

Detección de nefropatía como complicación en pacientes diabéticos de tipo II mediante el uso de la regresión logística

Man Kit Liao-Li¹, José María Celaya-Padilla¹, Carlos E. Galván-Tejada¹,
Jorge I. Galván-Tejada¹, Huizilopoztli Luna-García¹,
Hamurabi Gamboa-Rosales¹, Miguel Cruz²

¹ Universidad Autónoma de Zacatecas,
Maestría en Ciencias del Procesamiento de la Información,
México

² Instituto Mexicano del Seguro Social (IMSS),
Centro Médico Nacional Siglo XXI,
México

{27805072, jose.celaya, ericgalvan, gatejo,
h lugar, hamurabigr}@uaz.edu.mx

Resumen. Datos de la Organización Mundial de la Salud (OMS) muestran que en 2019 la Diabetes fue la novena causa más importante de muertes en el mundo, un efecto común de esta enfermedad no controlada derivada de la hiperglucemia, es que con el tiempo va dañando gravemente muchos órganos y sistemas del cuerpo, como los nervios y vasos sanguíneos, dando paso al desarrollo de complicaciones más comunes como infartos de miocardio o accidentes cerebrovasculares, neuropatías, retinopatías e insuficiencia renal, siendo la última nuestro interés en particular. En el estudio se realizó un análisis sobre un conjunto de datos para determinar las variables más significativas y predecir el desarrollo de Nefropatía en pacientes con diagnóstico confirmado de Diabetes a fin de generar un modelo de clasificación utilizando Regresión Logística como algoritmo de aprendizaje automático.

Palabras clave: Diabetes, nefropatía, regresión logística, selección de características, predicción.

Detection of Nephropathy as a Complication in Diabetic Patients Type II Using Logistic Regression

Abstract. According to data from the World Health Organization (WHO) in 2019, Diabetes was the ninth leading cause of death in the world, a common

effect of this uncontrolled disease derived from hyperglycemia, which over time seriously damages many organs and systems of the body, especially the nerves and blood vessels, giving way to the development of more common complications such as myocardial infarctions or strokes, neuropathies, retinopathy and kidney chronic disease, the last of which is of interest to us in this article. This research performed an analysis on a data set to determine the most significant variables to predict the development of nephropathy in patients with a confirmed diagnosis of Diabetes in order to generate a model using logistic regression as a Machine Learning (ML) algorithm.

Keywords: Diabetes, kidney disease, logistic regression, feature selection, prediction.

1. Introducción

Según datos de la Organización Mundial de la Salud (OMS) el periodo comprendido entre 2000 y 2016, la mortalidad prematura (que comprende a personas antes de los 70 años) originada por Diabetes aumentó en un 5%, y en 2019 este padecimiento causó la muerte en 1.5 millones de personas ocupando la novena causa de defunciones en el mundo [1].

La Federación Internacional de la Diabetes (FID) estimó que para ese mismo año había cerca de 463 millones de casos confirmados con esta enfermedad y se espera que para 2045 las cifras lleguen a los 700 millones, en México para el 2018 según la Encuesta Nacional de Salud y Nutrición se tenía un registro de 82,767,605 sujetos mayores de 19 años con Diabetes, además la tasa de mortalidad pasó de 8.60 en 2016 a 11.95 por cada 10 mil habitantes en 2020.

Para este mismo año, la diabetes pasó a ser la tercera causa de muertes en el país con 151,019 casos en el que 52% fueron hombres y 48% mujeres [2]. La Diabetes Mellitus (DM) mal controlada causa microangiopatía, una afección en los pequeños vasos sanguíneos que contribuye a cambios patológicos y lesión de múltiples órganos produciendo principalmente daños en el sistema nervioso, ocular y renal.

