



**UNIVERSIDAD NACIONAL DEL ALTIPLANO**  
**FACULTAD DE INGENIERÍA MECÁNICA ELÉCTRICA,**  
**ELECTRÓNICA Y SISTEMAS**  
**ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**



**DESARROLLO DE MODELOS DE REDES NEURONALES**  
**RECURRENTES PARA EL PRONÓSTICO DE NIVELES DE AGUA**  
**DEL LAGO TITICACA**

**TESIS**

**PRESENTADA POR:**

**RENE LUJANO LAURA**

**PARA OPTAR EL TÍTULO PROFESIONAL DE:**

**INGENIERO DE SISTEMAS**

**PUNO – PERÚ**

**2024**



## Reporte de similitud

NOMBRE DEL TRABAJO

**DESARROLLO DE MODELOS DE REDES NEURONALES RECURRENTE PARA EL PRONÓSTICO DE NIVELES DE AGUA DEL LAGO TITICACA**

AUTOR

**RENE LUJANO LAURA**

RECUENTO DE PALABRAS

**11690 Words**

RECUENTO DE CARACTERES

**63564 Characters**

RECUENTO DE PÁGINAS

**68 Pages**

TAMAÑO DEL ARCHIVO

**3.4MB**

FECHA DE ENTREGA

**Jan 19, 2024 9:28 AM GMT-5**

FECHA DEL INFORME

**Jan 19, 2024 9:29 AM GMT-5**

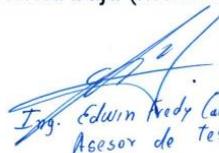
### ● 12% de similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos:

- 10% Base de datos de Internet
- Base de datos de Crossref
- 6% Base de datos de trabajos entregados
- 2% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossref

### ● Excluir del Reporte de Similitud

- Material bibliográfico
- Coincidencia baja (menos de 12 palabras)
- Material citado

  
Ing. Edwin Freddy Calderon Vilca  
Asesor de Tesis

  
Dr. Ing. Henry Iván Condori Alejo  
CIP: 56684  
Director de Postgrado y Fimccs  
Y.B.

Resumen



## DEDICATORIA

A mis padres, por haberme apoyado a lo largo de todos mis estudios, por ser los principales motores de alcanzar mis anhelos y sueños, por apoyarme incondicionalmente y por confiar siempre en mí.

A mis hermanos con mucha gratitud por su apoyo incondicional para que este trabajo sea realizado.

René Lujano Laura



## AGRADECIMIENTOS

A Dios, que con su bondad y amor me ha dado las fuerzas y la salud por haber podido concluir mis estudios y haberme permitido alcanzar uno de los mejores logros a lo largo de mi vida.

A la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional del Altiplano - Puno por la oportunidad que me brindó para realizar mis estudios.

A los docentes de la Escuela profesional de Ingeniería de Sistemas, por haber compartido sus conocimientos a lo largo de la vida universitaria.

Al Dr. Edwin Fredy Calderón Vilca, por aceptar la dirección de esta tesis, y su participación en el asesoramiento del desarrollo del presente trabajo de investigación.

Al D.Sc. Efrain Lujano Laura y M.Sc. Apolinario Lujano Laura por su invaluable orientación y apoyo durante el desarrollo de esta tesis. Su experiencia, paciencia y dedicación han sido fundamentales para la culminación de este trabajo.



# ÍNDICE GENERAL

	Pág.
<b>DEDICATORIA</b>	
<b>AGRADECIMIENTOS</b>	
<b>ÍNDICE GENERAL</b>	
<b>ÍNDICE DE FIGURAS</b>	
<b>ÍNDICE DE TABLAS</b>	
<b>ÍNDICE DE ANEXOS</b>	
<b>ACRÓNIMOS</b>	
<b>RESUMEN .....</b>	<b>13</b>
<b>ABSTRACT .....</b>	<b>14</b>
<b>CAPÍTULO I</b>	
<b>INTRODUCCIÓN</b>	
<b>1.1 PLANTEAMIENTO DEL PROBLEMA .....</b>	<b>15</b>
<b>1.2 FORMULACIÓN DEL PROBLEMA .....</b>	<b>16</b>
1.2.1 Problema general .....	16
1.2.2 Problema específico .....	16
<b>1.3 HIPÓTESIS DE LA INVESTIGACIÓN .....</b>	<b>17</b>
<b>1.4 JUSTIFICACIÓN DEL ESTUDIO .....</b>	<b>17</b>
<b>1.5 OBJETIVOS DE INVESTIGACIÓN .....</b>	<b>18</b>
1.5.1 Objetivo general .....	18
1.5.2 Objetivos específicos .....	18

## CAPÍTULO II

### REVISIÓN DE LITERATURA



<b>2.1 ANTECEDENTES .....</b>	<b>19</b>
2.1.1 Nivel internacional .....	19
2.1.2 Nivel nacional .....	20
2.1.3 Nivel regional/local .....	20
<b>2.2 MARCO TEORICO .....</b>	<b>21</b>
2.2.1 Inteligencia artificial .....	21
2.2.2 Aprendizaje automático .....	22
2.2.3 Aprendizaje profundo .....	23
2.2.4 Redes neuronales recurrentes (RNN) .....	25
2.2.5 Redes Long-Short Term Memory (LSTM) .....	27

### **CAPÍTULO III**

#### **MATERIALES Y MÉTODOS**

<b>3.1 ÁREA DE ESTUDIO .....</b>	<b>31</b>
<b>3.2 TIPO Y NIVEL DE LA INVESTIGACIÓN .....</b>	<b>32</b>
<b>3.3 TÉCNICA E INSTRUMENTOS .....</b>	<b>33</b>
3.3.1 Adquisición de datos .....	33
<b>3.4 PROCEDIMIENTO METODOLÓGICO .....</b>	<b>34</b>
3.4.1 Selección de las características importantes que influyen en los niveles de agua del lago Titicaca. ....	34
3.4.2 Pronóstico diario de niveles de agua del Lago Titicaca basada en el modelo RNN-LSTM .....	36
3.4.2.1 Preprocesamiento de datos .....	36
3.4.2.2 Dividir el conjunto de datos de entrenamiento y prueba .....	36
3.4.2.3 Definición de los hiperparámetros .....	36



3.4.3	Pronóstico mensual de niveles de agua del Lago Titicaca basada en el modelo RNN-LSTM .....	38
3.4.4	Métricas de bondad de ajuste .....	39
3.4.4.1	Raíz del error medio cuadrático (RMSE) .....	39
3.4.4.2	Eficiencia de Nash-Sutcliffe (NSE) .....	39
3.4.4.3	Eficiencia de Kling-Gupta (KGE') .....	40
3.4.4.4	Sesgo porcentual (PBIAS) .....	41

## CAPÍTULO IV

### RESULTADOS Y DISCUSIÓN

<b>4.1</b>	<b>CARACTERÍSTICAS IMPORTANTES QUE INFLUYEN EN LOS NIVELES DE AGUA DEL LAGO TITICACA .....</b>	<b>42</b>
4.1.1	Análisis de normalidad .....	42
4.1.2	Análisis de correlación .....	43
<b>4.2</b>	<b>PRONÓSTICO DIARIO DE NIVELES DE AGUA DEL LAGO TITICACA BASADA EN EL MODELO RNN-LSTM .....</b>	<b>46</b>
4.2.1	Efecto del tamaño de la ventana .....	46
<b>4.3</b>	<b>PRONÓSTICO MENSUAL DE NIVELES DE AGUA DEL LAGO TITICACA BASADA EN EL MODELO RNN-LSTM .....</b>	<b>52</b>
<b>V.</b>	<b>CONCLUSIONES .....</b>	<b>58</b>
<b>VI.</b>	<b>RECOMENDACIONES .....</b>	<b>59</b>
<b>VII.</b>	<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>60</b>
<b>ANEXOS</b>	<b>.....</b>	<b>66</b>

**ÁREA:** Inteligencia artificial

**TEMA:** Deep learning

**FECHA DE SUSTENTACIÓN:** 26 de enero del 2024



## ÍNDICE DE FIGURAS

	<b>Pág.</b>
<b>Figura 1</b> Diferentes tipos y algoritmos de aprendizaje automático.....	23
<b>Figura 2</b> Ilustración de la posición del aprendizaje profundo (DL), en comparación con el aprendizaje automático (ML) y la inteligencia artificial (IA).....	24
<b>Figura 3</b> Taxonomía de técnicas de DL, divididas en tres categorías principales (i) redes profundas para aprendizaje supervisado o discriminativo, (ii) redes profundas para aprendizaje no supervisado o generativo, y (ii) redes profundas para aprendizaje híbrido y otras relevantes. ....	25
<b>Figura 4</b> Procesamiento secuencial en una Red Neuronal Recurrente (RNN) .....	26
<b>Figura 5</b> Estructura de la red neuronal Long Short-Term Memory (LSTM) (Yan, ..... 2015) y adaptado por (Le et al., 2019). ....	27
<b>Figura 6</b> Distribución espacial del LT y ubicación de la estación hidrométrica muelle ENAFER.....	32
<b>Figura 7</b> Hidrograma de NALT – Estación muelle ENAFER a) diario y b) mensual. ....	34
<b>Figura 9</b> Análisis de normalidad para datos diarios (a-c) y mensuales (d-f) de NALT .....	43
<b>Figura 10</b> Matriz de correlación de Spearman entre diferentes ventanas de tiempo de niveles diarios de agua del LT.....	45
<b>Figura 11</b> Matriz de correlación de Spearman entre diferentes ventanas de tiempo de niveles mensuales de agua del LT.....	45
<b>Figura 12</b> Hidrograma de niveles de agua observado y pronosticado basada en el modelo RNN-LSTM con paso de tiempo de 1, 5, 10, 20, 30, 40, 50, 75, 100, 150, 250 y 365 días.....	49



<b>Figura 13</b>	Diagrama de dispersión entre niveles de agua pronosticado por la RNN-LSTM con ventanas de tiempo 1, 5, 10, 20, 30, 40, 50, 75, 100, 150, 250 y 365 días y valores observados. Izquierda (entrenamiento) y derecha (prueba). .....	50
<b>Figura 14</b>	Hidrograma de niveles de agua observado y pronosticado basada en el modelo RNN-LSTM con paso de tiempo de 1, 2, 3, 4, 5, 6, 8, 10, 12, 18, 24 y 48 meses. ....	55
<b>Figura 15</b>	Diagrama de dispersión entre niveles de agua pronosticado por la RNN-LSTM con ventanas de tiempo 1, 2, 3, 4, 5, 6, 8, 10, 12, 18, 24 y 48 meses, y valores observados. Izquierda (entrenamiento) y derecha (prueba). .....	56



## ÍNDICE DE TABLAS

	<b>Pág.</b>
<b>Tabla 1</b> Principales características hidrológicas del LT.....	33
<b>Tabla 2</b> Desempeño de la RNN-LSTM con diferentes ventanas de tiempo para el pronóstico de NALT diario .....	47
<b>Tabla 3</b> Desempeño de la RNN-LSTM con diferentes ventanas de tiempo para el pronóstico de NALT mensual .....	53



## ÍNDICE DE ANEXOS

	<b>Pág.</b>
<b>ANEXO 1:</b> Código fuente del modelo RNN-LSTM para pronóstico de niveles de agua del Lago Titicaca. ....	66
<b>ANEXO 2:</b> Declaración jurada de autenticidad de tesis.....	67
<b>ANEXO 3:</b> Autorización para el depósito de tesis en el repositorio institucional .....	68



## ACRÓNIMOS

<b>SENAMHI:</b>	Servicio Nacional de Meteorología e Hidrología
<b>RNN:</b>	Redes neuronales recurrentes
<b>LSTM:</b>	Memoria a corto y largo plazo
<b>NALT:</b>	Niveles de agua del lago Titicaca
<b>DE:</b>	Desviación estándar
<b>RMSE:</b>	Raíz del error cuadrático medio
<b>NSE:</b>	Eficiencia de Nash-Sutcliffe
<b>KGE:</b>	Eficiencia de Kling-Gupta
<b>PBIAS:</b>	Sesgo porcentual
<b>r:</b>	Coefficiente de correlación
<b>IA:</b>	Inteligencia artificial



## RESUMEN

El pronóstico preciso de los niveles de agua del Lago Titicaca (NALT) es esencial para la seguridad y la gestión eficiente de los recursos hídricos, especialmente en condiciones climáticas extremas. El objetivo de la investigación fue desarrollar modelos de Redes Neuronales Recurrentes (RNN) de memoria a corto y largo plazo (LSTM) para pronosticar niveles de agua a nivel diario y mensual, utilizando diferentes ventanas de tiempo. Se evaluaron los modelos utilizando métricas comunes en recursos hídricos, como el error cuadrático medio (RMSE), la eficiencia de Nash-Sutcliffe (NSE), la eficiencia de Kling-Gupta (KGE) y el sesgo porcentual (PBIAS). Para pronósticos diarios, la ventana de tiempo más efectiva fue de 1 día, con un RMSE de 0.0080 metros y un NSE de 0.9996. Para pronósticos mensuales, la mejor ventana de tiempo fue de 12 meses, con un RMSE de 0.103 metros y un NSE de 0.9741. En ambos casos, el modelo LSTM demostró un bajo sesgo y alta precisión. El número de épocas de entrenamiento no tuvo un impacto significativo en la precisión del modelo. La utilización de RNN-LSTM se revela efectiva en diferentes escalas temporales, siendo una herramienta valiosa para la predicción de niveles de agua del Lago Titicaca. Estos modelos pueden desempeñar un papel fundamental en la emisión de alertas tempranas, gestión del riesgo, regulación ecológica, suministro de agua, control de inundaciones y toma de decisiones en situaciones de cambio extremo en los NALT, con potenciales impactos en los ecosistemas.

**Palabras clave:** Altiplano peruano, Aprendizaje automático, Aprendizaje profundo, Aprendizaje supervisado, Cuenca del lago Titicaca, Inteligencia artificial, LSTM, RNN.



## ABSTRACT

The accurate forecast of water levels in Lake Titicaca (NALT) is essential for the safety and efficient management of water resources, especially under extreme weather conditions. The objective of the research was to develop Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) models for predicting water levels on a daily and monthly basis, using different time windows. The models were evaluated using common metrics in water resources, such as Root Mean Square Error (RMSE), Nash-Sutcliffe Efficiency (NSE), Kling-Gupta Efficiency (KGE), and Percentage Bias (PBIAS). For daily forecasts, the most effective time window was 1 day, with an RMSE of 0.0080 meters and an NSE of 0.9996. For monthly forecasts, the best time window was 12 months, with an RMSE of 0.103 meters and an NSE of 0.9741. In both cases, the LSTM model demonstrated low bias and high accuracy. The number of training epochs did not have a significant impact on the model's accuracy. The use of RNN-LSTM proves effective across different time scales, making them a valuable tool for Lake Titicaca water level prediction. These models can play a crucial role in issuing early warnings, risk management, ecological regulation, water supply, flood control, and decision-making in situations of extreme NALT fluctuations, with potential impacts on ecosystems.

**Keywords:** Artificial intelligence, Deep learning, Lake Titicaca Basin, LSTM, Machine learning, Peruvian Altiplano, RNN, Supervised learning.



# CAPÍTULO I

## INTRODUCCIÓN

### 1.1 PLANTEAMIENTO DEL PROBLEMA

Los lagos son recursos naturales valiosos y desempeñan un papel fundamental en el desarrollo social y económico (Zhu et al., 2020). Fueron considerados como uno de los paisajes más importantes del mundo, es bien sabido de un país a otro que las condiciones hidrológicas influyen significativamente en las funciones globales de los lagos (Ye et al., 2020). Uno de los cuerpos de agua más grande y profundo de Sudamérica es el lago Titicaca (LT), ubicado en la meseta andina entre Perú y Bolivia, y es de gran importancia para la región, ya que es una fuente de agua potable y un recurso clave para la agricultura y la pesca. Sin embargo, la variabilidad del clima y los patrones de lluvia pueden afectar significativamente los niveles de agua del lago Titicaca, lo que puede tener consecuencias graves para las comunidades locales. El nivel del agua es un indicador físico importante y su fluctuación puede afectar significativamente los ecosistemas de los lagos.

