Objective Assessment of Beat Quality in Transcranial Doppler Measurement of Blood Flow Velocity in Cerebral Arteries

Kian Jalaleddini [iD], Nicolas Canac [iD], Samuel G. Thorpe [iD], Michael J. O'Brien [iD], Mina Ranjbaran [iD],
Benjamin Delay [iD], Amber Y. Dorn [iD], Fabien Scalzo [iD], Corey M. Thibeault [iD],
Seth J. Wilk, *Senior Member, IEEE*, and Robert B. Hamilton

***Abstract*—*Objective*: Transcranial Doppler (TCD) ultrasonography measures pulsatile cerebral blood flow velocity in the arteries and veins of the head and neck. Similar to other real-time measurement modalities, especially in healthcare, the identification of high-quality signals is essential for clinical interpretation. Our goal is to identify poor quality beats and remove them prior to further analysis of the TCD signal. *Methods*: We selected objective features for this purpose including Euclidean distance between individual and average beat waveforms, cross-correlation between individual and average beat waveforms, ratio of the high-frequency power to the total beat power, beat length, and variance of the diastolic portion of the beat waveform. We developed an iterative outlier detection algorithm to identify and remove the beats that are different from others in a recording. Finally, we tested the algorithm on a dataset consisting of more than 15 h of TCD data recorded from 48 stroke and 34 in-hospital control subjects. *Results*: We assessed the performance of the algorithm in the improvement of estimation of clinically important TCD parameters by comparing them to that of manual beat annotation. The results show that there is a strong correlation between the two, that demonstrates the algorithm has successfully recovered the clinically important features. We obtained significant improvement in estimating the TCD parameters using the algorithm accepted beats compared to using all beats. *Significance*: Our algorithm provides a valuable tool to clinicians for automated detection of the reliable portion of the data. Moreover, it can be used as a preprocessing tool to improve the data quality for automated diagnosis of pathologic beat waveforms using machine learning.

***Index Terms*—Biomedical signal processing, ultrasonography, signal processing algorithms, change detection algorihms, heuristic algorithms, algorithm design and analysis.**

I. INTRODUCTION

TRANSCRANIAL *Doppler* (TCD) ultrasonography is a diagnostic technique for rapid, non-invasive assessment of cerebrovascular health [1]. TCD was first used in neurology in the 1980s and measures *cerebral blood flow velocity* (CBFV) [2]. TCD has demonstrated utility in the diagnosis of clinical conditions such as acute ischemic stroke [3], [4], intracranial pressure [5], [6], sickle cell disease [7], traumatic brain injury [8], dementia [9], cerebral emboli [10] and many others [11], [12].

CBFV beat waveform morphology has not been extensively studied mainly due to the noise level and lack of signal processing techniques developed for TCD signal analysis [5], [13]. Beat waveform morphology can exhibit a high degree of variability due to noise within the signal often caused by the movement of the subject or sonographer or the electronics, Fig. 9(A). However, clinically important parameters such as *mean Cerebral Blood Flow Velocity* (mCBFV) or *Pulsatility Index* (PI), which may be used to aid in diagnostic assessment, are often extracted from this same highly variable TCD beat waveform morphology [5], [8], [14], [15]. Thus, being able to reliably identify and exclude the low-quality portions of the signal in a consistent, quantitative manner is imperative for the accurate extraction of clinically relevant TCD parameters and any ensuing clinical interpretation based on those parameters.

The problem of variability and noise in the CBFV recording is often mitigated by manual classification of beats by an expert that "appear" different than others in a recording. This process can be time-consuming, labor intensive, subjective, and difficult to reproduce. In addition, it precludes real-time computation of clinically relevant parameters. Consequently, there is a need for objective beat quality criteria and an automated algorithm for classification of poor-quality beats to make the analysis comprehensive, reliable, and repeatable. The purpose of this paper is to present candidate features for objective quantification of beat quality, develop a classification algorithm to identify poor-quality beats in a recording, and to validate the resulting tool through comparison to human experts using a large dataset consisting of healthy and pathological beat waveforms.

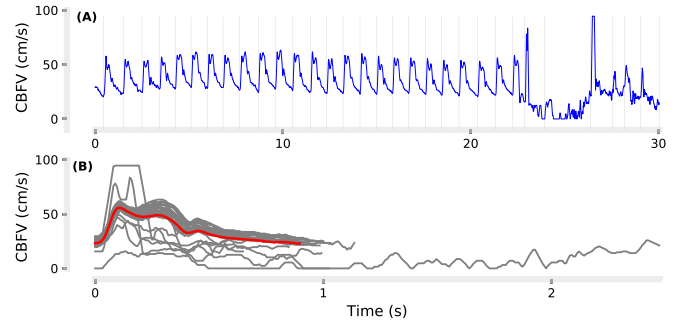|	
Subjects	82
TCD sessions	131
Recording intervals	1872
Recording interval per subject: average (standard deviation)	20.5(10.7)
Recording interval per subject: (minimum, maximum)	(1, 47)
Total beats	74,500
Beats rejected by rater1	8,472
Beats rejected by rater2	6,745
Beats rejected by algorithm (ED)	4,827
Beats rejected by algorithm (CC)	8,690
Beats rejected by algorithm (BL)	5,230
Beats rejected by algorithm (HFNP)	6,848
Beats rejected by algorithm (DV)	5,818

Fig. 1. Beat start/end times were identified from measurement of CBFV using our beat detection software (A), and the average beat waveform $\overline{CBFV}_b(k)$ (red beat) was computed from ensemble averaging of individual beat waveforms (grey beats) (B).

II. METHODS

In this section, we describe the experimental protocol for data acquisition, the features we used to characterize beat quality, and an algorithm to classify and remove low-quality beats.

A. Experiments

1) Subjects: We acquired TCD waveforms from 82 subjects enrolled at Erlanger Health Systems Southeast Regional Stroke Center in Chattanooga, TN. The subjects were either diagnosed with *Large Vessel Occlusion* (LVO) of intracranial arteries confirmed with *Computed Tomography Angiography* (CTA) or in-hospital control subjects who arrived at the hospital presenting with stroke symptoms, but were later confirmed negative for LVO by CTA imaging. We also performed a follow-up TCD recording on the LVO subjects within 72 hours of injury. In total, we acquired data from 131 sessions. The experiment protocols were approved by University of Tennessee College of Medicine Institutional Review Board (ID: 16-097).

