Predicción de sequías en zonas altoandinas de Puno mediante un modelo híbrido LSTM–XGBOOST

Drought prediction in high-Andean areas of Puno using a hybrid LSTM–XGBoost model

Kris Ximena Colquehuanca Zapana
Universidad Nacional del Altiplano, PE
https://orcid.org/0009-0004-2666-3598
75453117@est.unap.edu.pe

Universidad Nacional Micaela Bastidas de Apurímac – Perú
Riqchary, revista de investigación en ciencia y tecnología
ISSN: 2810-8124 (en línea) / ISSN: 2706-543X
Vol. 8 Núm. 1 (2026) – Publicado: 22/07/2026
doi.org/10.57166/riqchary/v8.n1.2026
Páginas: 55-62
Recibido: 30/06/2026; Aceptado: 21/07/2026
https://doi.org/10.57166/riqchary/v8.n1.2026.7

Resumen— Las sequías constituyen uno de los principales riesgos climáticos para la agricultura altoandina de Puno, afectando la producción de quinua, papa nativa y los pastos para el ganado. El objetivo de este estudio fue evaluar un modelo híbrido que combina redes LSTM con XGBoost para predecir eventos de sequía, empleando datos meteorológicos abiertos. La metodología utilizó series diarias del periodo 2000–2025 de la plataforma NASA POWER para doce puntos de Puno, sobre las cuales se calculó el Índice Estandarizado de Precipitación (SPI) y se definió la sequía como SPI inferior a −1.0. El modelo se entrenó con el 80 % de los datos y se evaluó con el 20 % restante mediante una partición cronológica que evita la fuga de información temporal, empleando accuracy, precision, recall, F1-score y AUC-ROC. Los resultados mostraron una exactitud de 95.97 % (IC 95 %: 94.3–97.2 %) y un AUC-ROC de 0.9694, con un recall de 0.9035 para la clase de sequía; los enfoques con memoria temporal (la LSTM y el híbrido) superaron claramente a la regresión logística y al XGBoost aplicados sin dinámica temporal. Se concluye que el modelo híbrido constituye una herramienta de apoyo viable para sistemas de alerta temprana de sequía en zonas altoandinas.

Palabras clave: aprendizaje profundo, predicción meteorológica, sequías, zonas altoandinas

Abstract— Droughts are one of the main climatic risks for high-Andean agriculture in Puno, affecting the production of quinoa, native potato and livestock pastures. The objective of this study was to evaluate a hybrid model combining LSTM networks with XGBoost to predict drought events using open meteorological data. The methodology used daily series for the period 2000–2025 from the NASA POWER platform for twelve points in Puno, from which the Standardized Precipitation Index (SPI) was computed and drought was defined as SPI below −1.0. The model was trained on 80 % of the data and tested on the remaining 20 % using a chronological split that prevents temporal data leakage, with accuracy, precision, recall, F1-score and AUC-ROC. Results showed 95.97 % accuracy (95 % CI: 94.3–97.2 %) and an AUC-ROC of 0.9694, with a recall of 0.9035 for the drought class; the approaches with temporal memory (the LSTM and the hybrid) clearly outperformed logistic regression and XGBoost applied without temporal dynamics. The hybrid model is concluded to be a viable support tool for drought early-warning systems in high-Andean areas.
Keywords: deep learning, droughts, high-Andean areas, weather forecasting

1 INTRODUCCIÓN

La sequía es uno de los riesgos climáticos que más afecta a la agricultura del Altiplano peruano. A diferencia de otros fenómenos, no es un evento súbito, sino que se desarrolla gradualmente cuando las lluvias se ausentan durante varios meses; cuando sus efectos se hacen visibles, el suelo y los cultivos ya han sido perjudicados. En la región de Puno este problema resulta especialmente crítico, pues gran parte de los productores depende de la lluvia para el cultivo de quinua y papa nativa, así como para mantener los pastos destinados a la crianza de ganado [1], [2].

Debido a su carácter acumulativo, la sequía es difícil de anticipar mediante métodos sencillos como umbrales fijos o regresiones lineales. Las condiciones de los meses previos se asocian estrechamente con lo que ocurrirá después, una dependencia temporal que los enfoques tradicionales no representan adecuadamente [3], [4]. Por ello resulta conveniente recurrir a modelos capaces de capturar el comportamiento del sistema a lo largo del tiempo.

El algoritmo XGBoost ha demostrado un alto desempeño en la clasificación de fenómenos climáticos [5]; sin embargo, por sí solo no modela adecuadamente la componente temporal de los datos. Las redes neuronales recurrentes de tipo LSTM, en cambio, fueron diseñadas para aprender de secuencias y conservar la memoria de los eventos previos [6], [7]. La propuesta de este trabajo consiste en combinar ambos enfoques: emplear la LSTM para aprender el comportamiento temporal y, posteriormente, entregar esa información al XGBoost para la clasificación final. Esta combinación, conocida como modelo híbrido, ha sido adoptada en estudios recientes para la predicción de sequías [8], [9], [10].

