criteria involving the curvature of the pulse, the result of which is determined using the second derivative and distance between points. 5. The assignment of three peaks based on peak density function criteria.	Compared to baseline peak designation algorithm also presented by Hu *et al* (Hu *et al* 2008a).	Under optimal parameters. For identifying non-artifactual pulses: TPR = 99.04% FPR = 20.08% Acc = 97.84% For identifying three peaks: TPR = 91.45%, 88.10%, and 86.62% FPR = 18.52%, 38.87%, and 52.94% Acc = 90.17%, 87.56%, and 86.53% Without reference library. Baseline achieved accuracy of 65%	Algorithm with optimized parameters was able to achieve string results in TPR and FPR for identifying the QRS wave peaks and non-artifactual ICP signals.	Baseline algorithm is reliant on reference library Need to further expand library for more accurate designation and potentially add further metrics for signal identification.

(Continued.)

Table C2. (Continued.)

Scalzo *et al* (2009)	66 patients with various ICP-related conditions (32 M/34 F), age demographic indicated as 'adult'	ICP (240 Hz or 400 Hz) ECG	Intra-parenchymal strain gauge placed in right frontal lobe	This model builds on the MOCAIP algorithm using machine learning. This paper evaluates different regression frameworks for the peak designation: MOCAIP is the basis of the algorithm used to identify locations of three peaks in ICP pulse. Typically, MOCAIP uses Gaussian priors to determine the position of the peaks from a set of peak candidates. To more aptly address the morphological variability of ICP signals, this article presents several regression based alternatives to improve peak designation and enable the use of machine learning algorithms in this application. This new algorithm initially conducts pre-processing to convert the ICP signals into normalized vector representations. The MOCAIP algorithm is used to determine candidate data points that are potentially peaks in the ICP pulses, the locations of the peaks are determined using a variety of proposed regression models. These include: 1. Multiple linear regression (MLR) (Chatterjee and Hadi 1986) 2. Spectral regression (SR) (Cai *et al* 2007) 3. Support vector machine (SVM) (Chang and Lin 2001) 4. Kernel spectral regression (KSR) (Cai *et al* 2007) 5. Extremely randomized decision trees (extra-trees) (Geurts *et al* 2006).	Each of the regression methodologies were analyzed on their abilities to properly label the peaks: 1. MLR 2. SVM 3. KSR 4. SR 5. Extra-trees They were compared to the 'actual' locations of the peaks which are presumed to have been labeled by trained human observers.	There were three comparisons for the regression frameworks that were evaluated, each had different methodologies and test/train splits: 1. Prediction accuracy (global)—accuracy of regression method to detect three peaks within ICP pulse. 5-fold cross validation of 11 080 ICP pulses randomly selected from dataset. Each fold of cross validation used 8864 pulses for training and 2216 pulses for testing. 2. Prediction accuracy on new patients (individual)—to train regression model for individual patients such that patients data is used as test set but not training set. 64-fold cross validation (one fold per patient, except for one) that training set of each fold had 4000 random pulses. 3. Number of training samples required—5-fold cross validation on 11 080 ICP pulses (trade-off between too few and too many training samples as it can increase computational requirements for learning time). For KSR: Global error $= 1.32$ ms $\pm$ 0.053 Individual error $= 3.63$ ms For extra-trees: Global error $= 2.64$ ms $\pm$ 0.21 Individual error = approx. 3 but not indicated	The best performance in global and individual prediction accuracies were of the KSR and extra-trees algorithms. If a large number of samples are used for KSR, SVM, and extra-trees, there is better performance but much slower efficiency.	KSR, SVM, and extra-trees models (leading methods) require more computational complexity.

(Continued.)

Table C2. (Continued.)

Reference	Subject information	Data type (sampling rate)	Sensor location (and sensor location, if provided)	Artifact removal method(s) used	Methods compared	Effectiveness	Study results and conclusions	Limitations
Megjhani *et al* (2019)	34 patients with EVD placed, recorded during 47 clamp trials, age demographic not indicated	ICP (125 Hz) ECG	EVD placed in the ventricle	ICP pulses were identified using morphological clustering analysis of ICP pulse (MOCAIP) by first identifying ICP pulses using ECG QRS waves. Hierarchal clustering is employed in short time segments where ICP pulses are clustered based on morphological distances. Dominant pulse identified. Proposed approach for waveform detection was 'active learning for valid ICP waveform detection (AL)'—there is a dominant ICP pulse of 30 s that is identified by the MOCAIP which is given a binary classification of either 'valid' or 'artifact'. The logistic regression model relies upon the assumption that the features are independently and identically distributed. The proposed framework uses the minimum determinant (variance reduction) of the Fisher information matrix and its inverse to estimate parameters. Three raters trained the model by the model showing them an example of a valid waveform and then adjusting their parameters. This process is repeated. It is trained until information gain plateaus.	Extracted 126 725 pulses, for error margin for 95% confidence level used 200 samples for test set and the rest for training set (3 raters manually annotated them) 3 approaches compared for valid ICP waveform detection 1. Template matching (TM)—compared to static reference library of ICP pulses validated by experts. 2. ICP stability algorithm (IS)—amplitude and duration of pulse is then compared with mean amplitude and duration of a double window comprising 3 previous and 3 future waves. 3. Threshold-based single wave criteria (TB)—thresholds defined based on amplitude and latency (presumed between ECG and ICP).	Cohen's kappa statistic for agreement between manually labeled test set, labels classified my active learning model on test set, and those classified by model for unlabeled dataset. Metrics used include PAUC, CCR, sensitivity, specificity, PPV, NPV all calculated at above 0.9 specificity and sensitivity, respectively. Also AUC was calculated. AUC: 1. AL (0.96 ± 0.02) 2. TM (0.71) 3. IS (0.51) 4. TB (0.5) CCR: 1. AL (0.94 ± 0.02) 2. TM (0.8) 3. IS (0.7) 4. TB (0.63) PPV and NPV 1. AL 2. TM 3. IS 4. TB Constraints that sensitivity and specificity must be above 0.9 indicated that the PAUC was better for TM than IS and TB, but AL was better than all three.	The AL model performed the best at identifying valid ICP signals, even if they do not exhibit the typical triphasic shape and it does not require a large dataset. It does not have some of the same pitfalls that the compared methods do, specifically in variability in ICP waveforms leading to false negatives or long stretches of artifacts.	Model relies upon having previously seen certain waveforms for removal. The more information that the model has, the more accurate the estimate. Small cohort used

(Continued.)

Table C2. (Continued.)