2) Recording: A trained sonographer acquired TCD scans using a 2 MHz hand-held probe. CBFV signals associated with the left or right *Middle Cerebral Arteries* (MCA) were identified by insonating through the transtemporal windows and recording the CBFV at the sampling rate of 125 Hz. The sonographer obtained recordings for as many depths as possible between 45-60 mm. Once a CBFV signal with a smooth fitting envelope trace was identified and optimized at a specific depth, a recording interval would begin, continuing for a total duration of 30 s. Two TCD experts manually inspected 1,872 recording intervals with a total of 74,500 beats (see Table I). Rater 1 was a peer-reviewed published scientist in the area of TCD waveform morphology. Rater 2 was a licensed registered vascular technologist.

B. Algorithm

1) Features: For each recording interval, we identified beat start and end times using the algorithm developed in [16] as shown in Fig. 1(A). Then, we normalized individual beat lengths to the median beat length. This was achieved by padding the beats whose lengths were shorter than the median beat length with the last value of the beat and truncating the beats whose length was longer than the median beat length. Then, from the l length-normalized beat waveforms $CBFV_{b1}(k), \ldots, CBFV_{bl}(k)$, we computed the average CBFV beat waveform (see Fig. 1(B)) denoted by $\overline{CBFV}_b(k)$. We extracted the following features for each individual beat.

Euclidean Distance (ED): ED_i is the Euclidean distance of the i-th length-normalized beat waveform from $\overline{CBFV}_b(k)$:

$$ED_i = \sqrt{\sum_{k=1}^{n} \left(CBFV_{bi}(k) - \overline{CBFV}_b(k)\right)^2} \quad (1)$$

where n is the number of samples in the length-normalized beats and $i \in \{1, \ldots, l\}$.

Cross-Correlation (CC): CC_i is the maximum of cross-correlation coefficient between $CBFV_{bi}(k)$ of the i-th beat and $\overline{CBFV}_b(k)$:

$$CC_i = \max \left(\sum_{m=1}^{n} CBFV_{bi}(m)\overline{CBFV}_b(m+k)\right) \quad (2)$$

Beat Length (BL) BL_i is the length of the i-th original (vs length-normalized) beat.

High Frequency Noise Power (HFNP) $HFNP_i$ is the ratio of the high-frequency to low-frequency power for the i-th beat.

$$HFNP_i = \frac{\sum_{f=f_t}^{f_N} |X_i(f)|^2}{\sum_{f=1}^{f_t} |X_i(f)|^2} \quad (3)$$

where $X(f)$ is the discrete Fourier transform of the CBFV, and f_t represents the threshold frequency set to 15 Hz and f_N is the Nyquist frequency.

Diastolic Variance (DV) DV_i is the variance of the diastolic portion of the i-th beat which was defined as the last 20% of the beat.

See Fig. 2 for an illustration of the distributions of these features for a sample recording.

2) Iterative Outlier Detection Algorithms: We developed an iterative version of the *InterQuartile Range* (IQR) outlier detection method. The IQR method, first proposed by Tukey [17], [18], works by calculating $IQR = q_3 - q_1$ where q_1 and q_3 are the first and third quadrants. It labels the data points that fall below $B_L = q_1 - 1.5IQR$ or above $B_H = q_3 + 1.5IQR$ as outliers. The range is shown by the location of the whiskers
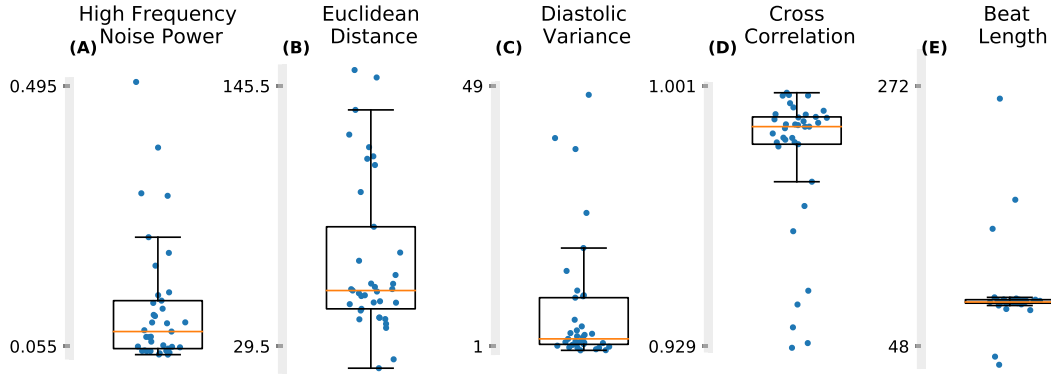
Fig. 2. Five features are assembled from individual beat CBFV waveforms: (A) *High-Frequency Noise Power* (HFNP) is the ratio of the high-frequency to low-frequency power; (B) *Euclidean Distance* (ED) is the euclidean distance from the average beat waveform; (C) *Diastolic Variance* (DV) is the variance of the diastolic portion of the beat waveform; (D) *Cross-Correlation* (CC) is the maximum of cross-correlation with the average beat waveform; (E) *Beat-Length* (BL) is the beat length in samples. The orange line indicates the median of the data. The boxes extends from the 25-th (q_1) to 75-th (q_3) percentiles of the data, indicating the interquartile range ($IQR = q_3 - q_1$) and the whiskers extends from $q_1 - 1.5IQR$ to $q_3 + 1.5IQR$.

in the box plot, Fig. 2. The *Iterative IQR* (IIQR) method works by identifying the farthest outlier from the boundary, if any, and removing one outlier at each iteration, continuing until no further outliers are present. This ensures that the boundaries B_L and B_H are not biased because of the presence of too many large outliers.

Iterative InterQuartile Range (IIQR) Algorithm: The following iterative algorithm works by identifying outliers from the set $F = \{f_1, f_2, \ldots, f_l\}$ for a beat feature described above (e.g. ED) where n is the total number of beats.

1) Denote the list of outlier indices as $I^o = \{\}$.
2) Calculate q_1 and q_3 that are the 25-th and 75-th percentiles of F.
3) Populate F^o with the outliers:

$$F_s^o = \{\forall x \in f, x < q_1 - 1.5IQR\}$$
$$F_l^o = \{\forall x \in f, x > q_3 + 1.5IQR\} \qquad (4)$$

If F_s^o and F_l^o are empty, then exit, returning the list of the outlier indices I^o.

4) Find the outlier that is the farthest from the boundaries:

$$\delta^s = f_s^o - (q_1 - 1.5IQR)$$
$$\delta^l = q_3 + 1.5IQR - f_l^o \qquad (5)$$

5) Append I^o with the index of the maximum, i^o:

$$i^o = \arg\max(\delta^s, \delta^l) \qquad (6)$$

6) Go to step 2.

We used the IIQR algorithm on each beat for all features in the feature set. The final rejected beats were the union of those identified as outliers for each feature.