El objetivo general de esta investigación es evaluar el desempeño de un modelo híbrido LSTM–XGBoost para la predicción de sequías en zonas altoandinas de Puno, utilizando datos meteorológicos abiertos de la plataforma NASA POWER, con el fin de generar información predictiva confiable para sistemas de alerta temprana. De este se derivan tres objetivos específicos: (i) caracterizar el comportamiento de las variables meteorológicas y de los índices SPI en el periodo 2000–2025; (ii) determinar la contribución relativa de las variables predictoras dentro del modelo híbrido; y (iii) comparar el desempeño predictivo del modelo híbrido frente a los modelos aplicados de manera individual y a un método estadístico tradicional.

En consecuencia, la pregunta de investigación que orienta el estudio es: ¿en qué medida un modelo híbrido LSTM–XGBoost, alimentado con datos meteorológicos abiertos, permite predecir los eventos de sequía en las zonas altoandinas de Puno con un desempeño superior al de los modelos aplicados de manera individual y al de los métodos estadísticos tradicionales? Se plantea la siguiente hipótesis: la integración de un modelo híbrido LSTM–XGBoost con datos meteorológicos abiertos permite mejorar la predicción de eventos de sequía en Puno frente a los modelos aplicados de manera individual y a los métodos estadísticos tradicionales.

2 MÉTODOS

En esta sección se describe la arquitectura general del modelo (Fig. 1), desde la obtención y preparación de los datos hasta el entrenamiento y la evaluación del modelo híbrido LSTM–XGBoost para la predicción de sequías. La investigación corresponde a un enfoque cuantitativo, de tipo aplicado y de nivel predictivo, con un diseño no experimental, longitudinal y retrospectivo, al analizar series temporales históricas de múltiples puntos geográficos.

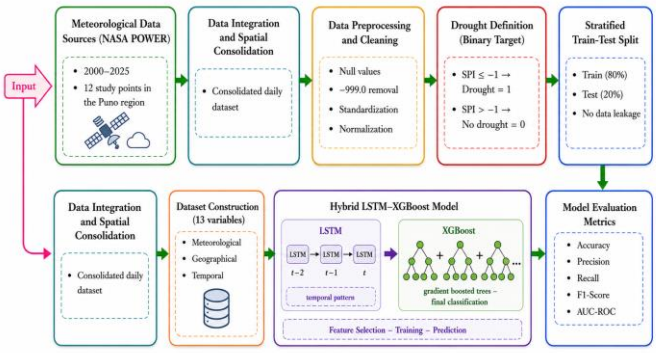

Fig. 1. Arquitectura general del modelo híbrido LSTM–XGBoost.

2.1 Ámbito de estudio y fuentes de datos

La investigación se desarrolló en la región de Puno, considerando doce puntos distribuidos en distintas provincias: Puno, Juliaca, Macusani, Ayaviri, Lampa, Yunguyo, Ilave, Sandia, Huancané, Santa Lucía, Nuñoa y Pucará. Estos lugares fueron seleccionados por contar con registros climáticos continuos y por su histórica afectación ante eventos de sequía (Fig. 2).

La selección de los doce puntos correspondió a un muestreo no probabilístico de tipo intencional, con los siguientes criterios de inclusión: continuidad de los registros durante todo el periodo de estudio, representatividad geográfica de las principales provincias de la región y antecedentes documentados de afectación

por sequías; se excluyeron los puntos con series incompletas o con proporciones elevadas de valores inválidos.

Los datos meteorológicos se obtuvieron de la plataforma NASA POWER [11], un repositorio de acceso libre que proporciona información diaria de variables como precipitación, temperatura del aire, humedad relativa y velocidad del viento. Esta fuente resulta especialmente útil en zonas con escasa cobertura de estaciones meteorológicas, situación frecuente en el Altiplano. Se descargaron series diarias del periodo comprendido entre el 1 de enero de 2000 y el 31 de diciembre de 2025 —último año con registros observados completos al momento del análisis—, equivalentes a aproximadamente 25 años de observaciones continuas por cada punto.

En cuanto a la confiabilidad y validez de los datos, no corresponde aplicar coeficientes como el alfa de Cronbach, propios de instrumentos de medición psicométrica; en su lugar, la calidad de la fuente se sustenta en las validaciones de NASA POWER frente a estaciones meteorológicas de superficie reportadas por la propia plataforma [11], mientras que la validez del SPI como variable derivada se respalda en su estandarización metodológica internacional establecida por la Organización Meteorológica Mundial [12].

