Sequential Emboli Detection from Ultrasound Outpatient Data

Blaise Kevin Guépié, Matthieu Martin, Victor Lacrosaz, Marilys Almar, Benoît Guibert, Philippe Delachartre

***Abstract*—This paper addresses the detection of emboli from signals acquired with a new miniaturized and portable transcranial Doppler ultrasound device. The use of this device enables outpatient monitoring but increases the number of artifacts. These artifacts usually come from the patient voice and motion and can be superimposed to emboli. For this reason and because of the scarcity of emboli compared to artifacts, reliably detect emboli is a challenging task. As an example, the 11809 s of signal used in this study contained 0.06 % of embolic events and 10.14 % of artifacts. Herein, we propose an automatic and sequential approach. The method is based on sequential determination of high intensity transient signal. We also define efficient features to describe emboli in the time frequency representation. On our database, the number of artifacts detected as emboli is divided by more than 10 compared to the other algorithms reported in the literature.**

***Index Terms*—Emboli detection, transcranial Doppler, ultrasound, time-frequency approach, artifact removal, outpatient data monitoring.**

I. Introduction

Emboli are solid or gaseous objects that circulate in the bloodstream until they become lodged in a blood vessel. The presence of emboli is related to the risk of strokes [1] and can cause body damage at several levels.

There are many techniques such as magnetic resonance imaging (MRI) and computed tomography (CT) [1] to detect emboli. Nevertheless, these techniques are expensive, invasive and they cannot be used to monitor the blood flow over a long period of time (half an hour to several hours). This last point is very important because emboli are usually events with a low occurrence. In this study we work with ultrasound data acquired with a portable transcranial Doppler technique. This technique is particularly interesting because it enables outpatient monitoring over several hours and is cheap and noninvasive compared to MRI and CT. On this type of data, emboli or artifacts are characterized by a high intensity transient signal (HITS). Thus, detecting emboli first requires the detection of HITSs, and then the discrimination of HITSs into emboli and artifacts.

Many authors have investigated the detection of emboli from transcranial Doppler ultrasound signals. In [2] and [3], [4], authors have worked on autoregressive modeling (AR) or short time Fourier transform (STFT) of acquired Doppler signals and, they use high order statistics to detect emboli. Other authors have also worked with STFT, fractional Fourier transforms or wavelet transforms and their derivatives [5], [6]. In papers [6], [7], [8], [9], [10], [11] and [12], [13], [14], [15], authors extracted features from large signal blocks in order to separate emboli and artifacts events. Herein, artifacts include high intensities due to the stochastic nature of the blood flow (or Doppler speckle).

Because of the scarcity of emboli, it is necessary to monitor the blood flow over a long period of time. Traditionally, monitoring is done in hospital setting. In this setting, few artifacts contaminate the signals acquired and the above-mentioned methods work relatively well. Today, thanks to the TCD-X device (Atys Medical, France) which is a new, miniaturized, mono-gate and portable device with a robotized probe, it is possible to monitor patients outside the hospital setting. Therefore, it reduces the hospital resources needed (rooms and medical staff) and more patient can be monitored. Nevertheless, because the patient are free, numerous artifacts (voice and motion) are superimposed on the acquired signals. These artifacts, like those appearing in other types of biomedical signals such as electroencephalography (EEG) and functional near-infrared spectroscopy [16], [17], affect signal interpretation and must be removed. The above-mentioned studies were designed for inpatient data which contains few artifacts. Therefore, they are not perfectly suited to our database which contains a large number of artifacts.

To the best of our knowledge, only [18] addresses emboli detection through artifact rejection for transcranial Doppler ultrasound outpatient data monitoring. It has proposed a procedure using parametric assumption (the spectral kurtosis approach) to detect a high-intensity transient signal (HITS), and a nonsequential threshold estimation (because estimating probability density function requires large blocks of data to work properly) to remove artifacts.

Detecting emboli from ultrasound outpatient data remains a challenging task. The main difficulty is related to the presence of numerous artifacts coming from patient motion or voice. These artifacts complicate the task of estimating nominal blood flow power because any part of signal has a high probability to contain an artifact. Moreover, After estimating the nominal blood flow power and then extracting HITSs, the presence of numerous artifacts also makes HITSs classification difficult. The difficulty comes from the fact that in many situations emboli and artifacts appear at the same time. This paper proposes an algorithm able to reliably detect emboli

B. Guépié is with Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, F-69621, Lyon, France and with Laboratoire de Modélisation et Sûreté des Systèmes, ICD, UMR 6281 CNRS, Université de Technologie de Troyes, France, e-mail: (blaise_kevin.guepie@utt.fr).

P. Delachartre is with Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, F-69621, Lyon, France, e-mail: (philippe.delachartre@creatis.insa-lyon.fr).

