End-to-End Automatic Morphological Classification of Intracranial Pressure Pulse Waveforms Using Deep Learning

Cyprian Mataczyński, Agnieszka Kazimierska, Agnieszka Uryga, Małgorzata Burzyńska, Andrzej Rusiecki, and Magdalena Kasprowicz

***Abstract*—Objective. Mean intracranial pressure (ICP) is commonly used in the management of patients with intracranial pathologies. However, the shape of the ICP signal over a single cardiac cycle, called ICP pulse waveform, also contains information on the state of the craniospinal space. In this study we aimed to propose an end-to-end approach to classification of ICP waveforms and assess its potential clinical applicability. Methods. ICP pulse waveforms obtained from long-term ICP recordings of 50 neurointensive care unit (NICU) patients were manually classified into four classes ranging from normal to pathological. An additional class was introduced to simultaneously identify artifacts. Several deep learning models and data representations were evaluated. An independent testing dataset was used to assess the performance of final models. Occurrence of different waveform types was compared with the patients' clinical outcome. Results. Residual Neural Network using 1-D ICP signal as input was identified as the best performing model with accuracy of 93% in the validation and 82% in the testing dataset. Patients with unfavorable outcome exhibited significantly lower incidence of normal waveforms compared to the favorable outcome group even at ICP levels below 20 mm Hg (median [first-third quartile]: 9 [1–36]% vs. 63 [52–88] %, $p = 0.002$). Conclusions. Results of this study confirm the possibility of analyzing ICP pulse waveform morphology in long-term recordings of NICU patients. Proposed approach could potentially be used to provide additional information on the state of patients with intracranial pathologies beyond mean ICP.**

***Index Terms*—Deep neural networks, intracranial pressure, intensive care unit.**

Manuscript received February 11, 2021; revised April 22, 2021 and June 6, 2021; accepted June 6, 2021. Date of publication June 11, 2021; date of current version February 4, 2022. This work was supported by the National Science Centre, Poland under Grant UMO 2019/35/B/ST7/00500. Agnieszka Uryga receives a scholarship from the Foundation for Polish Science. *(Cyprian Mataczyński and Agnieszka Kazimierska contributed equally to this work.) (Corresponding author: Cyprian Mataczynski.)*

Cyprian Mataczyński and Andrzej Rusiecki are with the Department of Computer Engineering, Faculty of Electronics, Wroclaw University of Science and Technology, 50-370 Wroclaw, Poland (e-mail: cyprian.mataczynski@pwr.edu.pl; andrzej.rusiecki@pwr.edu.pl).

Agnieszka Kazimierska, Agnieszka Uryga, and Magdalena Kasprowicz are with the Department of Biomedical Engineering, Faculty of Fundamental Problems of Technology, Wroclaw University of Science and Technology, 50-370 Wroclaw, Poland (e-mail: agnieszka.kazimierska@pwr.edu.pl; agnieszka.uryga@pwr.edu.pl; magdalena.kasprowicz@pwr.edu.pl).

Małgorzata Burzyńska is with the Department of Anaesthesiology and Intensive Care, Wroclaw Medical University, 50-367 Wroclaw, Poland (e-mail: malgorzata.burzynska@umed.wroc.pl).

Digital Object Identifier 10.1109/JBHI.2021.3088629

I. Introduction

INTRACRANIAL pressure (ICP) is frequently monitored in patients with brain pathologies as elevated ICP puts the patient at risk of cerebral ischemia or herniation of structures within the cranial vault. However, the clinical state of the patient cannot be fully characterized by mean ICP alone as the changes in intracranial volume which influence ICP can be buffered to a certain degree [1]. The ability of the craniospinal system to tolerate or compensate for volume increases is quantified by a parameter called 'brain compliance' [2]. As long as the compensatory mechanisms for adapting to increased volume are intact and the compliance of the system is normal, small increases in intracranial volume result in small increases in ICP. When brain compliance is decreased and the compensatory mechanisms are exhausted, small increases in volume lead to disproportionately large increases in ICP. The pressure–volume curve, i.e., the exponential relationship between pressure and volume in the intracranial space, has long been regarded as a potential source of useful information on the state of the craniospinal system [3]. Despite promising results published on that subject, direct measurement of compliance has, however, proven difficult to implement in clinical practice on a larger scale [2].

On the other hand, it has long been known that the ICP signal contains much more information than can be captured by simple mean value [1]. ICP pulse morphology, which refers to the shape of the pressure signal over a single cardiac cycle, is believed to contain indirect information about brain compliance [4]. Under normal conditions, the ICP pulse is characterized by three distinct subpeaks, denoted P1, P2, and P3, arranged in a saw-tooth pattern. As brain compliance decreases and ICP increases, the subpeaks gradually become less pronounced, eventually resulting in a 'rounded' or sinusoidal wave [5]. Rounding of the pulse wave is often observed at elevated ICP levels; however, the rate of changes in pulse morphology varies both across patients and over time [6]. As the patients exhibit decreased compliance at different pressure levels, the analysis of the shape of ICP pulse waveform is key to characterizing the state of the intracranial space.

Due to high variability of the ICP pulse waveform, identification of characteristic peaks is a highly complex task, which in turn requires highly complex algorithms. Several attempts have been made to automatically analyze the changes in ICP pulse waveform based on detection of peaks and notches [7]–[11]. However, in addition to high complexity, limiting their understanding and acceptance in the medical community, those methods often rely on averaged pulses or fail in case of pathologically rounded signals. In recent years, deep learning rose to prominence in the field of biomedical signal processing, including pattern recognition tasks. Deep learning models have been successfully applied to remove artifacts from the ICP signal [12] and to detect ICP elevation [13] but as far as we know, a deep learning approach has never been used for the task of ICP pulse morphology classification not only in terms of valid vs. invalid pulses, but also with separate categories reflecting the changes in the configuration and visibility of characteristic peaks, i.e., the progression from normal, triphasic waveform to a rounded, sinusoidal shape.

Therefore, in this work we aimed to develop an automated method for morphological classification of different shapes of ICP pulse waveforms using deep neural networks. An end-to-end approach was proposed for the purpose of analyzing long-term recordings collected from neurointensive care unit (NICU) patients with intracranial pathologies, comprising stages responsible for single pulse detection, artifact detection, and classification of non-artifactual waveforms. Deep learning models were developed, evaluated, and compared in terms of their accuracy in identifying four types of ICP pulses as well as artifacts in the ICP signal. Additionally, we investigated the link between the occurrence of different morphological types of ICP pulse waveform and treatment outcomes of the patients to assess the potential clinical usefulness of the proposed approach.

II. Related Work

Nucci *et al.* [14] introduced the classification criteria for ICP pulse waveform analysis based on overall shape of the signal, reflecting the changes in the configuration of characteristic peaks of the ICP pulse. The authors used a small neural network with coefficients of radial basis function (RBF) kernel approximation of the ICP signal serving as network input, and reported accuracy of 88.3% in data collected in normal pressure hydrocephalus patients during infusion studies. We attempted to replicate this method but were unable to choose universal RBF midpoints for our dataset that would allow for representation of artifacts as well as valid ICP pulses. We hypothesize that differences in data acquisition and general aim of the system make it unsuitable for our study.

