

Détection de micro-emboles par une approche de classification supervisée

Blaise Kévin GUÉPIÉ¹, Marilys ALMAR², Benoît GUIBERT², Philippe DELACHARTRE¹

¹ Université de Lyon; CREATIS; CNRS UMR5220; Inserm U1044; INSA-Lyon; Université Lyon 1, France

²Atys Medical, France

blaise-kevin.guepie@creatis.insa-lyon.fr, marilys.almar@atysmedical.com, benoit.guibert@atysmedical.com, philippe.delachartre@creatis.insa-lyon.fr

Résumé – La détection des micro-emboles dans l'artère cérébrale moyenne est un enjeu capital dans la mesure où ces micro-emboles sont un facteur prédominant d'accident vasculaire cérébral. Lorsqu'un appareil de Doppler transcrânien portable est utilisé pour le contrôle, la détection des micro-emboles se heurte à une difficulté majeure qui est la présence de nombreux artefacts dus notamment aux bruits de voix, de pas, d'origine électrique, etc. Dans cet article, une nouvelle approche de détection permettant de distinguer les micro-emboles des artefacts est proposée. Elle consiste à identifier les micro-emboles d'un patient test à partir d'une base de données contenant des micro-emboles et des artefacts d'autres patients.

Abstract – Detecting micro-emboli in the middle cerebral artery is a major challenge because these micro-emboli are related to the risk of stroke (cerebrovascular accident). When a transcranial Doppler portable device is used, the micro-emboli detection is difficult to achieve due to the presence of several artefacts such as voice, motion, of electrical origin, etc. In this paper, a new detection approach differentiating micro-emboli and artefacts is proposed. It consists in identifying a test patient micro-emboli from database containing other patients micro-emboli and artefacts.

1 Introduction et motivation

La détection de micro-emboles est un enjeu capital dans le domaine médical. Les micro-emboles conduisent parfois à l'obstruction des artères cérébrales [1]. Ce processus est l'une des causes majeures d'accidents vasculaires cérébraux (AVC). Détecter de façon fiable ces micro-emboles est donc une étape essentielle permettant une meilleure prise en charge des patients à risque.

Le Doppler transcrânien (TCD) est le système non-invasif généralement utilisé pour surveiller le flux sanguin et potentiellement détecter les emboles [2].

Plusieurs méthodes ont été proposées pour identifier les emboles de façon fiable à partir du TCD. La plupart sont basées sur une approche de traitement de signal. L'idée sous-jacente de ces méthodes est de définir une statistique et un seuil pour lequel toute valeur de la statistique dépassant le seuil équivaut à la présence d'un embole [3, 4]. Malheureusement ces méthodes sont peu fiables et donc inadaptées à la détection de micro-emboles [5].

Dans [5, 6], les auteurs proposent des détecteurs de micro-emboles à partir de la modélisation du signal en série tempo-

relle ou à partir de l'énergie du signal Doppler à chaque cycle cardiaque. Ces détecteurs ont été testés sur peu de signaux réels de micro-emboles.

Or, la présence aléatoire des micro-emboles dans le flux sanguin impose de surveiller ce dernier, c.à.d, de faire des acquisitions avec le TCD pendant un temps relativement long (plusieurs heures) afin d'espérer les identifier. Le signal alors acquis contient en plus des micro-emboles, de nombreux signaux transitoires de haute intensité (HITS) produit pas des artefacts de voix, de pas, d'origine électrique, etc.

Grâce à une approche de traitement d'image introduite dans [7], des variables permettant de mieux discriminer les artefacts et les micro-emboles ont été mis en évidence. A partir de ces variables, nous proposons dans ce document une approche de détection originale dans ce domaine basée sur une méthode de classification supervisée. L'idée principale est de se servir d'une base de données contenant des HITS (artefacts et micro-emboles) étiquetés de plusieurs patients comme données d'apprentissage et de construire un séparateur identifiant de façon fiable les micro-emboles d'un nouveau patient. De fait, cette approche créé des liens entre les HITS de différents patients alors que jusqu'ici cette idée n'avait pas été explorée dans une optique de détection de micro-emboles.