Según el diagnóstico de daños por eventos extremos realizado en el año 1993, para elaborar el plan Director Global Binacional de Protección – Prevención de Inundaciones y Aprovechamiento de los recursos del sistema endorreico Titicaca – Desaguadero – Poopó – Salar (TDPS), han caracterizado daños causados por eventos extremos (en el sector peruano y boliviano) a la agricultura, ganadería e infraestructura (vivienda, educación, saneamiento y transporte vial – ferroviario y lacustre) en los años 1982/83 (sequias), 1985/87 (inundaciones) y 1989/90 (sequias), con montos totales de daños US\$ 128.0 (agricultura y ganadería), 125.0 (agricultura e infraestructura) y 88.5 (agricultura)



millones respectivamente, siendo las inundaciones 1986/87 las más significativas en el sector peruano (TDPS, 1993).

Bajo estos acontecimientos, las posibilidades de presentarse eventos extremos (sequías e inundaciones) no son ajenos debido al cambio climático global, motivo por el cual, la importancia de contribuir con la implementación de modelos de redes neuronales recurrentes y redes LSTM que permitan dar a conocer el pronóstico anticipado de niveles de agua del Lago Titicaca.

## **1.2 FORMULACIÓN DEL PROBLEMA**

### **1.2.1 Problema general**

¿Es posible predecir los niveles de agua del lago Titicaca utilizando modelos RNN-LSTM?

### **1.2.2 Problema específico**

¿Cuáles son las características importantes que influyen en los niveles de agua del lago Titicaca?

¿Es posible implementar modelos de RNN-LSTM para el pronóstico diario de niveles de agua del lago Titicaca?

¿Es posible implementar modelos de RNN-LSTM para el pronóstico mensual de niveles de agua del lago Titicaca?



### **1.3 HIPÓTESIS DE LA INVESTIGACIÓN**

El modelo de RNN-LSTM es capaz de aprender patrones de variabilidad y tendencias en los datos y, por lo tanto, predecir con precisión los niveles de agua del LT en el futuro.

### **1.4 JUSTIFICACIÓN DEL ESTUDIO**

La predicción precisa del nivel de agua para lagos, es esencial para la alerta de inundaciones y gestionar los recursos de agua dulce (Zakaria et al., 2023). Las inundaciones son una causa importante de pérdidas sociales y económicas, así como de pérdidas de vidas humanas (Le et al., 2019).

Durante las últimas dos décadas, los algoritmos de aprendizaje automático se han utilizado ampliamente en el modelado hidrológico basado en datos en general como una alternativa a la técnica de modelado estadístico convencional (Zakaria et al., 2023). Se ha demostrado que los modelos de aprendizaje automático producen mejores resultados, especialmente cuando se trata de relaciones no lineales y de alta dimensión de variables hidrológicas (Zounemat-Kermani et al., 2020).

El uso de RNN-LSTM en la predicción hidrológica no se usaba con frecuencia (Hrnjica & Bonacci, 2019). Vacío de conocimiento: Aunque hay unos pocos estudios que han utilizado LSTM para la predicción de caudales diarios en un área semiárida en Marruecos (Nifa et al., 2023), mientras que algunos utilizaron modelos (ARIMA) para predecir los niveles de agua del LT en el pasado, esta investigación presenta, un primer intento en la aplicación de RNN-LSTM para el pronóstico de niveles del LT. Además, se desconoce si estos modelos pueden mejorar la precisión de las predicciones en comparación con otros modelos existentes. Por lo tanto, se justifica la necesidad de



investigar la capacidad de los modelos de RNN-LSTM para predecir los niveles de agua del LT.

## **1.5 OBJETIVOS DE INVESTIGACIÓN**

### **1.5.1 Objetivo general**

Desarrollar modelos de predicción de los niveles de agua del lago Titicaca utilizando RNN-LSTM.

### **1.5.2 Objetivos específicos**

Seleccionar las características importantes que influyen en los niveles de agua del lago Titicaca.

Implementar modelos de RNN-LSTM para el pronóstico diario de niveles de agua del lago Titicaca.

Implementar modelos de RNN-LSTM para el pronóstico mensual de niveles de agua del lago Titicaca.



## CAPÍTULO II

### REVISIÓN DE LITERATURA

#### 2.1 ANTECEDENTES

##### 2.1.1 Nivel internacional

En los últimos años, diversos estudios han desarrollado y aplicado algoritmos de aprendizaje automático, como el Multi-layer Perceptron Neural Networks (MLP-NN) y el Adaptive Neuro-Fuzzy Inference System (ANFIS) para el pronóstico de niveles horarios del río Muda en Malaysia (Nur Adli Zakaria et al., 2021), en el estudio realizado por (Morovati et al. en 2021), se demostró que el modelo LSTM desarrollado mostraba la capacidad de proporcionar predicciones precisas para los niveles diarios del agua en el Lago Tonle Sap en China, en la predicción del nivel del agua en tiempo real de canales basada en Multilayer Perceptron (MLP) y Recurrent Neural Network (RNN) en China (Ren et al., 2020), también se empleó (RNN-LSTM) en la construcción del modelo de predicción de variaciones de los niveles de agua del lago Poyang en China (Guo & Lai, 2020). Por otro lado (Zhu et al., 2020) realizaron el pronóstico del nivel del agua en varios lagos templados mediante modelos Feed forward neural network (FFNN), Recurrent neural network (RNN) y Deep learning (DL) en Polonia. (Agudelo Otálora et al., 2018) realizaron una comparación de modelos físicos y de inteligencia artificial para predicción de niveles de inundación en Colombia, (Kostić et al., 2016) utilizaron redes neuronales artificiales (ANN) para la estimación del caudal de agua del río Lim en el suroeste de Serbia, (Veintimilla-Reyes et al., 2016) utilizaron Artificial Neural Networks (ANN) para predecir el



caudal en el río Tomebamba en Ecuador. También, en Corea del Sur (Seo et al., 2015) utilizaron dos modelos híbridos, entre ellos, wavelet-based artificial neural network (WANN) y wavelet-based adaptive neuro-fuzzy inference system (WANFIS) para el pronóstico del nivel de agua en un embalse, (Chang et al., 2014) en Taiwán utilizaron ANN para el pronóstico del nivel del agua para el control de inundaciones urbanas, los algoritmos utilizados son una red neuronal artificial estática (ANN) (red neuronal de retropropagación-BPNN) y dos ANN dinámicas (red neuronal Elman-Elman NN; red autorregresiva no lineal con entradas exógenas-red NARX), mientras que (Filippo et al., 2012) utilizaron una red neuronal artificial (ANN) para mejorar la predicción del nivel del mar.

### **2.1.2 Nivel nacional**

En Perú, estas prácticas también fueron realizadas y en particular se utilizaron ANN para el modelado y pronóstico del Fenómeno El niño en Piura (Jiménez-Carrión et al., 2018), predicción de caudales medios mensuales en ríos de cuencas altoandinas: caso río Crisnejas (Vásquez-Ramírez & Vásquez-Paredes, 2023), modelos de pronóstico de caudales mensuales en el río Shullcas en Huancayo (Reyes & Asís, 2020), predicción del nivel del río: estudio de caso del río La Leche (Cárdenas et al., 2021) y RNN para el pronóstico de caudales en el río Chira (Lama & Sánchez, 2020)

### **2.1.3 Nivel regional/local**

En la región Puno, estas prácticas también fueron realizadas y en particular se utilizaron ANN para el pronóstico de caudales medios mensuales (Laqui, 2010) (Lujano et al., 2014), pronóstico de la evapotranspiración potencial (Laqui et al.,



2019), KNN para el pronóstico de caudales diarios (Lujano et al., 2023) y por otro lado se utilizó el enfoque box-Jenkins para identificar el modelo univariante ARIMA multiplicativo (1, 1, 0) (0, 1, 1) que permitie describir y pronosticar el comportamiento de las variaciones de niveles mensuales del lago Titicaca (Vargas, 2019). Sobre el desarrollo de modelos de RNN-LSTM para el pronóstico de niveles de agua del lago Titicaca, aun no se ha informado, lo que refuerza la importancia de realizar estudios enfocados al pronóstico preciso de los niveles de agua.

## 2.2 MARCO TEORICO

### 2.2.1 Inteligencia artificial

La inteligencia artificial (IA) es un concepto introducido en la década de 1940 y, en general, la IA se puede identificar como la ciencia del desarrollo de máquinas inteligentes o programas informáticos para imitar la inteligencia humana (Baduge et al., 2022). La IA es un método que se refiere a un sistema o una máquina que imita la inteligencia humana para realizar funciones en el mundo real.

La IA permite entrenar el sistema a partir de los datos y pensar y aprender de la experiencia para resolver problemas particulares (Saranya & Subhashini, 2023). La IA se ha convertido en un área importante para abordar la mayoría de problemas de sostenibilidad ambiental, como la biodiversidad, la energía, el transporte y la gestión del agua. Las aplicaciones de IA y los modelos de aprendizaje automático se utilizan cada vez más para predecir y optimizar la conservación de los recursos hídricos (Nti et al., 2022).

### 2.2.2 Aprendizaje automático

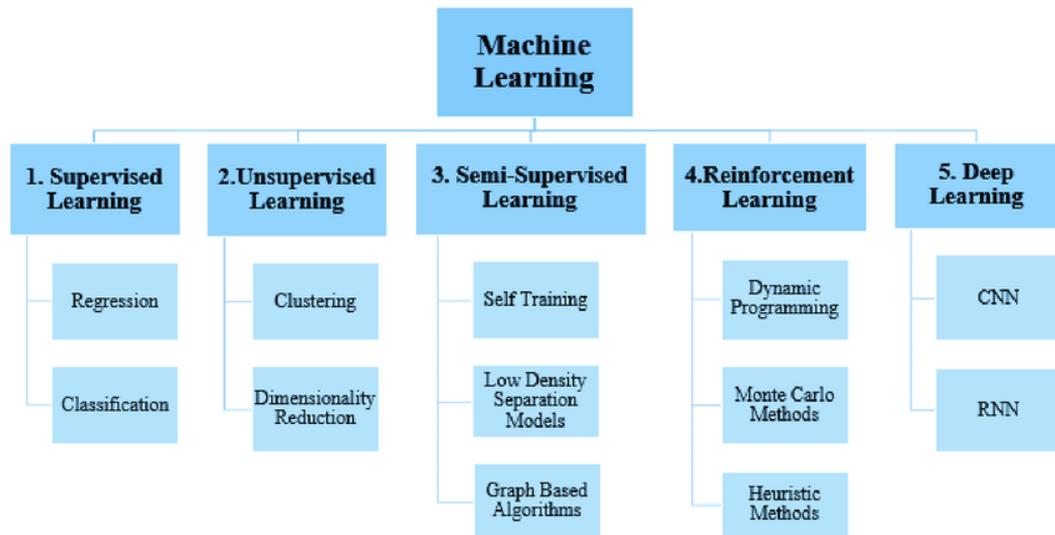
*El aprendizaje automático (ML)* es una técnica que permite que una computadora identifique patrones, haga predicciones más precisas y se perfeccione a sí misma a través de la experiencia sin estar programada con precisión para hacerlo. El aprendizaje automático se utiliza para crear una aplicación impulsada por IA. Este proceso se realiza utilizando las Metodologías ML (Saranya & Subhashini, 2023).

El aprendizaje automático se puede aplicar para soportar las limitaciones de los sistemas físicos en los que una gran cantidad de datos opera bajo relaciones complejas (Yaseen et al., 2015). En particular, la aplicabilidad del aprendizaje automático para predecir el proceso de elementos hidrológicos continuos y no lineales es muy alta (Choi et al., 2020). El aprendizaje automático puede ser una alternativa muy efectiva para áreas donde el sistema de observación es deficiente o donde el análisis físico no es posible debido a la baja densidad de la red de observación (Choi et al., 2020).

Existen cinco técnicas principales de aprendizaje automático que son: aprendizaje supervisado, aprendizaje no supervisado, aprendizaje semisupervisado, aprendizaje por refuerzo y finalmente aprendizaje profundo (Nassif et al., 2019). La Figura 1 muestra los diferentes tipos y algoritmos de aprendizaje automático.

**Figura 1**

*Diferentes tipos y algoritmos de aprendizaje automático.*



Nota: Imagen obtenida de Nassif et al. (2019)

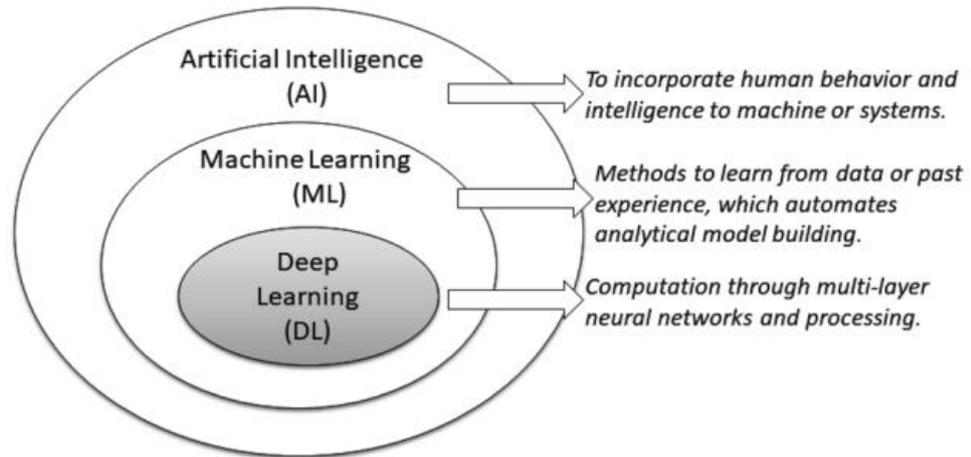
### 2.2.3 Aprendizaje profundo

*El aprendizaje profundo (DL)* trata con algoritmos influenciados por la estructura y función del cerebro humano. DL utiliza redes neuronales artificiales para crear un modelo inteligente y resolver problemas críticos. DL utiliza datos estructurados y no estructurados para entrenar un modelo (Saranya & Subhashini, 2023). Los modelos de aprendizaje profundo se dividen en tres módulos llamados preprocesamiento de datos, extracción y reconocimiento de características y optimización del modelo

Según la Figura 2, el aprendizaje profundo (DL) es una parte del ML, así como una parte del área amplia de la inteligencia artificial (Sarker, 2021).

## Figura 2

*Ilustración de la posición del aprendizaje profundo (DL), en comparación con el aprendizaje automático (ML) y la inteligencia artificial (IA)*



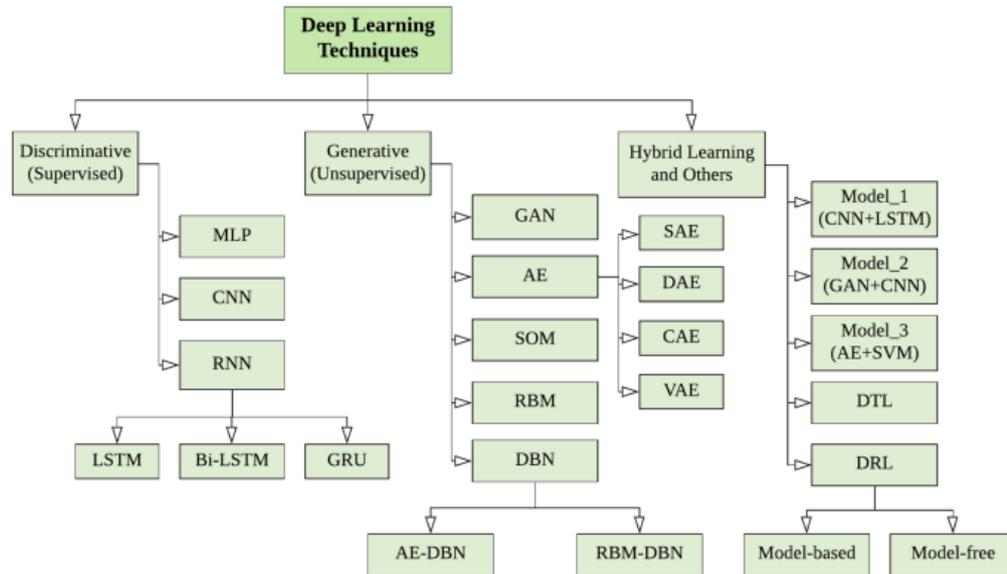
Nota: Imagen obtenida de Sarker (2021)

DL también representa métodos de aprendizaje a partir de datos donde el cálculo se realiza a través de procesamiento y redes neuronales multicapa. El término "profundo" en la metodología de aprendizaje profundo se refiere al concepto de múltiples niveles o etapas a través de las cuales se procesan los datos para construir un modelo basado en datos (Sarker, 2021).