Comúnmente, para diagnosticar en pacientes diabéticos si tienen daño renal se realiza un análisis de orina y medir sus niveles de albuminuria y creatinina con el fin de determinar el grado de filtración glomerular de los riñones

Las manifestaciones de un daño renal se detectan con estos parámetros, sin embargo, es importante considerar que no siempre se seguirá este patrón ya que solo se verá en pacientes con Diabetes Mellitus de tipo 1 (DM1), esta situación no es exactamente así en el caso de la Diabetes Mellitus tipo 2 (DM2) dado que muestran ausencia de albuminuria y a su vez experimentan un deterioro progresivo en la función renal [3].

Debido a las consecuencias a largo plazo que puede conducir esta patología y al impacto en las finanzas en los gobiernos nacionales e internacionales; la comunidad científica, médica y gubernamental han enfocado esfuerzos para reducir los índices de predisposición de este padecimiento en la población, tratando de diagnosticar prematuramente o en todo caso prevenirlo.

Esta investigación pretende impactar en el desarrollo de nuevas herramientas tecnológicas que contribuyan en esta necesidad, con las técnicas de Inteligencia Artificial es posible predecir con anticipación no solo a personas con altas probabilidades de contraer Diabetes si no, también en el desarrollo de complicaciones derivadas de esta enfermedad como Insuficiencia Renal dando la oportunidad de tratar con eficacia y antelación a dichos pacientes.

1.1. Trabajos relacionados

Recientemente la Inteligencia Artificial (IA) ha comenzado a incorporarse en la medicina para mejorar la atención de pacientes logrando una mayor precisión en el diagnóstico y abriendo camino a brindar una mejor atención médica en general. Existen proyectos en la actualidad dedicados a explorar las aplicaciones de la IA en la medicina, una de ellas es la asistencial que busca la prevención, diagnóstico, tratamiento y seguimiento de todo tipo enfermedades, la cual abordaremos con mayor énfasis en esta investigación [4].

En la siguiente sección se muestra un compendio de trabajos relacionados y colaboraciones sobre Nefropatía diabética para tener una perspectiva amplia de lo que se ha investigado en los últimos años. M. A. Makroum en su trabajo presentado en 2022 titulado “Machine Learning and Smart Devices for Diabetes Management: Systematic Review”, presenta una revisión sistemática de investigaciones realizadas en los últimos 10 años, comparando propuestas de aplicaciones para el control de Diabetes que tienen aplicación tanto de escasas técnicas de IA como los que si tienen implementación de algoritmos de aprendizaje automático.

Todas las implementaciones están limitadas exclusivamente al control de Diabetes, no se incluye para el seguimiento y predicción de complicaciones derivadas de esta patología [5]. En 2021 P. Chittora presenta su trabajo en el artículo “Prediction of Chronic Kidney Disease - A machine learning perspective”, una comparativa sobre 7 de diferentes modelos de clasificación: Redes Neuronales, C5.0, Regresión Logística, CHAID (Chi-square automatic interaction detection), LSVM, KNN y Random Forest. Genera en primera instancia un modelo con todas las características del conjunto de datos y posteriormente hace una selección de ellas por medio de diferentes métodos (filter, wrapper & embedded).

Evalúa cada algoritmo con las métricas de desempeño: Accuracy, Classification Error, Precisión, Recall, F-Measure, Coeficiente GINI, Curva ROC y Área bajo la Curva (AUC). El mejor modelo de clasificación fue LSVM con SMOTE (Sobremuestreo de las clases minoritarias) con todas las variables del conjunto alcanzando un Accuracy de 98.86%, seguido del modelo con el método embebido (LASSO FS SMOTE) utilizado para la selección de características, obtuvo un Accuracy de 98.46%.

También identificó que el estudio se realiza en un conjunto de datos de la Universidad de California Irvine de su repositorio de Machine Learning, en los que incluyen confirmación de padecimientos como Diabetes e hipertensión y estudios de laboratorios que suelen ser necesarios para el control de pacientes con Enfermedad Crónica Renal (ERC).