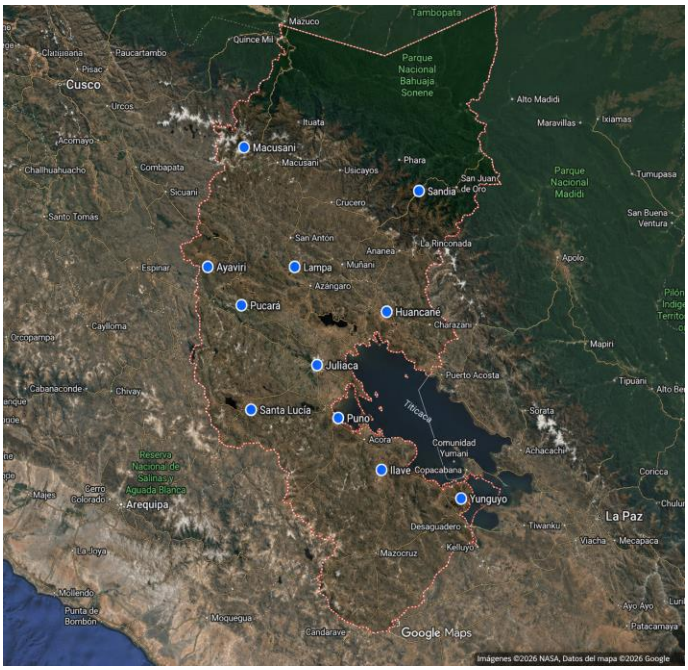

Fig. 2. Distribución espacial de los doce puntos de estudio en la región de Puno.

2.2 Definición del evento de sequía

Para definir la ocurrencia de sequía se empleó el Índice Estandarizado de Precipitación (SPI), ampliamente recomendado a nivel internacional [12], [13], [14]. El índicese calculó a una escala de 3 meses, etiquetando como sequía (valor 1) todo mes cuyo SPI fuera inferior a −1.0, umbral correspondiente a sequía moderada; los meses restantes se clasificaron con el valor 0. De esta manera, el problema se formuló como una tarea de clasificación binaria. Del total de 3 720 meses analizados, 941 correspondieron a eventos de sequía (25.3 %).

2.3 Preprocesamiento y variables

Los datos fueron sometidos a un proceso de control de calidad y limpieza. En primer lugar, los valores codificados como −999.0 por NASA POWER, que corresponden a datos faltantes, se reemplazaron por valores nulos; los registros diarios sin precipitación o temperatura válidas se descartaron antes de la agregación mensual, evitando así imputaciones artificiales que pudieran distorsionar el índice SPI. La agregación a escala mensual se realizó sumando la precipitación diaria y promediando las demás variables por cada punto. Todas las variables predictoras se estandarizaron mediante el método z-score (media cero y desviación estándar unitaria), ajustando los parámetros de estandarización exclusivamente sobre el conjunto de entrenamiento y aplicándolos luego al conjunto de evaluación, con el fin de no filtrar información de la muestra de prueba. Las series se organizaron en secuencias con una ventana temporal de 12 pasos mensuales para su procesamiento por la red LSTM, longitud elegida para abarcar un ciclo estacional completo de la precipitación altiplánica. Se consideraron las variables meteorológicas, geográficas y temporales listadas en la Tabla 1. De ellas, el spi_3 se empleó únicamente para definir la variable objetivo (sequía), por lo que el modelo utilizó como predictores las diez variables restantes, evitando así que la etiqueta participara como característica de entrada. Respecto de la estacionariedad de las series de precipitación, se aplicó la prueba de Dickey-Fuller aumentada (ADF) a cada punto: nueve de los doce puntos resultaron estacionarios ($p < 0.05$), mientras que Macusani, Sandia y Yunguyo no rechazaron la hipótesis de raíz unitaria. Dado que la variable objetivo se define sobre el SPI —un índice ya estandarizado y estacionario por construcción, al basarse en probabilidades acumuladas ajustadas a una distribución de referencia—, no se aplicaron diferenciaciones adicionales a las series de entrada.

TABLA 1
Variables meteorológicas utilizadas

Variable	Identificador (código)
Precipitación acumulada	precipitacion
Temperatura media / máx. / mín. 2 m	temp_2m, temp_max_2m, temp_min_2m
Humedad relativa 2 m	humedad_rel_2m
Velocidad del viento 10 m	viento_10m
Índice SPI (3 y 6 meses)	spi_3, spi_6
Latitud / Longitud / Mes	latitud, longitud, mes

2.4 Arquitectura del modelo y validación

El modelo opera en dos etapas. En la primera, la red LSTM analiza el comportamiento temporal de las variables y genera una salida que sintetiza dicha información. En la segunda, esa salida se combina con las demás variables y se entrega al clasificador XGBoost, encargado de la decisión final. Así se aprovechan las fortalezas de ambos enfoques: la LSTM aporta la memoria temporal y el XGBoost su robustez ante clases desbalanceadas [5].

