

Classification of patients into sane and diseased, using an Electrocardiogram by applying Machine Learning techniques

1st Jeffry Gondres Dannenberger, 2nd Diego Jiménez Alvarado, 3rd Allan Berrocal Rojas

Escuela de Ciencias de la Computación e Informática

Universidad de Costa Rica

San José, Costa Rica

{jeffry.gondres, diego.jimenezalvarado, allan.berrocal}@ucr.ac.cr

Abstract—Early detection of cardiovascular diseases is important as according to the World Health Organization (WHO), 17.9 million people die from these diseases each year.

Electrocardiograms (ECGs) allow individuals to assess the state of their hearts and are crucial for preventive detection. However, interpreting the results is not easy, even for cardiology specialists, especially considering the large number of patients that need to be analyzed in medical centers. Therefore, a system that can streamline the process and increase the probability of an accurate diagnosis would benefit both medical personnel and patients.

In this paper, a dataset with over twenty thousand pre-labeled ECG signals was used. Two machine learning techniques, LSTM and XGBoost, were explored to classify the presence or absence of disease in patients. The following article aims to present the results of both models for a binary classification task (sick or healthy) and a multi-class classification task (type of disease).

Index Terms—machine learning, ECG, deep learning, LSTM, XGBoost

I. INTRODUCCIÓN

La Inteligencia Artificial (IA) ha protagonizado un papel fundamental en avances científicos en los últimos años. La implementación de robots en la logística, vehículos autónomos y proyectos médicos, han venido a revolucionar el mundo en el cual vivimos. La búsqueda de la eficiencia genera una necesidad constante en entender el funcionamiento del ser humano al tomar decisiones, para replicarlo y buscar mejorarlo. Gracias al desarrollo computacional y la cantidad de datos disponibles, han permitido desarrollar mejores técnicas dentro de la IA. En particular, el Aprendizaje de Máquina, es uno de estos campos donde se ha generado una mejora considerable, por lo que ha surgido la posibilidad de hacer un uso efectivo de la misma a través del desarrollo de herramientas de diagnóstico, sin intervención humana [2]. De esta forma, el Aprendizaje Automático (ML, *Machine Learning* en inglés) se ha convertido en un área de estudio necesaria para la investigación en el sector salud.

ML es una rama de la IA y consiste en algoritmos matemáticos que aprenden de los datos introducidos, para luego generar respuestas inteligentes sin ser programadas, ni requerir intervención humana [4]. Es por esto que actualmente se ha ampliado su utilización en el análisis o tratamiento de

datos derivados de diagnósticos clínicos, cuyas fuentes pueden ser ensayos, reportes, imágenes, vídeos, texto, audio o archivos generados por dispositivos médicos [2].

La aplicación de ML en las ciencias de la salud es viable [13]. Así, por ejemplo, uno de los usos donde puede ofrecer grandes aportes al campo médico, es sobre la detección temprana de enfermedades [13]. De esta manera, ML puede ayudar a los médicos a diagnosticar una enfermedad lo antes posible [10].

II. ANTECEDENTES

Según [14] en el 2020, se publicaron más de veinte mil registros de formas de onda ECG, clasificadas por cardiólogos creando un conjunto de datos para aplicar ML, donde las etiquetas, son diagnósticos divididos en clases y subclases. Este conjunto abarca una amplia gama de clases de diagnóstico, incluyendo, en particular, una gran cantidad de registros de individuos sanos y enfermos.

Con la disponibilidad de PTB-XL, los investigadores científicos de datos, ahora tienen acceso a un gran conjunto de diversos diagnósticos, que puede impulsar significativamente el avance de algoritmos de IA aplicados a la interpretación del ECG [12], lo que a su vez mejoraría la detección y el diagnóstico temprano de enfermedades cardiovasculares, según estos autores.

III. MARCO TEÓRICO

En la actualidad, las técnicas de ML son ampliamente utilizadas en el campo de la IA debido a su capacidad para aprender. Entre estas técnicas, destacan algoritmos como las Redes Neuronales (RN), las cuales han demostrado excelentes resultados en diversas aplicaciones, como el caso del modelo ChatGPT. Las RN se inspiran en el funcionamiento de las redes neuronales en los organismos vivos, donde un conjunto de neuronas interconectadas trabajan en conjunto para resolver un problema, y cada neurona aprende de las demás [1].

