

# Aprendizaje máquina para la predicción del éxito de la desfibrilación mediante el análisis de la forma de onda de la fibrilación ventricular

G. Zubia Garea<sup>1</sup>, X. Jaureguiabeitia Lara<sup>1</sup>, U. Irusta Zarandona<sup>1</sup>, E. Aramendi Ecenarro<sup>1</sup>

<sup>1</sup> BioRes, Ingeniería de comunicaciones, Euskal Herriko Unibertsitatea UPV/EHU, Bilbo, España, [gzubia004@ikasle.ehu.eus](mailto:gzubia004@ikasle.ehu.eus), {unai.irusta, xabier.jaureguiabeitia, elisabete.arabemendi}@ehu.eus

## Resumen

Predecir el éxito de la terapia de desfibrilación eléctrica para revertir la fibrilación ventricular (FV) permitiría mejorar la supervivencia de los pacientes en parada cardiorrespiratoria extra-hospitalaria (PCREH). Los métodos de predicción se basan en el análisis de la forma de onda del electrocardiograma (ECG) de la FV usando predictores individuales. El objetivo del trabajo es desarrollar modelos de aprendizaje máquina para mejorar la precisión de los métodos. Se dispuso de una base de 1630 casos de PCREH anotados por clínicos, que incluían 3836 intentos de desfibrilación (1081 exitosos). Se diseñaron modelos predictores del éxito de la desfibrilación caracterizando con 27 parámetros el intervalo ECG pre-shock de 2.05s de duración. Se evaluaron los modelos en términos de sensibilidad (SE, éxito), especificidad (SP, no éxito) y la media de ambas (precisión balanceada, PB). Los modelos se ajustaron y evaluaron mediante validación cruzada en particiones por paciente y estratificadas, repitiendo el proceso 100 veces para caracterizar estadísticamente las métricas. Se comparó un modelo de regresión logística monoparamétrico con varios modelos de aprendizaje máquina: regresión logística multiparamétrica, bosques de árboles y máquinas de vectores soporte (SVM) con kernel gaussiano. El mejor modelo monoparamétrico produjo una PB mediana (rango interdecilo, IDR) de 79.5 (79.4-79.6)%. El mejor modelo multiparamétrico (SVM con 6 parámetros) resultó en una PB mediana de 81.4 (81.2-81.6)%, y SE, SP de 83.4 (83.1-84.0)% y 79.3 (79.1-80.0)%, respectivamente. Los modelos de aprendizaje máquina permiten mejorar la predicción monoparamétrica hasta en 2 puntos de PB. En el futuro deberán desarrollarse nuevos predictores, para lo que la extracción exhaustiva de características mediante redes convolucionales (CNN) sería una buena alternativa.

## 1. Motivación

Las enfermedades cardiovasculares (ECV) son la principal causa de mortalidad en Europa, representando el 39% de los fallecimientos. En el estado español, las ECV son responsables del 29% de las defunciones, es decir, de más de 122.000 personas-año. La parada cardiorrespiratoria extra-hospitalaria (PCREH) es un evento cardiovascular que constituye un problema de salud pública, con una supervivencia inferior al 10% y causando más de 30.000 muertes al año en este país [1].

Entre las arritmias, la fibrilación ventricular (FV) es causante del 85% de PCREHs [2,3], y su único tratamiento efectivo es la desfibrilación por descarga eléctrica. Desafortunadamente no todas las descargas son exitosas, y las que no lo son pueden deteriorar el miocardio por lo que los

métodos para la predicción del éxito en la desfibrilación son de gran interés. Estos métodos se basan en el análisis del ECG de la FV. El ECG durante la FV es irregular, caracterizándose por presentar latidos ventriculares de amplitud y duración dispares, tal y como se aprecia en la Figura 1. Durante la FV no hay contracción ventricular efectiva y los ventrículos se contraen desordenadamente.

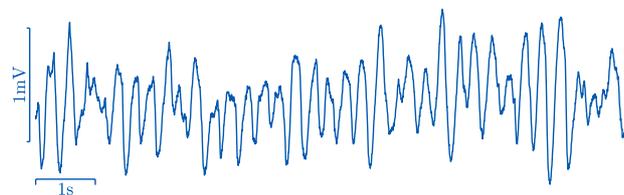


Figura 1. Ejemplo del ECG de una FV. La forma de la onda es irregular, con latidos ventriculares desorganizados

En las últimas décadas se ha avanzado hacia el cálculo de distintas características de la FV a partir de la señal de ECG que permitan predecir el éxito de la terapia eléctrica aplicada por un desfibrilador externo automático (DEA) al paciente en PCREH. Una de las más relevantes es AMSA (del inglés Amplitude Spectrum Area), que es una suma ponderada de las amplitudes en el dominio espectral [4,5]. Recientemente, Chicote et. al han desarrollado nuevos indicadores de características no lineales de las FV, como la entropía difusa y muestral [6,7], obteniendo resultados que mejoran los de los predictores clásicos como AMSA.