Sin embargo, no incluyen medicamentos o tratamiento que sigue cada paciente. En la selección de características se limita a considerar solamente 6 más significativas de cada método [6]. M. A. Islam muestra en su artículo “Risk Factor Prediction of Chronic Kidney Disease Based on Machine Learning Algorithms” del 2021, la implementación de algoritmos de clasificación como: Regresión Logística, Random Forest y Naive Bayes. Establece a la hemoglobina como la característica fundamental para el modelo de predicción y la menos significativa a la hipertensión. El mejor modelo de clasificación fue Random Forest con un Accuracy de 98.88%. El conjunto de datos que se analizó solo se limitó a estudiar los casos ya confirmados con ERC.

No hay antecedentes de Diabetes ni de tratamiento que siguen los pacientes, solo cuenta con historial de hipertensión [7]. M. Vásquez en su Tesis de Maestría de 2019 propone un modelo basado en Redes Neuronales (RN) complementando el análisis con SVM y Random Forest. Pronostica el desarrollo de Enfermedad Renal Crónica a partir de registro de historial médico de pacientes, no se incluyen pruebas de laboratorio solo padecimientos comunes que desarrollan las personas con ERC entre ellos la Diabetes e hipertensión.

El mejor modelo fue Redes Neuronales con un AUC de 98%, aunque Random Forest reporta un comportamiento similar muestra más falsos positivos. En el estudio no hay información referente a tratamientos o medicamentos que toman los pacientes ni datos sobre muestras de laboratorio [8].

H. Polat en su trabajo de investigación de 2017 titulado “Diagnosis of Chronic Kidney Disease Based on Support Vector Machine (SVM) by Feature Selection Methods”, realiza un modelo de clasificación en 5 diferentes perspectivas. El primero sin selección de características, y en las otras restantes con distintos métodos de selección con el fin de comparar las métricas de desempeño y el número de características significativas.

El mejor modelo de clasificación fue utilizando la selección por evaluador de subconjunto filtrado con un Accuracy del 98.5%. Su limitación es que utiliza el conjunto de datos de la Universidad de California Irvine que tiene el registro de 400 pacientes con 24 características para realizar el análisis, pero enfoca solamente a SVM como modelo clasificador [9].

Se puede observar que todas las propuestas se centran en comparar los algoritmos de predicción existentes, sin tomar en cuenta la importancia que tienen las características significativas resultantes, para determinar si el contexto del estudio es suficiente para abordar el tema en una perspectiva funcional.

Este artículo busca mejorar la evaluación del modelo de clasificación implementando validación cruzada, concepto que no se aborda en ninguna investigación anterior, y nos aseguran si las predicciones muestran algún sesgo o sobreajuste. Además, se pretende contribuir en el diagnóstico prematuro para disminuir el índice de incidencia de ERC en etapas terminales.

1.2. Objeto de estudio

De acuerdo con lo anterior el estudio busca predecir y diagnosticar prematuramente a pacientes con Diabetes a desarrollar potencialmente una Enfermedad Renal Crónica



Fig. 1. Diagrama que describe el flujo del desarrollo de la investigación para obtener el modelo de predicción de ERC.

(ERC) utilizando algoritmos de aprendizaje automático presentando un modelo categórico con las características más significativas previamente seleccionadas, a partir de un análisis sobre un conjunto de datos proporcionada por el Centro Médico Nacional Siglo XXI del Instituto Mexicano del Seguro Social (IMSS).

2. Materiales y metodología

Esta investigación se realizó mediante de software estadístico R Studio desarrollado en diferentes etapas que se describirá a continuación. En primera instancia, se efectuó un tratamiento y limpieza a los datos, en segunda se abordará la selección de características significativas, en tercera tendremos la generación del modelo de clasificación implementando los algoritmos más comunes encontrados en la literatura como Regresión Logística, Random Forest y KNN obteniendo su eficiencia en la predicción con las métricas de desempeño tema que se contempla en la cuarta etapa, por último se evaluará el modelo con la validación cruzada con el fin evitar un sesgo en las predicciones y la precisión del modelo; todo este desarrollo propuesto se puede apreciar en el diagrama de flujo que se muestra en la Fig. 1.