Scalzo *et al* (2012)	128 patients with various intracranial pressure related conditions, age demographic not included	ICP (400 Hz) IPP	Intra-parenchymal strain gauge placed in the right frontal lobe or sensor placed in the ventricle	This algorithm was proposed to be used to track ICP peaks in real-time by exploiting their temporal dependencies between the successive peaks. The tracking algorithm assigns a random variable to each of the three peaks in a typical ICP pulse and the relationship between the peaks were determined using a non-parametric belief propagation algorithm, which is a non-parametric Bayesian inference algorithm This algorithm was not developed directly for artifact identification; however, it is able to predict the temporal location and magnitude of ICP pulses despite induced dynamics acting as noise, as well as missing ICP peaks.	MOCAIP method on extracted dominant pulses	Using artificially distorted ICP signals, this algorithm was able to recover the intrinsic trend of the data set despite incidents of noise and missing inputs. Average prediction error in latency for first, second, and third peaks were 8.09 ± 2.0 ms, 6.90 ± 1.7 ms, and 7.46 ± 2.1 ms using proposed method. Average prediction error in latency for first, second, and third peaks were 11.88 ± 7.1 ms, 11.80 ± 7.0 ms, and 11.76 ± 7.0 ms using MOCAIP method.	Proposed algorithm can estimate the ICP elevation and latency of the three ICP peaks in a beat-by-beat fashion in the presence of noise/missing information.	Performance may worsen if lower sampling rate is used.
Scalzo *et al* (2013)	154 IH risk patients (108 patients treated for TBI, SAH, ICH), age demographic information not included	ICP (240 Hz) ECG	EVD placed in the ventricle	Regression model using supervised learning on labeled ICP datasets. First methodology for alarm detection had two stages: 1. Morphological clustering and analysis of ICP pulse algorithm (MOCAIP): for pulses immediately prior to an alarm. Looks at morphology of signals in 20 min prior to alarm. There are 24 morphological metrics examined. 2. Conditional discretization of morphological features (CDF): used to enhance the occurrences of false alarms. It processes 24 metrics of MOCAIP and mean ICP using adaptive joint occurrence frequency based on simultaneous occurrence. These are joined to make a feature vector called CDF. Second methodology was regression-based detection model based on input features extracted from segments of ICP data. Three methodologies were compared: 1. Spectral regression discriminant analysis (SR-DA), formulates solving DA as regularized regression problem. 2. Kernel Spectral Regression (SR-KDA), which generalizes SR-DA for kernel projection to be used and non-linearity to be obtained. 3. Support vector machine (SVM), machine learning technique to optimize the separation of hyperplane to limit misclassification while maximizing the sum of distances for the training samples from the hyperplane.	Comparison of SR-DA, SR-KDA, and SVM was conducted using 10-fold cross-validation. Outer 10-fold CV conducted for 5 independent runs. AUC-ROC was calculated for each. AUC-ROC reflects both sensitivity and specificity. Sensitivity also evaluated using true positive rate (TPR = TP/(TP + FN). Can be calculated for TPRs of 90%, 95%, and 97.5%. Also compared to simple threshold-based method.	AUC-ROC of threshold-based method = 55.9 Improvements in terms of AUC from morphological metrics to CDF features were respectively: 1. SR-KDA (69.2 ± 1.2–85.9 ± 1.1) 2. SVM ($64.1 \pm .9$–80.6 ± 2.6) 3. SR-DA (55.2 ± 7.1–79.4 ± 3.7) Reductions in FPR for 97.5% TPR was from 72.5 to 53 ± 0.3 for SR-KDA combined with CDF features.	Based on AUR-ROC and FPR, the SR-KDA combined with CDF features resulted in the best performance in the identification of certain signal morphologies and identifying false alarms.	False alarms are reduced but are not zero. Further work will be conducted on incremental training of algorithms to identify sections of signal recording that require annotations.

(Continued.)

Table C2. (Continued.)

Reference	Subject information	Data type (sampling rate)	Sensor location (and sensor location, if provided)	Artifact removal method(s) used	Methods compared	Effectiveness	Study results and conclusions	Limitations
Scalzo *et al* (2010)	Original data: 128 patients treated for various ICP conditions, no age demographic information provided. 'Challenging data' is subset of this data that was manually selected.	ICP (240 Hz or 400 Hz)	Intra-parenchymal strain gauge placed in the right frontal lobe.	Builds on MOCAIP framework from (Hu *et al* 2009) and builds on previous work from this primary author who replaced the use of Gaussian priors for a regression model. The new proposed method is MOCAIP++ which expands the number of peak regression techniques that can be used and exploit different ICP features. 1. ICP pulse segmentation to extract dominant pulse—uses the method presented by Hu *et al* (Hu *et al* 2008b) to detect the latency of the three peaks relative to the ECG QRS wave that is measured. The more robust pulse is located using the consecutive pulses and hierarchal clustering. 2. Detect peak candidates within dominant pulse—second derivative of ICP pulse used to determine convex and concave portions of the curve. 3. Extract ICP features from pulse—features included the first derivative of the time-series signal, second derivative, and the curvature. 4. Formally detects three peaks and extracts features—to choose the peaks amongst the peak candidates, three techniques were compared: a Gaussian model, Gaussian mixture models, linear spectral regression, and non-linear extension of spectral regression (kernel spectral regression).	Based on their performance on the challenging dataset: 1. Gaussian-based model (MOCAIP) 2. Gaussian Mixtures (GMM) 3. Spectral Regression (SR) 4. Kernel Spectral Regression (KSR). 72.57% ± 2.6, the use of the second derivative and curvature improves it to 80.26 ± 2.29 and 80.4% ± 2.2	Five-fold cross-validation with five iterations, four folds for training and one for evaluation. Average accuracy of the prediction of the three peaks: 1. KSR (88.78% ± 2.35) 2. SR (72.57 ± 2.6) 3. GMM (70.47% ± 2.64) 4. MOCAIP (65.83% ± 2.96) Computational cost for 2000 ICP pulses 1. MOCAIP (60 ms) 2. SR (90 ms) 3. KSR (1,340 ms) 4. GMM (33,940 ms) Computational cost for 1 pulse 1. SR (0.19 ms) 2. MOCAIP (6.7 ms) 3. GMM (2.3 ms) 4. KSR (19.6 ms) Using feature based peak detection improvements: 1. SR + L_{xx} = 80.26 ± 2.29 2. SR + curvature = 80.4% ± 2.2 3. GMM + L_x = 77.14% ± 1.85 4. KSR + L_x = 89.36% ± 2.51 Otherwise results remained the same or decreased.	SR-based extension to MOCAIP performs nearly as well as the KSR-based; however, SR is much more computationally efficient. Methodology does not result in overfitting, as when these methods were tested using normal data, results improved. This will allow for more reliable extraction of data in the ICU.	KSR requires all training pulses to keep as part of the model. Some data excluded because it was invalid.

