

Estimación de lluvias mensuales promedio con regresión lineal múltiple y redes neuronales artificiales en una cuenca semiárida

José Armando Rodríguez Carrillo, Julián González Trinidad,
Gamaliel Moreno Chávez, Carlos Francisco Bautista Capetillo,
Hugo Enrique Júnez Ferreira, Luis Fernando Castillo Martínez,
Sandra Dávila Hernández

Universidad Autónoma de Zacatecas,
Unidad Académica de Ingeniería Eléctrica,
México

{jarmando.rc, jgonza, gamalielmch, baucap,
hejunez, fercast, sandra.davila}@uaz.edu.mx

Resumen. En las últimas décadas, la estimación de lluvias es crucial debido a la importancia del recurso hídrico para el ser humano. Conocer su comportamiento y distribución es esencial para investigadores de la hidrología, permitiendo prevenir desastres como sequías e inundaciones, y aprovechar la lluvia en campos como agricultura, ganadería, industria, hidráulica, entre otros. Técnicas de Machine Learning han sido empleadas para la estimación de este fenómeno, por su capacidad de trabajar con grandes volúmenes de datos, presentando resultados más precisos que los métodos convencionales. El área de estudio es una cuenca semiárida en Valparaíso, Zacatecas, donde hay 4 estaciones pluviométricas con registros de 31 años (1990-2020). Los algoritmos empleados fueron Regresión Lineal Múltiple y Redes Neuronales Artificiales. Las métricas de evaluación fueron Error Absoluto Medio, Raíz del Error Cuadrático Medio y Coeficiente de Determinación. En la mayoría de los casos, los modelos de Redes Neuronales presentan mejores resultados que los de Regresión Lineal Múltiple, con valores de hasta 0.834, 0.255 y 0.378 en Coeficiente de Determinación, Error Absoluto Medio y Raíz del Error Cuadrático Medio respectivamente. Los modelos de Redes Neuronales son considerados buenos estimadores de lluvia en la cuenca.

Palabras clave: Estimación de lluvia, aprendizaje profundo, cuenca semiárida.

Estimation of Average Monthly Rainfall with Multiple Linear Regression and Artificial Neural Networks in a Semi-Arid Basin

Abstract. In recent decades, rainfall estimation has become a worldwide trend due to the importance of water resources for human beings. Knowing its behavior and distribution is one of the most important tasks for hydrology researchers, since this allows preventing natural disasters such as droughts and floods, as well as taking advantage of rainfall for different fields such as agriculture, livestock,

industry, hydraulics, among others. Machine Learning techniques have been used for the estimation of this hydrological phenomenon, because they allow working with a large amount of data, as well as presenting more accurate results than conventional hydrological methods. The study area is a semi-arid basin in the municipality of Valparaíso, Zacatecas, where there are 4 rainfall stations with 31-year records (1990-2020). The algorithms used for this work were Multiple Linear Regression and Artificial Neural Networks. The evaluation metrics were the Mean Absolute Error, the Root Mean Square Error and the Coefficient of Determination. In most cases, the Artificial Neural Networks models perform better than the Multiple Linear Regression models obtaining values up to 0.834, 0.255 and 0.378 in Coefficient of Determination, Mean Absolute Error and Root Mean Squared Error respectively. Artificial Neural Network models are considered good estimators of rainfall in the basin.

Keywords: Rainfall estimation, deep learning, semi-arid basin.

1. Introducción

La lluvia es la principal fuente de agua que tiene la sociedad para su desarrollo social y económico, teniendo una influencia directa en las actividades como la agricultura, sector de salud pública, industrial y la producción de energía eléctrica, se requiere seguir las investigaciones científicas y el manejo del agua para optimizar su aprovechamiento [8, 13]. Conocer el comportamiento espacio-temporal de la lluvia ayuda a crear planes y estrategias para la prevención de desastres naturales, como inundaciones y sequías [17].

En los últimos años, el cambio climático ha ocasionado que la cantidad de lluvia que llega a la superficie sea menor [8]. La medición de datos de lluvia se logra a través de una estación meteorológica, la cual registra la cantidad e intensidad de esta. Conocer la información registrada da ventajas para poder aplicar métodos que permitan determinar las relaciones y tendencia de la lluvia a lo largo del tiempo.

Generalmente, la predicción de lluvia se realiza a partir de modelos numéricos climatológicos en conjunto con datos de radares meteorológicos, basándose estos métodos en regresiones lineales, métodos numéricos, promedios y fórmulas empíricas, además de correlaciones entre diferentes variables meteorológicas y geográficas.