En este contexto, la experiencia adquirida por una neurona se transfiere a las demás, fortaleciendo así el aprendizaje de las neuronas anteriores. Esta dinámica de retroalimentación

continua permite que la RN aproveche la capacidad de "ajustarse automáticamente" propia de ML. Sin embargo, no solo las RN se benefician de esta técnica; el algoritmo XGBoost también la utiliza con éxito [5]. Cabe destacar que existen diferentes tipos de redes neuronales profundas (DNN), como las Redes Neuronales Convolucionales (CNN, *Convolutional Neural Network*) y las Redes Neuronales Recurrentes (RNN, *Recurrent Neural Network*).

Las RNN son muy efectivas para datos secuenciales, por tanto, una de sus aplicaciones más importantes es en el campo de las series de tiempo [9]. En las RNN, se asume que todas las entradas y salidas son dependientes [7], es decir, cada salida depende de la anterior, lo que les permite capturar dependencias en secuencias, como en las series de tiempo, donde el valor en un tiempo T depende del valor anterior (T-1) y así sucesivamente, como se muestra en la siguiente ecuación.

$$s[T] = f(s[T-1]) \quad (1)$$

El término *recurrente* en las RNN proviene del hecho de que tiene conexiones cíclicas (recursivas) y se realiza el mismo cálculo en cada elemento de la secuencia. Esto le permite aprender (o memorizar) partes de los datos para luego hacer inferencias a partir de datos previamente desconocidos. La ventaja de una RNN es que puede escalar a secuencias mucho más largas que los modelos que no se basan en secuencias [6].

Sin embargo, las RNN presentan un problema de dependencia a largo plazo debido al desvanecimiento del gradiente [3], lo que propició la aparición de otras arquitecturas relacionadas que enfrentaran dicho problema. De lo anterior surgen las redes de memoria a corto y largo plazo (*LSTM, Long Short-Term Memory*) que son un tipo de RNN especialmente diseñadas para modelar secuencias temporales. Las redes de tipo LSTM son capaces de aprender dependencias a más largo plazo y de "recordar" la información que se cree que es más importante, por lo que se consideran muy efectivas en análisis de secuencias [15].

Por otro lado, la series de tiempo contenidas en un ECG no son estacionales, por lo que no se recomiendan técnicas de pronóstico sobre estas, sino de clasificación. Son más bien sus diferencias con respecto a una señal normal lo que puede o no determinar la ausencia o presencia de la enfermedad, donde la señal normal, es normal para el paciente en sí, pues las características particulares de este, pueden hacer que las señales "normales" sean diferentes entre ellas. Es por lo anterior que como segunda opción a las redes LSTM, tenemos los algoritmos de clasificación mediante árboles de decisión (DT, *Decision Tree*). En estos algoritmos, las clasificaciones se hacen utilizando características extraídas de la señal de ECG, específicamente lo que se conoce como análisis de la variabilidad de los pulsos del corazón (HRV, *Heart Rate Variability*). Este análisis considera métricas con respecto a la forma del pulso del corazón, así por ejemplo: HRV SDNN, mide la variabilidad de la frecuencia cardiaca, por lo que mediante la obtención de estas métricas se puede extraer características de la serie de tiempo y analizar estas métricas

mediante algoritmos de clasificación, transformando así la serie de tiempo en una muestra de n características.

Para el procesamiento de los datos anteriores, XGBoost presenta un mejor desempeño si se comparan contra una DNN [8]. Esto puede explicarse porque este algoritmo combina los DT con las técnicas de las RN en cuanto a la consideración del error para el siguiente cálculo [5], según se demuestra a continuación.

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \quad (2)$$

donde:

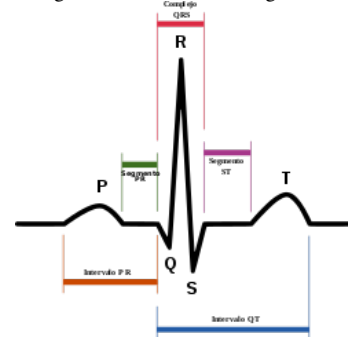
\hat{y}_i es la predicción para el modelo de la muestra x_i

K es el número total de árboles en el modelo.