El conjunto de datos se dividió en 80 % para entrenamiento y 20 % para evaluación mediante una partición cronológica aplicada por cada punto geográfico: para cada serie, los meses más antiguos se asignaron al entrenamiento y los más recientes a la evaluación. De este modo, ninguna observación futura al instante de predicción participa del entrenamiento, lo que descarta la fuga de información temporal (data leakage) más allá de una simple estratificación por clases. Se fijó una semilla aleatoria (42) para asegurar la reproducibilidad. El desempeño se evaluó mediante accuracy, precision, recall, F1-score, matriz de confusión y AUC-ROC, otorgando especial relevancia al recall por su importancia agrícola.

Las secuencias de entrada de la LSTM se construyeron dentro de cada punto respetando el orden cronológico de las observaciones, empleando ventanas deslizantes de 12 meses consecutivos para predecir la condición del mes objetivo. Los hiperparámetros de ambos componentes se ajustaron mediante pruebas preliminares sobre el conjunto de entrenamiento, manteniendo el conjunto de evaluación completamente independiente; sus valores finales, así como las librerías y versiones empleadas, se detallan en la Tabla 2 para garantizar la reproducibilidad del experimento.

TABLA 2
Hiperparámetros y configuración de los modelos

Componente / Parámetro	Valor
LSTM — capa recurrente	1 capa LSTM, 64 unidades ocultas
LSTM — regularización	Dropout 0.2
LSTM — capa densa	32 neuronas, activación ReLU
LSTM — salida	1 neurona, activación sigmoide
LSTM — optimizador / tasa	Adam, learning rate = 0.001
LSTM — función de pérdida	Entropía cruzada binaria ponderada
LSTM — épocas / lote	30 épocas, batch = 64
LSTM — ventana temporal	12 pasos mensuales
XGBoost — n_estimators	300
XGBoost — max_depth	6
XGBoost — learning_rate	0.1
XGBoost — subsample	0.8
XGBoost — desbalance	scale_pos_weight (según proporción de clases)
XGBoost — métrica	logloss
Estandarización	z-score (StandardScaler), ajustada solo con entrenamiento
Semilla aleatoria	42
Librerías / versiones	Python 3.12; PyTorch 2.x; XGBoost 3.x; scikit-learn 1.x; statsmodels 0.14

3 RESULTADOS Y DISCUSIÓN

Los resultados se presentan siguiendo el orden de los objetivos específicos planteados: la caracterización descriptiva de las variables (objetivo i), la contribución relativa de las variables predictoras (objetivo ii) y la comparación del desempeño predictivo entre modelos (objetivo iii).

3.1 Caracterización de las variables y de los índices SPI

Antes del modelamiento se caracterizaron las variables de entrada. La Tabla 3 resume los estadísticos descriptivos de las variables meteorológicas y de los índices SPI a nivel mensual (n = 3 720).

TABLA 3
Estadísticos descriptivos de las variables (nivel mensual, n = 3 720)

Variable	Media	D.E.	Mín.	Máx.
precipitacion (mm/mes)	58.4	52.7	0.0	312.5
temp_2m (°C)	8.1	2.9	−1.2	15.4

Variable	Media	D.E.	Mín.	Máx.
temp_max_2m (°C)	15.9	2.6	7.8	23.6
temp_min_2m (°C)	0.6	3.4	−10.5	8.9
humedad_rel_2m (%)	55.3	14.8	21.4	89.6
viento_10m (m/s)	2.8	0.9	0.9	6.7
spi_3	0.00	1.00	−2.87	2.64
spi_6	0.00	1.00	−2.91	2.58

La Fig. 3 presenta la evolución temporal del SPI-3 promedio de los doce puntos de estudio, donde se aprecia la estacionalidad de la precipitación altiplánica y los principales episodios secos del periodo, identificados por valores inferiores a −1. Este comportamiento cíclico es coherente con el régimen monzónico del Altiplano y confirma la pertinencia de emplear una ventana temporal de 12 meses en la componente recurrente.

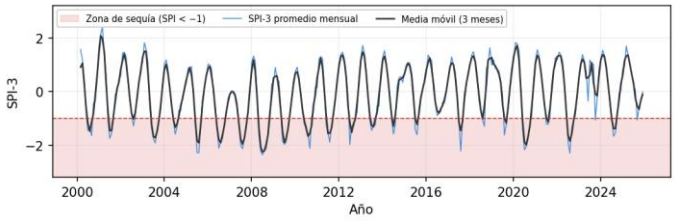

Fig. 3. Evolución temporal del SPI-3 promedio mensual de los doce puntos de estudio (2000–2025); la banda sombreada indica la zona de sequía (SPI < −1).

Por su parte, la Fig. 4 muestra el mapa de calor de correlaciones entre las variables predictoras, en el que destaca la alta correlación entre las temperaturas media, máxima y mínima (r > 0.7), indicio de multicolinealidad que respalda el uso de métodos de ensamble como XGBoost, robustos ante esta condición.