C. Clinical Metrics

We computed the following clinically significant TCD parameters from the average beat waveform for each recording interval. The average beat waveform was calculated using the technique described above II-B1 from all the beats as well as the beats accepted by a rater in a recording interval.

TCD Parameters
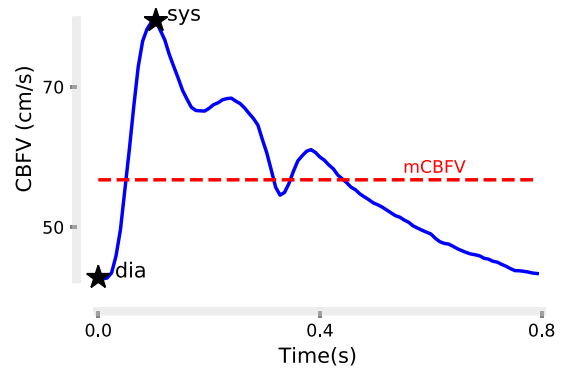
Fig. 3. Typical average beat $\overline{CBFV}_b(k)$ calculated from a recording interval. The figure is annotated to show the morphological points of the waveform that based on which the TCD parameters (e.g. mCBFV, PI) are defined.

See Fig. 3 for illustration.

- *Mean Cerebral Blood Flow Velocity* (mCBFV) is the mean of the average beat waveform.

$$mCBFV = \frac{1}{l} \sum_{k=1}^{l} \overline{CBFV}_b(k) \qquad (7)$$

- *Pulsatility Index* (PI) is defined as:

$$PI = \frac{sys - dia}{mCBFV} \qquad (8)$$

where sys is the systolic peak and dia is the diastolic valley of the average beat waveform. PI represents a measure of cerebrovascular resistance.

D. Statistics

We compared the agreement between raters and the algorithm at two levels: (i) beats; (ii) clinical parameters.

1) Beat Agreement: We compared the agreement of raters in acceptance and rejection of beats using confusion matrix and the Cohen's Kappa coefficient (κ). In order to determine the intra-rater agreement, the experts inspected 200 recordings, randomly selected from the data pool, again at a later time.

2) Clinical Parameter Agreement: We will use the following notation for comparison between raters at the level of clinical parameters:

$$D^{CM}_{rater1-rater2} = 100 \frac{CM_{rater1} - CM_{rater2}}{CM_{rater2}} \qquad (9)$$

where CM is the clinical metric: $mCBFV$ or PI. D is the percentage difference in the clinical metric of interest between two raters.

To compare two raters at this level, we first computed the percentage difference in the clinical metric of interest compared to when all beats are accepted: $D^{CM}_{all-rater1}$ and $D^{CM}_{all-rater2}$. Second, we aggregated the results for all recording intervals and computed the Pearson correlation coefficient between $D^{CM}_{all-rater1}$ and $D^{CM}_{all-rater2}$. Here a rater can be human or algorithm.

In order to compare the efficacy of the algorithm with a human rater and identify an optimal feature combination, we defined improvement I as the net reduction in error using the algorithm-accepted beats:

$$I^{CM} = \left| D^{CM}_{all-rater} \right| - \left| D^{CM}_{algo-rater} \right| \qquad (10)$$

If the rejected beats by a human rater shift the estimate of the clinical parameter considerably, then $\left| D^{CM}_{all-rater} \right|$ would be a significant positive value. If the algorithm is as effective as a human rater, then $\left| D^{CM}_{algo-rater} \right|$ would be close to zero and the improvement I^{CM} becomes positive. Otherwise, improvement I^{CM} is negative. For many recording intervals, improvement in the estimate of clinical parameters of using the algorithm-accepted beats over using all beats was positive. For some recordings, however, the improvement was negative, which is because of frequent misclassified beats in a recording and/or misclassified beats that had a disproportionately large contribution to the clinical parameter. We quantified the performance of each feature combination using the sum of the positive improvements (S_{pos} brown area under the curve in Fig. 4) and the sum of the negative improvements (S_{neg} absolute value of the cyan area under the curve in Fig. 4):

$$S^{CM}_{pos} = \sum \left[x : \forall x \in I^{CM} \wedge x \geq 0 \right]$$
$$S^{CM}_{neg} = \left| \sum \left[x : \forall x \in I^{CM} \wedge x < 0 \right] \right| \qquad (11)$$

Thus, the larger the positive improvement S_{pos} and smaller the negative improvement S_{neg}, the closer the algorithm to the human rater. See Fig. 4 for illustration of the case when we use all five features in the algorithm. For the purpose of identifying the optimal feature combination, we normalized and consolidated the clinical parameters mCBFV and PI:

$$S_{pos} = S^{mCBFV}_{pos} + S^{PI}_{pos}$$
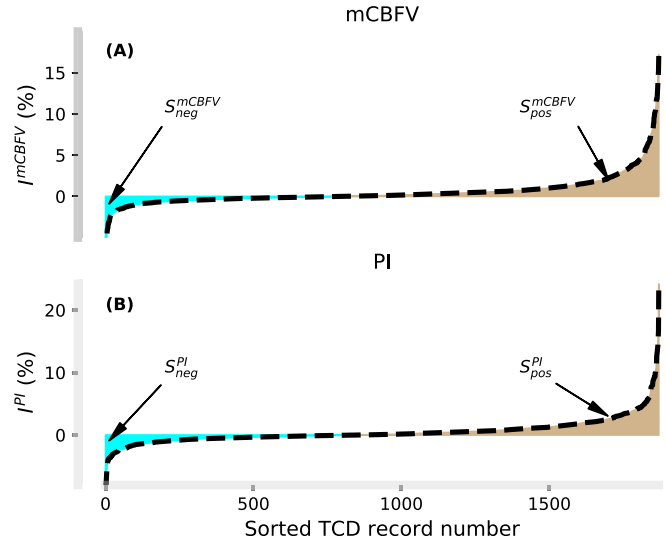$$S_{neg} = S^{mCBFV}_{neg} + S^{PI}_{neg} \qquad (12)$$

Fig. 4. Sorted improvement gained by using the algorithm-accepted beats over using all beats in the computation of clinical parameters: (A) *mean Cerebral Blood Flow Velocity* (mCBFV); (B) Pulsatility Index (PI). All five features ED-CC-BL-HFNP-DV were used for this figure. The area under the curve for positive improvement S_{pos} is shown in brown and for negative improvement S_{neg} in cyan. The larger S_{pos} and smaller S_{neg}, the closer the algorithm to human rater in filtering TCD beats for a more accurate estimate of TCD clinical parameters.

To identify the variability of improvement for each feature combination, we performed a bootstrap analysis using 1000 iterations for each feature combination. At each iteration, we randomly selected the data with replacement and calculated the mean and variability.