Sin embargo, tener modelos con exactitud y rapidez que predigan la cantidad e intensidad de la lluvia en una determinada área ayuda a entender mejor el fenómeno [13]. La Inteligencia Artificial se ha aplicado a problemas hidrológicos debido a que se pueden detectar patrones y tendencias de la lluvia con el análisis de los datos, además de que permite trabajar con una gran cantidad de información, a diferencia de los métodos convencionales [11].

Los algoritmos de Machine Learning (ML, por sus siglas en inglés), específicamente las Redes Neuronales Artificiales (ANN, por sus siglas en inglés) se han popularizado debido a que estas han demostrado tener una gran capacidad de modelar la no linealidad de los patrones [2]. En el 2014, [9] implementaron un algoritmo de ANN para modelar el comportamiento lluvia-escorrentía en el sur de Australia, además, implementaron correlaciones entre el flujo de la cuenca y la lluvia.

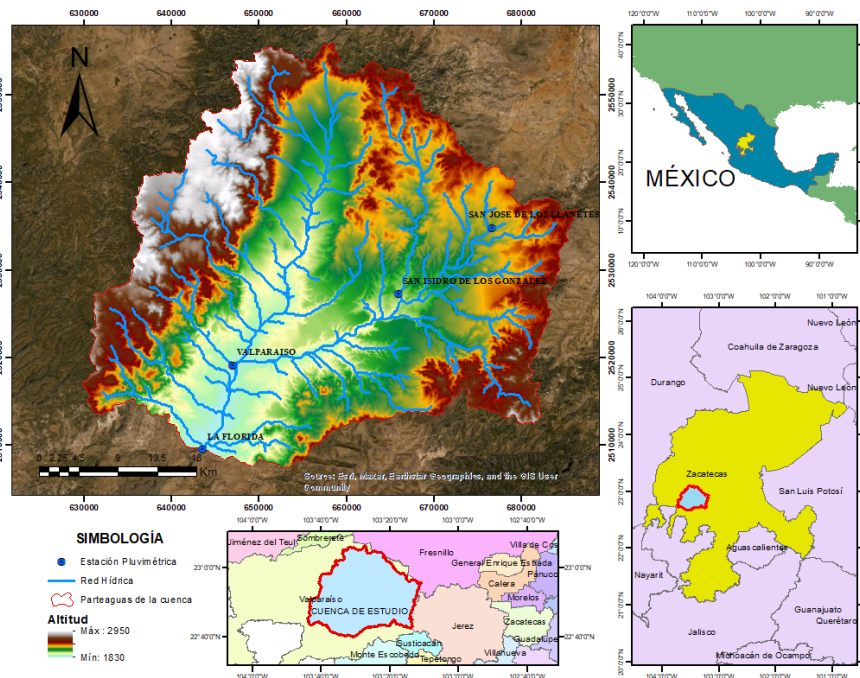


Fig. 1. Cuenca de Valparaíso, Zacatecas, México.

Por otro parte, [10] en 2018, propusieron un algoritmo de ANN para la predicción de lluvia, introduciendo como entradas del modelo índices climáticos, creando una red preliminar para identificar las relaciones de la lluvia con cada una de las variables para posteriormente, implementar un modelo con las características más significativas.

En 2019 [17], realizaron un modelo híbrido para la predicción de lluvia combinando dos modelos de pre-procesamiento de datos: un modelo ANN y Redes Neuronales Artificiales Estacionales evaluando el modelo en una ANN simple para posteriormente comparar los resultados con el Algoritmo Genético y el algoritmo de Reconocimiento Simulado con el modelo de Media Móvil Integrada Autorregresiva.

Igualmente, [13] en 2020 realizó una comparación de diferentes modelos de inteligencia artificial, los cuales fueron ANN, Máquina de Soporte Vectorial y Sistema de Inferencia Difusa basado en Red Adaptativa con Optimización de Enjambre de Partículas para la predicción diaria de lluvia en la provincia Hoa Binh, Vietnam, utilizando una base de datos de parámetros meteorológicos comprendidos entre la humedad relativa, velocidad del viento, temperatura máxima y temperatura mínima como datos de entrada y la lluvia diaria como salida.

De igual forma, [4] en 2020 implementaron algoritmos de aprendizaje automático para la predicción de lluvia diaria con término largo usando modelos lineales generalizados y algoritmos como Máquina de Soporte Vectorial, K- Vecino Más Próximo, Bosque Aleatorio, agrupamientos K-promedio y ANN, implementando patrones meteorológicos de bases de datos como entrenamiento del modelo.

