



# UN ENFOQUE DE APRENDIZAJE PROFUNDO PARA ESTIMAR LA FRECUENCIA RESPIRATORIA DEL FOTOPLETISMOGRAMA

## A DEEP LEARNING APPROACH TO ESTIMATE THE RESPIRATORY RATE FROM PHOTOPLETHYSMOGRAM

Lucas C. Lampier<sup>1,\*</sup> , Yves L. Coelho<sup>1</sup> , Eliete M. O. Caldeira<sup>1</sup> ,  
Teodiano F. Bastos-Filho<sup>1</sup> 

Recibido: 15-05-2021, Recibido tras revisión: 18-08-2021, Aceptado: 20-09-2021, Publicado: 01-01-2022

### Resumen

Este trabajo presenta una metodología para entrenar y probar una red neuronal profunda (*Deep Neural Network* – DNN) con datos de fotopleletismografías (*Photoplethysmography* – PPG), con la finalidad de llevar a cabo una tarea de regresión para estimar la frecuencia respiratoria (*Respiratory Rate* – RR). La arquitectura de la DNN está basada en un modelo utilizado para inferir la frecuencia cardíaca (FC) a partir de señales PPG ruidosas. Dicho modelo se ha optimizado a través de algoritmos genéticos. En las pruebas realizadas se usaron BIDMC y CapnoBase, dos conjuntos de datos de acceso abierto. Con CapnoBase, la DNN logró un error de la mediana de 1,16 respiraciones/min, que es comparable con los métodos analíticos reportados en la literatura, donde el mejor error es 1,1 respiraciones/min (excluyendo el 8 % de datos más ruidosos). Por otro lado, el conjunto de datos BIDMC aparenta ser más desafiante, ya que el error mínimo de la mediana de los métodos reportados en la literatura es de 2,3 respiraciones/min (excluyendo el 6 % de datos más ruidosos). Para este conjunto de datos la DNN logra un error de mediana de 1,52 respiraciones/min.

**Palabras clave:** redes neuronales profundas, fotopleletismografía, frecuencia respiratoria

### Abstract

This article describes the methodology used to train and test a Deep Neural Network (DNN) with Photoplethysmography (PPG) data performing a regression task to estimate the Respiratory Rate (RR). The DNN architecture is based on a model used to infer the heart rate (HR) from noisy PPG signals, which is optimized to the RR problem using genetic optimization. Two open-access datasets were used in the tests, the BIDMC and the CapnoBase. With the CapnoBase dataset, the DNN achieved a median error of 1.16 breaths/min, which is comparable with analytical methods in the literature, in which the best error found is 1.1 breaths/min (excluding the 8 % noisiest data). The BIDMC dataset seems to be more challenging, as the minimum median error of the literature's methods is 2.3 breaths/min (excluding 6 % of the noisiest data), and the DNN based approach achieved a median error of 1.52 breaths/min with the whole dataset.

**Keywords:** Deep Neural Networks, Photoplethysmography, Respiratory Rate

<sup>1,\*</sup>Postgraduate Program in Electrical Engineering, Universidade Federal do Espírito Santo (UFES), Vitória-ES, Brazil  
Autor para correspondencia ✉: lucas.lampier@hotmail.com.

Forma sugerida de citación: Lampier, L. C.; Coelho, Y. L.; Caldeira, E. M. O. y Bastos-Filho, T. F. "Un enfoque de aprendizaje profundo para estimar la frecuencia respiratoria del fotopleletismograma." *Ingenius, Revista de Ciencia y Tecnología*, N.º 27, pp. 96-104, 2022. DOI: <https://doi.org/10.17163/ings.n27.2022.09>.

## 1. Introducción

La frecuencia respiratoria (Respiratory Rate, RR) es un indicador importante del estado fisiológico de una persona, de manera especial en el monitoreo de enfermedades pulmonares. Esta señal fisiológica se mide tradicionalmente por medio de una espirometría, neumografía o electromiografía [1]. Estas metodologías son costosas y utilizadas en ambientes médicos. La fotoplethismografía (Photoplethysmography, PPG) es una técnica que mide la frecuencia cardíaca utilizando un oxímetro, mucho más económico que los métodos tradicionales de medición de la RR. Estudios recientes muestran que puede ser captada aun remotamente utilizando la cámara de un teléfono inteligente [2]. Considerando que la respiración tiene influencia sobre los ciclos cardíacos, la información de la RR también se presenta en la señal PPG [3].

Varias metodologías reportadas en la literatura han estimado exitosamente la RR con un error pequeño a partir de la señal PPG [1], [4–8], sin embargo, la atenuación del ruido es todavía un reto para obtener predicciones robustas a partir de señales PPG de baja calidad. Este artículo presenta una manera alternativa para inferir la frecuencia respiratoria a partir de señales PPG. Las técnicas de aprendizaje profundo (Deep Learning) han sido ampliamente utilizadas en problemas de procesamiento de imágenes, sin embargo, también son una herramienta poderosa en problemas de una dimensión (1D), limitadas principalmente por el tamaño y la calidad del conjunto de datos, y también por el poder computacional disponible durante el proceso de entrenamiento. A medida que crece la data disponible en línea, así como también el poder computacional del hardware, las DNN están siendo probadas en una gran variedad de problemas que anteriormente estaban dominados por procedimientos analíticos.

Con respecto al procesamiento de señales PPG, Ravichandran *et al.* [1] mostraron que una DNN puede utilizarse para extraer la señal RR a partir de señales PPG. Biswas *et al.* [9] extraen el pulso a partir de la PPG utilizando una DNN.

Este trabajo propone el uso de una DNN para extraer mediciones confiables de RR a partir de señales PPG, utilizando procedimientos analíticos para extraer señales secundarias de la PPG y una DNN para inferir la RR.