III. EXPERIMENTAL RESULTS

At the beat level, the agreement between human raters was in the moderate range ($\kappa_{12} = 0.58$). Rater 1 rejected 11.4% and Rater 2 rejected 9.0% of the beats. Overall, they disagreed on acceptance/rejection of 7.7% of the beats (Fig. 5). Similarly, the agreements of the raters with themselves were in the moderate range: $\kappa_{11} = 0.62$, $\kappa_{22} = 0.64$.

At the level of clinical parameters, the percentage differences in the clinical parameters were highly correlated between the two human raters ($d^{CM}_{all-rater1}$ versus $d^{CM}_{all-rater2}$), Fig. 6. Following the same analysis, we also compared the human raters to a random rater simulated as a Bernoulli distribution that rejected 10.2% of the beats (average of the rejection ratios of the human raters). Fig. 6 shows that the correlation between human raters and the random virtual rater was zero which further signifies the identified correlation between the human raters.

Fig. 7 illustrates the normalized positive and negative improvements for the 31 possible feature combinations. There was a trade-off between S_{pos} and S_{neg}, i.e. a feature combination with large S_{pos} (desired) typically resulted in larger S_{neg} (undesired). The optimal feature is one that maximizes $S_{pos} - S_{neg}$. There were three common combinations among the top five combinations for each rater: ED-LEN, CC, CC-LEN. Interestingly, for both raters, the poorest feature combination was DV. Perhaps more importantly though, is the fact that a large number of combinations containing different constituent
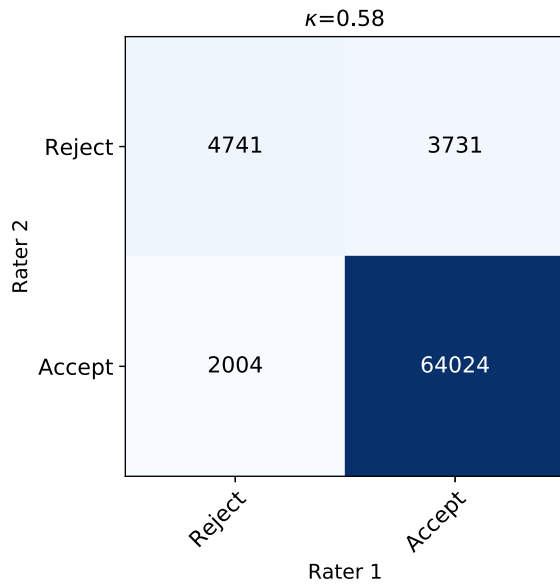
Fig. 5. Beat level agreement between the two human raters. Cohen's Kappa coefficient κ_{12} was 0.58.
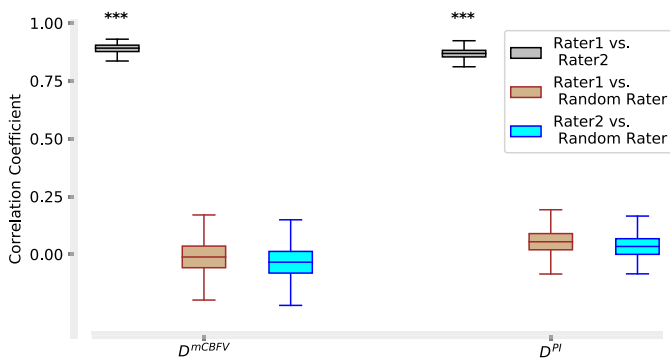

Fig. 6. Correlation coefficient of different raters between percentage differences in the clinical parameters because of beat rejection $D_{all-rater1}^{CM}$, $D_{all-rater2}^{CM}$, $D_{all-random\ rater}^{CM}$ where CM is the clinical metric of interest (mCBFV, PI). The correlation between human raters was high and the correlation between human rater and a random rater was insignificant. The boxes extend from the 25-th (q_1) to 75-th (q_3) percentiles of the data, and the whiskers extends from $q_1 - 1.5IQR$ to $q_3 + 1.5IQR$.

features all perform significantly better compared with not performing any beat rejection. In fact, even the worst performing feature combination (DV) had a significantly larger positive improvement over negative improvement for both raters. As seen in Fig. 7, the majority of other feature combinations comprise a significantly better performing cluster, which provides evidence for the robustness of the algorithm as a tool for generally differentiating between low and high quality segments of data, independent of the exact features selected.

We also assessed the agreement between the algorithm and the human raters at the level of estimation of clinical parameters. We used CC as the only feature for the algorithm and inspected the correlation coefficient between the percentage differences in the clinical parameters: $D_{all-rater1}^{CM}$, $D_{all-rater2}^{CM}$ and $D_{all-algorithm}^{CM}$. Fig. 8 demonstrates this high correlation.

Fig. 9 demonstrates typical performance of the beat rejection algorithm with CC as the only feature. It is evident that there is agreement between human rater-accepted beats (Fig. 9(A)) and those accepted using the algorithm (Fig. 9(B)). In this example, however, there are three beats that are subject to disagreement; two are rejected by the human rater but accepted by the algorithm and one rejected by the algorithm but accepted by the human rater. We elaborate on these disagreements in more detail in Section IV-C. Fig. 9(C-E) visualizes the average and 95% range of the beat waveforms calculated from all beats, human rater-accepted and algorithm-accepted beats. The average beat waveforms were consistent between human rater and algorithm-accepted beats as evident by the small Euclidean distance between them (7.48, Fig. 9(F)) and the consistent estimate of the clinical parameters $mCBFV$ and PI. The variability of the ensemble was also reduced dramatically as a result of removal of poor quality beats. Evidently, the beats that were the subject of disagreement had little effect on the final average beat waveform morphology and thus on the extracted clinical parameters. However, the difference in clinical parameters between all beats and human rater-accepted beat waveforms was quite large, highlighting the importance of filtering out poor quality beats for a more accurate estimate of clinical parameter.

IV. Discussion

A. Conclusion

In this paper, we proposed objective features to assess beat quality in TCD recordings from cerebral arteries, namely CC, ED, LEN, HFN, DV. We described the *Iterative InterQuartile Range* (IIQR) algorithm, an outlier rejection method that used these features to label beats that were different from others in a TCD recording. We assessed the performance of this tool on a dataset consisting of over 15 hours of TCD data with more than 74,000 beats recorded from healthy and pathological subjects. We identified that the best feature combinations for identifying poor-quality beats were CC, CC-LEN, and ED-LEN for our data. We demonstrated that the IIQR algorithm successfully identified low-quality beats such that the clinical parameters from the algorithm-accepted beats were consistent with those computed from expert-accepted beats.

