

Modelos de machine learning para la predicción del estado de salud prenatal y la prevención mediante cardiotocogramas.

William Fabian Arevalo Rodriguez

Ingrid Natalia Jimenez Prieto

José John Fredy González Veloza

Fundación Universitaria Los Libertadores

Fecha 1 de marzo 2023



PALABRAS CLAVE

Salud fetal, cardiotocogramas, prevención, aprendizaje automático, clasificación.

RESUMEN

La reducción de la mortalidad infantil se refleja en varios de los Objetivos de Desarrollo Sostenible de las Naciones Unidas y es un indicador clave del progreso humano. Los cardiotocogramas (CTG) son una opción simple y económicamente accesible para evaluar la salud fetal, lo que permite a los profesionales de la salud tomar medidas para prevenir la mortalidad infantil y materna. El objetivo principal de este trabajo es la implementación y evaluación de diversos modelos de aprendizaje automático con el fin de determinar el mejor en términos de construcción, costo computacional y precisión en la clasificación (diagnóstico) del estado del feto. Debido a ello se plantea el torneo de modelos de aprendizaje de máquina que permitan encontrar equilibrio entre un modelo fácil de replicar y aplicar y una sensibilidad y precisión alta frente a la predicción del estado del feto. Por lo cual se entrenan una lista de diversas técnicas de clasificación supervisada con un conjunto de datos provistos por un plan que impulsa la automatización de análisis automático de CTG. Aquellos modelos que se desempeñaron mejor en la precisión requieren de técnicas de aumento del gradiente en donde se alcanza un valor de alta precisión, dichos modelos revelan que las aceleraciones y la variabilidad anormal en cortos y largos plazos juegan un papel importante al determinar el estado de salud. Entre los modelos comprobados, el Extreme Gradient Boosting presenta los mejores resultados alcanzando una precisión del 96.0% para la categorización del estado del feto.

I. INTRODUCCIÓN

La muerte fetal es una de las tragedias más dolorosas para las futuras madres y sus familias. A pesar de los avances en la atención prenatal, aún no se puede prevenir completamente la muerte de un feto.

El efecto de morbilidad y mortalidad en el feto depende de la clase social y económica de la gestante, según un reporte en donde se informó la disimilitud entre países de altos ingresos con un promedio de 4 por cada 1000 nacidos y países de bajos/medios ingresos con un promedio de 33 por cada 1000 nacidos [1], tan solo en Colombia la mortalidad perinatal y neonatal en el 2019 fue de 15.1 muertes por cada 1000 nacidos [2], debido a estos datos se ha despertado el interés de profesionales en medicina de manera que crean eventos para la recolección de datos, sus análisis y detección.

El monitoreo fetal se ha convertido en una práctica común en la detección temprana de complicaciones durante el embarazo, el uso de pruebas exactas e intervenciones efectivas podrían evitar

que los neonatos con dificultades potenciales puedan ser identificados logrando excelentes resultados en la etapa del embarazo. La cardiotocografía (CTG) o monitorización electrónica de la frecuencia cardíaca fetal (MFE) es una forma de comprobar el bienestar del feto, consiste en el registro electrónico continuo de la frecuencia cardíaca fetal obtenida a través de un transductor de ultrasonido colocado sobre el abdomen materno, acoplando mediciones de presión y transmisión ultrasónica para detectar las aceleraciones y desaceleraciones y así clasificar estos registros como normales, sospechosos o anormales [3]. Sin embargo, la interpretación de los resultados del CTG es subjetiva y depende de la experiencia del observador.

Según un artículo publicado por la Organización Mundial de la Salud (OMS) en 2020, se están implementando varias estrategias para reducir la mortalidad fetal en todo el mundo, entre las medidas que más se destacan se encuentran: la mejora de acceso a atención prenatal lo que implica que las mujeres embarazadas reciban atención médica adecuada y regular

mediante la gestación, fortalecimiento del sistema de la salud esto incluye brindar suficientes profesionales de la salud capacitados para brindar atención prenatal, Reducción de la carga de enfermedades preexistentes es importante identificar y tratar estas condiciones antes y durante el embarazo, mejorar el acceso a tecnología médica como el monitoreo fetal electrónico donde se pueden detectar problemas en el feto antes de que se conviertan en una emergencia y la promoción de prácticas saludables durante el embarazo donde se busca incorporar educación sobre la nutrición adecuada, la actividad física y la reducción del consumo de alcohol y tabaco. [4]

Por otro lado, el desempeño de un modelo de inteligencia artificial es un proceso aparentemente sencillo pero la evaluación sistemática de múltiples modelos con el objetivo de seleccionar el mejor candidato resulta ser una tarea desafiante. Además, en esta labor, se deben tener en cuenta otros factores determinantes, como el contexto del problema que se intenta resolver con los modelos y los recursos disponibles tanto para la construcción como para el despliegue.