Sarker (2021) dividió las técnicas de aprendizaje profundo en tres categorías principales: (i) redes profundas para el aprendizaje supervisado o discriminativo; (ii) redes profundas para el aprendizaje generativo o no supervisado; y (ii) redes profundas para el aprendizaje híbrido que combinan ambas y otras relevantes, como se muestra en la Figura 3.

**Figura 3**

*Taxonomía de técnicas de DL, divididas en tres categorías principales (i) redes profundas para aprendizaje supervisado o discriminativo, (ii) redes profundas para aprendizaje no supervisado o generativo, y (ii) redes profundas para aprendizaje híbrido y otras relevantes.*



Nota: Imagen obtenida de Sarker (2021)

El aprendizaje supervisado incluye principalmente las redes neuronales recurrentes (RNN), junto con sus variantes (Figura 3). A continuación, se define brevemente la técnica RNN y LSTM.

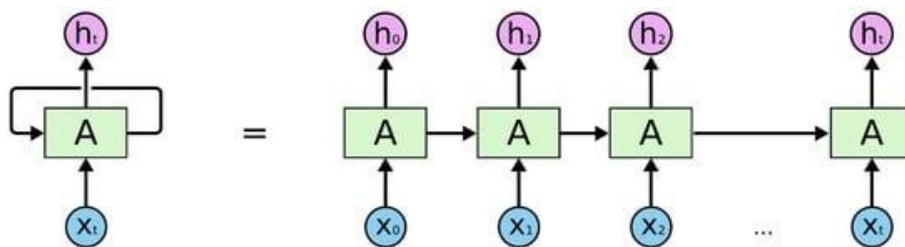
#### 2.2.4 Redes neuronales recurrentes (RNN)

La red neuronal recurrente se desarrolló por primera vez en la década de 1980 (Rumelhart et al., 1986; Werbos, 1988; Elman, 1990). Su estructura consta de una capa de entrada, una o más capas ocultas y una capa de salida. Los RNN tienen estructuras en forma de cadena de módulos repetitivos con la idea de usar

estos módulos como memoria para almacenar información importante de pasos de procesamiento anteriores. A diferencia de las redes neuronales feedforward, las RNN incluyen un circuito de retroalimentación que permite que la red neuronal acepte una secuencia de entradas. Esto significa que la salida del paso  $t-1$  se retroalimenta a la red para influir en el resultado del paso  $t$  y para cada paso subsiguiente. Por lo tanto, los RNN han tenido éxito en el aprendizaje de secuencias (Le et al., 2019). La Figura 4, muestra el procesamiento secuencial en RNN.

**Figura 4**

*Procesamiento secuencial en una Red Neuronal Recurrente (RNN)*



Nota: Imagen obtenida de Olah (2015); Le et al. (2019)

La previsión de series temporales hidrológicas se basa en los valores previos de la serie en función del número de componentes de persistencia (memoria). Las redes neuronales recurrentes (RNN) brindan esta función a través de varios bucles de retroalimentación. Un RNN generalizado puede enviar entradas en cualquier dirección desde y hacia todas las capas. Por lo tanto, la salida de la red no solo depende de las entradas externas que recibe, sino también del estado de la red en el paso de tiempo anterior (Nagesh Kumar et al., 2004). RNN es una estructura de red neuronal especial propuesta según el punto de vista de que la cognición humana se basa en la experiencia y la memoria pasadas. RNN no

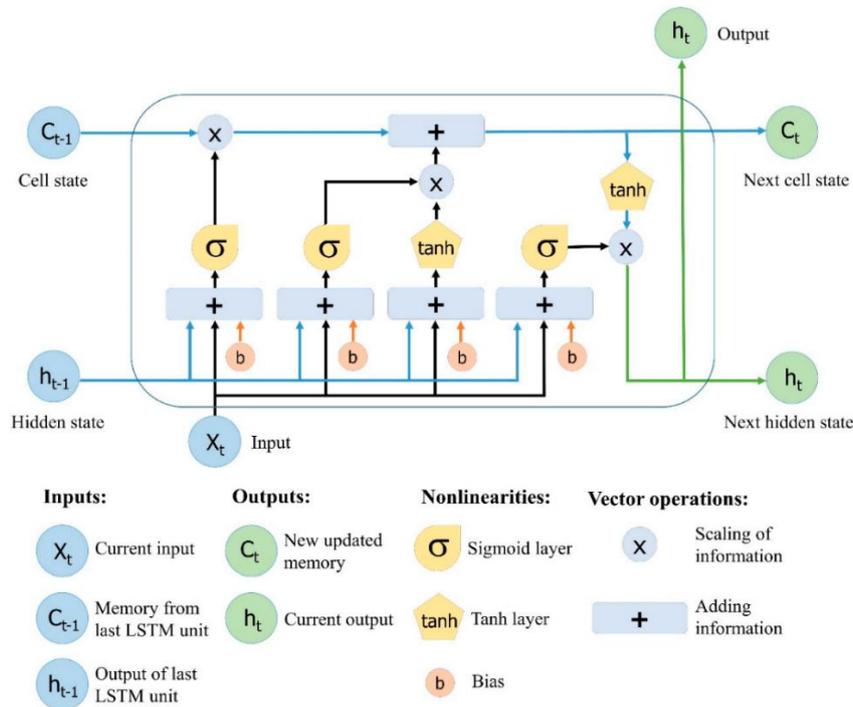
solo tiene en cuenta la entrada del momento anterior, sino que también le da a la red una función de memoria del contenido anterior (Ren et al., 2020).

### 2.2.5 Redes Long-Short Term Memory (LSTM)

Las redes de memoria a corto plazo (LSTM) es un modelo de aprendizaje automático que fue propuesto por Hochreiter & Schmidhuber (1997), para resolver la complejidad del almacenamiento de información en el proceso de retropropagación de secuencia larga. Debido al excelente desempeño en tareas de secuencia larga, LSTM se ha aplicado ampliamente en varios campos desde su inicio, especialmente en series de tiempo (Man et al., 2022).

**Figura 5**

*Estructura de la red neuronal Long Short-Term Memory (LSTM) (Yan, 2015) y adaptado por (Le et al., 2019).*



Nota: Imagen obtenida de Le et al. (2019)



Los LSTM son un tipo especial de RNN, capaces de aprender dependencias a largo plazo y recordar información durante períodos prolongados de forma predeterminada (Le et al., 2019). Según Olah (2015), el modelo LSTM está organizado en forma de estructura de cadena. Sin embargo, el módulo de repetición tiene una estructura diferente. En lugar de una única red neuronal como una RNN estándar, tiene cuatro capas que interactúan con un método único de comunicación. En la Figura 5 se muestra la estructura de la red neuronal LSTM.

La red LSTM se desarrolla a partir de la red neuronal recurrente (RNN). LSTM supera la debilidad de la RNN tradicional en el aprendizaje de dependencias a largo plazo (Fan et al., 2020). A comparación de la RNN, LSTM agrega un mecanismo de olvido que también puede resolver el problema de explosión de gradiente (Man et al., 2022).

Para Chen et al. (2021), LSTM es una red neuronal recurrente especial, que puede resolver eficazmente el problema de desaparición del gradiente de la red neuronal recurrente general (RNN). La teoría en la que se basa se refiere a la forma en que el cerebro humano recuerda las cosas y hace inferencias basadas en la memoria, y reemplaza la memoria cerebral con el núcleo recurrente para realizar la extracción de información de series temporales. Según Le et al. (2019), los LSTM están diseñados para evitar el problema de la dependencia a largo plazo porque utilizan puertas para controlar el proceso de memorización.

El LSTM ha sido estudiado para resolver problemas en hidrología y meteorología obteniendo resultados positivos (Lingling et al., 2020). Una celda LSTM tiene las siguientes puertas:

- a) Puerta de olvido: es la puerta que decide qué información se va a olvidar del estado oculto anterior e información de la entrada actual
- b) Puerta de entrada: decide qué información se utiliza para actualizar el estado de la celda.
- c) Puerta de salida: es la puerta que controla la información del estado de la celda que será entregada al nuevo estado oculto.

En la estructura de la red LSTM (Figura 5) el subíndice  $t$  es el tiempo,  $x$  es la entrada,  $h$  es el estado oculto,  $C$  el estado celular,  $\hat{C}$  es la salida de la parte reservada de la capa de bucle original,  $i$ ,  $o$  y  $f$  son las puertas de entrada, salida y olvido respectivamente.  $\sigma$  es la función de activación, como sigmoide y unidad lineal rectificadora (ReLU). LSTM tiene dos procesos 1) propagación hacia adelante y 2) propagación hacia atrás en la que este último utiliza el algoritmo de retro propagación para actualizar los parámetros y optimizar el modelo.

El proceso de propagación hacia adelante de la red LSTM  $i_t$ ,  $o_t$  y  $f_t$  se expresa mediante:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (1)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (2)$$

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (3)$$

La información de la celda  $C_t$  y la información oculta  $h_t$  se actualizan mediante:



$$\widehat{C}_t = \tan h (W_C [h_{t-1}, x_t] + b_C) \quad (4)$$

$$C_t = f_t \times C_{t-1} + i_t + \widehat{C}_t \quad (5)$$

$$h_t = o_t \times \tan h (C_t) \quad (6)$$

donde  $W$  es la matriz de pesos y  $b$  es el sesgo que se actualizan y optimizan durante el proceso.

Dado que resuelve los problemas de entrenamiento de una red recurrente, la red LSTM se considera una de las RNN más exitosas (Sarker, 2021).



## CAPÍTULO III

### MATERIALES Y MÉTODOS

#### 3.1 ÁREA DE ESTUDIO

El LT (Figura 6), se encuentra dentro de la región hidrográfica del Titicaca. Tiene un régimen permanente, y está ubicado entre los países de Perú y Bolivia. Los registros de NALT, se registran en la estación hidrométrica muelle ENAFER ubicado en las coordenadas WGS 84 15°50'16.8" en latitud sur, 70°1'24.8" en longitud oeste a una altitud de 3808 msnm. En la estación, el nivel de agua máximo promedio mensual se registró en abril de 1986, alcanzando los 3812.55 metros sobre el nivel del mar (m.s.n.m.), mientras que el mínimo, con 3806.24 m.s.n.m., se registró en diciembre de 1943, con un promedio general de 3809.43 m.s.n.m. (Tabla 1). Por otro lado, la superficie promedio del lago Titicaca se estima en 8,372 km<sup>2</sup>; sin embargo, la imagen de satélite Landsat 5 Thematic Mapper (TM) del 18 de mayo de 1986 revela un área de inundación que abarca un total de 9336.57 km<sup>2</sup> (Figura 6). El sistema de drenaje que fluye al LT se distribuye en forma dendrítica y los principales aportantes de agua de mayor a menor son los ríos Ramis, Ilave, Coata, Huancané, Suches, Illpa, Zapatilla y Salado. Aunque existen otros aportantes menores, la superficie total de la cuenca del lago Titicaca (CLT) es de 49,171.54 km<sup>2</sup> aproximadamente.

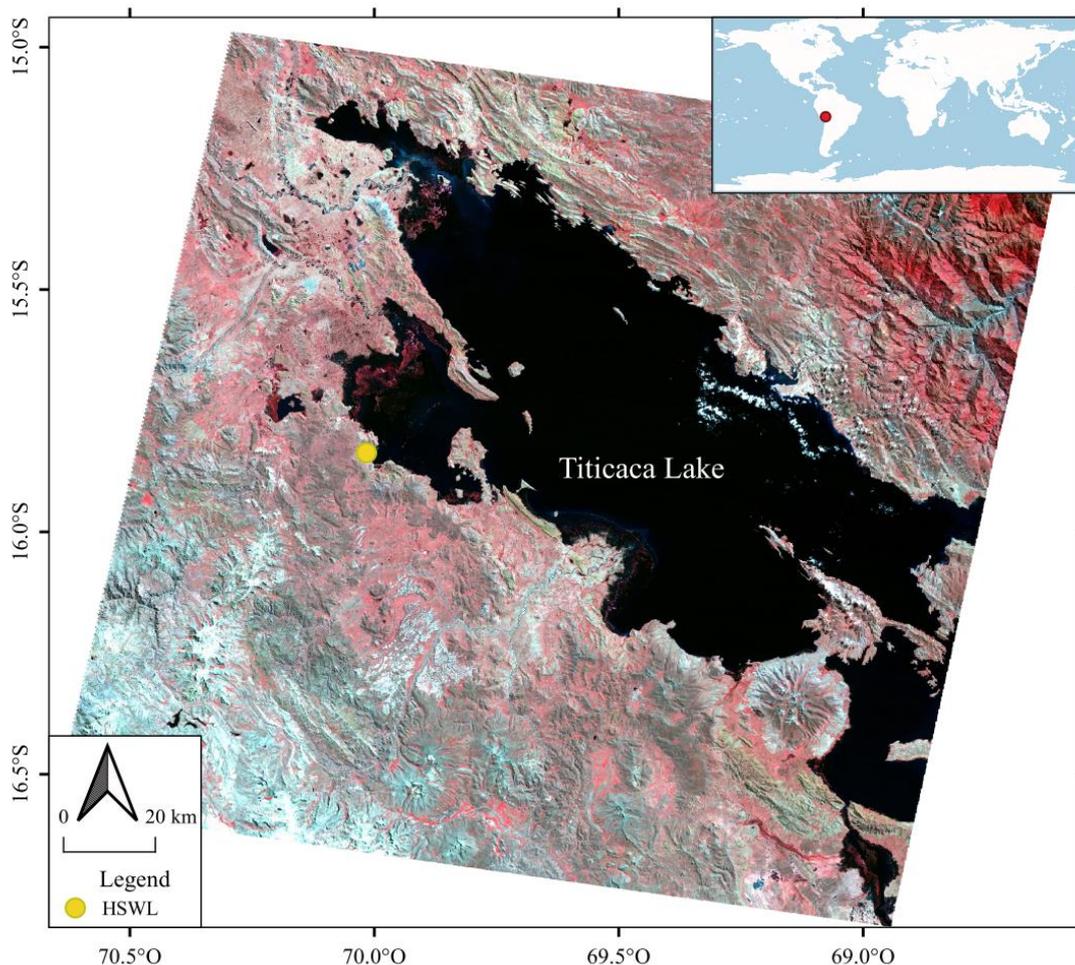
Los mayores acumulados de precipitaciones y aportes de caudales de los ríos se dan en los meses de diciembre a marzo, con temperaturas más bajas concentradas en los meses de junio y julio.

### 3.2 TIPO Y NIVEL DE LA INVESTIGACIÓN

La investigación es de tipo analítico, longitudinal y retrospectivo. Es analítico por que busca la asociación entre variables, es longitudinal y retrospectivo, por que registra los datos ocurridos en el pasado (Rojas, 2021). Dado que se busca comprobar la hipótesis previamente establecida, así como los objetivos trazados, el presente trabajo fue elaborado bajo el planteamiento metodológico del enfoque cuantitativo.