$f_k(x_i)$ es la predicción del árbol k para la muestra x_i

Según lo anterior, la interconexión entre los árboles y su evaluación permite un comportamiento más adecuado para los datos de enfermedades cardiacas, desprendidos del análisis HRV.

Fig. 1. ECG - Prueba diagnóstica



En la Fig. 1, se puede observar como los puntos PQRST representan el pulso del corazón, en un análisis HRV. Existen diferentes métricas sobre esos puntos, obtenidas a partir de la distancia entre cada uno de los puntos mencionados como instantes en el tiempo a lo largo de la señal ECG. Algunas de las mediciones que se pueden obtener son: promedio de tiempo entre latidos, variabilidad de la frecuencia cardiaca y porcentaje de puntos de inflexión de los intervalos.

Finalmente, para la extracción de métricas de desempeño, en este tipo de investigaciones, se utiliza una tabla como en la Fig. 2, pues permite realizar la comparación de los valores clasificados contra los valores esperados.

Fig. 2. Matriz de confusión

		Predicción	
		Negativo	Positivo
Valor Real	Negativo	VN	FP
	Positivo	FN	VP

Donde VP es: verdadero positivo, FP: falso positivo, FN: falso negativo y VN: verdadero negativo. Por otro lado, se conoce como Precisión Positiva (PP) a la proporción de

casos positivos que fueron identificados correctamente. La sensibilidad, o porcentaje de VP y se calcula con la siguiente ecuación:

$$PP = \frac{VP}{FN + VP}. \quad (3)$$

Mientras que la Precisión Negativa (PN), también conocida como la Especificidad, es la proporción de casos negativos que fueron identificados correctamente, la cual se puede calcular de la siguiente forma:

$$PN = \frac{VN}{VN + FP} \quad (4)$$

IV. JUSTIFICACIÓN

Debido a la relevancia del tema, por su tasa de mortalidad [16], es de suma importancia la detección precisa de las enfermedades del corazón [17] utilizando la técnica de IA más adecuada, para su aporte al estado del arte. El diagnóstico de este tipo de enfermedades, mediante el análisis de electrocardiogramas (ECG), es una herramienta no invasiva, que podría ser fácilmente replicable en Costa Rica y otros países.

A pesar de que muchas personas se realizan frecuentemente este examen, la interpretación de los resultados no es una tarea fácil, incluso para especialistas en cardiología [11]. Esto sumado a la gran cantidad de pacientes que se deben analizar, hacen que se vuelva una tarea bastante compleja de realizar. Por lo tanto, la creación de un sistema que pueda automatizar el proceso y que pueda interpretar correctamente los resultados del electrocardiograma podría significar un gran beneficio tanto para el personal médico como para las personas pacientes.

En este trabajo se examinan dos técnicas de ML cuyo objetivo es diagnosticar enfermedades del corazón de un paciente a partir de los datos obtenidos de electrocardiogramas de diez segundos a una frecuencia de 100 Hz. La primera es una técnica de aprendizaje profundo con Redes Neuronales Recurrentes (RNN) de corto y largo plazo (LSTM) y la segunda es una técnica de clasificación mediante árboles de decisión del tipo XGBoost.

Mencionado lo anterior, el desarrollo de esta investigación se estructurará de la siguiente manera: en la sección V se describe la metodología aplicada, en la sección VI se exponen los resultados en cuanto a la evaluación del sistema, mientras que en la sección VII se mencionan las conclusiones finales de la investigación.

V. METODOLOGÍA

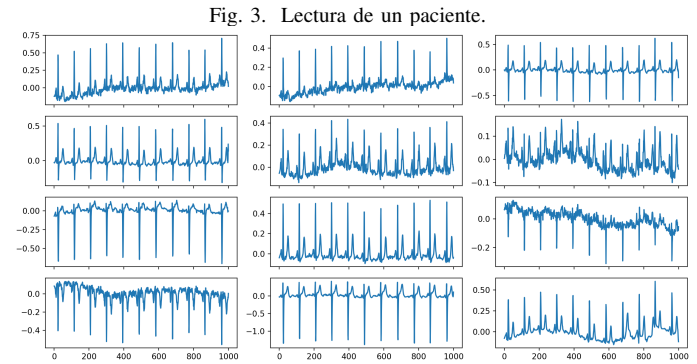
Esta sección describe la metodología seguida para poder determinar las variables, modelos e hiperparámetros que se utilizarán como ajuste dentro de los modelos de Aprendizaje Automático seleccionados para la resolución de la investigación.

