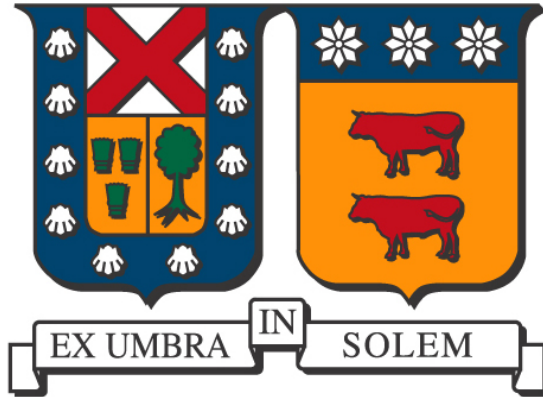


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE ELECTRÓNICA
SANTIAGO - CHILE



**“MULTI-LLM-AGENT PARA LA AGRICULTURA
INTELIGENTE”**

TABATA ANDREA AHUMADA AVENDAÑO

**MEMORIA DE TITULACIÓN PARA OPTAR AL TÍTULO DE INGENIERA CIVIL
TELEMÁTICA**

PROFESOR GUÍA:

NICOLÁS TORRES RUDLOFF

PROFESOR CORREFERENTE:

NICOLÁS GALVEZ RAMIREZ

JULIO - 2026



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título ☐ Tesis de Postgrado

Título del trabajo: Multi-llm-agent para la agricultura inteligente

Nombre del candidato(a): Tabata Andrea Ahumada Avendaño

Carrera / Grado: Ingeniería Civil Telemática

Campus: San Joaquín Departamento: Electrónica

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Nicolás Torres Rudloff, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por (**marcar una opción**):

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

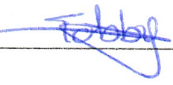
Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 03/09/2026

Firma: 

Estudiante o Candidato(a):

Fecha: 03-09-2026

Firma: 

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

Agradecimientos

A mi madre, por su apoyo incondicional, su paciencia y por escucharme durante tantos años hablar del estrés, las exigencias y los desafíos de la universidad. Su compañía y comprensión fueron fundamentales en cada etapa de este camino.

A mi padre, por su confianza constante y por esos cafés que me acompañaron en tantas noches de estudio. Su apoyo, incluso en los gestos más simples, fue una motivación importante para seguir avanzando.

A mi hermana, por estar presente en los momentos más difíciles y por entregarme ánimo, apoyo y cariño de manera permanente.

A mis amistades, por acompañarme durante esta etapa, por ayudarme a sobrellevar los momentos de mayor presión y, al mismo tiempo, por aportar risas, conversaciones y buenos recuerdos que hicieron este proceso mucho más llevadero.

A mis tres bebés, mis dos perros y mi gata, por su compañía incondicional y por hacer más cálidas las largas jornadas de estudio.

Finalmente, agradezco a los profesores y profesoras del Departamento de Electrónica y de la carrera de Ingeniería Civil Telemática por los conocimientos entregados, su disposición y la cercanía con la que contribuyeron a mi formación académica y profesional.

A mi madre, mi padre y mi hermana, por acompañarme con cariño, confianza y apoyo durante todos estos años. Este trabajo representa también una parte de su esfuerzo, paciencia y presencia constante en cada desafío que enfrenté.

Y a mí misma, por no rendirme incluso en los momentos difíciles, por seguir avanzando con constancia y por llegar hasta el final de esta etapa.

Resumen

La agricultura enfrenta una creciente exposición a eventos térmicos extremos, tales como heladas y olas de calor, cuya prevención requiere transformar información meteorológica dispersa en alertas comprensibles y oportunas. Esta memoria presenta el diseño, implementación y evaluación de un sistema de alertas agroclimáticas basado en inteligencia artificial y una arquitectura *Multi-LLM-Agent*, orientado a apoyar la toma de decisiones agrícolas en las 52 comunas de la Región Metropolitana de Chile.

La solución integra datos del Atlas Agroclimático de Chile, la red INIA/AGROMET, Open-Meteo y MeteoChile mediante una política de fusión jerárquica que prioriza observaciones directas de estación. La serie consolidada resultante contiene 368.082 pares únicos comuna–fecha. Sobre esta base se implementó un ensamble heterogéneo de siete clasificadores para la detección de heladas y olas de calor, complementado por modelos de regresión para estimar temperaturas mínimas y máximas. El sistema utiliza partición temporal, ponderación de clases, calibración de umbrales mediante el índice de éxito crítico (CSI), votación mayoritaria y una estrategia híbrida para los horizontes de predicción D+1 a D+5.