Aunque se han propuesto numerosos predictores [8–14], éstos se han empleado mayoritariamente de forma individual y excepcionalmente combinados en modelos de aprendizaje máquina (AM) [8]. Es por ello que el objetivo de este trabajo es analizar el poder predictivo de estos métodos y valorar la posibilidad de combinarlos de manera eficiente en modelos AM, empleando para ello una base de datos extensa con anotaciones clínicas contrastadas.

## 2. Materiales

La base de datos empleada está compuesta por los registros ECG de 1630 casos de PCREH de un estudio multicéntrico realizado entre 2008 y 2010 por la agencia italiana de la emergencia, Azienda Regionale Emergenza Urgenza. Los registros proceden de 3 marcas de DEA: Philips HeartStart FR2, Physio-Control Lifepak y ZOLL AED plus. El objetivo de este estudio era doble, por una parte, conocer el éxito en la desfibrilación con las técnicas actuales; y, por

otra, poder utilizar dichos registros para desarrollar nuevos métodos predictores del éxito en la desfibrilación.

Todos los registros contenían eventos de desfibrilación para los cuales además se dispuso de anotaciones clínicas sobre su éxito. Se definió como desfibrilación exitosa aquella en la que en el minuto posterior a la descarga se observaron latidos con complejos QRS visibles con frecuencia superior a 30 latidos por minuto [9]. Los registros se exportaron y convirtieron a un formato común en Matlab (MathWorks, Natick, MA), y se remuestrearon a 250 Hz. Los segmentos de análisis empleados fueron de 2.05s (512 muestras) tomados 0.5s antes del momento de la descarga. En total, se extrajeron 3836 segmentos de análisis, cuyo desglose puede observarse en la Tabla 1.

DEA	Shocks (pacientes)	Éxito	
		Sí	No
Philips	2066 (814)	591	1475
Physio	1581 (767)	405	1176
Zoll	189 (49)	85	104
<b>Total</b>	<b>3836 (1628)</b>	<b>1081</b>	<b>2755</b>

**Tabla 1.** Desglose de los registros por cada DEA en segmentos de análisis y pacientes, así como en descargas con y sin éxito.

### 3. Métodos

Para cada segmento se calcularon los 27 predictores principales del éxito de la desfibrilación descritos en la literatura [5–14], para los que existe una implementación de código abierto en formato Matlab accesible a través de <https://github.com/BChicote/shockOutcome>. Los métodos incluyen características de la FV calculadas en el dominio temporal, espectral, tiempo-frecuencia, de entropía, y basadas en el análisis de dinámica no lineal de la señal.

Para los análisis se tomaron las descargas exitosas como clase positiva ( $y = 1$ ) en modelos de decisión binaria. Se emplearon como métricas de desempeño la sensibilidad (SE, clase positiva), la especificidad (SP, clase negativa) y la precisión balanceada (PB), que es la media de SE y SP.

#### 3.1. Análisis monoparamétrico

Inicialmente se evaluó la capacidad discriminativa de cada característica mediante un análisis de Característica Operativa del Receptor (ROC), calculando el área bajo la curva (AUC en inglés). A su vez se ajustó un clasificador monoparamétrico basado en regresión logística (RL), como modelo base de clasificación.

#### 3.2. Análisis multiparamétrico

A continuación, se analizaron varios modelos de AM para combinar los predictores en modelos de decisión más elaborados. Para ello cada segmento  $i$  se representó mediante la tupla  $(\mathbf{x}_i, y_i)$ , en la que el vector  $\mathbf{x}_i$  agrupa las  $N$  características del segmento e  $y_i$  representa el diagnóstico clínico. Se analizaron los siguientes modelos:

- **Regresión logística multiparamétrica (RLM).** Se empleó como clasificador base por ser un modelo

lineal sencillo de decisión. En este caso se realizó un análisis de todas las combinaciones posibles de predictores, es decir se las combinaciones de  $N$  características de las 27 características. Dado que el número de combinaciones es muy elevado se realizó el análisis hasta  $N = 5$  ( $y N = 27 - 5 = 22$ ) y a partir de ahí se fijaron las 2, 3, ... características que aparecían en las combinaciones con mayor PB.