Un modelo de inteligencia artificial con cardiogramas puede mejorar la prevención de la mortalidad fetal debido a su capacidad para detectar de manera temprana patrones anormales en la frecuencia cardíaca fetal y en las contracciones uterinas de la madre. Los cardiogramas son una técnica ampliamente utilizada para la monitorización fetal, pero su interpretación depende de la experiencia del médico, lo que puede llevar a una evaluación inexacta y a una toma de decisiones tardía o errónea.

Con el propósito último de obtener un modelo óptimo que resuelva la problemática en cuestión, se consideran no sólo los aspectos de precisión aun siendo un punto importante puesto que acarrearía grandes costos si esta es baja, sino también los relacionados con la gestión eficiente de recursos, dado que la totalidad del proceso se lleva a cabo en la capa gratuita de un servicio en la nube. Una vez asegurado el mejor modelo, se profundiza en la investigación acerca de las propiedades del modelo en aras de resolver el problema específico de categorizar en uno de los múltiples estados la salud fetal, a partir de los valores resultantes del examen.

Con el fin de abarcar un proceso casi completo de construcción de inteligencia artificial, se traza como objetivo principal la implementación y evaluación de diversos modelos de aprendizaje automático con el fin de determinar el mejor en términos de construcción, costo computacional, precisión y sensibilidad en la clasificación (diagnóstico) del estado del feto con un mínimo de 70%, se llevarán a cabo diversas etapas que incluyen una exploración profunda del problema a resolver, definición de la arquitectura del modelo, recolección y el procesamiento de los datos, modelización de las técnicas y la evaluación de los resultados obtenidos.

II. ESTADO DEL ARTE

En virtud de los aspectos previamente expuestos y la relevancia inherente a los mismos, los especialistas en el ámbito de la salud optan por perseguir el progreso mediante diversas tácticas, tales como el empleo de computación e inteligencia artificial, con el propósito de brindar soporte a las decisiones que toman en relación al bienestar fetal. A la fecha, se han diseñado modelos fundamentados en técnicas de aprendizaje automático de competencias, los cuales han demostrado un rendimiento satisfactorio al momento de clasificar las distintas condiciones fetales posibles.

Los torneos de modelos de inteligencia artificial (IA) se han utilizado para abordar una variedad de problemas en la atención médica, incluida la prevención y predicción de mortalidad natal, además son una forma de abordar este problema mediante la competencia de varios equipos que desarrollan algoritmos para predecir el riesgo de mortalidad natal utilizando conjuntos de datos de historias clínicas electrónicas. Los modelos se evalúan en función de su precisión en la predicción del riesgo de mortalidad natal.

Desde una perspectiva alternativa, cabe destacar que el vertiginoso avance de la tecnología ha permitido un creciente poder de procesamiento computacional, con el surgimiento de capas gratuitas de sistemas en la nube que posibilitan la rápida creación de prototipos, ejecución y despliegue de algoritmos cada vez más complejos, como los ensamblados de árboles de decisión. En este contexto particular, el uso de la técnica de potenciación del gradiente representa el estado del arte en cuanto a la detección de patrones en datos tabulares, dado el manejo interno que se brinda a los datos. A través de la creación secuencial de un conjunto de modelos de predicción débiles que son generalizados, se logra una optimización a partir de una función de pérdida [5].

Un ejemplo notable de un torneo de modelos de IA para la mortalidad fetal es el Premio de Innovación del Instituto de Investigación de Perinatología de los Estados Unidos. Este premio, lanzado en 2015, desafió a los equipos a desarrollar modelos de IA para predecir el riesgo de mortalidad fetal utilizando datos de historias clínicas electrónicas. El equipo ganador utilizó una combinación de modelos de aprendizaje profundo y de aprendizaje automático para predecir el riesgo de mortalidad fetal con una precisión del 90%. [6]

Otro ejemplo de un torneo de modelos de IA para la mortalidad natal es el desafío "Data Science Bowl" organizado por Kaggle en 2018. En este desafío, los equipos desarrollaron modelos de IA para predecir el riesgo de mortalidad natal utilizando datos de ultrasonido y resonancia magnética fetal. El equipo ganador utilizó una combinación de técnicas de aprendizaje profundo y aprendizaje automático para predecir la mortalidad natal con una precisión del 92,5%. [7]

