

# Mesure du taux d'oxygène dans le sang par analyse vidéo : apport de l'intelligence artificielle

Alexis WUYART<sup>1</sup>, Imene KERTOUS<sup>1</sup>, Anis MERZOUK<sup>1</sup>, Laure ABENSUR VUILLAUME<sup>2</sup>, Frédéric BOUSEFSAF<sup>1</sup>

<sup>1</sup> Université de Lorraine, LCOMS, F-57000 Metz, France

<sup>2</sup> CHR Metz-Thionville, Metz, France

---

**Résumé** – Les techniques d'intelligence artificielle (IA) modernes ont récemment été utilisées pour mesurer la fréquence cardiaque et la tension artérielle. Les recherches dans ce domaine s'orientent désormais vers l'estimation de paramètres physiologiques tels que la saturation pulsée en oxygène (SpO<sub>2</sub>). Nous proposons dans cet article, une étude de la littérature scientifique dans le domaine de la mesure sans contact de la SpO<sub>2</sub> par caméra conventionnelle afin d'implémenter les méthodes prometteuses, et, in fine, de proposer une nouvelle approche basée sur l'IA. La méthode repose sur la transformée en ondelettes continue des signaux calculés par analyse vidéo. Ces représentations sont ensuite envoyées dans un réseau neuronal de type VGG19 par le biais d'une stratégie d'entraînement par transfert. Le réseau prédit une valeur binaire correspondant à une plage de SpO<sub>2</sub>. Les résultats sont très encourageants avec une classification de la SpO<sub>2</sub> atteignant une performance globale de 96,5%. Les perspectives futures s'articulent maintenant autour d'une solution pour lever la barrière de la classification binaire en développant le modèle pour qu'il puisse fournir des estimations continues de la SpO<sub>2</sub>.

**Abstract** – Modern artificial intelligence (AI) techniques have recently been used to measure heart rate and blood pressure. Research in this field is now moving towards the estimation of physiological parameters such as pulsed oxygen saturation (SpO<sub>2</sub>). In this paper, we propose a survey of the scientific literature in the field of non-contact measurement of SpO<sub>2</sub> by conventional camera in order to implement promising methods and, ultimately, to propose a new AI-based approach. The method is based on the continuous wavelet transform of signals calculated by video analysis. These representations are then fed into a VGG19 neural network via a transfer training strategy. The network predicts a binary value corresponding to a range of SpO<sub>2</sub>. The results are very encouraging with an overall SpO<sub>2</sub> classification performance of 96.5%. Future prospects now revolve around overcoming the barrier of binary classification by developing the model to provide continuous SpO<sub>2</sub> estimates.

---

## 1. Introduction

Ces dernières années ont été particulièrement marquées par l'essor de la mesure des signaux physiologiques par le biais d'une caméra numérique, notamment pour déterminer la fréquence cardiaque [2], la fréquence respiratoire [3], la tension artérielle [4] ou encore la saturation pulsée en oxygène de l'hémoglobine (SpO<sub>2</sub>) [5]. Ces techniques de mesure à distance ont permis de s'affranchir d'écueils qui étaient jusque-là inhérents aux capteurs de contact tels que les risques d'infection, de lésions cutanées et ont permis plus généralement la réduction des coûts liés à l'entretien des outils médicaux [13]. Ces dispositifs offrent également la possibilité d'un suivi ambulatoire renforcé dans le cadre d'actes de télé-médecine ou d'auto-suivi. En médecine courante, ils participent *in fine* à la réduction des maladies transmissibles en offrant des mesures sans contact permettant de diminuer les contaminations manuportées. Cependant, des améliorations sont à fournir au regard des bruits indésirables et artefacts présents dans les mesures, provenant du corps humain [1]. De plus, peu de dispositifs médicaux basés sur l'IA sont aujourd'hui

estampillés CE ou FDA du fait de cette imprécision [14].

C'est dans la continuité des avancées de la surveillance des fonctions vitales chez l'humain via des technologies sans contact, que s'inscrit ce travail. Il s'agira ici d'explorer des techniques d'intelligence artificielle (IA) – notamment par le biais de réseaux de neurones convolutifs 2D – sur la base de flux vidéo pour en extraire avec fiabilité la SpO<sub>2</sub> d'un individu.