#### Figura 6

*Distribución espacial del LT y ubicación de la estación hidrométrica muelle ENAFER*



Nota: Imagen obtenida de satélite Landsat 5 Thematic Mapper (TM)

### 3.3 TÉCNICA E INSTRUMENTOS

#### 3.3.1 Adquisición de datos

Se utilizaron datos del nivel de agua diario del LT. El nivel de agua, se mide en la estación muelle ENAFER durante todos los días a las seis de la mañana por el Servicio Nacional de Meteorología e Hidrología (SENAMHI) Perú. Los datos diarios de niveles de agua, estuvieron disponibles desde 01 de setiembre de 1982 hasta el 31 de diciembre de 2016 (35 años) (Figura 7a), mientras que los datos mensuales de niveles de agua estuvieron disponibles desde enero de 1915 (102 años) (Figura 7b). Para obtener datos mensuales para el periodo 1982-2016, los datos de niveles diarios se promediaron. En la Figura 7 se presenta la serie temporal diaria y mensual de los datos de NALT.

**Tabla 1**

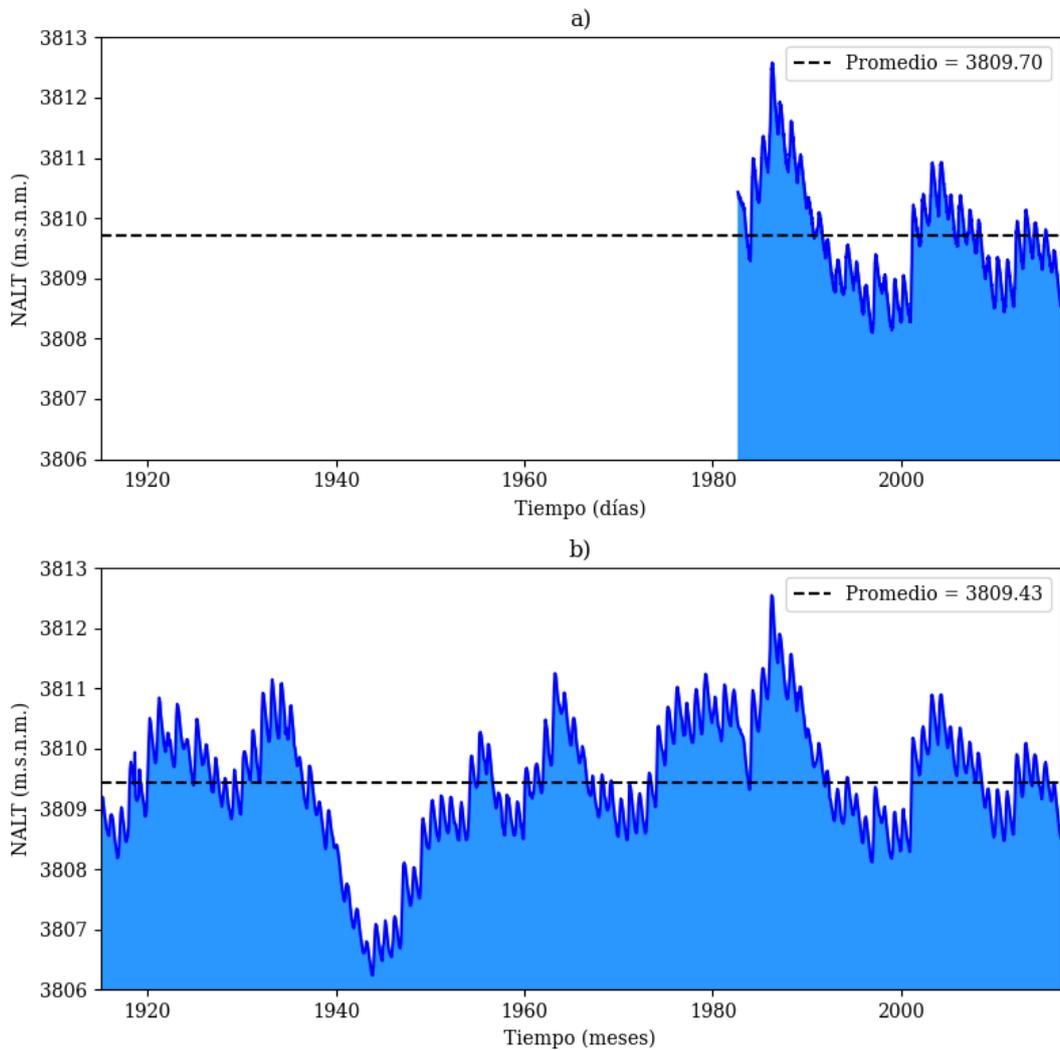
*Principales características hidrológicas del LT*

<b>Estadísticas descriptivas</b>	<b>NALT diario</b>	<b>NALT mensual</b>
Mínimo (m.s.n.m.)	-	3806.24 (diciembre 1943)
Máximo (m.s.n.m.)	3812.58 (17-22 abril 1986)	3812.55 (abril 1986)
Promedio (m.s.n.m.)	3809.44	3809.43
Rango (m)	-	6.31
Percentil 33 (m.s.n.m.)	3809.21	3809.04
Percentil 66 (m.s.n.m.)	3809.98	3809.90

Elaboración propia

**Figura 7**

*Hidrograma de NALT – Estación muelle ENAFER a) diario y b) mensual.*



Elaboración propia

### 3.4 PROCEDIMIENTO METODOLÓGICO

#### 3.4.1 Selección de las características importantes que influyen en los niveles de agua del lago Titicaca.

Se llevó a cabo un análisis exploratorio de datos (AED) tanto en la frecuencia diaria como mensual de los niveles de agua del lago Titicaca (NALT).



Este proceso involucró una variedad de métodos gráficos, incluyendo histogramas, gráficos cuantil-cuantil y diagramas de cajas, junto con análisis estadísticos descriptivos (método no gráfico), enfocados en identificar valores atípicos, tendencias y resultados inesperados. Los datos diarios que faltaban en la serie fueron de 0.24 % y se rellenaron mediante una interpolación lineal unidimensional.

A continuación, se realizó un análisis de normalidad para ambos conjuntos de datos, diarios y mensuales de NALT. Este análisis incluyó la evaluación mediante histogramas de frecuencia, gráficos cuantil-cuantil y diagramas de cajas, así como el cálculo de la asimetría. Además, se aplicó la prueba de normalidad de Anderson-Darling (AD). En el contexto de la prueba de normalidad de AD, la hipótesis nula ( $H_0$ ) establece que los datos siguen una distribución normal, mientras que la hipótesis alternativa ( $H_1$ ) establece que los datos no siguen una distribución normal. La estadística AD se comparó con un valor crítico correspondiente a un nivel de significancia (NS) del 5% ( $NS = 0.05$ ). Si la estadística AD resultaba mayor que el valor crítico, se rechazaba la hipótesis nula en favor de la  $H_1$ , indicando que los datos no se ajustan a una distribución normal.

Luego, se procedió con un análisis de correlación utilizando el coeficiente de Spearman. Este coeficiente se eligió por su capacidad de lidiar con datos continuos que no cumplen con la suposición de normalidad. No asume relaciones lineales y demuestra robustez frente a la falta de normalidad en los datos. Además, se construyó una matriz de correlación que visualizaba las relaciones entre diversas ventanas de tiempo, proporcionando una representación clara de las conexiones temporales en los NALT.

### 3.4.2 Pronóstico diario de niveles de agua del Lago Titicaca basada en el modelo RNN-LSTM

#### 3.4.2.1 Preprocesamiento de datos

Se identificaron datos faltantes de NALT desde el 1 al 31 de agosto de 2006, los mismos que fueron rellenados mediante una interpolación lineal unidimensional. Al reducir la cantidad de datos faltantes, se puede mejorar la precisión y la calidad de los resultados del modelado. Luego, para el entrenamiento del modelo RNN-LSTM, los datos de NALT se normalizaron entre 0 y 1 usando la ecuación:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (7)$$

donde  $X$  es el NALT,  $X_{min}$  y  $X_{max}$  son los niveles de agua mínimo y máximo, respectivamente.

#### 3.4.2.2 Dividir el conjunto de datos de entrenamiento y prueba

El total de datos de NALT diario fue de 12541. Los datos se separaron en un 80% (10032) para la fase de entrenamiento (1981-09-01 a 2010-02-17) y 20% (2509) para la fase de prueba (2010-02-18 a 2016-12-31).

#### 3.4.2.3 Definición de los hiperparámetros

Look\_back (considerar el pasado) En esta función, look\_back indica cuántos registros anteriores se usarán para predecir el registro siguiente (OpenAI, 2023). En el contexto de una red neuronal LSTM, el



parámetro "look\_back" se refiere a la ventana de tiempo o al número de pasos de tiempo anteriores que se utilizan para predecir el siguiente valor en una serie temporal. Si "look\_back" es demasiado pequeño, es posible que la red neuronal no capture suficientes patrones en la serie temporal y produzca predicciones inexactas. Por otro lado, si "look\_back" es demasiado grande, es posible que el modelo se sobreajuste a los datos de entrenamiento y no pueda generalizar bien a nuevos datos.

Para el entrenamiento del modelo RNN-LSTM se utilizaron ventanas de tiempo variables que van desde un día hasta 365 días, es decir, 1, 5, 10, 20, 30, 40, 50, 75, 100, 150, 250, 365 días. Estas ventanas de tiempo, cubren una amplia gama de escalas temporales, desde patrones diarios hasta patrones anual y permiten evaluar cómo el modelo maneja diferentes patrones y fluctuaciones en los datos. Para datos mensuales se utilizaron ventanas de tiempo que van desde un mes hasta 12 meses, es decir, 1, 2, 3, 4, 5, 6, 8, 10, 12, 18, 24 y 48 meses. Estas ventanas de tiempo cubren diferentes escalas temporales, desde patrones mensuales hasta patrones anuales.

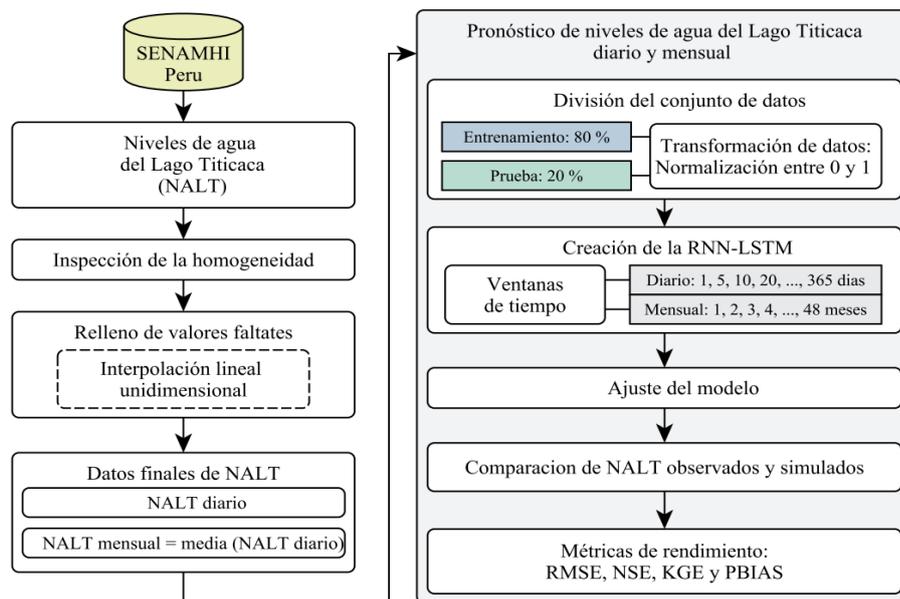
Se consideró 25 neuronas para todas las configuraciones de la red, una tasa de abandono (Dropout) 0.1 (mientras más elevado el RMSE se incrementa). El tamaño de épocas fue variable, para esto se consideró una parada automática tomando en cuenta las épocas y las pérdidas en el entrenamiento y prueba, con un límite máximo de 1000 épocas, un optimizador Adam y una función de pérdida de error cuadrático medio (MSE).

### 3.4.3 Pronóstico mensual de niveles de agua del Lago Titicaca basada en el modelo RNN-LSTM

Aquí se utilizaron registros de niveles medios mensuales del LT. Específicamente durante 1915 y 2016. Antes de separar los datos de entrenamiento y prueba, los datos de NALT se normalizaron mediante la ecuación 1. En consecuencia, para el entrenamiento del modelo RNN-LSTM se utilizaron ventanas variables que van desde 1 hasta 48 meses y el resto de la configuración fue similar de pronóstico diario.

**Figura 8**

*Diagrama de flujo del proceso de desarrollo de modelos RNN-LSTM para el pronóstico de NALT*



Elaboración propia

La Figura 8 muestra el esquema general para realizar el pronóstico de NALT diario y mensual mediante el uso de RNN-LSTM.

### 3.4.4 Métricas de bondad de ajuste

Para verificar la efectividad del modelo RNN-LSTM se seleccionaron cuatro métricas de bondad de ajuste, raíz del error cuadrático medio (RMSE), eficiencia de Nash-Sutcliffe (NSE), eficiencia de Kling-Gupta (KGE) y el sesgo porcentual (PBIAS), métricas estadísticas comúnmente utilizadas en estudios relacionados a los recursos hídricos. Para este proceso, se utilizó Google Colab y el paquete hydrostats. Hydrostats es un paquete disponible para Python y contiene más de 60 métricas para caracterizar los errores entre series de tiempo simuladas y observadas (Roberts et al., 2018).

#### 3.4.4.1 Raíz del error medio cuadrático (RMSE)

RMSE mide la precisión de la predicción que crea un valor positivo elevando al cuadrado los errores. Las puntuaciones de RMSE oscilan entre (0, Inf) y la predicción del modelo es ideal si RMSE es 0 (Zhang et al., 2018). RMSE se define de la siguiente manera:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=0}^n (S_i - O_i)^2} \quad (8)$$

donde,  $O_i$  es el valor observado,  $S_i$  es el valor simulado y  $n$  es el número de observaciones.

#### 3.4.4.2 Eficiencia de Nash-Sutcliffe (NSE)

Legates & McCabe (1999) sugieren utilizar el coeficiente NSE para evaluar la bondad del ajuste. NSE va desde menos infinito hasta 1.0, donde

los valores más altos indican una mejor relación (Nash & Sutcliffe, 1970). Los valores de NSE entre 0 y 1 indican que los valores de simulación son generalmente aceptables, mientras que los valores negativos de NSE indican que el rendimiento del valor de simulación es muy pobre (Moriasi et al., 2007). NSE es expresado por la siguiente ecuación:

$$NSE = 1 - \frac{\sum_{i=1}^n (S_i - O_i)^2}{\sum_{i=1}^n (S_i - \bar{O})^2} \quad (9)$$

donde,  $O_i$  es el valor observado,  $S_i$  es el valor simulado,  $\bar{O}$  es la media de los valores observados y  $n$  es el número de observaciones.

Moriasi et al. (2007) clasificaron el rendimiento de un modelo como muy bueno si  $NSE > 0.75$ , bueno si  $0.65 < NSE \leq 0.75$ , satisfactorio si  $0.50 < NSE \leq 0.65$  y malo si  $NSE \leq 0.50$ .

#### 3.4.4.3 Eficiencia de Kling-Gupta (KGE')

KGE' (Kling et al., 2012) es una versión modificada de KGE (Gupta et al., 2009) propuesta para evitar correlación cruzada entre las relaciones sesgo y variabilidad, el rango puede variar entre  $-\infty < KGE' < 1$ , valores cercanos a la unidad no indican sesgo. Está dado por la siguiente expresión:

$$KGE' = \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2}$$
$$\beta = \frac{\mu_S}{\mu_O}; \gamma = \frac{CV_S}{CV_O} = \frac{\sigma_S/\mu_S}{\sigma_O/\mu_O} \quad (10)$$

donde,  $\sigma$  es la desviación estándar en  $m^3/s$ ;  $r$  es el coeficiente de correlación entre el valor simulado y observado (adimensional);  $\beta$  es la razón de sesgo (adimensional);  $\gamma$  es la razón de variabilidad (adimensional);  $\mu$  es el valor medio en  $m$ ;  $CV$  es el coeficiente de variación (adimensional) y los subíndices  $s$  y  $o$  representan valores observados y simulados respectivamente.