III. MATERIALES Y MÉTODOS

Es importante destacar que no solo se busca un modelo que generalice de manera efectiva, sino que también sea fácil y económico en términos de recursos y tiempo, para su replicación e implementación, con la finalidad de alcanzar este propósito, se establece un plan metodológico que consistió en metas específicas, tales como llevar a cabo un análisis exploratorio de datos exhaustivo (EDA) con el fin de comprender algunos de los comportamientos y patrones de los datos; llevar a cabo un pre procesamiento adecuado de los datos para facilitar la ingestión eficiente de los insumos en las técnicas de entrenamiento; y evaluar los diversos modelos propuestos explorando la técnica con el mejor rendimiento.

En la ejecución de esta investigación, se aborda de manera parcial la metodología denominada Metodología Fundamental para la Ciencia de Datos, desarrollada por IBM y basada en la metodología CRISP-DM. Esta técnica comprende un total de diez (10) etapas, sin embargo, en este proyecto se trabajarán las primeras ocho (8) etapas con el propósito de establecer una base estratégica para la realización de proyectos relacionados con la ciencia de datos, independientemente de la tecnología utilizada,

el volumen de datos o el enfoque empresarial en el que se estén desarrollando [8].

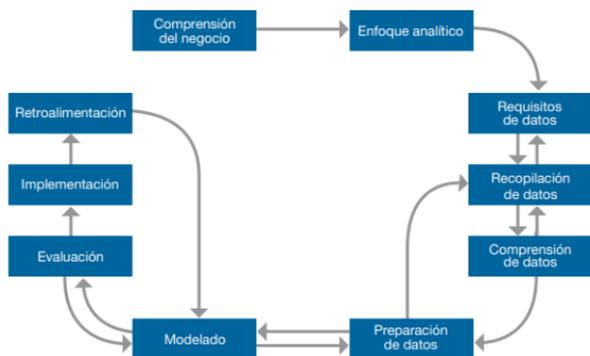


Fig. 1. Paso a paso de Metodología Fundamental para la Ciencia de Datos de IBM.

El proyecto se desarrolla siguiendo las etapas planteadas por la metodología, por lo que a continuación se profundiza en cada una de ellas:

Comprensión del negocio. En primer lugar, resulta fundamental comprender la importancia del problema que se busca abordar, ya que la morbilidad y mortalidad en el feto está estrechamente relacionada con la clase social y económica de la gestante.

Se ha constatado que los principales beneficiarios son los profesionales del área de la salud que habitualmente diagnostican este tipo de patologías, lo que les concede la posibilidad de liberar tiempo y contar con un consenso proveniente de un software que brinda soporte en el diagnóstico, detección y tratamiento en etapas tempranas de la posible enfermedad, lo que permitirá al paciente tomar acciones preventivas. Por ende, el paciente es uno de los beneficiarios con mayor peso, aunque no es el principal. Desde una perspectiva científica, la implementación del modelo que se presenta aquí permitirá una iteración y evolución ágil que admite un entrenamiento veloz y una baja demanda de recursos en esta fase.

De esta manera, los objetivos claves no solo se fundamentan en el rendimiento que pueda ofrecer el modelo seleccionado, sino también en su eficiencia en términos de recursos y tiempo invertido en su entrenamiento. Esto nos permite contar con un modelo altamente balanceado, que no representa una carga significativa en términos de costos durante su entrenamiento, incluso considerando que únicamente se utilizan recursos de capas gratuitas en el proyecto. Esto conlleva a un monitoreo constante durante la comparación de modelos en cuanto a la precisión en la predicción y la tasa de error.

Durante la elaboración del proyecto, se han presentado ciertas limitaciones al utilizar la plataforma de capa gratuita Kaggle. Por lo tanto, es necesario hacer un uso óptimo de los recursos disponibles para encontrar un modelo que se ajuste a las restricciones establecidas.

Enfoque analítico. La cardiocografía (CTG) o monitorización electrónica de la frecuencia cardíaca fetal (MFE) es un procedimiento médico utilizado para evaluar el bienestar del feto. Consiste en el registro electrónico constante de la frecuencia cardíaca fetal, mediante el uso de un transductor de ultrasonido que se coloca sobre el abdomen de la madre. Este proceso combina mediciones de presión y transmisión ultrasónica para detectar cualquier aceleración o desaceleración

en la frecuencia cardíaca fetal y clasificar los registros obtenidos en tres categorías: normales, sospechosos o anormales.