A similar approach to analysis of infusion study recordings was also previously proposed by Elixmann *et al.* [9] who identified five distinct waveform patterns. However, the study used a decision algorithm based on results of peak and notch detection, namely the number, relative height, and distance between subpeaks, to distinguish between waveform types. Various methods have also been presented for the task of binary classification of ICP waveforms into valid pulses versus artifacts, and multi-class classification as investigated in this study could theoretically be considered an extension of the previously proposed binary classification approaches. However, previous studies either also rely on peak identification [8], [10], [15], potentially producing higher number of false positives and false negatives related to difficulty in peak annotation in irregular but otherwise valid waveforms, are not a deep approach [16], or include a number of preprocessing steps that may significantly extend the computation time [12]. Consequently, in this study we propose a new approach taking into account recent developments in the field of deep learning. As we are not aware of any other studies that aimed to classify various pulse waveform patterns (i.e., beyond valid versus artifactual pulses) in long-term recordings obtained from NICU patients using deep neural networks, we compare our proposed model with the approach shown in [12], which was selected as the work with the most similar aim and methodology. However, it should be noted that [12] introduced a procedure for identification of artifacts in physiological signals, and for the purpose of comparison we used a modified version of that model allowing for multi-class instead of binary classification. Furthermore, we extended our analysis by investigating the possible relationship between the occurrence of different shapes of pulse ICP waves and clinical outcome in NICU patients.

III. Methodology

A. Problem Formulation

The aim of this study was to produce an end-to-end approach for detecting and annotating ICP pulse waveforms in long (i.e., lasting upwards of several hours) signals. Specifically, this means an algorithm capable of taking as input a raw, unprocessed recording and producing final results without manual preprocessing or human intervention during computations. The algorithm is therefore comprised of two steps: division of full signals into short pulse waveforms followed by classification of said pulse waveforms into one of five morphological classes: T1 – normal, T2 – potentially pathological, T3 – likely pathological, T4 – pathological, or A+E – artifacts and measurement errors (Figure 1). The classification expands upon criteria proposed previously by Nucci *et al.* [14] for analysis of infusion studies. First four classes reflect the changing relative height and visibility of characteristic subpeaks P1, P2, and P3 of the ICP pulse waveform, with class T1 (normal) representing the saw-tooth shape associated with normal compliance and class T4 (pathological) representing pathologically rounded pulse with unrecognizable peaks. The fifth class (A+E) was introduced to separate artifactual pulses related to loss of signal quality or failure to accurately identify pulse onset points without the need for introducing an additional step of artifact detection.

The problem can be formulated as obtaining a mapping m from full-length ICP signal to a set of tuples containing location and detected class of each of the pulses:

$$m : \mathbb{R}^L \rightarrow (Position, Class)^N$$

where $L \in \mathbb{N}$ is the number of data points in full-length ICP signal, $Position \in \{1, ..L\}$ is the number of data point in the full signal that marks the beginning of the pulse waveform, and

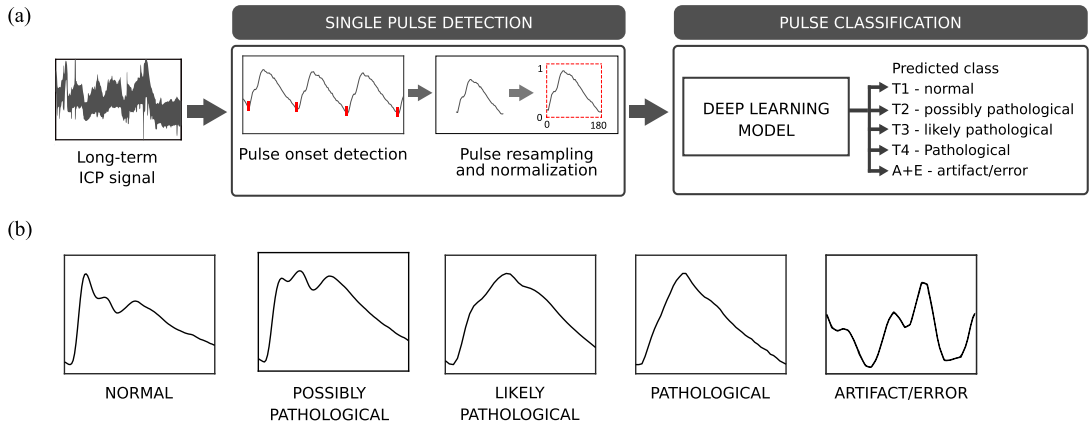

Fig. 1. (a) Overview of proposed approach to pulse waveform classification. (b) Illustrative examples of pulse waveform shapes in each of the five morphological classes.

TABLE I
PATIENT CHARACTERISTICS IN TOTAL GROUP OF 50 PATIENTS. DATA ARE PRESENTED AS N (% OF TOTAL GROUP) OR AS MEDIAN [FIRST-THIRD QUARTILE] UNLESS OTHERWISE INDICATED

Age	Mean: 45 years, Range: 20-85 years
Sex	Male: 32 (64%), Female: 18 (36%)
TBI	39 (78%)
aSAH	11 (22%)
GCS	7 [5-9]
GOS at discharge	unfavourable (I-III): 40 (80%), favourable (IV-V): 10 (20%)
GOS after 3 months	unfavourable (I-III): 26 (52%), favourable (IV-V): 24 (48%)

$Class \in \{T1, T2, T3, T4, A+E\}$ is the morphological class to which that pulse waveform belongs. The end of each pulse is also the beginning of a subsequent pulse, thus only one positional argument is required to unambiguously mark the whole pulse waveform location.

TABLE II
DETAILED PATIENT CHARACTERISTICS OF TBI PATIENTS (N=39). DATA ARE PRESENTED AS N (% OF TOTAL GROUP) OR AS MEDIAN [FIRST–THIRD QUARTILE]

ICU length of stay [days]	36 [16-63]
Marshall scale	3 [2-5]
Rotterdam scale	4 [3-4]
GOS at discharge	3 [2-3]
GOS after 3 months	3 [1-4]
30-days mortality	4 (10%)
ABP systolic [mm Hg]	105 [101-130]
ABP diastolic [mm Hg]	90 [70-101]
Arterial hypotension	17 (44%)
Pupils	anisocoria: 16 (41%), nonreactive: 13 (33%)
Hematoma	subdural: 13 (33%), epidural: 5 (13%), cerebral: 12 (31%)
Edema/cerebral contusion	23 (59%)
Axonal trauma	5 (13%)
Isolated head trauma	13 (33%)

B. Data Collection

Data from 50 patients admitted to the Neurointensive Care Unit (NICU) of University Hospital in Wroclaw, Poland between 2014 and 2019 were chosen for retrospective analysis in this study. The patients were selected out of all patients admitted to the NICU during this period on the basis of availability and acceptable quality of ICP recordings. The study was conducted with approval from the Bioethics committee at the Wroclaw Medical University, Poland (approvals no KB–624/2014 and KB–134/2014). All patients were adults over 18 years of age.

