

# Classification des signaux ECG par une transformée temps-fréquence optimisée et réseaux de neurones profonds

Yacine BELGUERMI<sup>1</sup>, Ali MOUKADEM<sup>1</sup>, Jean-Baptiste COURBOT<sup>1</sup>

<sup>1</sup>Institut de Recherche en Informatique, Mathématiques, Automatique et Signal (IRIMAS) UR 7499  
12 rue des Frères Lumière, 68 093 MULHOUSE Cedex  
yacine.belguermi@uha.fr

**Résumé** – Cet article présente une nouvelle approche pour la classification d'électrocardiogrammes (ECG) en 4 classes. Tout d'abord, le signal est segmenté en battements qui sont transformés en représentations temps-fréquence. Les images formées par le module des transformées temps-fréquence sont ensuite utilisées dans un réseau neuronal convolutif 2D (CNN) pour la classification. Trois transformations ont été utilisées et comparées en terme de taux de classification : la STFT (Short Time Fourier Transform), la ST (Stockwell Transform) et une ST optimisée. Les données utilisées pour l'apprentissage et le test sont obtenues à partir de deux bases de données de référence: MIT-BIH et INCART. Une précision de 99.72% est obtenue avec la ST optimisée sur un ensemble de test de référence, ce qui dépasse les résultats obtenus dans l'état de l'art.

**Abstract** – This article presents a novel approach for electrocardiogram (ECG) classification into 4 classes. First, the signal is segmented into beats which are transformed to time frequency representations. The resulting images from the module of the time frequency transforms are provided into a 2D convolutional neural network (CNN) for classification. Three transformations were used and compared in term of classification accuracy: the STFT (Short Time Fourier Transform), the ST (Stockwell Transform) and an optimized ST. Data used for training and testing are obtained from two reference databases : MIT-BIH and INCART. An accuracy of 99.72% was obtained using the optimized ST, which exceeds the results obtained from the state of art.

## 1 Introduction

Selon l'organisation mondiale de la santé [1], les maladies cardio-vasculaires sont la principale cause de décès dans le monde, où environ 17,9 millions de personnes sont décédées de maladies cardiaques en 2019. L'électrocardiogramme (ECG) est le processus d'enregistrement de l'activité électrique générée par les muscles cardiaques en mesurant le potentiel de tension entre un point spécifique et un autre par rapport à une référence, cette configuration est appelée une dérivation avec une configuration non invasive. Un signal ECG classique est un ensemble de motifs répétés appelés battements, c'est à dire un ensemble de 5 motifs qui sont P-QRS-T (figure 1) [2].

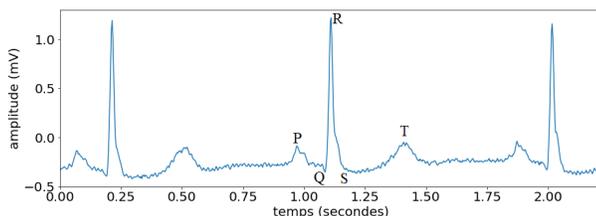


FIGURE 1 – exemple d'un complexe P-QRS-T

Il existe de nombreux des travaux de classification des signaux ECG qui diffèrent en termes de données utilisées, de descripteurs utilisés et de méthodes de classification. Dans cet

article, nous envisageons de classier les battements séparés à partir de deux bases de données de référence : MIT-BIH et INCART [3]. La plupart des recherches utilisent l'intervalle R-R de battements segmentés ou une certaine fenêtre temporelle autour du pic R, ce qui donne les informations nécessaires sur le rythme cardiaque et permet de détecter la maladie si elle existe [4, 5, 6, 7, 8]. Les intervalles R-R, les statistiques d'ordre supérieur et les coefficients d'ondelettes ont été utilisés comme caractéristiques par He. et al. [4], où la méthode de classification utilisé est une arbre de décision. Zhang. et al. [6] ont utilisés les descripteurs temporels du signal qui sont ensuite classifiés à l'aide d'une machine à vecteurs de support (SVM). Dans la méthode de De Chazal et al. [5], les caractéristiques sont basés sur la morphologie de l'ECG, les intervalles des battements cardiaques et les intervalles RR, puis les discriminants linéaires sont utilisés comme classificateur. Dans le travail de Zhou et al. [7], les battements ont été classés directement à l'aide d'un réseau neuronal auto-encodeur qui permettent même la reconstruction et le débruitage du signal d'entrée. Les données brutes (battements) ont également été utilisées par Acharya. et al. [8] avec un réseau neuronal convolutif 1D. Dans d'autres travaux [9], Huang et al. transforment l'ECG en images de spectrogramme STFT (Short Time Fourier Transform) qui sont ensuite introduites dans un réseau de neurones convolutif 2D, l'idée ici est de diviser le signal d'entrée en segments de 10 secondes sans localiser aucun pic, puis de les convertir en représentations temps-fréquence.