A. Datos utilizados

La presente investigación utilizará la base de datos PTB-XL mencionada en los antecedentes, cuyos datos provienen de personas que asistieron a consulta clínica para detección

de anomalías a nivel cardíaco entre octubre de 1989 y junio de 1996. Se utilizó un equipo electrónico de marca Schiller cuyo funcionamiento se basa en la conexión de doce electrodos distribuidos en tres extremidades, seis en el tórax y tres medidas entre cada extremidad (I, II y III) del paciente, permitiendo generar datos que describen una forma de onda por cada una de las doce entradas mencionadas anteriormente. Estas señales serán procesadas por diferentes algoritmos para generar modelos de clasificación utilizando como referencia los meta datos dentro del conjunto de datos por cada paciente.

La base de datos contiene 21,837 registros correspondientes a 18,885 pacientes únicos. Cada registro contiene información sobre la edad, el sexo, el peso y la altura del paciente. Cada registro contiene la ruta del archivo con los resultados del ECG. Cada archivo se compone de otros 1,000 registros. Esto se debe a que el procedimiento dura 10 segundos y el registro se almacena por milisegundo (100 Hz), y además, cada registro contiene doce columnas, dado que son las lecturas de cada electrodo que se le coloca al paciente.



En la Fig. 3 se muestra la lectura de un paciente, la cual incluye los mil registros para cada uno de los doce electrodos, formando así, 12 vistas de los latidos del corazón, durante los mismos 10 segundos, para este paciente.

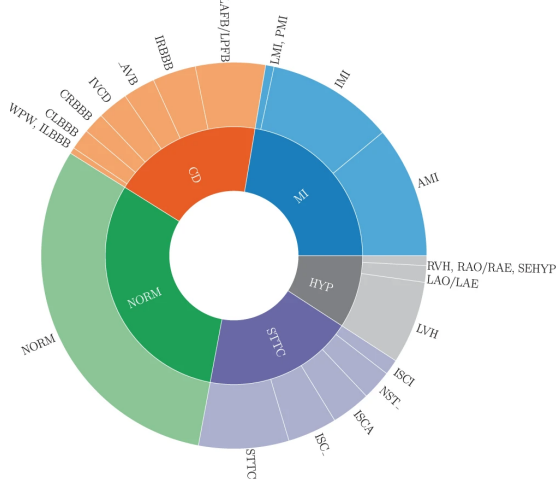
B. Variable a clasificar

Inicialmente las variables a clasificar, corresponden a las etiquetas con el diagnóstico, la cual se compone de 24 categorías, y para reducir las dimensiones del problema, se agrupan en 5 superclases. Luego se harán pruebas y finalmente en el experimento, se reducirán los datos a dos clases (enfermo o sano). Lo anterior mediante una organización jerárquica de la siguiente manera:

- NORM: Indica una lectura del electrocardiograma normal (sano).
- CD: Indica una alteración de la conducción (enfermo).
- MI: Indica una lectura de infarto de miocardio (enfermo).
- HYP: Indica una lectura de hipertrofia (enfermo).
- STTC: Indica cambios ST/T (enfermo).

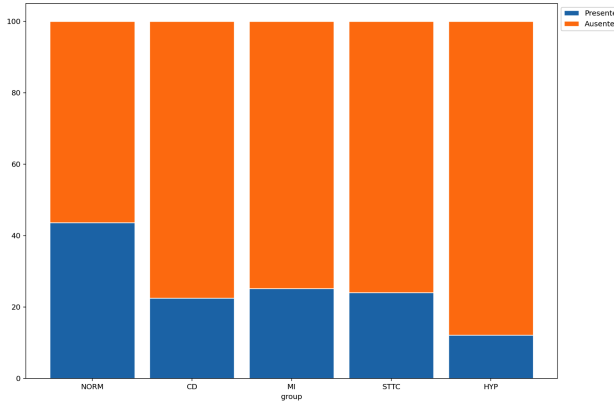
En la Fig. 4 puede observarse un resumen de la transformación de las clases a superclases siguiendo el artículo original donde los autores detallan el proceso realizado [14].