Out of the entire group of 50 patients (see Table I), 39 patients suffered from traumatic brain injury (TBI) and 11 had confirmed aneurysmal subarachnoid haemorrhage (aSAH). TBI and aSAH are two distinct clinical entities with different pathophysiology. However, both conditions are associated with changes in mean ICP and the shape of ICP pulse waveform due to disturbances in the intracranial volume equilibrium, and are subject to the same method of assessing the patient's outcome. As a result, previous studies on ICP pulse morphology combined various groups to increase the number of available recordings [7], [8]. In this study, the aSAH patients were selected in order to introduce a second, independent set of data that would allow for evaluation of final model performance.

TBI patients were treated according to the American Brain Trauma Foundation guidelines applicable at the time of admission [17], [18]. Patients with aSAH were treated according to guidelines from the American Heart Association/American Stroke Association [19]. The patients were classified in the NICU using the Glasgow Coma Scale (GCS). Outcome was assessed at discharge from the hospital and after three months using the Glascow Outcome Scale (GOS), with scores IV-V representing favourable outcome and scores I-III unfavourable outcome. The patient cohort was homogenous with regard to severity of the injury and treatment protocol. Table II presents detailed characteristics of the TBI group used in the assessment

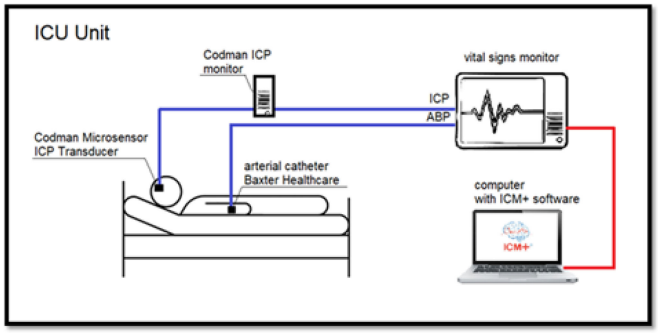

Fig. 2. Measurement setup. Intracranial pressure (ICP) was measured invasively using intraparenchymal probes. Arterial blood pressure (ABP) was measured invasively in the radial or femoral artery using standard monitoring kits. The blood pressure transducer was calibrated at the phlebostatic axis. Signals were monitored using standard bedside vital signs monitors and recorded on a portable computer using ICM+ software with custom-written measurement profile.

of the relationship between classification results and clinical outcome.

ICP was measured invasively using intraparenchymal probes (Codman MicroSensor ICP Transducer, Codman & Shurtleff, Randolph, MA, USA) inserted into the frontal cortex. Although the parenchymal ICP sensor is considered more invasive than the ventricular sensor due to implantation into the brain tissue, it was found to be more accurate, as the calibration and zeroing process only needs to be done once before insertion and the readings are not dependent on the patient's position in relation to the transducer. Arterial blood pressure (ABP) was measured in the radial or femoral artery using standard monitoring kits (Baxter Healthcare, CardioVascular Group, Irvine, CA, USA). The signals were recorded continuously and synchronously (Figure 2) using ICM+ software (Cambridge Enterprise Ltd, Cambridge, U.K.) with sampling frequency ranging from 50 Hz to 300 Hz. The signals were resampled to 50 Hz prior to further analyses in order to reduce computation time, but taking into account the minimum sampling frequency requirements reported in previous experimental studies [20].

The signals were monitored and recorded continuously starting on day 1 or day 2 after admission to the hospital, depending on the date of surgery, and in most cases the day of admission was the same date the injury occurred. Patients were monitored throughout their ICU stay with average recording length of 5 $\pm$ 3 days. In each case the decision to remove the sensor was made by the neurosurgeon and/or intensivist based on medical indications, particularly low mean value and stability of the ICP signal and progressive improvement of the patient's conditions.

C. Single Pulse Detection

A modified Scholkmann algorithm proposed by Bishop and Ercole [21] for analysis of neuroscience data was used for the purpose of single pulse detection. Pulse detection was performed in full long-term signals low-pass filtered with a cutoff frequency of 10 Hz. Individual ICP pulse onset points were defined as local minima preceding the first peak of the waveform occurring at intervals corresponding to the length of the cardiac cycle (i.e., around 1 s).

D. Classification Datasets

The full group of 50 patients was divided between the training/validation and test datasets, with 39 TBI patients assigned to the training and validation datasets and 11 aSAH patients assigned to the test dataset.

In the training and validation datasets, full long-term ICP signals from TBI patients were divided into pulse waveforms (see III-C), and a total of 21 390 pulses were randomly selected from all recordings. In addition to ICP, corresponding ABP pulse waves were selected to aid in manual classification as it has been previously shown that the systolic part of an ABP pulse correlates with the position of peak P1 in ICP pulse waveform and the slopes of ABP and ICP become increasingly divergent with higher waveform type [22]. Each example was then annotated by an expert researcher.

As pulses from the same patient within one waveform class are largely similar and could therefore influence generalization, in order to mitigate the correlation between examples the patients in the training set were selected in such a way that the group did not intersect with the validation dataset. A simple binary genetic algorithm was set up to divide the patients into two sets where one includes 2/3 of the total number of examples in each class and the other includes the remaining pulses. This created a split of the data into the training set consisting of 14 578 pulses and the validation set consisting of 6812 pulses.

The testing dataset was in turn extracted from 11 aSAH patients. Full ICP signals were again divided into pulse waveforms, and 650 pulses were randomly selected from all recordings. The examples were annotated by a panel of three expert researchers using ICP and corresponding ABP pulses. In cases of ambiguous waveform type (particularly signals exhibiting features of two adjacent classes) or disagreement between the experts' assessment, an additional label, the 'possible type', was added. This label was later used in an alternative scoring method (see III-F) and to test multi-label classification described in Appendix II. Inter-rater agreement between the three experts' primary type annotations was tested using Fleiss' kappa test [23] with significance level of 0.05. The reference classification provided by the experts showed statistically significant substantial agreement κ=0.700 (95% CI, 0.672 to 0.728), $p < 0.001$.

As shown in the class distribution in Figure 3, the resulting datasets were not balanced, making the classification task more challenging for smaller models. This issue is discussed in Appendix II.

E. Data Representations

Cubic resampling was used to unify the length of pulses to 180 samples. The pulses were then scaled to an interval between 0 and 1. This step was introduced to test classification based purely on the shape of the waveform and to remove the influence of ICP pulse amplitude which is strongly correlated with mean ICP level and may therefore vary across patients despite comparable pulse morphology. Other types of data representations, such as Fourier

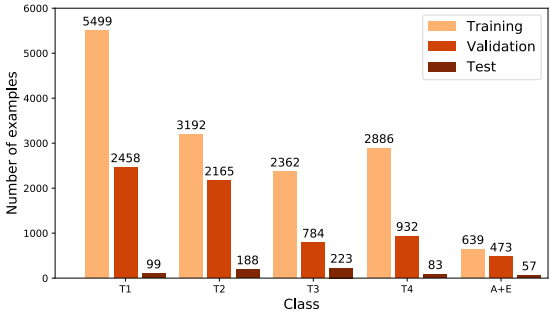

Fig. 3. Number of examples in each dataset.

transform coefficients or time-frequency representations, were investigated but did not result in improved classification accuracy (see Appendix II); therefore, only 1-D vectors of normalized signal samples were used as input. We did, however, investigate the generalization of the models using univariate ICP time series versus multivariate input including both ICP and ABP, as ABP was used in the process of manual annotation.

