

MODELOS DE APRENDIZAJE AUTOMÁTICO PARA LA PREDICCIÓN DE LA PROGRESIÓN DE LA ENFERMEDAD RENAL CRÓNICA

William David Amador Martínez, Leonardo Flórez Valencia

Abstract—Los modelos predictivos basados en técnicas de inteligencia artificial han evidenciado su capacidad de establecer predicciones del riesgo de diferentes patologías asociadas a variables clínicas con el fin de generar sistemas que faciliten la toma de decisiones médicas. Debido a que muchos de estos modelos de predicción no generan buenos resultados cuando faltan valores predictores, muchas veces no se pueden detectar de manera temprana la progresión de las enfermedades. El propósito de esta investigación es la implementación de un CDSS basado en métodos de imputación de datos para entrenar modelos de aprendizaje automático que permita la predicción de la TFG en un periodo de 2 años con intervalos de 6 meses. La metodología se dividirá en dos etapas (i) la preparación de los datos conformada por 482 pacientes con ERC en estadios 1 a 4 en un periodo mayor de 2 años de seguimiento en el programa, donde se realiza la limpieza filtrado e imputación de las variables seleccionadas para alimentar los modelos (ii) implementación de un modelo de predicción en el tiempo y evaluación métricas error MAE average en las diferentes configuraciones de los hiperparámetro de los modelos. Entre los enfoques evaluados las Redes Neuronales Convolucionales (CNN) se desempeñó mejor con un conjunto de testeo y validación con dos arquitecturas diferentes para una para los periodos de 6^a 12 y otra de 18 a 24 con un MAE 5.48, 6.46, 6.76, 6.53 para los períodos respectivamente en los datos de validación.

Proclamación del impacto—Los algoritmos de predicción en el área de la salud se convierten en una herramienta que puede facilitar la detección temprana de un deterioro en una patología clínica logrando alertar al profesional de la salud para hacer un tratamiento preventivo que evite en un agravio irreversible en un futuro

Índice de términos—Enter Deep Learning; Prediction; CKD.

I. INTRODUCTION

La enfermedad renal crónica (ERC) esta categorizada entre una de las 10 principales causas de muerte a nivel mundial, las muertes por ERC para el año 2017 fueron 1.2 millones casi igual al número de muertes por accidentes de tránsito y se provee que para el 2040 aumente a 2.2 millones [1]. Teniendo mayor presencia en países de ingresos bajos y medianos debido

a los pocos modelos de atención en salud que han sido implementados para esta patología, la ERC cuenta con comorbilidades asociadas como lo son: diabetes, enfermedades vasculares, hipertensión, obesidad; además existen causas que conllevan a la progresión de la enfermedad referentes a la raza, el sexo, ubicación y el estilo de vida [2] [3].

La ERC se ha convertido en una problemática a nivel de salud, ya que esta tiene implicaciones directas en las políticas en salud y el aumento en la carga financiera, al ser considerada una enfermedad de alto costo[4]. Esto sumado a que en Latinoamérica los países que cuentan con un sistemas de salud con mayor cobertura financiera tienen programas más robustos y asequibles, mientras que los países más pobres manejan servicios de atención que dependen en mayor medida de fuentes privadas para crear cobertura generando altos costos de tratamiento [5]. Con el fin de retrasar esta enfermedad para que no conlleve al requerimiento de terapias de remplazo renal y poder controlar los costos de manejo de esta población, se han establecido modelos de intervención de bajo costo como lo son tamizajes preventivos asociados a detección tempranas y gestión o intervención de casos conocidos [6].

La inteligencia artificial ha jugado un papel importante para ayudar a generar tamizajes preventivos a partir de la información clínica de los pacientes entre los cuales existen estudios que han demostrado que el aprendizaje automático supera a los métodos estadísticos convencionales debido a su capacidad para identificar mejor las variables relevantes para los resultados clínicos [7], técnicas como regresión logística, arboles de decisión, máquina de vector de soporte, k-vecino más cercano, clasificador de Bayes han sido altamente utilizadas en el diagnóstico de la ERC como lo muestran investigaciones [8] [9] otros enfoques han mostrado resultados prometedores en la estratificación del riesgo y predicción de la enfermedad renal utilizando técnicas de aprendizaje automático (ML) en estudios controlados [10].

Pese a esto existe el desafío lidiar con la gran cantidad de datos de diferentes tipos de estructuras, un problema que las técnicas de aprendizaje automático han logrado superar debido a su solidez, al ruido de datos y su capacidad para aprender de múltiples conjuntos de datos [11], Otro desafío que se presenta es el trabajo con valores predictores faltantes generando que los modelos de predicción no puedan proporcionar una estimación del riesgo efectiva, esto a causa de que la información de atención está incompleta por razones como costos de nuevos procedimientos, pertinencia médica, problemas con la

extracción de historia clínica, entre otras [12] [13]. Estos datos faltantes se convierten en un reto para las personas interesadas en estudiar la progresión de un paciente en su enfermedad, para lo cual se han realizado investigaciones [14] [15] que abordan este tema desde técnicas como la imputación media, imputación marginación sobre las variables faltantes, imputación múltiple con ecuaciones encadenadas (MICE), proceso gaussiano (GP) y 3D-MICE entre otras, todo esto teniendo en cuenta la importancia que los datos tienen y más si se habla de un modelo de predicción en el tiempo de un paciente [16].