M. Almar is with Atys Medical 17, Parc d'Arbora 69510 Soucieu en Jarrest, France, email: (marilys.almar@atysmedical.com)

despite the presence of numerous artifacts (voice, motion) related to the use of a portable and miniaturized transcranial Doppler ultrasound system, which provides greater mobility for a patient during the monitoring. Our algorithm is sequential and it is also more reliable (no parametric assumption is used) than the ones presented in [10], [11], [18]. It is composed of three main parts. Firstly, an adaptive HITS-to-blood ratio (HBR) is calculated to detect HITSs. Secondly, eight features are calculated using the HITS signals, the region of the spectrogram corresponding to the HITS called the HITS region (see the spectrogram region between the green band Fig. 1) and the most nonsymmetrical part of the HITS region spectrogram which is called the region of interest (ROI, see the red square Fig. 1). Finally, the features are given to a classification algorithm to discriminate emboli and artifacts.

The major contributions of this paper with respect to the previous ones [7], [9], [10], [11], [18] are the following :

- A nonparametric and robust estimation of the HBR. The new approach does not make any assumption on the distribution of the spectrogram values. It also avoids the calculation of the maximum velocity of the blood flow.
- The definition of six additional features based on the spectrogram that improve the discrimination of emboli and artifact. They are able to describe complex situations such as an embolus superimposed on nonsymmetrical voice harmonics.
- A sequential emboli detection procedure. In the previous papers, large blocks of signal were needed to estimate an artifact removal threshold. Thus, it was not possible to realize the emboli detection sequentially. The processing steps of our algorithm such as HITS detection and feature extraction are sequentially achieved.

This paper is organized as follows. Section II describes the proposed approach. The experimental results and the discussion are given in Section III. Conclusions are drawn in Section IV.

II. Proposed Method

A. Procedure Description

As outlined in Fig. 1, the procedure used to detect emboli is hierarchical. Firstly, HBR is calculated in the time domain. When its value exceeds a prescribed threshold h_1 a HITS is detected. We considered as a HITSs the contiguous events for which the HBR exceeds h_1. Secondly, for the time ranges where HITSs are detected, STFT is calculated and the least symmetrical part of the spectrogram (ROI) is extracted. It corresponds to the part of the spectrogram that has the highest probability of being a HITS. Eight features are then extracted from the HITS signal, the ROI and the HITS region. These features are particularly interesting to make the difference between emboli and artifacts. They describe the ROI and compare the HITS region with the ROI. Finally, using the previous features, a classification is achieved and it can be decided which HITS is an embolus and which one is an artifact.

B. Sequential Calculation of the HITS-to-Blood Ratio

In this subsection, we propose a method to sequentially calculate the HBR. The idea behind HBR is to calculate the ratio of the signal instantaneous power to the blood flow average power. Because cardiac cycle has a periodicity approximately equal to 1 s, we calculated the average blood flow power over 1 s in order to reduce its variability.

Let us define $x(n)$ the value of the received demodulated complex signal x at the discrete time $n \in \mathbb{N}^*$, $(\xi_n)_{n\in\mathbb{N}^*}$ a sequence that contains values corresponding to blood flow. Let K be the number of observations in 1 s and $(PF_{avg_n})_{n>K}$ a sequence that contains the blood flow a verage p ower values. For all $n \geq K+1$, $PF_{avg}(n)$ is defined a s f ollows :

$$PF_{avg}(n) = \frac{1}{K}\sum_{\ell=n-K}^{n-1} |\xi(\ell)|^2 \tag{1}$$

Where $|\cdot|$ represents the modulus.

To start HBR calculation, we must initialize the value of the blood flow average power. The first K values of the signal constitutes the initialization buffer of the blood flow ($\xi(n) = x(n)$ for $n \in [|1, K|]$). Beginning at $n = K+1$ and for all the discrete time $n \geq K+1$, the following steps are realized, firstly $HBR(n)$ is calculated :

$$HBR(n) \quad := \quad 10\log_{10}\frac{|x(n)|^2}{PF_{avg}(n)} \tag{2}$$

Secondly a value is attributed to $\xi(n)$ depending on the value of $HBR(n)$:

$$\xi(n) \quad = \quad \begin{cases} x(n) \text{ if } HBR(n) < h_1 \\ \xi(n-K) \text{ if } HBR(n) \geq h_1 \end{cases} \tag{3}$$

Finally $PF_{avg}(n+1)$ is calculated in order to be able to calculate the value of $HBR(n+1)$:

$$PF_{avg}(n+1) = PF_{avg}(n) + \frac{|\xi(n)|^2}{K} - \frac{|\xi(n-K)|^2}{K} \tag{4}$$

The interpretation of Eq. (4) is the following :

- If $HBR(n) \geq h_1$, $PF_{avg}(n+1) = PF_{avg}(n)$ since $\xi(n) = \xi(n-K)$, the blood flow average power is not changed.
- If $HBR(n) < h_1$, the term $\frac{|\xi(n-K)|^2}{K}$ is eliminated from $PF_{avg}(n)$ and the term $\frac{|\xi(n)|^2}{K}$ is added, the blood flow average power is changed.

Fig. 2 displays the number of artifacts per minute as function of the percentage of emboli detected for different values of h_1 by using three subsets of patients. In Fig. 2, a satisfying tradeoff given by the Elbow method [19] is $h_1 = 3$dB for each subset of patients. Then, we set $h_1 = 3$dB the maximum tolerated HBR of blood flow power in the rest of the paper.