F. Classification Models and Evaluation

A number of different models were investigated in this study; however, this paper focuses on models identified as best performing: 1-D Residual Neural Network (ResNet) [24], its two variants: dual channel network with joint ICP and ABP signal input or Siamese feature extractors, and Long-Short Term Memory Fully Convolutional Network (LSTM-FCN) [25]. The fully connected neural network was selected as the baseline for our results. Additionally, the approach introduced in [12] was adapted for multi-class classification for the purpose of comparison with proposed models and trained in the same way as the other models.

Convolutional neural networks (CNN) extract information based not only on a single sample but also on the sample's neighborhood, which allows them to easily extract morphological features and therefore makes them a perfect fit for the task of morphological classification. Additionally, in this case, due to the relatively short duration of processed signals (mostly less than 1 s in length), the networks are not required to overcome the challenge of modelling the long term dependencies. ResNets are deep convolutional models that use residual connections between layers for more stable error propagation. The hyperparameters were chosen through the empirical choice method across many conducted experiments with each of the proposed models. The architecture of residual models used in this study is presented in Figure 4, and the hyperparameters of the models are shown in Appendix I.

In addition to changing the number of channels in the first layer of the network, a Siamese architecture with residual feature extractors was tested (however, the latter was not trained on contrastive loss function as in the original paper [26], but through standard procedures). This approach was used to emulate the behaviour of manual annotators who used features of both ICP and ABP signals in the decision making process.

The LSTM-FCN models were employed to test the possibility that long-term dependencies are more relevant to the classification problem. The networks are composed of two different feature extractors, one with a CNN-based architecture, and the second consisting of LSTM cells. Based on the concatenated output of the two arms, an embedding is created and used by the fully connected network to make the final prediction. The main difference from the residual network lies in the LSTM layer which allows the embedding to consider the whole signal, as the memory cell of the LSTM is affected by all previous observations. The structure and hyperparameters of LSTM-FCN models used in this study are presented in Appendix I.

The model proposed in [12] is a combination of a stacked convolutional autoencoder (SCAE) and a CNN that takes 2-D images generated from 1-D signals as input. The first part converts segmented pulse waveforms into representative images. In the original paper, the second part classifies the input as either artifactual or valid. In our reproduction, the SCAE part was preserved while the last layer of the CNN classifier was modified to produce multinomial instead of binary classification.

An universal training loop using Python's PyTorch [27] package was created for all the models to ensure fair comparison of their scoring (see Appendix I). Additionally, taking into account that discrete classification employed in this study does not fully capture the gradual changes in the shape of the ICP pulse waveform cause by physiological and pathophysiological processes, two approaches to evaluation of the models' performance were investigated. First, the standard single-label accuracy score, denoted 'strict accuracy', and the second, denoted 'best accuracy', where the prediction is considered correct in the same cases as strict accuracy but also if it matches the class marked as 'possible' in manual annotations. The second scoring was proposed to test the models' performance in cases where it is acceptable to classify a waveform as belonging to more than one type.

G. Analysis of the Relationship Between Waveform Type and Outcome

Classification results were obtained from long-term recordings of 35 TBI patients. 4 patients from the original TBI group were excluded due to gaps in their recordings that did not have any impact on single ICP pulse classification but could influence the results of the analysis of the relationship between occurrence of ICP waveform types in long-term recordings and the patients' outcome. Each ICP pulse was assigned a waveform type based on classification results and mean ICP calculated as the average over the whole pulse. Classification results were compared with outcome assessed by GOS score after three months. GOS at discharge from the hospital was not used due to significant inequality of favourable versus unfavourable outcome groups that would have prevented reliable statistical analysis. Occurrence of different waveform types was calculated as the percentage of pulses classified as types T1-T4 in the recording. Pulses classified as artifacts (class A+E) were treated as noise and excluded from analysis.

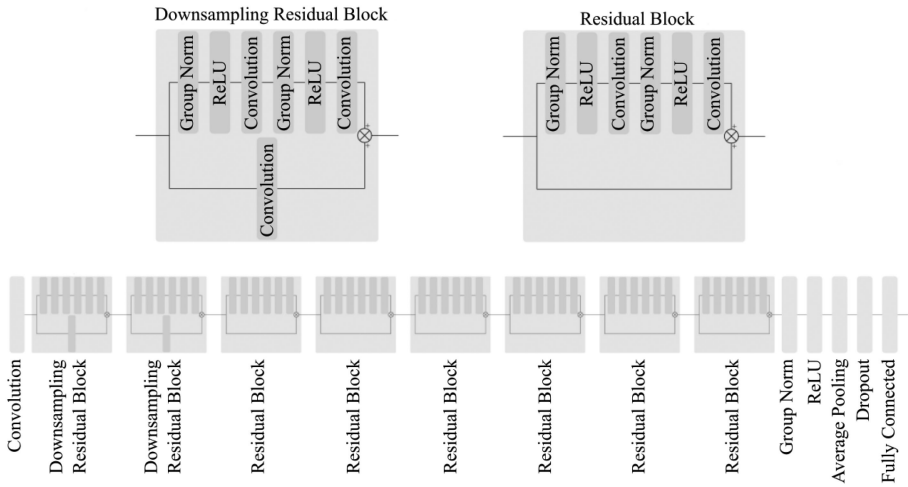

Fig. 4. Model of used Residual Network. The hyperparameters of the model are presented in Appendix I. The convolution nodes are one-dimensional convolution. The initial downsampling residual blocks are used to reduce the size of the processed tensors.

TABLE III
TRAINING RESULTS OF SELECTED MODELS IN DIFFERENT DATASETS. ALL OF THE MODELS SHOWN HERE USED SINGLE LABEL CLASSIFICATION (SEE APPENDIX II)

	N = 650		N = 6812		N = 14578	
	Test		Validation		Train	
Model	Best Accuracy [%]	Strict Accuracy [%]	Best Accuracy [%]	Strict Accuracy [%]	Best Accuracy [%]	Strict Accuracy [%]
Single channel ResNet	**86.00**	**81.85**	**95.17**	**92.95**	98.45	97.78
Siamese ResNet	83.54	80.62	94.10	92.18	98.15	97.24
Dual channel ResNet	81.69	78.92	94.85	92.56	**99.14**	**98.74**
Single channel LSTM-FCN	80.62	77.69	91.18	89.09	95.51	94.17
Reproduction of the approach from [12]	68.00	64.62	81.55	81.44	98.65	98.48
Single channel fully connected (baseline)	72.46	69.38	81.25	74.35	92.59	91.38

In order to separately analyze areas of the recordings where ICP falls within normal or increased range, the range of ICP values was subsequently divided into areas with ICP $<=$ 20 mm Hg and ICP $>$ 20 mm Hg based on moving average (window length: 5 minutes, window shift: 30 s) of single pulse mean ICP values.

Normality of all parameters used in the analyses was tested using the Shapiro-Wilk test. Upon rejection of the normality hypothesis for most of analyzed variables, non-parametric statistical tests were used to assess the difference between groups where applicable. Significance level of 0.05 was used in all analyses. All group-averaged results are presented as median [first-third quartile].

