



## Aplicación médica para la predicción clínica del índice de mortalidad en pacientes con trauma craneoencefálico

### Medical application for clinical prediction of mortality rate in patients with traumatic brain injury

 Kerry Kenton Kelly Kandler <sup>1 2</sup>  
[kerry.kelly@bicu.edu.ni](mailto:kerry.kelly@bicu.edu.ni)

 Orvin Daniel Reyes Reyes<sup>1</sup>  
[orvin.reyes@es.bicu.edu.ni](mailto:orvin.reyes@es.bicu.edu.ni)

 Deyvon Kestner Ordoñez Cuthbert <sup>1</sup>  
[deyvon.ordonez@bicu.edu.ni](mailto:deyvon.ordonez@bicu.edu.ni)

 Dexon-Mckensy Sambola <sup>1 2</sup>  
[dexon.sambola@bicu.edu.ni](mailto:dexon.sambola@bicu.edu.ni) / [desambola@outlook.com](mailto:desambola@outlook.com)

*Fecha de Recepción: 14-06-2022*

*Fecha de Aprobación: 15-07-2022*

#### RESUMEN

Uno de los campos con mayor crecimiento en los últimos años ha sido la inteligencia artificial, la cual se divide en subcampos como el aprendizaje automático, el cual provee técnicas y algoritmos para que los sistemas puedan aprender y mejorar de forma automática. El objetivo del proyecto fue desarrollar una aplicación médica cuya función radica en optimizar y facilitar la toma de decisiones referente a la atención brindada a pacientes de trauma craneoencefálico, por medio de un modelo predictivo que indique la probabilidad de muerte de estos pacientes. Se aplicó la metodología de programación extrema (XP), empleando técnicas de aprendizaje automático tomando como base el conjunto de datos CRASH-2, la cuál cuenta con 20,207 registros de pacientes aleatorizados de trauma craneoencefálico. Las redes neuronales artificiales fueron utilizadas para construir el modelo predictivo de supervivencia. La red neuronal con la arquitectura 6-(8-14-2)-1 alcanzó una exactitud del 76%, sensibilidad del 72.9% y especificidad del 94.1% sobre el conjunto de datos de prueba; demostrando una capacidad de discriminación prometedora con buena adaptación a la validación interna.

**Palabras claves:** Redes Neuronales; Aprendizaje Automático; Modelo Predictivo; Inteligencia Artificial; Medicina.

<sup>1</sup> Bluefields Indian & Caribbean University, Escuela de Informática, Bluefields, Nicaragua.

<sup>2</sup> Grupo de investigación de Inteligencia Artificial & Aplicaciones Médicas (GIAAM).



## ABSTRACT

One of the fastest growing fields in recent years has been artificial intelligence, which is divided into subfields such as machine learning, which provides techniques and algorithms so that systems can learn and improve automatically. The objective of the project was to develop a medical application whose function is to optimize and facilitate decision-making regarding the care provided to patients with head trauma, through a predictive model that indicates the probability of death of these patients. The extreme programming (XP) methodology was applied, using machine learning techniques based on the CRASH-2 data set, which has 20,207 records of randomized traumatic brain injury patients. Artificial neural networks were used to build the predictive survival model. The neural network with the 6-(8-14-2)-1 architecture achieved an accuracy of 76%, sensitivity of 72.9%, and specificity of 94.1% on the test data set; demonstrating a promising discrimination capacity with good adaptation to internal validation.

**Keywords:** Neural Networks; Machine Learning; Predictive Model; Artificial Intelligence, Medicine.

## BILA PRHNI RA

Las mani bilara langnika kau tara pali ba sika sins laka upla bui paskan ba, baku natkara baikanka langnika sirpi ra baiki sakisa baku ban diara kum lan taki ba baku, kan witin mihta sins laka natka nani yabisa baku sin yabal sirpi nani trabilka ba wapni mangki waia baku lika tanka sut ba sip kabia lan taki bara kau yamni daukaia silp natkara. Naha wark daukaia lukanka bapanka ba sika yarka warkka ba kau pawu waia kan ai warkka plamaika ba kau yamni daukaia bara yarka sip kabia aitani natka kat brin kaia kaikanka yamnira ai uplika nani ai lalra siknis bri nanira, bankra daukanka kum wina kaina manas ra kaiki ba wina baku natkara marikaia pruaia sip ba naha sat siknis ka bri uplika nani ba. Nahara alki daukan natka kum kau laiuhra brih waia ba XP wi ya ba, baku alki dauki lan takaia natka nani ba ul asla plamaika nani wahbi kira ba wal CRASH-2 wi ba, naha ba kulkisa 20,207 siknis uplika nani dakbi, dakbi ra ulbansa lal trabilka sikniska bri nani dukiara. Sins laka tanka pliki upla bui paskan ba yus munisa kaina manas paskanka ba laki kaiki tankira upla ba kau raya iwi kaia dukiara. Naha Sins laka pliki paskanka ba 6- (8-14-2)-1 alkan sa aitani kat 76% dahra walanka kum 72% bara tanka marikanka kum 94.1% purkara wal asla praki tanka sakan ba; mariki luki sakaia tankira bila kat kulkanka yan kaia mata.