#### 3.4.4.4 Sesgo porcentual (PBIAS)

PBIAS indica la tendencia promedio de los datos simulados basados en datos observados más grandes o más pequeños, siendo el mejor valor 0, los valores negativos indican una subestimación y los valores positivos una sobreestimación (Vu et al. (2018). El sesgo porcentual BIAS es expresado por la siguiente expresión:

$$PBIAS = \frac{\sum_{i=1}^N (S_i - O_i)}{\sum_{i=1}^N O_i} \times 100 \quad (11)$$

donde,  $O_i$  es el valor observado,  $S_i$  es el promedio de los valores simulados,  $S_i$  es el valor calculado,  $S$  es el promedio de los valores simulados y  $N$  es el número de observaciones.

Según los criterios de Moriasi et al. (2007), el desempeño estadístico de los modelos puede considerarse 'muy bueno' cuando  $PBIAS < \pm 5$ , 'bueno' cuando  $\pm 5 \leq PBIAS < \pm 10$ , 'satisfactorio' cuando  $\pm 10 \leq PBIAS < \pm 25$  y 'no satisfactorio' cuando  $PBIAS \geq \pm 25$ .



## CAPÍTULO IV

### RESULTADOS Y DISCUSIÓN

#### 4.1 CARACTERÍSTICAS IMPORTANTES QUE INFLUYEN EN LOS NIVELES DE AGUA DEL LAGO TITICACA

##### 4.1.1 Análisis de normalidad

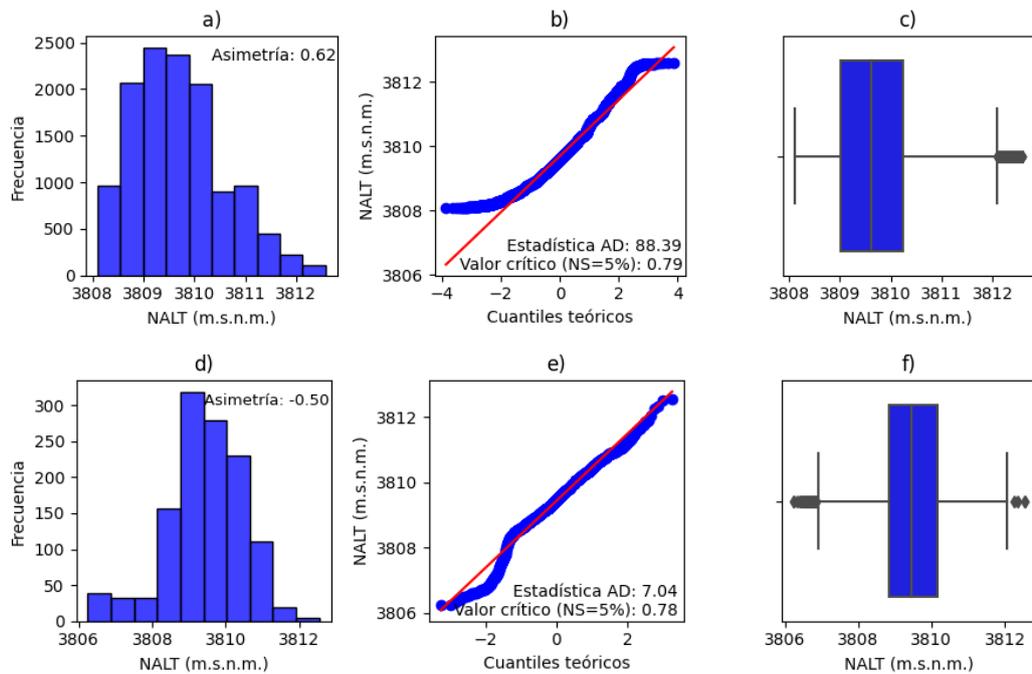
En el estudio realizado, se llevaron a cabo análisis de normalidad para evaluar la distribución de los datos diarios y mensuales de los NALT. En el análisis de normalidad de los datos diarios de NALT (Figura 9a-c), se observa una ligera asimetría positiva (0.62) en el histograma (Figura 9a) y una desviación de la línea diagonal en el gráfico cuantil-cuantil (Figura 9b), además, la estadística de Anderson-Darling de 88.39 supera significativamente el valor crítico de 0.79 para un nivel de significancia del 5%, lo que respalda la conclusión de que los datos no son normalmente distribuidos. Además, el diagrama de cajas (Figura 9c) muestra variabilidad en la distribución. Para los datos mensuales (Figura 9d-f), se detecta una ligera asimetría negativa (-0.50) en el histograma (Figura 9d) y una desviación en el gráfico cuantil-cuantil (Figura 9e). La estadística de Anderson-Darling de 7.06 supera el valor crítico de 0.78 para un nivel de significancia del 5%, indicando nuevamente una falta de normalidad. El diagrama de cajas (Figura 9f) revela una variabilidad considerable.

Los resultados de los análisis indican que tanto los datos diarios como los mensuales de los NALT no siguen una distribución normal. Esto se evidencia por la falta de simetría en los histogramas, las desviaciones en los gráficos cuantil-

cuantil y los valores significativamente altos de la estadística de AD en comparación con los valores críticos.

### Figura 8

*Análisis de normalidad para datos diarios (a-c) y mensuales (d-f) de NALT*



Elaboración propia

#### 4.1.2 Análisis de correlación

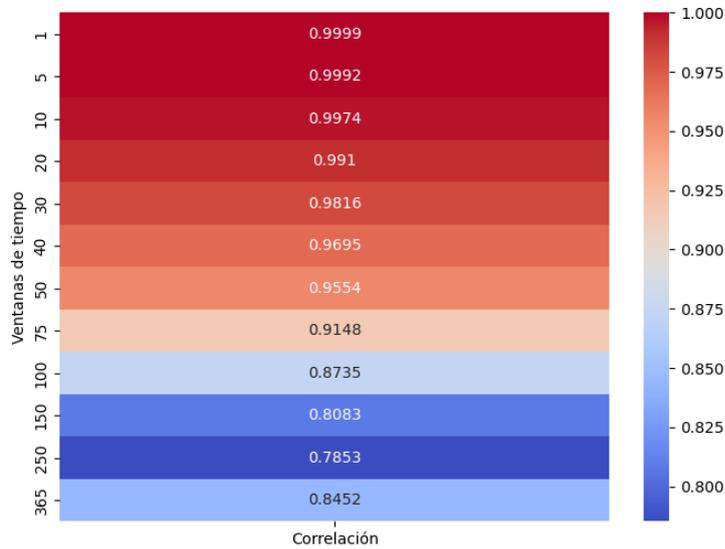
Dado que los datos diarios y mensuales de NALT no se ajustan a una distribución normal, se realizó un análisis de correlación utilizando el coeficiente de Spearman. El coeficiente de correlación de Spearman es una opción sólida cuando los datos son continuos, pero no siguen una distribución normal, mide la correlación entre dos variables sin asumir una relación lineal y es robusto frente a la falta de normalidad. Los resultados de este análisis se presentan en forma de una matriz de correlación, que ilustra las relaciones de correlación entre diversas



ventanas de tiempo (Figura 10 y 11). A medida que aumenta el tamaño de la ventana de tiempo, la correlación disminuye gradualmente. Esto indica que las variables están relacionadas, pero la fuerza de la relación disminuye a medida que se consideran intervalos de tiempo más largos. Por ejemplo, la correlación de 0.9999 (Figura 10) y 0.9823 (Figura 11) para la ventana de tiempo 1 sugiere una relación muy fuerte y casi perfectamente monotónica entre las variables consideradas en ese intervalo específico. Aunque la correlación disminuye con ventanas de tiempo más grandes, los valores siguen siendo altos en general. Para datos diarios hasta la ventana de tiempo 50, la correlación sigue siendo mayor que 0.95, lo que indica una relación muy fuerte. Esto implica que incluso con ventanas de tiempo más grandes, aún existe una fuerte relación monotónica entre las variables. Sin embargo, a medida que la ventana de tiempo aumenta más allá de 50, la correlación disminuye más rápidamente. Para datos mensuales hasta la ventana de tiempo 6, la correlación sigue siendo mayor que 0.8, lo que indica una relación fuerte. Esto implica que incluso con ventanas de tiempo mensuales más grandes, aún existe una relación bastante fuerte. Sin embargo, a medida que la ventana de tiempo aumenta más allá de 6, la correlación disminuye más rápidamente.

### Figura 9

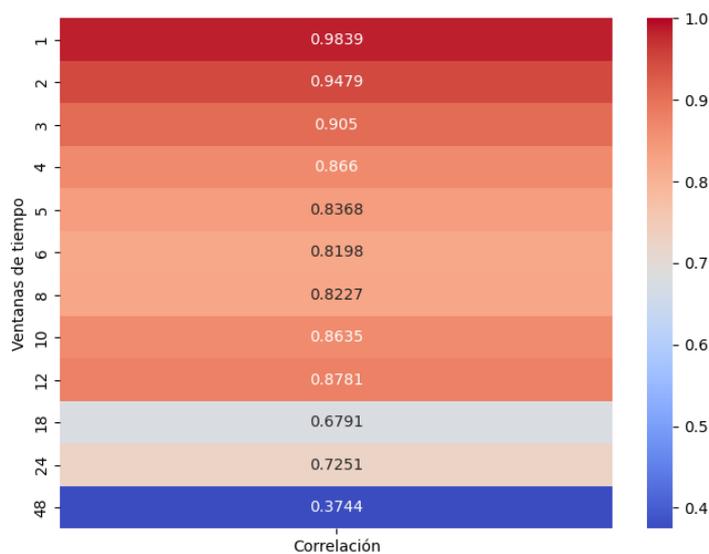
*Matriz de correlación de Spearman entre diferentes ventanas de tiempo de niveles diarios de agua del LT.*



Elaboración propia

### Figura 10

*Matriz de correlación de Spearman entre diferentes ventanas de tiempo de niveles mensuales de agua del LT.*



Elaboración propia

## 4.2 PRONÓSTICO DIARIO DE NIVELES DE AGUA DEL LAGO TITICACA BASADA EN EL MODELO RNN-LSTM

### 4.2.1 Efecto del tamaño de la ventana

Se analizaron los efectos del tamaño de la ventana en el rendimiento de la RNN-LSTM. Aquí los resultados se presentan para los tamaños de ventanas: 1, 5, 10, 20, 30, 40, 50, 75, 100, 150, 250, 365 días (Tabla 2). En general, los valores de RMSE son bastante bajos, lo que sugiere que el modelo RNN-LSTM es capaz de proporcionar predicciones precisas para los NALT. Sin embargo, se puede observar que los valores de RMSE aumentan a medida que se incrementa la ventana de tiempo, lo que sugiere que el modelo podría tener dificultades para predecir los niveles de agua a largo plazo. Esto es algo esperable, ya que, si se utilizan más datos de entrada para entrenar el modelo, aumenta la complejidad del mismo y es posible que se sobreajuste a los datos de entrenamiento, lo que puede llevar a una peor capacidad de generalización.

De igual forma los valores de NSE son altos para todas las ventanas de tiempo, indicando que el modelo RNN-LSTM tiene un buen desempeño para reproducir la variabilidad observada en los datos. Se puede observar que a diferencia de RMSE, los valores de NSE disminuyen a medida que se incrementa la ventana de tiempo. También se puede encontrar que el RMSE más bajo como el NSE más alto se obtienen con un tamaño de ventana de 1 día, lo que indica que un tamaño de ventana de 1 día puede ser la mejor opción para realizar el pronóstico de NALT en términos de precisión y también de eficiencia computacional. Sin embargo, para cualquiera de los casos, los valores de RMSE y NSE indican que el modelo RNN-LSTM es un enfoque prometedor para la predicción de NALT.

En términos de épocas de entrenamiento, se puede observar que no hay una clara tendencia que indique que un mayor número de épocas mejore significativamente el rendimiento del modelo. En algunos casos, incluso una menor cantidad de épocas puede llevar a resultados igualmente buenos.

**Tabla 2**

*Desempeño de la RNN-LSTM con diferentes ventanas de tiempo para el pronóstico de NALT diario*

Ventana de tiempo	Épocas	Entrenamiento				Prueba			
		RMSE (m)	NSE	KGE	PBIAS (%)	RMSE (m)	NSE	KGE	PBIAS (%)
1	64	0.0103	0.9999	0.9955	-0.0001	0.0080	0.9996	0.9943	0.0000
5	47	0.0322	0.9988	0.9909	0.0005	0.0232	0.9968	0.9976	0.0001
10	80	0.0161	0.9997	0.9942	-0.0002	0.0131	0.9990	0.9948	0.0000
20	69	0.0227	0.9994	0.9961	0.0003	0.0174	0.9982	0.9953	0.0001
30	67	0.0282	0.9991	0.9874	-0.0004	0.0223	0.9971	0.9841	-0.0001
40	54	0.0231	0.9994	0.9847	0.0000	0.0171	0.9983	0.9937	0.0000
50	53	0.0194	0.9996	0.9912	0.0001	0.0154	0.9986	0.9921	0.0000
75	58	0.0243	0.9993	0.9981	0.0003	0.0201	0.9977	0.9991	0.0001
100	62	0.0460	0.9976	0.9778	-0.0007	0.0330	0.9938	0.9807	-0.0001
150	61	0.0411	0.9981	0.9716	-0.0005	0.0233	0.9970	0.9712	-0.0001
250	42	0.0795	0.9931	0.9674	-0.0014	0.0598	0.9795	0.9813	-0.0003
365	28	0.0552	0.9967	0.9979	0.0008	0.0395	0.9895	0.9970	0.0001

Elaboración propia

De igual forma los valores de NSE son altos para todas las ventanas de tiempo, indicando que el modelo RNN-LSTM tiene un buen desempeño para reproducir la variabilidad observada en los datos. Se puede observar que a diferencia de RMSE, los valores de NSE disminuyen a medida que se incrementa la ventana de tiempo. También se puede encontrar que el RMSE más bajo como el NSE más alto se obtienen con un tamaño de ventana de 1 día, lo que indica que

un tamaño de ventana de 1 día puede ser la mejor opción para realizar el pronóstico de NALT en términos de precisión y también de eficiencia computacional. Sin embargo, para cualquiera de los casos, los valores de RMSE y NSE indican que el modelo RNN-LSTM es un enfoque prometedor para la predicción de NALT. En términos de épocas de entrenamiento, se puede observar que no hay una clara tendencia que indique que un mayor número de épocas mejore significativamente el rendimiento del modelo. En algunos casos, incluso una menor cantidad de épocas puede llevar a resultados igualmente buenos.