Tabla 1. Coordenadas en proyección WGS84 UTM Zona 13N de las estaciones pluviométricas.

Estación Pluviométrica	Coordenada N	Coordenada Y	Coordenada Z
La Florida	2509447.13	643557.02	1830
Valparaíso	2519079.76	646947.2	1890
San Isidro de los González	2527237.95	665968.4	2034
San José de Llanetes	2534778.33	676596.17	2187

El objetivo principal de este estudio es estimar la lluvia mensual promedio aplicando los algoritmos de Regresión Lineal Múltiple (MLR, por sus siglas en inglés) y ANN a partir de datos de lluvia exclusivamente de 4 estaciones pluviométricas en la cuenca y un modelo general, que es una combinación de las 4 estaciones.

En la investigación se define el capítulo 2 con el área de estudio, mostrando algunas características geomorfológicas y sociales de la misma; en el capítulo 3 se muestran los materiales y métodos utilizados para el desarrollo de la investigación, desde la obtención de la base de datos, así como los modelos de ML implementados y sus métricas de evaluación; en el capítulo 4 se presentan los resultados obtenidos con los modelos de ML seleccionados y finalmente en el capítulo 5 las conclusiones.

2. Zona de estudio

La cuenca se localiza en el municipio de Valparaíso, Zacatecas, México (Figura 1), con clima semiárido con temperaturas que varía entre 12-24 °C, rango de lluvias de 500 a 1000 mm anuales principalmente en verano, con una población de aproximadamente 24,000 habitantes; dentro de la cuenca existen varios reservorios de agua, como la presa Manuel Pelgueres y presa El Salitre, entre otros.

La cuenca tiene un área de aproximadamente 1,771.19 km², con centroide en el eje X de 657526.08 y en el eje Y de 2529970.50 en la proyección WGS84 UTM Zona 13N (Tabla 1). Su geología comprende mayormente riolita-toba ácida (59.8 %), arenisca-conglomerado (31 %), basalto (3.4 %), aluvial (3 %), andesita (2 %), toba ácida (0.6 %), pórfido andesítico (0.3 %) y caliza-lutita (0.1 %).

El uso de suelo de la cuenca está compuesto principalmente por vegetación secundaria arbustiva pastizal natural (20.9 %), agricultura de temporal anual (19 %), bosque de pino-encino (14.8 %), vegetación secundaria arbustiva de bosque de encino-pino (8.8 %), vegetación secundaria arbustiva de bosque de encino (6.6 %), bosque de encino-pino (5.4 %), vegetación secundaria arbustiva de bosque de pino-encino y bosque de encino (5.3 %), pastizal inducido (4.1 %), agricultura de riego anual (2.6 %), matorral crasicaule (2.3 %), pastizal natural (1.9 %), vegetación secundaria arbustiva de selva baja caducifolia (1 %), vegetación secundaria arbórea de bosque de pino-encino (0.6 %), asentamientos humanos (0.5 %), agricultura de temporal permanente (0.4 %), vegetación secundaria arbustiva de bosque de pino (0.2 %), agricultura de temporal anual y permanente con los cuerpos de agua (0.1 %), y vegetación secundaria arbórea de bosque de encino-pino y bosque de pino (0.044 %).

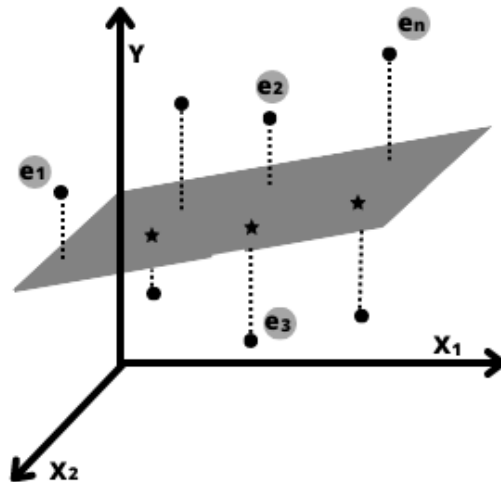


Fig. 2. Representación de MLR con dos variables independientes (x_1) y (x_2), donde se mapean las variables de entrada y se visualizan los residuos con respecto al modelo predicho [1].

La edafología está conformada por leptosol (26.02 %), luvvisol (20.80 %), phaeozem (18.96 %), cambisol (14.87 %), chernozem (9.52 %), kastañozem (5.12 %), Fluvisol (3.50 %), durizol (0.74 %), regosol (0.22 %), suelo ocupado por localidades (0.15 %) y cuerpos de agua (0.09 %).

