

Detección de fibrilación auricular en electrocardiogramas de corta duración.

Marco Scavino^a, Virginia Estragó^b, Matías Muñoz^c, Andrés Castrillejo^d, Ramón Álvarez-Vaz^e.

^{a,d,e} *Instituto de Estadística (IESTA), Facultad de Ciencias Económicas y de Administración, Universidad de la República, Uruguay.*

^{b,c} *Comisión Honoraria para la Salud Cardiovascular (CHSCV), Presidencia de la República, Uruguay.*

^a *mscavino@iesta.edu.uy*, ^b *vestrago@cardiosalud.org*, ^c *mmunoz@cardiosalud.org*,

^d *castrillejoandres@gmail.com*, ^e *ramon@iesta.edu.uy*.

Resumen

En este trabajo se presentan los lineamientos generales para la aplicación de dos algoritmos de aprendizaje automático para la identificación de una patología cardíaca, la fibrilación auricular (FA), a partir de la señal electrocardiográfica (ECG) de una derivación y de corta duración registrada con un dispositivo de tecnología electrónica móvil (DTEM). Visto la importancia del diagnóstico temprano de patologías cardiovasculares tales como la FA, nuestro objetivo es mejorar la capacidad de clasificación del DTEM el cual, en la práctica, deja sin clasificar un conjunto relevante de registros de ECG.

Se analiza el desempeño de técnicas de clasificación supervisada tales como árboles generados por partición recursiva y los bosques aleatorios, empleadas en combinación con métodos de extracción de características de la señal ECG, aplicadas a un banco de entrenamiento internacional y a un conjunto de prueba nacional de registros de ECG, generado en 2019, de una población adulta mayor del Uruguay, en virtud de una colaboración entre la CHSCV y el Plan Ibirapitá. Los resultados obtenidos se interpretan a través de los diagnósticos disponibles de los ECG realizados por expertos cardiólogos clínicos.

Palabras clave: electrocardiograma, fibrilación auricular, clasificación supervisada.

Introducción

La fibrilación auricular (FA) es un trastorno del ritmo cardíaco frecuente en el adulto y cuya prevalencia aumenta con la edad. Entre sus implicancias clínicas, la FA se asocia al accidente cardiovascular, entre 20% y 30% de los pacientes con ictus isquémico tienen un diagnóstico de FA antes, durante o después del evento inicial de accidente cerebrovascular. La FA puede cursar de forma asintomática pero exponer a quien la padece a un riesgo significativo de morbimortalidad y por esa razón es relevante diagnosticar esta arritmia. La contracción cardíaca requiere de un sistema de

estimulación y conducción eléctrica que está compuesto por fibras especializadas a nivel del corazón. El electrocardiograma (ECG) es el registro gráfico de los potenciales eléctricos que genera el corazón y representa el funcionamiento del corazón. Cualquier comportamiento anormal en la señal de ECG es una medida indicativa de un mal funcionamiento del corazón, denominado condición de arritmia. El ECG de 12 derivaciones se considera el patrón oro para el diagnóstico de muchas enfermedades cardiovasculares, entre ellas la FA. Sin embargo, cabe observar que es posible llegar al diagnóstico de FA con el registro de ECG de una sola derivación. El trazado de un ECG

representa el voltaje de la actividad eléctrica del corazón en función del tiempo. El registro de un ciclo cardíaco normal (ver Figura 1 más abajo) se compone de una onda P, que corresponde a la activación auricular, un complejo QRS, asociado a la fase de despolarización ventricular, y una onda T, que corresponde a la fase de repolarización ventricular. La medida del intervalo RR que transcurre desde el inicio de una onda R hasta el inicio de la onda R siguiente es clínicamente relevante.

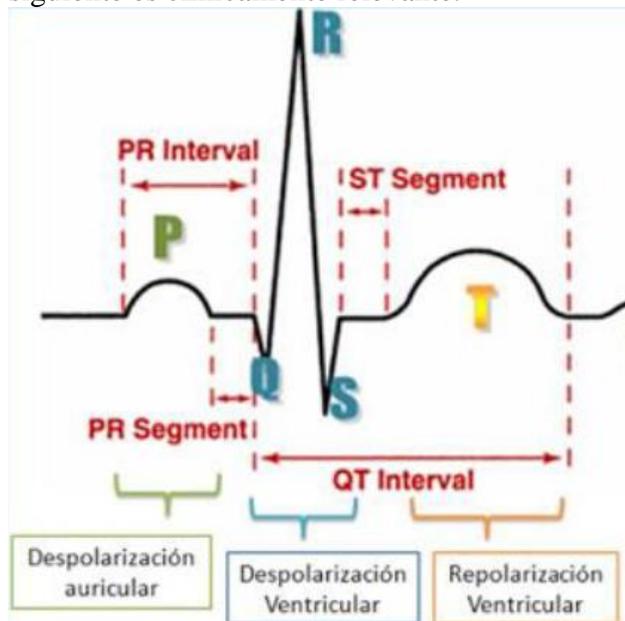


Figura 1.- Esquematación de un latido en un ritmo cardíaco normal.