(Continued.)

Table C2. (Continued.)

Yang *et al* (2012)	48 patients (15 TBI, 14 NPH, 19 hydro-cephalus), age demographic information was not provided.	ICP (400 Hz)	Parenchymal fiber-optic pressure transducer, location was not specified.	Referred to as "waveform descriptor" algorithm. Waveforms are transferred into the polar coordinate system using a log-polar transformation after the maximum and minimum of the signal segment have been determined. Shape context describes a given point using the 'distribution of the relative positions of that point and the remaining points on the image edge'. This was the basis for the waveform descriptor intended to determine the onset of ICP pulses using a segment before and after. The ICP signal of a particular pulse is scaled, and each point is transformed using a log-polar transformation. The signal is then normalized and then divided into bins. To detect the pulse onsets, there is a multi-stage process: 1. A template is constructed using representative ICP segments with two consecutive pulses with manually identified onset, then a template is constructed using feature extraction from onset by waveform descriptor. Automated segment selection is then conducted using hierarchal clustering to automatically select 43 representative segments. 2. For pulse onset detection of the signal, feature of each local minima is derived using the determined waveform descriptor. Feature extraction is conducted using average inter-neat interval of 6 nearest neighbors. 3. Similarity between local minimum and template is calculated using the χ^2 test statistic 4. Onset identified once similarity reaches a particular threshold. 5. 6 nearest neighbor pulses are incorporated into the template for future selection. Template built from dataset of 996 segments randomly selected from 22 patients (each segment with two consecutive beats starting from one onset and ending in another). Training set from 122 800 normal beats and 674 artifactual beats from 15 patients. The test dataset composed of 40 933 normal beats and 306 with artifacts from 11 patients.	ABP pulse onset detector algorithms: 1. Slope sum function (SSF) (Zong *et al* 2003). 2. Pulse waveforms delineator (PUD) (Li *et al* 2010). 3. Global minimum with sliding window (GM).	Predicted and ground truth were considered correct if the distance was equal or less than 30 ms. Sensitivity: 1. WD (0.9723) 2. SSF (0.9720) 3. PUD (0.9599) 4. GM (0.9226) PPV 1. WD (0.9475) 2. PUD (0.9327) 3. SSF (0.9136) 4. GM (0.8968)	The WD algorithm consistently had better PPV across tolerances of 10, 20, 30, 40, 50, and 60 ms. There were no statistical differences in terms of performance of WD compared to the other methods for specificity.	Required storage of templates Computational cost as high in WD. Its cost was 53.2132 s whereas the rest of the compared methods were ∼3 s or below.

ABP = arterial blood pressure, Acc = accuracy, AUC-ROC = area under the receiver operating characteristic curve, CCR = correct classification rate, CDF = conditional discretization of morphological features, EVD = extra-ventricular drain, FPR = false positive rate, GMM = Gaussian mixture models, IH = intracranial hypertension, ICH = intracerebral hemorrhage, ICP = intracranial pressure, IPP = intraparenchymal pressure, IS = ICP stability algorithm, KSR = kernel spectral regression, MOCAIP = morphological clustering and analysis of ICP pulse, NPV = negative predictive value, PAUC = partial area under the curve, PPV = positive predictive value, SVD = singular value decomposition, SVM = support vector machine, SR-DA = spectral regression discriminant analysis, SR-KDA = kernel spectral regression, TB = threshold-based single wave criteria, TPR = true positive rate, WD = waveform descriptor, χ^2 = chi-squared test.

Table C3. Empirical mode decomposition-based.

Reference	Subject information	Data type (sampling rate)	Sensor location	Artifact removal method(s) used	Methods compared	Effectiveness	Study results and conclusions	Limitations
Feng *et al* (2011)	59 patients admitted to neuro ICU continuously monitored for 24 h (avg. 203.5 h), age demographic not indicated	ICP	Sensor type and location not indicated	Artifact detection: uses empirical mode decomposition prior to conducting filtering to improve its efficacy as ICP signals are non-stationary. EMD decomposes signal into 16 IMFs. (IMFs describe unique modes of oscillations of the signal). Afterwards, the IMFs are filtered using a 3σ filter, anything outside of these filters is considered an artifacts. Iterations are conducted to estimate the width and exact locations of the identified artifacts. Artifact imputation: in order to not lose significant portions of datasets—impute detected artifact episodes based on ARMA which predicts values of the stationary time-series signal (relies upon stationarity in short portions of non-stationary ICP signal).	Not stated but assumed that the comparison was on manually cleaned signal.	Evaluated based on how well algorithm can detect artifacts and estimate widths. Precision (no useful signal misclassified): 100% Recall (artifacts missed): 73.6% This can be improved using more iterations, still need to determine optimal number. Width was detected with 82% accuracy (tends to overestimate).	AR and MA order of ARMA set to 20 *F*-score which describes overall performance of proposed method was 0.848 (higher *F*-score is better relationship between precision and recall). Method to remove artifacts is included (ARMA).	Higher computational overheads due to number of IMFs used. Poor performance with artifacts with especially large magnitude spikes. Seems to only target high frequency artifacts.

(Continued.)

Table C3. (Continued.)

Feng *et al* (2012)	82 TBI patients (monitored at least 12 h), age demographic not indicated	ICP	Sensor type and location not indicated	Artifact detection: 3 methods compared 1. Empirical mode decomposition (EMD) involves decomposing the ICP signal into IMFs. A MAD filter was used to identify position and duration of artifacts. 2. Wavelet transformation transforms ICP signal into "energy" using Haar wavelet. Based on aggregated energy, impulses will be used to detect artifacts, then calculate first derivative of aggregated energy and apply MAD filter. Robust MAD filter, adjusted 3σ filter where the cut-off is calculated based on an assumed normal distribution of artifacts. 3. Median filtering method involves applying a median filter and obtaining a trend component of the signal, calculating the residual components and applying a MAD filter. Artifact imputation: After identification, artifacts are imputed based on ARIMA model. Artifacts were often short and as such, ARIMA can normalize non-stationary.	EMD Wavelet Median filter	MSE: 1. Wavelet (21.8%) 2. Median (24.6%) 3. EMD (24.7%). RAE: 1. Wavelet (5.4%) 2. EMD (9.07%) 3. Median (10.2%). FE: 1. Median (6.6%) 2. Wavelet (8%) 3. EMD (13.6%) Time 1. Median (24 s) 2. Wavelet (57 s) 3. EMD (312 s)	Recommended the median filtering method due to similar degree of effectiveness but has dramatically higher computational efficiency. Method to remove artifacts is included (ARIMA).	Seemed to primarily target high frequency artifacts.