Fig. 4. Resumen de las categorías del diagnóstico.



Si mantenemos las cinco clases, un individuo puede ser diagnosticado con una o varias de estas superclases o categorías (enfermedades), por lo que es un problema de clasificación multi-clase. Además, como se puede notar en la figura 5, el conjunto de datos no se encuentra balanceado (división inferior enfermos, superior sanos), y también hay categorías que contienen más diagnósticos que otras.

Fig. 5. Distribución las categorías del diagnóstico.



Para corregir el problema del desbalanceo en caso de que los modelos utilizados no presenten un desempeño superior al 0.8 de precisión, se considerarán todas las clases que no son normales como "enfermos", y modelar el problema como uno de clasificación binaria, con una variable dicotómica, representando al paciente como sano o enfermo.

Por otro lado para la transformación de la señal del ECG a una tabla tabular se utilizará la librería de *neurokit2* que permitirá extraer más de ochenta y un parámetros sobre el análisis HRV. La extracción de estas características se harán sobre la señal del electrodo LA (brazo izquierdo), puesto que este es el electrodo más representativo, en caso de que se

quiera utilizar el modelo resultante, sobre ECG generados por relojes inteligentes en futuros estudios.

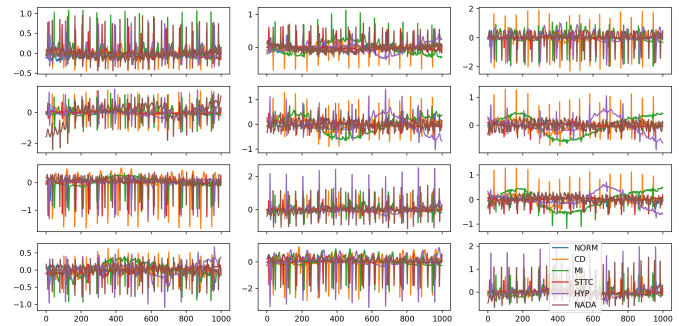
C. Modelos y técnicas a utilizar

Dado que una RNN de memoria a corto plazo (LSTM) trabaja muy bien con series temporales y XGBoost con datos tabulares, para el caso de LSTM se puede aprovechar su capacidad de modelar dependencias a largo plazo y para XGBoost se clasificará utilizando las métricas de HRV.

Para el modelo con LSTM se van a utilizar 2 capas de entrada y 256 ocultas, cuyas salidas se agregan mediante una capa de concatenación como proponen los autores en [14]. También los autores proponen utilizar la activación *tanh* y para la capa de concatenación la activación *sigmoid*. Esto último se debe a que la activación *softmax* aumenta la puntuación de una etiqueta, todas las demás se reducen (es una distribución de probabilidad) y en este caso no se quiere eso dado que se tienen múltiples etiquetas.

Adicionalmente, se aplica una selección de variables escogiendo los electrodos 1, 3, 4, 5, 8 y 9 para realizar otro modelo con las mismas características. Lo anterior debido a que observando algunos individuos con un caso de cada categoría, en los electrodos mencionados se visualiza una mayor separación entre cada categoría como se observa en la Fig. 6, lo que podría significar que tienen mayor poder predictivo. Esto puede ayudar a reducir las dimensiones del problema y obtener un mejor resultado.

Fig. 6. Distribución de las categorías del diagnóstico.



Como tercer y último modelo se propone modificar la probabilidad de corte para ayudar a mejorar la sensibilidad y especificidad de cada una de las categorías. También se espera que pueda ayudar a mejorar la precisión y minimizar el error.

Para el modelo XGBoost, la extracción de característica se hará mediante la librería para Python, *neurokit2*, la cual extrae los valores para las métricas HRV, utilizando al menos 5 métricas, puesto que muchas de las métricas tienen una auto correlación, lo cual se confirmó, al utilizar una matriz de correlación con todas las métricas. La validación cruzada se utilizará con 150 splits y cuatro repeticiones sobre un total de 6,850 observaciones con 5 variables HRV (ECG Rate Mean, SDNN, PIP, GI y SampEn), sexo, edad y etiqueta: sano = 1, enfermo = 0.