### 1.1. Trabajos afines

De acuerdo con Charlton *et al.* [10], los algoritmos PPG-RR pueden resumirse en cinco pasos:

- *Extraer las señales respiratorias.* Consiste en encontrar la variación de la señal PPG relacionada a los ciclos respiratorios.

- *Fusionar las señales respiratorias.* Las diferentes señales extraídas pueden combinarse para crear una señal única con mayor robustez ante el ruido (opcional).
- *Estimar la RR a partir de una ventana.* Se segmenta una ventana de la señal generada y se estima la RR a partir de esa aquella.
- *Combinar las estimaciones.* El resultado de diferentes ventanas puede utilizarse para generar una estimación final (opcional).
- *Filtrado de calidad.* Puede asignarse un score de calidad a la ventana PPG, para excluir predicciones de baja calidad (opcional).

Algunos de los primeros enfoques que estiman la RR a partir de PPG utilizan dos filtros Butterworth de segundo orden, uno pasa-alto y uno pasa-bajo con frecuencias de corte 0,1 Hz y 5 Hz, respectivamente, para eliminar el ruido de la PPG. Luego se estima la señal respiratoria aplicando un filtro pasa-bajo de 0,4 Hz a la señal PPG [11]. Madhav *et al.* [12] aplicaron una relación de valor singular (*Singular Value Ratio*, SVR) para extraer la periodicidad respiratoria de la PPG, y utilizaron análisis de los componentes principales (*Principal Component Analysis*, PCA) para estimar la actividad respiratoria del primer componente principal. En otro enfoque propuesto por Nam y colaboradores [13], se capturó una señal PPG utilizando la cámara de un teléfono inteligente, y luego se estimó la RR encontrando la frecuencia correspondiente al pico más alto del espectro generado por el periodograma de Welch [7].

Se compararon cinco metodologías diferentes en el trabajo realizado por Pimentel y colaboradores [4], incluyendo la desarrollada por los autores. La más antigua fue propuesta por el equipo encabezado por Nilsson [14], y consiste en aplicar un filtro Bessel pasa-banda de orden 16 con frecuencias de corte 0,13 Hz y 0,48 Hz (7,8 a 28,8 respiraciones/min). La segunda propuesta [8], utiliza la transformada de Fourier sobre la señal PPG y obtiene la frecuencia con la mayor amplitud en el ancho de banda entre 0,08 Hz y 0,4 Hz. En la metodología sugerida por Tarassenko [5], los modelos autorregresivos (AR) estiman un filtro utilizando la señal PPG, y de acuerdo con ellos la frecuencia del polo de mayor magnitud dentro del intervalo del ancho de banda de la RR (0,08 Hz a 0,7 Hz) corresponde a la RR. En lugar de utilizar la propia PPG [6] se estiman tres características temporales relacionadas con la RR: variación de la intensidad inducida por la respiración (*Respiratory-Induced Intensity Variation*, RIIV), variación de la amplitud inducida por la respiración (*Respiratory-Induced Amplitude Variation*, RIAV) y variación de la frecuencia inducida por la respiración (*Respiratory-Induced Frequency Variation*,

RIFV). Luego, se calcula el espectro de potencia de cada una y se estima la RR a partir de cada una tomando la frecuencia de mayor amplitud en la banda de frecuencias de la RR (0,067 Hz – 1,08 Hz) y calculando la media de las tres frecuencias. Finalmente, la metodología mostrada por Pimentel *et al.* [4] combina el modelado AR propuesto en Tarassenko [5], con la fusión de resultados de las tres características propuestas por Karlen y su equipo [6].

Los resultados presentados [4] muestran que para el conjunto de datos CapnoBase [15], el método que presentó el error más bajo [6], es decir, con un error absoluto de la mediana (*Median Absolute Error*, MdAE) de 0,8 respiraciones/min utilizando una ventana de 62 s; sin embargo, 46 % de los datos de PPG fueron excluidos debido al ruido. El segundo mejor en este conjunto de datos fue el citado por Tarassenko [5], que alcanzó un MdAE de 1,1 respiraciones/min y conservó 92 % de la data. El conjunto de datos BIDMC parece ser más desafiante, debido a que el mejor MdAE fue obtenido por el método propuesto por Shelley [8] (2,3 respiraciones/min), el cual también fue uno de los métodos que conservó la mayoría de la data (94 %).

Un trabajo reciente ha obtenido resultados aún más precisos [16], en el que un método deriva múltiples ondas de una señal PPG, mide su calidad y luego la utiliza como ponderación para combinarlas utilizando un suavizador de Kalman. Con este enfoque obtuvieron un error mediano de 0,2 respiraciones/min.

## 1.2. Redes neuronales profundas (Deep Neural Networks, DNN)

Para Goodfellow *et al.* [17], el aprendizaje profundo es un enfoque que puede manejar un paso importante en un problema de aprendizaje de máquina: la extracción de características. Su concepto consiste en concatenar múltiples capas de modelos simples, donde cada modelo «aprende» parte de los problemas-conceptos y, por tanto, un problema complejo puede dividirse en modelos más simples para ser resuelto de mejor manera. Un punto interesante de este enfoque es que realiza las tareas de aprendizaje automático (clasificación, regresión, agrupamiento, etc.) y también la extracción de características, en contraste con las técnicas poco profundas, en el que el paso de extracción de características debe hacerse de manera separada, en una etapa previa, y exige conocimiento del área. La limitación del enfoque de aprendizaje profundo es el conjunto de datos. Normalmente, los modelos de aprendizaje profundo tienen un gran número de valores a ser ajustados y, por tanto, necesitan grandes cantidades de datos para optimizarlos.