Esta necesidad de predecir y clasificar si una persona llegase a tener ERC nos lleva a investigaciones en las que el uso del aprendizaje profundo (DL) juega un papel importante para apoyar el análisis que los profesionales en salud hacen para su tomar decisiones que impactan directamente en la salud de los pacientes, mediante sistemas que permitan generar alerta temprana a los pacientes con comorbilidades como DM para predecir futuras fallas metabólicas que conlleven a otra enfermedad como ERC o enfermedades de las arterias coronarias (EAC)[17].

Los algoritmos de aprendizaje profundo han jugado un papel importante en el diagnóstico de enfermedades como la ERC debido a su alta eficiencia al identificar parámetros para una clasificación temprana, esto lo observamos en estudios como [18] que utilizan algoritmos como ANN, Simple RNN y MLP proporcionaron una alta precisión del 99%, 96%, 97% respectivamente. Un enfoque que toma relevancia en la parte clínica es la evaluación de la progresión de la enfermedad, donde modelos como Redes Neuronales Convolucionales (CNN) presentan resultados prometedores en la predicción de los próximos 6 o 12 meses de la ERC [19] utilizando períodos de observación de dos años utilizados para construir el modelo con un desempeño de AUROC de 0,957 y 0,954 respectivamente para cada periodo de tiempo.

II. METODOLOGÍA

Esta sección describe el procedimiento realizado para la consolidación de los datos, selección de variables objetivo del estudio, establecimiento de criterios de inclusión y exclusión de la población, preparación de los datos para su análisis, selección de los algoritmos, optimización de parámetros y evaluación de estos con los datos que se procesaran.

A. Población, cohorte del estudio

La población de estudio para esta investigación será analizada de los registros de atención primaria de un proveedor de servicios de salud de una cohorte retrospectiva de pacientes del Caribe Colombiano con ERC de atención primaria y secundaria. La población total en este estudio estuvo compuesta por 285 pacientes con diagnóstico de ERC estadios 1 a 4 información recopilada desde el año 2017 al 2021 obteniendo datos de laboratorio y de la historia clínica electrónica de los cuales se tiene disposición que son relevantes para esta enfermedad de los que se encuentra reporte en la literatura. Con este volumen de información se inició el proceso de establecer los criterios de inclusión, exclusión y cálculos correspondientes

para la determinación de la variable de estudio TFG como predictora de la enfermedad.

Los criterios inclusión y exclusión para este grupo de estudio fueron pacientes entre, 18-90 años, diagnosticados con enfermedad renal crónica y sin inicio de terapia de reemplazo renal que contaran con un tiempo de seguimiento mayor a 36 meses y con más de tres medidas de resultado de la TFG, no deberían contar con diagnóstico de otras enfermedades como VIH, Cáncer en el periodo de observación y trasplante renal en los registros médicos.

B. Datos predictores

El proceso de limpieza y preparación de datos inicia en la extracción de información desde la historia clínica, y otras fuentes de información del prestador del servicio, donde se identifican cuatro grupos de información, cada uno de estos grupos cuenta con diferentes variables asociadas las cuales se relacionan a continuación:

- A. Datos de laboratorios: 'Creatinina', 'HbA1c', 'Microalbuminuria', 'Ldl', 'Hdl', 'Colesterol Total', 'Fosforo', 'Albumina', 'Potasio', 'Calcio', 'Parathormona' con su respectiva fecha.
- B. Información básica: 'Nombres', 'Información de contacto', 'Sexo', 'Fecha de nacimiento', 'Etnia'.
- C. Datos de atenciones: 'Peso', 'Talla', 'Tas', 'Tad' con su respectiva fecha.
- D. Datos de antecedentes: CE10 con su respectiva fecha.

Para el grupo A solo se consideraron dos laboratorios creatinina y HbA1c debido a que por frecuencia de contratación en la IPS los pacientes no cuentan con una cantidad superior al 30% para la población de estudio y esta varía dependiendo del estudio, haciendo que estos laboratorios se descarten, no por importancia como variable de predicción, pero si por falta de información para alimentar el modelo.

Para el grupo B se toma la información relevante para el modelo, se descarta la variable Etnia por calidad de dato, del grupo C se toma la totalidad de las variables.

Por ultimo para el grupo D se realiza un agrupamiento de los CIE10 para generar las categorías de diabetes mellitus e hipertensión, los grupos asociados se relacionan a continuación:

DM = ["E105", "E106", "E107", "E108", "E109", "E110", "E115", "E116", "E117", "E118", "E119", "E146", "E147", "E149"]
HTA = ["I110", "I119", "I120", "I129", "I130", "I150", "I158", "I209", "I213", "I219", "I229", "I250", "I251", "I252", "I255", "I258", "I259"]

La Tabla I muestra una descripción de las variables resultantes después de la limpieza y rangos aceptables para esta patología