2.1. Preprocesamiento

El conjunto de datos utilizado para este estudio, fue proporcionado por la Unidad de Investigación Médica en Bioquímica del Hospital de Especialidades “Bernardo Sepúlveda” del Centro Médico Nacional Siglo XXI del Instituto Mexicano del Seguro Social (IMSS) por un convenio de colaboración con la Universidad Autónoma de Zacatecas, es un conjunto de datos, contiene 46 características y 1787 pacientes con diagnósticos confirmado de Diabetes y personas sanas; tomando en consideración solo las características que implican médicamente en el desarrollo de cualquier complicación derivado de la Diabetes como lo es la edad, índice de masa corporal, así como niveles de creatinina, urea, glucosa, lípidos y colesterol en sangre.

De la misma manera se añadió al análisis los medicamentos más utilizados por los pacientes para el control de la Diabetes para observar su comportamiento en la participación de complicaciones.

Este conjunto de datos ha sido utilizado en investigaciones previas para abordar la predicción de Diabetes [10] el cual, presenta valores faltantes por lo que se le sometió a un proceso de imputación de datos con la librería de R missForest, este método

iterativo consiste en asignar valores inexistentes realizando múltiples árboles de decisión sobre las muestras del del conjunto de datos, los valores asignados son un promedio de las observaciones, obteniendo un error en las predicciones del 1.55%, corroborando este dato se obtuvo el promedio de la variable HDLU antes y después de la imputación con 44.23 y 44.21 respectivamente.

Nuestro conjunto de datos tiene además de la Insuficiencia Renal otras complicaciones como Retinopatía, Neuropatía, Cardiopatías Isquémicas, entre otros; y para este estudio solo se consideraron a pacientes Nefrópatas con antecedente de diabetes descartando las demás, obteniendo de esta manera un conjunto de datos con un total de 26 pacientes de los cuales 13 son casos (Nefropatía) y 13 de control (Diabetes sin complicación).

A pesar de que el dataset final es demasiado pequeño, este estudio buscar obtener un primer acercamiento en la predicción con técnicas de Inteligencia Artificial desde un contexto aplicativo en la sociedad mexicana, todo con la premisa de generar una herramienta que mejore las estrategias de detección en etapas iniciales de la ERC derivada de la Diabetes.

2.2. Selección de características significativas

Para determinar las características más significativas para el modelo de clasificación se implementó el proceso de selección hacia adelante y hacia atrás. En el caso del primer método el proceso comienza con un modelo vacío que posteriormente va añadiendo variables significativas a partir de un criterio de similitud, el cual seguirá sumando términos al modelo hasta ya no encontrar variables que impacten en la respuesta del mismo.

En el segundo pasa lo contrario, parte de un modelo que tiene todas las características que pueden influir en la respuesta, y va eliminando términos menos influyentes dejando las variables más significativas, cabe resaltar que este procedimiento se realiza exclusivamente con Regresión Logística.

El criterio de similitud es la información de Akaike (AIC), un estimador muestral de la esperanza de log-verosimilitud, el cual se define en función de la máxima verosimilitud de las observaciones, cuanto más aumenta el número de parámetros aumenta el valor de AIC y el mejor modelo se va ajustando conforme este valor sea menor [11]. Con este proceso no se pretende identificar al modelo verdadero, sino al mejor de los modelos candidatos que se ajustan a las observaciones.

2.3. Generación de modelo de predicción

Una vez obtenido las características significativas se procede a generar el modelo de predicción implementando los algoritmos más usados en la literatura como Regresión Logística (RL), Random Forest y KNN.