Existen algunos enfoques que instrumentaron técnicas de aprendizaje profundo para resolver problemas relacionados con data fisiológica. Biswas *et al.* [9] implementaron una DNN para estimar la frecuencia

cardíaca (*Heart Rate*, HR), teniendo como data de entrada una señal PPG obtenida de un sensor colocado en la muñeca del usuario. El modelo fue probado en un conjunto de datos desafiante, en el cual la señal PPG fue capturada durante varias actividades físicas, las cuales insertan ruido y artefactos en ella. La arquitectura del modelo estaba constituida por dos capas convolucionales de una dimensión seguidas de dos capas de gran memoria de corto plazo (*Long Short-Term Memory*, LSTM) y, finalmente, una capa densa. Esta arquitectura alcanzó un error absoluto medio (*Mean Absolute Error*, MAE) de  $1,47 \pm 3,37$  latidos por minuto (*Beats Per Minute*, BPM).

En el trabajo [1] se presenta un enfoque de aprendizaje profundo para extraer la señal RR a partir de la señal PPG. Su modelo consta de varias capas convolucionales conectadas a varias capas deconvolucionales que transforman la señal PPG en la señal respiratoria, y luego calculan a partir de ella la RR. Su método fue probado en dos conjuntos de datos independientes, CapnoBase y Vortal [18], y alcanzó un error cuadrático medio (*Mean Squared Error*, MSE) de 0,262 en el primer conjunto de datos y de 0,145 en el segundo, lo que corresponde a una raíz cuadrada del error cuadrático medio (*Root Mean Squared Error*, RMSE) de aproximadamente 3,1 respiraciones/min y 2,3 respiraciones/min, respectivamente.

## 2. Materiales y métodos

### 2.1. Conjuntos de datos

En las pruebas se utilizaron dos conjuntos de datos: el benchmark CapnoBase [15] y el de respiración BIDMC PPG [19]. El primero consta de 42 registros [20], de 8 minutos cada uno, que contienen la PPG y la señal de dióxido de carbono (CO<sub>2</sub>) inhalado y exhalado, ambos recolectados a una frecuencia de muestreo de 300 Hz. El conjunto BIDMC PPG [21] contiene 8 minutos de registros de señales PPG y respiración de CO<sub>2</sub> de 53 voluntarios, recolectados a 125 Hz.

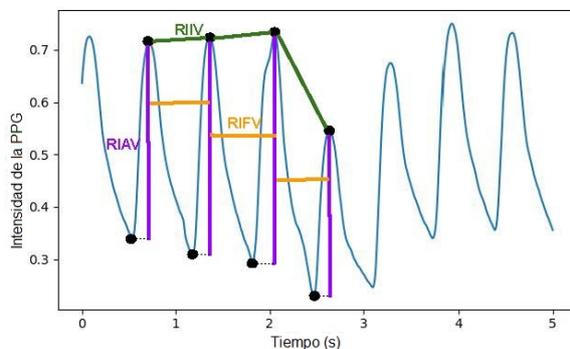
### 2.2. Características respiratorias del foletoletismograma

A partir de la PPG se generaron tres señales relacionadas con la respiración [6]:

1. Variación de la intensidad inducida por la respiración (*Respiratory-Induced Intensity Variation*, RIIV). A medida que la amplitud de la señal PPG varía en sincronía con el ciclo respiratorio, esta característica puede utilizarse para estimar la RR [3]. Esta serie es estimada tomando el valor de la intensidad de cada pico en la señal PPG [4].

2. Variación de la amplitud inducida por la respiración (*Respiratory-Induced Amplitude Variation*, RIAV). Los ciclos respiratorios inducen una variación de amplitud en las ondas PPG, la cual puede estimarse calculando la variación en la diferencia pico-valle en las ondas PPG [6].
3. Variación de la frecuencia inducida por la respiración (*Respiratory-Induced Frequency Variation*, RIFV). La variación en la HR también está altamente ligada a los eventos respiratorios de inspiración y expiración. Las inspiraciones aumentan la HR y las expiraciones la reducen. Esta señal puede representarse por el intervalo de tiempo entre picos consecutivos de la PPG [6].

La Figura 1 muestra una representación gráfica de las características RIIV, RIAV y RIFV.



**Figura 1.** Características relacionadas a la RR extraídas de la PPG

### 2.3. Índice de calidad de la señal (*Signal Quality Index, SQI*)

El SQI cuantifica la contaminación por ruido y artefactos en la ventana PPG analizada. Este índice se utiliza para excluir las señales que no alcanzan una calidad mínima. En este trabajo, el SQI se calcula abordando dos puntos: «líneas planas» en la señal, y concordancia entre diferentes detectores de picos [4]. Las «líneas planas» se definen como las partes de la señal PPG en las que la diferencia entre muestras consecutivas es cercana a cero. Para detectarlas se realizó el siguiente método:

1. La señal *StdDiff* es calculada como sigue:

$$StdDiff[i-wdw] = \sigma(diff(PPG[i-wdw : i])),$$

for  $i = wdw, wdw + 1, wdw + 2, \dots, N$ .

Donde  $\sigma(\cdot)$  es la operación que calcula la desviación estándar,  $diff(\cdot)$  es la primera derivada de la operación,  $PPG$  es la señal  $PPG$ ,  $wdw$  es la longitud de la ventana, que, en este caso, es la parte entera de  $1,5 s \times frecuencia de$

*muestreo* (1,5 s es el intervalo de tiempo entre picos consecutivos en el valor mínimo de HR, que normalmente es 40 BPM), y  $N$  es el número de muestras en la señal PPG.