TABLA I
VARIABLES EVALUADAS EN EL ESTUDIO

Variable	Descripción	Rango de evaluación
TFG	Tasa de filtración glomerular	120- 15
Sexo	Sexo del paciente (Masculino o Femenino)	1 o 0
Edad	Edad reportada en los diferentes intervalos de tiempo	18 - 110
Peso	Peso reportado en los diferentes intervalos de tiempo	40 - 140
Talla	Talla reportada en los diferentes intervalos de tiempo	1.0 - 2.0
Creatinina	Laboratorio de creatinina	0.09 - 15.0
Glicosilada	Laboratorio de hemoglobina glicosilada	3.8 - 18.5
TAS	Tensión arterial sistólica	0 - 300
TAD	Tensión arterial diastólica	0 - 300
DM	Diagnóstico de diabetes mellitus	1 o 0
HTA	Diagnóstico de hipertensión arterial	1 o 0

Al ser la TFG la variable de estudio se procede a calcular mediante la ecuación de CKD-EPI (1), la cual considera el sexo, edad, peso y creatinina, con estas variables se procede a calcular cada una de las observaciones en las que tenga estas variables en el mismo mes.

(1) if ['sexo'] == Hombre:

$$((140-X[\text{edad}]) * [\text{peso}]) / (72 * X[\text{creatinina}])$$
 else:

$$(((140-X[\text{edad}]) [\text{peso}]) / (72 * X[\text{creatinina}]) * 0.85)$$

Posterior a esto se realiza la clasificación de estadio, lo cual permite evidenciar el grado de afectación medido por el Filtrado Glomerular, la ERC cuenta con cinco estadios que en sus primeras fases puede pasar como enfermedad asintomática y en las últimas fases no hay posibilidad de retroceso de estadio requiriendo tratamiento con terapia renal sustitutiva.

TABLE II
 PROGRESIÓN DE LA ENFERMEDAD RENAL CRÓNICA

Estadio	Filtración Glomerular	Descripción
1	>90 + proteinuria	Daño renal FG normal o alto
2	89 - 60	Daño renal ligero con FG disminuido
3A	45 - 59	Daño medio a moderado
3B	30 - 44	Daño moderado a severo
4	29 - 15	Daño renal severo
5	< 15	Enfermedad renal crónica avanzada

C. Limpieza y preparación de datos

La preparación de datos es uno de los aspectos más importantes y, a menudo suele ocupar entre el 50 % y el 70 % del tiempo de un proyecto. Para esta fase se seleccionó la metodología CRISP – DM, que es un modelo de proceso que sirve como base para realizar ciencia de datos. Para el cumplimiento del proyecto y sus objetivos, en la preparación de datos se implementaron las siguientes tareas según esta metodología:

- Fusión de conjuntos de datos y/o registros
- Selección de un subconjunto de datos de muestra
- Agregando registros
- Derivación de nuevos atributos
- Ordenar los datos para el modelado

- Eliminar o reemplazar valores en blanco o faltantes
- División en conjuntos de datos de entrenamiento y prueba

Las 4 etapas de la metodología CRISP - DM y sus tareas se describen en la Fig. I:

Como primera etapa se realizó el proceso de unificación de las fuentes de información, puesto que el prestador de atención tiene tres fuentes de información distintas (historia clínica, información de laboratorio e información de gestión de atención), esto con el fin de recolectar las variables necesarias para alimentar el modelo de predicción a 2 años con costes a 6 meses, teniendo la información unificada se procede al segundo paso el cual consistió en la limpieza de datos, lo que implicó eliminar valores duplicados, faltantes e incorrectos.

Con el paso anterior se tiene un consolidado de 934 pacientes con los cuales se realiza la tercera etapa, en la cual se establecen 4 filtros para obtener la población de estudio, en el primer filtro se establece los rangos de normalización para peso y talla y laboratorios de estudio, el segundo filtro implementado consistió eliminar a los pacientes que no tenían una continuidad en el programa mayor de 2 años, en la tercera condición consistía en que los pacientes debían contar con mínimo con 3 resultados de creatinina en el periodo de observación y si contaban con diagnóstico de diabetes mellitus, los pacientes tenían que tener registrados 3 observaciones de hemoglobina glicosilada.

El cuarto y último filtro relacionado está basado en el cruce de los datos de creatinina e información personal y medidas antropométricas para obtener la TFG, con esta se eliminan los pacientes que se encuentren en estadio 5 puesto que esta población ya se encuentra en diálisis. Con estas etapas se obtuvo la población de estudio la cual está constituida por 482 pacientes en los cuales la predicción es oportuna y se puede realizar en los estadios 1, 2 y 3ª en los cuales es más fácil realizar un retroceso de estadio ya que en una etapa más avanzada no se podría lograr.

Como cuarta y última etapa se implementó una interpolación lineal, esta técnica se utilizó para determinar los valores de las funciones de cualquier punto intermedio a partir de los valores de dos puntos adyacentes con el objetivo de obtener la variable de estudio en el modelo, en un periodo de 2 años con intervalos de 6 meses, esto mediante la librería de Resampler.interpolate de pandas, donde se generaron grupos por cada paciente incorporando los datos de laboratorios que se querían interpolar, estos asociados a las fechas de muestra que se asocian como índices de entrada a la función de interpolación, a continuación se evidencia la estructura de la sintaxis :

```
W.index = pd.to_datetime(W.index)
Res=W.resample('s').interpolate().resample('262800T').asfreq()
```

Con los datos interpolados de cada uno de los pacientes se unifican con las otras variables con el fin que estos datos fueran utilizados como insumo para alimentar los modelos en la siguiente etapa de modelado y desarrollo.