(Continued.)

Table C3. (Continued.)

Reference	Subject information	Data type (sampling rate)	Sensor location	Artifact removal method(s) used	Methods compared	Effectiveness	Study results and conclusions	Limitations
Martinez-Tejada *et al* (2021)	Not indicated 26 h from 5 different monitoring sequences, patient age demographic not included	ICP	Not indicated	EMD algorithm proposed by Huang *et al* (1998) used to identify large spikes in ICP signal. EMD is a sifting method to convert non-stationary time-series data into a finite number of IMFs. This is an iterative process such that the original signal is expressed as the sum of all extracted IMFs and one residual. IMFs can be used to locate frequency of spiking artifact. EMD-Based Algorithm: 1. Break ICP into 16 separate IMFs based on physiological properties (using knowledge from (Feng *et al* 2011)). 2. Estimated IMFs used for spike detection. 3. Spike imputation in original signal.	1 s window size and p-values around 0.03, significance level equal to 0.05, data is non-stationary. ICP spikes manually inspected by expert using template (must be shorter than 0.5 s and abrupt magnitude increase).	Precision: 84% Recall: 77% TP = 114 FP = 21 FN = 34	Algorithm is able to identify most artifacts but precision and recall indicate that further work is required to improve results.	Functionality of adaptive threshold. Method is only able to identify certain artifacts, cannot remove. Small cohort used and no demographic information given

AR = auto-regressive, ARMA = auto-regressive moving average, EMD = empirical mode decomposition, FN= false negative, FP = false positive, F-score = performance measure balancing precision and recall, ICP = intracranial pressure, ICU = intensive care unit, IMF = intrinsic mode function, MA = moving average, MSE = mean squared error, RAE = relative absolute error, FE = forecast error, TP = true positive, TBI = traumatic brain injury.

Table C4. Machine learning-based.

Reference	Subject information	Data type (sampling rate)	Sensor location	Artifact removal method(s) used	Methods compared	Effectiveness	Study results and conclusions	Limitations
(Lee *et al* 2020)	309 subjects with traumatic brain injury. Recording within 24 h of admission. Age demographic was not included.	ICP ABP	Intra-parenchymal strain gauge	Stacked convolutional autoencoder (SCAE) and a convolutional neural network (CNN). Only 10 patients worth of data used for testing and training. Method detects ABP pulse onset to allow for ICP pulse onset and the segmentation of each pulse waveform. Segmented pulses were normalized by interpolation and converted into representative images by the SCAE without any loss of morphological features. The representative images were labeled as either 'valid' or 'artifactual' using a CNN.	Linear support vector machine (LSVM) learning method. Radial basis function kernel support vector machine (KSVM) learning	Effectiveness was calculated compared to cleaned ICP and ABP signals from two independent experts on 10 randomly selected patient data sets (p value as 0.01). Leave-one-out 10-fold cross validation used. Proposed method had highest net prediction rate for ABP and ICP artifact removal as 97.0% and 94.1%, respectively. Proposed method had highest sensitivity for ABP and ICP as 97.3% and 96.2%, respectively (lowest number of artifacts misclassified).	Effectiveness of removing artifacts included a variety of different high-frequency artifacts occurring in ABP and ICP signals. * Using this artifact removal tool there were a reduced proportion of patients that experienced 'clinical events' (ex. hypotension, hypoperfusion) during recording. * Changes in number of events calling for immediate attention (ex. external ventricular drainage etc) after artifact removal. * Reduced number of events in which ICP and CPP were outside of normal range * - changes in misdiagnoses.	Retrospective analysis with anonymized data. No indication whether reducing artifacts would improve outcomes in TBI treatment. Only tested on ICP patients, which could limit the presence of certain morphological artifacts present in other neurological disease cases. Proposed method can have some misclassification leading to high false positive rate, requires optimization in the SCAE.

(Continued.)

Table C4. (Continued.)

Reference	Subject information	Data type (sampling rate)	Sensor location	Artifact removal method(s) used	Methods compared	Effectiveness	Study results and conclusions	Limitations
Mataczynski *et al* (2022)	50 neuro-ICU patients, 39 of which were TBI patients (11 of those with SAH), age demographic was 45 (20–85). Continuous-ly monitored for 2–8 d	ICP (sampling rates between 50 and 300 Hz, but all resampled to 50 Hz) ABP (used for dual-channel methods)	Intra-parenchymal strain gauge placed in the frontal cortex.	Main method proposed was 2 variants of 1-D residual neural network (ResNet) (He *et al* 2016)—dual channel with joint ICP and ABP signal or Siamese feature extractors (which are meant to mimic manual annotators that consider both ABP and ICP). ResNets are CNN that use residual connections to minimize error propagation. Also a Long-Short Term Memory Fully Convolutional Network (LSTM-FCN) (Karim *et al* 2018). Use both CNN and LSTM feature extractors. 35 TBI patients (4/39 excluded due to large gaps in recording) were assigned to the training and validation datasets, 11 aSAH subjects were used for testing. The algorithm classified the ICP pulse segments as 'normal', 'possibly pathological', 'likely pathological', 'pathological', and 'artifact/error' Analysis was conducted with best method to evaluate real-time processing.	Standard single label accuracy score (strict accuracy). Another metric where if the algorithm also predicts 'possible' artifacts correct (best accuracy). Single channel fully connected was considered 'baseline'.	Using Test set best result 1. Single channel ResNet (best Acc = 86.00%, strict Acc = 81.85%) 2. Siamese ResNet 3. Dual channel ResNet 4. Single channel LSTM-FCN 5. single channel fully connected 6. Lee *et al* method (2020). Using validation set best result 1. Single channel ResNet (best Acc = 95.17%, strict Acc = 92.95%) 2. Dual channel ResNet 3. Siamese ResNet 4. Single channel LSTM-FCN 5. Lee *et al* method 6. Single channel fully connected Using train set best result 1. Dual channel ResNet (best Acc = 99.14%, strict Acc = 98.74%) 2. Single channel ResNet 3. Siamese ResNet 4. Single channel LSTM-FCN 5. Lee *et al* method 6. Single channel fully connected. Single channel ResN*et al*so performed well with artificially added artifacts.	Single channel ResNet can classify ICP waveforms with reasonably good accuracy in test data set. Not dependant on a particular length of recording.	The classification of signals is reliant on waves falling into particular 'classes. If a wave falls into two classes simultaneously it can lead to an error. Follows similar errors due to ambiguity in waveform shape as human classifiers. Authors believe that using a larger dataset could diminish this. A limitation of this method is that it classifies artifacts only after the single pulse waveform has been detected. Single pulses are not always present for every artifact (Huo *et al* 2024).