Debido a la necesidad de solucionar el desafío del apalancamiento de un diagnóstico mediante una herramienta de software, que permita lograr un consenso entre el profesional de salud y la máquina, se ha propuesto el uso de un modelo matemático capaz de extraer y comprender las características más relevantes que resulten en una clasificación precisa. Con este fin, se considera apropiado incursionar en el desarrollo de sistemas inteligentes, particularmente en el campo de la inteligencia artificial, en la cual se encuentra el aprendizaje automático, que es capaz de extraer las características o patrones más relevantes.

Requisitos de los datos. Se solicita que los datos obtenidos de los cardiocogramas contengan variables que permitan comprender el estado del feto, además de tener un volumen adecuado para garantizar una muestra representativa que simule las condiciones del mundo real. Por otra parte, resulta imperativo que la información esté debidamente etiquetada con el estado del feto, puesto que se llevará a cabo un enfoque bajo una metodología supervisada.

Se han hallado dos posibles conjuntos de datos obtenidos de diferentes fuentes y que podrían ser utilizados con fines analíticos.

- *CTU-CHB Intrapartum Cardiotocography Database:* Esta recopilación de 552 registros de cardiocografía (CTG) pertenece a la Universidad Técnica Checa (CTU) en Praga y al Hospital Universitario de Brno (UHB). Los registros se seleccionaron de entre un conjunto de 9164 grabaciones recopiladas recientemente en el UHB en la ciudad de Praga.

La prioridad consistió en crear un conjunto homogéneo mediante la inclusión de grabaciones que cumplieron con criterios tales como embarazo único, edad gestacional superior a 36 semanas, ausencia de defectos de desarrollo conocidos, etapa 2 del trabajo de parto de duración no mayor a 30 minutos, señal de FCF de calidad con un porcentaje superior al 50 % en cada ventana de 30 minutos, análisis disponible de parámetros bioquímicos de muestra de sangre arterial umbilical (pH), y predominancia de partos vaginales con escasa presencia de partos por cesárea (46 incluidos únicamente).

Cada grabación, con una duración de hasta 90 minutos, contiene una serie temporal de la frecuencia cardíaca fetal, una señal de contracción uterina y detalles clínicos relacionados con la madre, el parto y el feto, en este caso la variable objetivo es aquella evaluación experta del estado del feto que dieron el consenso de nueve obstetras expertos [9].

- *Sisporto 2.0: A program for automated analysis of cardiotocograms:* Los datos fueron obtenidos mediante la aplicación del programa SisPorto de análisis automatizado de cardiocogramas, el cual se adhiere rigurosamente a las directrices establecidas por la Federación Internacional de Ginecología y Obstetricia (FIGO), procediendo a la evaluación de los trazados tanto durante como después del parto sin llevar a cabo reducción de señal y con la capacidad de registrar simultáneamente gemelos. Este conjunto de datos se compone de 2126 registros caracterizados por la extracción de rasgos de cardiocogramas en la ciudad de Oporto, Portugal, que posteriormente fueron sometidos a la clasificación de tres obstetras expertos en tres categorías: Normal, sospechoso y patológico. [10].

En virtud de que el corpus de información que posibilita una comprensión y un aprendizaje óptimo es aquel que presenta una cantidad adecuada de registros, los cuales han sido verificados por diversos expertos del ámbito de la salud, se toma la decisión de optar por el dataset de Sisporto 2.0: A program for automated analysis of cardiocogram.

Comprensión de los datos. Se da inicio al proceso de familiarización con los datos en la plataforma Kaggle, la cual ofrece un capa gratuita de hardware potente para entrenamiento de modelos bajo el lenguaje de programación Python.

Los recursos proporcionados para cada sesión de edición de Notebook incluyen un tiempo de ejecución de 12 horas para las sesiones de Notebook de CPU y GPU, y 9 horas para las sesiones de Notebook de TPU. También se proporcionan 20 Gigabytes de espacio en disco de guardado automático (/kaggle/working) y un espacio de almacenamiento de borrador adicional (fuera de /kaggle/working) que no se guardará fuera de la sesión actual.