Let us know discuss about the effect of a bad initialization. Fig. 3 gives an example illustrating the HBR calculation. Fig. 3a) shows an embolus outside the initialization period. Fig. 3b) displays an embolus signal present during the initialization period. The same embolus is put in the penultimate cardiac cycle in order to give its HBR value. It can be observed in

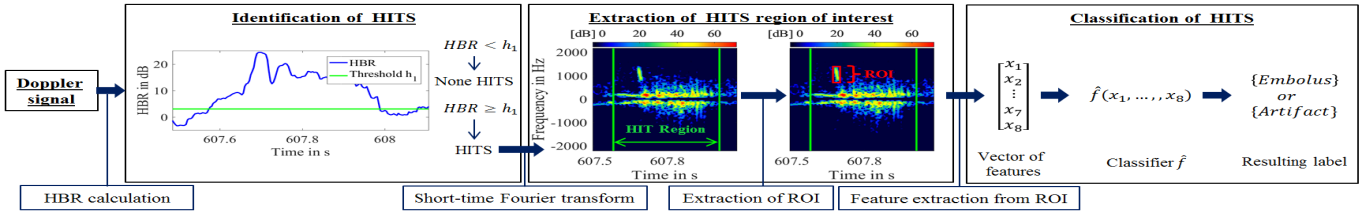

Fig. 1: Framework of the emboli detection process. The first part achieves HITS detection in the time domain. The second part concerns the features extraction from the HIT region and the ROI. The third part provides the HITS classification.

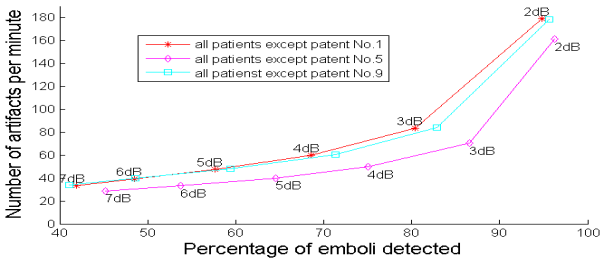

Fig. 2: Curves displaying the number of artifacts per minute as function of the percentage of emboli detected for three subsets of patients. The value $h_1 = 3$dB is a satisfying tradeoff using the Elbow method.

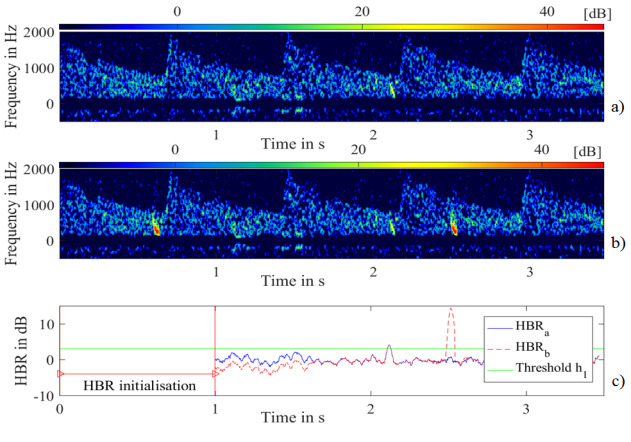

Fig. 3: a) no embolus is present during the initialization period, b) an embolus is present during the initialization period, c) HBR values over the time in both cases using our approach.

Fig. 3c) despite the presence of an embolus having high HBR value (approximatively 13dB) in the initialization period, the nominal value of the blood flow power is reliably estimated by our approach within less than 1 s.

C. Region of Interest in the Spectrogram

This part is very important to detect an embolus superimposed on an artifact in the time domain. According to [20], [21], emboli are unidirectional or nonsymmetrical in the spectrogram because they follow the same direction than the blood flow. It contrasts with several artifacts, particularly voice and motion, which are symmetrical. In this paper, the region of interest (ROI) is defined as the largest nonsymmetrical part belonging to the positive frequencies of the HITS spectrogram. The idea behind ROI is to separate emboli from symmetrical artifacts due to human activity such as voice and motion, even if they occur at the same time. In other words, the ROI is the part of the spectrogram that can potentially be an embolus. Three main cases can be observed in the HITS spectrogram.

- The current HITS is an embolus. Then, the extracted ROI is this embolus.
- The current HITS is an artifact. Then, the extracted ROI is the least symmetrical part of this artifact.
- The current HITS combines an embolus and an artifact. Then, the extracted ROI is the embolus.

To summarize, the ROI corresponds to an unique event (an embolus or an artifact) even if both are present in the spectrogram. The method to extract ROI from a HITS spectrogram is given by Algorithm 1.

Algorithm 1 ROI Extraction

1: Compute, from the current blood flow spectrogram, the ratio of the positive-frequency magnitude to the negative-frequency magnitude.
2: Set a threshold h_2 as the maximum of the previous ratio.
3: Compute, from the current HITS region spectrogram, the spectrogram ratio of the positive-frequency magnitude to the negative-frequency magnitude.
4: Threshold the previous ratio spectrogram by h_2 and define the largest part as the ROI.