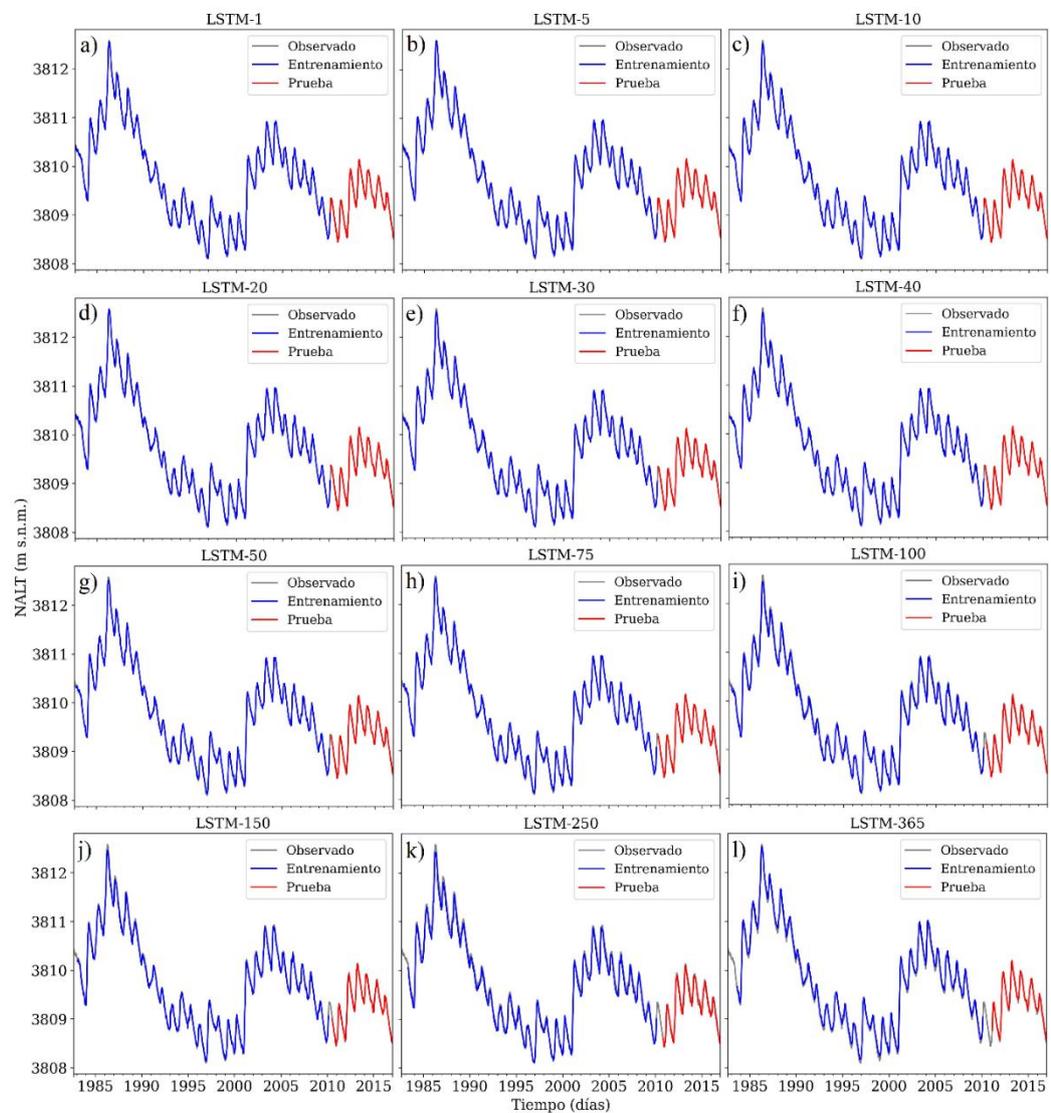
Los resultados de las métricas de desempeño para el rendimiento general de modelo RNN-LSTM para el período de entrenamiento (calibración) y prueba (validación) indican que se desempeñaron muy bien durante el período de entrenamiento, con una RMSE y NSE promedio de 0.0332 m y 0.9984, con una ligera diferencia en el periodo de prueba con una RMSE y NSE promedio de 0.0243 y 0.9954. La red RNN-LSTM funciona bastante bien en la predicción de los niveles de agua del lago Titicaca, ya que se obtienen valores muy altos de NSE y KGE, lo que indica que los valores predichos por el modelo son muy similares a los valores reales observados.

En la Figura 12 se presenta una comparación de la serie temporal de los valores de NALT observados y los valores simulados por RNN-LSTM durante el periodo de entrenamiento y prueba. Como se puede observar, RNN-LSTM puede capturar bien el patrón temporal en ambos periodos. Si bien la mayoría de NALT altos y bajos fueron capturados por la RNN-LSTM, esto se ve afectado cuando la ventana de tiempo se incrementa y esto también se nota el modelo subestima para niveles altos y sobrestima para niveles bajos. Los valores de PBIAS son cercanos

a 0 % para todas las ventanas de tiempo, lo que sugiere que el modelo RNN-LSTM no tiene un sesgo significativo en la reproducción de los valores observados. Estos resultados respaldan la utilidad del modelo RNN-LSTM como una herramienta de predicción de niveles de agua en el Lago Titicaca.

### Figura 11

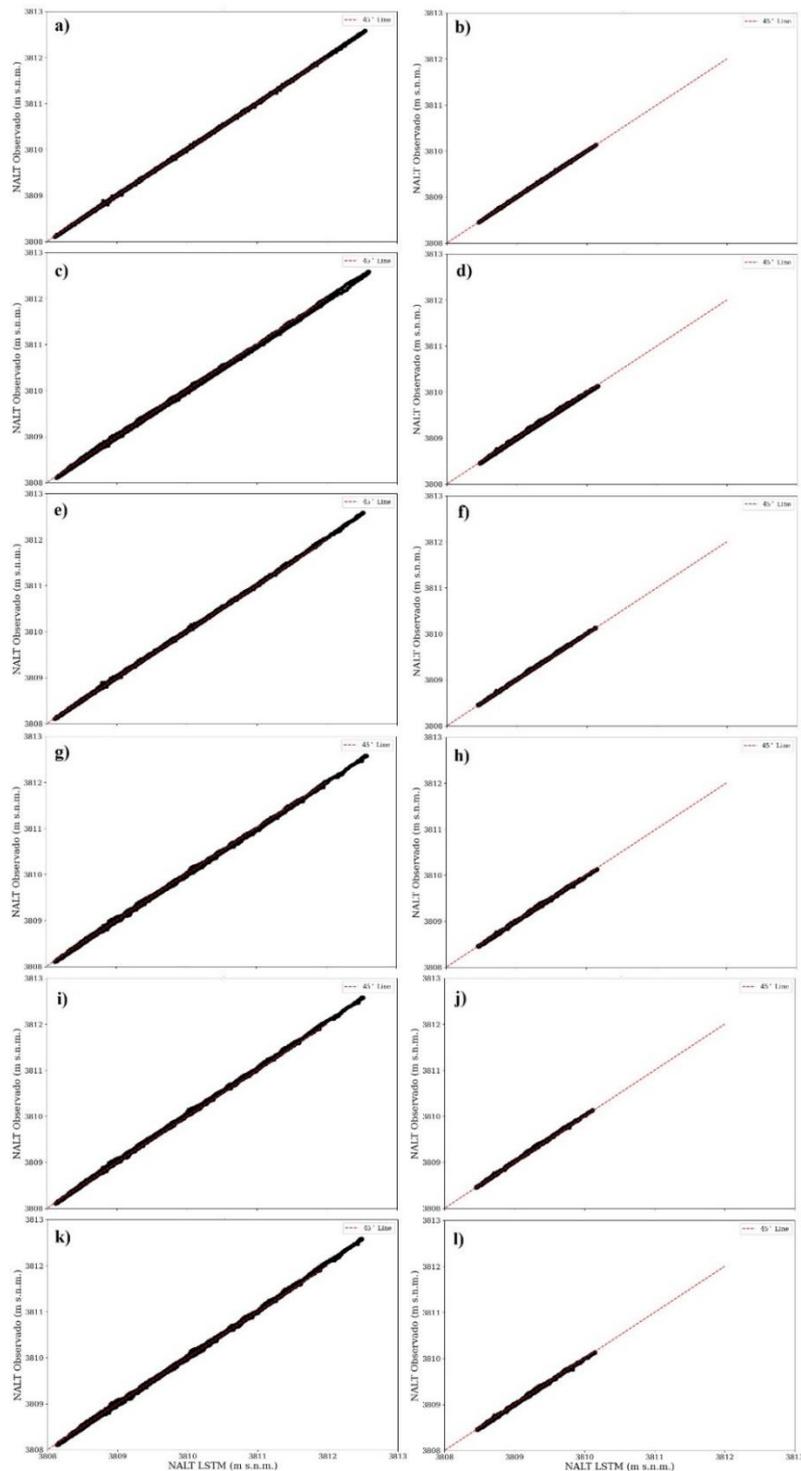
*Hidrograma de niveles de agua observado y pronosticado basada en el modelo RNN-LSTM con paso de tiempo de 1, 5, 10, 20, 30, 40, 50, 75, 100, 150, 250 y 365 días.*

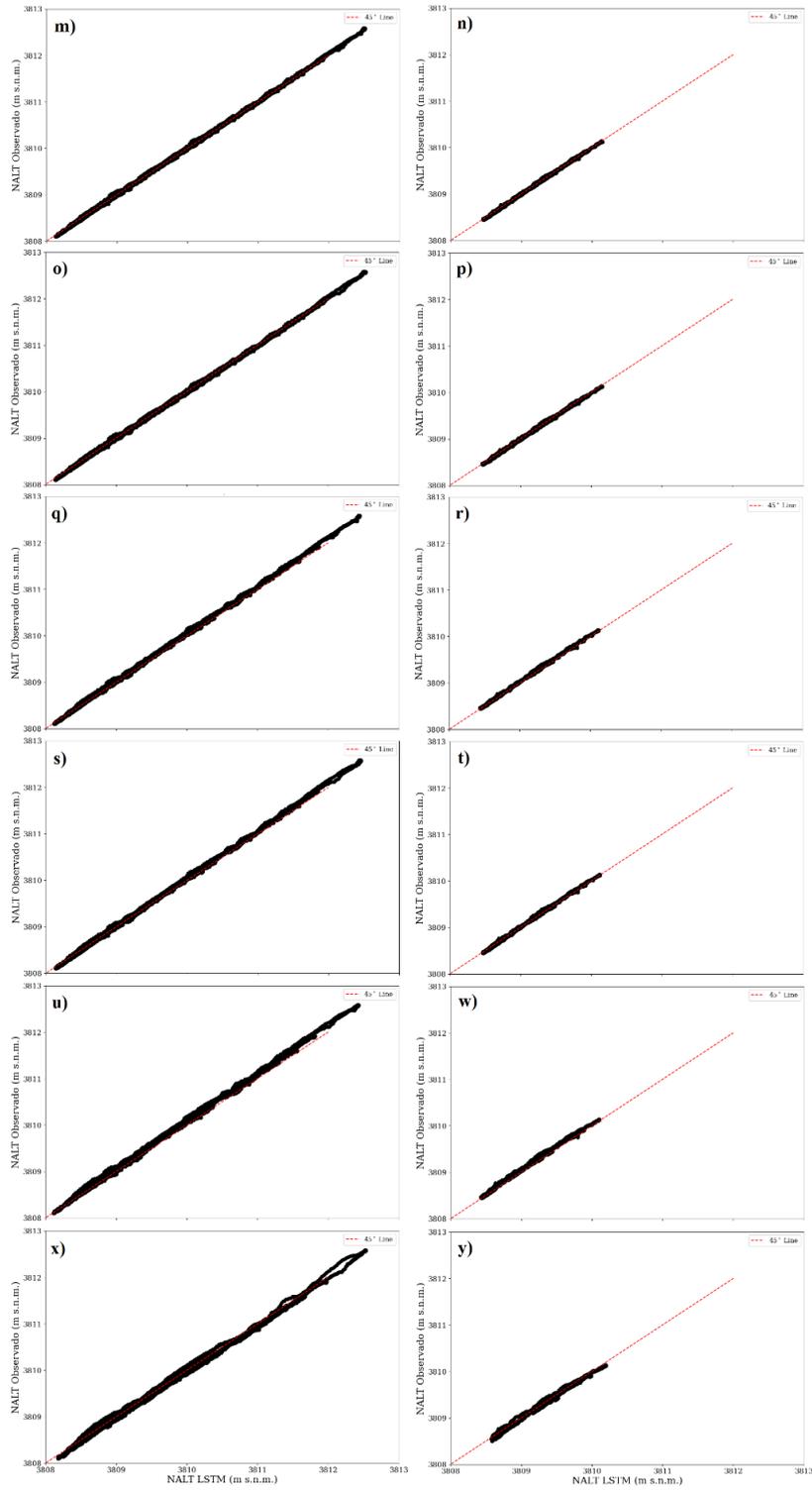


Elaboración propia

**Figura 12**

*Diagrama de dispersión entre niveles de agua pronosticado por la RNN-LSTM con ventanas de tiempo 1, 5, 10, 20, 30, 40, 50, 75, 100, 150, 250 y 365 días y valores observados. Izquierda (entrenamiento) y derecha (prueba).*





Elaboración propia



### **4.3 PRONÓSTICO MENSUAL DE NIVELES DE AGUA DEL LAGO TITICACA BASADA EN EL MODELO RNN-LSTM**

Se analizaron los efectos del tamaño de la ventana en el rendimiento de la RNN-LSTM. Aquí los resultados se presentan para los tamaños de ventanas: 1, 2, 3, 4, 5, 6, 8, 10, 12, 18, 24 y 48 meses (Tabla 3). El tamaño de la ventana de tiempo utilizada para el pronóstico mensual de niveles de agua en el lago Titicaca tuvo un efecto en el desempeño del modelo RNN-LSTM. Se encontró que el modelo tuvo un buen desempeño para todas las ventanas de tiempo evaluadas, con valores altos de NSE y KGE para todas las ventanas. Sin embargo, se observó que el RMSE y el PBIAS fueron menores cuando se utilizaron ventanas de tiempo más largas. Esto indica que el modelo fue capaz de producir pronósticos más precisos y reducir la tendencia a sobreestimar o subestimar los valores observados cuando se utilizó una ventana de tiempo más larga. En general, se recomienda utilizar una ventana de tiempo de al menos 12 meses para obtener los mejores resultados de pronóstico con el modelo RNN-LSTM.

En cuanto a las épocas de entrenamiento, aquí el número de épocas necesarias para alcanzar un alto rendimiento en el pronóstico mensual varía según la ventana de tiempo utilizada. Para ventanas de tiempo más pequeñas, se necesitan menos épocas para alcanzar un alto rendimiento, mientras que, para ventanas de tiempo más grandes, se necesitan más épocas para lograr un rendimiento comparable. La cantidad de épocas necesarias para alcanzar un alto rendimiento aumenta a medida que se aumenta la ventana de tiempo, lo que sugiere que la complejidad del modelo aumenta a medida que se consideran más datos históricos.

**Tabla 3**

*Desempeño de la RNN-LSTM con diferentes ventanas de tiempo para el pronóstico de NALT mensual*

Ventana de tiempo	Épocas	Entrenamiento				Prueba			
		RMSE (m)	NSE	KGE	PBIAS (%)	RMSE (m)	NSE	KGE	PBIAS (%)
1	78	0.1815	0.9736	0.9593	-0.0002	0.1626	0.9379	0.9668	-0.0002
2	72	0.2368	0.9551	0.9368	-0.0001	0.2193	0.8861	0.9349	-0.0002
3	295	0.1478	0.9825	0.9849	0.0002	0.1359	0.9557	0.9727	0.0000
4	387	0.1480	0.9825	0.9824	0.0003	0.1373	0.9543	0.9772	0.0001
5	377	0.1356	0.9853	0.9853	0.0002	0.1181	0.9657	0.9828	0.0000
6	440	0.1404	0.9843	0.9853	0.0005	0.1296	0.9583	0.9788	0.0001
8	267	0.1519	0.9816	0.9680	-0.0009	0.1370	0.9535	0.9689	-0.0002
10	402	0.1366	0.9851	0.9782	-0.0006	0.1198	0.9647	0.9732	-0.0002
12	476	0.1220	0.9881	0.9918	0.0004	0.1030	0.9741	0.9876	0.0001
18	347	0.1298	0.9866	0.9915	0.0008	0.1137	0.9686	0.9865	0.0002
24	285	0.1298	0.9866	0.9884	0.0005	0.1117	0.9698	0.9839	0.0001
48	227	0.1210	0.9886	0.9859	0.0004	0.1156	0.9606	0.9767	0.0001

Elaboración propia

Los resultados de las métricas de desempeño para el rendimiento general de modelo RNN-LSTM para el período de entrenamiento y prueba indican que se desempeñaron muy bien durante el período de entrenamiento, con una RMSE, NSE, KGE y PBIAS promedio de 0.1484 m, 0.9817, 0.9782 y 0.0001 %, con una ligera diferencia en el periodo de prueba con una RMSE, NSE, KGE y PBIAS promedio de 0.1336 m, 0.9541, 0.9742 y 0.0000 %. La red RNN-LSTM funciona mejor en la predicción diaria que en la predicción mensual. Esto se debe al hecho de que la predicción mensual utiliza ventanas



de tiempo más largas, lo que puede aumentar la complejidad del modelo y llevar a un peor rendimiento.

