

Comparación de Redes Neuronales Preentrenadas para la Evaluación de la Calidad en Registros Electrocardiográficos

A. Huerta Herraiz¹, A. Martínez Rodrigo¹, A. Puchol Calderón², M.I. Pachón Iglesias²,
J.J. Rieta Ibáñez³, R. Alcaraz Martínez¹

¹ Grupo de Investigación en Ingeniería Electrónica, Biomédica y de Telecomunicación, Universidad de Castilla-La Mancha, Cuenca, España, {alvaro.huerta, arturo.martinez, raul.alcaraz}@uclm.es

² Departamento de Arritmias Cardíacas, Hospital Virgen de la Salud, Toledo, España, apucholcal@gmail.com, mpachon@secardiologia.es

³ BioMIT.org, Dep. de Ingeniería Electrónica, Universitat Politècnica de Valencia, Valencia, España, jjrieta@upv.es

Resumen

Actualmente, es habitual el uso de las redes neuronales convolucionales para el análisis del registro electrocardiográfico (ECG). Sin embargo, su capacidad para evaluar la calidad con la que se registra esta señal fisiológica todavía no se ha examinado a fondo. Esta difícil tarea, cuyo objetivo es identificar los intervalos de ECG de baja calidad, podría ayudar a un diagnóstico más preciso de enfermedades cardíacas, especialmente si el ECG se ha registrado con un sistema de monitorización a largo plazo portátil o vestible. Así pues, en este trabajo se introduce un estudio comparativo en el que se analizan varios modelos de redes neuronales en la tarea de discernir entre intervalos de ECG de alta y de baja calidad. Al no existir bases de datos suficientemente extensas, anotadas, y con acceso público, se comparan tres redes neuronales preentrenadas, las cuales son AlexNet, GoogLeNet y VGGNet. Para ello, se han utilizado 2.000 segmentos de ECG de 5 segundos de duración, y clasificados en dos grupos de 1.000 segmentos cada uno. Se ha obtenido una validación robusta de los algoritmos realizando cinco procesos de entrenamiento y dividiendo el conjunto de datos de manera aleatoria, con un 80 % para entrenamiento y un 20 % para test. Los resultados muestran una exactitud generalizada en todos los modelos en torno a un 90 %. Sin embargo, en cuanto al tiempo de cálculo y memoria utilizada se han observado importantes diferencias. AlexNet ha mostrado ser el método más rápido para clasificar los segmentos de ECG, mientras que GoogLeNet ha requerido menos memoria. Por lo tanto, ambos algoritmos han demostrado un buen equilibrio entre la identificación de intervalos de ECG y carga computacional.

1. Introducción

La reciente evolución de las redes neuronales convolucionales ha permitido su uso en una amplia variedad de aplicaciones en campos tan diversos como la salud, el deporte, la robótica, etc. [1]. En el ámbito de la medicina se han utilizado en gran medida para la interpretación y el análisis de las señales electrocardiográficas (ECG) [2]. De hecho, algunas de las aplicaciones novedosas en las que ya se

están utilizando son la clasificación de arritmias, la detección automática de fibrilación auricular y la identificación de personas a través de biometría, entre otras [2]. Sin embargo, una aplicación muy interesante en el cual apenas se han utilizado hasta el momento es en la evaluación de la calidad con la que se registra el ECG.

Lamentablemente, hasta el momento la presencia de niveles de ruido elevados en el ECG dificulta su interpretación, limitando así el posterior diagnóstico clínico [3]. Además, la presencia de artefactos de movimiento y otras interferencias provocan muchas falsas alarmas en sistemas de monitorización de ECG en tiempo real [4]. La presencia de ruido y artefactos también es especialmente problemática para los dispositivos de ECG vestibles o portátiles, los cuales tienen la difícil tarea de monitorizar y almacenar señales de ECG durante varias semanas e incluso meses, en los que el paciente sigue realizando sus tareas cotidianas [5]. Así pues, estos dispositivos tendrán que trabajar en entornos muy ruidosos y cambiantes, de forma que las señales captadas estarán ampliamente corrompidas por artefactos de movimiento, interferencias de la red eléctrica, desplazamiento de la línea de base, y ruido de alta frecuencia [5]. Por lo tanto, en estos dispositivos la identificación automática de los tramos de ECG de alta calidad no contaminados por ruido es muy importante, con el fin de extraer de manera fiable la información clínica [3].