B. Raters

Two expert human raters manually inspected 74,500 beats to make acceptance/rejection decisions. The rating process was highly time-consuming and took several days. The agreement between the raters at the beat level was moderate $\kappa_{12} = 0.58$ [19]. We visually inspected the first hundred ratings with the largest disagreement between the raters. It was evident the major source of discrepancy between raters was disagreement over beats with "average" quality. We also found that their threshold for accepting an "average" beat was variable. This finding was consistent with our finding of moderate intra-rater agreements: $\kappa_{11} = 0.62$, and $\kappa_{22} = 0.64$ comparable to the inter-rater agreement. The shifts in the clinical parameters because of the rejected beats were, however, highly correlated for the raters. This is presumably due to the insensitivity of the clinical metrics
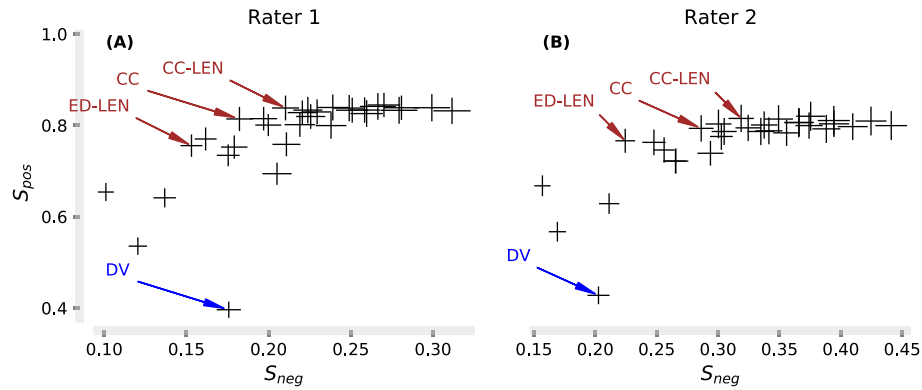
Fig. 7. Cumulative positive and negative improvements S_{pos}, S_{neg} for 31 possible feature combinations. Each combination is shown by a cross; the center of the cross represents the mean bracketed by vertical and horizontal bars representing standard deviations in S_{pos} and S_{neg} directions associated with the bootstrap study. The optimal combinations which maximized $S_{pos} - S_{neg}$ were CC, and CC-DV.
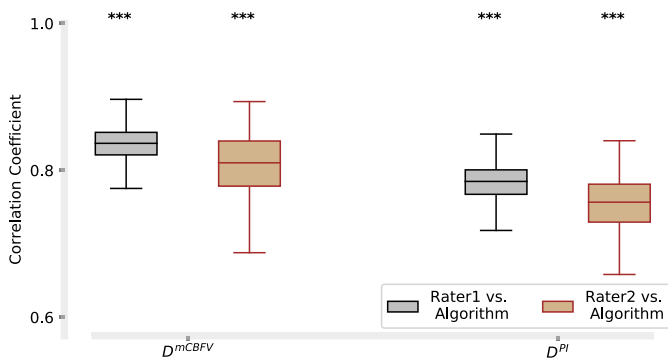

Fig. 8. Correlation coefficient of the algorithm versus human raters between percentage differences in the clinical parameters because of beat rejection: $D_{all-rater1}^{CM}$, $D_{all-rater2}^{CM}$ and $D_{all-algorithm}^{CM}$.

to the acceptance/rejection of the "average" quality beats while both raters agreed on rejecting beats that were very poor and whose rejection did make a difference in the estimate of the clinical metrics. This points to the much-needed rejection of very poor beats, the lack of objective guidelines for beat classification and the subjective nature of manual beat classification.

C. Data

In order to properly interpret the results of this study, a number of facets pertaining to the data require discussion, specifically with regard to the expert annotations.

We found the optimal features for the beat rejection model by reproducing the shifts in the clinical parameters that expert human raters made. We did *not* attempt to tune our model to reproduce the accepted/rejected beats directly. The major reason was the inherently subjective and unrepeatable nature of manually evaluating data quality, evident from the moderate agreement between our human raters. This limitation likely persists even for the most experienced TCD experts, and it is important to keep this in mind when drawing comparisons to the expert-accepted beats, e.g. Fig. 9. Furthermore, while the difference between poor and good beats is generally self-evident, no objective classification criteria exists. An unavoidable consequence of this is low inter-and intra-rater reliability for intermediate range of

signals. As a result, no true gold standard can exist for this type of work. Thus, beat level disagreement between the raters does not necessarily imply inaccuracy and can be simply the side affect of rater variability and lack of objective classification criteria.

Nevertheless, some method of evaluation is required, and some metric of performance, however imperfect, is needed. It is clear from the examples shown here, comparison to random raters (Fig. 6), and based on inspection by other independent experts that the set of manually accepted beats used in this study represents a significant improvement in signal quality over simply using the set of all detected beats. While exact agreement is not necessarily desirable or even achievable, the percentage shift in clinical parameters estimated from accepted beats compared to using all beats provides an objective way of quantifying the discrepancies to compensate for the lack of a gold-standard. This metric was particularly useful because it is clinically meaningful, sensitive to the poor beats whose acceptance/rejection makes a difference in the estimate of the clinical parameters, and insensitive to average quality beats whose acceptance/rejection does not necessarily shift the clinical parameters considerably.

Finally, to facilitate the manual inspection process, a software tool was developed to display the data along with the detected beats, which would allow the user to add, delete, or shift beat start and stop indices. In addition, this tool included a feature to aid the user in identifying potentially bad beats by flagging beats that were abnormal according to their cross correlation with the average beat waveform. The user was ultimately responsible for manually labeling the data. We turned off this feature for the second rater but CC still stood out and was among the top features for both raters together with ED-LEN and CC-LEN. Nevertheless, we feel that one of the more important takeaways from this work is that any combination of the features discussed resulted in significantly improved performance, a result which appears to be robust in spite of the limitations or biases present in this study.

