

Detección de fibrilación ventricular mediante técnicas de aprendizaje profundo

A. Picon¹, U. Irusta², A. Alvarez-Gila¹, E. Aramendi², E. Garrote¹, U. Ayala³, F. Alonso-Atienza⁴, C. Figuera⁴

¹ Smart Systems Department, TecNALIA, Derio, Spain, {artzai.picon,aitor.alvarez,estibaliz.garrote}@tecnalia.com

² Departamento de Ingeniería de Comunicaciones, Universidad del País Vasco (UPV-EHU), Bilbao, Spain, {unai.irusta,elisabete.aramendi}@ehu.eus

³ Tratamiento de la Señal y Comunicaciones, Mondragon Unibertsitatea, Arrasate, Spain, uayala@mondragon.edu

⁴ Teoría De La Señal y Comunicaciones y sistemas Telemáticos y Computación, Universidad Rey Juan Carlos, Madrid, Spain, {felipe.alonso,carlos.figuera}@urjc.es

Resumen

La detección de arritmias ventriculares, en particular la fibrilación ventricular (FV), es parte fundamental de los algoritmos de clasificación de arritmias de los desfibriladores. Dichos algoritmos deciden si administrar la descarga de desfibrilación, para lo que clasifican los ritmos en desfibrilables (Sh) o no desfibrilables (NSh). Este trabajo propone un nuevo abordaje para la clasificación Sh/NSh de ritmos basado en un sistema de aprendizaje profundo.

Para el trabajo se emplearon tres bases de datos públicas de la plataforma Physionet (CUDB, VFDB y AHADB), y se extrajeron segmentos de 4 y 8 segundos. Se anotaron los segmentos como Sh y NSH en base a las anotaciones de las bases de datos, que fueron auditadas por expertos. Los datos se dividieron por paciente en 80% para desarrollar los algoritmos y 20% para evaluación. El sistema de aprendizaje profundo emplea dos etapas convolucionales seguidas de, una red long-short-term-memory y una etapa final de clasificación basada en red neuronal. A modo de referencia se optimizó un clasificador SVM basado en las características de detección de arritmias ventriculares más eficientes publicadas en la literatura. Se calculó la sensibilidad (Se), ritmos desfibrilables, especificidad (Sp), ritmos no desfibrilables, y la precisión (Acc).

El método de aprendizaje profundo proporcionó Se, Sp y Acc de 98.5%, 99.4% y 99.2% para segmentos de 4 segundos y 99.7%, 98.9%, 99.1% para segmentos de 8 segundos. El algoritmo permite detectar FV de forma fiable con segmentos de 4 segundos, corrigiendo un 30% de los errores del método basado en SVM.

1. Introducción

La incidencia anual de la parada cardiorrespiratoria extrahospitalaria (PCREH) en España se estima en 18.6 casos de por cada 100 000 habitantes, con un índice de supervivencia inferior al 12% [1]. Una de las principales causas de la PCREH son las arritmias ventriculares malignas, y en particular la fibrilación ventricular (FV). Durante la FV la ausencia de perfusión provoca la muerte en pocos minutos por lo que la pronta intervención es esencial. La única terapia efectiva para revertir la FV es administrar una descarga eléctrica que desfibrile el corazón y lo devuelva a un ritmo normal con perfusión. El Desfibrilador Externo Automático (DEA) es un equipo sencillo que permite que personas mínimamente entrenadas puedan desfibrilar y tratar a pacientes en

PCREH. Los DEAs incorporan un algoritmo automático que mediante el análisis del ECG detecta la FV y suministra la descarga eléctrica.

Los algoritmos DEA analizan segmentos cortos del ECG, del orden de 3-10 segundos, calculan parámetros para la detección de arritmias ventriculares y los combinan en algoritmos de decisión. El algoritmo diagnostica por lo tanto el segmento como desfibrilable (Sh, shockable) o no desfibrilable (NSH, nonshockable). La American Heart Association (AHA) establece que dichos algoritmos deben detectar la FV con sensibilidades superiores al 90%, y deben identificar los ritmos NSH con especificidades superiores al 95% (99% en caso de ritmos sinusales normales, RSN) [2]. En las últimas tres décadas se han estudiado numerosos parámetros de discriminación Sh/NSH [3], y se han comparado sus desempeños empleando bases de datos ECG de acceso público [3,4]. Recientemente se han propuesto técnicas de aprendizaje automático que emplean, seleccionan y usan varios de dichos parámetros [3,5] para la detección de ritmos desfibrilables.