Para acometer este objetivo se han presentado diversos algoritmos durante los últimos años [3]. La mayoría de ellos están basados en la detección de puntos fiduciales y eventos morfológicos de la señal de ECG, y en el posterior análisis de parámetros como la media entre intervalos RR, la consistencia en el tiempo de las formas de onda PQRST, la coherencia de los complejos QRS, etc. [3]. Sin embargo, en segmentos de ECG muy ruidosos, no siempre es posible una detección exacta de los puntos fiduciales, así como tampoco lo es en señales limpias que presentan formas de onda muy cambiantes u ondas T muy puntiagudas y de gran amplitud [6]. Estos aspectos son comunes en registros de ECG de larga duración, por lo que en este contexto

los métodos presentados hasta la fecha reducen mucho su rendimiento [3]. Por lo tanto, existe una fuerte demanda de algoritmos novedosos que no estén basados en características morfológicas de la señal y que puedan trabajar satisfactoriamente con los registros de larga duración adquiridos por dispositivos vestibles [3].

Como respuesta a esta necesidad, la alternativa más viable en la actualidad son las redes neuronales convolucionales. Éstas son capaces de extraer las características más importantes del ECG sin la necesidad de delinear puntos fiduciales, y sin requerir además ningún tipo de intervención externa [7]. Sin embargo, hasta la fecha solamente existen unos pocos trabajos que han utilizado este tipo de algoritmos para evaluar la calidad del registro de ECG [8]. Desafortunadamente, estos han entrenado modelos desde cero con un número muy reducido de muestras y, por lo tanto, podrían estar sobreentrenados. Además, teniendo en cuenta que no hay muchas bases de datos de ECG públicas con anotaciones que sean suficientemente largas para entrenar una red desde cero [9], la utilización de redes preentrenadas es una opción muy interesante. En este caso, se emplea el concepto de *aprendizaje de transferencia*, para aprovechar el conocimiento adquirido por una red y en una determinada tarea, ahorrando tiempo de aprendizaje al solo requerirse un pequeño entrenamiento específico. A este respecto, en un reciente estudio piloto ya se ha evaluado la calidad del ECG empleando una red preentrenada, la cual ha obtenido un mayor rendimiento que otros modelos entrenados desde cero [10]. Por lo tanto, el presente trabajo tiene como objetivo comparar el rendimiento de varias redes neuronales convolucionales preentrenadas para discernir entre segmentos de ECG de alta y baja calidad.

2. Base de datos

La base de datos empleada en el presente trabajo fue del conjunto de registros de ECG propuesto para entrenamiento en el PhysioNet/CINC Challenge 2017 [11]. Se trata de un conjunto público de 8.528 registros de ECG de una sola derivación, y una duración de entre 9 y 60 segundos. Se adquirieron utilizando un dispositivo portátil de la marca AliveCor TM conectado a un teléfono móvil, con una frecuencia de muestreo de 300 Hz y 16 bits de resolución sobre un rango dinámico de 5 mV.

Para entrenar y validar cada uno de los modelos preentrenados que se analizan, las señales de ECG se dividieron en segmentos de 5 segundos, y se agruparon en dos categorías. En primer lugar, los episodios de fibrilación auricular, ritmo sinusal y otros ritmos se consideraron señales de alta calidad, mientras que los segmentos de ruido se consideraron de baja calidad. Debido al gran desequilibrio entre ambos grupos (47.439 segmentos de alta calidad frente a 1.168 intervalos de baja calidad, ver Tabla 1), se creó un conjunto de datos específico tomando 1.000 fragmentos de ECG de cada clase. En el grupo de alta calidad se mantuvo la representatividad de todos los ritmos cardíacos, escogiendo 340 fragmentos de ritmo sinusal, 330 fragmentos de fibrilación auricular y 330 fragmentos de otros ritmos.