Ce travail est financé par l'ANR-13-LAB3-0006-01 LabCom AtysCrea et soutenu par le LABEX CeLyA (ANR-10-LABX-0060) de l'Université de Lyon dans le cadre du programme " Investissements d'Avenir" (ANR-11-IDEX-0007) piloté par l'Agence Nationale de Recherche (ANR)

2 Matériel de contrôle

L'effet Doppler traduit le changement de fréquence d'une onde ultrasonore réfléchiée par une cible en mouvement. La vitesse de la cible se calcule par l'expression [7] :

$$v = \frac{f_D c}{2f_0 \cos(\theta_D)} \quad (1)$$

où f_D est la fréquence de décalage Doppler, f_0 est la fréquence de l'onde émise, c est la vitesse de propagation de l'onde sonore (ou célérité du son) et θ_D est l'angle entre l'insonation et la direction de mouvement de la cible.

Le Doppler transcranien (TCD) est généralement utilisé pour détecter des micro-embolies. Sa technique est basée sur l'effet Doppler. La vitesse du flux sanguin est évaluée à partir d'un système ultrasons. La détection de micro-embolies nécessite d'insonifier l'artère cérébrale moyenne par un faisceau d'ultrasons émis à travers une région mince de l'os au dessus de l'oreille [2]. Les ondes émises ont une fréquence centrale comprise entre 1.5 et 2 MHz et ont une durée de 10 μs . La fréquence de répétition des impulsions (PRF) est comprise entre 4 et 12 kHz. A partir d'une profondeur choisie, les signaux ultrasonores réfléchis sont démodulés à la fréquence du signal émis par une démodulation en quadrature pour obtenir le signal Doppler audio complexe.

3 Extraction des caractéristiques

La définition exacte d'un micro-embolie n'est pas précise à ce jour. Mais, la 9^e conférence internationale de l'hémodynamique cérébrale a donné quelques caractéristiques permettant son identification :

- la durée du signal est inférieure à 300 ms ;
- l'amplitude du signal est supérieure de 3 db au signal de référence du flux sanguin (signal nominal) ;
- la vitesse est unidirectionnelle dans le sonogramme du Doppler ;
- le signal audio est caractéristique.

L'apparition d'un micro-embolie est un événement aléatoire qui est peu fréquent dans la plupart des enregistrements. Pour espérer détecter ces événements, il est indispensable de réaliser une surveillance longue durée de l'artère cérébrale moyenne avec un TCD. Un nouveau dispositif portable de TCD a été développé dans ce but.

Cependant, les acquisitions du TCD portable font souvent apparaître des artefacts qui se traduisent par la superposition de signaux transitoires de haute intensité (HITS) sur les données collectées. Ces artefacts sont des bruits de voix, de pas, d'origine électrique, etc. Dès lors, détecter de façon fiable des micro-embolies signifie extraire des paramètres caractérisant les micro-embolies et les artefacts.

Les auteurs de l'article [7] ont extrait certains de ces paramètres à travers l'étude de signaux réels de TCD portable. Notre procédure d'extraction de paramètres caractéristiques est une modification de celle de cet article. Elle est décrite dans les

lignes suivantes. Premièrement, une image est créée grâce à la transformation dans le plan temps-fréquence du signal audio Doppler. Les HITS sont identifiés par une approche de traitement d'image et une boîte englobante est réalisée pour chacun d'eux. Ensuite, pour chaque HITS, le vecteur de paramètres extraits est le suivant :

- I_{max} : l'intensité max de la boîte englobante ;
- V_{max} : la vitesse max de la boîte englobante ;
- V_{min} : la vitesse min de la boîte englobante ;
- D : la durée de la boîte englobante ;
- R : le ratio (vitesse max - vitesse min)/ durée ;
- IR_{max} : l'intensité max dans le reflux.

L'intensité max dans le reflux IR_{max} correspond à l'intensité max dans la portion d'image dans le reflux comprise entre l'instant d'apparition et de disparition du HITS. Les détails du calcul de ces paramètres peuvent être consultés dans [7].