En un ritmo normal se observa la regularidad de los intervalos RR, cuya duración depende de la frecuencia cardíaca. Mientras que, en términos generales, la FA puede ser detectada debido a la presencia de intervalos RR irregulares, y a la ausencia de onda P (entre otros aspectos).

Hoy en día es posible detectar la FA utilizando dispositivos de tecnología electrónica móvil (DTEM) que permiten grabar, almacenar y transmitir los registros de ECG en forma remota. Nuestro trabajo tiene el objetivo de validar la utilización de un DTEM para la detección de fibrilación auricular en un grupo de adultos mayores de Uruguay.

La colaboración entre la Comisión Honoraria para la Salud Cardiovascular (CHSCV) y el Plan Ibirapitá, que promueve la inclusión digital al servicio de las personas mayores con el fin de mejorar la inclusión social, la participación y la equidad, ha posibilitado la realización de talleres en los cuales centenares de personas han participado del estudio de tamizaje de FA apoyando el pulpejo de los dedos sobre las almohadillas del dispositivo y recibiendo su diagnóstico validado por expertos cardiólogos.



Figura 2.- Dispositivo de tecnología móvil.

Banco de datos del plan piloto

En el ámbito de un plan piloto previo a la realización de los talleres CHSCV-Plan Ibirapitá, se obtuvieron los registros de ECG de 114 participantes tanto con el DTEM de una derivación y de corta duración como con el ECG de doce derivaciones.

| | Possible FA | Normal | Sin Clasificar | Total |
|--------------|-------------|-----------|----------------|------------|
| Con FA | 4 | 0 | 1 | 5 |
| Sin FA | 3 | 86 | 20 | 109 |
| Total | 7 | 86 | 21 | 114 |

18%

Tabla 1.- Resultados del plan piloto.

Se ha constatado que el DTEM dejó sin clasificar el 18% de los registros, los cuales sí fueron clasificados por el equipo de cardiólogos

de la CHSCV, quienes emplearon el patrón de oro dado por el ECG de doce derivaciones.

De aquí la importancia de desarrollar algoritmos de clasificación que, utilizando la salida del DTEM, permitan mejorar la capacidad de clasificación del dispositivo.

A los efectos del posterior análisis estadístico, cada registro de salida del DTEM, de duración 30 segundos y frecuencia de muestreo 30 Hz, está sometido a un procedimiento de conversión, del original PDF a TSV (valores separados por tabuladores).



Figura 3.- Salida original PDF del DTEM: parte del registro 110305.

De esta forma se obtuvieron 9000 pares de puntos por cada registro de ECG. A continuación, se muestran algunos ejemplos de registros de ECG, clasificados y no clasificados por el DTEM, junto con el correspondiente diagnóstico de los cardiólogos.

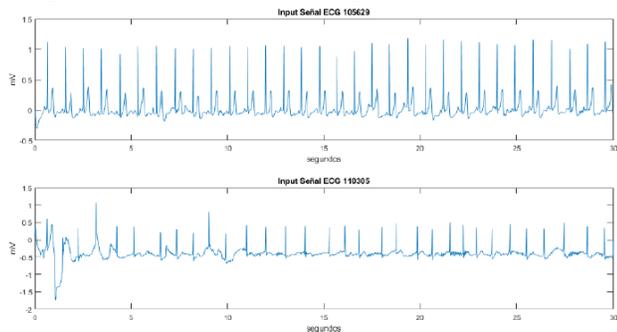


Figura 4.- Cuadro superior: registro (105629) de ECG clasificado como normal por el DTEM y los expertos. Cuadro inferior: registro (110305) de ECG clasificado como FA por el dispositivo y los expertos.