Notre objectif est d'effectuer une classification automatique

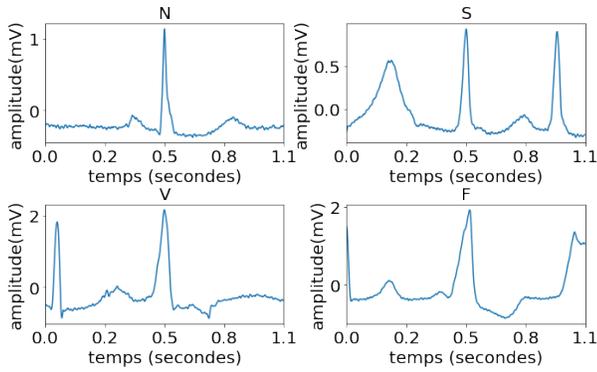


FIGURE 2 – Exemples typiques de signaux par classe.

des battements cardiaques en 4 classes (figure 2) selon les recommandations de l’Advancement of Medical Instrumentation (AAMI) : normal (N), supraventriculaire ectopique (S), ventriculaire ectopique (V) et fusion entre ventriculaire et normal (F). L’approche utilisée dans cet article consiste à projeter le signal temporel qui correspond à chaque battement cardiaque dans le plan temps-fréquence afin de les classifier avec un CNN. La différence principale avec l’approche citée dans [9] réside dans le fait que l’on cherche à classifier les signaux par battements et non pas par durée de 10 secondes pour détecter des éventuelles arythmies. Les données ont été obtenues à partir des bases de données MIT-BIH et INCART [3], puis nous évaluons nos modèles qui correspondent aux transformés temps-fréquence appliqués, en utilisant un ensemble de données de référence et une technique de validation croisée.

## 2 Traitement et classification

### 2.1 Les ensembles de données et leur structure

La base de données MIT-BIH contient 48 extraits d’une demi-heure d’enregistrements d’ECG ambulatoires à deux dérivations avec une fréquence d’échantillonnage de 360 Hz, obtenus auprès de 47 sujets étudiés par le laboratoire BIH [3]. Tout d’abord, dans notre implémentation, nous avons éliminé la classe non-classifiée (Q) en raison de sa rareté. En suivant le schéma de P. et al. [5], nous obtenons deux ensembles de données également divisés (DS1 et DS2) qui ont été choisis pour effectuer un test réaliste ; ce schéma est devenu une norme pour les tests et le choix des données. La base de données INCART, à 12 dérivations, est également utilisée pour consolider notre modèle [3]. Cette base de données comprend 75 enregistrements annotés. Chaque enregistrement dure 30 minutes et contient 12 dérivations échantillonnées à 257Hz [3]. Après avoir rééchantillonné la base de données INCART de 257Hz à 360 Hz, nous avons finalement utilisé deux distributions de données :

1. *distribution 1* : ajouter l’ensemble de données INCART à l’ensemble d’apprentissage DS1 pour consolider l’apprentissage. L’ensemble de test DS2 reste le même.

TABLE 1 – Nombre d’exemple par classe pour l’ensemble des données utilisées. Notons le déséquilibre en faveur des battements normaux "N".