**Bila bak sakan nani:** Sins laka pliki paskan ka, bak saki lan takanka, Kaina manas wina dia takaia ba aisanka marikanka, Sins laka upla bui paskan dukia, Sika.

**Para citar este artículo (APA):** Kelly K., Kerry K., Reyes R., Orvin D., Ordoñez C., Deyvon K. y Sambola, Dexon-Mckensy (2022). Aplicación médica para la predicción clínica del índice de mortalidad en pacientes con trauma craneoencefálico. *Wani* (77),4-17. <https://doi.org/10.5377/wani.v38i77.14659>



## INTRODUCCIÓN

Los traumatismos se consideran un conjunto de lesiones provocadas por violencias externas al organismo, pueden variar desde una lesión única no complicada hasta lesiones múltiples en extremo complejas (Montero, 2012). Según OPS/OMS (s. f.) cada día en el mundo, cerca de 16,000 personas mueren a causa de traumatismos, estos representan 12% de la carga mundial de morbilidad, la tercera causa más importante de mortalidad general y la principal causa de muerte en el grupo de edades de 1 a 40 años. Aun así, en la mayoría de los países el traumatismo es típicamente considerada como accidente, con poco esfuerzo comprometido en investigación para el estudio y la reducción de esta enfermedad (Montero, 2012). Esto presenta un gran problema al cuerpo médico y comunidad sanitaria, quienes deben buscar métodos o alternativas a esta situación que afecta la vida de millones de personas alrededor del mundo. Especialmente en lugares que no cuentan con la instrumentación necesaria para la realización de la Tomografía Axial Computarizada (TAC), cuya falta dificulta el proceso de la correcta diagnosticación de este tipo de pacientes.

El correcto diagnóstico del estado de pacientes con Traumas Craneoencefálico (TCE) se ha convertido en una problemática de gran impacto en la comunidad médica, debido a la incertidumbre de este proceso al utilizar la Escala de Coma de Glasgow (ECG) como única herramienta para determinar el estado de un paciente de trauma, ya que el puntaje se basa en la habilidad del observador, es decir, dos personas pueden puntuar la evaluación de un paciente de manera diferente, especialmente si no están entrenadas. Esto sugiere una problemática, por lo cual se debe abordar este inconveniente, con la idea de brindar soluciones que propicien resultados más objetivos y certeros a la hora de llevar a cabo la evaluación de pacientes de trauma. El proyecto se centra en la búsqueda de soluciones disruptivas capaces de contribuir positivamente en el desarrollo del campo médico; para ello, se inclina a la aplicación de la Inteligencia Artificial (IA) en la medicina. En esta ocasión se pretende atender la problemática del TCE, mediante tecnologías de AA y Redes Neuronales (RN); por lo que, la finalidad del proyecto fue desarrollar una aplicación médica que permita estimar la probabilidad de riesgo de mortalidad en pacientes con TCE a partir de parámetros clínicos con el fin de proporcionar una mejor atención.

La IA es una disciplina que comprende un espectro amplio de funcionalidades, entre las que se destaca las predicciones de probabilidad de riesgo de mortalidad en pacientes. Los objetivos principales de la IA en el campo médico abarcan desde el auxilio al médico, estimar la probabilidad de riesgo, la optimización de la imagen, reconocimiento de estructuras, segmentación de lesiones y transcripción del informe (Aguirre et al., 2021). Según un informe de CB Insights, el 86% de las organizaciones proveedoras de asistencia utilizan IA, se espera que para 2025 haya sistemas de IA que puedan responder a cuestiones concretas de los pacientes y facilitar la gestión de la salud en la población (APD, 2021). Para ello, se utilizan técnicas como AA y Aprendizaje Profundo (AP). Estos son campos de las ciencias computacionales y la IA, la cual refiere a la ciencia e ingeniería por medio del cual las máquinas pueden analizar datos y aprender basada en esos datos.

La aplicabilidad de estas tecnologías en el campo médico ha ido en ascenso en los últimos años, en Houston, Texas a través del Instituto de investigación quirúrgica del ejército de los Estados Unidos, Liu & Salinas (2017) relizaron un estudio con el fin de identificar las aproximaciones del

AA para predecir los resultados de trauma, con la hipótesis de que los modelos similares para la predicción de resultados pueden compartir características comunes, el estudio demostró la eficacia de AA para predecir resultados en pacientes con trauma y mostraron que se puede determinar una base común de características o variables de AA para predecir los resultados del trauma. En esta misma línea, Raj et al. (2019) apuntaron al desarrollo de un algoritmo dinámico totalmente automatizado y objetivo basado en presión intracraneal, presión arterial media y presión de perfusión cerebral para predecir la mortalidad a 30 días.