(Continued.)

Table C4. (Continued.)

Taco *et al* (2022)	39 severe TBI patients, age demographic information not indicated	ICP (125 Hz) ABP (125 Hz) ECG (500 Hz)	Sensor type and location not indicated	Proposed methodology: 1. Data windowing (break into 10 s windows). 2. Down sampling ECG data. 3. Fourier transformation of signals into frequency domain. 4. Standardization of each signal (easier for ANN). 5. Data splitting (70%–15%–15% ratio of training, validation, test data, respectively). 6. Stratified cross validation. 7. Deep learning algorithm used to train final model using either CNN, LSTM, BiLSTM, and Transformer. Comparison of which ML method was most effective and whether time-series, frequency-series, or time–frequency-series inputs were better.	Time-series data CNN, LSTM, BiLSTM, Transformer. Frequency-series data CNN, LSTM, BiLSTM, Transformer. Time–frequency-series data CNN, LSTM, BiLSTM, Transformer.	Effectiveness was measured using Acc, TP, FP, FN, TN, GPU time. Time series Acc: LSTM (94.81%) TP: CNN (91.55%) FP: LSTM (0.00%) FN: CNN (8.45%) TN: LSTM (100%) GPU time: CNN (7 min 24 s). Frequency series Acc: LSTM/BiLSTM (92.96%) TP: LSTM (96.48%) FP: CNN (3.91%) FN: LSTM (4.23%) TN: CNN (96.09%) GPU time: CNN (7 min 3 s). Time–frequency series Acc: BiLSTM (92.59%) TP: CNN (95.07%) FP: Transformer (5.47%) FN: CNN (4.93%) TN: Transformer (94.53%) GPU time: CNN (8 min).	CNN has best computational efficiency in frequency domain. LSTM has best accuracy in time domain.	Different methods are better for different applications.

(Continued.)

Table C4. (Continued.)

Reference	Subject information	Data type (sampling rate)	Sensor location	Artifact removal method(s) used	Methods compared	Effectiveness	Study results and conclusions	Limitations
Huo *et al* (2024)	40 neurocritical care patients from first site (develop) 20 neurocritical care patients (external validation) Age demographic indicated as 'adult'	ICP (125 Hz, down sampled to 25 Hz due to data processing restrictions)	EVD placed in the ventricle	Feature generation used. 94 descriptive features were used including: median, maximum, minimum, in/max difference, entropy, root mean sum of squared distance, frequency domain features like power spectral density and derived features. This was conducted with and without normalization). Reapplied after high-pass filtration. 2 processing pipelines compared: 1. Quantile transformation to normalize feature set, then applied PCA, then performed hyperparameter optimization based on a nested 5-fold cross-validation and grid search approach. 2. Sequential feature selection (SFS) and hyperparameter optimization from feature subset in 'greedy fashion'. These were each combined with estimators—histogram-based gradient boosting (HGB) and extremely randomized trees (ETs) (which were internally validated).	Internal model validation was done using 5-fold cross-validation and averaged multiclass AUCROC and AUPRC (precision recall curve) in "one-versus-rest" method (for identification of normal signal, artifacts, and drainage incidents).	OvR AUCROC and AUPRC (precision recall curve) were highest in PCA & HGB (0.893 (0.871–0.918)) and PCA & HGB (0.610 (0.518–0.683)). The rest of the ranking is as follows: 1. PCA & ET 2. SFS & HGB 3. SFS & ET.	Algorithm able to classify minor artifacts as erroneous. This model was noted to inadequately label several segments of data with 56% of artifacts and 8.2% of drainages as normal signals as well as 14% of drainages falsely labeled as artifacts and 3.6% as false positives.	Indicates model can be used for artifact removal but does not specify how. This method seemed to struggle with the middle sections of drainages and artifacts occurring when the drain is opened or closed.

(Continued.)

Table C4. (Continued.)

Rajagopal *et al* (2016)	25 patients suffering from intercranial pressure related conditions, age demographic not included	ICP (400 Hz) ECG	Intra-parenchymal strain gauge placed in the right frontal lobe	Method presented is Iterative Causal Subspace Tracking (I/CST) which uses manifold learning techniques which considers the average ICP of the pulses and their relative position in time as having a relationship with the overall shape of the pulse. This method uses a subspace learning algorithm to determine a kernel discriminant analysis of reference pulses that can be solved using spectral regression. SR-KDA is used to find a regression model capable of generating similar projections for data samples. Tracking is used on consecutive samples such that the signal can be denoised. Prediction algorithm used is *k*-nearest neighbor regression. Values are constrained to the average of similar waveforms of its *k*-nearest neighbors. This framework is RajagopalScalzo01 (RS01) denoising algorithm.	Between baseline and various denoising kernels with variances from the normalized ICP amplitude of 5% (LPF1), 10% (LPF2), and 20% (LPF3) to the compared.	Removal of additive white-Gaussian noise and Poisson noise using signal to noise ratio (SNR). The RS01 filter performed the best when the additive white Gaussian noise variance was 15%, 25%, and 35% (SNR = 25.8013, 25.5826, 25.8380, respectively) and when the Poisson noise variance was 25% and 35% (SNR = 25.8184, 25.9031, respectively). LPF1 performed between than the other algorithms when the white Gaussian noise variance was 5% (SNR = 38.3530) and Poisson noise was 5% and 15% (SNR = 41.0770, and 26.2788, respectively). The RS01 seems to function well when noise is of high variance, and not as well when the noise is smaller.	RS01 does not require previous knowledge of the noise profile in order to be effective. RS01 is more robust across frequencies whereas specific kernels can function better at a particular noise. Performs well when noise is above 5% of signal magnitude. RS01 deals well with maxima and minima.	The size of averaging kernel is not easily determined solely using input data, it requires initial calibration. Some useful high frequency data can be inadvertently masked using a Gaussian filter. RS01 performs poorly when the noise is less than 3% of the signal magnitude.

ABP = arterial blood pressure, AUC-ROC = area under the receiver operating characteristic curve, ANN = artificial neural network, Acc = accuracy, CNN = convolutional neural network, CPP = cerebral perfusion pressure, ECG = electrocardiogram, EVD = external ventricular drain, FN = false negative, FP = false positive, GPU = graphics processing unit, ICP = intracranial pressure, ICU = intensive care unit, KSVM = radial basis function kernel support vector machine, LSTM = long short-term memory, LSVM = linear support vector machine, ResNet = residual neural network, SCAE = stacked convolutional autoencoder, TP = true positive, TN = true negative, TBI = traumatic brain injury, AUPRC = area under the precision–recall curve, ET = extremely randomized trees, HGB = histogram-based gradient boosting, I/CST = iterative causal subspace tracking, LPF = low pass filter, PCA = principal component analysis, RS01 = RajagopalScalzo01 (denoising algorithm), SFS = sequential feature selection, SNR = signal to noise ratio, SR-KDA = spectral regression—kernel discriminant analysis, sub-arachnoid hemorrhage.