Desarrollos recientes en el ámbito del aprendizaje automático permiten que el propio algoritmo aprenda y seleccione las características de discriminación, explotando así toda la información disponible en las señales. Este es el principio que subyace a los sistemas basados en redes neuronales profundas, y que han sido adaptadas ya para trabajar con señales unidimensionales como el ECG. De hecho existen sistemas de aprendizaje profundo basados en el ECG para la clasificación de latidos [6] y de algunas arritmias [7]. Sobre estos antecedentes, el objetivo de este trabajo es desarrollar un algoritmo de aprendizaje profundo para la detección de FV para su uso en DEAs.

2. Materiales y métodos

2.1. Bases de registros ECG

Para el presente trabajo se emplearon tres bases de datos de acceso público: la bases de datos de arritmias ventriculares del MIT-BIH (VFDB) y de la Creighton University (CUDB), disponibles ambas en la plataforma Physionet [8], y la serie 1 de la base de datos de la AHA

(AHADB). En total se analizaron 77 registros, 22 registros de 30 min de la VFDB (canal 1), 35 registros de 8 min de la CUDB, y 10 registros (8201-8210) de 35 min de la AHADB (canal 1). La frecuencia de muestreo del ECG en todos los casos es de 250Hz.

El ECG se preprocesó siguiendo el esquema propuesto en [4]: (1) sustracción de la media, (2) filtro de media móvil de orden cinco y (3) filtrado pasobanda típico en DEAs (1-30 Hz). Las anotaciones de ritmo de las bases originales se revisaron para adaptarlas al formato AHA [2], es decir se suprimieron intervalos de ruido o artefactados, intervalos con ausencia de ritmo (asistolia), o ritmos intermedios para los que los beneficios de la desfibrilación no son claros. Se anotaron los ritmos en dos clases, Sh y NSh. Los ritmos Sh incluyen las arritmias ventriculares malignas como la FV y taquicardia ventricular rápida, y los ritmos NSh el resto de ritmos/arritmias, como RSN, fibrilación atrial, taquicardias supraventriculares o contracciones ventriculares prematuras. Los registros se segmentaron en intervalos de 4 y 8 seg, por ser éstas las duraciones típicas usadas en los trabajos de referencia [4], y se excluyeron los segmentos que incluyeran transiciones de ritmo.

Finalmente las bases de segmentos se dividieron (por pacientes) en dos subconjuntos, uno para el entrenamiento de los algoritmos con aproximadamente el 80% de los pacientes y otro para evaluar el desempeño con el resto. La tabla 1 muestra un resumen de los segmentos empleados en el trabajo, subrayar que el acceso a los mismos es libre a través de [3].

| | Desarrollo | | | Evaluación | | |
|--------------|------------|-------------------|------------------|------------|------------------|------------------|
| | <i>n</i> | Sh/NSh (4seg) | Sh/NSh (8seg) | <i>n</i> | Sh/NSh (4seg) | Sh/NSh (8seg) |
| VFDB | 18 | 1257/6462 | 596/3146 | 4 | 329/1299 | 150/634 |
| CUDB | 28 | 598/2374 | 272/1149 | 7 | 118/612 | 51/297 |
| AHADB | 8 | 989/3050 | 486/1520 | 2 | 287/698 | 141/340 |
| Total | 54 | 2844/11886 | 1354/5815 | 13 | 734/2609 | 342/1271 |

Tabla 1. Bases de datos (BBDD) de desarrollo y evaluación para segmentos de 4 y 8 segundos, por número de registros (*n*), y tipo de ritmo shockable (Sh) y no shockable (NSh).

2.2. Algoritmos de clasificación

El sistema de aprendizaje profundo empleado en este trabajo se compone tres bloques, tal y como se muestra en la Figura 1. En el bloque inicial una red neuronal convolucional (CNN) extrae los descriptores de alto nivel del ECG. Estos se integran según la dimensión temporal mediante una red LSTM (Long Short-Term Memory) lo que permite una representación de la variabilidad temporal. Finalmente, existe una etapa de clasificación para producir una salida binaria a partir de la predicción (en el rango 0-1) de la red LSTM.

Para la red CNN se emplearon dos bloques convolucionales cada uno con: una capa convolucional de tamaño 3, una función de activación lineal rectificada

(RELU) y una etapa de max-pooling que permite la compresión temporal de la señal de forma progresiva. Cada capa convolucional se configuró con 32 filtros, y la longitud de la etapa de maxpooling se fijó en 3. Para la red LSTM el número de proyecciones de la red fue de 20. Finalmente, la etapa de clasificación se optimizó mediante el método del descenso del gradiente incremental. Detalles de la red se ilustran en la Figura 1 incluyendo las dimensiones de cada capa.