El uso de algoritmos de aprendizaje automático para el diagnóstico de enfermedades puede reducir los tiempos de consulta e incluso evitar la consulta con el especialista. Son un área del conocimiento donde convergen diferentes técnicas, en particular las RN, los Árboles de Decisión (AD) y las redes bayesianas teniendo un gran desarrollo e impacto en la medicina (Montero Rodríguez et al., 2019). El modelo para la predicción de supervivencia en pacientes con trauma basado AA presentado por Rau et al. (2019), demostró que tanto los modelos de regresión lineal, máquina de vectores de soporte, RN y puntaje de gravedad de lesiones y traumas exhibieron una alta precisión similar en la predicción de la supervivencia de los pacientes con trauma. Mientras que, en el conjunto de datos de prueba, el modelo de RN tenía la máxima precisión equilibrada y especificidad predictiva.

## MATERIALES Y MÉTODOS

El proyecto se centra en la investigación aplicada de tipo descriptiva, se llevó a cabo una observación no participativa, sirvió para valorar las técnicas de IA utilizadas en la solución de problemas similares a la planteada en esta investigación. Posteriormente se procedió a la recolección de datos médicos referentes a pacientes con TCE, enfatizando en obtener la mayor cantidad de datos posibles para la creación del modelo predictivo. Este proceso fue llevado a cabo a través de la búsqueda en línea en el repositorio freeBird, cuyos conjuntos de datos de ensayos clínicos están disponibles de forma gratuita con el objetivo de facilitar el proceso de investigación y desarrollo para mejorar la atención a pacientes y promover el conocimiento médico.

Con la finalidad de cumplir con esos criterios, se seleccionó el conjunto de datos denominado CRASH-2 (Clinical Randomization of an Antifibrinolytic in Significant Hemorrhage), esta recoge datos de 20,207 pacientes aleatorizados de TCE producto de accidentes de tráfico, de 274 hospitales de 40 países que estaban dentro de las 8 horas posteriores a la lesión (Roberts et al., 2013).

### *Metodología de desarrollo*

El desarrollo del modelo se centró en la metodología XP, permitió un desarrollo constante, ordenado y adecuado al contar con interacciones continuas con expertos en la materia. El primer paso en el desarrollo del modelo consistió en establecer una comunicación directa con personas del sector salud que interactúan directa e indirectamente con la problemática, se definieron las historias de usuarios para describir la funcionalidad del modelo. Las historias de usuario describen la funcionalidad del software desde el punto de vista del usuario (Cohn, 2004). Se crearon las tarjetas CRC (Clase, Responsabilidad y Colaboración) por cada historia de usuario, estas proporcionan una funcionalidad directa al modelo. Las tarjetas CRC Representan explícitamente varios objetos simultáneamente (Beck & Cunningham, 1989).

Se dividió en procesos de forma secuencial. El primer proceso lógico estuvo constituido por la limpieza y depuración del conjunto de datos seleccionado para el entrenamiento del modelo predictivo. Según Ilyas & Chu (2019), la limpieza de datos se utiliza para referirse a todo tipo de tareas y actividades para detectar y reparar errores en los datos. En esta ocasión, no se identificaron predictores de varianza cero. Luego, se analizó las columnas que contenían pocos valores únicos, identificando columnas con únicamente 2 valores únicos. Podemos referirnos a estas columnas como predictores de varianza casi-cero, ya que su varianza no es cero, sino un número muy pequeño cercano a este. Kuhn & Johnson (2019) menciona «aunque los predictores de varianza casi-cero probablemente contienen poca información predictiva valiosa, es posible que no deseemos filtrarlos».

Seguidamente se analizaron las 44 columnas presentes en el conjunto de datos, con el objetivo de identificar aquellas que no se relacionaban directamente con la predicción de mortalidad a causa de TCE, identificando 14 columnas que no reflejaban información relevante para el modelado. De esta manera, el conjunto de datos se vio reducido a 30 características a proseguir en el proceso de depuración y limpieza de datos. También se logró identificar valores nulos en 28 de las 30 columnas del conjunto de datos.

El método de eliminación de filas se usa comúnmente para manejar los valores nulos. Se elimina una fila particular si tiene un valor nulo para una característica o columna particular o en caso de que una columna contenga un 70% o más de valores nulos, este es el que se elimina (Maladkar, 2018). Ninguna columna contuvo más del 70% de valores nulos, se procedió a eliminar las filas que contenían valores nulos, los cuales sumaban el total de 3,643. Por ende, el conjunto de datos que inicialmente contenía 20,207 registros quedó reducido a 16,564, los cuales representan una cantidad considerable para seguir adelante con el modelado.