2. Se inicializa el arreglo  $Q$  con unos y el mismo tamaño que la  $PPG$ :  $Q = ones(N)$
3. Se calcula el valor umbral:  $th = \mu(StdDiff) - \alpha \times \sigma(StdDiff)$ , donde  $\mu(\cdot)$  es la operación que calcula la media y  $\alpha$  es un parámetro ajustable.
4. Los valores de  $k$  se hacen cero en las ubicaciones correspondientes donde  $StdDiff$  es muy bajo:
 

```
for i = wdw, wdw + 1, wdw + 2.....N : {
    if (StdDiff[i - wdw] < (th)) : {
        Q[i + wdw] = 0
    }
}
```
5. El valor  $K$  es la proporción entre los valores iguales a  $Q$  dentro de la ventana PPG analizada, y su tamaño total.

La concordancia entre dos detectores de picos fue realizada de la siguiente manera: si la diferencia entre la posición encontrada por los detectores en el mismo pico es menor que 150 ms, se considera que ambos detectores concuerdan en ese pico y, por lo tanto, su posición es correcta; de lo contrario, el pico se considera como un ruido [4]. Entonces se aplica un score F1 para medir la concordancia entre dos detectores de picos en el segmento analizado de la señal. Los métodos de detección de picos fueron proporcionados por dos librerías diferentes de Python: Neurokit2 [22], que implementa el algoritmo de detección de picos PPG descrito en [23] y la librería Heartpy [24], en la que los autores desarrollan su propia técnica de detección de picos.

Finalmente, el SQI se calcula como:  $SQI = F1 \times K$ , donde F1 es la concordancia entre picos y K es la proporción de «línea plana».

### 2.4. Arquitectura DNN

La arquitectura del modelo se inspiró en la CorNET [9], que tiene dos bloques convolucionales, cada uno de los cuales contiene una capa convolucional (32 filtros con un kernel de tamaño 40), una capa de normalización *batch*, una capa de activación con unidades lineales rectificadas (*Rectified Linear Units*, ReLU) y, finalmente, una capa *maxpooling* con un tamaño de pool de 4 (para el entrenamiento se utilizó una capa *dropout* con una tasa de eliminación (*drop rate*) de 0.1). Luego de los bloques convolucionales se concatenan dos capas LSTM, cada una con 128 unidades y función de activación tangente hiperbólica (*tanh*). Para la predicción se utilizó una neurona con activación lineal.

La arquitectura propuesta para predecir la RR es el resultado de la optimización mediante algoritmos

genéticos, de los parámetros de la CorNET que se presentan en la Tabla 1.

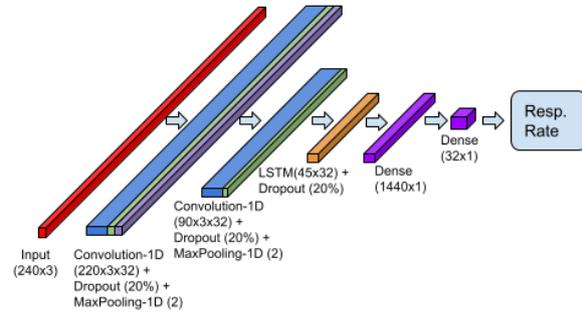
**Tabla 1.** Parámetros probados en la arquitectura DNN

Parámetros	Posibilidades
# de bloques convolucionales	0/1/2
# de filtros convolucionales por capa	128/64/32
Longitud de los filtros convolucionales	11/21/31
# de capas LSTM	1/2/3
# de unidades LSTM por capa	128/64/32
# de capas densas	0/1/2
# de neuronas por capa	128/64/32/16

Los parámetros mencionados arriba son optimizados, mientras que para los parámetros no mencionados se utiliza la definición por defecto en la librería Keras [25]. Considerando que el número de combinaciones es muy grande (2916), los parámetros se optimizan utilizando algoritmos genéticos. Esta estrategia comienza definiendo una población de  $L$  modelos con parámetros aleatorios, y luego todos los modelos son entrenados, probados y categorizados de acuerdo con el resultado. Los  $M$  modelos con mejor desempeño se conservan mientras que los otros se descartan. El conjunto de los modelos conservados se denomina padres, y luego se generan  $M - J$  nuevos modelos tomando valores de tres modelos diferentes (un padre para los parámetros convolucionales, uno para las capas LSTM, y otro para las capas densas). Adicionalmente, existe una probabilidad de  $H$  % de que los parámetros de los nuevos modelos cambien aleatoriamente (estos cambios se denominan mutaciones); estos nuevos modelos se denominan descendientes. Asimismo, existen  $J$  modelos nuevos generados de forma completamente aleatoria, y a este conjunto se le asigna el nombre de extranjeros. Luego, los conjuntos de padres, descendientes y extranjeros se unen comienza una nueva interacción.

El procedimiento se repite por  $E$  épocas (*epochs*). La idea es que los parámetros que disminuyen el error de estimación prevalezcan en el conjunto de padres, y su combinación pueda generar un modelo en el conjunto de descendientes que mejore el resultado de sus padres. Asimismo, nuevos parámetros aleatorios se insertan con las mutaciones y los extranjeros creados en cada iteración. Estas adiciones aleatorias establecen una posibilidad de que la población de modelos salga de un mínimo local subóptimo. Para este experimento, los parámetros de la optimización mediante algoritmos genéticos fueron los siguientes:  $L = 45$ ,  $M = 20$ ,  $J = 5$ ,  $E = 20$  y  $H = 20$ , lo cual redujo el número de modelos entrenados de 2916 a 520 (25 nuevos modelos/época  $\times$  20 épocas + los primeros 20 padres). El mejor modelo encontrado es una DNN con 2 bloques convolucionales con 32 filtros y un MaxPooling de una dimensión (tamaño de kernel 2) en cada capa convolucional, un tamaño de kernel de 21, y activación ReLU. Luego, una capa LSTM con 32 neuronas, y finalmente,

2 capas densas, una con 32 neuronas y activación ReLU y la última con una neurona y activación lineal. La Figura 2 presenta un resumen de la arquitectura DNN final.