H. Analysis of the Potential for Real-Time Processing

Finally, in order to assess if proposed end-to-end approach could be realistically used to process ICP signals in real-time, an additional experiment was performed using a single illustrative ICP recording and the best performing 1-D ResNet model. In order to simulate real-life continuous measurement, but taking into account that the classification stage requires individual pulses to be detected first, the recording was divided into 10-seconds-long chunks. Each chunk was then processed with the hardware specification described in Appendix I. and the computation times for both single pulse detection and classification step were recorded.

IV. RESULTS

A. Classification Results

Table III shows classification accuracy of selected best performing models compared to baseline accuracy of the fully connected network. The results shown are the best results obtained for each model after empirically choosing hyperparameters through a series of experiments. All three variants of the ResNet model as well as the LSTM-FCN model outperformed the fully connected network, with the highest accuracy registered for the single channel ResNet using only the ICP signal as input. The addition of the ABP signal in the dual channel and Siamese ResNet did not improve classification accuracy in the validation and test datasets, although dual channel ResNet achieved the highest accuracy in the training dataset. As discussed in Appendix II, the single channel ResNet model also retained the highest accuracy in experiments including addition of artificially created A+E examples, weighing of the gradients, or unsupervised pretraining. The modified version of the model proposed in [12] achieved comparable accuracy in the training dataset, but performed considerably worse than the ResNet models in the other two datasets.

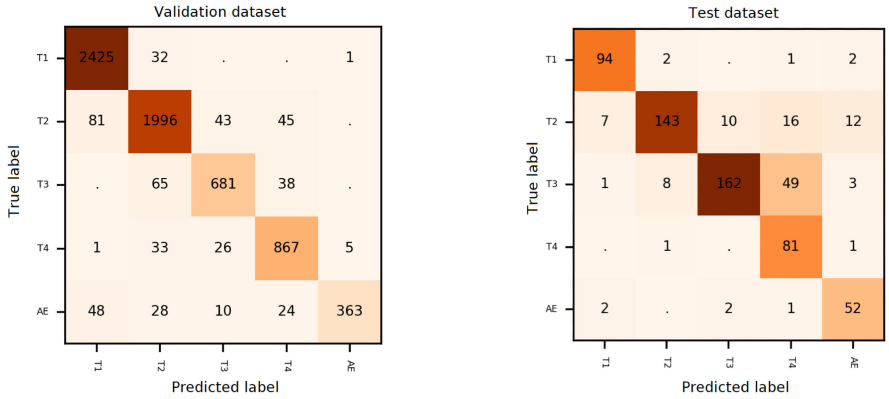

Fig. 5. Confusion matrices for the best performing ResNet model. The number in each tile shows how many examples with given true label were classified as given predicted label.

TABLE IV
DETAILED CLASSIFICATION SCORES FOR THE BEST-PERFORMING RESNET MODEL

Class	T1	T2	T3	T4	A+E
Validation dataset $N = 6812$					
Precision	0.95	0.94	0.75	0.76	0.73
Recall	0.95	0.72	0.90	0.93	0.99
F1-score	0.95	0.82	0.82	0.84	0.84
Specificity	0.97	0.88	0.99	0.99	0.99
Test dataset $N = 650$					
Precision	0.95	0.76	0.73	0.98	0.91
Recall	0.90	0.93	0.93	0.55	0.74
F1-score	0.93	0.84	0.82	0.70	0.82
Specificity	0.98	0.98	0.97	0.88	0.97

The use of best accuracy scoring instead of standard strict accuracy showed improved classification accuracy of all models (e.g., from 82% to 86% in the test dataset in case of the best performing single channel ResNet). However, attempts at multi-label classification did not boost the models' performance (see Appendix II). The confusion matrices for the single channel ResNet model (Figure 5) show that while in the validation dataset the main problem was presented by the artifact class characterized by the lowest number of examples, in the test dataset the errors primarily concern likely pathological and pathological pulses (types T3 and T4), and detailed classification scores (Table IV) in the latter dataset show that type T4, although characterized by high precision, showed markedly lower recall compared to other types.

B. Relationship Between Waveform Type and Outcome

Mean ICP and ICP waveform type. Figure 6 shows average ICP in each waveform class based on data from patients separated into favourable and unfavourable outcome groups. In both cases, mean ICP increased with progressively more pathological waveform type. There were no statistically significant differences in mean ICP between patients with favourable and unfavourable outcome. However, while mean ICP was slightly lower in the favourable outcome group in waveform type T1, it grew to higher values than in the unfavourable outcome group in more pathological types T3 and T4.

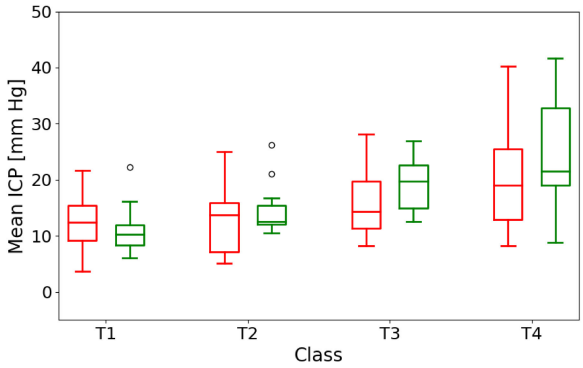

Fig. 6. Mean ICP in each ICP waveform type for unfavourable (red, left-hand side boxes) vs. favourable (green, right-hand side boxes) outcome groups. Central box line: median, box edges: first-third quartile, whiskers: most extreme data points not including outliers (circle signs).

TABLE V
GROUP-AVERAGED OCCURRENCE OF ICP WAVEFORM TYPES FOR UNFAVOURABLE VS. FAVOURABLE OUTCOME GROUPS. RESULTS ARE PRESENTED AS MEDIAN [FIRST-THIRD QUARTILE] WITH P-VALUE OF MANN-WHITNEY U TEST. NS - RESULTS NOT STATISTICALLY SIGNIFICANT

ICP <= 20 mm Hg			
Class	Unfavourable outcome	Favourable outcome	p-value
T1 [%]	9 [1-36]	63 [52-88]	0.002
T2 [%]	27 [6-49]	25 [11-36]	ns
T3 [%]	5 [0-41]	5 [0-8]	ns
T4 [%]	1 [0-38]	0 [0-0]	0.032
ICP > 20 mm Hg			
Class	Unfavourable outcome	Favourable outcome	p-value
T1 [%]	2 [0-9]	34 [10-39]	< 0.001
T2 [%]	11 [2-47]	43 [31-54]	ns
T3 [%]	29 [0-40]	16 [2-26]	ns
T4 [%]	13 [1-46]	2 [0-7]	0.028

Occurrence of different ICP waveform types. Occurrence of different waveform types in the favourable and unfavourable outcome groups is presented in Table V. Significantly higher

incidence of normal waveforms (type T1) as well as lower incidence of pathological waveforms (type T4) was observed in patients with favourable outcome for ICP <= 20 mm Hg. An inverse relationship was found for ICP > 20 mm Hg, with markedly increased occurrence of pathological waveforms in the unfavourable outcome group.

C. Potential for Real-Time Processing

In a subset of 1000 chunks of length equal 10 seconds extracted from an illustrative ICP recording the average processing time was 0.027 [0.026-0.029] seconds per pulse for full analysis, with 0.020 [0.019-0.021] seconds per pulse for single pulse detection and 0.0075 [0.0070-0.0080] seconds per pulse for waveform shape classification.