Classe	N	S	V	F
Nombre	243157	4729	26965	1021

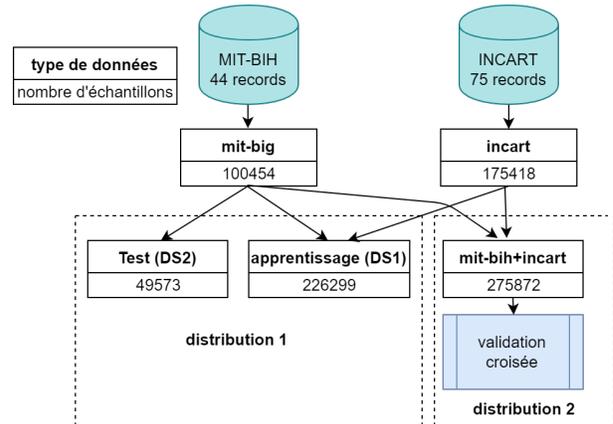


FIGURE 3 – Distributions des données utilisées dans cet article, à partir des bases MIT-BIH et INCART [3].

2. *distribution 2* : mélanger tous les enregistrements des deux bases puis utiliser la technique de validation croisée à 5 blocs.

La figure 3 illustre les distributions des données proposées dans cette étude. Avec les bases de données utilisées, le nombre d’exemples par classe est illustré dans le tableau 1.

En utilisant les annotations des bases de données, nous avons pu segmenter et diviser tous les battements en utilisant une fenêtre de 400 échantillons centrés autour de chaque pic R. Enfin, nous normalisons les données de la façon suivante :

$$\bar{x} = \frac{x - \mu_x}{x_{max} - x_{min}} \quad (1)$$

$x$  est le signal d’entrée qui représente un seul battement,  $\mu_x$  est sa moyenne,  $x_{min}$  et  $x_{max}$  représentent son minimum et le maximum, respectivement.

### 2.2 Représentations temps-fréquence

Dans ce travail, nous comparons trois transformées temps-fréquence : la STFT, la ST (Stockwell Transform) [10] et une ST optimisée [11]. Le principe de ces transformées temps-fréquence linéaires consiste à appliquer une fenêtre glissante sur le signal afin de calculer le spectre local. La différence entre les trois méthodes est essentiellement liée à la façon dont la largeur de la fenêtre varie en fonction de la fréquence. Pour la STFT, la transformée est dite mono-résolution, car la largeur de la fenêtre ne varie pas en fonction de la fréquence. Pour les deux versions de la ST, elles peuvent être vues comme des versions hybrides

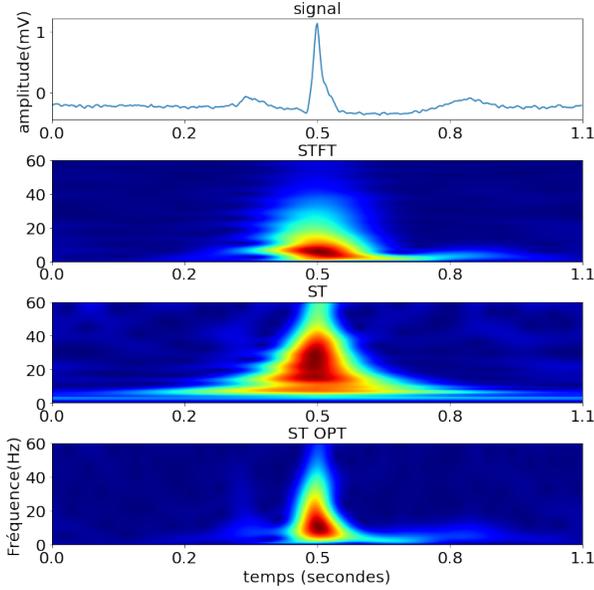


FIGURE 4 – Exemple d’un signal et ses représentations temps-fréquence (STFT, ST et ST optimisée)

entre la transformée en ondelettes et la STFT, où la largeur de la fenêtre dépend fonction de la fréquence d’analyse. Soient  $x(t)$ ,  $w(t) \in L^2(\mathbb{R})$  le signal et la fenêtre d’analyse, respectivement. Les transformées temps-fréquences appliquées dans cet article seront notées  $S_x^\omega(t, f)$ , elles peuvent être exprimées comme suit :

$$S_x^\omega(t, f) = \int_{-\infty}^{+\infty} x(\tau)w(\tau - t)e^{-2j\pi f\tau} d\tau, (2)$$

