

Détection de l'insuffisance cardiaque à l'aide d'un réseau neuronal convolutif via les signaux ECG

Jad BOTROS¹, Farah MOURAD-CHEHADE¹, David LAPLANCHE^{1,2}

¹Laboratoire Informatique et Société Numérique (LIST3N), Université de Technologie de Troyes (UTT)
12 rue Marie Curie, CS 42060, 10004 Troyes CEDEX, France

²Pôle Santé publique, Hôpitaux Champagne Sud (HCS)
Troyes, France

jad.botros@utt.fr, farah.chehade@utt.fr, david.laplanche@hcs-sante.fr

Résumé – L'insuffisance cardiaque (IC) est une maladie cardiovasculaire chronique qui entraîne une augmentation de la mortalité, de la morbidité et des coûts de santé. L'électrocardiogramme (ECG) est un outil de diagnostic simple et non invasif qui peut révéler des changements différentiels dans l'IC. En raison de leur faible amplitude et de leur durée, ces changements peuvent être subtils et éventuellement mal classés lors d'une interprétation manuelle ou d'une analyse clinique. Cet article décrit un modèle de réseau neuronal convolutif profond (CNN) à 7 couches pour la détection automatique de l'IC. Le modèle CNN proposé ne nécessite ni prétraitement des signaux ECG, ni extraction de caractéristiques. Le modèle est entraîné et testé à l'aide d'un ensemble équilibré de données extraites des bases de données MIT-BIH et BIDMC, et il atteint une précision de 99,33%, une sensibilité de 99,10% et une spécificité de 99,57%.

Abstract – Heart failure (HF) is a chronic cardiovascular disease that causes increased mortality, morbidity, and healthcare costs. The electrocardiogram (ECG) is a simple and non-invasive diagnostic tool that can unveil differential changes in HF. Because of the small amplitude and duration of these changes, they can be subtle and possibly misclassified during manual interpretation or clinical analysis. A 7-layer deep convolutional neural network (CNN) model for HF automatic detection is described in this paper. The proposed CNN model requires neither pre-processing of ECG signals, nor engineered features. The model is trained and tested using a balanced dataset extracted from the MIT-BIH and BIDMC databases, and it achieves an accuracy of 99.33%, sensitivity of 99.10%, and specificity of 99.57%.

1 Introduction

L'insuffisance cardiaque (IC) est l'incapacité du cœur à pomper suffisamment de sang pour répondre aux demandes métaboliques ou pour assurer le retour veineux. Il s'agit d'un syndrome clinique causé par des lésions du tissu musculaire du cœur. Les cardiopathies ischémiques, la maladie coronarienne, l'hypertension et le diabète sont les principales causes de ces lésions [1].

Selon les statistiques, l'IC touche 2 à 3% de la population dans les pays développés : environ 15 millions d'européens sont affectés par ce syndrome [2]. À l'échelle mondiale, plus de 26 millions de personnes ont été diagnostiquées avec une IC, ce qui constitue un problème de santé publique majeur ainsi qu'un fardeau économique important [3].

Une combinaison de signes et de symptômes, corroborée par des tests, est nécessaire au diagnostic de l'IC. L'électrocardiogramme (ECG) est une mesure non invasive permettant d'enregistrer l'activité électrique du cœur. L'IC a un effet sur ce signal. Cependant, l'évaluation visuelle du signal ECG obtenu du patient prend du temps, et l'interprétation manuelle est sujette à la variabilité inter-observateurs. Les changements qui affectent le signal ECG sont de nature différentielle, ce qui les rend dif-

ficiles à être détectés à l'œil nu. D'où l'importance d'utiliser des systèmes de diagnostic assisté par ordinateur (DAO) pour détecter ces anomalies, en raison de leur précision et de leur fiabilité accrues. Le problème de diagnostic de l'IC est défini comme un problème de classification binaire qui prend en entrée un signal d'ECG et rend en sortie une décision concernant la présence ou non de l'IC. Un système de DAO typique comprend ainsi quatre étapes-clés : le pré-traitement du signal, l'extraction de caractéristiques, la sélection de caractéristiques significatives et la classification.