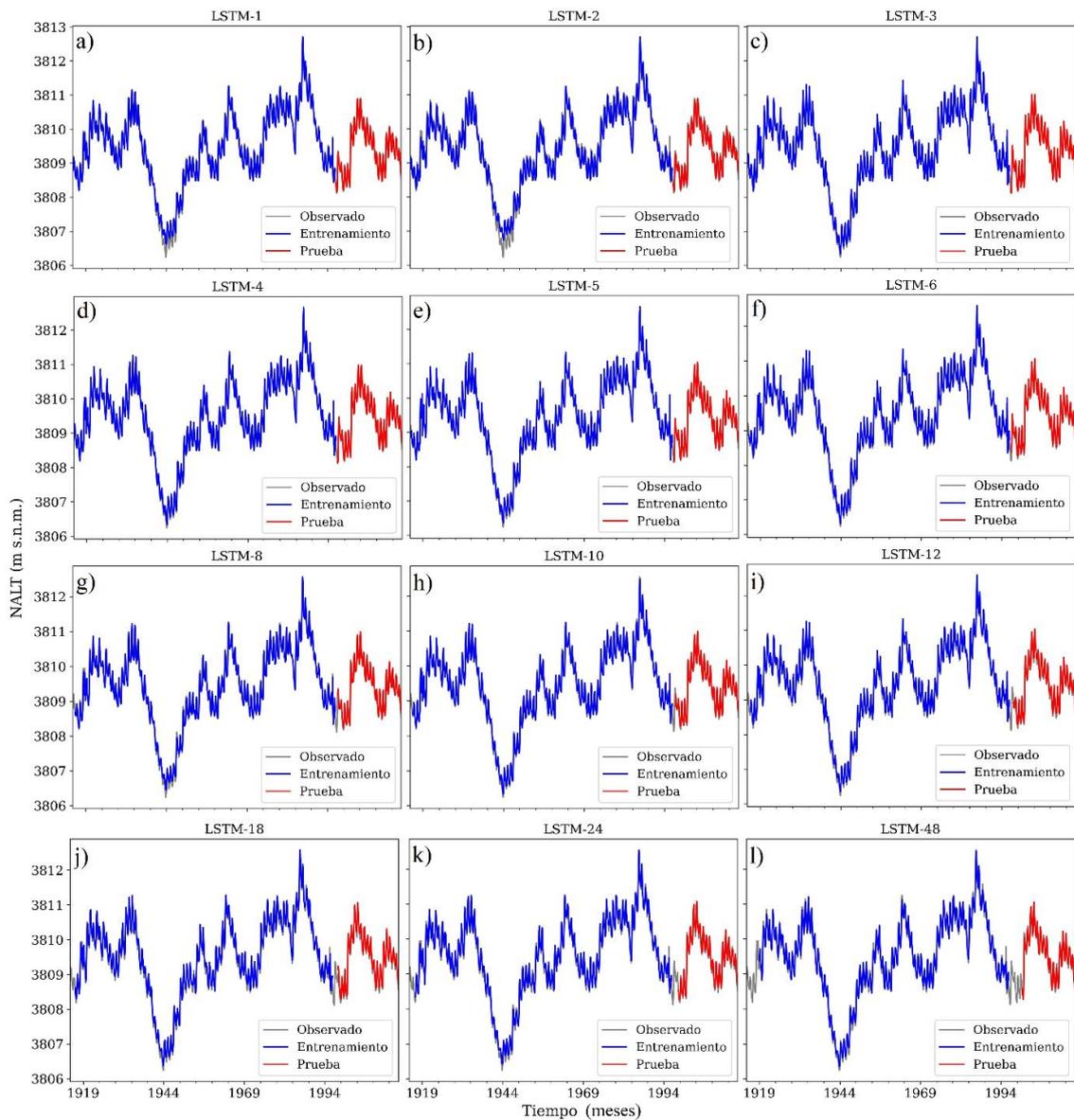
La Figura 14 muestra la serie de tiempo de los datos observados de NALT junto con el período de entrenamiento y prueba del modelo RNN-LSTM utilizado para pronosticar los niveles de agua. El período de entrenamiento del modelo RNN-LSTM está representado por la línea de azul, y fue utilizado para entrenar el modelo con los datos históricos de

niveles de agua. El período de prueba se muestra con línea roja, y fue utilizada para evaluar la capacidad del modelo para pronosticar los niveles de agua. Al comparar los valores observados con los valores pronosticados, se puede observar que el modelo RNN-LSTM fue capaz de ajustarse bien a los datos observados durante el período de entrenamiento. Durante el período de prueba, el modelo también logró pronosticar los niveles de agua con precisión, ya que la línea roja se ajusta bastante bien a los valores observados. Si bien la mayoría de NALT altos y bajos fueron capturados por la RNN-LSTM, esto se ve afectado cuando la ventana de tiempo es menor (1 y 2 meses) y se puede observar que RNN-LSTM no logra representar adecuadamente los niveles más bajos durante el periodo de entrenamiento (año 1944).

**Figura 13**

*Hidrograma de niveles de agua observado y pronosticado basada en el modelo RNN-*

*LSTM con paso de tiempo de 1, 2, 3, 4, 5, 6, 8, 10, 12, 18, 24 y 48 meses.*

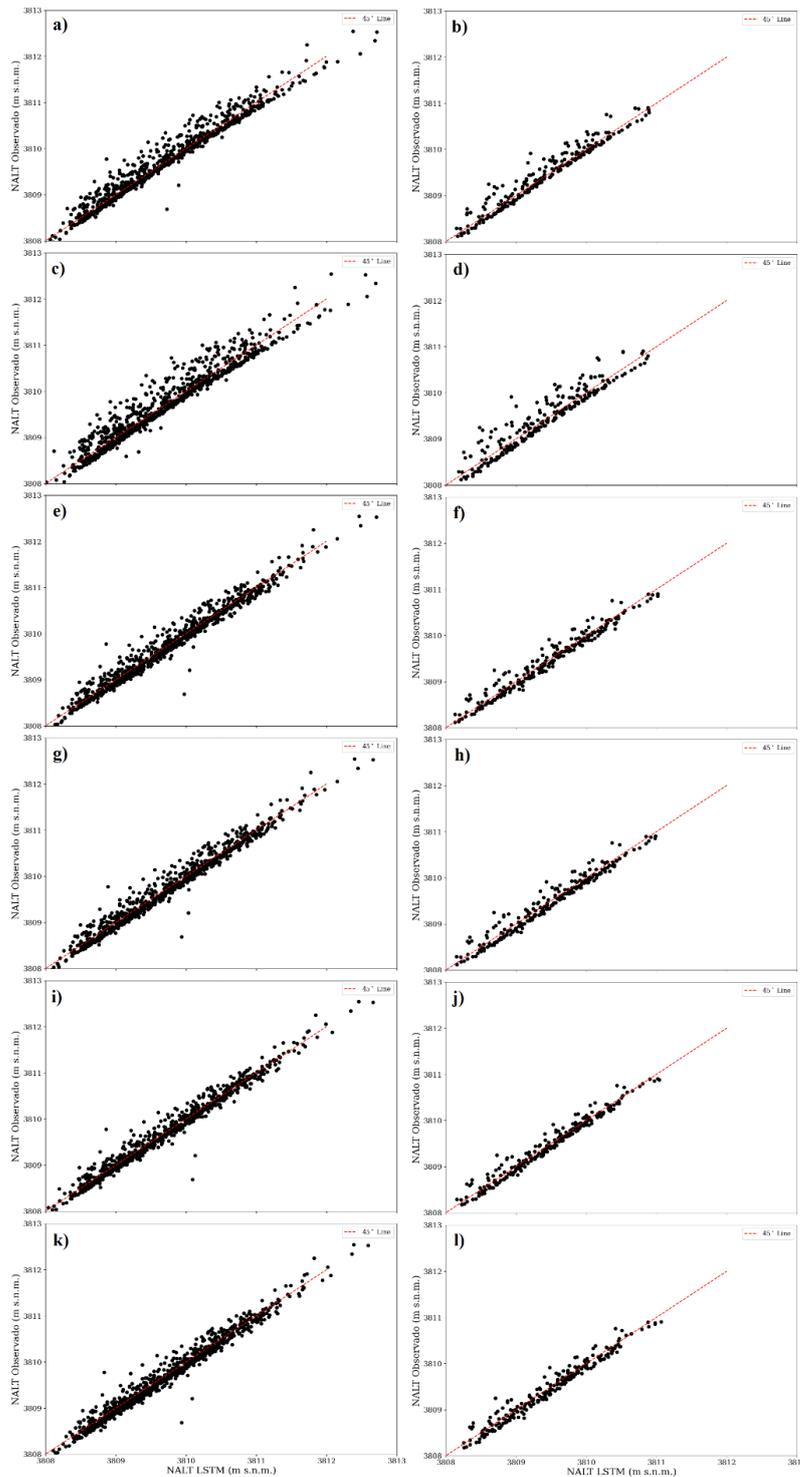


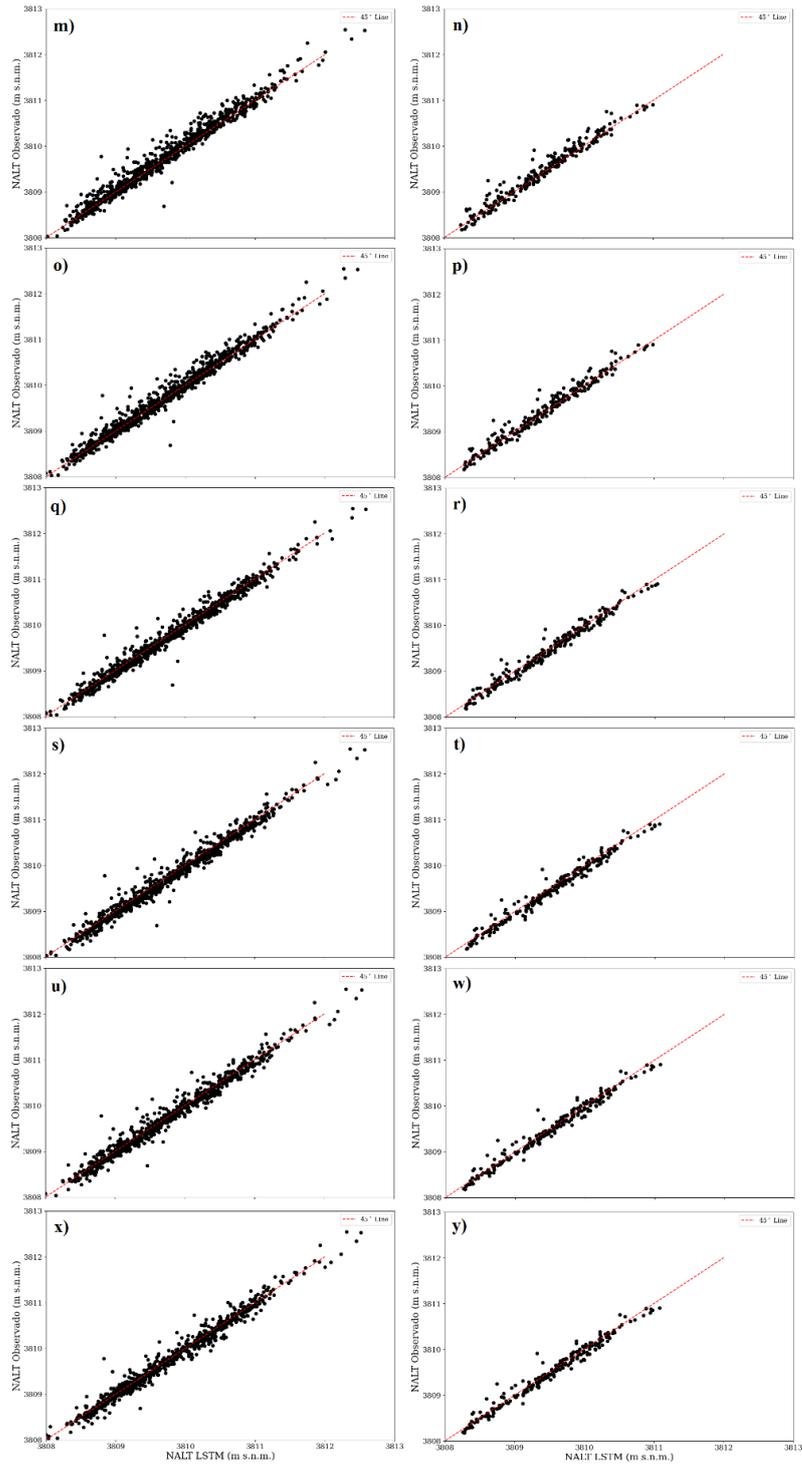
Elaboración propia

### Figura 14

Diagrama de dispersión entre niveles de agua pronosticado por la RNN-LSTM con ventanas de tiempo 1, 2, 3, 4, 5, 6, 8, 10, 12, 18, 24 y 48 meses, y valores observados.

Izquierda (entrenamiento) y derecha (prueba).





Elaboración propia



## V. CONCLUSIONES

En esta investigación, se utilizó una RNN-LSTM para pronosticar los niveles diarios y mensuales del LT utilizando diferentes ventanas de tiempo. Los resultados llevaron a las siguientes conclusiones:

- Es interesante notar que existe una fuerte relación monotónica entre las variables consideradas en diferentes ventanas de tiempo, pero la fuerza de la relación disminuye gradualmente a medida que se consideran ventanas de tiempo más largas. Además, los resultados indican que la relación puede verse afectada por factores externos o estacionales, lo que se refleja en una disminución en la correlación en ciertos intervalos específicos.
- El rendimiento de la red RNN-LSTM se ve afectado por el tamaño de la ventana. Una ventana de tamaño grande deteriora el rendimiento del modelo y la eficiencia computacional. Para el pronóstico diario de NALT, se puede lograr un mejor rendimiento del modelo con un tamaño de ventana igual a 1 día siendo capaz de capturar los valores altos y bajos con mayor precisión. Además, la precisión del modelo no necesariamente mejora con un mayor número de épocas de entrenamiento, y que el sesgo de los resultados es muy bajo.
- Para el pronóstico mensual de NALT la red RNN-LSTM muestra que la mejor precisión se logra con una ventana de tiempo de 12 meses. Además, se observa que el modelo tiene un sesgo muy bajo y que el número de épocas de entrenamiento no tiene un efecto significativo en la precisión del modelo. Estos hallazgos pueden ser de gran utilidad para la gestión sostenible del recurso hídrico del lago Titicaca y otros cuerpos de agua similares.



## VI. RECOMENDACIONES

- Se recomienda utilizar los resultados para la predicción de los niveles de agua del lago Titicaca, debido a que produjeron estimaciones cercanas a los datos observados y tomar decisiones para la solución de problemas.
- Se recomienda realizar nuevas investigaciones utilizando modelos alternativos basados en aprendizaje profundo que permitan realizar pronósticos en un mayor intervalo de tiempo para así lograr obtener pronósticos a largo plazo.
- Se recomienda realizar nuevas investigaciones como la integración de variables adicionales, como la precipitación y la temperatura, para evaluar la influencia de estas variables en los niveles de agua del lago Titicaca, la aplicación de diferentes modelos de aprendizaje automático para comparar su desempeño en la predicción de los niveles de agua del lago Titicaca, determinar la mejor combinación de hiperparámetros para este modelo en particular y la evaluación de la incertidumbre en los pronósticos del modelo y su propagación a través de modelos de gestión del agua para mejorar la toma de decisiones.