Dado que las variables de entrada en su totalidad son numéricas, se utilizó el método de análisis de varianza (ANOVA) con un nivel de confianza del 95 %, tomando en cuenta 29 factores, el cuál es una prueba de hipótesis estadística paramétrica para determinar si las medias de las muestras de datos provienen de la misma distribución o no (Broenlee, 2020). Se utilizó la librería Scikit-learn del lenguaje de Python, que proporciona una implementación de la prueba F de ANOVA en la función `f_classif()`. Se llevó a cabo una prueba F de ANOVA de una vía para cada predictor continuo, el cuál prueba si todas las diferentes clases de  $Y$  tienen la misma media que  $X$ . Teniendo en cuenta que:

$N_j$  = El número de casos con  $Y = j$

$\bar{X}_j$  = La media muestral del predictor  $X$  para la clase objetivo  $Y = j$

$S_{j^2}$  = La varianza muestral del predictor  $X$  para la clase objetivo  $Y = j$

$$S_{j^2} = \sum_{i=1}^{N_j} \frac{(X_{ij} - \bar{X}_j)^2}{(N_j - 1)}$$

$\bar{X}$ : La gran media del predictor X

$$\bar{X} = \sum_{j=1}^J \frac{N_j \bar{X}_j}{N}$$

Las notaciones anteriores están basadas en pares completos de  $(X, Y)$ ; entonces, el valor p basado en el estadístico F es calculado por el valor  $p = Prob \{F (J-1, N-J) > F$  en donde,

$$F = \frac{\sum_{j=1}^J N_j (\bar{X}_j - \bar{X})^2 / (J - 1)}{\sum_{j=1}^J (N_j - 1) S_j^2 / (N - 1)}$$

F (J-1, N-1) es una variable aleatoria que sigue una distribución con grado de libertad J-1 y N-J, si el denominador de un predictor es cero, establece el valor  $p = 0$  para el predictor. Se pudo probar la relación entre el predictor categórico y la respuesta continua, logrando comprobar si existe la misma varianza entre los grupos de características categóricas con respuesta continua; Teniendo en cuenta que, si hay una varianza igual entre los grupos, indica que esta característica no tiene impacto en la respuesta o variable de salida y no se considera para el entrenamiento del modelo. Se seleccionó diez variables (iage, isbp, irr, icc, ihr, igcseye, igcsmotor, igcsverbal, igcs y bheadinj). Seguidamente se procedió a identificar la existencia de multicolinealidad entre estas variables. La multicolinealidad ocurre cuando las variables independientes están correlacionadas, si el grado de correlación entre las variables es lo suficientemente alto, puede causar problemas al ajustar el modelo e interpretar los resultados (Frost, 2021).

Mediante la función *corr()* de la librería Pandas del lenguaje de programación Python, se procedió a calcular el coeficiente de correlación de Pearson, en base a la siguiente ecuación,

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{[\sum (x - \bar{x})^2 (y - \bar{y})^2]}}$$

Donde  $x$  e  $y$  representan a cada una de las variables, de forma cíclica hasta que cada variable haya sido evaluada con las otras del conjunto. De esta manera se obtuvo el coeficiente de correlación entre todas las variables restantes, estas se reflejan en la figura 1.

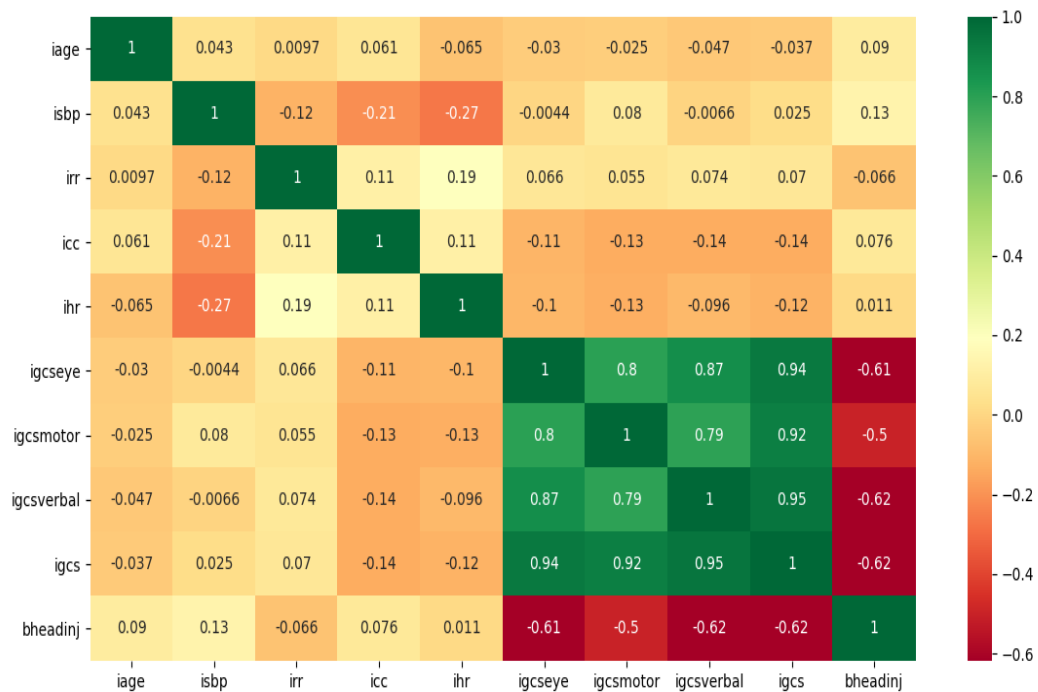


Figura 1. Mapa de calor en base a los coeficientes de correlación de las variables

Las variables que presentaban una correlación positiva (negativa) basándose en la regla empírica para la interpretación de los coeficientes de correlación (tabla 1), fueron las variables igcseye, igcsmotor, igcsverbal y igcs. Sin embargo, solo se necesrto una para incluir al modelo, se uso la variable igcs, dado que el valor de esta variable está basado en la sumatoria de las demás variables.