D. Features & Algorithm

To ensure fidelity of reported clinical parameters from TCD measurement, an expert typically inspects the data and
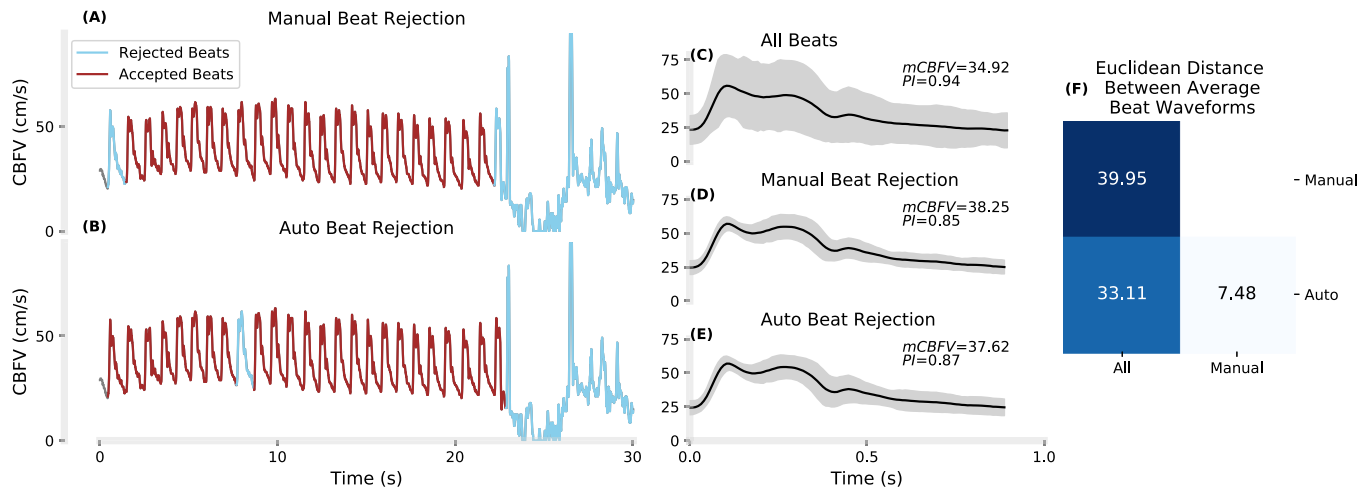
Fig. 9. Typical TCD recording annotated manually by a human rater (A) and by the algorithm (B). Comparison of the average beat waveforms from all beats accepted (C), human-rater accepted beats (D), and algorithm accepted beats (E) along with the TCD clinical parameters $mCBFV$ and PI estimated from the average beat waveforms. Euclidean-distance between the identified average beat waveforms (F).

labels/removes the low-quality portion of the data, a process which is fraught with problems, as previously detailed. Quantitative measures of TCD signal quality are of paramount importance for the accurate extraction of clinical parameters. However, the TCD beat waveform is complex and can change based on subject, age, race, health, vessel, and sonographer skill, among other things. Therefore, absolute feature criteria for beat classification can be inappropriate as they can change from recording to recording independent of data quality. Consequently, we chose to classify beats by comparing their features relative to others in a recording and developed the IIQR algorithm to identify those with substantially different values.

In this work, we have presented quality features for each beat to be compared against those from other beats in a recording for decision making. We selected these features with physiological, clinical and signal processing considerations. We presented two features to assess the beat waveform morphology. Clinical features such as systole, diastole, mean and peaks velocities, systolic flow acceleration, oscillating flow, curvature among others extracted from the TCD beat waveform morphology have been shown to be a major source of clinical information. We used Euclidean distance and cross-correlation to characterize the difference of a beat waveform morphology from the average waveform. They are different in the sense that cross-correlation (the normalized version that we used here) is not sensitive to velocity scaling or shifting, and to the temporal shifting of the beat waves and is a measure of synchrony between the waveforms. Euclidean distance is more effective when it is desired to reject beats based on the variation in morphological points of beats. We also used beat length. Deviations in beat length can be attributed to both the inherent healthy and pathologic variability of the heart. More importantly, beat length deviations can also be due to errors in identifying the beat onset and end point. It is also common to fail to detect any beats in portions of very low quality signal, resulting in abnormal lengths. Thus, it is in our interest to label beats that

are either outliers in terms of beat length, i.e. statistically longer or shorter than their adjacent beats. Inclusion of these beats may compromise the quality of the average beat waveform. We used high frequency noise power and diastolic variance for signal processing and instrumentation reasons. High frequency noise power is a measure of undesired high-frequency noise in the acquired signal since the CBFV should have power primarily in low frequencies only ($f < 15Hz$). Therefore, beats with larger high frequency powers are more likely to be of low quality and subject to environmental noise and signal processing artifacts. The diastolic portion of a CBFV beat waveform is more prone to lower signal-to-noise ratio compared to the systolic portion because of the lower power of the diastole. Therefore, we proposed to use diastolic variance as a measure of power in the diastolic portion of the beat which is expected to have low variability but occasionally has high variability due to noise.

With the five proposed features, we had 31 different combinations of features to choose from. In order to find the optimal feature combination, we systematically assessed each feature, inspected the results and compared the clinically important parameters against those obtained from expert-accepted beats. We identified a trade-off: the higher the positive improvement for a combination, the higher the negative improvement, Fig. 7. An important factor was the number of features. As the number of features in a feature combination increased, with the exception of a few, both the positive and negative improvements increased, as illustrated in Fig. 10. Therefore, a combination with more features, on average, resulted in large positive improvement when quality was poor and many beats needed to be rejected. However, the same combination also rejected beats in recordings where beat quality was not poor which resulted in negative improvements. Therefore, we selected the combination that maximized the "score function" defined as the negative improvement subtracted from positive improvement. We identified that CC, CC-LEN, ED-LEN were the best feature combinations to make beat rejection decisions. It is important to note that these
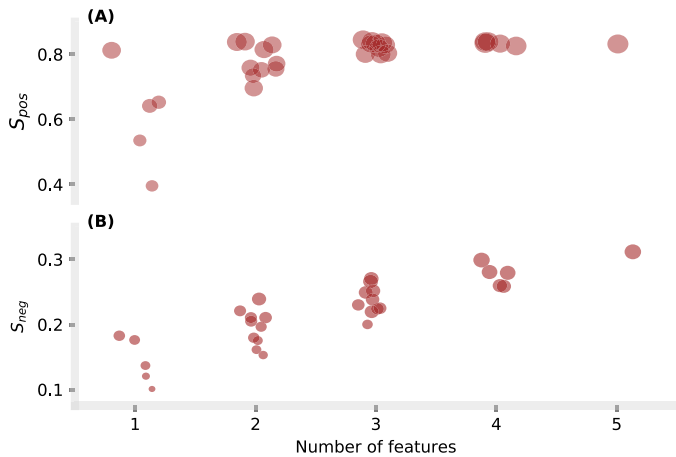
Fig. 10. Changes in cumulative positive S_{pos} (A) and negative S_{neg} (B) improvements as a function of number of features in a feature combination. The improvement of a feature combination is represented by a circle centered at (number of feature + jitter, improvement) with radius representing the 95% range. Small jitter in horizontal axis was introduced for visual separation of circles. On average, a combination with many features had a large S_{pos} (desired) at the cost of having a large S_{neg} (undesired).

choices were optimal only for our labeled data. Theoretically, the choice could be different depending on the experimental conditions, for example with a different ultrasound probe, data acquisition system, sonographer, etc. Nevertheless, as a general rule of thumb, we believe that cross-correlation is a good first feature to inspect. In addition, if there are strong *a priori* reasons to choose certain features based on knowledge of the data or particular study, then the appropriate features can be selected.