La arquitectura Multi-LLM-Agent separa el proceso de generación de alertas en tres agentes especializados: Monitor, que activa alertas mediante reglas deterministas; Razonador, que interpreta el riesgo a partir de predicciones, umbrales agrícolas e información histórica; y Comunicador, que genera mensajes orientados a WhatsApp. Se evaluaron los proveedores Claude, Gemini y Llama 3 mediante métricas de latencia, robustez de formato y una evaluación cualitativa retrospectiva de consistencia interpretativa, pertinencia agronómica, claridad y accionabilidad.

En el corte operativo analizado, el sistema evaluó 3.206 predicciones D+1 y alcanzó, para heladas, un CSI de 0,395, una probabilidad de detección de 0,773 y un F1-score de 0,566. La comparación descriptiva con MeteoChile, sobre 3.034 pares comuna–fecha comunes, mostró una mayor capacidad de detección para el sistema propuesto, a costa de una mayor cantidad de falsas alarmas. La arquitectura Multi-LLM-Agent procesó 199 alertas y logró respuestas JSON finales válidas

en el 100 % de los casos tras aplicar mecanismos de validación y recuperación. Los resultados entregan evidencia inicial de la factibilidad técnica y operacional de integrar predicción meteorológica, interpretación agroclimática y comunicación automatizada. No obstante, se requiere ampliar la validación estacional, territorial y con usuarios reales antes de considerar un despliegue productivo a mayor escala.

Palabras clave: agricultura inteligente; alerta temprana agroclimática; predicción de heladas; eventos térmicos extremos; modelos de lenguaje de gran escala; sistemas multiagente.

Abstract

Agriculture is increasingly exposed to extreme thermal events, such as frost and heat waves, whose prevention requires transforming fragmented meteorological information into timely and understandable alerts. This thesis presents the design, implementation, and evaluation of an agroclimatic alert system based on artificial intelligence and a *Multi-LLM-Agent* architecture, aimed at supporting agricultural decision-making across the 52 municipalities of the Metropolitan Region of Chile.

The solution integrates data from the Chilean Agroclimatic Atlas, the INIA/AGROMET network, Open-Meteo, and MeteoChile through a hierarchical data-fusion policy that prioritizes direct weather-station observations. The resulting consolidated series contains 368,082 unique municipality–date pairs. On this basis, a heterogeneous ensemble of seven classifiers was implemented to detect frost and heat-wave events, complemented by regression models for minimum and maximum temperature estimation. The system uses temporal data splitting, class weighting, Critical Success Index (CSI)-based threshold calibration, majority voting, and a hybrid strategy for D+1 to D+5 prediction horizons.

The Multi-LLM-Agent architecture separates alert generation into three specialized agents: a Monitor agent, which activates alerts through deterministic rules; a Reasoner agent, which interprets risk using predictions, agricultural thresholds, and historical information; and a Communicator agent, which produces WhatsApp-oriented messages. Claude, Gemini, and Llama 3 were evaluated using latency, output-format robustness, and a retrospective qualitative assessment of interpretive consistency, agronomic relevance, clarity, and actionability.

Within the analyzed operational period, the system evaluated 3,206 D+1 predictions and achieved, for frost detection, a CSI of 0.395, a probability of detection of 0.773, and an F1-score of 0.566. A descriptive comparison with MeteoChile, conducted on 3,034 common municipality–date pairs, showed greater detection capability for the proposed system at the cost of a higher number of false alarms. The Multi-LLM-Agent architecture processed 199 alerts and achieved valid final

JSON outputs in 100 % of cases after applying validation and recovery mechanisms. The results provide initial technical and operational evidence of the feasibility of integrating meteorological prediction, agroclimatic interpretation, and automated communication. However, broader seasonal, territorial, and real-user validation is required before considering large-scale production deployment.

Keywords: smart agriculture; agroclimatic early warning; frost prediction; extreme thermal events; large language models; multi-agent systems.

Glosario

Agente Comunicador Componente encargado de transformar el análisis estructurado en mensajes claros y accionables orientados al usuario final.