De este proceso RL resultó ser el más sencillo y eficiente de implementar mientras los otros dos presentaban un sobre ajuste en las métricas de desempeño por lo que el estudio se centrará en este modelo, además de que resulta muy útil para los casos en que se desea predecir la presencia o ausencia de una característica según los valores de

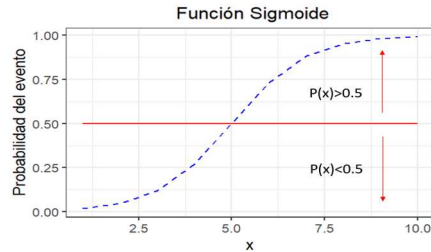


Fig. 2. Gráfica de la función sigmoide que describe al modelo de Regresión Logística.

un conjunto de predictores previamente analizadas y seleccionadas en donde se eliminan atributos poco relacionados con la salida.

Este algoritmo es similar al modelo de regresión lineal, pero está adaptado para modelos en que la variable dependiente es dicotómica o binaria [12]. Este modelo predice los eventos en probabilidades de ocurrencia de un suceso, en función del valor que toman las variables independientes de 0 a 1. Matemáticamente es posible utilizar el modelo de la regresión lineal de mínimos cuadrados, pero los valores extremos del predictor serían menores que 0 o mayores a 1.

Por ello es común utilizar una función que responda a estos requisitos que en la práctica suele ser sigmoide, dada por la ecuación 1. Si en la función anterior, le damos valores a z obtendremos una curva como la que se muestra en la Fig. 2, la cual es utilizada para clasificar 1 para cuándo $y > 0.5$ y 0 para $y < 0.5$:

$$p(x) = \frac{1}{1 + e^{-z}} \quad (1)$$

Por esto, z viene siendo la representación de los coeficientes del modelo de regresión que después de realizar un desarrollo algebraico sobre la ecuación sigmoide obtenemos la expresión matemática que describe el modelo de Regresión Logística que se muestra en la ecuación 2 [13]:

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \quad (2)$$

2.4. Medidas de desempeño

Un algoritmo se evalúa para determinar si realizará una buena predicción, basándose en cómo funciona el modelo con nuevos y futuros datos, para ello se utilizan varias métricas que se emplearon en la investigación para validar de manera empírica el desempeño del modelo y que se describen a continuación:

- **Accuracy (Exactitud):** Mide la fracción de predicciones correctas y se expresa debido a probabilidad de 0 a 1, entre más se acerque a 1, mayor será su exactitud predictiva [14].
- **Curva ROC:** La curva característica operativa del receptor (Receiver Operating Characteristic) es una representación de la sensibilidad frente a la

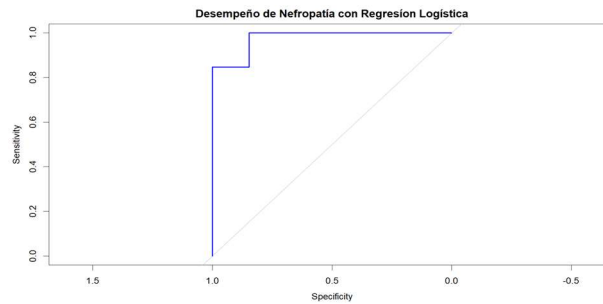


Fig. 3. Curva ROC del modelo de Regresión Logística con Creatinina, Insulina y Sexo.

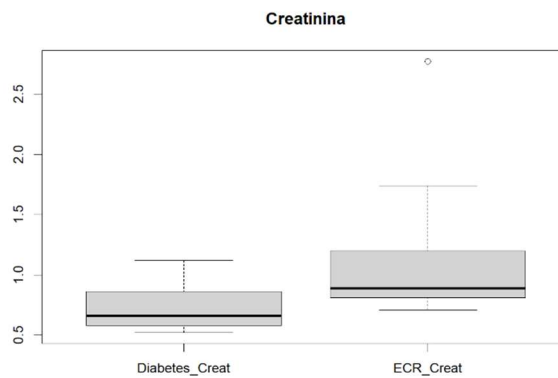


Fig. 4. Gráficas de caja y bigotes que describen la creatinina en pacientes diabéticos y los que tienen ERC.