De nombreuses techniques d'apprentissage automatique telles que le classifieur bayésien [5], les arbres de régression (CART) [6], le classifieur SVM [7] et le random forest [8] ont été proposées dans la littérature pour résoudre le problème du diagnostic de l'IC. Toutefois, ces méthodes nécessitent un pré-traitement des données pour une extraction de caractéristiques, la qualité de ces dernières affectant directement la capacité d'apprentissage du réseau. Une autre étape importante de l'apprentissage automatique est la sélection des caractéristiques. Ce processus est également contraignant en temps et en précision. Des modèles d'apprentissage profond utilisant des réseaux tels que le réseau neuronal profond (DNN) [9], le réseau neuronal convolutif (CNN) [11], [13] [14] et le réseau neuro-

nal récurrent à mémoire court et long terme (LSTM) [10], [12], [14] ont été proposés pour optimiser les performances d'un système de DAO afin d'éviter les écueils de l'apprentissage automatique traditionnel.

Ce papier propose un nouveau modèle basé sur l'apprentissage profond pour diagnostiquer l'IC à partir d'un court enregistrement de signaux ECG. Le modèle proposé est un réseau neuronal convolutif (CNN) profond comportant sept couches, avec des exigences minimales de pré-traitement. Le principal avantage de l'utilisation d'un CNN par rapport aux méthodes traditionnelles d'apprentissage automatique est qu'il détecte automatiquement les caractéristiques importantes sans aucune intervention humaine. Avec ses sept couches, l'architecture proposée est plus petite que les architectures actuellement disponibles dans la littérature, et elle est donc moins gourmande en mémoire et en temps de calcul. Cette approche est illustrée à l'aide de signaux ECG réels acquis à partir de bases de données holter ambulatoires et d'enregistreurs ECG. Les résultats démontrent de meilleures performances, même avec des segments d'ECG de 2s en entrée.

Le reste de ce papier est organisé comme suit. La section 2 décrit le jeu de données utilisé pour l'apprentissage et le test. La section 3 décrit la méthode proposée. La section 4 présente et discute les résultats obtenus. Enfin, la section 5 présente la conclusion.

2 Matériels et jeu de données

Les signaux ECG utilisés dans cette étude rétrospective proviennent de deux bases de données accessibles au public : la base de données MIT-BIH sur le rythme sinusal normal (MIT-BIH NSR) et la base de données BIDMC sur l'insuffisance cardiaque congestive (BIDMC CHF). Le tableau Tableau 1 résume les caractéristiques de chaque base de données.

Pour s'assurer que tous les signaux ECG sont échantillonnés à la même fréquence, les enregistrements MIT-BIH NSR sont sur-échantillonnés à 250 Hz. Les signaux sont ensuite segmentés en fragments de 2 secondes sans chevauchement afin de garantir les différences entre les échantillons d'entraînement et les échantillons de test. On obtient ainsi 36 000 segments ECG de 2 secondes par signal. Ensuite, les segments sont régularisés à l'aide de la normalisation du score Z afin qu'ils contribuent tous de manière égale à l'analyse.

Afin d'aborder le problème de classification, un jeu de données équilibré est créé à partir des segments de 2s disponibles. Ainsi, un nombre égal de segments ECG est sélectionné dans les bases de données MIT-BIH NSR et BIDMC CHF, pour éviter le problème du déséquilibre des données et donc de biaiser la classification. Le jeu de données considéré est composé de 60 000 segments d'ECG de 2s, avec 30 000 pour chaque classe, normale ou IC. Davantage de fragments pourraient être pris, au prix d'un temps de calcul et d'une consommation de mémoire plus élevés. Afin d'impliquer également tous les sujets pour chaque classe, avec 18 sujets normaux et 15 patients IC, envi-

TABLE 1 – Caractéristiques des bases de données

Bases de données	Nombre de signaux	Propriétés
MIT-BIH NSR	18 signaux ECG à long terme	05 hommes entre 26 et 45 ans 13 femmes entre 20 et 50 ans fs = 128 Hz
BIDMC CHF	15 signaux ECG à long terme	11 hommes entre 22 et 71 ans 04 femmes entre 54 et 63 ans fs = 250 Hz

ron 1 667 fragments d'ECG 2s sont sélectionnés aléatoirement par sujet normal et 2 000 par signal de patient IC.

Le jeu de données obtenu sera utilisé dans la suite pour l'entraînement et le test du modèle.