Les paramètres précédents diffèrent fortement d'un patient à un autre. Dans l'approche de classification supervisée envisagée, c'est à dire, afin de classifier les HITS d'un nouveau patient à partir d'une base de données de HITS d'autres patients, il convient de normaliser ces paramètres. Pour ce faire, la vitesse moyenne du flux sanguin V_{moy} , l'intensité moyenne dans le flux I_{moy} et l'intensité moyenne dans le reflux IR_{moy} sont déterminées. Le vecteur de paramètres caractéristiques retenu pour la classification est alors

$$\left[\frac{I_{max}}{I_{moy}}, \frac{V_{max}}{V_{moy}}, \frac{V_{min}}{V_{moy}}, D, \frac{R}{V_{moy}}, \frac{IR_{max}}{IR_{moy}} \right]. \quad (2)$$

4 Classification de HITS

L'objectif de cette partie est de définir un séparateur permettant de discriminer les HITS parmi les classes de micro-embolies et d'artefacts. Avant toute chose, il est important de noter que certains paramètres discriminants de la section 3 sont obtenus comme fonctions d'autres paramètres. Il convient alors de vérifier l'existence ou non d'un espace de représentation de dimension inférieure qui déformerait le moins possible le nuage d'origine des HITS. L'analyse en composante principale (ACP) est une méthode généralement utilisée pour réaliser cette tâche [8]. Dans la pratique, le nouvel espace sera défini par l'ACP sur les données d'apprentissage.

La tâche suivante est celle de la construction d'un séparateur de HITS dans le nouvel espace de représentation. Plusieurs méthodes existent dans la théorie de la classification. Pour les données étiquetées, le séparateur à vaste marge (SVM) qui réalise de bonne performance [9] sera utilisé.

5 Résultats

Plusieurs patients ont été contrôlés par le TCD-X de Atys Medical qui est un appareil Doppler transcranien portable. Après analyse de ces acquisitions, trois patients ont été retenus pour tester notre approche de classification supervisée de HITS. Les

HITS détectés sur les trois patients ont été écoutés puis leurs représentations temps-fréquence ont été visualisées par un expert dans le but de les étiqueter. Le tableau 1 présente le nombre d'artefacts et de micro-emboles identifiés sur chaque patient.

TABLE 1 – HITS détectés sur les trois patients.

Patients	HITS	
	Artefacts	Micro-emboles
n°1	118	52
n°2	81	32
n°3	105	13

Les paramètres caractéristiques des HITS sont ceux de (2). L'ACP est effectuée sur les données d'apprentissage constituées des HITS des patients n°1 et n°2.

TABLE 2 – Pourcentage d'inertie des projections sur les composantes principales.

Axes	n°1	n°2	n°3	n°4	n°5	n°6
% d'inertie	43	37	13	9	7	1

Le tableau 2 montre que les deux premiers axes conservent 80% de l'inertie totale. Ils semblent donc assez représentatifs pour discriminer les artefacts et les micro-emboles. Cette hypothèse est visuellement confirmée par la figure 1 qui présente les HITS projetés sur les 2 premiers axes (ou composantes principales). En effet, on constate que les micro-emboles et les artefacts sont assez bien séparés. Par ailleurs, d'après cette figure, un séparateur avec un noyau non-linéaire semble être plus approprié pour la discrimination des HITS. Un SVM avec un noyau gaussien est utilisé dans la suite.

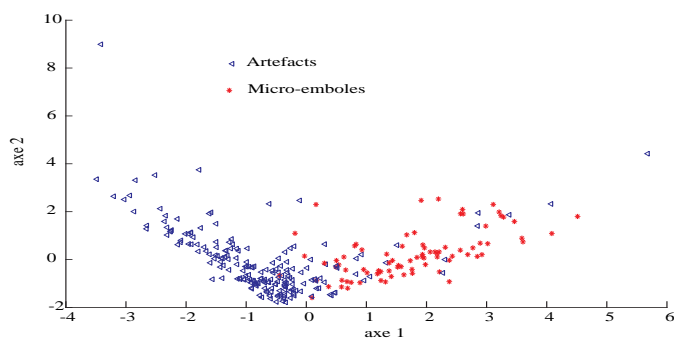


FIGURE 1 – Hits projetés sur les deux premiers axes de l'ACP.