- **Bosques Aleatorios (RF en inglés).** Estos algoritmos se basan en la idea de que es posible ajustar muchas clasificadores débiles (individualmente imprecisos) que deciden independientemente. Cuando las decisiones de estos clasificadores se agregan (por mayoría de votos) el clasificador se vuelve robusto y muy preciso. Por lo tanto, un RF es un conjunto de  $B$  árboles de decisión  $\{T_1(\mathbf{x}), \dots, T_B(\mathbf{x})\}$  que produce predicciones  $B$  cuasi-incorreladas  $\{\hat{y}_1 = T_1(\mathbf{x}), \dots, \hat{y}_B = T_B(\mathbf{x})\}$  sobre el éxito en la desfibrilación de cada segmento de análisis del ECG [15], siendo la predicción del modelo  $\hat{y}$  la decisión mayoritaria de los  $\hat{y}_i$ . Para diseñar el clasificador RF se tuvieron en cuenta 3 hiperparámetros: número de predictores por rama, el número de árboles y el número mínimo de observaciones en la última división. Finalmente, se asignaron probabilidades previas uniformes para cada clase durante el entrenamiento para abordar el desequilibrio de clase.
- **Maquinas de vectores de soporte (SVM en inglés).** Se empleó una SVM con kernel gaussiano para modelar relaciones no lineales entre parámetros [16], mediante la siguiente transformación métrica:

$$K(\mathbf{x}, \mathbf{x}_i) = \exp(-\gamma \|\mathbf{x} - \mathbf{x}_i\|^2)$$

en la que  $\|\cdot\|^2$  es la norma  $L_2$ . Se ajustaron dos hiperparámetros, el ancho del kernel,  $\gamma$ , y el margen o la flexibilidad del límite de decisión,  $C$ . Se buscaron los valores de  $C$  y  $\gamma$  que maximizaban la PB, para ello se realizó una búsqueda logarítmica exhaustiva previa a cada entrenamiento, de  $40 \times 40$  en los rangos  $10^{-1} \leq C \leq 10^{1.5}$  y  $10^{-3} \leq \gamma \leq 10^1$  respectivamente.

#### 3.3. Evaluación de los análisis

Los 3 modelos propuestos fueron evaluados usando una estrategia de validación cruzada de 10 iteraciones (10-fold CV). La partición de los datos se realizó por paciente (patient-wise) y cuasiestratificada, de modo que los segmentos de análisis de cada iteración presentaran una distribución similar a la del conjunto de datos. Como los resultados podían depender de cómo se dividieran los datos, se realizaron 100 particiones aleatorias utilizando el Servidor de Números Cuánticos de la Universidad Nacional de Australia (<https://qrng.anu.edu.au>). Esto permitió la caracterización estadística de las métricas de desempeño de los modelos.

### 4. Resultados

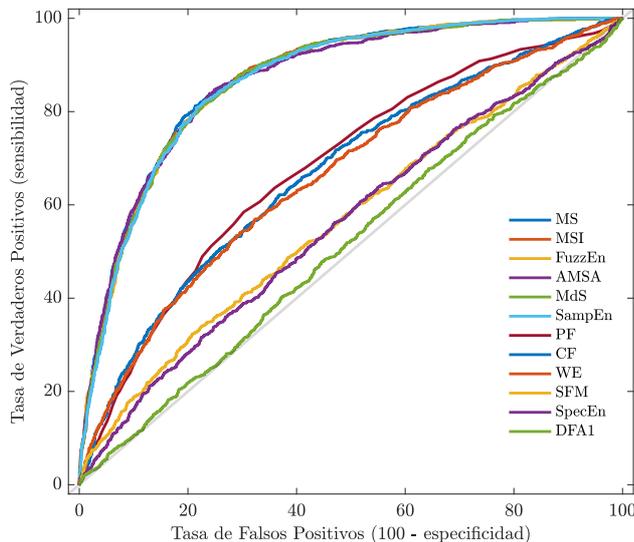
Los resultados del análisis monoparamétrico se muestran en la Tabla 2. Dicha tabla resume por un lado los resultados del análisis de curva ROC para todos los datos mediante el AUC, y se muestra a su vez la PB obtenida para el modelo LR monoparamétrico. En este último caso, se muestra el

valor mediano para las 100 particiones 10-fold CV, así como el rango interdecil (IDR, entre 10%-90%). Se observa que los valores de AUC y PB son muy similares para las mejores características (el resto no se muestra). Éstas se corresponden con las que mejores resultados han reportado en la literatura [7,8], e incluyen predictores basados en la pendiente (MS, MdS), entropías muestral y difusa (SmpEn, FuzzyEn), AMSA, y el análisis de la forma de onda basados en diagramas de Poincaré (MSI).