V. Discussion

In this work we investigated the feasibility of using deep neural networks for identification of different shapes of ICP pulse waveform as well as artifacts in recordings obtained from patients with intracranial pathologies. Given the small number of previous studies on the subject of morphological classification of ICP pulse waveforms, none of which–to the best of our knowledge–used deep learning methods, we tested different approaches to solving this task, including different models, data representations, and training and evaluation methods. The single channel ResNet model using 1-D vector of ICP signal samples as input was identified as the best performing model, achieving best classification accuracy of 86% (82% strict accuracy) in an independent test dataset which suggests good generalization ability. This shows the potential of classifying ICP pulse waveform with relatively high accuracy. Additionally, proposed approach is robust to recordings of different length than the ones used in this study as ICP waveform classification is performed on a single-pulse level and does not require information about the position of each pulse in the full recording. Assessment of changes in pulse morphology over time (i.e., changes in the occurrence of different waveform classes) is secondary to classification of individual pulses and could therefore be performed on both short and long-term measurements.

The fact that neither different models nor alternative data representations resulted in further improvement in accuracy could be to a certain degree explained by the complexity of the task. Despite the use of four distinct non-artifactual classes of ICP pulse waveforms in this study, it should be noted that the changes in brain compliance represented by changes in the shape of the waveform are continuous rather than discrete. As a result, some of the pulses may exhibit the features of more than one class or fall into the 'gray area' between two classes. This phenomenon is visible in the test dataset confusion matrix of the single channel ResNet model where the majority of classification errors occurred between two adjacent classes, and very few were observed between types that are far apart. While this reduces the accuracy score of the models, from the viewpoint of clinical utility this type of mistake is less severe than erroneous classification of normal pulses as pathological or vice versa. Additionally, the increase in best accuracy score compared to strict accuracy shows that a number or errors stems from ambiguity of the waveform shape that has also been noted by the experts performing manual annotations. Taking into account the accuracy recorded in the validation set (best accuracy of over 95% for the single channel ResNet model), we hypothesize that classification accuracy of the model could be further improved by providing a bigger and more balanced training dataset. It should be noted that the test dataset was not only independent from the training and validation datasets, but included data from a different distribution. Whereas all datasets were selected to provide examples of different shapes of ICP pulse waveforms, the test dataset was collected in a separate group of patients for the purpose of assessing the models' real-life applicability. Therefore, despite the decrease in accuracy between the validation and test datasets, the models' performance can still be considered acceptable as the model correctly classifies a large percentage of pulses from a completely different data distribution.

Interestingly, while manual classification of ICP pulses by expert researchers was based on the ICP signal as well as corresponding ABP signal, taking into account previous studies which showed the correlation between the systolic part of ABP and the P1 portion of ICP pulse waveform [22], inclusion of the ABP signal did not improve the performance of deep learning models, and single channel ResNet outperformed both its dual channel and Siamese counterpart. As indicated by comparison of results achieved in the training and test datasets, the use of an additional signal only made the models prone to overfitting. This, however, is an advantage, as a simpler model means that the computation time is lower, and the classification could theoretically be performed in a continuous and real-time manner.

The potential for real-time processing is further supported by the use of an end-to-end pipeline including single pulse detection and artifact exclusion steps in addition to pulse waveform classification. ICP recordings are subject to a variety of disturbances that manifest, for instance, as very short spikes in mean value or waveform deformations. Those disturbances do not carry clinically useful information but instead are related to the purely technical aspect of collecting the signal. It has been shown that artifacts are the cause of a significant number of false positive alarms in the intensive care units [28] and that artifact removal improves the performance of other measures used in TBI management [12]. Various methods of reducing the impact of artifacts on ICP signal analysis have been proposed so far [29], [30], also using deep learning models [12]. Our approach, instead of introducing further algorithms for artifact detection, treats it as part of the classification stage, with artifacts such as noise, sensor calibration signal, or distorted waveforms regarded as an additional class. This also allows for mitigation of false positives resulting from errors in the single pulse detection step. Furthermore, within the proposed pipeline average processing time of a single 10-seconds-long fragment of the ICP signal is over 10 times shorter than the collection time. This shows that the algorithm is capable of effectively working in real-time if adapted to process the full signal in smaller chunks as soon as they are recorded. Given the large difference between computation and collection time, the windows could also be

overlapping, removing the potential problem of losing data from pulses located at the edges of the window that would be discarded by the single pulse detection step.

To investigate the potential clinical applicability of proposed approach we obtained classification results in a group of patients with intracranial pathologies using the best performing single channel ResNet model. In accordance with the relationship described by the pressure–volume curve, higher waveform types were associated with progressively higher mean ICP. While within each waveform class we did not register statistically significant differences in mean ICP between the favourable and unfavourable outcome groups, in the favourable outcome group mean ICP rose more steeply. Given the assumption that higher waveform types are associated with decreasing compliance, this suggests that reduction of the compensatory reserve occurs more slowly in patients with favourable outcome. On the other hand, in patients with unfavourable outcome larger variance of results in each waveform type suggests weaker dependence on mean ICP.

The latter is supported by further analysis of results divided between two ICP levels. The threshold used in this study to separate low and high ICP ranges, 20 mm Hg, reflects the value adopted in the clinical setting as the level above which therapeutic interventions should be introduced [31]. In our study the difference between outcome groups was already pronounced at ICP levels below the threshold for clinical intervention, where the unfavourable outcome group exhibited significantly lower incidence of type 1 waveforms in favour of higher waveform types suggesting diminished compensatory reserve. Over the threshold both groups moved towards higher waveform types, although in terms of average number of pulses of given type the change was more pronounced in patients with unfavourable outcome, with markedly increased incidence of pathological waveforms.

The occurrence of pathologically changed waveforms at lower ICP should be treated as a warning sign, indicating that despite normal levels of pressure the compensatory reserve is already reduced and the system may not be able to tolerate further increases in volume. Evaluation of intracranial compliance based on analysis and interpretation of ICP pulse morphology would also represent a method free of any additional risks to the patient as it is not additionally invasive. In most clinical settings, ICP is monitored continuously (as recommended by The American Brain Trauma Foundation guidelines [18]) and at sufficiently high sampling frequency to analyze the pulse shape in detail. Furthermore, in contrast to imaging techniques, monitoring of cerebral compliance by means of ICP pulse morphology analysis can be performed in a continuous manner during the entire time of ICP monitoring. Consequently, this approach would avoid the frequently cited constraints of the standard method of assessing the full pressure–volume curve by injection or withdrawal of fluid from the cerebrospinal fluid space, namely the intermittent nature of the procedure and the risk of causing potentially dangerous increases in ICP through changes in intracranial volume [32].