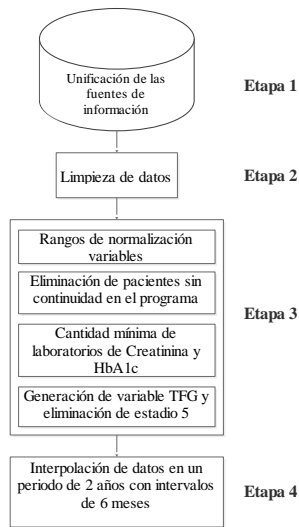


Fig. I. Etapas de limpieza y preparación de datos.

D. Modelos desarrollados

Se utilizaron diferentes algoritmos y configuraciones con el fin de evaluar sus ventajas sobre el problema de predicción propuesto en el documento. Para esto se seleccionaron 3 modelos: una técnica clásica de predicción XGBoost y dos técnicas de Deep learning: Perceptrón Multicapa (MLP) y redes neuronales convolucionales (CNN).

Para evaluar el mejor rendimiento de cada una de las técnicas de Deep learning se inició estableciendo configuraciones sencillas para evaluar su rendimiento al generar la predicción y se fue modificando la arquitectura de estas con el fin de obtener un mejor resultado ver Fig. II, para el caso de la modelo del Perceptrón Multicapa se generaron 3 configuraciones:

- Se establece una capa de entrada con las variables seleccionadas, una capa oculta donde se hace una evaluación de la cantidad de neuronas en un rango de (4-128), seguida de una capa de dropout en un rango de (0.2-0.5) y finaliza con una capa densa con activación lineal.
- Para esta configuración se adiciona una capa adicional a la estructura A acompañada de otra capa de dropout ambas con los mismos rangos de evaluación.
- Se ingresa un nuevo bloque con otra capa oculta al evidenciar una mejora en el comportamiento de la arquitectura B respecto a la A, pero no se obtiene una mejora significativa por lo que no se realiza configuraciones adicionales con más capas ocultas.

Para la arquitectura de las redes neuronales convolucionales se generaron dos configuraciones:

- Se establece la entrada con las variables seleccionadas, a una capa convolucional de una dimensión, esta capa crea un kernel de convolución que se convoluciona con la entrada de la capa sobre una sola dimensión espacial evaluada en un rango de (4-128), la salida pasa por un Maxpool de una dimensión y este a su vez por un Flatten para terminar en una capa densa con activación lineal.

- Para obtener mejores características se toma la arquitectura A de CNN y al finalizar se le incorpora una capa densa adicional en un rango de (4-128), seguida de una dropout antes de la capa densa de salida.

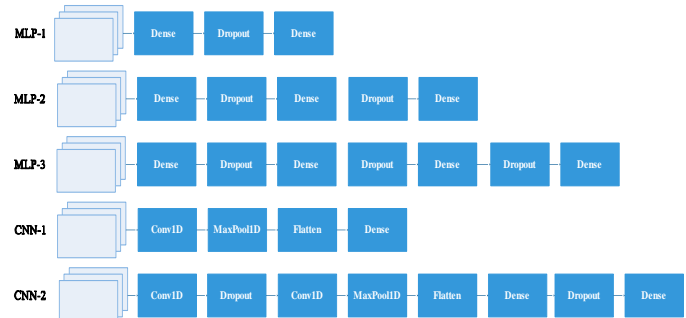


Fig. II. Arquitectura de los modelos evaluados.

Para medir el desempeño de cada uno de los modelos se utilizó la librería Optuna un framework de optimización de código abierto para automatizar la búsqueda de hiperparámetros, para esto en cada modelo fue utilizado este framework con el fin de optimizar esta labor de búsqueda para evaluar las mejores la configuración creando un objeto de estudio que invoque el método de optimización en 400 iteraciones, donde se pasa a una función objetiva y se proporciona una sugerencias de parámetros en la cual se pueden administrar el estado de la prueba para establecer u obtener atributos definidos de la mejor configuración obtenida, esto sin importar que sean parámetros continuos, discretos, flotantes o enteros, con la posibilidad de poder seleccionar el valor de muestreo del rango de parámetros a modificar. En la Tabla III se relacionan los hiperparámetros seleccionados para su evaluación en cada modelo y los mejores resultados obtenidos en el parámetro de objeto de estudio.