Agente Monitor Componente que activa alertas mediante reglas deterministas y prepara el contexto meteorológico, predictivo y agrícola requerido por las etapas posteriores.

Agente Razonador Componente responsable de interpretar predicciones, umbrales térmicos y contexto agrícola para evaluar el riesgo asociado a una alerta.

alerta mixta Alerta activada cuando se detectan simultáneamente distintos tipos de eventos térmicos de riesgo en el área de cobertura del sistema.

alerta temprana Notificación emitida con anticipación para advertir la ocurrencia probable de un evento de riesgo.

corrección adaptativa de sesgo Ajuste aplicado a una predicción de temperatura utilizando el error medio reciente observado para una comuna o para el sistema completo.

downscaling espacial Proceso de derivar información climática o meteorológica local a partir de modelos, pronósticos o análisis de mayor escala espacial.

ensamble heterogéneo Combinación de modelos predictivos de distintas familias algorítmicas para generar una decisión conjunta más robusta.

estrés térmico Respuesta fisiológica adversa de un cultivo cuando la temperatura supera o desciende de sus rangos óptimos.

evento no detectado Evento térmico observado que no fue identificado por el sistema predictivo.

evento raro Evento de baja frecuencia y alto impacto, cuya detección requiere métricas específicas debido al desbalance de clases.

evento térmico extremo Condición climática con temperaturas fuera del rango habitual que puede generar impactos significativos en sistemas agrícolas.

fallback Mecanismo de recuperación aplicado cuando una respuesta generada no cumple el formato o las validaciones requeridas por el sistema.

falsa alarma Predicción positiva de un evento térmico que no fue confirmada posteriormente por la observación disponible.

fenología Estudio de las etapas de desarrollo de los cultivos y de su relación con las condiciones ambientales.

fusión jerárquica de fuentes Procedimiento de integración de datos meteorológicos que conserva, para cada par comuna–fecha, el registro disponible de mayor prioridad según una regla predefinida.

helada Evento meteorológico caracterizado por temperaturas iguales o inferiores a 0°C, capaz de provocar daño fisiológico en tejidos vegetales.

helada radiativa Tipo de helada asociado a la pérdida intensa, durante la noche, del calor acumulado en el suelo durante el día, lo que enfría las capas bajas de la atmósfera cercanas a la superficie.

horizonte de predicción Cantidad de días entre la ejecución del sistema y la fecha para la cual se genera una predicción. D+1 representa el día siguiente y D+2 a D+5 los días posteriores.

impacto agrícola Efecto productivo, fisiológico o económico generado por una condición climática sobre un cultivo.

latencia Tiempo transcurrido entre el envío de una solicitud a un servicio o modelo y la recepción de una respuesta procesable.

modelo predictivo Modelo computacional capaz de estimar la ocurrencia futura de un evento o el valor futuro de una variable a partir de información histórica y reciente.

Multi-LLM-Agent Arquitectura compuesta por múltiples agentes basados en modelos de lenguaje que colaboran para realizar tareas especializadas dentro del sistema.

ola de calor Período prolongado de temperaturas anormalmente altas respecto de los valores climatológicos normales, capaz de generar estrés en los cultivos.

pre-gathering Preparación previa y estructurada del contexto entregado a un modelo de lenguaje mediante consultas a fuentes internas antes de realizar la invocación al modelo.

reanálisis meteorológico Producto climático que combina observaciones y modelos numéricos para reconstruir condiciones atmosféricas históricas de manera espacialmente continua.

serie meteorológica consolidada Serie temporal resultante de integrar observaciones de distintas fuentes mediante una política de prioridad y trazabilidad.

trazabilidad Capacidad de reconstruir el origen, procesamiento, modelo utilizado, predicción generada y alerta emitida dentro del sistema.

umbral operacional de alerta de helada Criterio preventivo utilizado por el sistema para activar una alerta cuando la temperatura mínima prevista es inferior a 3°C. No equivale a un umbral universal de daño agrícola.

umbral operacional de alerta de ola de calor Criterio preventivo utilizado por el sistema para activar una alerta cuando la temperatura máxima prevista alcanza o supera 30°C. No equivale a un umbral universal de daño agrícola.

umbral térmico Valor de temperatura a partir del cual pueden producirse efectos adversos sobre un cultivo o una etapa fenológica.

votación mayoritaria Regla de decisión en la que la salida final se activa cuando una cantidad mínima de modelos individuales emite un voto positivo.