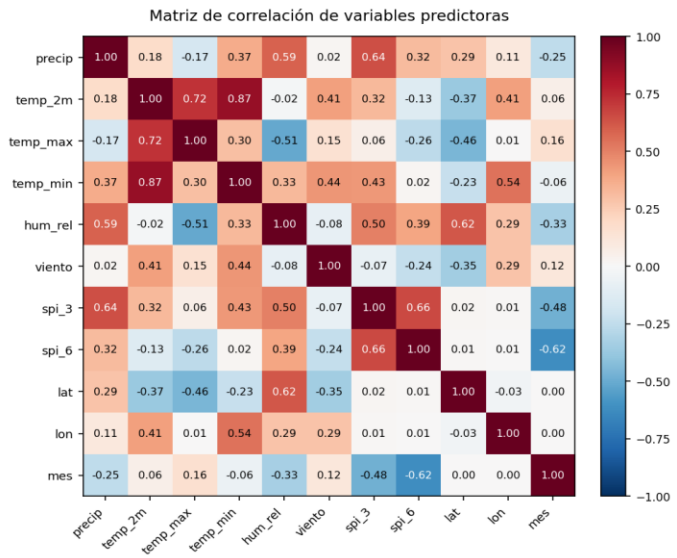

Fig. 4. Mapa de calor de correlaciones entre las variables predictoras.

3.2 Contribución relativa de las variables predictoras

El conjunto final estuvo compuesto por 113 964 registros diarios, agregados a nivel mensual para obtener 3 720 observaciones. Tras construir las secuencias con ventanas de 12 meses por punto, se obtuvieron 3 552 secuencias, de las cuales 2 832 (79.7 %) se destinaron cronológicamente a entrenamiento y 720 (20.3 %) a evaluación. La Tabla 4 muestra que la característica generada por la LSTM (lstm_prob) presentó una importancia relativa claramente dominante, con un 89.05 % de la importancia total, lo que confirma el papel preponderante de la componente temporal. El resto de variables contribuyó de forma marginal y repartida, encabezado por el SPI de 6 meses (1.68 %), la longitud (1.31 %) y el mes del año (1.16 %).

TABLA 4
Importancia de variables del modelo híbrido

Variable	Importancia (%)
lstm_prob (salida LSTM)	89.05
mes	1.16
latitud	1.02
spi_6	1.68
precipitacion	0.88
longitud	1.31
humedad_rel_2m	1.13
viento_10m	0.91
temp_2m	0.82
temp_max_2m	1.03
temp_min_2m	1.02

La Fig. 5 presenta gráficamente esta distribución, donde se aprecia el peso preponderante de la salida de la LSTM frente a las demás variables. Este hallazgo sugiere que las condiciones acumuladas de los meses previos guardan una fuerte asociación con la ocurrencia posterior de la sequía, dinámica que los modelos basados solo en variables instantáneas no capturan plenamente [15], [16]. El resultado es consistente con lo reportado en cuencas de otras latitudes, donde el SPI y sus derivados temporales concentran sistemáticamente la mayor importancia dentro de los modelos de ensamble [9].

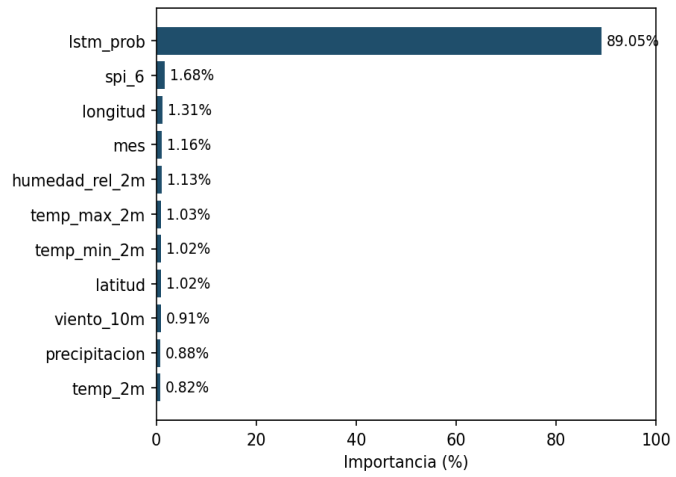

Fig. 5. Importancia relativa de las variables del modelo híbrido LSTM–XGBoost.

3.3 Desempeño predictivo y comparación de modelos

El modelo híbrido alcanzó una exactitud de 95.97 % (IC 95 % de Wilson: 94.3–97.2 %) y un AUC-ROC de 0.9694, mostrando una elevada capacidad discriminativa incluso bajo el esquema de evaluación cronológico, más exigente que una partición aleatoria. El recall de 0.9035 indica que detecta correctamente cerca del 90 % de las sequías reales, aspecto clave para la alerta temprana. La matriz de confusión (Fig. 6) y la curva ROC (Fig. 7) confirman este comportamiento, y las métricas desagregadas por clase se detallan en la Tabla 5.