Tabla 1. Regla empírica para la interpretación del tamaño de coeficiente de correlación

Tamaño de la correlación	de la Interpretación	Inserte lo valores de significancia
.90 a 1.00 (-.90 a -1.00)	Correlación positiva (negativa) muy alta	Altamente significativo
.70 a .90 (-.70 a -.90)	Correlación positiva (negativa) alta	Muy significativo
.50 a .70 (-.50 a -.70)	Correlación positiva (negativa) moderada	Significativo
.30 a .50 (-.30 a -.50)	Correlación positiva (negativa) baja	Ligeramente significativo

.00 a .30 (.00 a -.30)

Correlación insignificante

Levemente  
significativo

---

Fuente: adaptado de (Hinkle et al., 2003)

La aproximación tomada para el modelado de las redes neuronales giró en torno al reconocimiento de patrones y el procesamiento de señales a partir de la información recolectada y depurada del conjunto de datos. Los datos para prueba y validación constataron el 26% y 20% del total del conjunto de datos, y los datos de entrenamiento el 54%. La estandarización se llevó a cabo utilizando la siguiente fórmula,

$$x = \frac{x - X_{min}}{X_{max} - X_{min}}$$

Donde los valores *min* y *max* de *X* representan los valores mínimo y máximo en el conjunto de datos respectivamente y la *x* representa un dato particular en el conjunto de datos. Para definir la arquitectura de la RN, se descompuso el problema al nivel más básico para una mejor comprensión de la misma. Al tratarse de una clasificación binaria, la aproximación adoptada fue la del aprendizaje supervisado, en la cual los algoritmos trabajan con datos ‘etiquetados’, intentando encontrar una función que, dadas las variables de entrada, les asigne la etiqueta de salida adecuada. El aprendizaje supervisado incluye varias técnicas, en dependencia del problema que se pretende resolver. En este caso, se utilizó la regresión logística. Salas Velasco (1996) afirma que es un algoritmo de clasificación empleado para predecir el resultado de una variable categórica binaria en función de las variables independientes.

Se definió al modelo como un Modelo de Perceptrón Multicapa (MLP), el cual está compuesto por capas de nodos, en el cual cada nodo está conectado a todas las salidas de la capa anterior, y la salida de cada nodo está conectada a todos los nodos de la siguiente capa. Todo el proceso de aprendizaje se encapsula en la arquitectura de RN, la cual luego de diferentes pruebas controladas, fue definida como una red 6-(8-14-2)-1 que se representa en la figura 2. Esto implica que la red cuenta con tres capas ocultas de ocho, catorce y dos neuronas respectivamente, así como una capa de entrada de seis variables, las cuales son las características previamente identificadas con mayor impacto en el resultado de pacientes con TCE, y una capa de salida con una sola variable.



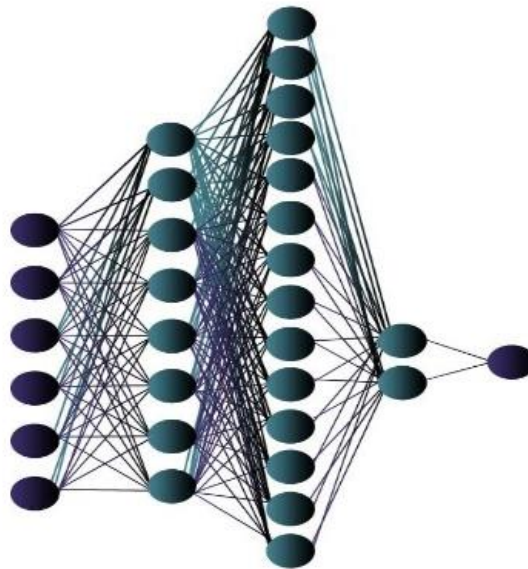


Figura 2. Arquitectura RN modelo predictivo

Se usó la función de activación de unidad lineal rectificada (ReLU) en las primeras cuatro capas, y la función sigmoide en la capa de salida. Para entrenar el modelo se definió una función de pérdida para evaluar el conjunto de pesos y un algoritmo de optimización para encargarse de la búsqueda de diferentes pesos para la red. Estos pesos son parámetros dentro de la red neuronal que se encuentran definidos dentro de las capas ocultas, los cuáles transforman los datos de entrada antes de retornarlos. En este caso, dado que se trata de una clasificación binaria, se seleccionó la entropía cruzada como función de pérdida, ya que matemáticamente, es la función de pérdida preferida en el marco de inferencia de máxima verosimilitud.