## 2. Background

### 2.1 Photopléthysmographie

L'oxygène est principalement véhiculé dans le sang lorsque celui-ci est lié à l'hémoglobine formant le complexe HbO<sub>2</sub>, autrement appelé oxyhémoglobine (97 à 98% de la teneur totale en oxygène) [6]. La saturation pulsée en oxygène est ainsi définie comme le rapport entre l'HbO<sub>2</sub> et la concentration totale d'hémoglobine dans le sang (composée ainsi d'HbO<sub>2</sub> et de désoxyhémoglobine notée Hb), tel que :

$$SpO_2 = \frac{[HbO_2]}{[HbO_2] + [Hb]} \cdot 100\% \quad (1)$$

Cet indicateur joue ainsi un rôle physiologique majeur puisqu'il témoigne du bon fonctionnement de la fonction respiratoire qui assure l'apport d'oxygène aux organes. Ce dernier est mesuré par oxymétrie de pouls, permettant le contrôle rapide en continu du statut d'oxygénation du patient.

Ceci est rendu possible grâce à la photopléthysmographie (PPG), technique non invasive qui permet de mesurer les variations de la lumière transmise ou réfléchiée. L'HbO<sub>2</sub> et l'Hb présentent des différences significatives dans l'absorption de la lumière rouge (autour des 650 nm) et du proche infrarouge (autour des 950 nm). Ces deux longueurs d'onde ont pour propriété de bien pénétrer dans les tissus, *a contrario* de la lumière bleue, verte et infrarouge lointaine qui sont elles absorbées par les tissus non vasculaires et l'eau [7]. L'HbO<sub>2</sub> va ainsi absorber davantage de rayonnement infrarouge tandis que l'Hb absorbera mieux dans le rouge comme en témoigne la *figure 1*.

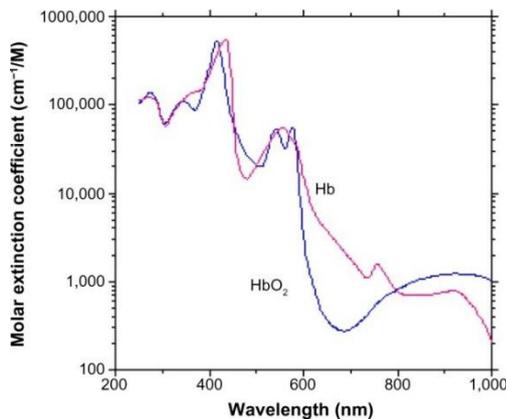


Figure 1 – Spectre d'absorption de l'hémoglobine oxygénée et désoxygénée [6].

## 2.2 Estimation de la SpO<sub>2</sub>

Le rapport des ratios est la méthode utilisée par les oxymètres de pouls. Représentée en *figure 2*, celle-ci repose sur la mesure des intensités lumineuses émergentes moyenne et maximale des longueurs d'onde rouge (respectivement  $DC_R$  et  $AC_R$ ) et infrarouge (respectivement  $DC_{IR}$  et  $AC_{IR}$ ) telle que :

$$\text{Rapport des ratios (R)} = \frac{\frac{AC_R}{DC_R}}{\frac{AC_{IR}}{DC_{IR}}} \quad (2)$$

Cette méthode a également été employée de manière sans contact et à distance par le biais d'une caméra monochromatique [9].

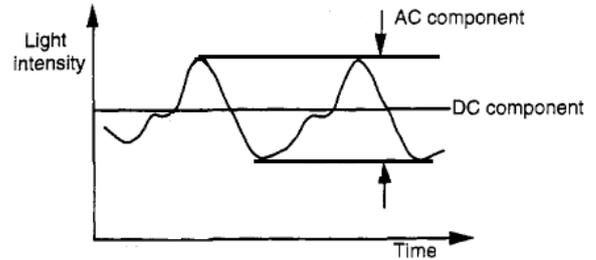


Figure 2 – Signal PPG et ses composants pulsatiles (AC) et continus (DC). Le composant DC est interprété comme la moyenne de l'onde [8].