A modo de referencia se compararon los resultados con los obtenidos siguiendo un esquema de aprendizaje automático clásico, basado en una maquina de vectores soporte (SVM) con función de base radial gaussiana. Los parámetros de la SVM se determinaron mediante optimización bayesiana minimizando el error total ponderado (ver sección 2.3). Para el clasificador SVM se emplearon 6 características típicas que han sido ya identificadas como las de mejor desempeño en trabajos anteriores [3]. Las características seleccionadas fueron: SampEn [9], vFleak [4], x1 [10], x2 [10], bCP [11] y Li [12]. La descripción matemática de los parámetros se encuentra en las referencias citadas, y el código para su cálculo está disponible en [3].

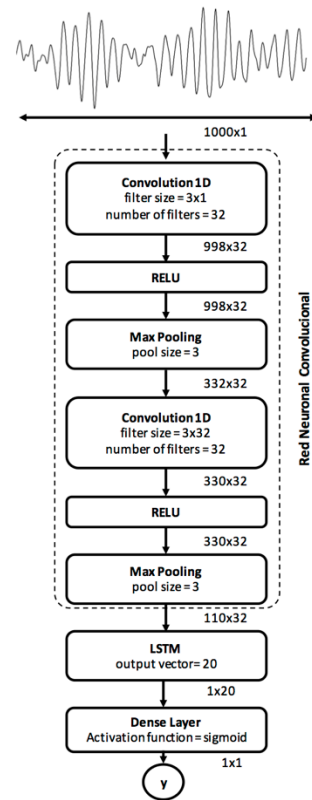


Figura 1. Arquitectura básica del discriminador Sh/NSh basado en técnicas de aprendizaje profundo (segmentos de 4-seg). Se indican en cada bloque las dimensiones de los datos.

2.3. Métricas de desempeño y optimización

Para la definición de las métricas de desempeño de los algoritmos de detección de ritmos desfibrilables se define una matriz de contingencia 2x2 con los ritmos Sh como clase positiva y los NSh como negativa. Se comparan las predicciones de los algoritmos con las anotaciones de ritmo y se obtienen así la sensibilidad (Se), especificidad

(Sp), los valores predictivos positivo y negativo (VPP y VPN) y la precisión (Acc).

Tal y como se observa en la Tabla 1, en la base de datos existe un desequilibrio importante de clases ya que la prevalencia de los ritmos NSh es cuatro veces mayor. Para obtener resultados equilibrados de Se y Sp, se asignaron pesos a las clases positiva y negativa de forma que ambas tuvieran la misma importancia en la optimización de los clasificadores.

3. Resultados

En primer lugar se caracterizó la convergencia del proceso de optimización del clasificador basado en aprendizaje profundo. Los resultados se muestran en la figura 2 para el clasificador basado en segmentos de 4-seg. Se observa que a partir de 50 épocas se obtienen valores aceptables de la función de coste, la entropía cruzada, y que el coste es muy próximo a 0 para más de 250-300 épocas, estabilizándose el coste del bloque de test a un valor de 0.05. Se escogió un valor de 500 épocas para finalizar el proceso de entrenamiento de las redes profundas, tanto con segmentos de 4-seg como de 8-seg.

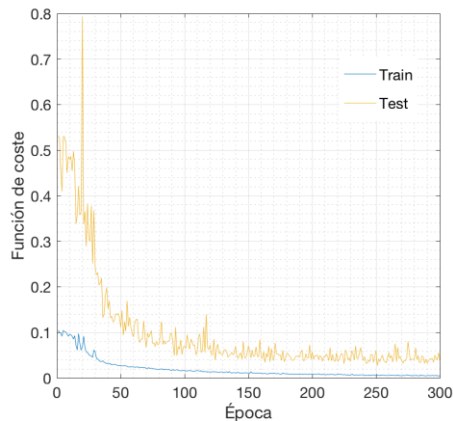


Figura 2. Convergencia de la red para segmentos de 4-seg.