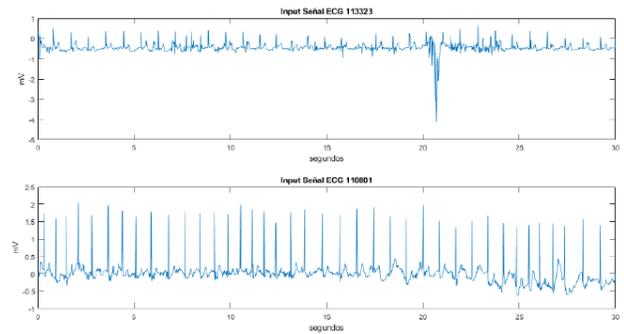


Figura 5.- Cuadro superior: registro (113323) de ECG no clasificado por el dispositivo y clasificado como NO FA por los expertos. Cuadro inferior: registro (110801) de ECG no clasificado por el dispositivo y clasificado como FA por los expertos.

Visto la relevancia clínica de los intervalos RR hemos empleado el algoritmo de Pan y Tompkins, que consta de una etapa de preprocesamiento y de otra basada en reglas de decisiones para la individuación del pico de la onda R en cada latido. Posteriormente, hemos utilizado dos técnicas de clasificación supervisada entrenadas con 3326 registros procedentes de la competencia Physionet 2017, y aplicadas al conjunto de prueba del banco de datos del plan piloto nacional descrito anteriormente.

En este trabajo se introducen las principales características para el análisis de señales ECG de una derivación y de corta duración con la finalidad de detectar automáticamente un tipo de arritmia, la fibrilación auricular (FA). En la presente sección 1 se ha planteado el problema de interés. La sección 2 contiene la descripción de los datos empleados en este estudio y de los aspectos metodológicos esenciales. En la sección 3 se presentan los resultados obtenidos, luego las conclusiones del estudio.

Metodología

Banco de datos de entrenamiento

Se considera un conjunto de entrenamiento de 3326 registros de ECG de una sola derivación de duración 30 segundos (frecuencia de muestreo 30 Hz) seleccionados a partir del banco público de entrenamiento internacional de 8528 registros constituido para la competencia 2017 “PhysioNet/Computing in Cardiology (CinC)”,

<https://physionet.org/content/challenge-2017/1.0.0/>, Clifford et al (2017).

De estos 3326 registros de ECG, 2932 son etiquetados como normales, los restantes 394 como FA.

Detección del complejo QRS

A los efectos de proponer un conjunto de rasgos característicos adecuado para discriminar entre registros de ECG normales y registros con FA, es necesario detectar el ciclo cardíaco, lo cual corresponde a la detección del complejo QRS de cada latido. La detección del complejo QRS, que permite ubicar el punto (o pico) R de cada latido, puede llevarse a cabo aplicando el algoritmo de Pan y Tompkins (1985). Este algoritmo consiste primero en una etapa de preprocesamiento de la señal ECG y en otra etapa decisional basadas en umbrales adaptativos. La etapa de preprocesamiento consiste en los siguientes cuatro pasos:

- Filtrado pasa-banda: pasa-abajo y pasa-alto en cascada, para aislar la energía QRS predominante centrada en 10 Hz, atenuando las bajas frecuencias características de las ondas P y T y la deriva de la línea base.
- Filtro derivador: destacar las pendientes rápidas de la señal, que, en el caso de la señal ECG, coincide con el complejo QRS.
- Transformación no lineal: se eleva al cuadrado la señal, convirtiéndola en positiva. Se destacan las frecuencias más altas en la señal, que se deben principalmente al complejo QRS.

- Integrador con ventana móvil para incluir una medida del ancho del complejo QRS en la señal procesada.

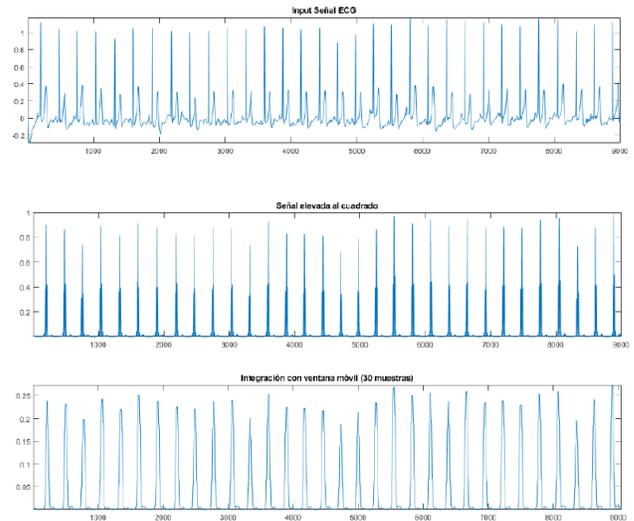


Figura 6.- Etapa de preprocesamiento del algoritmo de Pan y Tompkins.