Dans la continuité des mesures sans contact, la méthode Adaptive Pulse Blood Volume (APBV), se basant sur la signature unique et intrinsèque du signal d'impulsion du volume sanguin sous forme vectoriel, a montré un certain intérêt [10]. La méthode consiste en une phase (d'entraînement et de création d'un modèle) basée sur le calcul d'une collection de rapports sur bruit (SNR) pour différentes plages de SpO<sub>2</sub>. L'estimation d'une SpO<sub>2</sub> inconnue à partir d'un signal PPG se fait par l'observation de la réponse des SNR. Le SNR le plus élevé retranscrit la bonne valeur de la SpO<sub>2</sub>. Cette méthode permet en outre de s'affranchir des artefacts provoqués par les mouvements naturels du sujet, qui sont les principales sources d'erreur.

Enfin, et au regard des méthodes susmentionnées jusqu'ici, les réseaux de neurones convolutifs sont apparus comme une méthode novatrice avec des résultats supérieurs quant à l'estimation de la SpO<sub>2</sub> [11], [12]. C'est donc dans ce pan que s'inscrit naturellement cette étude en proposant une approche originale basée sur la représentation en ondelettes continue du signal PPG (calculé à partir d'un flux vidéo) couplée à un modèle conventionnel d'IA pour l'estimation sans contact de la SpO<sub>2</sub>.

## 3. Méthodologie

Trois vidéos d'une durée de 10 à 12 minutes ont été enregistrées auprès d'un échantillon de 15 participants, à une fréquence d'images de 15 fps, à l'aide d'une caméra RGB. Cet échantillon est composé de 14 hommes et 1 femme d'âge et catégories socioprofessionnelles hétérogènes. Les mesures ont été réalisées sous seing privée chez des volontaires sains. Les données sont traitées par des algorithmes de traitement d'images et du signal visant à transformer la vidéo en un signal PPG. La *figure 3* résume la préparation des données dans cette étude.

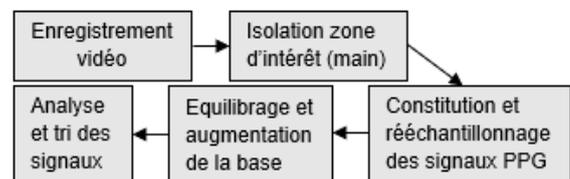
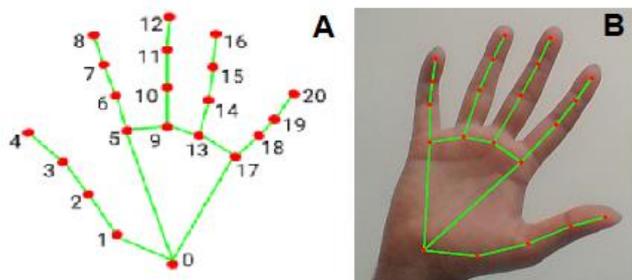


Figure 3 – Organigramme de la préparation des données.

Afin d'obtenir des signaux PPG épurés, il est important de déterminer une zone d'intérêt ; dans le cas présent, la main a été sélectionnée. Plusieurs méthodes de tracking et de segmentation ont été testées.



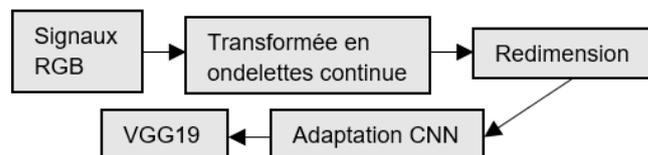
**Figure 4** – Description du modèle de main proposé par la librairie MediaPipe (a) et son application directe sur une image d'une base de données publique (b).

La solution MediaPipe a été retenue comme le montre la *figure 4*. C'est une bibliothèque multiplateforme qui fournit des solutions d'apprentissage automatique prêtes à l'emploi pour les tâches de vision par ordinateur. MediaPipe Hands permet ainsi un suivi des mains et des doigts très qualitatif. Elle utilise l'apprentissage automatique pour déduire 21 points de repère 3D d'une main à partir d'une seule image. Le tracking est une étape primordiale pour élaguer le bruit d'un signal PPG brut [5].

Une classification binaire des données de la SpO<sub>2</sub> a été réalisée dans le cadre de cette étude. Les signaux PPG ont été découpés en morceaux selon deux plages de SpO<sub>2</sub> : (1) une première nommée "élevée" qui désigne une SpO<sub>2</sub> de 95% ±2%, (2) une deuxième "basse" avec des valeurs de SpO<sub>2</sub> plus faibles. Ce travail a été réalisé manuellement, en vérifiant respectivement chaque signal avec les données de référence fournies. Un second découpage a eu lieu dans chacune des plages afin d'obtenir des signaux de taille (3,300), à savoir 3 signaux RGB et 300 échantillons. La technique de l'overlapping a permis d'accroître artificiellement le nombre de données pour bénéficier d'une meilleure efficacité du réseau neuronal convolutif (CNN). Enfin, pour assurer l'homogénéité des données acquises dans ces deux plages, d'autres overlapping ont été appliqués et divers signaux, notamment de la base "élevée", ont été supprimés.