3 Méthode proposée

Ce papier propose un réseau CNN à 7 couches pour la classification binaire des signaux ECG. Le principal avantage d'utiliser un CNN par rapport à ses prédécesseurs est qu'il peut détecter automatiquement des caractéristiques importantes sans intervention humaine. Sa faible dépendance à l'égard du pré-traitement réduit le besoin d'effort humain tout en augmentant ses capacités. Le modèle se compose de deux couches convolutives, de deux couches de max-pooling et de trois couches entièrement connectées. La couche convolutive est composée d'un nombre fixe de filtres et est utilisée pour extraire les cartes des caractéristiques. La couche de max-pooling, qui prend la valeur maximale dans une région de filtre spécifique, est utilisée pour réduire la dimensionnalité du réseau. La couche entièrement connectée, qui est généralement placée avant la couche de sortie, est utilisée pour agréger les données de la carte finale des caractéristiques et générer la classification finale.

La première couche prend en entrée le segment d'ECG de 500 points, de durée 2 secondes. Il s'agit d'une couche convolutive avec cinq filtres 1×13 appliqués avec un stride de 1, c'est-à-dire que cinq cartes de caractéristiques sont générées en convoluant les différents filtres avec le signal ECG de 500 points d'entrée. La deuxième couche est une couche de mise en commun maximale (max-pooling) avec une taille de pool de deux et un stride de 4. Cette couche réduit les dimensions des cartes de caractéristiques en convoluant un filtre de 1×2 avec chacune des cartes de caractéristiques générées précédemment. Par conséquent, le nombre de paramètres à apprendre est réduit, tout comme la quantité de traitement dans le réseau. Par conséquent, le modèle est plus robuste aux changements de position des caractéristiques dans l'entrée. Ensuite, une autre couche convolutive composée de dix filtres d'une taille de 1×9 chacun est appliqué avec un stride de 1. Cet ensemble de filtres est appliqué aux cartes de caractéristiques réduites en dimension

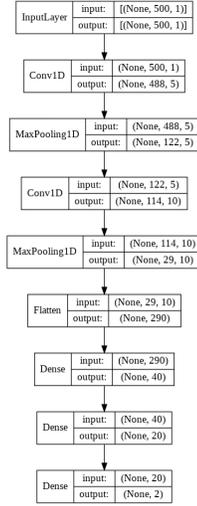


FIGURE 1 – Structure détaillée du modèle proposé.

pour extraire des caractéristiques de plus haut niveau. La quatrième couche est une couche de max-pooling avec les mêmes propriétés et tâches que la deuxième couche. Après cela, il y a trois couches entièrement connectées, chacune avec 40, 20 et 2 unités. L'entrée de la première couche n'est rien d'autre qu'une version aplatie de la sortie de la couche précédente. À l'exception de la dernière couche, qui utilise la fonction d'activation softmax, toutes les couches utilisent la fonction d'activation Leaky ReLU. La Figure 1 montre la structure détaillée et les dimensions de chaque couche du réseau proposé.

Les poids du modèle sont initialisés en utilisant l'initialisation normale de Glorot, et ils sont mis à jour en utilisant la rétro-propagation avec une taille de lot de 10. Le modèle est construit sur une période de soixante époques. Soit y_i l'étiquette du segment i de l'ensemble de données d'apprentissage, c'est-à-dire $y_i = 1$ si le segment i est extrait d'un signal IC et $y_i = 0$ sinon. On considère \hat{p}_i la probabilité prédite que le segment i est IC, obtenue à la sortie du réseau. Pour le problème de classification binaire, la fonction d'entropie croisée binaire est utilisée pour calculer la perte du modèle, comme le montre l'équation suivante :

$$\mathcal{L}(q) = \frac{-1}{M} \sum_{i=1}^M y_i \cdot \log(\hat{p}_i) + (1 - y_i) \cdot \log(1 - \hat{p}_i),$$

où M est le nombre total de segments pris dans la phase d'apprentissage et q est l'indice de l'époque. L'entropie croisée calcule un score qui représente la différence moyenne entre les valeurs réelles et prédites. Ce score doit être minimisé, la valeur 0 correspondant à une entropie croisée parfaite.

4 Résultats et discussion

Le modèle est validé en utilisant une validation croisée stratifiée à 10 plis sur le jeu de données de 60 000 segments. L'ensemble de données est divisé en dix parties égales avec un

pourcentage égal de chaque classe dans chaque partie, et l'apprentissage et la validation sont effectués dix fois, en utilisant chaque fois 9 plis pour l'apprentissage et le pli restant pour la validation. Trois mesures d'évaluation sont utilisées dans ce document pour évaluer la performance du modèle proposé, à savoir la précision, la sensibilité et la spécificité, comme indiqué ci-dessous :

$$Précision = \frac{TP + TN}{TP + FN + TN + FP}$$

$$Sensibilité = \frac{TP}{TP + FN}$$

$$Spécificité = \frac{TN}{TN + FP}$$

où TP , TN , FP et FN désignent les taux de vrais positifs, de vrais négatifs, de faux positifs et de faux négatifs.