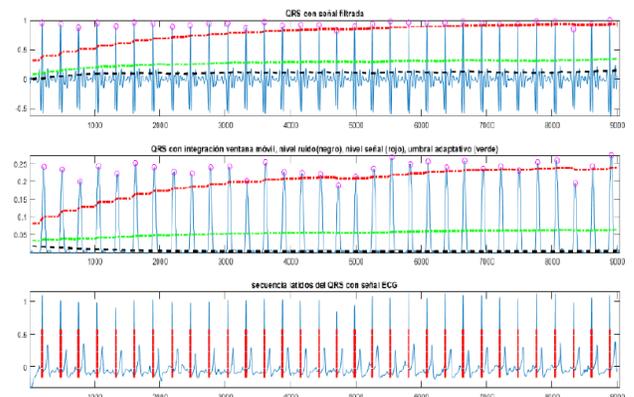


Figura 7.- Detección de los picos R a través del algoritmo de Pan y Tompkins.

Conjunto de rasgos característicos y técnicas de clasificación supervisada

Una vez ubicados los puntos R que corresponden a cada latido, se puede obtener un conjunto de rasgos característicos de cada señal ECG basados en los intervalos RR. En este estudio se han considerado los siguientes nueve rasgos:

- Primeras diferencias de los intervalos RR;

- Proporción de intervalos RR que difieren por más de 50 ms del intervalo precedente;
- máximo, mínimo y desviación típica de los intervalos RR;
- cociente entre mínimo y máximo intervalo RR;
- media de los intervalos RR mayores al percentil 90;
- media de los intervalos RR comprendidos entre los percentiles 80 y el 90;
- diferencia entre las dos medias anteriores.

Finalmente, al conjunto de rasgos característicos basados en los intervalos RR, se incorporan otros aspectos de la señal ECG obtenidos a partir del análisis en el dominio de frecuencia (periodograma).

El conjunto de rasgos característicos así determinado constituye la información de entrada para dos algoritmos de clasificación supervisada, árboles generados por partición recursiva (paquete R 'rpart') y bosques aleatorios (paquete R 'randomForest') que se entienden adecuado aplicar, en primera instancia, al conjunto de entrenamiento seleccionado de Physionet 2017 y, en segunda instancia, al conjunto de prueba formado por los registros de ECG del plan piloto, descrito en la introducción.

Resultados y Discusión

Se aplica el modelo de árbol de clasificación empleando 'rpart' en R a todo el conjunto de rasgos característicos descrito en la sección anterior. La importancia de cada variable se visualiza en el siguiente gráfico:

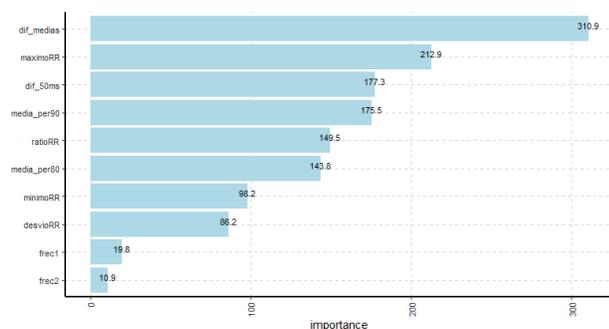


Figura 8.- Importancia de las variables en el modelo árbol de clasificación.

La siguiente tabla resume el desempeño del modelo árbol de clasificación aplicado al banco de datos de entrenamiento:

Tabla 2.- Desempeño árbol de clasificación.

| | | Clase verdadera | |
|-----------------------------|--------|-----------------|-----|
| | | Normal | FA |
| Resultados del clasificador | FA | 19 | 338 |
| | Normal | 2913 | 56 |

El puntaje F_1 respecto de FA definido como el cociente $(2 \cdot \#verdaderos\ positivos) / (2 \cdot \#verdaderos\ positivos + \#falso\ positivos + \#falsos\ negativos)$ es igual a 0,9001.