TABLE III
HIPERPARÁMETROS PARA SU EVALUACIÓN EN CADA MODELO

Modelo	Hiperparámetro	6 M	12M	18M	24 M	
XgBoost	colsample_bytree	1.0	1.0	1.0	1.0	
	gamma	0.5	0.5	0.5	1.5	
	learning_rate	0.01	0.1	0.1	0.01	
	max_depth	3	3	3	3	
	min_child_weight	1	1	5	1	
	n_estimators	500	50	50	500	
	subsample	0.8	0.8	0.8	0.6	
	Capa Densa	77	42	120	116	
MLP-1	Dropout	0.47668	0.20118	0.42117	0.45093	
		5417279	9455694	0927278	4628442	
		7115	95186	67095	23364	
	Learning Rate	0.00500	0.03690	0.00482	0.00339	
		8717226	1318668	1499521	0799093	
		219981	8181	830509	828015	
	MLP-2	Capa Densa	98	121	72	83
	Dropout	0.31194	0.31328	0.21147	0.21118	
		1823706	8694653	7339586	6747187	
		9129	41863	05136	2813	
MLP-3	Capa Densa	89	100	101	105	
	Dropout	0.48599	0.23044	0.49986	0.43966	
		8220526	9807587	6547084	8958330	
		8915	00762	86145	1298	
	Learning Rate	0.00305	0.00184	0.00196	0.00466	
		5207981	1162066	9933564	5967178	
		4559173	5679635	5745847	445578	
	MLP-3	Capa Densa	73	43	78	31
	Dropout	0.22514	0.46603	0.26005	0.24085	
		9073103	2577635	1771625	4534826	
CNN-1		16895	7962	1181	48078	
	Capa Densa	70	17	40	20	

	Dropout	0.47713	0.27669	0.43697	0.28032
		8745078	4686510	3273883	8484862
		8335	99033	20456	589
	Capa Densa	28	74	17	71
	Dropout	0.26436	0.31231	0.25288	0.21145
		1068009	4824520	3841585	2088988
		02004	1573	0234	50287
	Learning Rate	0.00237	0.00040	0.00352	0.21145
		4122116	3610181	8206118	2088988
		5409264	2002484	063573	50287
CNN-1	CNN 1D	18	105	25	65
	Learning Rate	0.00274	0.00786	0.00575	0.01264
		0882354	3194514	5759774	5763240
		8693813	254913	429311	824717
CNN-2	CNN 1D	94	91	72	109
	Dropout	0.28623	0.20880	0.28315	0.20711
		2375449	6870816	5934602	8454248
		3529	60865	40757	43078
	CNN 1D	14	58	15	21
	Capa Densa	37	112	17	22
	Dropout	0.27264	0.34755	0.47651	0.31677
		2911147	0878814	1089755	2586692
		5399	28953	26313	37274
	Learning Rate	0.00088	0.00397	0.00709	0.00076
		7848709	6322331	2553811	0103978
		2026174	653507	567256	1691324

TABLE IV
RESULTADOS EVALUACIÓN DE LOS MODELOS

Técnica	GRUPO DE Datos	Mae 6 meses	Mae 12 meses	Mae 18 meses	Mae 24 meses
XGBoost	Entrenamiento	5.43	6.79	5.80	7.63
XGBoost	Testeo	5.66	6.74	6.74	7.76
XGBoost	Validación	5.81	6.51	5.94	8.32
ANN-1	Entrenamiento	5.99	6.46	6.76	7.71
ANN-1	Testeo	5.40	5.96	6.29	6.66
ANN-1	Validación	5.61	5.67	6.32	6.34
ANN-2	Entrenamiento	6.11	6.65	6.35	7.72
ANN-2	Testeo	5.38	6.03	6.27	6.62
ANN-2	Validación	5.49	6.16	5.90	6.37
ANN-3	Entrenamiento	6.09	7.54	6.84	7.68
ANN-3	Testeo	5.43	7.07	6.33	6.73
ANN-3	Validación	5.38	8.07	6.30	6.49
CNN-1	Entrenamiento	6.05	6.48	6.24	7.74
CNN-1	Testeo	5.23	5.98	6.20	6.60
CNN-1	Validación	6.26	5.68	5.70	6.45
CNN-2	Entrenamiento	6.17	6.62	6.56	7.84
CNN-2	Testeo	5.77	6.02	6.32	6.65
CNN-2	Validación	6.37	5.99	6.42	6.29

La calibración de los hiperparámetros mediante sesión de optimización, es realizada mediante la evaluación del conjunto de datos de entrenamiento y evaluada en los datos de testeo, teniendo como métrica de medición el error absoluto medio en cada una de las iteraciones de búsqueda. Como resultado de los diferentes entrenamientos y configuraciones realizadas mediante la función objetivo se seleccionaron las mejores configuraciones obtenidas en cada modelo a partir de las métricas, posterior se procedió a usarlas para calcular el rendimiento en los diferentes grupos de datos como se evidencia en la Tabla IV.

III. RESULTADOS

En esta sección se presentan los resultados obtenidos en esta investigación referente a los mejores modelos, junto con una arquitectura e interfaz para la puesta en marcha.

A. Modelo

Para determinar si los resultados de cada uno de los modelos obtenidos en el paso anterior eran representativos se procedió a evaluar esto con una validación cruzada k-10 sobre los datos de testeo y entrenamiento véase Fig. III y Fig. IV.