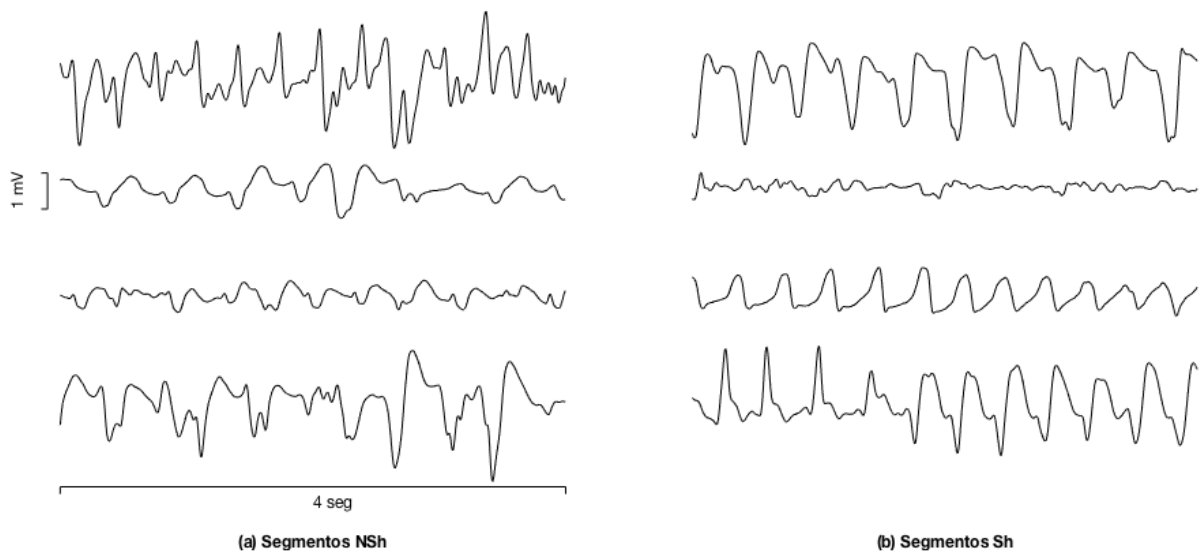


Figura 3. Ejemplos de segmentos de 4seg con errores de clasificación para el clasificador basado en aprendizaje profundo, para ritmos no shockables (izquierda) y ritmos shockable (derecha). Se aprecia que los segmentos incorrectamente clasificados corresponden a ritmos sin complejos QRS bien definidos (no shockables) y a ritmos ventriculares de baja amplitud y/o frecuencia (shockables).

Los resultados de clasificación sobre la base de datos de test se muestran en la Tabla 2. Esta tabla contiene también el análisis de curva ROC de los clasificadores obtenido a partir de la variable continua devuelta por la función de activación de la capa densa (aprendizaje profundo) o de la función de discriminación (SVM). El análisis se hace en base al área bajo la curva (AUC). En los resultados se observa que tanto la Se como la Sp son muy elevadas en ambos casos, pero que se obtienen mejores resultados empleando el sistema de aprendizaje profundo y segmentos de 4-seg para la clasificación.

| Método | Métricas de desempeño | | | | | |
|--------|-----------------------|------|------|------|------|-------|
| | Se | Sp | VPP | VPN | Acc | AUC |
| P-4 | 99.6 | 99.1 | 96.7 | 99.9 | 99.2 | 99.35 |
| SVM-4 | 98.0 | 99.2 | 97.0 | 99.4 | 98.9 | 99.03 |
| P-8 | 100 | 99.2 | 97.2 | 100 | 99.4 | 99.96 |
| SVM-8 | 98.2 | 99.4 | 97.7 | 99.5 | 99.1 | 99.81 |

Tabla 2. Métricas de desempeño de los clasificadores sobre las bases de evaluación. Las abreviaturas de los métodos se leen como P-4 (aprendizaje profundo, con segmentos de 4seg), o SVM-8 (SVM, con segmentos de 8 seg).

Finalmente, la figura 3 muestra ejemplos de errores en la clasificación empleando los algoritmos de aprendizaje profundo. Los ejemplos mostrados son segmentos de 4-segundos de la base de evaluación. Se observa que los segmentos no shockables errados se corresponden con ritmos que no presentan complejos QRS bien definidos, y están en la frontera entre ritmos ventriculares y no ventriculares. Los segmentos shockables errados corresponden a arritmias ventriculares de baja amplitud y/o frecuencia ventricular.

4. Discusión y conclusiones

Este trabajo presenta el primer abordaje, hasta donde los autores conocen, basado en aprendizaje profundo para la detección de arritmias ventriculares. Los métodos se han comparado con las soluciones más avanzadas propuestas hasta el momento en base a la extracción de características y métodos de aprendizaje automático [3]. El abordaje basado en aprendizaje profundo proporcionó mejores resultados que el basado en aprendizaje automático, especialmente para segmentos de 4-seg. Esto es relevante teniendo en cuenta que: (1) los métodos de aprendizaje profundo tratan la señal ECG directamente sin concepciones a priori sobre las características de las arritmias a discriminar, y (2) los resultados de partida proporcionaban ya valores de precisión cercanos o por encima del 99%. Así, por ejemplo, con segmentos de 4-seg se pasa de una precisión del 98.89% a 99.22% que se corresponde con eliminar un 29.7% de los errores cometidos por el sistema clásico basado en aprendizaje automático.