Après la validation croisée stratifiée à 10 plis, le modèle présenté atteint une précision moyenne de 99,3334%, avec une sensibilité moyenne de 99,1% et une spécificité moyenne de 99,57%, avec des écarts types respectivement faibles de 0,6254, 0,0136 et 0,0047. Le nombre d'époques choisi est suffisant pour que la fonction de perte converge et atteigne un minimum stable. La perte moyenne du modèle est de 0,0473, avec un écart type de 0,0260. Le code a été écrit en Python et implémenté dans Google Colab.

Afin d'illustrer les performances de la méthode proposée, une comparaison est établie entre les résultats obtenus et ceux d'autres méthodes de diagnostic de l'IC utilisant les mêmes bases de données. Trois méthodes existantes sont ainsi considérées. Les auteurs de [6] ont exploré le pouvoir de discrimination de diverses mesures de la variabilité de la fréquence cardiaque (VRC) et ont choisi trois combinaisons de ces mesures pour former l'entrée d'un classifieur CART. La méthode de [11] a utilisé un CNN à 11 couches pour classer des fragments d'ECG de 2s, avec un pré-traitement minimal, en normal ou IC. Les auteurs de [13] ont sur-échantillonné les signaux ECG à une fréquence commune, les ont filtrés en utilisant la décomposition en mode empirique et ont effectué une détection des pics R avant la segmentation des battements. Pour la classification, les battements ont été introduits dans un CNN à 11 couches. Le tableau 2 montre les métriques d'évaluation des méthodes de la littérature comparées au modèle proposé. Les résultats de la méthode de [11] donnés dans le tableau 2 sont obtenus avec un jeu de données non équilibré dans la rangée supérieure. Lorsque des données équilibrées sont considérées, la précision tombe à 94,40%, tout comme la sensibilité et la spécificité, qui passent respectivement de 98,87% à 94,68% et de 99,01% à 94,12%, comme le montre la ligne inférieure. Les résultats montrent que le modèle proposé est plus performant que les méthodes existantes. Une spécificité plus élevée est obtenue avec la méthode de [6], mais au prix d'une sensibilité et d'une précision inférieures à celles obtenues avec le modèle proposé. L'utilisation de l'apprentissage profond dans la méthode présentée, associée à un apprentissage efficace et à une

TABLE 2 – Métriques d'évaluation des différentes méthodes trouvées dans la littérature.

Méthode	Précision	Sensibilité	Spécificité
[6]	96.36	89.74	100
[11]-non équilibré	98.97	98.87	90.01
[11]-équilibré	94.40	94.68	94.12
[13]	80.10	81	79.30
Modèle proposé	99.33	99.10	99.57

architecture adaptée, garantit des performances élevées par rapport aux méthodes traditionnelles d'apprentissage automatique. En effet, les capacités d'apprentissage profond permettent une abstraction plus fiable du signal dans un espace à haute dimension sans nécessiter une intervention humaine. Comparé aux architectures des modèles profonds de [11] et [13], le modèle proposé est plus petit, avec 7 couches au lieu de 11, et plus rapide lorsque le temps d'apprentissage est pris en compte.

5 Conclusion

Le présent papier propose un système de classification automatique des ECG pour la détection de l'insuffisance cardiaque. Le modèle proposé se compose d'un CNN à 7 couches et est entièrement automatique, sans étapes de pré-traitement ou de sélection de caractéristiques. Les enregistrements d'ECG normaux sont obtenus à partir de la base de données NSR du MIT BIH, et les enregistrements d'ECG d'insuffisance cardiaque sont obtenus à partir de la base de données CHF du BIDMC. Ces deux bases de données sont librement accessibles sur Internet. Pour éviter les problèmes de déséquilibre des données, un jeu de données est créé avec 30000 segments d'ECG normaux et 30000 segments d'ECG de patients souffrant d'insuffisance cardiaque. Avec sa grande performance, le classifieur CNN proposé joue un rôle important dans la détection et la classification des signaux ECG, comme en témoignent sa précision de 99,3334%, sa sensibilité de 99,1% reflétant le taux de vrais positifs et sa spécificité de 99,57%, qui représente le taux de vrais négatifs. Les travaux futurs incluront la fusion de diverses données médicales, telles que des séries temporelles, des mesures ponctuelles et des notes cliniques, afin d'assurer un suivi efficace de l'état d'un sujet dans le temps et d'offrir une classification plus précise avec une stratification des patients en différents niveaux d'insuffisance cardiaque. Le problème de la classification serait alors défini comme une classification hybride multi-classes, combinant des données de natures différentes.