D. Features extraction

To classify HITS into emboli and artifacts, eight features are extracted from the HITS signal, the ROI and the HITS region. As we do not need other parts of the signal to calculate the features, it can be done sequentially. The rationale and definition of these features are given below.

a) Duration of HITS (D) : this feature is equal to the duration of the event : HBR greater than h_1 (see feature $x_1 = D$ on Fig. 4). When this duration is short, the HITS is likely an artifact, i.e., an outlier due to the stochastic nature of the ultrasound scattering of the blood flow (Fig. 4f and Table I). If the duration is long, we can not prejudge the type of HITS. It can be an artifact (Fig. 4a and Table I) or an embolus superimposed on an artifact (Fig. 4e and Table I).

b) HBR_{avg} : this feature is calculated in the time-domain. It corresponds to the average value of the HBR calculated over the HITS signal (see section II-B). Its value is

TABLE I: Value of features extracted from HITSs in Fig. 4.

		Feature name							
		Time domain		Time-frequency domain					
HITS type	Fig.	$D(ms)$	$HBR_{avg}(dB)$	$D_{ROI}(ms)$	$HBR_{ROI}(dB)$	F_{max} (Hz)	F_{spr} (Hz)	R_{TF} Hz/ms	R_{siz}
Artifact	4a	76.62	9.46	65.68	12.11	308.32	205.55	3.13	1.18
Artifact	4b	288.26	13	7.3	-0.39	376.84	137.03	18.78	5.5
Artifact	4d	456.1	13.65	36.49	4.24	479.61	137.03	3.76	0.03
Artifact	4f	29.19	4.93	18.24	3.37	993.48	171.29	9.39	2.11
Embolus	4c	36.49	10.23	14.6	11.48	1644.38	685.16	46.94	1.31
Embolus	4e	120.41	9.99	25.54	11.6	1096.25	274.06	10.73	1.59

R_{siz} has no unit of measurement. It is defined, in the spectrogram, as the ratio of the HITS ROI size to the HITS size.

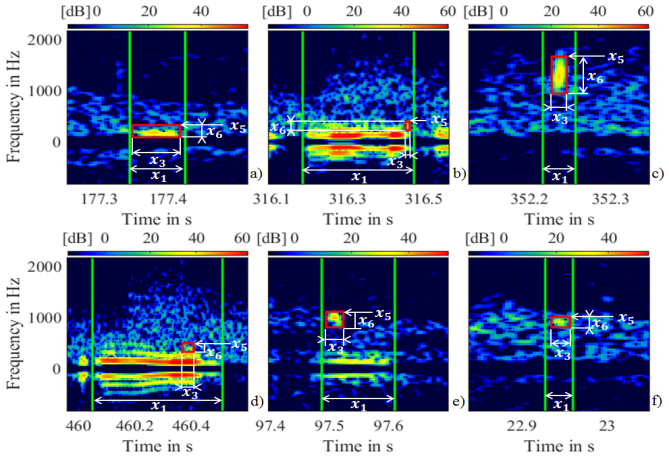

Fig. 4: Representation of the features for different types of HITSs. a) a nonsymmetrical artifact, b) a symmetrical artifact whose ROI corresponds to blood flow, c) an embolus, d) a symmetrical artifact whose ROI corresponds to the less symmetrical part of the artifact, e) an embolus superimposed on an artifact, f) a blood flow outlier. Parameters shown: $x_1 = D$, $x_3 = D_{ROI}$, $x_5 = F_{max}$, $x_6 = F_{spr}$, $R_{TF} = x_6/x_3$.

often high for symmetrical artifacts (Fig. 4b and Table I) and emboli (Fig. 4c and Table I). This characteristic can be used to eliminate many blood flow outliers (Fig. 4f and Table I). For nonsymmetrical HITSs, the probability to be an embolus increases as HBR_{avg} increases.

c) Duration of the HITS ROI (D_{ROI}) : this feature is equal to the duration of the HITS ROI (see feature $x_3 = D_{ROI}$ on Fig. 4). This value is particularly useful to analyze superimposed HITSs. As previously mentioned in section II-C, in this situation, the ROI corresponds to the part of the spectrogram with the highest probability of being an embolus. Therefore, contrary to D, D_{ROI} is only designed to characterize events that look like embolus. Thus, when D_{ROI} is higher or lower than its common value for emboli, the HITS is potentially an artifact [20] (see respectively Fig. 4a, Fig. 4b and Table I).

d) HBR of the HITS ROI (HBR_{ROI}) : this feature is defined as the ratio of the maximum magnitude in the ROI to the maximum magnitude of the blood flow. It is used to remove symmetrical artifacts. In fact, when a HITS is a symmetrical artifact, its associated ROI almost always corresponds to blood flow (Fig. 4b and Table I). In this case, HBR_{ROI} is lower than its common value for emboli.

e) Maximum frequency of the HITS ROI (F_{max}) : this parameter corresponds to the maximum frequency inside the ROI (see feature $x_5 = F_{max}$ on Fig. 4). It is used to eliminate artifacts corresponding to human activity such as voice (Fig. 4d and Table I) and motion. On the spectrogram, these artifacts appear traditionally at a lower frequency than emboli.