D. Hiperparámetros

Los hiperparámetros clave de XGBoost son el número de árboles, la cantidad de grupos y la tasa de aprendizaje según las repeticiones. En LSTM, los hiperparámetros importantes son el número de unidades, capas, tasa de dropout y longitud de secuencia de entrada. Estos hiperparámetros afectan la complejidad y el riesgo de sobreajuste de los modelos. La mejor elección varía según los datos y requiere experimentación.

E. Evaluación y Métricas

Finalmente, para evaluar los modelos se van a utilizar tablas de aprendizaje y prueba con un 80% de observaciones para la tabla de entrenamiento y un 20% para la tabla de pruebas. Como métrica de rendimiento se va a utilizar la precisión (ACC), que en este caso se obtiene promediando el ACC de cada una de las clases a predecir. Esto se debe a que los autores mencionan que

“En nuestro entorno, se prefiere el macro-promedio, ya que esperamos un desequilibrio de clases y no queremos que la puntuación esté dominada por unas pocas clases grandes. Además, la distribución de las patologías en el conjunto de datos no sigue la distribución natural en la población, sino que refleja el proceso de recogida de datos...” [12].

También se incluye la matriz de confusión como métrica, para obtener la precisión positiva y negativa de cada categoría. Para disminuir el sesgo en las pruebas, se utilizará validación cruzada.

VI. RESULTADOS Y DISCUSIÓN

Tabla I
TABLA DE RESULTADOS ACC (MÁXIMOS)

	ACC
XGBoost	0.84
LSTM-PC	0.62
LSTM	0.60
LSTM-SV	0.33

Como se aprecia en la Tabla I, se presentan los resultados de los cuatro modelos aplicados. Los modelos LSTM no logran superar un rendimiento del 0.80, lo que destaca la complejidad del problema al trabajar con series de tiempo. Sin embargo, el modelo XGBoost alcanza un rendimiento aceptable al utilizar datos tabulares, lo que sugiere que merece una mayor exploración en futuras investigaciones.

El modelo propuesto por los autores [12], denominado LSTM, tiene un rendimiento ACC de 0.60, mientras que utilizando selección de variables, denominado de ahora en adelante LSTM-SV, se obtiene un rendimiento ACC de 0.33 y finalmente el rendimiento del modelo con selección de la probabilidad de corte, denominado LSTM-PC, obtuvo un ACC de 0.62. Por tanto, a pesar de que el rendimiento no es el adecuado, se logró demostrar que aplicando las técnicas de probabilidad de corte se puede mejorar el rendimiento de los modelos, también que es importante utilizar todas las variables

a pesar de que visualmente pareciese que algunos electrodos no tienen poder predictivo.

Adicionalmente, la Tabla II muestra la precisión positiva y negativa para cada una de las clases o categorías. Los mejores resultados nuevamente se observan con el modelo LSTM propuesto por los autores [12] y con la selección de la probabilidad de corte. Aunque con la selección de la probabilidad de corte no mejora significativamente la precisión global, sí mejora la especificidad y sensibilidad para cada una de las categorías.

Tabla II
TABLA DE RESULTADOS PRECISIÓN POSITIVA (PP) Y NEGATIVA (PN)

CD		HYP		MI		NORM		STTC	
PP	PN	PP	PN	PP	PN	PP	PN	PP	PN
0.73	0.94	0.44	0.98	0.62	0.93	0.86	0.88	0.67	0.92
0.17	0.97	0.00	1.00	0.00	1.00	0.39	0.80	0.00	1.00
0.75	0.90	0.60	0.77	0.74	0.87	0.86	0.88	0.76	0.85