Ritmo	Registros	Intervalos de 5 s.
Ritmo sinusal	5.154	28.413
Fibrilación auricular	771	4.329
Otros ritmos	2.557	14.697
Ruido	46	1.168

Tabla 1. Número de registros y de intervalos de 5 segundos para cada ritmo pertenecientes a la base de datos de PhysioNet/CinC Challenge 2017 [11].

3. Métodos

3.1. Transformación de los segmentos de ECG

La mayoría de redes preentrenadas se caracterizan por recibir una señal en dos dimensiones como entrada. Por lo tanto, cada segmento de ECG de 5 segundos se convirtió a una representación tiempo-frecuencia, empleando para ello la transformada wavelet continua (TWC). Esta transformación es especialmente adecuada para señales no estacionarias, como lo es el ECG [12]. La representación gráfica de los coeficientes wavelet obtenidos mediante esta herramienta es conocido como escalograma y se empleó como entrada para las redes analizadas. En este caso, la TWC se realizó con una wavelet madre llamada *Morlet* y se emplearon 48 voces por octava para determinar el número de escalas.

3.2. Arquitecturas preentrenadas

En el presente trabajo se evaluaron tres arquitecturas preentrenadas muy conocidas, tales como son AlexNet, VGGNet y GoogLeNet. AlexNet se hizo muy popular por ganar una competición de identificación automática de imágenes en el año 2012 [13], y es probablemente la arquitectura más conocida dentro de las redes neuronales convolucionales. De hecho, este algoritmo se ha utilizado en una gran variedad de aplicaciones [1]. Se compone de ocho capas con capacidad de aprendizaje, de las cuales cinco son convolucionales y tres son totalmente conectadas. Además, también incluye dos capas de agrupación o *pooling* para reducir las dimensiones espaciales del mapa de características, sin perder información relevante. Como función de activación utiliza una función lineal en todas las capas convolucionales y en las totalmente conectadas. También incluye dos funciones de *dropout* para reducir el sobreentrenamiento después de las dos primeras capas totalmente conectadas.

Por otro lado, VGGNet se ha hecho muy popular debido a su sencillez y al empleo de convoluciones de pequeño tamaño [14]. La principal aportación de esta red es el uso de muchas capas convolucionales con filtros de pequeño tamaño para poder lograr redes muy profundas, las cuales suelen obtener un buen rendimiento. Existen diferentes variantes dentro de VGGNet con diferentes niveles de profundidad, pero la alternativa escogida en este trabajo es VGG16 por ser una de las más comunes. Este modelo emplea filtros de convolución de tamaño 3×3 en una arquitectura que se compone de trece capas convolucionales, obteniendo una eficiencia muy alta [14]. Todas las capas de convolución van seguidas por funciones de activación lineales. Además, en el centro de la red se emplean cinco capas de agrupación para extraer características. Por

último, el modelo se completa con tres capas totalmente conectadas, las cuales van seguidas de funciones de activación y *dropout*.

Por último, GoogLeNet [15] está basado en un diseño totalmente diferente a los dos modelos anteriores. Mientras que AlexNet y VGG16 se basan en una estructura de capas colocadas de manera secuencial, GoogLeNet está diseñado mediante una compleja estructura con diferentes ramas en paralelo. Concretamente, esta red utiliza el *Módulo Inception* como base de su modelo. Este módulo se compone de una serie de capas de convolución y de agrupación, las cuales trabajan en paralelo combinando las características extraídas. Conectando nueve *Módulos Inception* en cascada junto con otras capas y funciones necesarias, GoogLeNet consigue una arquitectura con una profundidad de veintidós capas y más de cien bloques computacionales.