Siglas

AdaBoost Adaptive Boosting.

AGROMET Red Agrometeorológica Nacional.

ANN red neuronal artificial.

API interfaz de programación de aplicaciones.

ARClim Atlas de Riesgos Climáticos de Chile.

CNN red neuronal convolucional.

CSA Climate-Smart Agriculture.

CSI Critical Success Index.

DL aprendizaje profundo.

DNN redes neuronales profundas.

EMA Estación Meteorológica Automática.

F1 F1-Score.

FAO Food and Agriculture Organization of the United Nations.

FIA Fundación para la Innovación Agraria.

FN False Negative.

FP False Positive.

IA inteligencia artificial.

INIA Instituto de Investigaciones Agropecuarias.

IoT Internet de las Cosas.

JSON JavaScript Object Notation.

KNN K-Nearest Neighbors.

LLM modelos de lenguaje de gran escala.

MAE Mean Absolute Error.

MAS sistema multiagente.

ML aprendizaje automático.

MLP Multilayer Perceptron.

NLP procesamiento de lenguaje natural.

POD Probability of Detection.

PR Precision-Recall.

RAG generación aumentada por recuperación.

RF Random Forest.

RMSE Root Mean Square Error.

ROC Receiver Operating Characteristic.

SVM Support Vector Machine.

TN True Negative.

TP True Positive.

WMO World Meteorological Organization.

XGBoost Extreme Gradient Boosting.

Tabla de Contenido

1	Introducción	1
1.1	Problemática	7
1.2	Objetivos	10
1.2.1	Objetivo Principal	10
1.2.2	Objetivos Específicos	10
2	Estado del Arte	11
2.1	Predicción de eventos de heladas mediante aprendizaje automático . . .	11
2.2	Modelos avanzados de predicción meteorológica y climática basados en inteligencia artificial	17
2.3	Impacto de las temperaturas extremas y umbrales térmicos en los cultivos	20
2.4	Servicios agroclimáticos y sistemas de alerta temprana	21
2.5	Sistemas conversacionales y recuperación de conocimiento para ases- oría agrícola	23
2.6	Síntesis y brecha de la literatura	26
3	Propuesta de Solución	29
3.1	Visión general de la solución	29
3.2	Justificación del Enfoque Multi-LLM-Agent	30
3.3	Enfoque conceptual de la solución	32
3.4	Componentes de la solución	33
3.4.1	Capa 1 — Adquisición e integración de datos	33
3.4.2	Capa 2 — Predicción e interpretación	33
3.4.3	Capa 3 — Pipeline multiagente y comunicación	35
3.5	Pipeline de operación y validación	37
3.6	Ejemplo de salida del sistema	38
3.7	Ámbito de aplicación y alcance	38
3.8	Síntesis de la propuesta	39

TABLA DE CONTENIDO

4	Requisitos del Sistema	41
4.1	Requisitos Funcionales	41
4.2	Requisitos No Funcionales	43
4.3	Interfaces del Sistema	44
4.4	Trazabilidad de Requisitos	45
4.5	Supuestos del Sistema	45
5	Diseño de Arquitectura del Sistema	47
5.1	Visión General de la Arquitectura	47
5.2	Flujo Operacional del Sistema	48
5.3	Componentes del Sistema	50
5.3.1	Módulo de Adquisición de Datos	50
5.3.2	Módulo de Preparación de Datos	50
5.3.3	Módulo de Predicción	51
5.3.4	Módulo de Interpretación Agroclimática	51
5.3.5	Módulo de Generación de Alertas	53
5.3.6	Módulo de Comunicación	53
5.4	Fuentes de Datos Meteorológicos	54
5.5	Modelo Lógico de Datos	55
5.6	Infraestructura Tecnológica	56
6	Metodología	59
6.1	Enfoque Metodológico y Diseño Experimental	60
6.2	Área de Estudio y Alcance Territorial	61
6.3	Fuentes de Datos	63
6.3.1	Atlas Agroclimático de Chile	63
6.3.2	Red AGROMET del Instituto de Investigaciones Agropecuarias	64
6.3.3	Open-Meteo	65
6.3.4	Dirección Meteorológica de Chile	65
6.4	Gestión y Trazabilidad de la Información	66