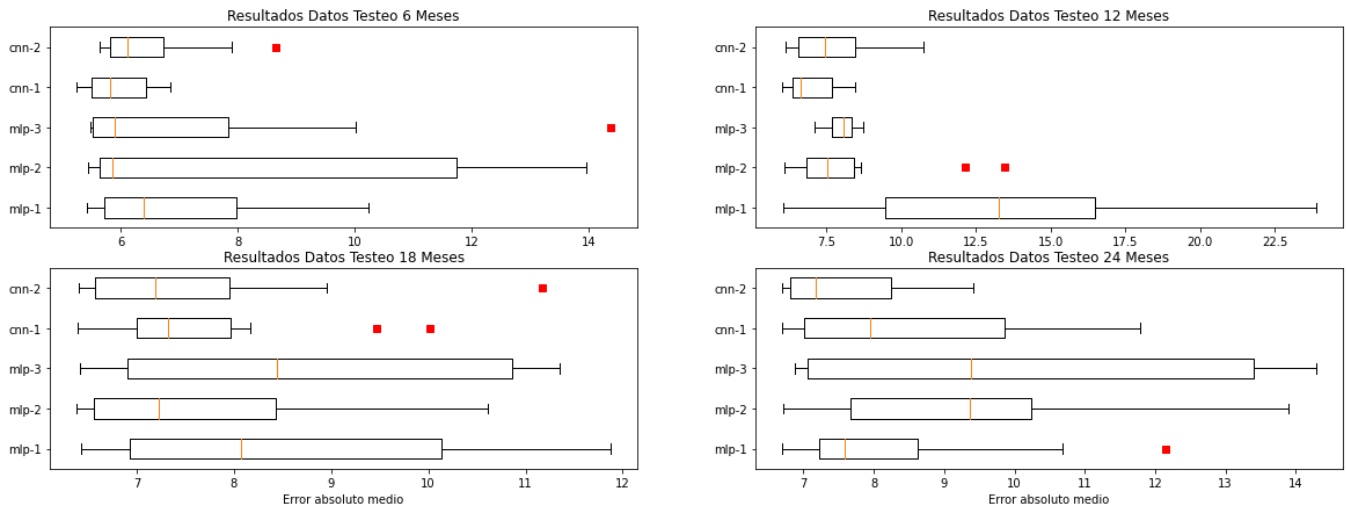
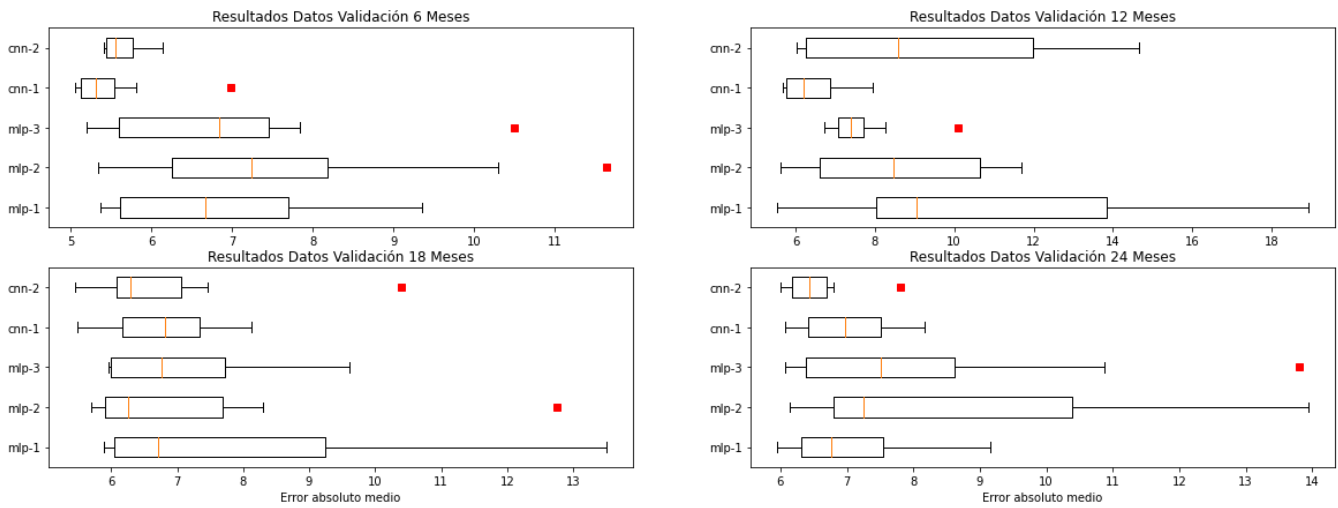


Fig. III. Resultados de modelos con datos de testeo en los 4 periodos de estudio.



Tras analizar los resultados se evidencia que los mejores rendimientos tanto para los datos de entrenamiento como de prueba se presentan en las configuraciones de los modelos de CNN, para el periodo de 6 a 12 meses el modelo CNN-1 es el que presenta mejor rendimiento con una media (mean) = 5.95%, una desviación estándar (std) = (+/- 0.54%) en 6 meses y mean = 7.00%, std= (+/- 0.84%) para 12 meses en prueba mientras que en validación mean=5.48%, std=(+/- 0.55%) en 6 meses y mean = 6.46%, std= (+/- 0.79%) a 12 meses. Mientras que en la ventana de 18 a 24 el modelo CNN-2 es el que presenta mejores resultados con un rendimiento de mean=7.63%, std= (+/- 1.41%) en 18 meses y mean=7.54%, std= (+/- 0.89%) para 24 meses en prueba mientras que en validación mean= 6.76%, std= (+/- 1.35%) en 18 meses y mean=6.53% , std= (+/- 0.50%) para 24 meses.

B. Desarrollo de interfaz de usuario

La Fig. V muestra la arquitectura implementada para el desarrollo de la API REST utilizada en esta investigación, el desarrollo del backend esta implementado en una lambda en Python.