3. Materiales y métodos

La metodología empleada en la investigación consistió en la obtención de los datos de lluvia registrada en los pluviómetros, reportados por la Comisión Nacional del Agua (CONAGUA), así como la aplicación de los métodos de ML para la estimación y sus métricas de evaluación. Se utilizó una validación cruzada de K-fold = 5, con un conjunto de entrenamiento de 70 % de los datos y el resto para prueba de acuerdo al criterio de muestreo aleatorio simple [19].

3.1. Obtención de los datos

Para delimitar el área de estudio se partió del Sistema Digital de Elevaciones del Instituto Nacional de Estadística y Geografía [7], posteriormente se descargó el índice de estaciones meteorológicas reportado por CONAGUA [3] para el software Google Earth Pro, esto con el fin de identificar las estaciones que se encontraban dentro de la cuenca (La Florida (LF), Valparaíso (VP), San Isidro de los González (SI) y San José de Llanetes (LL)), se extrajeron los datos de lluvia mensual promedio para el período 1990-2020, como se observó que faltaban datos, estos se calcularon con el método racional deductivo. Completados los datos se generó la base de datos con una combinación de entradas (3 estaciones) y una como salida.

Tabla 2. Resultados de los modelos de MLR.

Estación Pluviométrica	R^2	MAE	RMSE
Modelo General	0.757	0.295	0.499
La Florida	0.772	0.274	0.443
Valparaíso	0.759	0.293	0.485
San Isidro de los González	0.790	0.268	0.494
San José de Llanetes	0.767	0.274	0.455

3.2. Regresión lineal múltiple

El modelo MLR (ecuación 1) fue el primer algoritmo de ML y el más utilizado debido a su facilidad de aplicación, considera la relación lineal entre la variable dependiente e independiente [18]. La Figura 2 representa el mapeo de las variables y sus residuales con respecto al modelo MLR:

$$y = \beta_0 + \beta_1 \cdot x_1 + \dots + \beta_n \cdot x_n + \epsilon, \quad (1)$$

donde y es la variable dependiente, x_i se refiere a la variable independiente, β_i es el parámetro y ϵ el error del dependiente de (x).

3.3. Redes neuronales artificiales

Una ANN es un modelo de Aprendizaje Profundo que está inspirado en la estructura del cerebro humano para la clasificación y reconocimiento de patrones, con el fin de realizar predicciones. Generalmente se utiliza una arquitectura de Perceptrón Multi-Capa para aumentar la exactitud del modelo [6].

Las neuronas son unidades computacionales las cuales contienen una función de activación que discriminan la información obtenida de otras neuronas y la selección de estas funciones dependen del problema y pueden cambiar en las diferentes capas de los modelos. Las neuronas están conectadas por pesos, estos representan el axón de las neuronas biológicas, que se estiman aleatoriamente mediante propagación hacia adelante y se actualizan con la propagación hacia atrás, utilizando un optimizador, el cual se encarga de buscar el mínimo de la función, con el fin de reducir el error que generó el modelo [5, 15] y esto se expresa matemáticamente en la ecuación 2, donde b es el bias del modelo, x_n las variables de entrada y w_n los pesos sinópticos de la red. En la figura 3 se muestra el diagrama de una ANN:

$$y = f\left(b + \sum_{i=1}^n x_n w_n\right). \quad (2)$$

3.4. Parámetros de ANN

Todos los modelos de ANN fueron entrenados con la función de pérdida de “Mean Squared Error”, el optimizador aplicado fue “Adam”, el cual tuvo los parámetros de tasa de aprendizaje de 0.001, un valor en β_1 de 0.9, β_2 de 0.999 y un valor de condición de paro epsilon de 1 e-07.

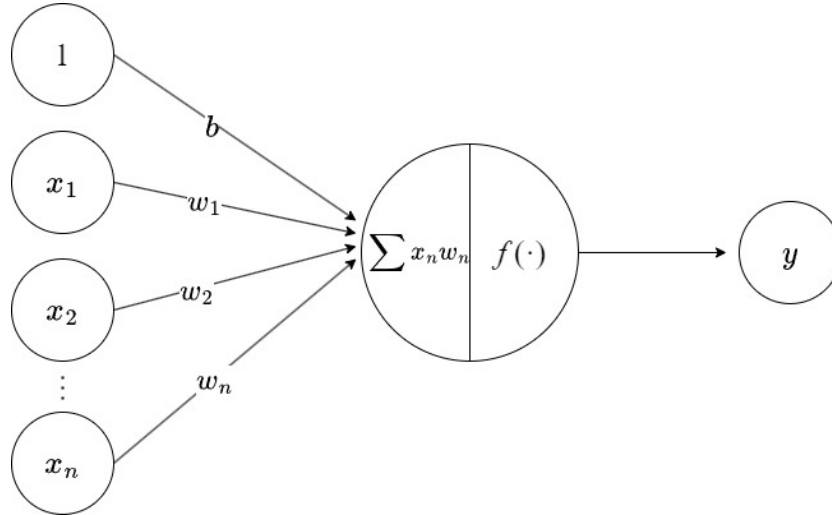


Fig. 3. Representación de una unidad de una ANN [16].