Table C5. Filter-based.

Reference	Subject information	Data type (sampling rate)	Sensor location	Artifact removal method(s) used	Methods compared	Effectiveness	Study results and conclusions	Limitations
Feng *et al* (2012)	82 TBI patients (monitored at least 12 h), age demographic information not included	ICP	Sensor type and location not indicated	Artifact detection: 3 methods compared 1. Empirical mode decomposition (EMD) involves decomposing the ICP signal into IMFs. A MAD filter was used to identify position and duration of artifacts. 2. Wavelet transformation transforms ICP signal into "energy" using Haar wavelet. Based on aggregated energy, impulses will be used to detect artifacts, then calculate first derivative of aggregated energy and apply MAD filter. Robust MAD filter, adjusted 3σ filter where the cut-off is calculated based on an assumed normal distribution of artifacts. 3. Median filtering involves applying a median filter and obtaining a trend component of the signal, calculating the residual components and applying a MAD filter. Artifact imputation: After identification, artifacts are imputed based on ARIMA model. Artifacts were often short and as such, ARIMA can normalize non-stationary.	EMD Wavelet Median filter	MSE: 1. Wavelet (21.8%) 2. Median (24.6%) 3. EMD (24.7%) RAE: 1. Wavelet (5.4%) 2. EMD (9.07%) 3. Median (10.2%) FE: 1. Median (6.6%) 2. Wavelet (8%) 3. EMD (13.6%) Time: 1. Median (24 s) 2. Wavelet (57 s) 3. EMD (312 s)	Recommended the median filtering method due to similar degree of effectiveness but has dramatically higher computational efficiency.	Seemed to primarily target high frequency artifacts

ARIMA = auto regressive integrated moving average, EMD = empirical mode decomposition, FIR = finite impulse response, ICP = intracranial pressure, MAD = mean absolute deviation, MSE = mean squared error, PCA = principal component analysis, PSD = power spectral density, RAE = relative absolute error, SA = signal averaging, TBI = traumatic brain injury.

Table C6. Wavelet-based.

Reference	Subject information	Data type (sampling rate)	Sensor location	Artifact removal method(s) used	Methods compared	Effectiveness	Study results and conclusions	Limitations
Megjhani *et al* (2023)	229 patients who had EVDs after SAH (155 F/74 M) with 62 developing cerebral ischemia, age demographic of patients was 56 years (46-67)	ICP (240 Hz)	EVD placed in the ventricle	Any ICP values greater than 80 mmHg were removed. Finite impulse response bandpass filter was applied with cutoff frequencies of 0.1 and 40 Hz. Wavelet transformation (Morlet CWT) was completed to convert the ICP time-series signals into the time–frequency domain. Synchrosqueezing transform (SST) was performed to calculate instantaneous frequencies. Determined absolute power between 0.33 and 4.16 Hz. Low pass filter applies at 0.1 and Otsu thresholding was used to identify thresholds for absolute power to determine good ICP. Small unwanted noise was removed using morphological operation (on resulting binary image).	Validation was conducted using a graphical interface where random segments were assigned to evaluators such that they could examine the algorithms performance. Performance was labeled as "Good", "Ok with Missing", "False Positive", "Missing", and "Not Sure"	86% labeled "Good" 2.0% were 'Ok with Missing', i.e. some was labeled correctly. 6.2% were 'Missing' 5.4% were 'False Positive'	The presented algorithm was intended to mimic the human reviewer procedure of identifying artifacts. Algorithm able to identify ICP waveforms when EVD was clamped.	Use more robust thresholding like adaptive thresholding, maximum entropy thresholding, iterative thresholding, iterative thresholding, histogram-based methods, and region-based methods. Not all evaluators agreed on 100% of the segments. Sudden changes in patient state were a common factor leading to false positives. Wavelet was sensitive to abrupt changes in ICP magnitude. Otsu's method is not always suitable when multiple thresholds are required to be used.

(Continued.)

Table C6. (Continued.)

Reference	Subject information	Data type (sampling rate)	Sensor location	Artifact removal method(s) used	Methods compared	Effectiveness	Study results and conclusions	Limitations
Feng *et al* (2012)	82 TBI patients (monitored at least 12 h), age demographic information was not included	ICP	Sensor type and location not indicated	Artifact detection: 3 methods compared 1) Empirical mode decomposition (EMD), involves decomposing the ICP signal into IMFs. A MAD filter was used to identify position and duration of artifacts. 2) Wavelet transformation, transforms ICP signal into "energy" using Haar wavelet. Based on aggregated energy, impulses will be used to detect artifacts, then calculate first derivative of aggregated energy and apply MAD filter. Robust MAD filter, adjusted 3σ filter where the cut-off is calculated based on an assumed normal distribution of artifacts. 3) Median filtering method involves applying a median filter and obtaining a trend component of the signal, calculating the residual components and applying a MAD filter. Artifact imputation: After identification, artifacts are imputed based on ARIMA model. Artifacts were often short and as such, ARIMA can normalize non-stationary.	EMD Wavelet Median filter	MSE: 1. Wavelet (21.8%) 2. Median (24.6%) 3. EMD (24.7%) RAE: 1. Wavelet (5.4%) 2. EMD (9.07%) 3. Median (10.2%) FE: 1. Median (6.6%) 2. Wavelet (8%) 3. EMD (13.6%) Time: 1. Median (24 s) 2. Wavelet (57 s) 3. EMD (312 s)	Recommended the median filtering method due to similar degree of effectiveness but has dramatically higher computational efficiency.	Haar wavelet relies on the right selection of the basis function parameters. Seemed to primarily target high frequency artifacts.

ARIMA = auto regressive integrated moving average, CWT = continuous wavelet transform, EMD = empirical mode decomposition, EVD = extra-ventricular drain, FE = forecast error, FIR = finite impulse response, ICP = intracranial pressure, MSE = mean squared error, RAE = relative absolute error, SAH = aneurysmal sub-arachnoid hemorrhage, SST = synchrosqueezing transform.