L'approche proposée dans cette étude est schématisée en *figure 5*. La transformée en ondelettes continue du signal PPG de chaque canal RGB est indépendamment calculée. L'ondelette mère de Morlet a été utilisée. Une partie réelle et imaginaire sont calculées pour chaque signal. La représentation en ondelettes est redimensionnée pour atteindre une taille

de 256 par 256 pixels. *In fine*, les signaux PPG de dimension (300,3) sont transformés en représentation en ondelettes continue de taille (256, 256, 6). Elle servira d'entrée à un réseau CNN de type VGG19. Il s'agit d'un réseau neuronal convolutif de 19 couches, préentraîné sur des millions d'images issues de la base de données publique ImageNet. La fonction d'activation est de type softmax, ce qui permet d'avoir en sortie une classification binaire : "SpO<sub>2</sub> élevée", ou "SpO<sub>2</sub> basse".



**Figure 5** – Architecture de l'approche développée dans cette étude, nommée « Modèle LCOMS ».

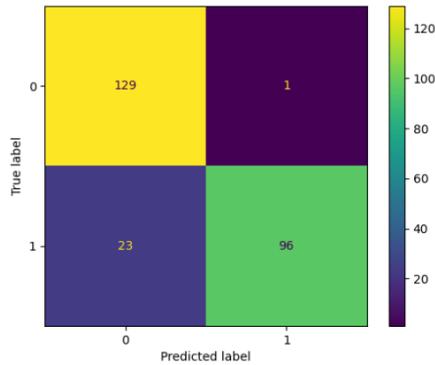
## 4. Résultats

Après réimplémentation des modèles issus de la littérature (cf. [11], [12]) ainsi que du modèle proposé (LCOMS), plusieurs apprentissages ont été lancés afin de tester la classification des données de SpO<sub>2</sub>. La distribution des données est de 70% en entraînement, 15% en validation et 15% en test. Nous proposons une seconde répartition des données visant à vérifier les capacités de généralisation du modèle d'IA en excluant totalement les données d'un participant des bases d'entraînement et de validation (stratégie du patient out). Les résultats des apprentissages sont présentés dans le *tableau 1*. Les premiers apprentissages ont été réalisés sur des données sans stratégie de recouvrement (1772 signaux dont 908 données "élevées" et 864 données "basses").

**Tableau 1** – Résultats des apprentissages à partir d'une base de données formée sans stratégie de recouvrement.

	Val accuracy	Test accuracy	Patient out
Model [11]	0,65	0,46	0,47
1 <sup>st</sup> model [12]	0,60	0,43	0,34
2 <sup>nd</sup> model [12]	0,52	0,50	0,51
3 <sup>rd</sup> model [12]	0,69	0,40	0,46
LCOMS model	0,90	0,90	0,80

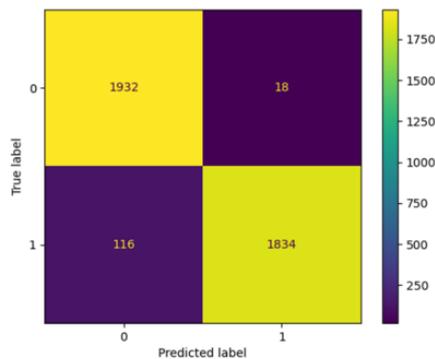
Les quatre modèles issus de la littérature scientifique qui ont été testés ne délivrent pas de performances satisfaisantes. Le modèle LCOMS présente de bien meilleures performances avec 90% de précision en phase d'entraînement et de test, et 80% pour la stratégie du patient out. La matrice de confusion inhérente à la performance du modèle LCOMS présentée en *figure 6* corrobore ces résultats.



**Figure 6** – Matrice de confusion du modèle LCOMS après apprentissages à partir d'une base de données formée sans stratégie de recouvrement.