La Figura 2 muestra el análisis de curva ROC para los mejores y peores predictores. Se observa que predictores propuestos en la literatura como la entropía wavelet (WE), el análisis de fluctuación sin tendencia (DFA), o la medida de planicie espectral (SFM) y entropía espectral (SpecEn), que miden características asociadas exclusivamente a la no linealidad y desorden de la forma de onda FV, no producen resultados satisfactorios. Es importante indicar que las implementaciones de entropía muestral y difusa con altos valores de AUC y PB (Tabla 2) se modificaron para incluir información de amplitud de la FV [7,8].

Característica	AUC (%)	PB (%)
MS	87.0	79.5 (79.4-79.6)
SmpEn	86.4	79.2 (79.1-79.3)
FuzzyEn	86.6	79.1 (79.0-79.2)
AMSA	86.5	79.0 (78.9-79.0)
MdS	86.5	78.8 (78.7-78.9)
MSI	86.6	78.5 (78.4-78.6)

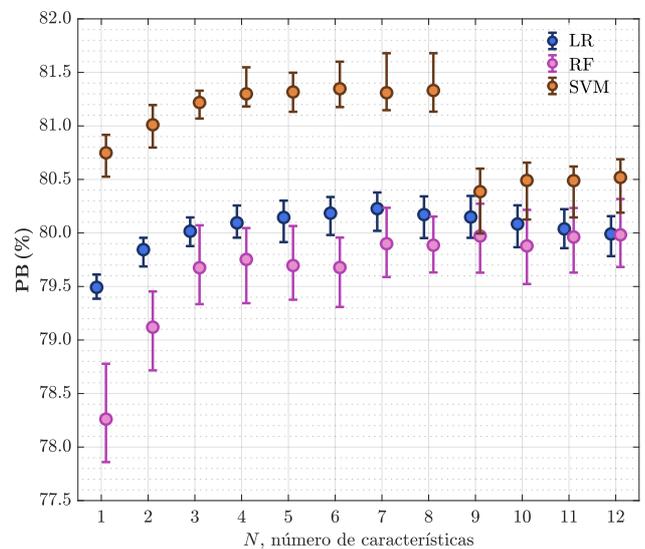
**Tabla 2.** Ranking de las mejores características para el análisis de curva ROC (AUC), y para el clasificador RL monoparamétrico (PB). En la PB se muestra el valor mediano (IDR) para las 100 particiones.



**Figura 2.** Curvas ROC para los 6 parámetros con mejor y peor poder discriminativo.

Al combinar los parámetros en los tres modelos de AM se obtuvieron los resultados mostrados en la Figura 3 donde se muestra el valor mediano (IDR) para las 100 particiones 10-fold CV, y se compara la PB de los tres modelos de AM estudiados en función del número de características de los

modelos. Para los modelos basados en RF se emplearon 150 árboles,  $\sqrt{N}$  predictores por rama, y 40 observaciones en la última división, ya que en los experimentos preliminares produjeron valores de PB más elevados. Tal y como se observa en la figura los mejores resultados se obtuvieron para los modelos basados en SVM, que superaron en más de 1.2-puntos de PB al resto de modelos, y en 1.9-puntos al mejor modelo monoparamétrico (Tabla 2). Además, se observa que los modelos no mejoran con la inclusión de nuevas características a partir de un número bajo de características ( $N = 6$  para el caso de la SVM). Por ejemplo, los mejores resultados se obtienen para la SVM con 6 características, que resultan en valores de 81.4 (81.2-81.6)%, 83.4 (83.1-84.0)% y 79.3 (79.1-80.0)% para la PB, SE y SP, respectivamente. Los mejores resultados para LRM y RF se obtuvieron para  $N = 7$  y  $N = 9$ , con valores de PB, SE y SP de 80.2 (80.0-80.4)%, 81.4 (81.0-81.6)% y 79.0 (78.9-79.2)% para LRM y de 80.0 (79.7-80.3)%, 81.6 (80.9-82.2)% y 78.4 (78.1-78.8)% para RF.