Références

[1] Kemp, C. & Conte, J. The pathophysiology of heart failure. *Cardiovascular Pathology*. **21**, 365-371 (2012)

[2] Braunschweig, F., Cowie, M. & Auricchio, A. What are the costs of heart failure?. *Europace*. **13**, ii13-ii17 (2011)

[3] Ponikowski, P., Anker, S., AlHabib, K., Cowie, M., Force, T., Hu, S., Jaarsma, T., Krum, H., Rastogi, V., Rohde, L. & Others Heart failure : preventing disease and death worldwide. *ESC Heart Failure*. **1**, 4-25 (2014)

[4] Berry, C., Murdoch, D. & McMurray, J. Economics of chronic heart failure. *European Journal Of Heart Failure*. **3**, 283-291 (2001)

[5] Asyali, M. Discrimination power of long-term heart rate variability measures. *Proceedings Of The 25th Annual International Conference Of The IEEE Engineering In Medicine And Biology Society (IEEE Cat. No. 03CH37439)*. **1** pp. 200-203 (2003)

[6] Melillo, P., Fusco, R., Sansone, M., Bracale, M. & Pecchia, L. Discrimination power of long-term heart rate variability measures for chronic heart failure detection. *Medical Biological Engineering Computing*. **49**, 67-74 (2011)

[7] Liu, G., Wang, L., Wang, Q., Zhou, G., Wang, Y. & Jiang, Q. A new approach to detect congestive heart failure using short-term heart rate variability measures. *PLoS One*. **9**, e93399 (2014)

[8] Masetic, Z. & Subasi, A. Congestive heart failure detection using random forest classifier. *Computer Methods And Programs In Biomedicine*. **130** pp. 54-64 (2016)

[9] Chen, W., Liu, G., Su, S., Jiang, Q. & Nguyen, H. A CHF detection method based on deep learning with RR intervals. *2017 39th Annual International Conference Of The IEEE Engineering In Medicine And Biology Society (EMBC)*. pp. 3369-3372 (2017)

[10] Wang, L., Zhou, W., Liu, N., Xing, Y. & Zhou, X. CHF Detection with LSTM Neural Network. *2018 40th Annual International Conference Of The IEEE Engineering In Medicine And Biology Society (EMBC)*. pp. 514-517 (2018)

[11] Acharya, U., Fujita, H., Oh, S., Hagiwara, Y., Tan, J., Adam, M. & San Tan, R. Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals. *Applied Intelligence*. **49**, 16-27 (2019)

[12] Wang, L. & Zhou, X. Detection of congestive heart failure based on LSTM-based deep network via short-term RR intervals. *Sensors*. **19**, 1502 (2019)

[13] Padmavathi, C. & Veenadevi, S V. Heart Disease Recognition from ECG Signal Using Deep Learning. *International Journal of Advanced Science and Technology*. **29**, pp. 2303-2316 (2020)

[14] Lih, O., Jahmunah, V., San, T., Ciaccio, E., Yamakawa, T., Tanabe, M., Kobayashi, M., Faust, O. & Acharya, U. Comprehensive electrocardiographic diagnosis based on deep learning. *Artificial Intelligence In Medicine*. **103** pp. 101789 (2020)

[15] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P., Mark, R., Mietus, J., Moody, G., Peng, C. & Stanley, H. PhysioBank, PhysioToolkit, and PhysioNet : components of a new research resource for complex physiologic signals. *Circulation*. **101**, E215-20 (2000,6)

[16] Baim, D., Colucci, W., Monrad, E., Smith, H., Wright, R., Lanoue, A., Gauthier, D., Ransil, B., Grossman, W. & Braunwald, E. Survival of patients with severe congestive heart failure treated with oral milrinone. *J. Am. Coll. Cardiol.* **7**, 661-670 (1986,3)