El desempeño predictivo del algoritmo árbol de clasificación para los registros de ECG en el plan piloto respecto de la clasificación adoptada por el DTEM y respecto del diagnóstico de los cardiólogos (basado en el ECG de 12 derivaciones) se resume, respectivamente, en las dos siguientes tablas:

Tabla 3.- Predicción árbol de clasificación vs diagnóstico del DTEM.

| | | Resultados DTEM | | |
|-----------------------------|--------|-----------------|------------|----------------|
| | | Normal | Posible FA | Sin clasificar |
| Resultados del clasificador | FA | 1 | 4 | 2 |
| | Normal | 85 | 3 | 19 |

Tabla 4.- Predicción árbol de clasificación vs diagnóstico de los cardiólogos.

| | | Resultado cardiólogos | |
|-----------------------------|--------|-----------------------|----|
| | | Normal | FA |
| Resultados del clasificador | FA | 3 | 4 |
| | Normal | 106 | 1 |

Se aplica el modelo de bosques aleatorios empleando 'randomForest' en R a todo el conjunto de rasgos característicos descrito en la sección anterior. La importancia

de cada variable se visualiza en el siguiente gráfico:

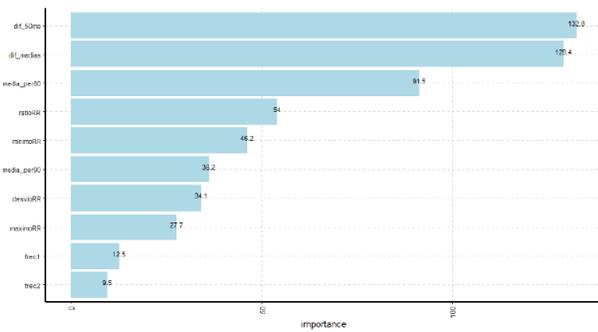


Figura 9.- Importancia de las variables en el modelo bosques aleatorios.

Las dos siguientes tablas resumen el desempeño predictivo del algoritmo de clasificación basado en bosques aleatorios para los registros de ECG en el plan piloto respecto de la clasificación adoptada por el DTEM y respecto del diagnóstico de los cardiólogos:

Tabla 5.- Predicción bosques aleatorios vs diagnóstico del DTEM.

| Resultados del clasificador | | Resultados DTEM | | |
|-----------------------------|--------|-----------------|------------|----------------|
| | | Normal | Posible FA | Sin clasificar |
| Resultados del clasificador | FA | 1 | 5 | 2 |
| | Normal | 85 | 2 | 19 |

Tabla 6.- Predicción bosques aleatorios vs diagnóstico de los cardiólogos.

| Resultados del clasificador | | Resultado cardiólogos | |
|-----------------------------|--------|-----------------------|----|
| | | Normal | FA |
| Resultados del clasificador | FA | 4 | 4 |
| | Normal | 105 | 1 |

Los resultados evidencian el potencial de técnicas automáticas de detección de FA basadas en un número relativamente pequeño de características extraídas de una señal de ECG de una derivación y de corta duración, y en la aplicación de técnicas de aprendizaje automático.

Conclusiones

Los resultados muestran que el desempeño predictivo de los dos algoritmos de clasificación supervisada aquí empleados y aplicados al conjunto seleccionado de rasgos característicos de señales ECG convertidas de manera automática de la salida de un DTEM, genera únicamente un falso negativo. De hecho, se trata de un registro que resultó difícil de interpretar también para los cardiólogos.

El estudio aquí presentado considera únicamente los registros de ECG obtenidos en el ámbito de un plan piloto que precedió un diseño de muestreo de carácter nacional llevado a cabo en Uruguay en una colaboración entre la CHSCV y el Plan Ibirapitá.

En la actualidad, se está realizando la etapa de calibración y validación de los diagnósticos informados por los cardiólogos basados en la salida del DTEM. De esta manera, una vez construidos procedimientos automáticos de clasificación de los registros de ECG con elevada capacidad predictiva, se entiende posible poner las premisas para un estudio de la estimación de la prevalencia de la FA a nivel nacional.

Referencias

- Clifford G. D. et al. (2017). AF Classification from a Short Single Lead ECG Recording: the PhysioNet/Computing in Cardiology Challenge 2017. **Computing in Cardiology**. Vol. 44. 4 Pp.
- Pan, J. y Tompkins, W. J. (1985). A real-time QRS detection algorithm. **IEEE Trans. Biomed. Eng.**, Vol. 32. Pp. 230-236.
- Cryer, J. D. y Chan, Kung-Sik. (2008). **Time Series Analysis: with application in R**. Second Edition. Springer, New York.
- Therneau, T., Atkinson, B. y Ripley, B. (2015). rpart: Recursive Partitioning and Regression Trees. URL <https://CRAN.R-project.org/package=rpart>. R package version 4.1-15.
- Liaw, A. y Wiener, M. (2002). Classification and Regression by random Forest. **R News**. Vol. 2. Pp. 18-22.