La arquitectura se basó en una capa de entrada con 12 neuronas usando la función de activación “ReLU”, una capa oculta de 6 neuronas con la función de activación “ReLU” y una capa de salida con una neurona usando la función de activación “lineal” con 150 épocas para cada modelo.

3.5. Métricas de evaluación

Las métricas utilizadas para la evaluación de los modelos son el Error Absoluto Medio (MAE, por sus siglas en inglés), en la ecuación 3, la Raíz del Error Cuadrático Medio (RMSE, por sus siglas en inglés), en la ecuación 4, finalmente utilizando el Coeficiente de Determinación (R^2), reportado en la ecuación 5 [12]:

$$\text{MAE} = \frac{\sum_{i=1}^n |E_i|}{n}, \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n E_i^2}{n}}, \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (P_i - A_i)}{\sum_{i=1}^n (A_i - \bar{A}_i)}, \quad (5)$$

donde A_i es el valor verdadero, P_i es el valor predicho, E_i es la diferencia entre el valor predicho y el valor verdadero, \bar{A}_i es el promedio de los valores verdaderos y n se refiere al número total de valores.

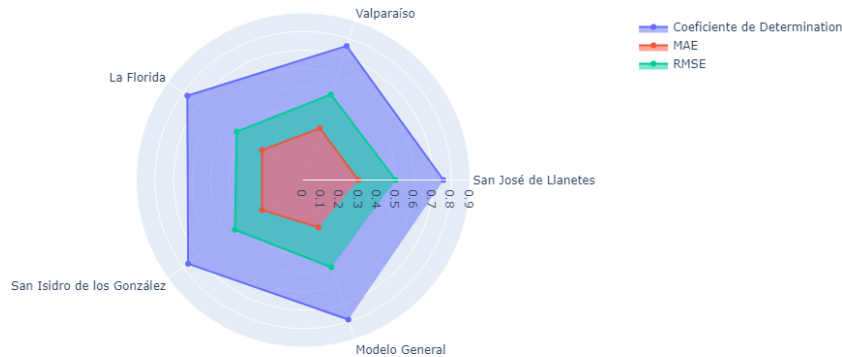


Fig. 4. Comportamiento de R^2 , MAE y RMSE para los modelos de MLR.

4. Resultados

Se crearon 5 bases de datos diferentes, la primera fue con la estación de LF y las 3 estaciones restantes como entrada; la segunda fue implementada como salida la estación VP y las 3 restantes como entrada; la tercera tuvo como salida la estación SI y las 3 restantes como entrada; la cuarta con los datos de LL como salida y las tres restantes como entradas; el quinto modelo es la unión de las cuatro anteriores, de tal forma que los datos se cuadruplican, llamándolo Modelo General (MG). Para cada base de datos se creó un modelo MLR y un modelo ANN.

4.1. Resultados de MLR

Los resultados se pueden visualizar en la Tabla 2. Para el algoritmo MLR el modelo que mejor resultados dio fue en la estación SI con un R^2 de **0.790** y un MAE de **0.268**. El mejor puntaje en RMSE fue la estación de LF con **0.443**. El MG fue de los mejores valores más bajos R^2 de 0.757, un MAE de 0.295 y un RMSE de 0.499. Las variaciones entre modelos fue del rango de 0.033 en R^2 , 0.027 en MAE y 0.056 en RMSE Figura 4, donde se puede visualizar que existe una uniformidad entre los resultados, pues estos varían muy poco entre ellos, en cuanto a sus métricas de evaluación.

4.2. Resultados de ANN

Los resultados se muestran en la Tabla 3. El modelo que mejor resultados fue para la estación de LL con R^2 de **0.834**, **0.255** en MAE y **0.378** en RMSE. El modelo con los rendimientos más bajos en R^2 y MAE fue la estación de VP con 0.758 y 0.295 respectivamente mostrando la variación espacial y temporal de la lluvia.