It has to be noted, however, that this work was conducted as a retrospective study in a relatively small group of patients and with certain limitations. During identification of ICP ranges in full recordings we did not differentiate patients with continuous hypertensive episodes from patients with high instability of the signal, and information about medical interventions affecting mean ICP was not included. The length of analyzed recordings was not standardized and although patients who did not exhibit values in both ICP ranges were excluded from further analysis, no lower length limit for low/high ICP portion of the recording was used. Furthermore, the parameter used to assess the occurrence of different waveform types, i.e., the percentage of all pulses, was global in nature, derived from all data points in given ICP range, not taking into consideration whether they occurred prior to or following ICP increases or at which stage of monitoring. Finally, outcome as assessed by the Glasgow Outcome Scale provides information on the patient's general condition following injury and is a commonly accepted metric [33], but it does not take into account the diverse character of brain injury in terms of type, severity, and comorbidities. Specifically, it may be influenced by extracranial injury, such as damage to the spinal cord or limb amputations [34] which will not result in changes in ICP pulse morphology, and it assigns a disproportionate weight to physical disability over cognitive impairment [35]. The GOS scores were chosen in this study taking into account its pilot nature in order to assess the potential utility of proposed method of ICP pulse analysis and with full awareness that a more exhaustive description of the patient cohort, with regard to both clinical assessment and other physiological factors, would be required to definitively show the benefits of this approach in the clinical setting.

TABLE VI
HYPERPARAMETERS OF USED RESNET ARCHITECTURE. BOTH RESIDUAL AND DOWNSAMPLING RESIDUAL BLOCKS ARE STANDARD RESIDUAL BLOCKS WITH CONVOLUTIONS OF SIZE 3 AND GROUP NORMALIZATION WITH 32 GROUPS. THE DOWNSAMPLING BLOCK IS ALSO SCALING THE OUTPUT BY ADDING STRIDE 2 TO THE FIRST CONVOLUTION IN THE MAIN BRANCH WHILE ADDING SIZE 1 CONVOLUTION WITH THE SAME NUMBER OF FILTERS AS MAIN BRANCH AND STRIDE 2 TO TO THE SKIP CONNECTION BRANCH

Layer Number	Layer Type	Hyperparameters
1	1D-Convolution	input_channels, filters=64, window=3, stride=1
2 - 3	Downsampling Residual Block	filters=64, stride=2
4 - 9	Residual Block	filters=64, stride=1
10	Group Normalization	groups=32, channels=64
11	ReLU	-
12	Average Pooling	Adaptive to size 1
13	Dropout	0.6 probability during training
14	Fully connected layer	64 inputs, 5 neurons

APPENDIX I

The hyperparameters of used networks are shown in Tables VI (ResNet) and VII (LSTM-FCN). The fully connected baseline network consisted of 3 Layers with 64, 32 and 5 neurons with dropout between layers with small probability of 0.3 and ReLU activation functions between hidden layers. The models were trained on a machine with AMD's Ryzen 9 3900XT (3.8 - 4.7 GHz) 12 core CPU and Nvidia's GeForce RTX 3090 GPU with 24 GB of VRAM. The training was performed for 100

TABLE VII
HYPERPARAMETERS OF USED LSTM-FCN ARCHITECTURE. ARCHITECTURE PRESENTED IN THE TABLE PRODUCES THE EMBEDDING WHICH IS LATER CLASSIFIED BY A FULLY CONNECTED LAYER OF 128 INPUTS AND 5 NEURONS

Convolutional branch		Recurrent branch	
Layer type	Hyperparameters	Layer Type	Hyperparameters
1D-Convolution	channels=32, size=8, stride=1	LSTM	hidden_size = 64, layers=1
Batch Norm.	-		
ReLU	-		
1D-Convolution	channels=64, size=5, stride=1		
Batch Norm.	-		
ReLU	-	Dropout	probability = 0.8
1D-Convolution	channels=64, size=3, stride=1		
Batch Norm.	-		
ReLU	-		
Global Average Pooling	-		

epochs optimized by Stochastic Gradiant Descent with Nesterov momentum of 0.95 and starting learning rate of 0.01. Learning rate was lowered to 0.001 on epoch 33 and to 0.0001 in epoch 66. Training times were relatively short - around 4 minutes for ResNet, 2 minutes for LSTM-FCN and less than 1 minute for fully connected baseline. Single label classfication and multi-label classification experiments used Cross Entropy Loss and Binary Cross Entropy, respectively. Model performance in the validation dataset was logged at the end of every epoch and running averages of the training dataset every 10 steps. Batch training with batch size of 256 was used. The source code used for analysis is available https://github.com/MaczekO/ICP_NN in this GitHub repository.

APPENDIX II

This section describes alternative approaches to the classification task that did result in improved accuracy.

Models. In an attempt to simplify the model without a reduction in accuracy, networks based on Long-Short Term Memory cell (LSTMs) [36], Gated Recurrent Unit cell (GRUs) [37], and Shallow Convolutional Neural Networks (CNNs) [38] with different configurations were tested. However, the results were not satisfactory, showing at least 5% drop in accuracy scores.

Data representations. In an attempt to boost the performance of the models, different representations of ICP pulse waveforms were tested: Fourier transform coefficients, spectrograms, approximation by orthogonal Chebyshev polynomials, empirical mode decomposition, and RBF approximation coefficients. However, none of the methods resulted in an improvement over the 1-D vector of signal samples, and some of them resulted in loss of information during the approximation step, posing an additional challenge for the model.

Secondly, the effect of using the ABP signal as a second input was tested by attaching the signal as a second channel for input to convolutional layers and by training a shared weights model of two univariate feature extractors, then concatenating the results and performing classification through a fully connected network. Neither method, however, improved the classification accuracy.

Addressing class imbalance. To reduce the great imbalance between the A+E class and non-artifactual classes, artificial examples were created by choosing a number of examples from other classes and heavily obscuring them with a composition of multiple sine waves with different parameters, as it is similar to the experimental data. This did not improve the results, possibly due to the fact that the method of creating artificial examples introduced new distribution of data into the class that did not match the distribution of existing examples.

Pretraining. The possibility of pretraining on a large, unannotated dataset was investigated in place of starting the classification from a random distribution of weights. A dataset consisting of all pulses from all patients in the training dataset was created and used in pretraining based on an autoencoder structure with a ResNet model without the classfication layer as encoder and a simple fully connected network as decoder. The pretraining lasted 15 epochs with Adam as optimizer, reducing the MSE of signal reconstruction. Unfreezing of only the classification layer during supervised training resulted in a large reduction in accuracy (66% in the validation dataset), probably due to high correlation between pulses from the same patient. Unfreezing of all layers with trained weights used as a starting point for the network resulted in slightly lower (90.24% in the validation dataset) accuracy than starting from random weights.

Multi-label classification. The multi-label approach was based on the same types of networks but with separate sigmoids instead of softmax as output. The classification threshold was unified to 0.5 for all classes. Binary cross entropy for minimization with SGD optimizer was used. Achieved results were similar to single label classification, e.g., ResNet with multi-label output achieved Jaccard score of 82.44% and best accuracy of 95.59% in the validation dataset.

ResNet depth ablation. In a second attempt to reduce the complexity of the network, additional experiments with ResNet architecture were performed. The change in the number of residual blocks lowered the number of parameters as well the complexity of the network. While this procedure did not affect the results in the validation dataset, it significantly affected the generalization to the test dataset, with removal of blocks resulting in a decrease of 4 to 6% in best accuracy and 3 to 6% in strict accuracy. This indicates that all of the layers are important for the generalization ability of used ResNet.