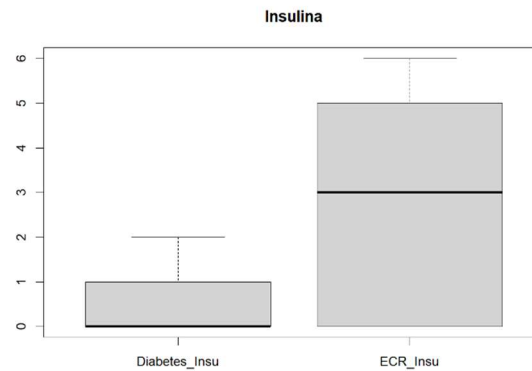


Fig. 5. Gráficas de caja y bigotes que describen el comportamiento de la Insulina en pacientes diabéticos y ERC.

especificidad, expresada como una probabilidad de predicción de un evento con los verdaderos positivos y los falsos positivos dada por el Área Bajo la Curva (AUC) establecida entre valores de 0 al 1.

- **Sensibilidad:** Se define como la capacidad de un algoritmo para predecir un resultado positivo verdaderos, es decir que tanto la predicción como el resultado real es positivo. Se dice que un algoritmo tiene una sensibilidad perfecta cuando predice que todo es positivo pase lo que pase. Por tanto, esta métrica no es suficiente para evaluar el desempeño del algoritmo.
- **Especificidad:** Esta métrica se define como la capacidad que tiene un algoritmo para predecir un resultado en positivo falso, es decir que teniendo el resultado real en negativo la predicción obtiene un positivo [15].

2.5. Validación cruzada

Para garantizar un buen modelo de clasificación es necesario asegurar que las predicciones sean precisas y evitar sobreajustes en sus resultados por ello la validación cruzada cumple un papel importante.

Este proceso consiste en dividir el conjunto de datos original en k subconjuntos (en nuestro caso $k=3$) en el que cada uno se les hará el entrenamiento y prueba con el modelo de predicción resultante, el proceso se repetirá k veces y en cada iteración se tomará un conjunto de prueba diferente.

Al finalizar las iteraciones se calculan las métricas de desempeño de cada subconjunto y se obtienen sus promedios, con esto nos da un estimado de la precisión del modelo sin un sobreajuste en la predicción [16].

3. Resultados

Para el análisis de predicción a desarrollar una ERC con precedentes de Diabetes se tomaron en cuenta las 42 características y al implementar la selección hacia adelante y hacia atrás se obtuvo un modelo con 32 variables de las cuales la mayoría no aportaba una significancia importante en el modelo.

Por lo que se procedió a realizar una reducción manual eliminando las variables menos significativas descritas por la prueba t de student, un parámetro que determina si las medias de dos grupos (características) tienen una diferencia importante [17], consiguiendo de esta manera solamente 3 características significativas (Creatinina, Insulina y Sexo) con un desempeño AUC de 97.63% cuya Curva ROC se muestra en la Fig. 3.

Creatinina e Insulina fueron las variables que más aportan en la predicción, implementando la primera característica en un modelo univariado con Regresión Logística nos da un AUC de 80.47%, se puede ver su comportamiento en las gráficas de caja y bigotes que muestran una presencia mayor en los pacientes con Enfermedad Renal Crónica (ERC) que en los diabéticos (ver Fig. 4).

Realizando el mismo procedimiento con Insulina, el AUC del modelo univariado fue de 77.51% se puede observar en la Fig. 5 las gráficas de caja muestran una diferencia considerable en los pacientes con ERC respecto a los diabéticos.