Las especificaciones de la CPU incluyen 4 núcleos de CPU y 30 Gigabytes de RAM. Las especificaciones de la GPU P100 incluyen 1 Nvidia Telsa P100 GPU, 2 núcleos de CPU y 13 Gigabytes de RAM. Las especificaciones de la GPU T4 x2 incluyen 2 Nvidia Telsa T4 GPUs, 2 núcleos de CPU y 13 Gigabytes de RAM. Las especificaciones de la TPU incluyen 4 núcleos de CPU y 16 Gigabytes de RAM. También hay una especificación TPU 1VM que incluye 96 núcleos de CPU y 330 Gigabytes de RAM. [13]

En primera instancia comprender la composición esencial del conjunto. El preprocesamiento previamente ejecutado por el software implica varios pasos entre ellos se destaca la eliminación de picos de la FHR y la detección de contracciones uterinas. Se eliminan los artefactos picos en la FHR que tengan una diferencia entre latidos adyacentes superior a 25 latidos por minuto (lpm) y se sustituye cada señal de contracción uterina por la media de 17 valores centrados en ella para eliminar el ruido de alta frecuencia.

En este caso, se trabaja con 22 variables que recopilan estadísticas relativas a la frecuencia cardiaca fetal, así como histogramas y su movimiento, cada registro representa un paciente.

VARIABLES	
Variable	Descripción
baseline value	Frecuencia cardiaca fetal basal (latidos por minuto)
accelerations	Número de aceleraciones por segundo
fetal_movement	Número de movimientos fetales por segundo
uterine_contractions	Número de contracciones uterinas por segundo
light_decelerations	Número de deceleraciones leves por segundo
severe_decelerations	Número de deceleraciones graves por segundo
prolongued_decelerations	Número de deceleraciones prolongadas por segundo
abnormal_short_term_variability	Porcentaje de tiempo con variabilidad a corto plazo anormal

mean_value_of_short_term_variability	Valor medio de la variabilidad a corto plazo
percentage_of_time_with_abnormal_long_term_variability	Porcentaje de tiempo con variabilidad a largo plazo anormal
mean_value_of_long_term_variability	Valor medio de la variabilidad a largo plazo
histogram_width	Ancho del histograma de la frecuencia cardiaca fetal
histogram_min	Mínimo (frecuencia baja) del histograma de la frecuencia cardiaca fetal
histogram_max	Máximo (frecuencia alta) del histograma de la frecuencia cardiaca fetal
histogram_number_of_peaks	Número de picos del histograma de la frecuencia cardiaca fetal
histogram_number_of_zeroes	Número de ceros del histograma de la frecuencia cardiaca fetal
histogram_mode	Moda del histograma de la frecuencia cardiaca fetal
histogram_mean	Media del histograma de la frecuencia cardiaca fetal
histogram_median	Mediana del histograma de la frecuencia cardiaca fetal
histogram_variance	Varianza del histograma de la frecuencia cardiaca fetal
histogram_tendency	Tendencia del histograma de la frecuencia cardiaca fetal (ascendente, descendente)
fetal_health	Salud fetal, etiquetada 1 (Normal), 2 (Suspect) and 3 (Pathological)

Tabla. 1. Variables del conjunto de datos.

Se procede al análisis del comportamiento de las características exploradas mediante la utilización de gráficos, como lo son la matriz de correlación, gráficos de esparcimiento, barras y cajas de bigotes, en los cuales se observan datos que se consideran saludables, sin que se presenten datos atípicos. Sin embargo, se determina que es necesario llevar a cabo la estandarización de los datos, lo cual se efectúa en fases posteriores del proyecto, además, aunque existe un desbalance en las clases de la variable objetivo no se intenta balancear las clases pero si utilizar una métrica robusta frente a dicho desbalance como lo es el área bajo la curva (AUC).

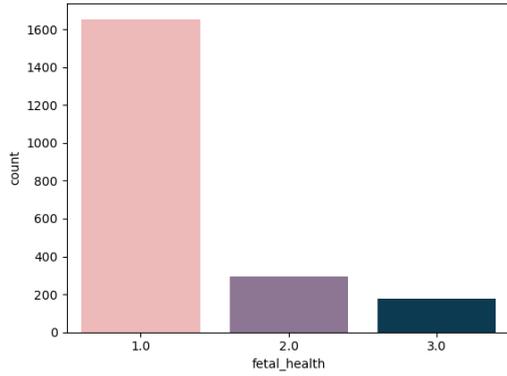


Fig. 2. Cantidad de registros de cada clase en la variable salud fetal.