3.3. Evaluación del rendimiento

Para el entrenamiento y validación de cada uno de los modelos preentrenados, se utilizó un enfoque basado en repeticiones. Así pues, en 5 iteraciones consecutivas la base de datos se dividió, empleándose el 80 % para entrenamiento y el 20 % para test. El proceso de re-entrenamiento de la red se realizó con un algoritmo de descenso de gradiente estocástico con un momento de 0,9 y una tasa de aprendizaje de 0,0001, fijados de manera experimental. En todos los casos se consideraron 10 épocas con un tamaño de lote de 10 muestras. Además, en todos los modelos se adaptó la capa de salida de la red para distinguir solamente dos clases: alta y baja calidad.

El rendimiento de las diferentes redes discerniendo entre intervalos de ECG de alta y baja calidad pertenecientes al conjunto de test se evaluó en términos de sensibilidad (Se), especificidad (Es) y exactitud (Ex). Además, para las cinco iteraciones se obtuvo la media, desviación estándar (std), y los valores máximo y mínimo de estos índices. La sensibilidad se definió como la tasa de segmentos de ECG de alta calidad correctamente clasificados, la especificidad como el porcentaje de intervalos de baja calidad correctamente identificados, y por último, la exactitud, como el número total de fragmentos de ECG correctamente detectados. Además, para comparar el comportamiento de los algoritmos, también se analizó el tiempo, el uso de la CPU y la memoria requerida durante la fase de validación. Los experimentos se desarrollaron en una estación de trabajo all-in-one HP ProOne 600 G2 con un procesador Intel i7 a 3,41 GHz, utilizando Matlab R2019a. Cabe mencionar que también se obtuvieron los valores medio, estándar, máximo y mínimo de estas medidas para las cinco iteraciones conducidas.

4. Resultados

En la Tabla 2 se observa una exactitud media similar en todos los modelos, oscilando ligeramente entre el 89,65 % para VGG16 y el 90,75 % para GoogLeNet. Además, los resultados medios de sensibilidad y especificidad se encuentran bastante balanceados para los tres modelos, existiendo diferencias de tan solo un 4–8 % entre ambas métri-

Modelo	Valor	Se. (%)	Es. (%)	Ex. (%)
AlexNet	Media	88,90	92,50	90,70
	Std	3,27	4,00	0,96
	Máximo	93,50	95,50	91,75
	Mínimo	84,50	85,50	89,50
VGG16	Media	85,60	93,70	89,65
	Std	8,98	3,87	3,42
	Máximo	94,00	97,00	91,50
	Mínimo	71,00	89,00	84,00
GoogLeNet	Media	88,8	92,7	90,75
	Std	2,77	1,72	0,75
	Máximo	92,00	94,00	91,50
	Mínimo	85,00	90,00	89,50

Tabla 2. Media, std, valores máximo y mínimo obtenidos de las métricas de Sensibilidad, Especificidad y Exactitud para cada uno de los tres modelos en la fase de test.

Modelo	Valor	Tiempo (s)	CPU (%)	Mem. (MB)
AlexNet	Media	12,76	35,42	2.814,8
	Std	0,21	0,52	75,08
	Máximo	12,95	36,20	2.909
	Mínimo	12,43	34,80	2,724
VGG16	Media	78,84	39,27	4.854,2
	Std	0,57	0,29	58,87
	Máximo	79,75	39,62	4.935
	Mínimo	78,05	38,86	4.784
GoogLeNet	Media	25,33	37,03	1.589,8
	Std	0,45	0,31	8,87
	Máximo	25,86	37,40	1.596
	Mínimo	24,78	36,72	1.575

Tabla 3. Media, std, valores máximo y mínimo obtenidos para el tiempo, consumo de CPU y utilización de memoria para cada uno de los tres modelos durante la fase de test.

cas. Sin embargo, en los tres casos se observa una mayor capacidad para identificar segmentos de ECG de baja calidad, siendo los valores de especificidad y sensibilidad alrededor del 92 % y el 88 %, respectivamente. Por último, no se aprecia una gran dispersión en los tres índices para las cinco iteraciones realizadas, ya que los valores de std son inferiores al 4 % y las diferencias entre los valores máximos y mínimos inferiores al 10 % para la mayoría de los casos.