**Figura 2.** Arquitectura de red neuronal seleccionada

## 2.5. Procedimiento de entrenamiento y prueba

Se utilizaron dos conjuntos de datos para entrenar y evaluar los modelos BIDMC [19] y CapnoBase [15]. Se empleó el método de validación cruzada con cuatro grupos: para cada base de datos, sus archivos se dividieron en cuatro grupos, cada uno de los cuales contenía el 25 % de los datos en cada conjunto. Luego, se realizó el mismo procedimiento en cada grupo: primero, se encontraron los picos de la PPG utilizando la librería Neurokit2 [22]. Posteriormente, utilizando una ventana deslizante de 64 s con un paso de 4 s, se calcularon las tres señales relacionadas a la RR (RIIV, RIAV y RIFV), que luego se interpolaron a una frecuencia de muestreo artificial de 4 Hz y se extrajeron las primeras 240 muestras (60 s) de la ventana, generando así una muestra con 240 líneas (pasos de tiempo) y tres columnas (correspondientes a RIIV, RIAV y RIFV). El paso final consiste en normalizar la data, de manera que cada canal de la nueva muestra tenga media 0 y varianza 1:  $Y = (X - \mu(X))/\sigma(X)$ , donde  $Y$  es la serie normalizada,  $X$  la serie de entrada,  $\mu(\cdot)$  es la operación que calcula la media y  $\sigma(\cdot)$  es la operación que calcula la desviación estándar.

Para obtener el valor de referencia de la RR, se localiza el pico de  $\text{CO}_2$  correspondiente a cada muestra utilizando el módulo Respiratorio de Neurokit2, y luego se calcula la mediana de la RR de la muestra de 64 s, tomando la inversa del valor de la mediana en el intervalo de tiempo entre picos consecutivos.

Posteriormente se dividen los cuatro grupos de muestras normalizadas en conjuntos de entrenamiento y prueba, donde el conjunto de entrenamiento tiene tres grupos y el de prueba tiene uno. Los procedimientos de entrenamiento y prueba se repitieron cuatro veces, para que cada grupo pudiera ser evaluado una vez en el conjunto de prueba. Cada uno de los 4 modelos (uno para cada grupo de prueba) fue entrenado por 1400 épocas utilizando el optimizador Adam con

parámetros estándar de Keras [25]. El procesamiento se realizó en un computador Ryzen 5 3500X con 16 GB de memoria RAM y un NVIDIA GeForce RTX 2060 SUPER con 8 GB de memoria dedicada. Se utilizó la métrica del error cuadrático medio para calcular la pérdida (loss) en cada época. La Figura 3 presenta la curva de pérdida resultante.

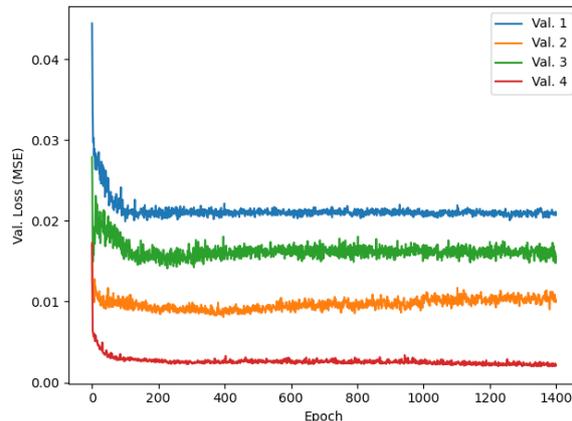


Figura 3. Curva Ts para las 4 instancias de la red

Los grupos de validación 1 y 4 tuvieron una pérdida decreciente hasta la época 1400, el grupo 2 tuvo su pérdida mínima alrededor de la época 400 y el grupo de validación 3 tuvo su mínimo alrededor de la época 200, y luego decreció nuevamente después de la época 800. El gráfico se utilizó para determinar el número de épocas de entrenamiento de las DNN, dado que la mitad de los grupos de validación se mantienen decreciendo hasta la época 1400 y la capa *dropout* parece evitar el sobreajuste en los otros grupos, debido a que la pérdida no se incrementa mucho luego de los mínimos. El proceso se repitió revolviendo la data y los grupos, y entrenando cuatro DNN (una para cada grupo de entrenamiento-prueba) durante 1400 épocas para realizar la validación cruzada.

### 3. Resultados y discusión

La Tabla 2 y la Tabla 3 presentan el MdAE y la desviación estándar del error, así como también la raíz cuadrada del error cuadrático medio y el umbral SQI utilizado para filtrar las muestras ruidosas. Probando diferentes valores de SQI es posible analizar la influencia de la calidad de la señal en el resultado.

Los resultados de MdAE alcanzados por nuestro modelo sobre el conjunto de datos CapnoBase son compatibles con los de las mejores metodologías comparadas [4], en las cuales el mejor MdAE para una ventana de 64 s fue 0,8 respiraciones/min, obtenidas utilizando el método [6]. Sin embargo, la metodología excluye el 36 % de las muestras que no alcanzan el

mínimo SQI. El segundo mejor fue el método [5], alcanzando un MdAE de 1,1 respiraciones/min para el 92 % de las mejores muestras.