En esa medida, el optimizador empleado para el entrenamiento del modelo fue el algoritmo descenso de gradiente estocástico eficiente. El optimizador fue configurado con una tasa de aprendizaje de 0.01. Este valor determina la rapidez con la que la RN actualiza los conceptos que ha aprendido, lo que implica que, en el proceso de entrenamiento, fue necesario una mayor cantidad de iteraciones para reflejar los cambios en los pesos de cada una de las neuronas. Se realizaron pruebas controladas en dónde se experimentaron diferentes valores de *epoch* (100, 200, 300, 400 y 500) y *batch* (2, 4, 8, 16, 32, 64), en dónde la configuración de un *epoch* de 100 y un *batch* de 32 predominaron sobre los demás logrando los mejores resultados, quedando desarrollado de esta manera el modelo predictivo.

Posteriormente se procedió a la implementación del modelo predictivo en una aplicación web. Para ello se utilizó el *framework* Django del lenguaje de programación Python, el cual permitió generar una interfaz gráfica para la utilización del modelo predictivo. Se hizo uso de las tarjetas CRC para desarrollar cada uno de los apartados de la aplicación web, con la finalidad de abordar las necesidades previamente identificadas en el proceso de planificación. Por lo que fue necesario la elaboración de una base de datos, en la cual se almacenaron los datos obtenidos de cada predicción que se realice a través del modelo, así como información relevante para futuros estudios. Para las

pruebas se seleccionaron 100 registros reales del conjunto de datos, de forma aleatoria, previamente separados con esta finalidad.

## RESULTADOS Y DISCUSIÓN

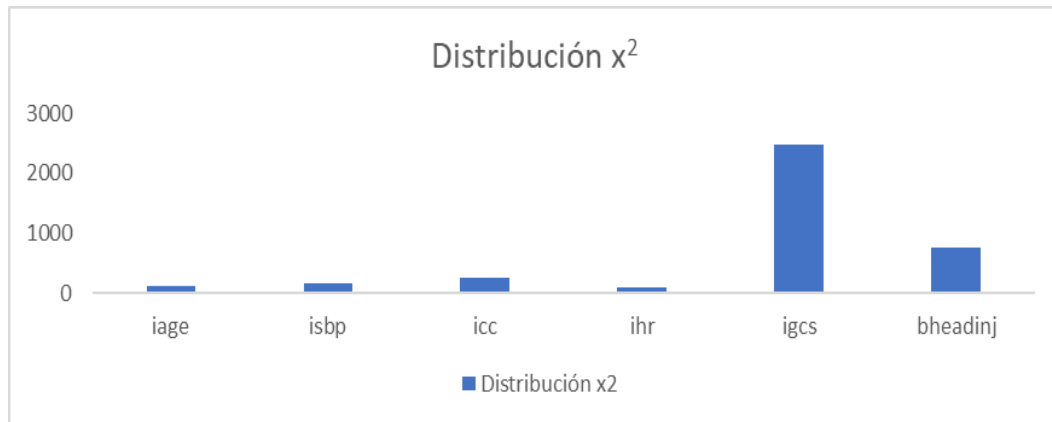
Se logró una exactitud de 76% que refleja la cantidad de predicciones realizadas, sobre el conjunto de datos de prueba, que fueron correctas. Por otro lado, se logró una sensibilidad de 72.9%, esta probabilidad indica la proporción de casos positivos que fueron correctamente identificados por el modelo, en este caso la capacidad que tiene el modelo de detectar correctamente las probabilidades positivas de muerte entre los pacientes de TCE. El modelo logró una especificidad de 94.1%, lo cual refleja la probabilidad de los casos negativos que el modelo ha logrado clasificar correctamente. En este caso, la capacidad del modelo de identificar a los pacientes que sobrevivieron con una baja probabilidad de muerte. Así mismo, el modelo fue capaz de lograr un 27.1% de probabilidad de tasa de error, en el caso de falsos negativos.

**Tabla 2. Métricas de evaluación del modelo predictivo**

Métrica	%
Exactitud	76
Sensibilidad	72.9
Especificidad	94.1
Tasa de falsos negativos	27.1

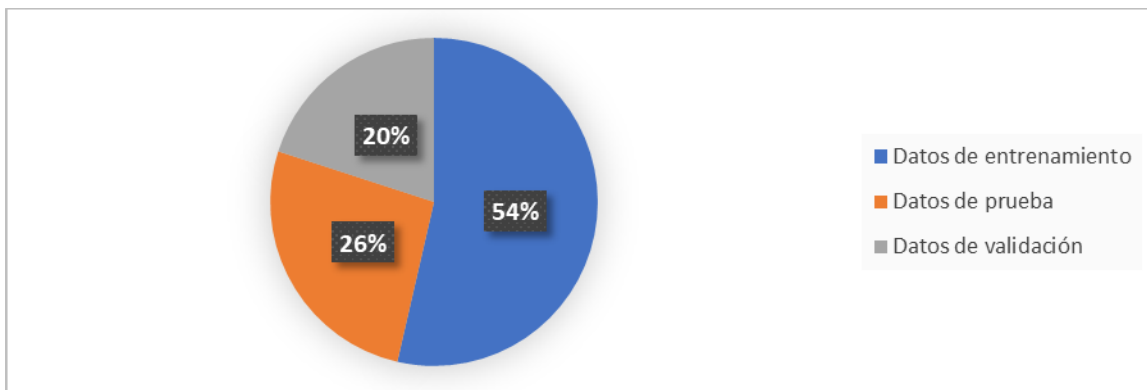
También fue capaz de brindar información fiable referente al estado de los pacientes de TCE en tiempo real con datos disponibles al momento. Por medio de las historias de usuario, se lograron identificar 10 requerimientos funcionales para la aplicación, así como 3 requerimientos no funcionales, los cuales sirvieron de base para el desarrollo del proyecto. Se recolectaron 20,207 registros con 44 variables de pacientes de 274 hospitales en 40 países, en donde fueron incluidos aquellos pacientes adultos (18 años o más) con trauma, presentes dentro de las ocho horas posterior al incidente causante de la lesión, que presentaban hemorragia significativa o se consideraba algún riesgo de este.