Preparación de los datos. El primer paso que se lleva a cabo es la transformación de los datos con el fin de estandarizarlos. Se comienza con el escalamiento de todos los datos utilizando una función de escalado estándar ampliamente conocida como Puntaje Z. A continuación, se procede a la realización de dos particiones: una de entrenamiento y otra de validación, las cuales se dividen en un 70% y un 30%, respectivamente. Finalmente, se lleva a cabo la división de las dos particiones en características y en la columna objetivo, en este caso, el estado del feto.

Modelamiento. Este proyecto se centra en la utilización de herramientas basadas en las librerías de Scikit-Learn y XGBoost. Se sostiene el objetivo planteado de fácil implementación, buen desempeño y bajos recursos.

En primer lugar, se plantea un torneo de modelos bajo hiperparámetros estándar, en el cual participan técnicas como regresiones logísticas, máquinas de soporte de vectores, árboles de decisiones, bosques aleatorios, Extreme Gradient Boosting y Hist Gradient Boosting.

Con el fin de proporcionar métricas de rendimiento robustas, se trabaja bajo un método de remuestreo conocido como validación cruzada. Además, se monitorea la métrica AUC SCORE, también conocida como puntaje del área bajo la curva [11].

Tras la obtención de los resultados del campeonato de modelado, se lleva a cabo la selección del modelo que alcance una calificación óptima, para posteriormente explorar y ajustar los hiperparámetros de dicho modelo, comúnmente denominado como proceso de "Fine Tuning" [12].

Evaluación. Se procede a realizar una evaluación del mejor modelo seleccionado utilizando el conjunto de datos de prueba, los cuales no han sido previamente vistos por el modelo. Además, con el objetivo de validar la precisión de las predicciones realizadas por el modelo, se emplea una matriz de confusión, la cual permite visualizar de forma clara y concisa si el modelo ha sido capaz de acertar en las clases correspondientes a las muestras evaluadas. En resumen, se utiliza la matriz de confusión para validar la capacidad de clasificación del modelo y determinar su rendimiento en la tarea a resolver.

Una vez que se han entrenado y evaluado los diferentes modelos, se compilan los resultados en una tabla comparativa que incluye el nombre del modelo y la métrica de precisión utilizada, en este caso el puntaje de Área Bajo la Curva (AUC score). Dicha tabla permite una fácil comparación de los diferentes modelos y la selección del mejor modelo para la tarea de clasificación correspondiente. En resumen, la tabla comparativa de los resultados de los modelos permite seleccionar el modelo con la métrica de precisión más alta, en este caso, el modelo con el AUC score más alto.

RESULTADOS DE LOS MODELOS	
Modelo	AUC Score
Regresión Logística	89.72%
Arbol de decision	91.67%
Bosque aleatorios	93.82%
Máquinas de soporte de vectores	90.66%
HistGradientBoosting	95.03%
Extreme Gradient Boosting	97.73%

Tabla. 2. Precisión de modelos.

La tabla previamente presentada demuestra que el modelo de Extreme Gradient Boosting es el que proporciona un mejor rendimiento en términos de precisión, medido a través del puntaje de Área Bajo la Curva (AUC score). A pesar de que la complejidad del modelo se encuentra entre las más altas de los modelos entrenados, su capacidad para modelar relaciones no lineales y capturar interacciones complejas entre variables lo hace una buena opción para la tarea de clasificación correspondiente. Después de realizar el proceso de ajuste fino de los parámetros del modelo de Extreme Gradient Boosting, se obtiene un mejor rendimiento en términos de precisión, medida a través del puntaje de AUC Score, el cual alcanza un valor de 96.08% en el conjunto de datos de prueba.

	precision	recall	f1-score	support
0	0.97	0.98	0.98	496
1	0.92	0.85	0.89	101
2	0.97	0.95	0.96	41
accuracy			0.96	638
macro avg	0.96	0.93	0.94	638
weighted avg	0.96	0.96	0.96	638

Fig. 3. Reporte de clasificación

El reporte presentado demuestra que el modelo se desempeña muy bien de manera general, aclarando que para la clase 1 (Sospechoso) tiende a perder un poco más de sensibilidad que las otras clases.

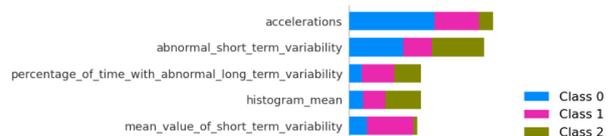


Fig. 4. Variables predictoras con mayor peso

Además como se evidencia en la Figura 4 luego del proceso de ajuste fino de los parámetros del modelo de Extreme Gradient Boosting, también se realiza una validación de los pesos asignados a las variables predictoras en el modelo con el paquete Shap. Como resultado de esta validación, se evidencian las cinco características más importantes en el momento de clasificar las muestras, estas características son aceleraciones, variabilidad anormal a corto plazo, porcentaje de tiempo con variabilidad anormal a largo plazo, media del histograma y el valor medio de la variabilidad a corto plazo.