El mayor error en RMSE se encontró fue la estación SI con 0.496, lo que indica la poca relación de los años analizados. Los modelos varían en un rango de 0.076 en R^2 , 0.040 en MAE y 0.118 en RMSE Figura 5, donde se puede visualizar de una mejor manera la relación que existe entre R^2 con MAE y RMSE, ya que mientras los últimos dos mencionados disminuyen, el primer R^2 aumenta.

Tabla 3. Resultados de los modelos ANN.

Estación Pluviométrica	R^2	MAE	RMSE
Modelo General	0.791	0.269	0.490
La Florida	0.810	0.266	0.404
Valparaíso	0.758	0.295	0.486
San Isidro de los González	0.788	0.257	0.496
San José de Llanetes	0.834	0.255	0.378

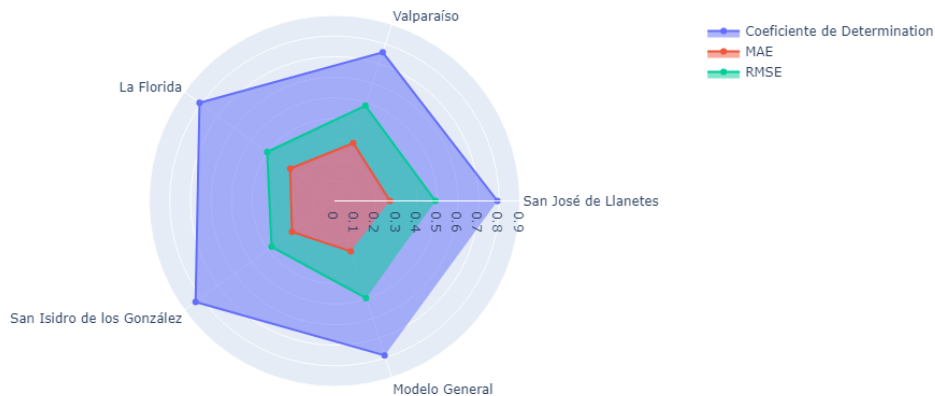


Fig. 5. Comportamiento de R^2 , MAE y RMSE para los modelos de ANN.

Autores como [6, 14], han reportado valores similares a los encontrados en esta investigación, sobre todo en los valores de R^2 . Sin embargo difieren de [11, 5], los cuales indican que se pueden mejorar algunas métricas de las utilizadas para regiones similares a esta, aunque puede influir que ellos utilizan datos diarios y otras características de entrada.

4.3. Comparación de los modelos de MLR y ANN

En la figura 6 se muestran las comparaciones de las métricas para ambos modelos. Para R^2 de los modelos MLR y ANN (a), donde se puede observar que para el MG, el modelo ANN superó al modelo MLR con 3.4 %; para la estación de LF, el modelo ANN también mejoró su rendimiento con respecto al modelo MLR con 3.8 %; caso contrario con la estación de VP y SI, donde los modelos ANN bajaron ligeramente su rendimiento con 1 % y 2 % respectivamente; finalmente la estación de LL, mostró el mejor desempeño para esta métrica, aumentando un 6.7 %.

Cabe destacar que para esta métrica, los modelos ANN mejoraron considerablemente en la estación LL, LF y en el MG. Los resultados de las diferencias de MAE de los modelos (b) demuestran que el resultado MG-ANN-MAE fue mejor que MG-MLR-MAE con 0.269 y 0.295 respectivamente; LF-ANN-MAE lanzó el resultado 0.266, siendo mejor que LF-MLR-MAE, pues este valor fue de 0.274; comparando VP-ANN-MAE y VP-MLR-MAE, el mejor fue el último modelo mencionado con 0.293, a diferencia de 0.295; se obtuvo un valor de 0.268 en SI-MLR-MAE, siendo