Las variables edad (iage), presión arterial sistólica (isbp), tiempo de recarga capilar central (icc), frecuencia cardiaca por minuto (ihr), puntaje de coma de Glasgow (igcs), y la presencia de lesión significativa en la cabeza (bheadinj) demostraron ser las características con mayor impacto para el desarrollo del modelo. El puntaje de ECG demostró ser la característica más significativa y con mayor impacto en la predicción de mortalidad en pacientes de trauma craneoencefálico, mostrando una distribución lineal referente a este.



**Figura 3.** Variables del conjunto de datos con mayor valor predictivo sobre la probabilidad de muerte de paciente

El 54% del conjunto de datos para el entrenamiento del modelo estuvo compuesto por 8,878 registros. Los datos de prueba estuvieron conformados por 4,372 registros que representan el 26% del total y los datos para validación fueron 3,314 registros equivalentes al 20% respectivamente.



**Figura 4.** División del conjunto de datos depurado

La arquitectura 6-(8-14-2)-1 presentó el mejor rendimiento entre las pruebas controladas realizadas. El modelo presenta una prometedora capacidad de discriminación y una buena adaptación a la validación interna, es de destacar el gran tamaño de la muestra en relación con el número de variables de pronóstico examinadas. Las fortalezas del estudio radican en el uso de un conjunto de pacientes bien descrita, la recopilación de datos prospectivos y estandarizados sobre los factores de pronóstico, la facilidad de seguimiento y el uso de una medida de resultado validada en un punto fijo en el tiempo después de la lesión. Sin embargo, el proyecto también presenta algunas limitantes. Los datos a partir del cual se desarrolló el modelo provienen de un ensayo clínico y, por lo tanto, puede limitar la validez externa. Por ejemplo, los pacientes fueron reclutados dentro de las ocho horas posteriores a la lesión y no es posible estimar la precisión del modelo para los pacientes evaluados más allá de ese periodo. No obstante, el ensayo pragmático del que se

obtuvieron los datos no requirió pruebas adicionales y, por lo tanto, incluyó una diversidad de pacientes.

El proceso de depuración de datos y la selección del origen de datos jugaron un papel muy importante en el desempeño final del modelo predictivo; el seleccionar únicamente seis variables de conjunto original de datos, incrementó en gran medida la capacidad de predecir acertadamente la probabilidad de muerte de pacientes de trauma craneoencefálico en base al conjunto de datos trabajado, dado la cantidad de datos faltantes, aunque esto implica una desventaja al limitar la cantidad de predictores que podrían beneficiar el desempeño del modelo. La arquitectura empleada en la RN, no presentó la mayor exactitud, sí presentó un mejor rendimiento en cuanto a la sensibilidad y especificidad, lo cual en términos generales son métricas más deseables para el modelo.

## CONCLUSIONES

Se desarrolló la aplicación médica para la predicción del índice de mortalidad en pacientes con TCE, alimentado por un modelo de aprendizaje basado en RN de perceptrón multicapa, el cual pone a disposición la posibilidad de aprovechar las tecnologías de la IA para incidir en áreas de la salud. El procesamiento del conjunto de datos constituyó una etapa fundamental en el proceso de desarrollo del modelo predictivo, resaltando la importancia de seleccionar de una fuente confiable los datos a ser utilizados en busca de un óptimo desempeño del modelo. El dividir el conjunto de datos en subconjuntos de entrenamiento, prueba y validación permitieron realizar las validaciones internas del modelo, en el proceso de búsqueda de la configuración con mejor desempeño a la problemática.

Demostó una exactitud y especificidad prometedora en la predicción de la probabilidad de mortalidad de pacientes con TCE. Adicionalmente, hay que destacar que es posible lograr un mejor desempeño en modelos de esta índole con datos mejor estructurados y con una mejor agrupación demográfica. Los resultados de este estudio pueden proporcionar información alentadora para el desarrollo de un nuevo modelo de predicción de RN que puedan ser integrado en los sistemas de atención de TCE para predecir la supervivencia de los pacientes. No obstante, es importante enfatizar que este modelo predictivo, aunque puede ser complementario a la toma de decisiones clínicas, no pueden, ni deben reemplazar el juicio clínico.