Pour pouvoir comparer avec la ST et sa version optimisée, la fenêtre gaussienne est appliquée pour la STFT,  $w(t) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{t^2}{2\sigma^2}}$ , avec  $\sigma = 0.1$ . Pour la ST, la largeur de la fenêtre contrôlée par la variable  $\sigma$  sera en fonction de la fréquence,  $\sigma = \frac{1}{|f|}$ ,  $f \in \mathbb{R}$ . Pour la ST optimisée, nous utilisons la version proposée dans [11] avec  $\sigma(f) = \frac{mf^p+k}{f^r}$  et  $m, p, k$  et  $r \in \mathbb{R}$  sont choisis pour maximiser la concentration d’énergie dans le plan temps-fréquence :

$$\arg \max_{m,p,k,r} \left( 1 / \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \overline{|S_x^\omega(t, f)|} dt df \right) (3)$$

$$\text{avec } \overline{|S_x^\omega(t, f)|} = \frac{S_x^\omega(t, f)}{\sqrt{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |S_x^\omega(t, f)|^2 dt df}}$$

Dans cet article, nous appliquons aux complexes P-QRS-T les trois transformées temps-fréquence entre 0 et 60 Hz. Ensuite, nous calculons la module de chaque transformée pour obtenir 3 images pour chaque signal. La figure 4 illustre la procédure avec un exemple d’un complexe P-QRS-T et ses représentations temps-fréquences.

## 2.3 Méthode de classification

Comme nos données sont maintenant représentées sous forme d’images qui représentent la module de la transformé temps-fréquence, nous pouvons profiter des avantages d’un CNN 2D pour les classifier [12]. Les architectures CNN ont été présentées pour la première fois par LeCun et al. [13] pour la reconnaissance de chiffres manuscrits. Les CNNs sont connues pour leur capacité à extraire automatiquement des caractéristiques en appliquant des filtres non linéaires [12]. Notre classifieur se compose de deux parties, la première est un réseau CNN pour l’extraction automatique des caractéristiques, puis un réseau peu profond entièrement connecté est ajouté pour classifier les résultats du CNN. Trois types principaux de couches sont utilisés dans notre architecture : Conv2D, MaxPooling2D et BatchNormalization. Ensuite trois couches sont ajoutées (2 avec activation de relu et une couche softmax) ; les types des couches mentionnées sont présentées dans [12]. L’ensemble de l’architecture est résumé dans le tableau 2.

TABLE 2 – Architecture de réseau neuronal proposée.

couche	forme de sortie
Entrée	120,30,3
Conv2D+Batch_norm	120,30,64
Max_pool2D	60,15,64
Conv2D+Batch_norm	60,15,64
max_pool2D	30,7,64
Conv2D+Batch_norm	30,7,128
Conv2D+Batch_norm	30,7,256
Conv2D+Batch_norm	30,7,256
Max_pool2D	15, 4, 256
flatten	15360
relu+Batch_norm	256
relu+Batch_norm	128
softmax	4

## 3 Résultats et discussion

Nous avons utilisé l’optimiseur SGD (Descente de Gradient Stochastique) avec un taux d’apprentissage de 0.001, le nombre d’époques est 100 avec la possibilité d’arrêt précoce si la précision d’apprentissage et de validation sont supérieures à 98%. Les résultats sont présentés dans le tableau 2. L’utilisation de distribution 2 permet d’obtenir de meilleurs résultats, sans sur-apprentissage ni sous-apprentissage, avec une excellente sensibilité pour les classes individuelles et une précision globale de 99%. Avec la distribution 1 nous avons obtenu des résultats faibles pour la STFT et la ST surtout avec les classes S et F. Cependant, en utilisant la ST optimisée, nous avons obtenu d’excellents résultats pour les deux distributions.

Globalement, nous remarquons que les résultats obtenus avec la ST optimisée dépassent les résultats de l’état de l’art comme

TABLE 3 – Résultats obtenus à l’aide de transformées temps-fréquence (en bleu) et comparaison avec l’état de l’art.