Tabla 2. Medidas de error en las predicciones de la RR utilizando la DNN

Conjunto de datos BIDMC				
SQI	N (%)	MdAE (latidos/min*)	STD (latidos/min)	RMSE (latidos/min)
0,90	87	1,39	6,55	6,59
0,93	80	1,39	6,57	6,62
0,97	66	1,26	6,51	6,55
1,00	37	1,04	4,47	4,47
0,00	100	1,52	6,86	6,94

\* latidos/min es igual a respiraciones/min

Para el mismo conjunto de datos, nuestro método obtuvo un MdAE de 1,16 respiraciones/min (calculado concatenando los resultados de todas las cuatro pruebas), y un MdAE de 1,11 respiraciones/min con el 93 % de las muestras con el SQI más alto. Sin embargo, el RMSE del modelo DNN propuesto no alcanzó el error del RespNET [1], que es 3,1 respiraciones/min, indicando que nuestro modelo todavía puede ser refinado para alcanzar mejores resultados.

Tabla 3. Medidas de error en las predicciones de la RR utilizando la DNN

Conjunto de datos CapnoBase				
SQI	N (%)	MdAE (latidos/min*)	STD (latidos/min)	RMSE (latidos/min)
0,90	98	1,16	5,93	5,94
0,93	97	1,16	5,94	5,95
0,97	93	1,11	5,65	5,66
1,00	73	1,06	4,46	4,97
0,00	100	1,16	5,88	5,90

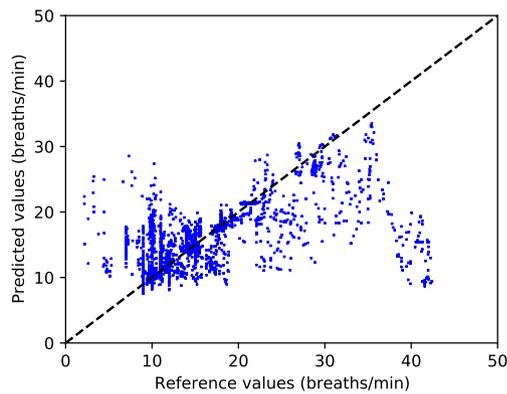
\* latidos/min es igual a respiraciones/min

El modelo DNN propuesto alcanzó errores menores en el conjunto de datos BIDMC, que los otros métodos comparados [4]. Con todas las muestras, nuestro método obtuvo un MdAE de 1,52 respiraciones/min en este conjunto de datos, mientras que el error más pequeño obtenido [4] para este conjunto de datos fue para la metodología propuesta [8], con un MdAE de 2,3 respiraciones/min. Esta mejora en el resultado muestra el éxito de los métodos DNN para procesar data fisiológica. En otras palabras, con un conjunto de datos suficientemente grande y con la complejidad adecuada, puede extraerse información importante de la data, mientras se descartan el ruido y los artefactos.

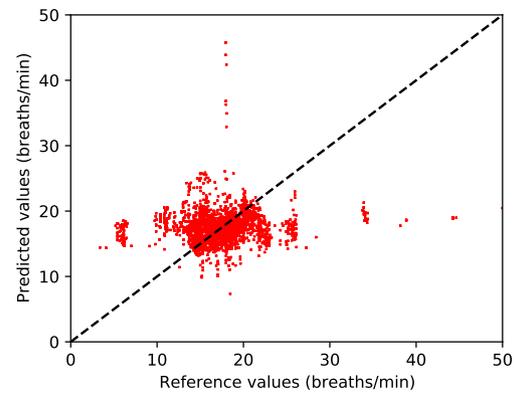
También puede notarse en la Tabla 2 y en la Tabla 3 que el MdAE por sí mismo no mide un buen resultado. Además, al tener un MdAE bajo los resultados presentan un RMSE considerable, ya que se refuerza la alta variación de los resultados y el cálculo de la mediana no toma en cuenta lo mucho que se desvían los resultados extremos del valor deseado, solo se calculan los valores en el centro. Esto puede observarse

en el gráfico de dispersión entre las predicciones y los valores verdaderos presentados en la Figura 4 y en la Figura 5.

Para comparar mejor el desempeño de nuestro método contra la literatura revisada, la Tabla 4 presenta un resumen de los resultados de cada método. El mejor resultado en el conjunto de datos CapnoBase obtenido por el método propuesto por Khreis *et al.* [16], que alcanzó un MdAE de 0,2 respiraciones/min, mientras que los métodos de referencia previos obtuvieron uno de 0,8 respiraciones/min en el mismo conjunto de datos (excluyendo de la data el 36 % más ruidoso). El método desarrollado alcanzó un MdAE de 1,2 respiraciones/min utilizando la data completa.



**Figura 4.** Dispersión de las predicciones y los valores verdaderos de la RR en el conjunto de datos CapnoBase. Las muestras graficadas tienen un SQI por encima de 0,9.



**Figura 5.** Dispersión de las predicciones y los valores verdaderos de la RR en el conjunto de datos BIDMC. Las muestras graficadas tienen un SQI por encima de 0,9

La base de datos BIDMC parece más desafiante, dado que los métodos probados sobre este conjunto de datos mostraron un MdAE mínimo de 2,3 respiraciones/min, y nuestro método alcanzó un MdAE de 1,5 respiraciones/min. Khreis *et al.* [16] no utilizaron el conjunto de datos BIDMC en su estudio, por lo que no fue posible comparar los resultados con su método sobre este conjunto de datos.