Debido a que el modelo se desarrolló con datos de un ensayo clínico y se validaron únicamente internamente, se necesita una mayor validación prospectiva en conjuntos independientes para fortalecer la generalización del modelo. En virtud de la naturaleza del conjunto de datos, el modelo no es capaz de responder correctamente a datos más allá de las ocho horas posterior a la lesión del paciente, por lo que actualizar el modelo respecto a estas circunstancias agregaría un mayor valor. A futuro se pretende evaluar diferentes formas o formatos para presentar los modelos a los médicos; su uso en la práctica clínica y verificar en última instancia si tendría algún impacto en el tratamiento y el desenlace de los pacientes con TCE.

## CONFLICTO DE INTERESES

Los autores declaran no tener conflicto de intereses.

## REFERENCIAS

- Aguirre, F., Carballo, L., González, X., & Gigirey, V. (2021). Inteligencia Artificial aplicada a imágenes médica. *Derecho y Realidad*, 47-58. <https://doi.org/10.19053/16923936.v18.n35.2020.9638>
- APD. (2021). *Efectos de la inteligencia artificial en la medicina y sus aplicaciones más novedosas*.
- Beck, K., & Cunningham, W. (1989). A laboratory for teaching object oriented thinking. *ACM SIGPLAN Notices*, 24(10), 1-6. <https://doi.org/10.1145/74878.74879>
- Broenlee, J. (2020). *How to Perform Feature Selection With Numerical Input Data*.
- Cohn, M. (2004). *User Stories Applied: For Agile Software Development*.
- Frost, J. (2021). *Multicollinearity in Regression Analysis: Problems, Detection, and Solutions*.
- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). Rule of Thumb for Interpreting the Size of a Correlation Coefficient. En *Applied Statistics for the Behavioral Sciences* (5.<sup>a</sup> ed., Vol. 5). [http://oak.ucc.nau.edu/rh232/courses/EPS525/Handouts/Correlation Coefficient Handout - Hinkle et al.pdf](http://oak.ucc.nau.edu/rh232/courses/EPS525/Handouts/Correlation%20Coefficient%20Handout%20-%20Hinkle%20et%20al.pdf)
- Ilyas, I., & Chu, X. (2019). *Data Cleaning*. Nueva York: Association for Computing Machinery.
- Kuhn, M., & Johnson, K. (2019). *Feature Engineering and Selection: A Practical Approach for Predictive Models*. Chapman and Hall/CRC.
- Liu, N. T., & Salinas, J. (2017). Machine Learning for Predicting Outcomes in Trauma. *Shock*, 48(5), 504-510. <https://doi.org/10.1097/SHK.0000000000000898>
- Maladkar, K. (2018). *5 Ways To Handle Missing Values In Machine Learning Datasets*.
- Montero Rodríguez, J. C. de J., Roshan Biswal, R., & la Cruz, E. S. de. (2019). Algoritmos de aprendizaje automático de vanguardia para el diagnóstico de enfermedades. *Research in Computing Science*, 148(7), 455-468. <https://doi.org/10.13053/rcs-148-7-34>
- Montero, T. (2012). Traumatismos. *Revista Cubana de Medicina Militar*, 41(1), 1-3.
- Navarro Cadavid, A., Fernández Martínez, J. D., & Morales Vélez, J. (2013). Revisión de metodologías ágiles para el desarrollo de software. *Prospectiva*, 11, 33.
- OPS/OMS. (s. f.). *Discapacidades y traumatismos causados por el tránsito*.
- Raj, R., Luostarinen, T., Pursiainen, E., Posti, J. P., Takala, R. S. K., Bendel, S., Konttila, T., & Korja, M. (2019). Machine learning-based dynamic mortality prediction after traumatic brain injury. *Scientific Reports*, 9(1), 17672. <https://doi.org/10.1038/s41598-019-53889-6>

- Rau, C. S., Wu, S. C., Chuang, J. F., Huang, C. Y., Liu, H. T., Chien, P. C., & Hsieh, C. H. (2019). Machine learning models of survival prediction in trauma patients. *Journal of Clinical Medicine*, 8(6), 1-13. <https://doi.org/10.3390/jcm8060799>
- Roberts, I., Shakur, H., Coats, T., Hunt, B., Balogun, E., Barnetson, L., Cook, L., Kawahara, T., Perel, P., Prieto-Merino, D., Ramos, M., Cairns, J., & Guerriero, C. (2013). The CRASH-2 trial: a randomised controlled trial and economic evaluation of the effects of tranexamic acid on death, vascular occlusive events and transfusion requirement in bleeding trauma patients. *Clinical Governance: An International Journal*, 18(3). <https://doi.org/10.1108/CGIJ.2013.24818CAA.005>
- Salas Velasco, M. (1996). La regresión logística: Una aplicación a la demanda de estudios universitarios. *Dialnet*, 193-217.