En cuanto al coste computacional de los algoritmos, la Tabla 3 muestra los resultados obtenidos. Respecto al tiempo y memoria consumida, se observan diferencias considerables para los tres modelos. Así pues, el tiempo medio de cálculo requerido varía entre 12,76 y 78,84 segundos por iteración, por lo que AlexNet es 6,5 veces más rápida que VGG16. La memoria empleada por cada uno de los modelos también es muy variable, necesitando GoogLeNet 1.589,8 MB y VGG16 4.854,2 MB. Finalmente, en cuanto al consumo de CPU, se observa una menor dispersión entre valores medios, con una diferencia inferior al 4 % entre las tres redes. Al igual que en los resultados de clasificación, en este caso no se observa gran dispersión entre los valores obtenidos para las cinco iteraciones conducidas.

5. Discusión y conclusiones

En el presente trabajo se han comparado tres conocidos modelos de redes neuronales convolucionales preentrenadas para la tarea específica de evaluar la calidad del ECG. Aunque recientemente se han realizado algunas comparaciones sobre cómo funcionan diferentes modelos de redes preentrenadas en aplicaciones relacionadas con el ECG, ta-

les como la detección de morfologías atípicas [16], la clasificación de arritmias [17], o la identificación de personas [18], sus resultados no pueden compararse directamente con la tarea de discernir entre segmentos de ECG de alta y de baja calidad. Además, en estos trabajos los algoritmos se compararon principalmente en función de los resultados de clasificación obtenidos en un conjunto de test, sin tener en cuenta otra información sobre su eficacia, como el porcentaje de uso de la CPU, el consumo de memoria y el tiempo de cálculo. Sin embargo, estas métricas de rendimiento pueden servir de referencia y mostrar los recursos necesarios para embeber cada algoritmo en sistemas de monitorización de ECG portátiles o vestibles.

En cuanto a los resultados de rendimiento obtenidos, se observa que VGG16 requirió más tiempo que AlexNet y GoogLeNet para clasificar todos los segmentos de ECG incluidos en el conjunto de test. A pesar de tener una estructura simple, el diseño secuencial hace que este modelo sea menos productivo que otras redes basadas en una arquitectura paralela [15]. Además, para un uso similar de CPU que el resto de redes, VGG16 requirió mucha más memoria. Esto podría ser debido al enorme número de hiperparámetros que emplea. Así pues, mientras que VGG16 tiene más de 138 millones de hiperparámetros, AlexNet utiliza 65 millones y GoogLeNet sólo 6 millones [14]. Por lo tanto, VGG16 no parece ser una opción adecuada para ser incorporado en dispositivos portátiles para la monitorización de ECG a largo plazo, ya que requiere una fuerte carga computacional y, por lo tanto, un consumo elevado de batería. En cuanto a los otros dos modelos, AlexNet necesitó menos tiempo para clasificar todos los intervalos de ECG, mientras que GoogLeNet utilizó menos memoria para ejecutarse. Por lo tanto, la implementación de ambos algoritmos para la monitorización de ECG en tiempo real podría implicar requerimientos computacionales similares.

Por otro lado, a pesar de las diferencias entre las tres arquitecturas, han obtenido valores muy similares de exactitud, sensibilidad y especificidad. De hecho, solamente se han registrado diferencias entre 1 y 3 % en estas tres métricas a la hora de analizar los segmentos de alta y baja calidad del conjunto de test. Este resultado sugiere que modelos no muy complejos, tales como AlexNet y GoogLeNet, presentan suficiente capacidad de aprendizaje para obtener un buen rendimiento en la evaluación la calidad del ECG.