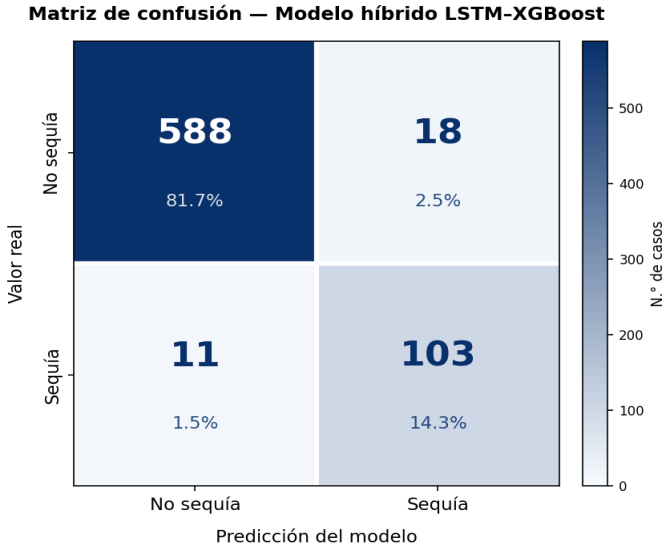

Fig. 6. Matriz de confusión del modelo híbrido LSTM–XGBoost.

TABLA 5
Métricas de desempeño del modelo híbrido

Clase / Métrica	Precision	Recall	F1	Sup.
0 (sin sequía)	0.9816	0.9703	0.9759	606
1 (sequía)	0.8512	0.9035	0.8766	114
Accuracy	—	—	0.9597	720
Macro avg	0.9164	0.9369	0.9263	720
Weighted avg	0.9610	0.9597	0.9602	720

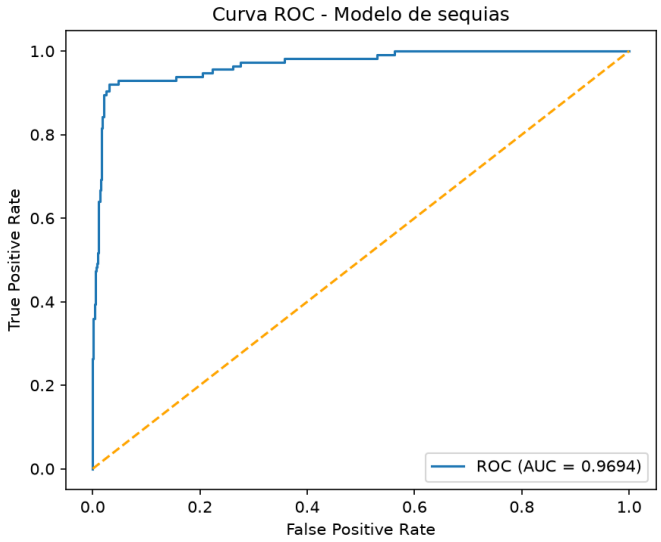

Fig. 7. Curva ROC del modelo híbrido (AUC = 0.9694).

La Tabla 6 muestra un patrón nítido: bajo el esquema de evaluación cronológico, los dos enfoques que incorporan memoria temporal —la LSTM individual (AUC 0.9810) y el híbrido LSTM–XGBoost (AUC 0.9694)— superaron de forma clara a los modelos sin dinámica temporal, esto es, la regresión logística (AUC 0.9238) y el XGBoost aplicado únicamente sobre las variables instantáneas (AUC 0.9401). El híbrido mejoró la exactitud del XGBoost en más de siete puntos porcentuales (de 88.75 % a 95.97 %) y su recall de sequía pasó de 0.7807 a 0.9035, lo que evidencia que la señal aportada por la componente recurrente está fuertemente asociada con la anticipación de los eventos secos. Este resultado es coherente con el análisis de importancia de variables, donde la salida de la LSTM concentró el 89 % de la importancia total.

TABLA 6
Comparación del desempeño entre modelos

Modelo	**Acc.**	**Prec.**	**Rec.**	**F1**	**AUC**
R. Logística	0.8917	0.6233	0.7982	0.7000	0.9238
XGBoost	0.8875	0.6138	0.7807	0.6873	0.9401
LSTM	0.9611	0.8468	0.9211	0.8824	0.9810
LSTM–XGBoost	0.9597	0.8512	0.9035	0.8766	0.9694

La LSTM individual y el modelo híbrido alcanzaron un desempeño muy próximo entre sí (AUC 0.9810 y 0.9694, respectivamente), sin diferencias prácticas relevantes: los intervalos de confianza del 95 % de sus exactitudes se superponen, por lo que no se aprecia evidencia de una diferencia estadísticamente significativa entre ambos enfoques temporales. Este comportamiento es esperable, ya que el híbrido se apoya de forma preponderante en la probabilidad generada por la LSTM, de manera que ambos capturan esencialmente la misma dinámica temporal. La aportación específica del híbrido radica en integrar esa señal recurrente dentro de un clasificador de ensamble robusto ante el desbalance de clases y la multicolinealidad, lo que le permite superar de manera consistente al XGBoost aplicado sin dinámica temporal, aun cuando no aventaje a la LSTM en exactitud global.