La categoría con el diagnóstico NORM obtuvo los mejores resultados con una precisión de 0.86 en los casos donde la condición sí estaba presente, y un 0.88 en los casos donde no estaba presente. Para esta no fue necesario aplicar un cambio en la probabilidad de corte debido a que desde un inicio ya tenía una buen especificidad y sensibilidad. Para el caso del diagnóstico CD también se obtuvo resultados aceptables, con un 0.73 de precisión positiva, y un 0.94 de precisión negativa. Con la ayuda de la probabilidad de corte se mejoró la precisión positiva para alcanzar el 0.75 aunque esto disminuyó la precisión negativa a 0.90. En cuanto a la clase STTC, también se logra predecir por lo menos por encima del 50% en ambas categorías con un 0.67 de precisión positiva y un 0.92 de precisión negativa. En este caso también se logró mejorar la precisión de la categoría positiva con la probabilidad de corte alcanzando un 0.76 en esta categoría, pero disminuyendo la precisión de la clase negativa a un 0.85. De forma similar la clase MI obtiene resultados muy parecidos con un 0.62 de precisión positiva y un 0.93 de precisión negativa, que de igual manera con la ayuda de la probabilidad de corte se logra mejorar la precisión positiva hasta alcanzar un 0.74 de precisión a cambio de disminuir la precisión negativa a un 0.87. La última categoría HYP es la que peor resultados a nivel general obtiene. De forma tal que se tiene en la precisión positiva un 0.44, lo cual es incluso peor que un 0.5 (predicción aleatoria), mientras que para la precisión negativa se obtiene una precisión de 0.98 que es la precisión más alta obtenida. Por tanto, con la ayuda nuevamente de la probabilidad de corte se logra alcanzar una precisión positiva de 0.60 con una precisión negativa de 0.77, y aunque la precisión negativa es mucho pero con esto se logra obtener un mejor balance entre la especificidad y la sensibilidad.

Luego del análisis con LSTM utilizando para series de tiempo, se extrajeron las variables HRV. Inicialmente sin los valores de sexo y edad se obtuvieron rendimientos de hasta 0.61; sin embargo, utilizando estas variables dentro del modelo, se lograron máximos de hasta 0.84 utilizando 25,

30, 35, 40 y 45 árboles con combinaciones de validación cruzada de 50, 100, 150, 200 y 250 grupos con 1,2,3,4,5 y 6 repeticiones. Los mejores valores se obtuvieron con 40 árboles, 150 grupos y 4 repeticiones. Cabe destacar que para las redes LSTM como para XGBoost se probaron con Multi-Etiqueta, Multi-Clase y Binario, donde debido a la brevedad de las ondas, los modelos Multi-Etiqueta y Multiclase no superaron el valor de 0.6 para precisión, y para LSTM el modelo de clasificación binario obtuvo los mismos rendimientos que el Multi-Etiqueta.

Cabe resaltar que la precisión positiva, es decir, los casos donde sí esta presente la enfermedad, para cada una de las categorías es donde se obtiene peores resultados. Esto no se debe al desbalance de las clases ya que los casos positivos son menos que los casos negativos como se observa aún cuando los datos se balancearon mediante la salida binaria (sanos o enfermos).

Finalmente los mejores rendimientos para XGBoost sí tienen relación con el balanceo de las clases mediante la salida binaria y la consideración de las variables de sexo y edad fueron determinantes para mejorar el rendimiento del modelo.

VII. CONCLUSIONES

Como los autores mencionan, los electrocardiogramas son de los procesos más comunes hoy en día, y dada la gran cantidad de pacientes que se lo aplican [12], existe un gran potencial para aplicar modelos de predicción automatizada que puedan ayudar tanto a profesionales médicos, para que puedan apoyarse en esta tecnología para su diagnóstico, como a pacientes con el fin de que tengan resultados más acertados y rápidos. Debido a la complejidad de los datos y a que no existe una forma estandarizada para manejarlos, todavía hay mucho por mejorar y por hacer, tal como se pudo observar en los resultados obtenidos. También de esta investigación, podemos decir que la aplicación de técnicas como la selección de corte de probabilidad muestran mejores resultados y pueden ser aprovechadas por futuros investigadores en modelos LSTM.

Aunque inicialmente se supuso que las redes neuronales del tipo LSTM arrojarían mejores resultados que otras técnicas, por su capacidad de procesamiento de series de tiempo, las particularidades de los pulsos de corazón, son sin duda algunos de los principales obstáculos para poder determinar una diagnóstico confiable a partir de estas ECG, sin embargo, la caracterización de la señal, mediante el análisis HRV, resultó más apropiado, que el análisis de la serie de tiempo.