REFERENCES

[1] K. Germon, "Interpretation of ICP pulse waves to determine intracerebral compliance," *J. Neurosci. Nurs.*, vol. 20, no. 6, pp. 344–351, 1988.
[2] M. Czosnyka and G. Citerio, "Brain compliance: The old story with a new 'et cetera'," *Intensive Care Med.*, vol. 38, no. 6, pp. 925–927, 2012.
[3] A. Marmarou, K. Shulman, and J. Lamorgese, "Compartmental analysis of compliance and outflow resistance of the cerebrospinal fluid system," *J. Neurosurg.*, vol. 43, no. 5, pp. 523–534, 1975.
[4] E. R. Cardoso, J. O. Rowan, and S. Galbraith, "Analysis of the cerebrospinal fluid pulse wave in intracranial pressure," *J. Neurosurg.*, vol. 59, no. 5, pp. 817–821, 1983.
[5] J.-Y. Fan, C. Kirkness, P. Vicini, R. Burr, and P. Mitchell, "Intracranial pressure waveform morphology and intracranial adaptive capacity," *Amer. J. Crit. Care*, vol. 17, no. 6, pp. 545–554, 2008.

[6] T. Ellis, J. McNames, and B. Goldstein, "Residual pulse morphology visualization and analysis in pressure signals," in *Proc. IEEE 27th Annu. Conf. Eng. Med. Biol.*, 2006, pp. 3966–3969.
[7] F. Scalzo, P. Xu, M. Bergsneider, and X. Hu, "Nonlinear regression for sub-peak detection of intracranial pressure signals," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2008, pp. 5411–5414.
[8] X. Hu, P. Xu, F. Scalzo, P. Vespa, and M. Bergsneider, "Morphological clustering and analysis of continuous intracranial pressure," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 3, pp. 696–705, Mar. 2009.
[9] I. M. Elixmann, J. Hansinger, C. Goffin, S. Antes, K. Radermacher, and S. Leonhardt, "Single pulse analysis of intracranial pressure for a hydrocephalus implant," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 3939–3942.
[10] A. Calisto, M. Galeano, S. Serrano, A. Calisto, and B. Azzerboni, "A new approach for investigating intracranial pressure signal: Filtering and morphological features extraction from continuous recording," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 3, pp. 830–837, Mar. 2013.
[11] H.-J. Lee, E.-J. Jeong, H. Kim, M. Czosnyka, and D.-J. Kim, "Morphological feature extraction from a continuous intracranial pressure pulse via a peak clustering algorithm," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 10, pp. 2169–2176, Oct. 2016.
[12] S.-B. Lee *et al.*, "Artifact removal from neurophysiological signals: Impact on intracranial and arterial pressure monitoring in traumatic brain injury," *J. Neurosurgery*, vol. 132, no. 6, pp. 1952–1960, 2019.
[13] B. Quachtran, R. Hamilton, and F. Scalzo, "Detection of intracranial hypertension using deep learning," in *Proc. 23 rd Int. Conf. Pattern Recognit.*, 2016, pp. 2491–2496.
[14] C. G. Nucci *et al.*, "Intracranial pressure wave morphological classification: Automated analysis and clinical validation," *Acta Neurochirurgica*, vol. 158, no. 3, pp. 581–588, 2016.
[15] P. K. Eide, "A new method for processing of continuous intracranial pressure signals," *Med. Eng. Phys.*, vol. 28, no. 6, pp. 579–587, 2006.
[16] M. Megjhani *et al.*, "An active learning framework for enhancing identification of non-artifactual intracranial pressure waveforms," *Physiol. Meas.*, vol. 40, no. 1, 2019, Art. no. 0 15002.
[17] T. F. Brain *et al.*, "Guidelines for the management of severe traumatic brain injury. introduction," *J. Neurotrauma*, vol. 24, p. S 1, 2007, doi: 10.1089/neu.2007.9999.
[18] N. Carney *et al.*, "Guidelines for the management of severe traumatic brain injury," *Neurosurgery*, vol. 80, no. 1, pp. 6–15, 2017.
[19] E. S. Connolly Jr. *et al.*, "Guidelines for the management of aneurysmal subarachnoid hemorrhage: A guideline for healthcare professionals from the American Heart Association/American Stroke Association," *Stroke*, vol. 43, no. 6, pp. 1711–1737, 2012.
[20] S. Holm and P. K. Eide, "Impact of sampling rate for time domain analysis of continuous intracranial pressure (ICP) signals," *Med. Eng. Phys.*, vol. 31, no. 5, pp. 601–606, 2009.
[21] S. M. Bishop and A. Ercole, "Multi-scale peak and trough detection optimised for periodic and quasi-periodic neuroscience data," in *Proc. Intracranial Press. Neuromonitoring XVI*, 2018, pp. 189–195.
[22] E. Carrera *et al.*, "What shapes pulse amplitude of intracranial pressure?," *J. Neurotrauma*, vol. 27, no. 2, pp. 317–324, 2010.
[23] R. Artstein and M. Poesio, "Inter-coder agreement for computational linguistics," *Comput. Linguistics*, vol. 34, no. 4, pp. 555–596, 2008.
[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[25] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2017.
[26] R. Hadsell, S. Chopra, and Y. LeCun, "*Dimensionality reduction by learning an invariant mapping*," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, 2006, pp. 1735–1742.
[27] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems* vol. 32 (H. H. Wallach, A. Larochelle Beygelzimer, F. d Alché-Buc, E. Fox, and R. Garnett, Eds.), pp. 8024–8035, Curran Associates, Inc., 2019.
[28] M. Imhoff and S. Kuhls, "Alarm algorithms in critical care monitoring," *Anesth. Analg.*, vol. 102, no. 5, pp. 1525–1537, 2006.
[29] M. Feng, L. Y. Loy, F. Zhang, and C. Guan, "Artifact removal for intracranial pressure monitoring signals: A robust solution with signal decomposition," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2011, pp. 797–801.
[30] F. Scalzo, D. Liebeskind, and X. Hu, "Reducing false intracranial pressure alarms using morphological waveform features," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 1, pp. 235–239, Jan. 2012.
[31] C. Hawthorne and I. Piper, "Monitoring of intracranial pressure in patients with traumatic brain injury," *Front. Neurol.*, vol. 5, pp. 121–137, 2014.
[32] C. S. Robertson *et al.*, "Clinical experience with a continuous monitor of intracranial compliance," *J. Neurosurg.*, vol. 71, no. 5, pp. 673–680, 1989.
[33] T. McMillan, L. Wilson, J. Ponsford, H. Levin, G. Teasdale, and M. Bond, "The glasgow outcome scale-40 years of application and refinement," *Nature Rev. Neurol.*, vol. 12, no. 8, pp. 477–485, 2016.
[34] N. Stocchetti and E. R. Zanier, "Chronic impact of traumatic brain injury on outcome and quality of life: A narrative review," *Critical Care*, vol. 20, no. 1, pp. 1–10, 2016.
[35] S. I. Anderson, A. M. Housley, P. A. Jones, J. Slattery, and J. D. Miller, "Glasgow outcome scale: An inter-rater reliability study," *Brain Inj.*, vol. 7, no. 4, pp. 309–317, 1993.
[36] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[37] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *NIPS 2014 Workshop Deep Learning*, Dec. 2014, *arXiv:1412.3555*.
[38] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.