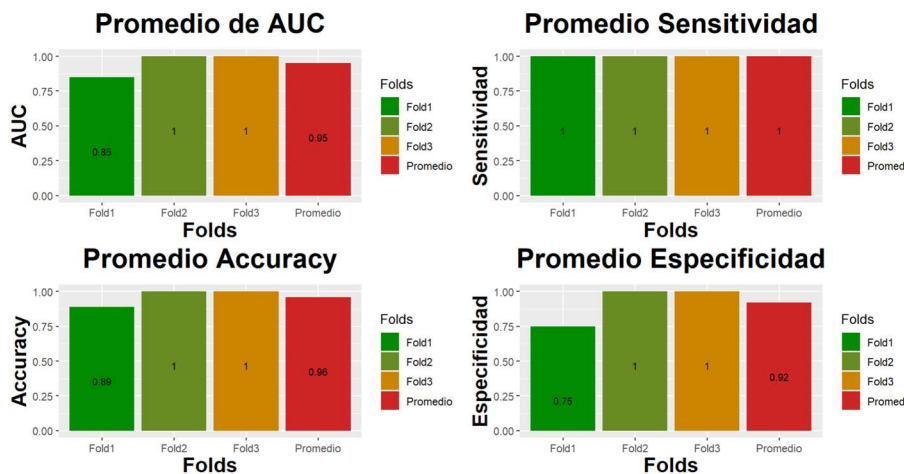


Fig. 6. Promedios generales de las métricas de desempeño en la validación cruzada.

3.1. Validación cruzada

Esta evaluación se realizó con la librería de Caret para hacer la separación del conjunto de 26 pacientes en subconjuntos k-folds ($k=3$) obteniendo así particiones diferentes de las observaciones originales y con ello realizar entrenamientos y pruebas del modelo propuesto, en este caso se dividió el data set en 3 particiones, en cada una se evaluaron las métricas de desempeño en prueba y se obtuvo el promedio general de cada métrica, mismas que se muestran en la Fig. 6.

4. Conclusiones y trabajo a futuro

Después de aplicar la metodología propuesta, esta generó un modelo conformado por Creatinina, Insulina y Sexo, el cual obtuvo un desempeño en la prueba de 97.63% para AUC, 92.31% en Accuracy, 84.62% de Especificidad y 100% en Sensibilidad; el nivel de precisión fue un tanto cercano a los mencionados anteriormente en la sección de trabajos relacionados presenta una ventaja, nuestro modelo tiene únicamente 3 variables en comparación al mejor modelo abordado en la sección 1.1 que utiliza 24 características.

Se encontró a la Creatinina por sí sola como una variable muy buena para la predicción de una ERC, en un modelo univariado esta característica representa más del 80% de eficiencia; aunque no es algo desconocido por la comunidad médica tener a la Creatinina como factor determinante en la detección de una ERC se puede afirmar que este parámetro resultó ser más significativa que la misma Urea (característica también presente en nuestro conjunto de datos), desecho que comúnmente también suele presentar niveles muy altos en pacientes con daño renal [18], además considerando la investigación médica referida al inicio de este artículo en donde describen que la

presencia de albuminuria como medida de diagnóstico no es constante en todos los pacientes con Diabetes de tipo II [2] y con los resultados obtenidos en este estudio, se propone un seguimiento constante de niveles de este residuo en el torrente sanguíneo como prioritario, más que la presencia de proteína en la orina, para detectar el nivel de riesgo a predisponer o en su caso diagnosticar una Nefropatía Diabética en pacientes con este tipo de Diabetes.

Este argumento se tendría que ser valorada y validada por un experto en el área para ser considerada como aportación a la comunidad médica y académica.

Otra variable importante encontrada fue el medicamento de Insulina el cual se complementa en el modelo de clasificación multivariado, y según la práctica médica este parámetro suele ser más alto en pacientes con Insuficiencia Renal, debido a este fármaco se metaboliza en el riñón y al existir una falla considerable en este órgano, la Insulina permanece más tiempo en el sistema, por ello, considerar esta variable permite tener una perspectiva complementaria en la predicción de este padecimiento.

Cabe mencionar que el conjunto de datos que se utilizó para el estudio no cuenta con niveles albuminuria para añadir en el análisis, sería de gran interés considerarla en un trabajo a futuro para observar su comportamiento con el modelo de predicción ya que es muy utilizada en la comunidad médica para diagnosticar a los pacientes con Enfermedad Renal Crónica seguido de la creatinina, así también incluir el tiempo de diagnóstico de Diabetes para ver la relación del daño paulatino de los vasos sanguíneos con la falla renal.