**Tabla 4.** Medidas de error de diferentes métodos presentados en la literatura. Se incluye N, que es el porcentaje de ventanas PPG utilizadas, el error absoluto mediano (*Median Absolute Error*, MdAE) y sus percentiles P<sub>25</sub> y P<sub>75</sub>, la raíz cuadrada del error cuadrático medio (*Root Mean Squared Error*, RMSE) y error absoluto medio (*Mean Absolute Error*, MAE) y su desviación estándar (*standard deviation*, STD)

Método	Conjunto de datos							
	Capnobase				BIDMC			
N (%)	MdAE (25TH - 75TH)	RMSE	MAE (STD)	N (%)	MdAE (25TH - 75TH)	RMSE	MAE (STD)	
Propuesto	100	1,2 (0,4-3,4)	5,9	3,1 (5,0)	100	1,5 (0,6-3,6)	6,9	3,4 (6,0)
Khreis (2020)	100	0,2 (0,1-0,9)	-	-	-	-	-	-
Ravichandra (2019)	-	-	3,1	-	-	-	-	-
Pimentel (2017)	92	1,9 (0,3-3,4)	-	-	94	2,7 (1,5-5,3)	-	-
Shelley (2016)	92	2,2 (0,2-8,3)	-	-	94	2,3 (0,9-7,9)	-	-
Karlen (2013)	64	0,8 (0,3-2,7)	-	-	34	5,7 (1,5-9,7)	-	-
Fleming (2007)	92	1,1 (0,4-3,5)	-	-	94	5,5 (2,7-8,1)	-	-
Nilsson (2000)	92	10,2 (4,8-12,4)	-	-	94	4,6 (2,5-8,5)	-	-

## 4. Conclusiones

Este trabajo presenta una metodología para utilizar un enfoque DNN para estimar la RR utilizando señales PPG. El modelo se inspiró en una arquitectura exitosa previa, que fue optimizada para nuestro problema utilizando una optimización genética adaptada. Las entradas a la DNN son tres señales respiratorias relacionadas con la RR extraídas de la PPG, y la salida es el valor de RR correspondiente. Para probar la metodología, los archivos de dos conjuntos de datos de acceso abierto se dividieron en cuatro grupos, manteniendo la misma proporción de los archivos de cada conjunto de datos en cada grupo. Se repitió cuatro veces un procedimiento de entrenamiento-prueba, cada uno con tres grupos para entrenar y uno para probar, por lo que cada grupo se utilizó una vez como prueba. Los resultados alcanzados son comparables con la mayoría de las metodologías analíticas de referencia (benchmark) utilizando el conjunto de datos CapnoBase, y la DNN los superó en el conjunto de datos BIDMC, que es más desafiante, mostrando el éxito de la metodología para procesar data fisiológica.

Aunque exitoso, usualmente, el desempeño de los métodos analíticos se acompaña por umbrales heurísticos o un número grande de parámetros libres entonados por un experto, los cuales pueden impedir la generalización de las metodologías desarrolladas. Sin embargo, los enfoques DNN están diseñados para generalizar la data lo mayor posible y tener tan poco preprocesamiento como sea posible.

## Agradecimientos

Los autores agradecen al CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), una agencia del Ministerio Brasileño de Ciencia, Tecnología, Innovaciones y Comunicaciones que apoya el desarrollo científico y tecnológico, FAPES (Fundação de Amparo à Pesquisa e Inovação do Espírito Santo), una agencia del Estado de Espírito Santo, Brasil, que apoya el desarrollo científico y tecnológico, y CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) por el apoyo financiero otorgado a este trabajo.

## Referenciass

[1] V. Ravichandran, B. Murugesan, V. Balakarthikeyan, K. Ram, S. P. Preejith, J. Joseph, and M. Sivaprakasam, “RespNet: A deep learning model for extraction of respiration from photoplethysmogram,” in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 5556–5559. [Online]. Available: <https://doi.org/10.1109/EMBC.2019.8856301>

[2] A. Floriano, L. Lampier, R. S. Rosa, E. Caldeira, and T. Bastos-Filho, “Remote vital sign monitoring in accidents,” *Polytechnica*, vol. 4, no. 1, pp. 26–32, Apr. 2021. [Online]. Available: <https://doi.org/10.1007/s41050-020-00027-1>

[3] D. J. Meredith, D. Clifton, P. Charlton, J. Brooks, C. W. Pugh, and L. Tarassenko, “Photoplethysmographic derivation of respiratory rate: a review of relevant physiology.” *Journal of medical engineering & technology*, vol. 36, no. 1, pp. 1–7, jan 2012. [Online]. Available: <https://doi.org/10.3109/03091902.2011.638965>

[4] M. A. Pimentel, A. E. Johnson, P. H. Charlton, D. Birrenkott, P. J. Watkinson, L. Tarassenko, and D. A. Clifton, “Toward a robust estimation of respiratory rate from pulse oximeters,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1914–1923, 2017. [Online]. Available: <https://doi.org/10.1109/TBME.2016.2613124>

[5] S. G. F. L. Tarassenko, “A comparison of signal processing techniques for the extraction of breathing rate from the photoplethysmogram,” *International Journal of Biological and Medical Sciences*, vol. 1, no. 6, pp. 366–370, 2007. [Online]. Available: <https://bit.ly/3Cprqqg>

[6] W. Karlen, S. Raman, J. M. Ansermino, and G. A. Dumont, “Multiparameter respiratory rate estimation from the photoplethysmogram.” *IEEE transactions on bio-medical engineering*, vol. 60, no. 7, pp. 1946–1953, jul 2013. [Online]. Available: <https://doi.org/10.1109/TBME.2013.2246160>

[7] L. Nilsson, T. Goscinski, S. Kalman, L. G. Lindberg, and A. Johansson, “Combined photoplethysmographic monitoring of respiration rate and pulse: A comparison between different measurement sites in spontaneously breathing subjects,” *Acta Anaesthesiologica Scandinavica*, vol. 51, no. 9, pp. 1250–1257, 2007. [Online]. Available: <https://doi.org/10.1111/j.1399-6576.2007.01375.x>