Méthode	Sensibilité				Précision	Test
	N	S	V	F		
STFT	91.84	17.77	87.91	21.32	88.30	DS2
ST	90.43	03.82	86.08	00.37	86.24	
ST opt.	98.60	<b>99.92</b>	<b>98.93</b>	<b>98.90</b>	<b>98.68</b>	
STFT	<b>99.54</b>	88.01	96.99	74.25	98.91	validation croisée
ST	98.50	80.67	63.67	65.24	97.57	
ST opt.	99.38	<b>99.62</b>	<b>98.90</b>	<b>94.25</b>	<b>99.72</b>	
P. [5]	86.86	83.83	77.74	89.43	86.19	DS2
H. [4]	92	91	89	/	91.50	DS2
L. [7]	<b>99.76</b>	82.29	98.31	87.71	99.34	DS2
Z. [6]	88.94	79.06	85.48	93.81	88.34	DS2
A. [8]	88.39	84.21	91.67	88.53	89.03	validation croisée

l’illustre le tableau 3. Un autre point à signaler dans cet article est l’utilisation de deux bases de données avec des fréquences d’échantillonnage différentes, sur lesquelles nous avons réussi à garder un taux de classification élevé, ce qui permet d’ajouter d’autres bases de données.

## 4 Conclusion

Nous avons mis en place un système de classification robuste des signaux ECG basé sur les représentations temps-fréquence et les réseaux de neurones convolutifs à partir de deux bases de données : MIT-BIH et INCART [3]. Nous avons pu atteindre avec la ST optimisée des résultats qui sont largement supérieurs aux résultats obtenus dans la littérature. En outre, les résultats montrent que l’optimisation de la représentation temps-fréquence améliore nettement les performances de CNN, ce qui confirme qu’il ne suffit pas d’introduire les données au modèle pour obtenir des résultats satisfaisants indépendamment de la qualité des données fournies. Concernant l’application abordée dans ce papier, d’autres transformées temps-fréquence [14] peuvent être également testées et comparées dans des travaux ultérieurs.

## Références

- [1] *Cardiovascular diseases (CVDs)*. who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). Accédé le : 17.03.2022.
- [2] Akanksha AGRAWAL et Dhanashri H. GAWALI. “Comparative study of ECG feature extraction methods”. In : *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*. 2017, p. 2021-2025.
- [3] Ary L GOLDBERGER et al. “PhysioBank, PhysioToolkit, and PhysioNet : components of a new research resource for complex physiologic signals”. In : *circulation* 101.23 (2000), e215-e220.
- [4] Jinyuan HE et al. “A pyramid-like model for heartbeat classification from ECG recordings”. In : *PloS one* 13.11 (2018), e0206593.
- [5] Philip DE CHAZAL, Maria O’DWYER et Richard B REILLY. “Automatic classification of heartbeats using ECG morphology and heartbeat interval features”. In : *IEEE transactions on biomedical engineering* 51.7 (2004), p. 1196-1206.
- [6] Z. ZHANG et al. “Heartbeat classification using disease-specific feature selection”. In : *Comput Biol Med* 46 (mars 2014), p. 79-89.
- [7] Lin ZHOU et al. “Deep learning-based classification of massive electrocardiography data”. In : *2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*. IEEE. 2016, p. 780-785.
- [8] U. R. ACHARYA et al. “A deep convolutional neural network model to classify heartbeats”. In : *Comput Biol Med* 89 (oct. 2017), p. 389-396.
- [9] Jingshan HUANG et al. “ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network”. In : *IEEE access* 7 (2019), p. 92871-92880.
- [10] Robert Glenn STOCKWELL, Lalu MANSINHA et R.P. LOWE. “Localization of the complex spectrum : the S transform”. In : *IEEE transactions on signal processing* 44.4 (1996), p. 998-1001.
- [11] Ali MOUKADEM et al. “A new optimized Stockwell transform applied on synthetic and real non-stationary signals”. In : *Digital Signal Processing* 46 (2015), p. 226-238.
- [12] Arohan AJIT, Koustav ACHARYA et Abhishek SAMANTA. “A review of convolutional neural networks”. In : *2020 international conference on emerging trends in information technology and engineering (ic-ETITE)*. IEEE. 2020, p. 1-5.
- [13] Yann LECUN et al. “Backpropagation applied to handwritten zip code recognition”. In : *Neural computation* 1.4 (1989), p. 541-551.
- [14] Thomas OBERLIN, Sylvain MEIGNEN et Valérie PERRIER. “The Fourier-based synchrosqueezing transform”. In : *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE. 2014, p. 315-319.