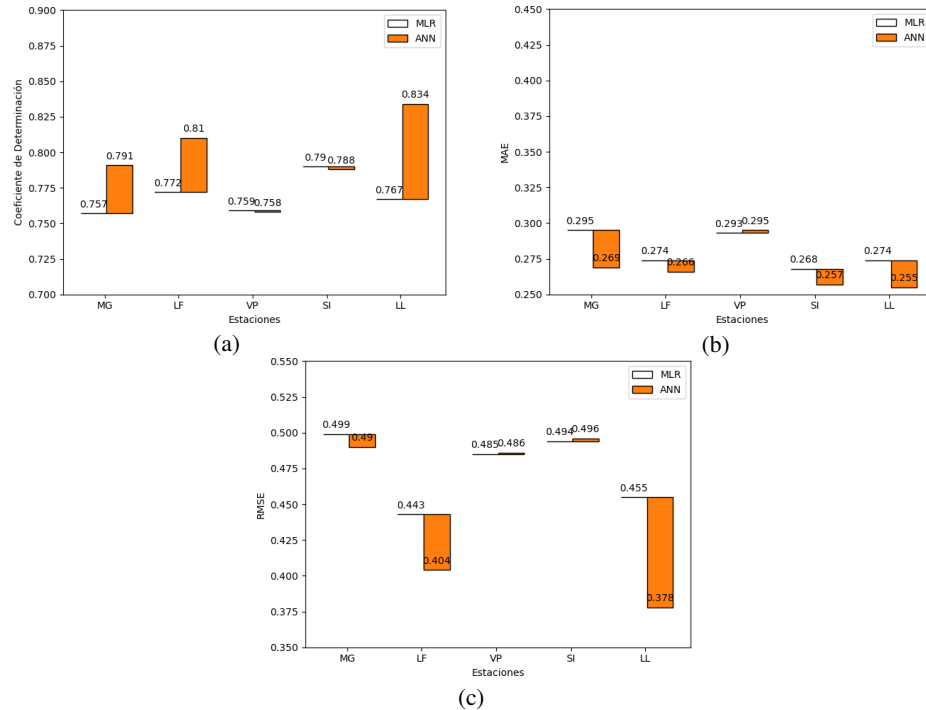


Fig. 6. Comparación de las métricas de los modelos MLR vs ANN.

mejorado por SI-ANN-MAE con 0.257; finalmente con LL-MLR-MAE el valor fue de 0.274, comparándolo con LL-ANN-MAE, que fue mejor con 0.255, además de que fue la mejor evaluación en MAE para todos los modelos implementados. En MAE los modelos ANN se demostraron muy similares e incluso en una estación siendo superado por los modelos MLR.

En la métrica MG-ANN-RMSE y MG-MLR-RMSE los resultados fueron 0.490 y 0.499, denotando una ligera mejora en el modelo ANN; para LF-MLR-RMSE el valor fue de 0.443, superado por LF-ANN-RMSE con 0.404; en VP-MLR-RMSE (0.485) y VP-ANN-RMSE (0.486) los valores fueron muy similares, con una diferencia mínima de 0.001; en SI-MLR-RMSE el valor fue de 0.494 y con menor rendimiento SI-ANN-RMSE fue de 0.496; en la estación LL, los resultados con LL-MLR-RMSE y LL-ANN-RMSE fueron 0.455 y 0.378, respectivamente, donde el modelo ANN para esta estación fue el mejor resultado en RMSE de todos los modelos. Para esta métrica, las estaciones LF y LL mejoraron considerablemente, en comparación con el resto de estaciones, esto se puede observar en (c).

Los datos obtenidos dan elementos para entender el comportamiento espacio-temporal de la lluvia en un área estudiada de 1,771 km², mostrando que los registros de lluvia en las estaciones son diferentes, además, influenciados por la altitud. Por otro lado, el modelo ANN, ajusta mejor los datos para estimar con menor incertidumbre la lluvia, logrando con ello obtener datos más confiables para el cálculo del volumen de agua disponible.

5. Conclusiones

En esta investigación se planteó como objetivo estimar la lluvia mensual promedio aplicando los algoritmos de MLR y ANN a partir de datos de lluvia registrados en 4 estaciones pluviométricas de una cuenca semiárida y un modelo general, que es una combinación de las 4 estaciones.

El modelo ANN mostró los mejores resultados con respecto a R^2 de 0.834 para la estación LL, con una mínima diferencia (0.049), entre las estaciones VP, LF y SI, esta pequeña diferencia se puede asumir que es debida a la variación espacio-temporal de la lluvia. El MAE y RMSE mostró un comportamiento similar a R^2 con respecto a las estaciones, con un rango de 0.255-0.295 y 0.378-0.496 respectivamente.

El modelo general de ANN mejora la estimación de lluvia en la cuenca, con respecto al MLR, por lo tanto se sugiere que la estimación de la lluvia podría realizarse con el método ANN con menor incertidumbre. Se recomienda que trabajos futuros se enfoquen a implementar esta metodología con más algoritmos de ML, con intervalos de tiempo más cortos (diarios y anuales).

Agradecimientos. Los autores agradecen al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico a través de la convocatoria “Becas Nacional (Tradicional) 2021-3”, para el alumno José Armando Rodríguez Carrillo, facilitando el desarrollo de la presente investigación y a la Comisión Nacional del Agua (CONAGUA) por los datos proporcionados.