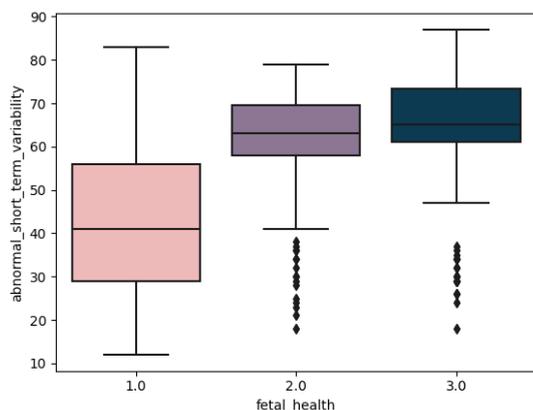


Fig. 5. Diagrama de caja y bigotes para el porcentaje de tiempo con variabilidad a corto plazo anormal en el estado de salud.

Al validar la relación que poseen las variables más relevantes con la variable objetivo podemos crear una hipótesis la cual parte de que aquellos fetos con pocas aceleraciones por minuto, un alto valor en el porcentaje de tiempo con variabilidad a corto o largo plazo anormal pueden categorizarse como un feto con un estado de la salud fuera de lo normal.

VI. DISCUSIÓN

Basándonos en los resultados obtenidos por el mejor modelo en términos de precisión y sensibilidad, se abre una oportunidad para implementar un software que pueda conectarse al sistema desde donde se extrajeron los datos y brindar un apoyo anticipado y preventivo al diagnóstico médico, especialmente para aquellos menos experimentados. El modelo logra una tasa de decisión acertada del 96%, lo que significa que puede ser una herramienta valiosa en países con ingresos bajos donde la disponibilidad de especialistas médicos es limitada. Sin embargo, se debe considerar que en países con altos ingresos, la tasa de acierto del modelo es menor que la de los especialistas, lo que puede limitar su utilidad. Es importante destacar que para implementar este proyecto en países de bajos ingresos, es necesario recolectar muestras bajo el mismo sistema, ya que los registros y comportamientos de los pacientes pueden variar significativamente.

Al permitir la toma de decisiones preventivas por parte del profesional, la tasa de falsos positivos para los estados sospechosos y patológicos puede no ser un problema grave, siempre y cuando las recomendaciones del profesional no tengan un impacto drástico en la salud y desarrollo del paciente y su feto. En el caso más crítico, cuando se diagnostica a un feto con un estado de salud normal cuando en realidad tiene alguna patología, la probabilidad de que esto ocurra es del 4%. La consideración de la decisión proporcionada por el modelo y la

opinión final del profesional dependerá de la posible patología y su gravedad.

Para que este estudio tenga efectividad en Colombia, es necesaria la implementación tanto del software como del hardware desde donde se extrajeron las muestras. Además, se debe realizar un análisis de la población, ya que es posible que los comportamientos difieran de aquellos con los que se realizó este proyecto. Es importante considerar que la variabilidad en las muestras puede afectar la precisión del modelo y, por lo tanto, la capacidad de proporcionar recomendaciones precisas al profesional. Por lo tanto, es necesario validar la precisión del modelo en la población colombiana antes de implementarlo en la práctica clínica.

VII. CONCLUSIONES

Debido a la alta capacidad de generalización y robustez del modelo final implementado utilizando la técnica de Extreme Gradient Boosting, el modelo no presenta sobreajuste hacia ninguna clase en particular del problema. Esto se debe a que el modelo ha sido entrenado en un conjunto de datos de entrenamiento diverso y suficientemente grande, lo que le permite aprender patrones y características generales de los datos. Además, el proceso de ajuste fino de los parámetros del modelo contribuye a evitar el sobreajuste y mejora su capacidad de generalización.