Es así, que se recomienda para futuros trabajos, el procesamiento de señales ECG mediante características de variabilidad del ritmo cardíaco (HRV), agregando variables de normalidad, o demográficas, como la edad y el sexo del paciente, dentro del conjunto de datos, pues esto ayudará al modelo a obtener mejores desempeños, ya que eso fue evidenciado empíricamente, cuando se agregaron tales variables al modelo de árboles de decisión, XGBoost.

REFERENCES

[1] R. Atienza, *Advanced Deep Learning with TensorFlow 2 and Keras*. Packt Publishing Ltd, Feb. 2020.

[2] L. C. N. Barbosa, A. H. J. Moreira, V. Carvalho, J. a. L. Vilaça, and P. Morais, "Biosignal databases for training of artificial intelligent systems," in *Proceedings of the 9th International Conference on Bioinformatics Research and Applications*, ser. ICBRA '22. New York, NY, USA: Association for Computing Machinery, 2023, p. 74–81. [Online]. Available: <https://doi.org.ezproxy.sibdi.ucr.ac.cr/10.1145/3569192.3569218>

[3] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.

[4] A. Burkov, *The Hundred-Page Machine Learning Book*. Andriy Burkov, 2019.

[5] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 785–794. [Online]. Available: <https://doi.org.ezproxy.sibdi.ucr.ac.cr/10.1145/2939672.2939785>

[6] J. Dawani, *Hands-On Mathematics for Deep Learning*. Packt Publishing Ltd, Jun. 2020.

[7] T. Donkers, B. Loepp, and J. Ziegler, "Sequential user-based recurrent neural network recommendations," in *Proceedings of the Eleventh ACM Conference on Recommender Systems*, ser. RecSys '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 152–160. [Online]. Available: <https://doi.org.ezproxy.sibdi.ucr.ac.cr/10.1145/3109859.3109877>

[8] F. Giannakas, C. Troussas, A. Krouska, C. Sgouroulou, and I. Voyiatzis, "Xgboost and deep neural network comparison: The case of teams' performance," in *Intelligent Tutoring Systems*, A. I. Cristea and C. Troussas, Eds. Cham: Springer International Publishing, 2021, pp. 343–349.

[9] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks," in *Proceedings of the 23rd International Conference on Machine Learning*, ser. ICML '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 369–376. [Online]. Available: <https://doi.org.ezproxy.sibdi.ucr.ac.cr/10.1145/1143844.1143891>

[10] B. Ramsundar, P. Eastman, P. Walters, and V. Pande, *Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy, Drug Discovery, and More*. O'Reilly Media, 2019.

[11] A. P. C. Salerno, S. M. and H. S. Waxman, "Competency in interpretation of 12-lead electrocardiograms: a summary and appraisal of published evidence," *Annals of Internal Medicine*, vol. 138, no. 59, pp. 751–760, 2003.

[12] N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek, "Deep learning for ecg analysis: Benchmarks and insights from ptb-xl," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1519–1528, 2021.

[13] S. k. Vangipuram and R. Appusamy, "Machine learning framework for covid-19 diagnosis," in *International Conference on Data Science, E-Learning and Information Systems 2021*, ser. DATA'21. New York, NY, USA: Association for Computing Machinery, 2021, p. 18–25. [Online]. Available: <https://doi.org.ezproxy.sibdi.ucr.ac.cr/10.1145/3460620.3460624>

[14] S. N. B. R. D. K. D. L. F. I. S. W. Wagner, P. and T. Schaeffter, "Pt看xl, a large publicly available electrocardiography dataset. scientific data," *Scientific Data*, vol. 7, no. 1, pp. 1–15, 2020.

[15] P. A. Warrick, V. Lostonlen, M. Eickenberg, J. Andén, and M. N. Homsí, "Arrhythmia classification of 12-lead electrocardiograms by hybrid scattering-lstm networks," in *2020 Computing in Cardiology*, 2020, pp. 1–4.

[16] E. Wilkins, L. Wilson, K. Wickramasinghe, P. Bhatnagar, J. Leal, R. Luengo-Fernandez, R. Burns, M. Rayner, and N. Townsend, *European Cardiovascular Disease Statistics 2017*. Belgium: European Heart Network, Feb. 2017.

[17] —, *European Cardiovascular Disease Statistics 2017*. Belgium: European Heart Network, Feb. 2017.