En vista de todas estas observaciones, AlexNet y GoogLeNet son los modelos que muestran un mejor compromiso entre exactitud, discerniendo entre segmentos de ECG de alta y baja calidad, y carga computacional. Por lo tanto, estos modelos parecen más adecuados que VGG16 para evaluar la calidad del ECG, así como para ser embebidos en sistemas de monitorización de ECG vestibles o portátiles.

Agradecimientos

Este trabajo ha sido cofinanciado por los proyectos de investigación competitiva DPI2017-83952-C3 de MINECO-AEI-FEDER-UE, SBPLY/17/180501/000411 de la JCCLM, AICO/2019/036 de la GVA y FEDER 2018/11744.

Referencias

- [1] Patil T, Pandey S, Visrani K. A review on basic deep learning technologies and applications. In Kotecha K, Piuri V, Shah H, Patel R (eds.), *Data Science and Intelligent Applications*, volume 52. Springer, 2021; 565–573.
- [2] Hong S, Zhou Y, Shang J, Xiao C, Sun J. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. *Comput Biol Med Jul 2020*; 122:103801.
- [3] Satija U, Ramkumar B, Manikandan MS. A review of signal processing techniques for electrocardiogram signal quality assessment. *IEEE Rev Biomed Eng 2018*;11:36–52.
- [4] Bashar SK, Ding E, Walkey AJ, McManus DD, Chon KH. Noise detection in electrocardiogram signals for intensive care unit patients. *IEEE Access 2019*;7:88357–88368. ISSN 2169-3536.
- [5] Vavrinsky E, Subjak J, Donoval M, Wagner A, Zavodnik T, Svobodova H. Application of modern multi-sensor holter in diagnosis and treatment. *Sensors Basel May 2020*;20(9).
- [6] Martínez A, Alcaraz R, Rieta JJ. Application of the phasor transform for automatic delineation of single-lead ECG fiducial points. *Physiol Meas Nov 2010*;31(11):1467–85.
- [7] Xu X, Wei S, Ma C, Luo K, Zhang L, Liu C. Atrial fibrillation beat identification using the combination of modified frequency slice wavelet transform and convolutional neural networks. *J Healthc Eng 2018*;2018:2102918.
- [8] Zhang Q, Fu L, Gu L. A cascaded convolutional neural network for assessing signal quality of dynamic ECG. *Computational and Mathematical Methods in Medicine 2019*;2019.
- [9] Mincholé A, Rodríguez B. Artificial intelligence for the electrocardiogram. *Nat Med 01 2019*;25(1):22–23.
- [10] Huerta A, Martínez-Rodrigo A, González VB, Quesada A, Rieta J, Alcaraz R. Quality assessment of very long-term ECG recordings using a convolutional neural network. In *2019 E-Health and Bioengineering Conference (EHB)*. 2019; 1–4.
- [11] Clifford GD, Liu C, Moody B, Li-wei HL, Silva I, Li Q, Johnson A, Mark RG. AF classification from a short single lead ECG recording: the PhysioNet/Computing in Cardiology Challenge 2017. In *2017 Computing in Cardiology (CinC)*. 2017; 1–4.
- [12] Shoeb A, Clird G. Chapter 16—wavelets; multiscale activity in physiological signals. *Biomedical Signal and Image Processing 2005*;–.
- [13] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*. 2012; 1097–1105.
- [14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv14091556 2014*;1–14.
- [15] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition*. 2015; 1–9.
- [16] Diker A, Cömert Z, Avcı E, Toğaçar M, Ergen B. A novel application based on spectrogram and convolutional neural network for ECG classification. In *1st IEE International Informatics and Software Engineering Conf*. 2019; 1–6.
- [17] Alquran H, Alqudah A, Abu-Qasmieh I, Al-Badarneh A, Almashaqbeh S. ECG classification using higher order spectral estimation and deep learning techniques. *Neural Network World 2019*;29(4):207–219.
- [18] Byeon YH, Pan SB, Kwak KC. Intelligent deep models based on scalograms of electrocardiogram signals for biometrics. *Sensors Basel Feb 2019*;19(4).