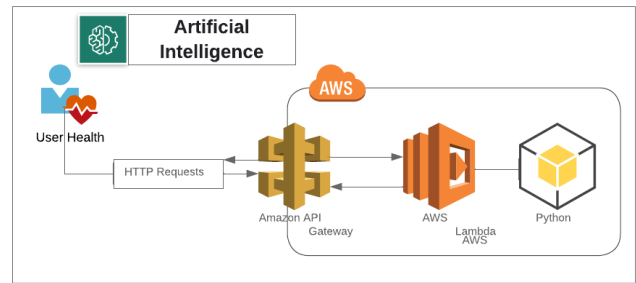


Fig. V. Arquitectura en AWS depuesta en marcha de modelo.

Como resultado de la interfaz del usuario se genera una aplicación Progressive Web App (pwa) la cual permite al usuario ingresar los valores actuales de las variables utilizadas para alimentar los modelos y esta realiza la petición a la Api, que procesa los 4 algoritmos para brindar una respuesta con los

valores predichos para 6, 12 ,18 y 24 meses los cuales se pueden visualizar en la gráfica, encontrando el estadio final predicho y la probabilidad de avance de estadio acompañada de una gráfica del daño renal acorde al estadio final del modelo.

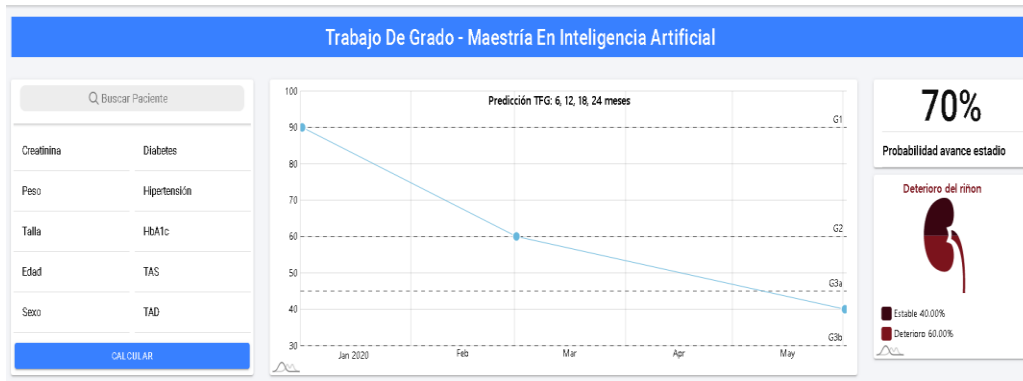


Fig. VI. Desarrollo interfaz de usuario de modelo de predicción.

IV. CONCLUSIONES

Tras analizar y evaluar los modelos de inteligencia artificial y considerar las variables mínimas para la predicción de enfermedad renal crónica, se obtuvieron dos modelos que generan una predicción de la enfermedad de un paciente en una ventana de dos años con un periodo de 6, 12, 18 y 24 meses. Esto mediante la implementación de dos arquitecturas de la técnica CNN con las cuales se evidenció que para los periodos de 6 a 12 meses se puede establecer una configuración reducida mientras que para los periodos 18 y 24 es necesario incluir una estructura más compleja para reconocer las características, debido a que al aumentar el tiempo la progresión de la enfermedad en los pacientes puede presentar comportamientos más aleatorios.

Entre las variables más destacadas del estudio se incluyeron la edad, diagnósticos de diabetes mellitus, hipertensión arterial y laboratorios clínicos los cuales pueden utilizarse como desenlace destacando entre ellos la hemoglobina glicosilada, creatinina y TFG, todas estas enfocadas como desencadenantes de la ERC, puesto que al presentar alguna comorbilidad la progresión de enfermedad es más rápida. El uso de modelos de aprendizaje automático podría usarse de manera eficiente en la gestión de pacientes en entidades prestadoras de servicio para mantener un control oportuno y así generar una vigilancia estrecha y detección temprana de ERC de los pacientes detectados con una progresión acelerada o como puerta de entrada para clasificación de los pacientes en las rutas de atención de cada prestador, con el fin de hacer detección y de esta manera cumplir con las metas exigidas por la cuenta de alto costo de la que hace parte esta patología.

Para la aplicación de modelos en la práctica clínica, el conjunto de variables utilizadas se tendría que ampliar para incluir otras mediciones de laboratorio y posiblemente se requeriría información sobre el estilo de vida de cada individuo a contemplar en un nuevo estudio, implementando un caso clínico para tener una observación en el tiempo y evaluar las predicciones propuestas.

V. LIMITACIONES Y TRABAJOS FUTUROS

Este estudio se basó en los datos extraídos de fuentes de información de un prestador de servicio del Caribe Colombiano, esto genera que no sea replicable por la zona geográfica, raza, hábitos entre otros, generando restricciones al momento de transferir los modelos para una población diferente.

La adaptación de los modelos a otra población requiere la incorporación de nuevos datos característicos de la población y la aplicación clínica requeriría más análisis y evaluación en un ensayo clínico, también se tendrían que evaluar los modelos en datos más recientes de la población para tener un seguimiento y comprobar la eficiencia con datos reales.