Al evaluar el modelo en un conjunto de datos de prueba independiente, se obtiene un rendimiento preciso y uniforme en todas las clases del problema. Sin embargo, si el modelo estuviera sobre ajustado a alguna clase en particular, podría presentar un rendimiento deficiente en las demás clases del problema, lo que indicaría una falta de generalización y una limitada robustez del modelo. Es importante tener en cuenta que, en el caso de clases desbalanceadas, las métricas de desempeño del modelo pueden ser afectadas negativamente. Para abordar este problema, se pueden aplicar técnicas para manejar el desbalance de clases, como SMOTE, sobremuestreo, submuestreo, o en última instancia, recopilar más datos de las clases desbalanceadas. La implementación de estas técnicas puede mejorar las métricas de desempeño del modelo y aumentar su capacidad para generalizar y hacer recomendaciones precisas.

Además se demuestra que los modelos de aprendizaje automático pueden ser entrenados con éxito con un bajo consumo de recursos, utilizando un entorno de nube que permite un despliegue rápido y efectivo.

Mediante la aplicación de metodologías de modelado de inteligencia artificial de vanguardia, se puede lograr una reducción significativa en la tasa de falsos positivos, lo cual resulta de suma importancia para la implementación efectiva de un modelo en situaciones prácticas.

Este proyecto también destaca la importancia del preprocesamiento de los datos, ya que una correcta limpieza y transformación de los datos puede mejorar significativamente la precisión y eficacia del modelo final. Además, la utilización de herramientas y técnicas de machine learning permite descubrir patrones y relaciones entre los datos que pueden ser difíciles de detectar con enfoques más tradicionales.

REFERENCIAS

- [1] Alfirevic Z, Devane D, Gyte GML, (2018) Cardiotocografía continua (CTG) como forma de monitorización fetal electrónica para la evaluación fetal durante el trabajo de parto, Doi: parto10.1002/14651858.CD006066.pub3
- [2] Ministerio de Salud Colombia, Instituto nacional de salud (2020). Mortalidad perinatal y neonatal y morbilidad materna extrema neonatal. Febrero de 2020. Recuperado 28 de febrero 2023 <https://www.ins.gov.co/Paginas/Inicio.aspx>
- [3] Rodríguez J, Páez J, Correa S, Villamizar M, Prieto S, Cortés J, et al, (Jun, 2022) Análisis de la cardiotocografía anteparto mediante las relaciones S/k y la probabilidad en 20 minutos <https://www.scielo.br/j/rbsmi/a/Ssk5V8qm4y4GxgCBH35KTzS/?format=pdf&lang=es>
- [4] World Health Organization (2020). Stillbirths. Recuperado el 2 de abril de 2023, de <https://www.who.int/health-topics/stillbirth>
- [5] Vera E, (2021) Impacto del registro cardiotocografico en la morbimortalidad neonatal <http://repositorio.ug.edu.ec/bitstream/redug/62376/1/CD%20812->
- [6] National Institutes of Health (2015). Perinatal Data and Biomarker Repository and Perinatal Health and Disease Conditions (R01). Retrieved from Recuperado 28 de marzo 2023 <https://grants.nih.gov/grants/guide/rfa-files/RFA-HD-15-032.html>
- [7] Kaggle. (2018). Data Science Bowl 2018. Recuperado 29 de marzo de 2023 Retrieved from <https://www.kaggle.com/c/data-science-bowl-2018>
- [8] Rollins, J. (2015). Why we need a methodology for data science. Recuperado 10 marzo 2023, de <https://www.ibmdatahub.com/blog/why-we-need-methodology-and-data-science>.
- [9] Universidad Técnica Checa (CTU) en Praga, Hospital Universitario de Brno (UHB), Base de datos de cardiotocografía intraparto CTU-CHB (Feb, 2014) <https://doi.org/10.13026/C22013>
- [10] Campos D, Bernardes J, Garrido A, Marques J, Pereira L (Nov, 2000) Sisporto 2.0: Un programa para el análisis automatizado de cardiotocogramas, [https://doi.org/10.1002/1520-6661\(200009/10\)9:5<311::AID-MFM12>3.0.CO;2-9](https://doi.org/10.1002/1520-6661(200009/10)9:5<311::AID-MFM12>3.0.CO;2-9)
- [11] Cross-Validation : definición e importancia en Machine Learning, Recuperado 1 de abril de 2023 <https://datascientest.com/es/cross-validation-definicion-e-importancia>
- [12] Hinton G, Vinyals O, Dean J. (9 Marzo 2015).Distilling the Knowledge in a Neural Network. Cornell University. <https://arxiv.org/abs/1503.02531>
- [13] Notebooks Documentation,” Kaggle. Recuperado 3 de Abril 2023 <https://www.kaggle.com/docs/notebooks>