[8] K. H. Shelley, A. A. Awad, R. G. Stout, and D. G. Silverman, “The use of joint time frequency analysis to quantify the effect of ventilation on the pulse oximeter waveform.” *Journal of clinical monitoring and computing*, vol. 20, no. 2, pp. 81–87, apr 2006. [Online]. Available: <https://doi.org/10.1007/s10877-006-9010-7>

[9] D. Biswas, L. Everson, M. Liu, M. Panwar, B.-E. Verhoef, S. Patki, C. H. Kim, A. Acharyya, C. Van Hoof, M. Konijnenburg, and N. Van Helleputte, “Cornet: Deep learning framework for PPG-based heart rate estimation and biometric identification in ambulant environment,” *IEEE Transactions on*

- Biomedical Circuits and Systems*, vol. 13, no. 2, pp. 282–291, 2019. [Online]. Available: <https://doi.org/10.1109/TBCAS.2019.2892297>
- [10] P. H. Charlton, D. A. Birrenkott, T. Bonnici, M. A. F. Pimentel, A. E. W. Johnson, J. Alastruey, L. Tarassenko, P. J. Watkinson, R. Beale, and D. A. Clifton, “Breathing rate estimation from the electrocardiogram and photoplethysmogram: A review,” *IEEE Reviews in Biomedical Engineering*, vol. 11, pp. 2–20, 2018. [Online]. Available: <https://doi.org/10.1109/RBME.2017.2763681>
- [11] L. G. Lindberg, H. Ugnell, and P. A. Oberg, “Monitoring of respiratory and heart rates using a fibre-optic sensor,” *Medical & Biological Engineering & Computing*, vol. 30, no. 5, pp. 533–537, 1992. [Online]. Available: <https://doi.org/10.1007/BF02457833>
- [12] K. V. Madhav, M. R. Ram, E. H. Krishna, K. N. Reddy, and K. A. Reddy, “Estimation of respiratory rate from principal components of photoplethysmographic signals,” *Proceedings of 2010 IEEE EMBS Conference on Biomedical Engineering and Sciences, IECBES 2010*, no. December, pp. 311–314, 2010. [Online]. Available: <https://doi.org/10.1109/IECBES.2010.5742251>
- [13] Y. Nam, Y. Kong, B. Reyes, N. Reljin, and K. H. Chon, “Monitoring of heart and breathing rates using dual cameras on a smartphone,” *PLoS ONE*, vol. 11, no. 3, mar 2016. [Online]. Available: <https://doi.org/10.1371/journal.pone.0151013>
- [14] L. Nilsson, A. Johansson, and S. Kalman, “Monitoring of respiratory rate in postoperative care using a new photoplethysmographic technique,” *Journal of Clinical Monitoring and Computing*, vol. 16, no. 4, pp. 309–315, 2000. [Online]. Available: <https://doi.org/10.1023/A:1011424732717>
- [15] W. Karlen, “CapnoBase IEEE TBME Respiratory Rate Benchmark,” 2021. [Online]. Available: <https://doi.org/10.5683/SP2/NLB8IT>
- [16] S. Khreis, D. Ge, H. A. Rahman, and G. Carrault, “Breathing Rate Estimation Using Kalman Smoother with Electrocardiogram and Photoplethysmogram,” *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 3, pp. 893–904, 2020. [Online]. Available: <https://doi.org/10.1109/TBME.2019.2923448>
- [17] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. [Online]. Available: <https://bit.ly/3Eh4Twb>
- [18] P. H. Charlton, T. Bonnici, L. Tarassenko, D. A. Clifton, R. Beale, and P. J. Watkinson, “An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram,” *Physiological measurement*, vol. 37, no. 4, pp. 610–626, apr 2016. [Online]. Available: <https://doi.org/10.1088/0967-3334/37/4/610>
- [19] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23, pp. e215–e220, jun 2000. [Online]. Available: <https://doi.org/10.1161/01.cir.101.23.e215>
- [20] CapnoBase. (2020) Capnobase is a collaborative research project that provides an online database of respiratory signals and labels obtained from capnography, spirometry and pulse oximetry. [Online]. Available: <https://bit.ly/3EjffHKm>
- [21] M. A. F. Pimentel, A. E. W. Johnson, P. H. Charlton, D. Birrenkott, P. J. Watkinson, L. Tarassenko, and D. A. Clifton, “Toward a robust estimation of respiratory rate from pulse oximeters,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1914–1923, 2017. [Online]. Available: <https://doi.org/10.1109/TBME.2016.2613124>
- [22] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. H. A. Chen, “Neurokit2: A python toolbox for neurophysiological signal processing,” *Behavior Research Methods*, vol. 53, no. 4, pp. 1689–1696, Feb 2021. [Online]. Available: <https://doi.org/10.3758/s13428-020-01516-y>
- [23] M. Elgendi, I. Norton, M. Brearley, D. Abbott, and D. Schuurmans, “Systolic peak detection in acceleration photoplethysmograms measured from emergency responders in tropical conditions,” *PLOS ONE*, vol. 8, no. 10, pp. 1–11, 10 2013. [Online]. Available: <https://doi.org/10.1371/journal.pone.0076585>
- [24] P. van Gent, H. Farah, N. van Nes, and B. van Arem, “Heartpy: A novel heart rate algorithm for the analysis of noisy signals,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 66, pp. 368–378, 2019. [Online]. Available: <https://doi.org/10.1016/j.trf.2019.09.015>
- [25] F. Chollet, “Keras: Deep learning for humans,” GitHub. Inc, 2015. [Online]. Available: <https://bit.ly/3dA0g57>