La incorporación de datos con una ventana más amplia de observación permitiría evaluar nuevas técnicas e incluir nuevas variables de estudio determinadas con los siguientes laboratorios: microalbuminuria, colesterol LDL, proteinuria,

entre otros. Esto permitiría evaluar no solo la progresión de la enfermedad de un paciente nuevo, si no contemplar una ventana de históricos para tener menor error en la predicción de los pacientes que ya se encuentran en el programa.

REFERENCES

- [1] B. Bikbov *et al.*, «Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017», *The Lancet*, vol. 395, n.º 10225, pp. 709-733, 2020.
- [2] J. Coresh, «Update on the burden of CKD», *J. Am. Soc. Nephrol.*, vol. 28, n.º 4, pp. 1020-1022, 2017.
- [3] J. W. Stanifer, M. Von Isenburg, G. M. Chertow, y S. Anand, «Chronic kidney disease care models in low-and middle-income countries: a systematic review», *BMJ Glob. Health*, vol. 3, n.º 2, p. e000728, 2018.
- [4] V. A. Luyckx, M. Tonelli, y J. W. Stanifer, «The global burden of kidney disease and the sustainable development goals», *Bull. World Health Organ.*, vol. 96, n.º 6, p. 414, 2018.
- [5] M. Wainstein *et al.*, «International Society of Nephrology Global Kidney Health Atlas: structures, organization, and services for the management of kidney failure in Latin America», *Kidney Int. Suppl.*, vol. 11, n.º 2, pp. e35-e46, 2021.
- [6] M. Tonelli y J. A. Dickinson, «Early detection of CKD: implications for low-income, middle-income, and high-income countries», *J. Am. Soc. Nephrol.*, vol. 31, n.º 9, pp. 1931-1940, 2020.
- [7] A. L. Beam y I. S. Kohane, «Big data and machine learning in health care», *Jama*, vol. 319, n.º 13, pp. 1317-1318, 2018.
- [8] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng, y B. Chen, «A machine learning methodology for diagnosing chronic kidney disease», *IEEE Access*, vol. 8, pp. 20991-21002, 2019.
- [9] K. Twarish Alhamazani, J. Alshudukhi, S. Aljaloud, y S. Abebaw, «Implementation of Machine Learning Models for the Prevention of Kidney Diseases (CKD) or Their Derivatives», *Comput. Intell. Neurosci.*, vol. 2021, 2021.
- [10] I. Ali y P. Kalra, «Risk prediction in chronic kidney disease», *Curr. Opin. Nephrol. Hypertens.*, vol. 28, n.º 6, pp. 513-518, 2019.
- [11] A. Smiti, «When machine learning meets medical world: Current status and future challenges», *Comput. Sci. Rev.*, vol. 37, p. 100280, 2020.
- [12] S. W. J. Nijman *et al.*, «Real-time imputation of missing predictor values improved the application of prediction models in daily practice», *J. Clin. Epidemiol.*, vol. 134, pp. 22-34, 2021.
- [13] C. Chang, Y. Deng, X. Jiang, y Q. Long, «Multiple imputation for analysis of incomplete data in distributed health data networks», *Nat. Commun.*, vol. 11, n.º 1, pp. 1-11, 2020.
- [14] B. K. Beaulieu-Jones, D. R. Lavage, J. W. Snyder, J. H. Moore, S. A. Pendergrass, y C. R. Bauer, «Characterizing and managing missing structured data in electronic health

records: data analysis», *JMIR Med. Inform.*, vol. 6, n.º 1, p. e8960, 2018.

[15] Y. Luo, P. Szolovits, A. S. Dighe, y J. M. Baron, «3D-MICE: integration of cross-sectional and longitudinal imputation for multi-analyte longitudinal clinical data», *J. Am. Med. Inform. Assoc.*, vol. 25, n.º 6, pp. 645-653, 2018.

[16] J. Hoogland *et al.*, «Handling missing predictor values when validating and applying a prediction model to new patients», *Stat. Med.*, vol. 39, n.º 25, pp. 3591-3607, 2020.

[17] A. T. Mohamed, S. Santhoshkumar, y V. Varadarajan, «INTELLIGENT DEEP LEARNING BASED PREDICTIVE MODEL FOR CORONARY HEART DISEASE AND CHRONIC KIDNEY DISEASE ON PEOPLE WITH DIABETES MELLITUS», *Malays. J. Comput. Sci.*, pp. 88-101, 2022.

[18] S. Akter *et al.*, «Comprehensive Performance Assessment of Deep Learning Models in Early Prediction and Risk Identification of Chronic Kidney Disease», *IEEE Access*, vol. 9, pp. 165184-165206, 2021.

[19] S. Krishnamurthy *et al.*, «Machine learning prediction models for chronic kidney disease using national health insurance claim data in Taiwan», 2021, vol. 9, n.º 5, p. 546.



William Amador. Ingeniería Electrónica en 2014 de la Universidad Manuela Beltrán en Bogotá – Colombia. Áreas de interés: Automatización, Machine learning y procesamiento de imágenes.



Leonardo Flórez Valencia. Ingeniería de Sistemas y Computación, Universidad de los Andes, Bogotá, Colombia, Maestría en Ingeniería de Sistemas, y Computación, Universidad de los Andes, Bogotá, Colombia, DEA Images et Systèmes INSA de Lyon, Francia, Doctorado en Imágenes y Sistemas, INSA de Lyon, Francia