La classification avec un SVM à noyau Gaussien sur l'espace formé des deux premiers axes est donnée par le tableau 3.

Les 5 premiers axes conservent 99% de l'inertie totale. Cela signifie que l'information sur la dispersion des observations est quasiment contenue dans un espace de dimension 5. La classi-

TABLE 3 – Matrice de confusion du patient n°3 à partir du SVM à noyau Gaussien appris sur les 2 premières composantes principales des patients n°1 et n°2.

Groupes réels	Groupes prédits	
	Artefacts	Micro-emboles
Artefacts	91	14
Micro-emboles	3	10

fication avec un SVM à noyau Gaussien sur l'espace formé des 5 premiers axes est donnée par le tableau 4.

TABLE 4 – Matrice de confusion du patient n°3 à partir du SVM à noyau Gaussien appris sur les 5 premières composantes principales des patients n°1 et n°2.

Groupes réels	Groupes prédits	
	Artefacts	Micro-emboles
Artefacts	95	10
Micro-emboles	2	11

Les résultats montrent que l'erreur de mauvaise classification passe de 14,4% sur l'espace des deux premiers axes à 10,1% sur l'espace des cinq premiers axes. L'augmentation de l'erreur de 4,2% est raisonnable vu que l'on passe de 80% à 99% de conservation de l'inertie totale.

L'étude des HITS mal classés révèle que certains artefacts sont semblables à des micro-emboles. La seule distinction se fait de manière auditive ; ce qui les rend difficiles à détecter à partir des paramètres définis. De même, certains micro-emboles sont mal classés parce qu'ils ont été séparés en deux parties : l'une de haute fréquence (ou vitesse) et l'autre de basse fréquence qui est considérée comme un artefact.

L'étude de l'impact des paramètres caractéristiques sur les deux premiers axes de l'ACP se fait généralement grâce au cercle de corrélation. La figure 2 présente le cercle de corrélation entre les paramètres caractéristiques et les deux premiers axes. Le constat à partir de cette figure est que tous les paramètres caractéristiques de (2) contribuent activement à la formation de l'axe 1. L'axe 1 apparaît comme une moyenne de tous ces paramètres. A contrario, pour l'axe 2, l'intensité maximale dans le flux, la vitesse minimale dans le flux et la durée du HITS sont les premiers facteurs de contribution. Cela traduit un effet de discrimination : les artefacts sont en général plus intense, de basse fréquence (ou vitesse) et de durée plus longue comparativement aux micro-emboles.

Le tableau 5 présente le résultat de la classification d'un SVM à noyau linéaire appris sur les 2 premières composantes principales. En comparant ce tableau avec le tableau 3 du SVM à noyau Gaussien appris sur les 2 premières composantes principales, on constate un accroissement de l'erreur de 14,4% à 18,6% en défaveur du noyau linéaire. Ce résultat est cohérent avec la figure 1 où une séparation non-linéaire entre les artefacts et les

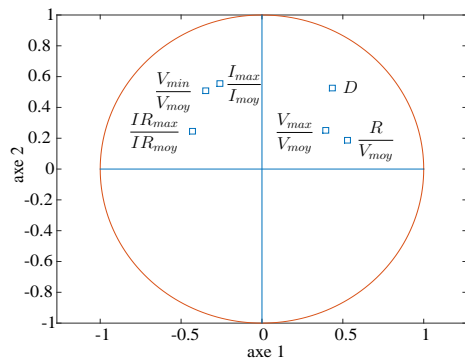


FIGURE 2 – Cercle de corrélation entre les paramètres caractéristiques et les deux premières composantes principales.

micro-emboles semble la plus appropriée.

TABLE 5 – Matrice de confusion du patient n°3 à partir du SVM à noyau linéaire appris sur les 2 premières composantes principales des patients n°1 et n°2.