**Figura 3.** Mediana (IDR) de las puntuaciones de rendimiento para los modelos LRM, RF y SVM en función del número de características utilizadas.

### 5. Discusión

En este estudio se ha analizado el poder discriminativo de los predictores de éxito de la desfibrilación basados en análisis de la forma de onda FV pre-descarga, empleando para ello la base de datos con anotaciones clínicas más extensa reportada en la literatura para este problema. Se ha analizado también la mejora en la predicción asociada al uso de modelos avanzados de AM.

Los mejores predictores descritos en la literatura muestran valores similares de AUC y PB, corroborando los resultados reportados en estudios previos [7,8], aunque algunos de ellos no demostraron poder discriminativo sobre la base de datos de estudio. Además, los mejores predictores (Tabla 2) contienen información asociada a la amplitud, bien directamente (AMSA, y pendiente), o mediante versiones modificadas de los predictores con información de amplitud (SmpEn, FuzzyEn) [6,7]. Es bien conocido que el valor de amplitud de la FV está asociado a la duración de la misma, y predice en gran medida el éxito de la desfibrila-

ción [17]. La combinación de parámetros en modelos de AM (Figura 3) demuestra que es posible obtener mejoras de la PB, por ejemplo, el mejor modelo multiparamétrico mejora en 1.9-puntos la PB del mejor modelo LR monoparamétrico. Sin embargo, esta mejora es limitada, ya que supone solo un recorte de un 10% sobre los 20-puntos de margen para la PB de 80% del modelo monoparamétrico. Se concluye así que los parámetros de discriminación proporcionan información muy parecida de la forma de onda ECG de la FV. Dado el tamaño de la base de datos disponible, es posible que en el futuro se puedan desarrollar métodos eficientes de extracción de características de la FV basadas en aprendizaje profundo (CNN). Este tipo de soluciones ya han sido demostradas para el análisis del ECG y la detección de FV [18]. Además, otra línea de trabajo para la mejora de la predicción son los modelos multimodales que integren información de varias señales, en particular información sobre el CO<sub>2</sub> expirado que proporciona una medida subrogada de perfusión. Estos modelos han demostrado ya mejoras en la predicción del éxito de la desfibrilación [19], pero la disponibilidad de las señales es limitada y no se podrían analizar con los datos del presente estudio.

Finalmente, estos nuevos métodos se podrían implementar en los DEAs para mejorar la supervivencia en la PCREH. Actualmente, los DEA analizan segmentos cortos del ECG para decidir si es necesario desfibrilar; sin embargo, si se predice que la desfibrilación será no exitosa conviene posponer el shock y continuar con la reanimación.

## 6. Conclusiones

Como conclusión, los mejores predictores de la literatura reportan valores muy parecidos de AUC y PB. Además, las técnicas de AM multiparamétrico mejoran la PB, siendo la SVM con kernel gaussiano y 8 predictores la que mejores resultados reporta. Sin embargo, la mejora es leve respecto de modelos monoparamétricos porque muchos de los predictores cuantifican características similares de la FV. Por lo tanto, sería interesante introducir la extracción exhaustiva de características usando CNNs.

## Agradecimientos

Este trabajo ha recibido ayuda financiera del Ministerio de Ciencia, Innovación y Universidades, proyecto RTI-2018-101475-BI00, junto con el Fondo Europeo de Desarrollo Regional (FEDER), y del Gobierno Vasco por medio de la subvención a grupos de investigación IT-1229-19 y la beca del programa Ikerbasque IKAS-C\_2019\_1\_0275.