We have developed the *Iterative InterQuartile Range* (IIQR) method to detect outlier beats. We chose Tukey's IQR method because it is non-parametric, robust to outliers, requires no *a priori* information about the data, and is not guaranteed to always label a portion of the data as outliers [18], [20]. In this method, outliers are the points falling outside of the whiskers. We used Tukey's standard boxplot where whiskers are at $q_1 - \beta(q_3 - q_1)$ and $q_3 + \beta(q_3 - q_1)$ with $\beta = 1.5$ where q_1 and q_3 are the 25-th and 75-th percentiles. One can change the sensitivity of the algorithm by changing β. The lower the β, the higher the sensitivity of the algorithm in rejecting outliers, i.e. the user is specifically interested in the regions of the data with the highest quality because of their specific use case. In fact, there are some studies demonstrating optimal β other than 1.5 if the number and distribution of the observations are known [21], [22]. However, since these are not known *a priori* for our application, we used the conventional $\beta = 1.5$ in this paper. Outlier detection algorithms typically work by comparing a data point to the statistics gathered from all the available data points to make a decision about whether the data point is an outlier. Importantly, the outliers themselves are used to compute the statistic, which under the right circumstances, can bias the resulting statistics enough to prevent the detection of all outliers. To get around this problem, we chose to develop an iterative algorithm to recompute the statistics each time an outlier is removed. At each iteration, we remove only the most prominent outlier and

recompute the statistics, iterating until no further outlier can be detected. While this approach is more accurate in the computation of the statistics due to minimizing the contribution of outliers in the statistics, it can be computationally more expensive due to the number of the iterations.

The IQR method is a two-sided outlier rejection technique and identifies both too low and too high outliers. However, for some features, it is more appropriate to have a single-sided rejection rule. Thus, we also developed and used the single-sided rejection versions of the IQR method as needed. For high frequency noise, Euclidean distance, and diastolic variance, we only rejected too high outliers as low values for these features are desired by definition. For cross-correlation, we only rejected low outliers as high correlation is desired by definition. For beat length, however, we used the two-sided version as both low and high beat lengths can be inappropriate.

The final rejected beats were the union of beats identified by each of the features since each feature inspected a different physiological aspect of the beats. While there was overlap between the features, this is a simple and the least conservative approach. In the future, it is of interest to study more complex feature voting algorithms.

Since the proposed beat rejection mechanism is based on outlier detection, it can possibly break if the overall data quality is extremely poor. For example, the algorithm will likely pass the poor beats if they are the majority. As such, the TCD signal is of such low quality to begin with that any attempt at analysis should be avoided at the first place. Future work is needed to filter out the TCD scans that do not retain a minimum absolute signal quality.

The IIQR algorithm, in the current implementation, is *not* designed for real-time implementation in a target processor. The main application is to refine the TCD data at the beat level for other post hoc analyses and diagnostic algorithms that are sensitive to data quality.

E. Other Algorithms

There are a few other relevant algorithms that can be used to assess and improve the quality of the TCD signal. Gunn's recent technique requires an auxiliary signal [23]. They have developed a system identification algorithm and trained a model between *Arterial Blood Pressure* (ABP) signal and CBFV which detected and corrected for artifacts in CBFV. It assumed that artifacts are sparse events that increase the complexity of the dynamics of the system between ABP and CBFV. This approach requires measurements of additional signals i.e. ABP. Moreover, the identified model will presumably need to be tuned for each TCD recording to account for intersubject and intrasubject variabilities and differences in the waveform between health and disease.

Similar to TCD recordings, Intracranial Pressure (ICP) is a triphasic pulsatile waveform that can be contaminated with noise and artifacts. A few methods have been developed to improve the ICP/CBFV data quality as well as the extraction of additional features [24], [25]. As one example, the MOrphological Clustering nd Analysis of Intracranial Pulses (MOCAIP) algo-

rithm utilizes a five-step process, including a hierarchical clustering method to construct a representative non-artifactual beat known as a dominant pulse [12], [24], [26]–[28]. Following the clustering of the pulsatile beats, the dominant pulse is compared to a pulse library to determine if it is a spurious pulse and sub-peak landmarks are identified for post-hoc feature extraction. Although this work was originally used for ICP pulses, it was later adapted for the analysis of CBFV waveforms [12]. In comparison to the IIQR algorithm developed in this work, there are a few advantages and disadvantages. First, the identified clusters in the MOCAIP algorithm could be meaningful, e.g. cluster with pathologic pulse waveform or high-noise cluster, etc. Second, the use of the validated pulse library in MOCAIP provides insight into the quality of the signal collected, e.g. if there is no signal, the MOCAIP algorithm will not return valid results. However, it is this pulse library and post hoc analysis that does not allow MOCAIP to be used in a near real-time use, such as within the clinical environment for fast diagnostic purposes where the IIQR could be. Furthermore, the IIQR algorithm is more general in the sense that we can label and reject beats based on a select feature (five proposed in this paper) whereas MOCAIP is based on Euclidean distance between the beats.

F. Clinical Implication

This paper presents a tool that can be used as a pre-filtering stage to improve data quality for assessment of many cerebrovascular pathologies including *Acute Ischemic Stroke* (AIS), [3], [29], [30]. The limited number of available effective treatments and interventions are highly time-sensitive. Therefore, to avoid unnecessary delays, pre-hospital diagnostic tools are critical for the fast triage and transfer of patients to specialized and appropriate treatment centers. Because TCD measures blood flow through the cerebral vasculature directly in a fast, portable, and noninvasive way, it is a strong candidate technology for pre-hospital diagnosis and assessment [31]. Furthermore, analysis of subtle, clinically significant changes in CBFV waveform morphology has shown promise for potentially discovering new biomarkers and developing machine learning algorithms to aid in performing objective diagnostic assessments for diseases which affect the cerebrovasculature [14], [15], [32].

Nevertheless, difficulties associated with acquiring high quality TCD data have hindered its adoption for this purpose. Due to the ability of the algorithm presented here to reliably assess the quality of local segments of data based on objective criteria, we have taken yet another step forward toward alleviating the obstacles preventing the adoption of TCD as a pre-hospital diagnostic tool. In being able to quickly and accurately extract the high quality portions of signal from a TCD recording, this algorithm may be able to aid in diagnostic assessments for conditions such as AIS as well as many other cerebrovascular conditions, which often rely heavily on obtaining an accurate, clean picture of the beat waveform present in the signal. Significantly, the algorithm was able to closely replicate the performance of a TCD expert in terms of measured clinical parameters, yet it has none of the disadvantages normally associated with manual inspection.