Otro aspecto importante para considerar es contar con un conjunto de datos más amplio de casos confirmados de ERC dado que en este estudio solo se tenía 13 pacientes positivos y 13 con solo Diabetes, contar con más observaciones ayudaría tener un modelo con mayor precisión. Es importante resaltar que este estudio es un pequeño acercamiento en la predicción de la Enfermedad Renal Crónica Diabética, en próximas investigaciones se pretende implementar otros algoritmos de predicción como Support Vector Machine (SVM) y Redes Neuronales, para comparar sus comportamientos respecto a los obtenidos con Regresión Logística y obtener un mejor modelo de clasificación.

Agradecimientos. Esta investigación pudo ser realizada gracias al apoyo otorgado por el Consejo Nacional de Ciencia y Tecnología al alumno Man Kit Liao-Li con el número de becario 901279 de la Maestría en Ciencias del Procesamiento de la Información, que sin su aporte el desarrollo sobre este tema en la comunidad científica no sería posible.

Referencias

1. Organización Mundial de la Salud: Diabetes. Centro de Prensa, Organización Mundial de la Salud (2021)
2. Estadísticas a Propósito del Día Mundial de la Diabetes. Datos Nacionales (2021)
3. Góriz, J., Terrádez, L.: Clínica y Anatomía Patológica de la Nefropatía Diabética 1. Clínica de la Enfermedad Renal Diabética Introducción, Valencia (2021)

4. Ávila-Tomás, J.F., Mayer-Pujadas, M.A., Quesada-Varela, V.J.: Artificial intelligence and its applications in medicine II: Current importance and practical applications *Atencion Primaria*, vol. 53, no. 1, pp. 81–88 (2021)
5. Makroum, M.A., Adda, M., Bouzouane, A., Ibrahim, H.: Machine Learning and Smart Devices for Diabetes Management: Systematic Review. *Sensors*, vol. 22, no. 5, p. 1843 (2022)
6. Chittora, P, et.al.: Prediction of Chronic Kidney Disease - A Machine Learning Perspective. *IEEE Access*, vol. 9, pp. 17312–17334 (2021)
7. Islam, M. A., Akter, S., Hossen, M.S., Keya, S.A., Tisha, S.A., Hossain, S.: Risk factor prediction of chronic kidney disease based on machine learning algorithms. In: *Proceedings of the 3rd International Conference on Intelligent Sustainable Systems, ICISS 2020*, pp. 952–957 (2020)
8. Vásquez, M. Gabriel, R.: Clasificador con redes neuronales para el pronóstico de la enfermedad renal crónica en la población colombiana (2022)
9. Polat, H., Danaei Mehr, H., Cetin, A.: Diagnosis of Chronic Kidney Disease Based on Support Vector Machine by Feature Selection Methods. *Journal of Medical Systems*, vol. 41, no. 4 (2017)
10. Alcalá-Rmz, V.: Identification of People with Diabetes Treatment through Lipids Profile Using Machine Learning Algorithms (2021)
11. González, A.: Selección de variables: una revisión de métodos existentes (2015)
12. IBM: *Regresión Logística. SPSS Statistics* (2021)
13. Páez, O., Sangrador, O., Arias, M.: *Regresión logística binaria simple* (2022)
14. Amazon Web Services: *Amazon Machine Learning Guía para desarrolladores* (2016)
15. Irizarry, R. A.: Introducción a la ciencia de los datos. pp. 529–582 (2021)
16. Laura-Ochoa, L.: Evaluation of classification algorithms using cross validation. In: *Proceedings of the LACCEI international Multi-conference for Engineering, Education and Technology*, vol. 2019 (2019)
17. Portal de formación estadística: *La prueba t* (2022)
18. Salabert, E.: *Urea alta: causas, síntomas, y como bajar sus niveles* (2018)