f) Frequency spreading of HITS ROI (F_{spr}) : this feature is equal to the difference between the maximum and the minimum frequency of the ROI $x_6 = F_{spr}$. Its goal is to reliably identify blood flow outliers. Blood flow outliers can appear at any frequency. Therefore, F_{max} can also be high for several blood flow outliers. Nevertheless, the frequency spreading of blood flow outliers (Fig. 4f and Table I) is lower than the frequency spreading of emboli (Fig. 4c and Table I).

g) Ratio of frequency spreading in ROI over ROI duration (R_{TF}) : the idea behind this feature is to use the shape of the ROI. F_{spr} and D_{ROI} are respectively equal to the height and width of the ROI. Therefore, the parameter $R_{TF} = F_{spr}/D_{ROI}$ gives information about the ROI shape. HITSs for which the parameter R_{TF} is lower or higher than its common value for emboli (Fig. 4c and Table I) can be considered as artifacts (Fig. 4a and Table I).

h) Ratio of HITS ROI size over HITS size in spectrogram (R_{siz}) : this parameter is designed to eliminate partially symmetrical artifacts. The ROI of a partially symmetrical artifact is traditionally a small part of this artifact (Fig. 4d and Table I). Therefore, for this kind of artifacts R_{siz} tends toward zero.

E. Feature selection

The previous section outlined eight features to make the difference between artifacts and emboli. Nevertheless, it is important to know if all the parameters are useful, i.e., if a subset of parameters contains as much information as all the parameters to describe HITSs.

Three main sets of algorithms can be used to select features with respect to their importance in class prediction : Wrapper, Embedded and Filter. Wrapper and Embedded algorithms carried out a procedure of classification during feature selection. Therefore, they are time-consuming depending on the classifier used. Herein, to reduce computational time, we used Filter algorithms called minimal-redundancy-maximal-relevance criterion (mRMR) and Relief-F [22]. These algorithms are both based on the predictors intrinsic characteristics for the ranking task. Therefore, the number of features chosen after the ranking task depends on the classification algorithm used in Section II-F.

F. Machine Learning Algorithms

The three supervised learning algorithms we used to discriminate between emboli and artifacts are described below.

1) Support Vector Machine (SVM): this algorithm is traditionally used to linearly separate binary-labeled data. Its principle is to maximize the separation margin, the margin being the distance between the closest observations of the two groups. To address nonlinearly separable data, the trick is to project them onto high-dimensional space. This projection usually enables the linear separation of the data. In order to deal with imbalanced classes, as in our case, the SVM can be weighted [23].

2) Naive Bayes (NB): this algorithm is based on the Bayes theorem. It considers that predictors are independent given an observation class member. Nevertheless, even if this strong assumption is not true, the NB classifier can work very well [24].

3) Decision Tree (DT) [25]: This algorithm is an iterative method that partition data. It realizes successive binary test, using data characteristics, in order to build homogeneous classes. DT rules are easy to understand because they can be shown graphically.

III. Experimental Results and Discussion

A. Data Acquisition

Our database was composed of twelve patients with a carotid stenosis. Carotid stenosis is a partial obstruction of the carotid artery caused by atheromatous deposits (fat or calcareous) on the wall of the artery. It increases the risk of stroke because atheromatous deposits can break off and block blood circulation into the middle cerebral artery. This event is very serious knowing that 80% of the cerebral blood flow circulates through this artery.

Patients were monitored with the TCD-X (Atys Medical, France) device. The pulse repetition frequency (PRF) was set between 4 and 5 kHz. The emitted waves had a central frequency equal to 1.5 MHz and lasted 10 μs. A quadrature demodulation was applied on the back-scattered signal giving audible two-channel signals.

HITSs were labeled on the acquired signals by two experts. They used the definition of embolus given in [20]. The steps of the labeling procedure are the following. Firstly, the HBR and D_{ROI} parameters of HITS are checked. If their values are respectively lower than 3 dB or greater than 300 ms, the HITS is labeled as an artifact. Then, the expert looked for symmetry in the spectrogram and listened to the audio data. If no symmetry was observable in the spectrogram and if a snap, chirp, or moan sound was heard on the audio data, the HITS was labeled as an embolus. If not, the HITS was labeled as an artifact. Finally, the second expert confirmed or refuted the HITSs labeled as emboli by the first expert. Table II gives the number of emboli found for each patient.

B. Experimental Results

In this study, we used the following parameters. The STFT was carried out with a 128 points Blackman window and had a 91% overlap. The DC component of the signals was eliminated with a fourth-order butterworth highpass filter with a cut-off frequency equal to 150 Hz. A Gaussian kernel was used for SVM. Moreover, its kernel parameter pair and its constant of tolerance were searched in a grid $\left[10^{-3}, 10^{3}\right] \times \left[10^{-3}, 10^{3}\right]$ containing 49 equidistant pairs. DT used Gini impurity.