ORCID iDs

Tobias Bergmann https://orcid.org/0009-0001-1082-6485
Nuray Vakitbilir https://orcid.org/0000-0003-2764-145X
Alwyn Gomez https://orcid.org/0000-0002-3737-2065
Abrar Islam https://orcid.org/0009-0006-6669-3374
Kevin Y Stein https://orcid.org/0000-0002-5983-008X
Amanjyot Singh Sainbhi https://orcid.org/0000-0003-3231-5683
Noah Silvaggio https://orcid.org/0009-0003-3860-8134
Logan Froese https://orcid.org/0000-0002-6076-0189
Frederick A Zeiler https://orcid.org/0000-0003-1737-0510

References

Aboy M, McNames J and Goldstein B 2001 Automatic detection algorithm of intracranial pressure waveform components *2001 Conf. Proc. 23rd Annual Int. Conf. IEEE Engineering in Medicine and Biology Society* vol 3 (IEEE) pp 2231–4 (available at: http://ieeexplore.ieee.org/document/1017216/)

Aboy M, McNames J, Thong T, Tsunami D, Ellenby M S and Goldstein B 2005 An automatic beat detection algorithm for pressure signals *IEEE Trans. Biomed. Eng.* **52** 1662–70

Asgari S, Xu P, Bergsneider M and Hu X 2009 A subspace decomposition approach toward recognizing valid pulsatile signals *Physiol. Meas.* **30** 1211–25

Bishop S M and Ercole A 2018 Multi-scale peak and trough detection optimised for periodic and quasi-periodic neuroscience data *Intracranial Pressure & Neuromonitoring XVI* (*Acta Neurochirurgica Supplement*) vol 126, ed T Heldt (Springer International Publishing) pp 189–95 (available at: http://link.springer.com/10.1007/978-3-319-65798-1_39)

Cai D, He X and Han J 2007 Spectral regression for efficient regularized subspace learning *2007 IEEE 11th Int. Conf. on Computer Vision* (IEEE) pp 1–8 (available at: http://ieeexplore.ieee.org/document/4408855/)

Calisto A, Galeano M, Bramanti A, Angileri F, Campobello G, Serrano S and Azzerboni B 2010 Analysis of intracranial pressure recordings: comparison of PCA and signal averaging based filtering methods and signal period estimation *2010 Annual Int. Conf. IEEE Engineering in Medicine and Biology Society, EMBC'10* pp 3638–41 (available at: www.scopus.com/inward/record.uri?eid=2-s2.0-84903863469&doi=10.1109%2fIEMBS.2010.5627420&partnerID=40&md5=be06ba6dab26c1dfae968251867c647e)

Calisto A, Galeano M, Serrano S, Calisto A and Azzerboni B 2013 A new approach for investigating intracranial pressure signal: filtering and morphological features extraction from continuous recording *IEEE Trans. Biomed. Eng.* **60** 830–7

Carney N *et al* 2017 Guidelines for the management of severe traumatic brain injury, fourth edition *Neurosurgery* **80** 6–15

Chang C-C and Lin C-J 2001 LIBSVM—a library for support vector machines (available at: www.csie.ntu.edu.tw/*cjlin/libsvm)

Chatterjee S and Hadi A S 1986 Influential observations, high leverage points, and outliers in linear regression *Stat. Sci.* **1** 379–93

Chaudhari H, Nalbalwar S L and Sheth R 2016 A review on intrensic mode function of EMD *2016 Int. Conf. on Electrical, Electronics, and Optimization Techniques (ICEEOT)* (IEEE) pp 2349–52 (available at: http://ieeexplore.ieee.org/document/7755114/)

Choi S, Adnane M, Lee G-J, Jang H, Jiang Z and Park H-K 2010 Development of ECG beat segmentation method by combining lowpass filter and irregular R–R interval checkup strategy *Expert Syst. Appl.* **37** 5208–18

Czosnyka M 2004 Monitoring and interpretation of intracranial pressure *J. Neurol. Neurosurg.* **75** 813–21

Dai H, Jia X, Pahren L, Lee J and Foreman B 2020 Intracranial pressure monitoring signals after traumatic brain injury: a narrative overview and conceptual data science framework *Front. Neurol.* **11** 959

Feng M, Loy L Y, Sim K, Phua C, Zhang F and Guan C 2012 Artifact correction with robust statistics for non-stationary intracranial pressure signal monitoring *Proc. -Int. Conf. on Pattern Recognition* pp 557–60 (available at: www.scopus.com/inward/record.uri?eid=2-s2.0-84874564964&partnerID=40&md5=fa11fe4ca71c1de6aac0494d139f1e54)

Feng M, Loy L Y, Zhang F and Guan C 2011 Artifact removal for intracranial pressure monitoring signals: a robust solution with signal decomposition *Proc. Annual Int. Conf. IEEE Engineering in Medicine and Biology Society, EMBS* pp 797–801 (available at: www.scopus.com/inward/record.uri?eid=2-s2.0-84861950416&doi=10.1109%2fIEMBS.2011.6090182&partnerID=40&md5=e8b1b6663da363985b6460f787679e1b)

Geurts P, Ernst D and Wehenkel L 2006 Extremely randomized trees *Mach. Learn.* **63** 3–42

Hawryluk G W J *et al* 2019 A management algorithm for patients with intracranial pressure monitoring: the Seattle International Severe Traumatic Brain Injury Consensus Conference (SIBICC) *Intensive Care Med.* **45** 1783–94

He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (IEEE) pp 770–8 (available at: http://ieeexplore.ieee.org/document/7780459/)

Holm S and Eide P K 2009 Impact of sampling rate for time domain analysis of continuous intracranial pressure (ICP) signals *Med. Eng. Phys.* **31** 601

Hu X, Xu P, Lee D J, Paul V and Bergsneider M 2008a Morphological changes of intracranial pressure pulses are correlated with acute dilatation of ventricles *Acta Neurochirurgica Supplements* vol 102, ed H-J Steiger (Springer Vienna) pp 131–6

Hu X, Xu P, Lee D J, Vespa P, Baldwin K and Bergsneider M 2008b An algorithm for extracting intracranial pressure latency relative to electrocardiogram R wave *Physiol. Meas.* **29** 459–71

Hu X, Xu P, Scalzo F, Vespa P and Bergsneider M 2009 Morphological clustering and analysis of continuous intracranial pressure *IEEE Trans. Biomed. Eng.* **56** 696–705

Huang N E, Shen Z, Long S R, Wu M C, Shih H H, Zheng Q, Yen N-C, Tung C C and Liu H H 1998 The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis *Proc. R. Soc.* A **454** 903–95

Huo S *et al* 2024 A supervised, externally validated machine learning model for artifact and drainage detection in high-resolution intracranial pressure monitoring data *J. Neurosurg.* **141** 509–17

Karim F, Majumdar S, Darabi H and Chen S 2018 LSTM fully convolutional networks for time series classification *IEEE Access* **6** 1662–9