Nuestro trabajo incide en la línea de algunos resultados publicados en los últimos años, que muestran el potencial de los métodos de aprendizaje profundo en problemas de clasificación de señales biomédicas unidimensionales, como el ECG. Así existen ya propuestas de clasificadores de arritmias que mejoran el desempeño de clínicos especialistas [7], o clasificadores de latidos de alta precisión [6]. Se ha conseguido en este trabajo un primer algoritmo de aprendizaje profundo que puede implementarse en un desfibrilador para la detección precisa de la FV.

Este trabajo tiene dos limitaciones importantes. La primera es que, aunque el número de segmentos empleados es muy grande, los registros provienen de pocos pacientes. Esta limitación es común a los estudios basados en bases de datos públicas [3-5], ya que existen pocos casos con arritmias ventriculares a pesar de disponer de segmentos largos de señal (varios minutos de FV). La segunda es que los resultados no se han obtenido para pacientes en PCREH, ya que las bases públicas se obtuvieron de pacientes con un Holter, en situaciones intrahospitalarias. En los pacientes PCREH, los ritmos observados por el desfibrilador ocurren 5-10 minutos después de haberse producido la parada, a la llegada de los servicios de emergencias. Esto tiene implicaciones importantes en el diseño de los clasificadores [3], ya que las características de los ritmos observados pueden ser considerablemente distintas [3].

Agradecimientos

Este trabajo ha sido financiado por el Ministerio de Economía y Competitividad mediante el proyecto TEC2015-64678R junto con el Fondo Europeo de

Desarrollo Regional (FEDER), así como por la UPV-EHU mediante el proyecto EHU16/18.

Referencias

- [1] Rosell-Ortiz F et al. Out-of-hospital cardiac arrest (OHCA) attended by mobile emergency teams with a physician on board. Results of the Spanish OHCA Registry (OSHCAR). *Resuscitation*, vol 113, 2017, pp 90-5.
- [2] Kerber R et al. Automatic External Defibrillators for Public Access Defibrillation: Recommendations for Specifying and Reporting Arrhythmia Analysis Algorithm Performance, Incorporating New Waveforms and Enhancing Safety. *Circulation*, vol 95 (6), 1997, pp 1677-81.
- [3] Figuera C, Irusta U, Morgado E, Aramendi E, Ayala U, Wik L, Kramer-Johansen J, Eftestol T. Machine Learning Techniques for the Detection of Shockable Rhythms in Automated External Defibrillators. *PloS ONE*, 11 (7), 2016; e0159654.
- [4] Amman A, Tratnig R, Unterkofler K. Reliability of old and new ventricular fibrillation detection algorithms for automated external defibrillators. *Biomed Eng Online*, 4 (60), 2005.
- [5] Li Q, Clifford GD. Ventricular fibrillation and tachycardia classification using a machine learning approach. *IEEE Trans Biomed Eng*, 61 (6), 2014, pp 1607-13.
- [6] Kiranyaz S, Ince T, Gabbouj M. Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. *IEEE Trans Biomed Eng*, 63 (3), 2016, pp 664-75.
- [7] Rajpurkar P, Hannun AY, Haghpanahi M, Bourn C, Ng AY. Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks. *arXiv: 1707.01836*, 2017.
- [8] Goldberger et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation*, 101(23), 2000, pp e215-e220.
- [9] Lake DE, Richman JS, Griffin MP, Moorman JR. Sample entropy analysis of neonatal heart rate variability. *Am J Physiol Regul Integr Comp Physiol*, 283(3), 2002, pp R789-797.
- [10] Ayala U, Irusta U, Ruiz J, Eftestol T, Krame-Johansen J, Alonso-Atientza F, Alonso E, González-Otero D. A reliable method for rhythm analysis during cardiopulmonary resuscitation. *Biomed Res Int*, 2014:872420.
- [11] Irusta U, Ruiz J, Aramendi E, Ruiz de Gauna S, Ayala U, Alonso E. A high-temporal resolution algorithm to discriminate shockable from nonshockable rhythms in adults and children. *Resuscitation*, 83(9), 2012, pp 1090-7.
- [12] Li Y, Bisera J, Weil MH, Tang W. An Algorithm Used for Ventricular Fibrillation Detection Without Interrupting Chest Compression. *IEEE Trans Biomed Eng*, 59(1), 2012, pp 78-86.