1) Feature selection procedure:

Table III gives the combined score of the proposed features. This score was obtained as follows. A training set was associated to each patient. This training set was composed of all the other patients. Then mRMR and Relief-F independently established the rank of the features from each patient's training set. The addition of the attributed ranks for both methods and all patients gave the combined score. The smaller the combined score the more informative the feature is to classify HITS.

After ordering the features, we sought if all of them were useful to discriminate between emboli and artifacts. For each patient, the same training set was used and the three classification algorithms were independently trained on eight feature subsets. The feature subsets were composed of the n first features for $n = 1, \cdots, 8$.

Table IV gives the average misclassification error of each algorithm. The DT and NB algorithms had the best classification performance when the first six features were used. The SVM algorithm classification results were optimal with the eight features.

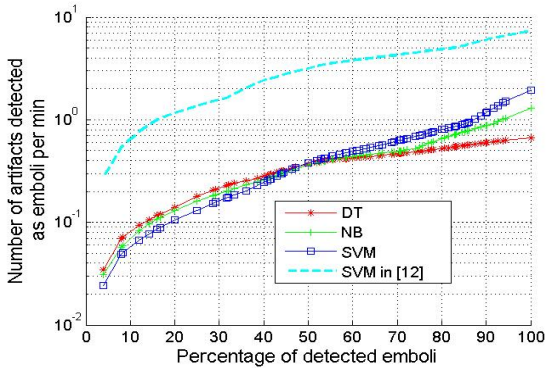

Fig. 5: Comparison of the algorithms performance : Decision Tree (DT), Naive Bayes (NB) and Support Vector Machine (SVM).

2) Comparison of Machine Learning Algorithms:

Figure 5 shows the curves of the detected emboli percentage versus the number of artifacts detected as emboli per minute. The algorithms presented in this paper are better than the one introduced in [18]. Our new method always gives a lower number of artifacts for a same percentage of detected emboli.

C. Discussion

The fact that NB and DT algorithms have optimum performance using six parameters out of eight can be explained as follows. The parameter HBR_{avg} is expected to enable the elimination of nonsymmetrical artifacts with low intensity such as blood flow outliers. Nevertheless, the database

TABLE II: Database description.

Patients	$No.1$	$No.2$	$No.3$	$No.4$	$No.5$	$No.6$	$No.7$	$No.8$	$No.9$	$No.10$	$No.11$	$No.12$
Signal duration (s)	323	480	111	157	1620	1121	3624	1100	601	300	1700	672
Number of emboli	65	16	27	36	25	9	5	7	41	12	13	7

TABLE III: Features ranking.

		Feature name							
		HBR_{ROI}	R_{TF}	F_{max}	F_{spr}	D_{ROI}	R_{siz}	D	HBR_{avg}
mRMR and Relief-F	Combined feature score	**24**	52	78	86	122	142	179	181
	Rank	1	2	3	4	5	6	7	8

TABLE IV: Error for each subset of features.

		Subset of the n most informative features							
		$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$
Average misclassification error	SVM	0.313	0.209	0.164	0.150	0.144	0.09	0.087	**0.081**
	DT	0.205	0.154	0.108	0.101	0.08	**0.069**	0.073	0.074
	NB	0.35	0.215	0.189	0.211	0.162	**0.111**	0.128	0.123

TABLE V: Features values of the three misclassified HITSs in Fig. 6.

			Feature name							
			Time domain		Time-frequency domain					
Fig.	HITS type	Prediction	$D(ms)$	$HBR_{avg}(dB)$	$D_{ROI}(ms)$	$HBR_{ROI}(dB)$	F_{max} (Hz)	F_{spr} (Hz)	R_{TF} Hz/ms	R_{siz}
6a)	Embolus	Artifact	32.84	5.41	10.95	2.26	719.41	205.55	18.78	1.91
6b)	Artifact	Embolus	58.38	8.54	54.73	8.81	548.13	376.84	6.89	1.63
6c)	Embolus mislabeled	Embolus	784.49	14.04	18.24	7.77	1541.6	376.84	20.66	1.65

R_{siz} has no unit of measurement. It is defined, in the spectrogram, as the ratio of the HITS ROI size to the HITS size.

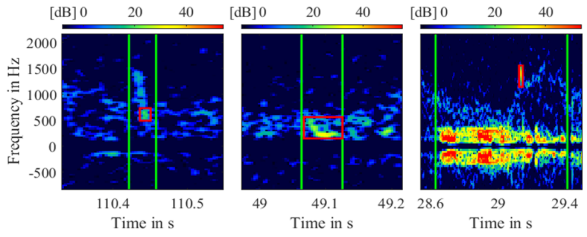

Fig. 6: Three misclassified HITSs. a) an embolus not detected, b) an artifact classified as an embolus, c) a HITS labeled as an artifact by the experts but that is undoubtedly an embolus.

contained few nonsymmetrical artifacts compared to emboli and symmetrical artifacts. Therefore, this parameter contains little information to classify HITS. Moreover, for most emboli (i.e., which are not superimposed on artifacts), HBR_{avg} is highly correlated to HBR_{ROI} because in the presence of an embolus, ROI and embolus region are the same. As a result, for NB and DT algorithms, information provided by HBR_{ROI} is useless, it is already given by HBR_{avg}. However, the information provided by HBR_{ROI} is useful for the SVM algorithm because its performance is less sensitive to correlated subsets of features. The fact that D_{ROI} is useless for DT and NB algorithms can be explained the same way. For nonsymmetrical artifact and emboli, D_{ROI} is highly correlated with D.