Referencias

1. Carrasquilla-Batista, A., Chacón-Rodríguez, A., Núñez-Montero, K., Gómez-Espinoza, O., Valverde, J., Guerrero-Barrantes, M.: Regresión lineal simple y múltiple: Aplicación en la predicción de variables naturales relacionadas con el crecimiento microalgal. *Revista Tecnología en Marcha*, vol. 29, pp. 33–45 (2016) doi: 10.18845/tm.v29i8.2983
2. Chiacchiera, A., Sai, F., Salvetti, A., Guariso, G.: Neural structures to predict river stages in heavily urbanized catchments. *Water*, vol. 14, no. 15, pp. 2330 (2022) doi: 10.3390/w14152330
3. CONAGUA: Comisión nacional del agua. (2023)
4. Diez-Sierra, J., Del Jesus, M.: Long-term rainfall prediction using atmospheric synoptic patterns in semi-arid climates with statistical and machine learning methods. *Journal of Hydrology*, vol. 586, pp. 124–789 (2020) doi: 10.1016/j.jhydrol.2020.124789
5. Endalie, D., Haile, G., Taye, W.: Deep learning model for daily rainfall prediction: Case study of Jimma, Ethiopia. *Water Supply*, vol. 22, no. 3, pp. 3448–3461 (2022) doi: 10.2166/ws.2021.391
6. Gu, J., Liu, S., Zhou, Z., Chalov, S. R., Zhuang, Q.: A stacking ensemble learning model for monthly rainfall prediction in the Taihu Basin, China. *Water*, vol. 14, no. 3, pp. 492 (2022) doi: 10.3390/w14030492
7. INEGI: Instituto nacional de estadística, geografía e informática (2023)
8. Kala, A., Vaidyanathan, S. G.: Prediction of rainfall using artificial neural network. In: *International Conference on Inventive Research in Computing Applications*, pp. 339–342 (2018)

9. Kamruzzaman, M., Shahriar, M. S., Beecham, S.: Assessment of short term rainfall and stream flows in South Australia. *Water*, vol. 6, no. 11, pp. 3528–3544 (2014) doi: 10.3390/w6113528
10. Lee, J., Kim, C. G., Lee, J. E., Kim, N. W., Kim, H.: Application of artificial neural networks to rainfall forecasting in the Geum River Basin, Korea. *Water*, vol. 10, no. 10, pp. 1448 (2018) doi: 10.3390/w10101448
11. Liyew, C. M., Melese, H. A.: Machine learning techniques to predict daily rainfall amount. *Journal of Big Data*, vol. 8, pp. 1–11 (2021) doi: 10.1186/s40537-021-00545-4
12. Naser, M. Z., Alavi, A.: Insights into performance fitness and error metrics for machine learning (2020) doi: 10.48550/arXiv.2006.00887
13. Pham, B. T., Le, L. M., Le, T. T., Bui, K. T. T., Le, V. M., Ly, H. B., Prakash, I.: Development of advanced artificial intelligence models for daily rainfall prediction. *Atmospheric Research*, vol. 237 (2020) doi: 10.1016/j.atmosres.2020.104845
14. Ridwan, W. M., Sapitang, M., Aziz, A., Kushiar, K. F., Ahmed, A. N., El-Shafie, A.: Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. *Ain Shams Engineering Journal*, vol. 12, no. 2, pp. 1651–1663 (2021) doi: 10.1016/j.asej.2020.09.011
15. Sarasa-Cabezuelo, A.: Prediction of rainfall in Australia using machine learning. *Information*, vol. 13, no. 4, pp. 163 (2022) doi: 10.3390/info13040163
16. Sharma, P.: Introduction to the neural network model, glossary, and backpropagation (2022) www.analyticsvidhya.com/blog/2022/01/introduction-to-the-neural-network-model-glossary-and-backpropagation/
17. Tran-Anh, D., Duc-Dang, T., Pham-Van, S.: Improved rainfall prediction using combined pre-processing methods and feed-forward neural networks. *J — Multidisciplinary Scientific Journal*, vol. 2, no. 1, pp. 65–83 (2019) doi: 10.3390/j2010006
18. Uyanık, G. K., Güler, N.: A study on multiple linear regression analysis. *Procedia-Social and Behavioral Sciences*, vol. 106, pp. 234–240 (2013) doi: 10.1016/j.sbspro.2013.12.027
19. Verma, S. P.: *Estadística básica para el manejo de datos experimentales: Aplicación en la geoquímica (geoquimiometría)*. Universidad Nacional Autónoma de México (2005)