## Referencias

- [1] Consejo Español de Resucitación Cardiopulmonar (CERCP), «Día Europeo de Concienciación del Paro Cardíaco, 16 de octubre». Disponible en: [https://www.cerpc.org/images/stories/recursos/2015/varios/NOTA\\_PRENSA\\_DIA\\_PARO\\_CARDIACO\\_2015.pdf](https://www.cerpc.org/images/stories/recursos/2015/varios/NOTA_PRENSA_DIA_PARO_CARDIACO_2015.pdf).
- [2] M. Rubart, «Mechanisms of sudden cardiac death», *J. Clin. Invest.*, vol. 115, n.º 9, pp. 2305-2315, ago. 2005.
- [3] N. Rodríguez de Viguri, J. López Mesa, y M. Ruano Campos M, *Manual de soporte vital avanzado.*, 4.ª ed. Madrid, España: Masson, 2007.
- [4] G. Ristagno, Y. Li, F. Fumagalli, A. Finzi, y W. Quan, «Amplitude spectrum area to guide resuscitation—A retrospective analysis during out-of-hospital cardiopulmonary resuscitation in 609 patients with ventricular fibrillation cardiac arrest», *Resuscitation*, vol. 84, n.º 12, pp. 1697-1703, dic. 2013.
- [5] G. Ristagno *et al.*, «Amplitude Spectrum Area to Guide Defibrillation: A Validation on 1617 Patients With Ventricular Fibrillation», *Circulation*, vol. 131, n.º 5, pp. 478-487, feb. 2015.
- [6] B. Chicote *et al.*, «Application of Entropy-Based Features to Predict Defibrillation Outcome in Cardiac Arrest», *Entropy*, vol. 18, n.º 9, p. 313, ago. 2016.
- [7] B. Chicote *et al.*, «Fuzzy and Sample Entropies as Predictors of Patient Survival Using Short Ventricular Fibrillation Recordings during out of Hospital Cardiac Arrest», *Entropy*, vol. 20, n.º 8, p. 591, ago. 2018.
- [8] M. He *et al.*, «Combining multiple ECG features does not improve prediction of defibrillation outcome compared to single features in a large population of out-of-hospital cardiac arrests», *Crit. Care*, vol. 19, n.º 1, p. 425, dic. 2015.
- [9] C. G. Brown y R. Dzwonczyk, «Signal Analysis of the Human Electrocardiogram During Ventricular Fibrillation: Frequency and Amplitude Parameters as Predictors of Successful Countershock», *Ann. Emerg. Med.*, vol. 27, n.º 2, pp. 184-188, feb. 1996.
- [10] I. Jekova, F. Mougeolle, y A. Valance, «Defibrillation shock success estimation by a set of six parameters derived from the electrocardiogram», *Physiol. Meas.*, vol. 25, n.º 5, pp. 1179-1188, oct. 2004.
- [11] A. Neurauter *et al.*, «Prediction of countershock success using single features from multiple ventricular fibrillation frequency bands and feature combinations using neural networks», *Resuscitation*, vol. 73, n.º 2, pp. 253-263, may 2007.
- [12] T. Eftestøl, K. Sunde, S. Ole Aase, J. H. Husøy, y P. A. Steen, «Predicting Outcome of Defibrillation by Spectral Characterization and Nonparametric Classification of Ventricular Fibrillation in Patients With Out-of-Hospital Cardiac Arrest», *Circulation*, vol. 102, n.º 13, pp. 1523-1529, sep. 2000.
- [13] L. D. Sherman, T. D. Rea, J. D. Waters, J. J. Menegazzi, y C. W. Callaway, «Logarithm of the absolute correlations of the ECG waveform estimates duration of ventricular fibrillation and predicts successful defibrillation», *Resuscitation*, vol. 78, n.º 3, pp. 346-354, sep. 2008.
- [14] Y. Gong, Y. Lu, L. Zhang, H. Zhang, y Y. Li, «Predict Defibrillation Outcome Using Stepping Increment of Poincare Plot for Out-of-Hospital Ventricular Fibrillation Cardiac Arrest», *BioMed Res. Int.*, vol. 2015, pp. 1-7, 2015.
- [15] T. Hastie, J. Friedman, y R. Tibshirani, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2001.
- [16] C. A. González, «SVM: Máquinas de Vectores Soporte», presentado en Escuela de Ingeniería Informática de Valladolid.
- [17] W.D. Weaver, L.A. Cobb, D. Dennis, R. Ray, A.P. Hallstrom, M.K. Copass, «Amplitude of ventricular fibrillation waveform and outcome after cardiac arrest», *Ann Intern Med* 1985;102(1):53–5.
- [18] X. Jaureguibeitia, G. Zubia, U. Irusta, E. Aramendi, et al. «Shock Decision Algorithms for Automated External Defibrillators Based on Convolutional Networks», *IEEE Access*, vol. 8, pp. 154746-58, 2020.
- [19] B. Chicote, E. Aramendi, P. Owens, M. Daya, A. Idris, «Value of capnography to predict defibrillation success in out-of-hospital cardiac arrest», *Resuscitation*, vol. 138, pp. 74-81, 2019