No matter which classification algorithm was selected among SVM, NB and DT, it gave better performance than the algorithms presented in [10], [11], [18]. This reflects the importance of the new processing introduced in this paper. Moreover, the new features are able to describe more complex situations than in the previous papers, including the superposition of an embolus and artifacts. In comparison, both features used in [18], i.e., R_{TF} and HBR_{avg}, were respectively the second and the last most informative features in this study.

Figure 5 shows that when the percentage of detected emboli is 90%, the previous paper, which had up to that point the best result among the other outpatient emboli detection algorithms, gives six artifacts per minute. In this study, using the DT, NB and SVM algorithms, we obtained respectively 0.6, 0.88 and 1.17 artifacts per minute for the same percentage of emboli detected. Compared to [10], [11], [18], the gain is considerable, the number of artifact detected per minutes is divided by more than 10 using the DT separator.

It is not obvious to choose the best classifier among SVM, NB and DT. On the one hand, if we want to maximize the percentage of emboli detected (i.e, near 100%) and minimize the number of artifacts, DT is the best. On the other hand, if we want to obtain the lowest number of artifacts (i.e, near 0 per min) and the highest number of emboli, SVM is the best. Nevertheless, if we consider the medical application, the aim of the algorithm is to detect all the emboli with the minimum number of artifacts. Therefore, DT is superior to NB which is superior to SVM.

The algorithm presented in this paper also differs from [10], [11], [18] because all its steps from the detection of HITSs up to the ROI features extraction are sequential. As described in Section II.B, a signal buffer updated at each sampling time is the only thing required to detect an embolus and extract the features used by the classification algorithm.

Let us finally discuss misclassified HITSs. Figure 4 shows three different cases. On Figure 6a, we can see an embolus classified as an artifact. If we look at the values of the associated features in Table V, we can conclude that it is a small or micro-embolus ($HBR_{ROI} < 3\,dB$). It is often tricky to differentiate micro-emboli from artifacts so it could be an explanation of this false negative. On Figure 6b, an artifact classified as an embolus can be seen. The values of

the associated features presented in Table V are close to the values that characterize an emboli. This is probably why it was a false positive. The major difference between this HITSs and an emboli is that its main frequency is low. Its sound is not the snap, chirp, or moan characterizing an embolus as mentioned in [20]. Figure 6c) shows a HITS classified as an artifact by the human experts. This HITS is probably classified as an artifact because it is superimposed on a voice artifact and almost outside of the part of the spectrogram corresponding to the blood flow. The human experts cannot hear the characteristics sound of an embolus. However, the other characteristics of an embolus given in [20] which are, lasts less than 300 ms, is nonsymmetrical in spectrogram, has an HBR higher than 3 dB, are verified. This means that the definition of embolus given in [20] needs to be upgraded in order to improve emboli detection performance. We could introduce probabilistic labels in order to take into account uncertain cases.

IV. Conclusion

This paper presents an automatic emboli detection algorithm (i.e., no human intervention is required) designed for outpatient transcranial ultrasound data. The new features introduced are reliable to eliminate most artifacts intrinsically related to outpatient data. The algorithm proposed outperforms the actual state of the art. Tested on patients with carotid steneosis, the number of artifacts detected per minute was divided by 10. Our method is also sequential. Therefore, it opens the possibility to make device able to detect emboli online (during monitoring). In a future work, attention will be paid to uncertain emboli (i.e, emboli with probabilistic labeling) in order to take into account small or micro-emboli.

Acknowledgment

This study was funded by the ANR-13-LAB3-0006-01 LabCom AtysCrea and was supported by the LABEX CELYA (ANR-10-LABX-0060) and PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the "Investissements d'Avenir" program (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

References

[1] S. Wallace, G. Døhlen, H. Holmstrøm, C. Lund, and D. Russell, "Cerebral microemboli detection and differentiation during transcatheter closure of atrial septal defect in a paediatric population," *Cardiology in the Young*, vol. 25, pp. 237–244, 2 2015.

[2] M. Geryes, S. Ménigot, W. Hassan, M. Almar, B. Guibert, C. Gautier, J. Charara, and J. M. Girault, "A new energy detector of micro-emboli using a time-varying threshold," in *2015 International Conference on Advances in Biomedical Engineering (ICABME)*, Sept 2015, pp. 89–92.

[3] M. Geryes, S. Ménigot, W. Hassan, A. Mcheick, J. Charara, and J.-M. Girault, "Detection of doppler microembolic signals using high order statistics," *Computational and Mathematical Methods in Medicine*, vol. 2016, pp. 1 – 8, 2016.

[4] J. M. Girault, D. Kouame, A. Ouahabi, and F. Patat, "Micro-emboli detection: an ultrasound doppler signal processing viewpoint," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 11, pp. 1431–1439, Nov 2000.