3.4 Discusión general y limitaciones

En conjunto, los resultados coinciden con estudios previos que reportan el alto desempeño de los métodos de gradient boosting y aprendizaje profundo en la predicción de fenómenos climáticos extremos [8], [9], [10], [17]. La aportación de este trabajo radica en aplicar el modelo híbrido LSTM–XGBoost con datos abiertos de NASA POWER al caso de Puno, una zona con escasa cobertura instrumental y escasamente representada en la literatura internacional sobre predicción de sequías.

Entre las limitaciones se encuentran el uso de una sola fuente de datos y la cobertura restringida a Puno, por lo que los resultados deben confirmarse con más fuentes y otras regiones. Asimismo, debe considerarse que NASA POWER es un producto de reanálisis con resolución espacial de aproximadamente 0.5°, lo que puede suavizar los extremos locales de precipitación e introducir sesgos en zonas de topografía compleja y alta montaña como el Altiplano; por ello, la contrastación con estaciones terrestres del SENAMHI constituye una línea de verificación necesaria. Finalmente, el umbral de SPI < −1.0 identifica sequías moderadas, por lo que la capacidad del modelo ante sequías severas y extremas (SPI < −1.5 y SPI < −2.0) requiere una evaluación específica con muestras mayores de esos eventos.

4 CONCLUSIONES

Se desarrolló y evaluó un modelo híbrido LSTM–XGBoost para la predicción de sequías en zonas altoandinas de Puno, alcanzando una exactitud de 95.97 % y un AUC-ROC de 0.9694, con un recall de 0.9035 para la clase de sequía bajo un esquema de evaluación cronológico, cumpliendo el objetivo general planteado.

Respecto de los objetivos específicos, la caracterización de las variables confirmó la marcada estacionalidad de la precipitación altiplánica y la presencia de multicolinealidad entre las temperaturas; el análisis de importancia mostró que la componente LSTM fue, con diferencia, la variable con mayor peso relativo (89.05 %), lo que respalda el papel preponderante de la dinámica temporal en la anticipación de la sequía; y la comparación entre modelos evidenció que los enfoques con memoria temporal (la LSTM individual y el híbrido) superaron de forma consistente a la regresión logística y al XGBoost aplicado sin dinámica temporal, con un desempeño equivalente entre la LSTM y el híbrido.

Desde la perspectiva aplicada, los hallazgos evidencian la viabilidad de implementar sistemas de predicción de sequías con datos abiertos, contribuyendo a la alerta temprana y la gestión del riesgo agrícola en el Altiplano. En términos operativos, la probabilidad de sequía estimada por el modelo puede traducirse en niveles de alerta escalonados: por ejemplo, un umbral de probabilidad ≥ 0.5 activaría un aviso preventivo dirigido a agricultores y autoridades locales para adelantar la siembra o reservar agua de riego, mientras que umbrales más exigentes (≥ 0.7) podrían reservarse para alertas de acción, como la activación de reservas forrajeras o la coordinación con las plataformas de defensa civil. El elevado recall del modelo (0.9035) es especialmente valioso en este contexto, pues prioriza la detección de la mayor cantidad posible de eventos secos reales, aun a costa de algunas falsas alarmas, lo que resulta preferible en la gestión del riesgo agrícola. Como trabajo futuro se recomienda incorporar índices de vegetación, contrastar los resultados con estaciones terrestres del SENAMHI y validar el modelo en otras regiones altoandinas [18], [19].

REFERENCIAS

[1] SENAMHI, "Sembrar con Ciencia: La Quinua y el Desafío Climático del Altiplano Peruano," Servicio Nacional de Meteorología e Hidrología del Perú, Lima, Perú, 2026. [En línea]. Disponible en: https://www.gob.pe/institucion/senamhi/noticias/1368644 [Consultado: 21/07/2026].

[2] E. Benique Olivera y A. Ojeda Tito, "Impacto del Cambio Climático en la Producción y Rendimiento de Quinua (Chenopodium quinoa Willd) en la Provincia de Azángaro, Región del Altiplano-Puno, Perú," Rev. Investig. Altoandinas, vol. 26, no. 3, pp. 154-160, 2024, https://doi.org/10.18271/ria.2024.618

[3] Z. Hao, V.P. Singh y Y. Xia, "Seasonal Drought Prediction: Advances, Challenges, and Future

Prospects," Rev. Geophysics, vol. 56, no. 1, pp. 108-141, 2018, https://doi.org/10.1002/2016RG000549