Kawoos U, McCarron R, Auker C and Chavko M 2015 Advances in intracranial pressure monitoring and its significance in managing traumatic brain injury *Int. J. Mol. Sci.* **16** 28979–97

Kim H, Lee S-B, Son Y, Czosnyka M and Kim D-J 2018 Hemodynamic instability and cardiovascular events after traumatic brain injury predict outcome after artifact removal with deep belief network analysis *J. Neurosurg. Anesthesiol.* **30** 347–53
Lee S-B, Kim H, Kim Y-T, Zeiler F A, Smielewski P, Czosnyka M and Kim D-J 2020 Artifact removal from neurophysiological signals: impact on intracranial and arterial pressure monitoring in traumatic brain injury *J. Neurosurg.* **132** 1952–60
Levick J R 1991 Electrocardiography *An Introduction to Cardiovascular Physiology* (Elsevier) pp 45–54 (available at: https://linkinghub.elsevier.com/retrieve/pii/B978075061028550007X)
Li B N, Dong M C and Vai M I 2010 On an automatic delineator for arterial blood pressure waveforms *Biomed. Signal Process. Control* **5** 76–81
Martinez-Tejada I, Wilhjelm J E, Juhler M and Andresen M 2021 Empirical mode decomposition-based method for artefact removal in raw intracranial pressure signals *Intracranial Pressure and Neuromonitoring XVII* (*Acta Neurochirurgica Supplement*) vol 131, ed B Depreitere, G Meyfroidt and F Güiza (Springer International Publishing) pp 201–5 (available at: https://link.springer.com/10.1007/978-3-030-59436-7_39)
Mataczynski C, Kazimierska A, Uryga A, Burzynska M, Rusiecki A and Kasprowicz M 2022 End-to-end automatic morphological classification of intracranial pressure pulse waveforms using deep learning *IEEE J. Biomed. Health Inform.* **26** 494–504
Megjhani M *et al* 2019 An active learning framework for enhancing identification of non-artifactual intracranial pressure waveforms *Physiol. Meas.* **40** 015002
Megjhani M *et al* 2023 Automatic identification of intracranial pressure waveform during external ventricular drainage clamping: segmentation via wavelet analysis *Physiol. Meas.* **44** 064002
Miller J D, Stanek A and Langfitt T W 1972 Concepts of cerebral perfusion pressure and vascular compression during intracranial hypertension *Progress in Brain Research* vol 35 (Elsevier) pp 411–32 (available at: https://linkinghub.elsevier.com/retrieve/pii/S0079612308601028)
Mokri B 2001 The Monro–Kellie hypothesis: applications in CSF volume depletion *Neurology* **56** 1746–8
Munakomi S and Das J M 2024 Intracranial pressure monitoring *StatPearls* (StatPearls Publishing) (available at: www.ncbi.nlm.nih.gov/books/NBK542298/)
Nourallah B, Zeiler F A, Calviello L, Smielewski P, Czosnyka M and Menon D K 2018 Critical thresholds for intracranial pressure vary over time in non-craniectomised traumatic brain injury patients *Acta Neurochir.* **160** 1315–24
Page M J *et al* 2021 The PRISMA 2020 statement: an updated guideline for reporting systematic reviews *BMJ* **372** n71
Raj R *et al* 2022 Dynamic prediction of mortality after traumatic brain injury using a machine learning algorithm *npj Digit. Med.* **5** 96
Rajagopal A, Hamilton R B and Scalzo F 2016 Noise reduction in intracranial pressure signal using causal shape manifolds *Biomed. Signal Process. Control* **28** 19–26
Rozanek M, Skola J, Horakova L and Trukhan V 2022 Effect of artifacts upon the pressure reactivity index *Sci. Rep.* **12** 15131
Scalzo F, Asgari S, Kim S, Bergsneider M and Hu X 2010 Robust peak recognition in intracranial pressure signals *Biomed. Eng. OnLine* **9** 61
Scalzo F, Asgari S, Kim S, Bergsneider M and Hu X 2012 Bayesian tracking of intracranial pressure signal morphology *Artif. Intell. Med.* **54** 115–23
Scalzo F, Liebeskind D and Hu X 2013 Reducing false intracranial pressure alarms using morphological waveform features *IEEE Trans. Biomed. Eng.* **60** 235–9
Scalzo F, Xu P, Asgari S, Bergsneider M and Hu X 2009 Regression analysis for peak designation in pulsatile pressure signals *Med. Biol. Eng. Comput.* **47** 967–77
Shiogai Y, Stefanovska A and McClintock P V E 2010 Nonlinear dynamics of cardiovascular ageing *Phys. Rep.* **488** 51–110
Sorrentino E *et al* 2012 Critical thresholds for cerebrovascular reactivity after traumatic brain injury *Neurocrit. Care* **16** 258–66
Taco J, Foreman B, Kundu P and Lee J 2022 A novel technique for data quality improvement in human health PHM *PHM_CONF* vol 14 (available at: https://papers.phmsociety.org/index.php/phmconf/article/view/3177)
Tricco A C *et al* 2018 PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation *Ann. Intern. Med.* **169** 467–73
Tsai C-J, Lee C-I and Yang W-P 2008 A discretization algorithm based on class-attribute contingency coefficient *Inf. Sci.* **178** 714–31
Xu X, Xu S, Jin L and Song E 2011 Characteristic analysis of Otsu threshold and its applications *Pattern Recognit. Lett.* **32** 956–61
Yang L, Zhao M, Peng C, Hu X, Feng H and Ji Z 2012 Waveform descriptor for pulse onset detection of intracranial pressure signal *Med. Eng. Phys.* **34** 179–86
Zakrzewska A P, Placek M M, Czosnyka M, Kasprowicz M and Lang E W 2021 Intracranial pulse pressure waveform analysis using the higher harmonics centroid *Acta Neurochir.* **163** 3249–58
Zeiler F A *et al* 2019 Comparison of performance of different optimal cerebral perfusion pressure parameters for outcome prediction in adult traumatic brain injury: a Collaborative European Neurotrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) study *J. Neurotrauma* **36** 1505–17
Zeiler F A, Iturria-Medina Y, Thelin E P, Gomez A, Shankar J J, Ko J H, Figley C R, Wright G E B and Anderson C M 2021 Integrative neuroinformatics for precision prognostication and personalized therapeutics in moderate and severe traumatic brain injury *Front. Neurol.* **12** 729184
Zong W *et al* 2003 An open-source algorithm to detect onset of arterial blood pressure pulses *Computers in Cardiology, 2003* (IEEE) pp 259–62 (available at: http://ieeexplore.ieee.org/document/1291140/)