REFERENCES

[1] M. Y. Kassab *et al.*, "Transcranial Doppler: An introduction for primary care physicians," *J. Amer. Board Family Med.*, vol. 20, no. 1, pp. 65–71, 2007.

[2] R. Aaslid, T.-M. Markwalder, and H. Nornes, "Noninvasive transcranial Doppler ultrasound recording of flow velocity in basal cerebral arteries," *J. Neurosurg.*, vol. 57, no. 6, pp. 769–774, 1982.

[3] A. M. Demchuk *et al.*, "Accuracy and criteria for localizing arterial occlusion with transcranial Doppler," *J. Neuroimag.*, vol. 10, no. 1, pp. 1–12, 2000.

[4] S. G. Thorpe *et al.*, "Velocity curvature index: A novel diagnostic biomarker for large vessel occlusion," *Transl. Stroke Res.*, pp. 1–10, Oct. 2018. [Online]. Available: https://doi.org/10.1007/s12975-018-0667-2

[5] S. Kim *et al.*, "Inter-subject correlation exists between morphological metrics of cerebral blood velocity and intracranial pressure pulses," *Neurocritical Care*, vol. 14, no. 2, pp. 229–237, 2011.

[6] P. Raboel *et al.*, "Intracranial pressure monitoring: invasive versus non-invasive methods—A review," *Critical Care Res. Pract.*, vol. 2012, 2012, Art. no. 950393.

[7] R. J. Adams, "TCD in sickle cell disease: an important and useful test," *Pediatric Radiol.*, vol. 35, no. 3, pp. 229–234, 2005.

[8] C. M. Thibeault *et al.*, "A cross-sectional study on cerebral hemodynamics after mild traumatic brain injury in a pediatric population," *Frontiers Neurol.*, vol. 9, 2018, Art. no. 200.

[9] H. A. Keage *et al.*, "Cerebrovascular function in aging and dementia: A systematic review of transcranial Doppler studies," *Dementia Geriatric Cogn. Disorders Extra*, vol. 2, no. 1, pp. 258–270, 2012.

[10] H. Markus, A. Pereira, and G. Cloud, *Stroke Medicine*. London, U.K.: Oxford Univ. Press, 2016.

[11] S. Purkayastha and F. Sorond, "Transcranial Doppler ultrasound: technique and application," *Seminars Neurol.*, vol. 32, no. 4, pp. 411–420, 2012.

[12] S. Kim *et al.*, "Noninvasive intracranial hypertension detection utilizing semisupervised learning," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 4, pp. 1126–1133, Apr. 2013.

[13] J.-X. Wang, X. Hu, and S. C. Shadden, "Data-augmented modeling of intracranial pressure," *Ann. Biomed. Eng.*, vol. 47, no. 3, pp. 714–730, Mar. 2019.

[14] A. Kurji *et al.*, "Differences between middle cerebral artery blood velocity waveforms of young and postmenopausal women," *Menopause*, vol. 13, no. 2, pp. 303–313, 2006.

[15] S. Aggarwal *et al.*, "Noninvasive monitoring of cerebral perfusion pressure in patients with acute liver failure using transcranial Doppler ultrasonography," *Liver Transplantation*, vol. 14, no. 7, pp. 1048–1057, 2008.

[16] N. Canac *et al.*, "Algorithm for reliable detection of beat onsets in cerebral blood flow velocity signals," *BioRxiv*, 2018. [Online]. Available: https://www.biorxiv.org/content/early/2018/07/09/364356

[17] I. Ben-Gal, "Outlier detection," in *Data Mining and Knowledge Discovery Handbook*. New York, NY, USA: Springer, 2005, pp. 131–146.

[18] D. C. Hoaglin, B. Iglewicz, and J. W. Tukey, "Performance of some resistant rules for outlier labeling," *J. Amer. Statist. Assoc.*, vol. 81, no. 396, pp. 991–999, 1986.

[19] A. J. Viera *et al.*, "Understanding interobserver agreement: The kappa statistic," *Family Med.*, vol. 37, no. 5, pp. 360–363, 2005.

[20] J. W. Tukey, *Exploratory Data Analysis*. Reading, MA, USA: Addison-Wesley, 1977.

[21] B. Iglewicz and S. Banerjee, "A simple univariate outlier identification procedure," in *Proc. Annu. Meeting Amer. Statist. Assoc.*, 2001, pp. 5–9.

[22] D. C. Hoaglin and B. Iglewicz, "Fine-tuning some resistant rules for outlier labeling," *J. Amer. Statist. Assoc.*, vol. 82, no. 400, pp. 1147–1149, 1987.

[23] C. A. Gunn, "Convex optimization methods for system identification with applications to noninvasive intracranial pressure estimation," Ph.D. dissertation, Elect. Comput. Eng. Dept., Univ. California, Los Angeles, Los Angeles, CA, USA, 2018.

[24] X. Hu *et al.*, "Morphological clustering and analysis of continuous intracranial pressure," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 3, pp. 696–705, Mar. 2009.

[25] P. K. Eide, "A new method for processing of continuous intracranial pressure signals," *Med. Eng. Phys.*, vol. 28, no. 6, pp. 579–587, 2006.

[26] R. Hamilton *et al.*, "Forecasting intracranial pressure elevation using pulse waveform morphology," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2009, pp. 4331–4334.

[27] N. R. Gonzalez *et al.*, "Cerebral hemodynamic and metabolic effects of remote ischemic preconditioning in patients with subarachnoid hemor-rhage," in *Cerebral Vasospasm: Neurovascular Events After Subarachnoid Hemorrhage*. New York, NY, USA: Springer, 2013, pp. 193–198.

[28] X. Hu *et al.*, "Intracranial pressure pulse morphological features improved detection of decreased cerebral blood flow," *Physiol. Meas.*, vol. 31, no. 5, pp. 679–695, 2010.

[29] S. G. Thorpe *et al.*, "Decision criteria for large vessel occlusion using transcranial Doppler waveform morphology," *Frontiers Neurol.*, vol. 9, 2018, Art. no. 847.

[30] G. Tsivgoulis *et al.*, "Validation of transcranial Doppler with computed tomography angiography in acute cerebral ischemia," *Stroke*, vol. 38, no. 4, pp. 1245–1249, 2007.

[31] A. M. Demchuk *et al.*, "Specific transcranial Doppler flow findings related to the presence and site of arterial occlusion," *Stroke*, vol. 31, no. 1, pp. 140–146, 2000.

[32] C. J. Lockhart *et al.*, "Nitric oxide modulation of ophthalmic artery blood flow velocity waveform morphology in healthy volunteers," *Clin. Sci.*, vol. 111, no. 1, pp. 47–52, 2006.