Comme rappelé en amont de cette étude, nous avons fait une classification binaire de la SpO<sub>2</sub>. Pour cela, les valeurs 1 et 0 ont été respectivement assignées à une valeur de SpO<sub>2</sub> correspondant à 95%, ce qui reflète un état normal, et à des valeurs de SpO<sub>2</sub> plus faibles. La diagonale de la matrice de confusion montre dans un premier temps que le système a pu reconnaître 129 signaux “bas” en ne se trompant qu’une unique fois. Et relate dans un second temps, que ce système a pu reconnaître 96 signaux “élevés” bien qu’il se soit trompé 23 fois, ce qui présente ici déjà une avancée majeure par rapport aux résultats des modèles comparatifs.

Les seconds apprentissages ont été réalisés avec des données présentant une stratégie de recouvrement, permettant de passer de 1700 à 20000 signaux, dont le corollaire résulte en un gain significatif des performances du modèle.



**Figure 7** – Matrice de confusion du modèle LCOMS après apprentissages à partir d'une base de données formée avec une stratégie de recouvrement.

Les scores étaient très intéressants avec jusqu’à 96.5 % de performance globale comme le démontre la figure 7 contre 90% de performance auparavant.

## 5. Conclusion

Ce travail a mis en lumière un système mesurant la SpO<sub>2</sub> par analyse vidéo RGB via une approche basée sur la représentation en ondelettes des signaux PPG et un modèle neuronal convolutif. Cette dernière a permis une classification binaire de la SpO<sub>2</sub> et a montré

des résultats très encourageants. Il convient de mentionner que la norme ISO en vigueur pour les oxymètres de pouls suggère une erreur tolérable jusqu’à 3%, ce en quoi s’inscrit pratiquement les estimations de SpO<sub>2</sub> à distance par l’intermédiaire du modèle LCOMS présenté dans cette étude. Les perspectives futures s’articulent maintenant autour d’une solution pour lever la barrière de la classification binaire en développant le modèle pour qu’il puisse fournir des estimations continues de la SpO<sub>2</sub>.

## 6. Références

- [1] Bousefsaf, F., et al. (2019). 3D convolutional neural networks for remote pulse rate measurement and mapping from facial video. *Applied Sciences*, 9(20), 4364.
- [2] Bousefsaf, F. (2014). *Mesure sans contact de l'activité cardiaque par analyse du flux vidéo issu d'une caméra numérique: extraction de paramètres physiologiques et application à l'estimation du stress* (Doctoral dissertation, Université de Lorraine).
- [3] Rossol, S. L., et al. (2020). Non-contact video-based neonatal respiratory monitoring. *Children*, 7(10), 171.
- [4] Sugita, Norihiro, et al. "Contactless technique for measuring blood-pressure variability from one region in video plethysmography." *Journal of Medical and Biological Engineering* 39 (2019): 76-85.
- [5] Van Gastel, M., et al. (2016). New principle for measuring arterial blood oxygenation, enabling motion-robust remote monitoring. *Scientific reports*, 6(1), 38609.
- [6] Nitzan, M., et al. (2014). Pulse oximetry: fundamentals and technology update. *Medical Devices: Evidence and Research*, 231-239.
- [7] Sinex, J. E. (1999). Pulse oximetry: principles and limitations. *The American journal of emergency medicine*, 17(1), 59-66.
- [8] Webster, J. G. (Ed.). (1997). *Design of pulse oximeters*. CRC Press.
- [9] Humphreys, K., et al. (2007). Noncontact simultaneous dual wavelength photoplethysmography: a further step toward noncontact pulse oximetry. *Review of scientific instruments*, 78(4), 044304.
- [10] De Haan, G. (2014). Improved motion robustness of remote-ppg by using the blood volume pulse signature. *Physiological Measurement*.
- [11] Ding, X., et al. (2018). Measuring oxygen saturation with smartphone cameras using convolutional neural networks. *IEEE journal of biomedical and health informatics*, 23(6), 2603-2610.
- [12] Mathew, J., et al. (2023). Remote blood oxygen estimation from videos using neural networks. *IEEE Journal of Biomedical and Health Informatics*.
- [13] Tarassenko, L., et al. (2014). Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiological measurement*, 35(5), 807.
- [14] Muehlematter, U. J., et al. (2021). Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *The Lancet Digital Health*, 3(3), e195-e203.