## VII. REFERENCIAS BIBLIOGRÁFICAS

- Agudelo Otálora, L. M., Moscoso Barrera, W. D., Paipa Galeano, L. A., & Sciarrotta, C. M. (2018). Comparison of physical models and artificial intelligence for prediction of flood levels. *Tecnología y Ciencias Del Agua*, 9(4). <https://doi.org/10.24850/j-tyca-2018-04-09>
- Baduge, S. K., Thilakarathna, S., Perera, J. S., Arashpour, M., Sharafi, P., Teodosio, B., ... & Mendis, P. (2022). Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Automation in Construction*, 141, 104440. <https://doi.org/10.1016/j.autcon.2022.104440>
- Cardenas, D. C., Altamirano, F. S., & Bernaola, M. Y. (2021, December). Using Neural Networks in River Level Prediction-case study of the river La Leche-Peru. In *2021 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT)* (pp. 1-5). IEEE.
- Chang, F. J., Chen, P. A., Lu, Y. R., Huang, E., & Chang, K. Y. (2014). Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control. *Journal of Hydrology*, 517, 836–846. <https://doi.org/10.1016/j.jhydrol.2014.06.013>
- Chen, S., & Qiao, Y. (2021, August). Short-term forecast of Yangtze River water level based on Long Short-Term Memory neural network. In *IOP Conference Series: Earth and Environmental Science* (Vol. 831, No. 1, p. 012051). IOP Publishing.
- Choi, C., Kim, J., Han, H.; Han, D., Kim, H.S. (2020). Development of Water Level Prediction Models Using Machine Learning in Wetlands: A Case Study of Upo Wetland in South Korea. *Water*, 12, 93. <https://doi.org/10.3390/w12010093>
- Elman, J.L. (1990). Finding structure in time. *Cogn. Sci.* 1990, 14, 179–211.
- Fan, H., Jiang, M., Xu, L., Zhu, H., Cheng, J., & Jiang, J. (2020). Comparison of Long Short Term Memory Networks and the Hydrological Model in Runoff Simulation. *Water* 2020, Vol. 12, Page 175, 12(1), 175. <https://doi.org/10.3390/W12010175>



- Filippo, A., Rebelo Torres, A., Kjerfve, B., & Monat, A. (2012). Application of Artificial Neural Network (ANN) to improve forecasting of sea level. *Ocean & Coastal Management*, 55, 101–110. <https://doi.org/10.1016/J.OCECOAMAN.2011.09.007>
- Guo, Y., & Lai, X. J. (2020). Water level prediction of Lake Poyang based on long short-term memory neural network. *Journal of Lake Sciences*, 32(3), 865-876.
- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1–2), 80–91. <https://doi.org/10.1016/J.JHYDROL.2009.08.003>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
- Hrnjica, B., & Bonacci, O. (2019). Lake Level Prediction using Feed Forward and Recurrent Neural Networks. *Water Resources Management*, 33(7), 2471–2484. <https://doi.org/10.1007/s11269-019-02255-2>
- Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, 424–425, 264–277. <https://doi.org/10.1016/J.JHYDROL.2012.01.011>
- Kostić, S., Stojković, M., & Prohaska, S. (2016). Hydrological flow rate estimation using artificial neural networks: Model development and potential applications. *Applied Mathematics and Computation*, 291, 373–385. <https://doi.org/10.1016/j.amc.2016.07.014>
- Lama, G. L. R., & Sánchez, I. (2020, October). Hybrid models based on mode decomposition and recurrent neural networks for streamflow forecasting in the Chira river in Peru. In 2020 IEEE Engineering International Research Conference (EIRCON) (pp. 1-4). IEEE.
- Laqui, W., Zubieta, R., Rau, P., Mejía, A., Lavado, W., & Ingol, E. (2019). Can artificial neural networks estimate potential evapotranspiration in Peruvian highlands?. *Modeling Earth Systems and Environment*, 5, 1911-1924.



- Laqui, W. (2010). Aplicación de redes neuronales artificiales a la modelización y previsión de caudales medios mensuales del río Huancané. *Revista Peruana Geo-Atmosférica RPGA* (2), 30-44.
- Le, X.-H., Ho, H.V., Lee, G., & Jung, S. (2019). Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting. *Water*, 11, 1387. <https://doi.org/10.3390/w11071387>
- Legates, D.R., & McCabe, G.J. (1999). Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.*, 35, 233–241.
- Lujano, E., Lujano, A., Quispe, J. P., & Lujano, R. (2014). Pronóstico de caudales medios mensuales del río Ilave utilizando modelos de redes neuronales artificiales. *Rev. Investig. Altoandin.*, 16(1), 89–100.
- Lujano, E., Lujano, R., Huamani, J. C., & Lujano, A. (2023). Hydrological modeling based on the KNN algorithm: An application for the forecast of daily flows of the Ramis river, Peru. *Tecnología y Ciencias Del Agua*, 14(2), 169–203. <https://doi.org/10.24850/j-tyca-14-02-05>
- Man, Y., Yang, Q., Shao, J., Wang, G., Bai, L., & Xue, Y. (2022). Enhanced LSTM Model for Daily Runoff Prediction in the Upper Huai River Basin, China. *Engineering*. <https://doi.org/10.1016/J.ENG.2021.12.022>
- Moriasi, D., Arnold, J., Van Liew, M., Bingner, R.; Harmel, R., & Veith, T. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE*, 50, 885–900.
- Morovati, K., Nakhaei, P., Tian, F., Tudaji, M., & Hou, S. (2021). A machine learning framework to predict reverse flow and water level: A case study of Tonle Sap Lake. *Journal of Hydrology*, 603, 127168. <https://doi.org/10.1016/j.jhydrol.2021.127168>
- Nagesh Kumar, D., Srinivasa Raju, K., & Sathish, T. (2004). River flow forecasting using recurrent neural networks. *Water resources management*, 18, 143-161.



- Nash, J.E. & Sutcliffe, J.V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol*, 10, 282–290.
- Nassif, A. B., Shahin, I., Attili, I., Azzeh, M., & Shaalan, K. (2019). Speech recognition using deep neural networks: A systematic review. *IEEE access*, 7, 19143-19165.
- Nifa, K., Boudhar, A., Ouatiki, H., Elyoussfi, H., Bargam, B., & Chehbouni, A. (2023). Deep Learning Approach with LSTM for Daily Streamflow Prediction in a Semi-Arid Area: A Case Study of Oum Er-Rbia River Basin, Morocco. *Water*, 15(2), 262. <https://doi.org/10.3390/w15020262>
- Nti, E. K., Cobbina, S. J., Attafuah, E. E., Opoku, E., & Gyan, M. A. (2022). Environmental sustainability technologies in biodiversity, energy, transportation and water management using artificial intelligence: A systematic review. *Sustainable Futures*, 100068. <https://doi.org/10.1016/j.sftr.2022.100068>
- Nur Adli Zakaria, M., Abdul Malek, M., Zolkepli, M., & Najah Ahmed, A. (2021). Application of artificial intelligence algorithms for hourly river level forecast: A case study of Muda River, Malaysia. *Alexandria Engineering Journal*, 60(4), 4015–4028. <https://doi.org/10.1016/j.aej.2021.02.046>
- Olah, C. (2015). Understanding LSTM Networks. Available online: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/> (accessed on 26 June 2023).
- Plan Director Global Binacional de Protección - Prevención de Inundaciones y Aprovechamiento de los Recursos del Lago Titicaca, río Desaguadero, lago Poopó y lago Salar de Coipasa (Sistema T.D.P.S.). (1993). Diagnóstico de daños por eventos extremos.
- Remesan, R., & Mathew, J. (2015). Hydrological Data Driven Modelling. In *Hydrological Data Driven Modelling*. <https://doi.org/10.1007/978-3-319-09235-5>
- Ren, T., Liu, X., Niu, J., Lei, X., & Zhang, Z. (2020). Real-time water level prediction of cascaded channels based on multilayer perception and recurrent neural network. *Journal of Hydrology*, 585. <https://doi.org/10.1016/j.jhydrol.2020.124783>



- Roberts, W., Williams, G. P., Jackson, E., Nelson, E. J., & Ames, D. P. (2018). Hydrostats: A Python package for characterizing errors between observed and predicted time series. *Hydrology*, 5(4). <https://doi.org/10.3390/hydrology5040066>
- Rojas, CM. 2021. Gestión holística de la Redacción científica. 8va ed. electrónica. Consultado el 10 de mar 2022. Disponible en [https://www.vetcomunicaciones.com.ar/uploadsarchivos/gestion\\_de\\_la\\_redaccion\\_cientifica\\_2021\\_pdf.pdf](https://www.vetcomunicaciones.com.ar/uploadsarchivos/gestion_de_la_redaccion_cientifica_2021_pdf.pdf)
- Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536.
- Saranya, A., & Subhashini, R. (2023). A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends. *Decision Analytics Journal*, 100230. <https://doi.org/10.1016/j.dajour.2023.100230>
- Sarker, I. H. (2021). Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2(6), 420.
- Seo, Y., Kim, S., Kisi, O., & Singh, V. P. (2015). Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *Journal of Hydrology*, 520, 224–243. <https://doi.org/10.1016/j.jhydrol.2014.11.050>
- Vargas Valverde, C. M. (2019). Modelo univariante para describir y predecir el comportamiento de los niveles medios de agua del Lago Titicaca – Puno.
- Veintimilla-Reyes, J., Cisneros, F., & Vanegas, P. (2016). Artificial Neural Networks Applied to Flow Prediction: A Use Case for the Tomebamba River. *Procedia Engineering*, 162, 153–161. <https://doi.org/10.1016/j.proeng.2016.11.031>
- Vu, T.T., Li, L., & Jun, K.S. (2018). Evaluation of multi-satellite precipitation products for streamflow simulations: A case study for the Han River Basin in the Korean Peninsula, East Asia. *Water*, 10, 642.



- Werbos, P.J. (1988). Generalization of backpropagation with application to a recurrent gas market model. *Neural Netw.* 1, 339–356.
- Yan, S. (2015). Comprender LSTM y sus diagramas. Disponible en línea: <https://medium.com/mlreview/understanding-lstm-and-its-diagrams-37e2f46f1714> (consultado el 01 de junio de 2023).
- Yaseen, Z. M., El-Shafie, A., Jaafar, O., Afan, H. A., & Sayl, K. N. (2015). Artificial intelligence based models for stream-flow forecasting: 2000–2015. *Journal of Hydrology*, 530, 829-844.
- Ye, X. C., Liu, F. H., Zhang, Z. X., & Xu, C. Y. (2020). Quantifying the impact of compounding influencing factors to the water level decline of China's largest freshwater lake. *Journal of Water Resources Planning and Management*, 146(6), 05020006.
- Zakaria, M. N. A., Ahmed, A. N., Malek, M. A., Birima, A. H., Khan, M. M. H., Sherif, M., & Elshafie, A. (2023). Exploring machine learning algorithms for accurate water level forecasting in Muda river, Malaysia. *Heliyon*, 9(7).
- Zhu, S., Hrnjica, B., Ptak, M., Choiński, A., & Sivakumar, B. (2020). Forecasting of water level in multiple temperate lakes using machine learning models. *Journal of Hydrology*, 585. <https://doi.org/10.1016/j.jhydrol.2020.124819>
- Zounemat-Kermani, M., Matta, E., Cominola, A., Xia, X., Zhang, Q., Liang, Q., & Hinkelmann, R. (2020). Neurocomputing in surface water hydrology and hydraulics: A review of two decades retrospective, current status and future prospects. *Journal of Hydrology*, 588, 125085.



## ANEXOS

**ANEXO 1:** Código fuente del modelo RNN-LSTM para pronóstico de niveles de agua del Lago Titicaca.

El código fuente se encuentra almacenado como repositorio de GitHub: [rlujano/LSTM-NALT: RNN-LSTM for dayli water level forecasting \(github.com\)](https://github.com/rlujano/LSTM-NALT)



## ANEXO 2: Declaración jurada de autenticidad de tesis



Universidad Nacional  
del Altiplano Puno



Vicerrectorado  
de Investigación



Repositorio  
Institucional

### DECLARACIÓN JURADA DE AUTENTICIDAD DE TESIS

Por el presente documento, Yo Rene Lujano Laura, identificado con DNI N° 47263576 en mi condición de egresado de:

**Escuela Profesional**,  **Programa de Segunda Especialidad**,  **Programa de Maestría o Doctorado** de Ingeniería de Sistemas, informo que he elaborado el/la  **Tesis** o  **Trabajo de Investigación** denominada:

“DESARROLLO DE MODELOS DE REDES NEURONALES RECURRENTE PARA EL PRONÓSTICO DE NIVELES DE AGUA DEL LAGO TITICACA”

Es un tema original.

Declaro que el presente trabajo de tesis es elaborado por mi persona y **no existe plagio/copia** de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso, o similar) presentado por persona natural o jurídica alguna ante instituciones académicas, profesionales, de investigación o similares, en el país o en el extranjero.

Dejo constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no asumiré como tuyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o Internet.

Asimismo, ratifico que soy plenamente consciente de todo el contenido de la tesis y asumo la responsabilidad de cualquier error u omisión en el documento, así como de las connotaciones éticas y legales involucradas.

En caso de incumplimiento de esta declaración, me someto a las disposiciones legales vigentes y a las sanciones correspondientes de igual forma me someto a las sanciones establecidas en las Directivas y otras normas internas, así como las que me alcancen del Código Civil y Normas Legales conexas por el incumplimiento del presente compromiso

Puno 18 de Enero del 2024

FIRMA (obligatoria)



Huella



### ANEXO 3: Autorización para el depósito de tesis en el repositorio institucional



Universidad Nacional  
del Altiplano Puno



Vicerrectorado  
de Investigación



Repositorio  
Institucional

#### AUTORIZACIÓN PARA EL DEPÓSITO DE TESIS O TRABAJO DE INVESTIGACIÓN EN EL REPOSITORIO INSTITUCIONAL

Por el presente documento, Yo Rene Lujano Laura,  
identificado con DNI 47263576 en mi condición de egresado de:

Escuela Profesional,  Programa de Segunda Especialidad,  Programa de Maestría o Doctorado

de Ingeniería de Sistemas, informo que he elaborado el/la  Tesis o  Trabajo de  
Investigación denominada:

“DESARROLLO DE MODELOS DE REDES NEURONALES RECURRENTE PARA EL PRONÓSTICO  
DE NIVELES DE AGUA DEL LAGO TITICACA”

para la obtención de  Grado,  Título Profesional o  Segunda Especialidad.

Por medio del presente documento, afirmo y garantizo ser el legítimo, único y exclusivo titular de todos los derechos de propiedad intelectual sobre los documentos arriba mencionados, las obras, los contenidos, los productos y/o las creaciones en general (en adelante, los “Contenidos”) que serán incluidos en el repositorio institucional de la Universidad Nacional del Altiplano de Puno.

También, doy seguridad de que los contenidos entregados se encuentran libres de toda contraseña, restricción o medida tecnológica de protección, con la finalidad de permitir que se puedan leer, descargar, reproducir, distribuir, imprimir, buscar y enlazar los textos completos, sin limitación alguna.

Autorizo a la Universidad Nacional del Altiplano de Puno a publicar los Contenidos en el Repositorio Institucional y, en consecuencia, en el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, sobre la base de lo establecido en la Ley N° 30035, sus normas reglamentarias, modificatorias, sustitutorias y conexas, y de acuerdo con las políticas de acceso abierto que la Universidad aplique en relación con sus Repositorios Institucionales. Autorizo expresamente toda consulta y uso de los Contenidos, por parte de cualquier persona, por el tiempo de duración de los derechos patrimoniales de autor y derechos conexos, a título gratuito y a nivel mundial.

En consecuencia, la Universidad tendrá la posibilidad de divulgar y difundir los Contenidos, de manera total o parcial, sin limitación alguna y sin derecho a pago de contraprestación, remuneración ni regalía alguna a favor mío; en los medios, canales y plataformas que la Universidad y/o el Estado de la República del Perú determinen, a nivel mundial, sin restricción geográfica alguna y de manera indefinida, pudiendo crear y/o extraer los metadatos sobre los Contenidos, e incluir los Contenidos en los índices y buscadores que estimen necesarios para promover su difusión.

Autorizo que los Contenidos sean puestos a disposición del público a través de la siguiente licencia:

Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visita: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

En señal de conformidad, suscribo el presente documento.

Puno 18 de Enero del 2024

FIRMA (obligatoria)



Huella