Groupes réels	Groupes prédits	
	Artefacts	Micro-emboles
Artefacts	84	21
Micro-emboles	1	12

TABLE 6 – Matrice de confusion du patient n°3 à partir du SVM à noyau Gaussien appris sur les 2 premières composantes principales du patient n°2.

Groupes réels	Groupes prédits	
	Artefacts	Micro-emboles
Artefacts	89	16
Micro-emboles	4	9

Afin de comparer l'influence de l'enrichissement de la base de données sur la classification, l'ensemble d'apprentissage constitué des patients n°1 et n°2 est remplacé par un nouvel ensemble d'apprentissage constitué uniquement du patient n°2. Le tableau 6 montre le résultat de la classification suite à cette modification. Le taux d'erreur de mauvaise classification passe de 14,4% dans le cas du tableau 3 à 17% dans le cas présent. Ceci conforte l'idée qu'un enrichissement de la base de données améliore les performances de la détection des micro-emboles.

6 Conclusion

Cet article traite de la classification de signaux transitoires de haute intensité (HITS) détectés lors du contrôle de l'artère

cérébrale moyenne par un dispositif de Doppler transcranien portable. La première étape a consisté à définir des paramètres caractéristiques normalisés des HITS. Le but de la normalisation était de réduire la dépendance de ceux-ci aux patients. Par la suite, nous avons réalisé une séparation des micro-emboles et des artefacts d'un patient test à partir d'une base de données contenant les HITS étiquetés d'autres patients. La classification a été effectuée grâce au séparateur à vaste marge (SVM). L'ensemble des 6 paramètres caractéristiques peut être réduit à 2 paramètres dans un nouvel espace de projection sans perte significative des performances de la classification. Par ailleurs, le taux de mauvaise classification décroît lorsque l'ensemble d'apprentissage s'enrichit. Ces résultats montrent que notre approche est prometteuse.

Références

- [1] Kahlert, P. and Al-Rashid, F. and Döttger, P. and Mori, K. and Plicht, B. and Wendt, D. and Bergmann, L. and Kottenberg, E. and Schlamann, M. and Mummel, P. and others. *Cerebral Embolization During Transcatheter Aortic Valve Implantation : A Transcranial Doppler Study*. *Circulation*, vol 126, num 10, page 1245–1255, 2012.
- [2] Mackinnon, Andrew D. and Aaslid, Rune and Markus, Hugh S. *Long-Term Ambulatory Monitoring for Cerebral Emboli Using Transcranial Doppler Ultrasound*. *Stroke*, vol 35, num 1, page 73–78, 2004.
- [3] Aydin, N. and Marvasti, F. and Markus, H. *Embotic Doppler ultrasound signal detection using discrete wavelet transform*. *Information Technology in Biomedicine, IEEE Transactions on*, vol 8, num 2, page 182–190, 2004.
- [4] Girault, J. and Kouame, D. and Ouahabi, A. and Patat, F. *Micro-emboli detection : an ultrasound Doppler signal processing viewpoint*. *Biomedical Engineering, IEEE Transactions on*, vol 47, num 11, page 1431-1439, 2000.
- [5] Girault, J.-M. and Ménigot, S. and Guibert, B. *Détection automatique de micro-emboles cérébraux grâce à un nouveau détecteur de variance conditionnelle*. *IRBM*, vol 33, num 3, page 217 – 222, 2012.
- [6] Girault, J. and Zhao, Z. *Synchronous detector as a new paradigm for automatic microembolus detection*. *International Journal of Biomedical Engineering and Technology*, vol 14, num 1, page 60–70, 2014.
- [7] Fanglue, L. and Saglio, C. and Almar, M. and Guibert, B. and Delachartre, P. *Micro-embolic signal characterization based on long time Doppler time-frequency image processing and statistics*. *Signal Processing (ICSP), 12th International Conference on*, page 1046–1051, 2014.
- [8] Jolliffe, I.T. *Principal Component Analysis*. Springer, 2002.
- [9] Caruana, R. and Niculescu-Mizil, A. *An Empirical Comparison of Supervised Learning Algorithms*. *ACM*, page 161–168, 2006.