[5] M. Gencer, G. Bilgin, and N. Aydin, "Embolic doppler ultrasound signal detection via fractional fourier transform," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, July 2013, pp. 3050–3053.

[6] P. Sombune, P. Phienphanich, S. Muengtaweepongsa, A. Ruamthanthong, and C. Tantibundhit, "Automated embolic signal detection using adaptive gain control and classification using anfis," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2016, pp. 3825–3828.

[7] Y. Chen and Y. Wang, "Doppler embolic signal detection using the adaptive wavelet packet basis and neurofuzzy classification," *Pattern Recogn. Lett.*, vol. 29, no. 10, pp. 1589–1595, Jul. 2008.

[8] S. Marvasti, D. Gillies, F. Marvasti, and H. S. Markus, "Online automated detection of cerebral embolic signals using a wavelet-based system," *Ultrasound in Medicine & Biology*, vol. 30, no. 5, pp. 647 – 653, 2004.

[9] G. Serbes and N. Aydin, "Denoising performance of modified dual-tree complex wavelet transform for processing quadrature embolic doppler signals," *Medical & biological engineering & computing*, vol. 52, no. 1, pp. 29–43, 2014.

[10] A. Karahoca and M. A. Tunga, "A polynomial based algorithm for detection of embolism," *Soft Computing*, vol. 19, no. 1, pp. 167–177, 2014.

[11] A. Karahoca, T. Kucur, and N. Aydin, "Data mining usage in emboli detection," in *Bio-inspired, Learning, and Intelligent Systems for Security, 2007. BLISS 2007. ECSIS Symposium on*, Aug 2007, pp. 159–162.

[12] P. Sombune, P. Phienphanich, S. Phuechpanpaisal, S. Muengtaweepongsa, A. Ruamthanthong, and C. Tantibundhit, "Automated embolic signal detection using deep convolutional neural network," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2017, pp. 3365–3368.

[13] N. Aydin, S. Padayachee, and H. S. Markus, "The use of the wavelet transform to describe embolic signals," *Ultrasound in Medicine & Biology*, vol. 25, no. 6, pp. 953 – 958, 1999.

[14] N. Aydin, F. Marvasti, and H. S. Markus, "Embolic doppler ultrasound signal detection using discrete wavelet transform," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, no. 2, pp. 182–190, June 2004.

[15] E. Roy, P. Abraham, S. Montrsor, M. Baudry, and J.-L. Saumet, "The narrow band hypothesis: An interesting approach for high-intensity transient signals (hits) detection," *Ultrasound in Medicine & Biology*, vol. 24, no. 3, pp. 375 – 382, 1998.

[16] K. T. Sweeney, H. Ayaz, T. E. Ward, M. Izzetoglu, S. F. McLoone, and B. Onaral, "A methodology for validating artifact removal techniques for physiological signals," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 5, pp. 918–926, Sept 2012.

[17] S. Abbaspour and A. Fallah, "Removing ecg artifact from the surface emg signal using adaptive subtraction technique," *Journal of Biomedical Physics & Engineering*, vol. 4, no. 1, p. 33, 2014.

[18] B. K. Guépié, B. Sciolla, F. Millioz, M. Almar, and P. Delachartre, "Discrimination between emboli and artifacts for outpatient transcranial doppler ultrasound data," *Medical & Biological Engineering & Computing*, pp. 1–11, 2017.

[19] F. A. Zeiler, J. Donnelly, D. Cardim, D. K. Menon, P. Smielewski, and M. Czosnyka, "Icp versus laser doppler cerebrovascular reactivity indices to assess brain autoregulatory capacity," *Neurocritical Care*, Oct 2017.

[20] Consensus Committee of the Ninth International Cerebral Hemodynamic Symposium, "Basic identification criteria of doppler microembolic signals," *Stroke*, vol. 26, no. 6, p. 1123, 1995.

[21] G. Serbes and N. Aydin, "Denoising performance of modified dual-tree complex wavelet transform for processing quadrature embolic doppler signals," *Medical & Biological Engineering & Computing*, vol. 52, no. 1, pp. 29–43, 2014.

[22] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A survey on evolutionary computation approaches to feature selection," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 4, pp. 606–626, Aug 2016.

[23] E. J. L. Pulgarin and J. I. S. Esmeral, "Data-driven fdi for wind farms using w-svm," in *2016 IEEE International Symposium on Intelligent Control (ISIC)*, Sept 2016, pp. 1–6.

[24] H. Zhang, "Exploring conditions for the optimality of naïve bayes," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 19, no. 02, pp. 183–198, 2005.

[25] M. Tayefi, H. Esmaeili, M. S. Karimian, A. A. Zadeh, M. Ebrahimi, M. Safarian, M. Nematy, S. M. R. Parizadeh, G. A. Ferns, and M. Ghayour-Mobarhan, "The application of a decision tree to establish the parameters associated with hypertension," *Computer Methods and Programs in Biomedicine*, vol. 139, pp. 83 – 91, 2017.