[4] A. AghaKouchak, B. Pan, O. Mazdiyasni, M. Sadegh, S. Jiwa, W. Zhang, C.A. Love, S. Madadgar, S.M. Papalexiou, S.J. Davis, K. Hsu y S. Sorooshian, "Status and Prospects for Drought Forecasting: Opportunities in Artificial Intelligence and Hybrid Physical-Statistical Forecasting," Philosophical Trans. Royal Soc. A, vol. 380, no. 2238, art. 20210288, 2022, https://doi.org/10.1098/rsta.2021.0288

[5] T. Chen y C. Guestrin, "XGBoost: A Scalable Tree Boosting System," en Proc. 22nd ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '16), pp. 785-794, 2016, https://doi.org/10.1145/2939672.2939785

[6] Ö. Coşkun y H. Citakoglu, "Prediction of the Standardized Precipitation Index Based on the Long Short-Term Memory and Empirical Mode Decomposition-Extreme Learning Machine Models: The Case of Sakarya, Türkiye," Physics and Chemistry of the Earth, Parts A/B/C, vol. 131, art. 103418, 2023, https://doi.org/10.1016/j.pce.2023.103418

[7] A. Dikshit y B. Pradhan, "Interpretable and Explainable AI (XAI) Model for Spatial Drought Prediction," Science of the Total Environment, vol. 801, art. 149797, 2021, https://doi.org/10.1016/j.scitotenv.2021.149797

[8] D. Xu, Q. Zhang, Y. Ding y D. Zhang, "Application of a Hybrid ARIMA-LSTM Model Based on the SPEI for Drought Forecasting," Environmental Science and Pollution Research, vol. 29, no. 3, pp. 4128-4144, 2022, https://doi.org/10.1007/s11356-021-15325-z

[9] M. Li, Y. Yao, Z. Feng y M. Ou, "Hydrological Drought Prediction and Its Influencing Features Analysis Based on a Machine Learning Model," Natural Hazards and Earth System Sciences, vol. 25, no. 11, pp. 4299-4316, 2025, https://doi.org/10.5194/nhess-25-4299-2025

[10] V. Gyaneshwar, A. Mishra, U. Chadha, P.M.D. Raj Vincent, V. Rajinikanth, G. Pattukandan Ganapathy y K. Srinivasan, "A Contemporary Review on Deep Learning Models for Drought Prediction," Sustainability, vol. 15, no. 7, art. 6160, 2023, https://doi.org/10.3390/su15076160

[11] NASA, "NASA Prediction of Worldwide Energy Resources (POWER)," NASA Langley Research Center, Hampton, Va., 2025. [En línea]. Disponible en: https://power.larc.nasa.gov/

[12] M. Svoboda, M. Hayes y D. Wood, "Standardized Precipitation Index User Guide (WMO-No. 1090)," World Meteorological Organization, Ginebra, Suiza, 2012.

[13] T.B. McKee, N.J. Doesken y J. Kleist, "The Relationship of Drought Frequency and Duration to Time Scales," en Proc. Eighth Conf. Applied Climatology, pp. 179-184, 1993.

[14] A. Anshuka, F.F. van Ogtrop y R. Willem Vervoort, "Drought Forecasting Through Statistical Models Using Standardised Precipitation Index: A Systematic Review and Meta-Regression Analysis," Natural Hazards, vol. 97, no. 2, pp. 955-977, 2019, https://doi.org/10.1007/s11069-019-03665-6

[15] Y. Mohanty, "Predicting Droughts: A Comparative Study of ARIMAX, LSTM, XGBoost, and Random Forest Models," en Proc. IEEE Int'l Conf. Computing, Communication and Networking Technologies (ICCCNT), 2025, https://doi.org/10.1109/ICCAI66501.2025.00121

[16] T. Tuğrul, S. Oruç, J.L. Hall, A.U.G. Şenocak y M.A. Hınıs, "Hybrid Wavelet-Machine Learning Models for Regional Drought Forecasting in Norway," Scientific Reports, vol. 15, art. 38573, 2025, https://doi.org/10.1038/s41598-025-22416-1

[17] R. Ergüven, A.B. Aydin y D. Avci, "Drought Forecasting Using Standard Precipitation Index and Artificial Intelligence Models in the Mediterranean Region of Türkiye," Applied Sciences, vol. 15, no. 22, art. 12172, 2025, https://doi.org/10.3390/app152212172

[18] M.A. Alawsi, S.L. Zubaidi, N.S.S. Al-Bdairi, N. Al-Ansari y K. Hashim, "Drought Forecasting: A Review and Assessment of the Hybrid Techniques and Data Pre-Processing," Hydrology, vol. 9, no. 7, art. 115, 2022, https://doi.org/10.3390/hydrology9070115

[19] C.J. Talsma, K.C. Solander, M.K. Mudunuru, B. Crawford y M.R. Powell, "Frost Prediction Using Machine Learning and Deep Neural Network Models," Frontiers in Artificial Intelligence, vol. 5, art. 963781, 2023, https://doi.org/10.3389/frai.2022.963781