TABLA DE CONTENIDO

6.5	Preparación de Datos y Construcción de Características	67
6.5.1	Fusión jerárquica de fuentes	67
6.5.2	Tratamiento de valores faltantes	68
6.5.3	Variables de entrada	69
6.6	Definición de Eventos Térmicos	70
6.7	Entrenamiento del Sistema Predictivo Propuesto	71
6.7.1	Partición temporal de los datos	71
6.7.2	Normalización de variables	72
6.7.3	Ensamble heterogéneo de clasificadores	73
6.7.4	Manejo del desbalance de clases	73
6.7.5	Calibración de umbrales mediante CSI	74
6.7.6	Votación mayoritaria	74
6.7.7	Regresión de temperatura	75
6.8	Corrección Adaptativa de Sesgo	75
6.9	Predicción Multi-Horizonte	76
6.10	Protocolo de la Arquitectura Multi-LLM-Agent	77
6.10.1	Agente Monitor	78
6.10.2	Agente Razonador	79
6.10.3	Agente Comunicador	80
6.11	Evaluación Comparativa de Modelos de Lenguaje	81
6.12	Operación Automatizada y Reentrenamiento	84
6.13	Métricas de Evaluación del Sistema Predictivo	86
6.13.1	Critical Success Index	86
6.13.2	Probabilidad de Detección	87
6.13.3	Raíz del Error Cuadrático Medio	87
6.14	Síntesis Metodológica	87
7	Resultados	89
7.1	Cobertura, integración y calidad de las series meteorológicas	89

TABLA DE CONTENIDO

7.2	Desempeño del sistema predictivo para el horizonte D+1	92
7.2.1	Variación del desempeño entre comunas	93
7.3	Desempeño de regresión de temperatura	95
7.4	Evaluación multi-horizonte D+1 a D+5	97
7.5	Comparación operacional con fuentes externas	99
7.6	Evaluación de la arquitectura Multi-LLM-Agent	101
7.6.1	Evaluación cualitativa retrospectiva	102
7.6.2	Evidencia de entrega de alertas mediante WhatsApp	105
7.7	Casos representativos del pipeline	106
7.8	Síntesis de resultados y limitaciones observadas	108
8	Conclusiones	111
8.1	Conclusiones generales	111
8.2	Cumplimiento de objetivos	114
8.3	Limitaciones del estudio	115
8.4	Trabajo futuro	116

Referencias

Anexos

Anexo A Infraestructura Tecnológica

A.1	Arquitectura física	
A.2	Servicios implementados	
A.3	Configuración de despliegue	

Anexo B Modelo Físico de Datos

B.1	Diagrama entidad-relación	
B.2	Entidades principales	

Anexo C Automatización del Pipeline

TABLA DE CONTENIDO

C.1	Flujo de automatización
C.2	Etapas del pipeline
C.3	Programación de tareas

Anexo D Integración con WhatsApp

D.1	Arquitectura de comunicación
D.2	Proceso de envío
D.3	Formato de mensajes

Anexo E Configuración Experimental de los Modelos

E.1	Variables empleadas
E.2	Modelos y configuración de entrenamiento
E.3	Entorno de ejecución y reproducibilidad

Anexo F Catálogo de Variables Meteorológicas y Trazabilidad

F.1	Variables disponibles por fuente
F.2	Política de fusión

Anexo G Protocolo Reproducible de Evaluación Multi-LLM-Agent

G.1	Contexto entregado al Agente Razonador
G.2	Esquema de salida esperado
G.3	Rúbrica y procedimiento de evaluación cualitativa retrospectiva

Anexo H Resultados Complementarios

H.1	Métricas de helada por comuna
H.2	Composición de la serie consolidada
H.3	Cumplimiento de formato JSON por proveedor
H.4	Evaluación cualitativa retrospectiva
H.5	Ejemplos anonimizados de alertas generadas

Índice de figuras

1.1	Proyección de cambios en el rendimiento de cultivos (maíz, trigo y arroz) en función del aumento de la temperatura media, considerando escenarios con y sin adaptación. Fuente: (Tran et al., 2025).	1
1.2	Uso habitual de plataformas digitales en Chile. Fuente: Elaboración propia a partir de Subsecretaría de Telecomunicaciones y Cadem (2024)	6
1.3	Efecto del uso de servicios de mensajería móvil en la adopción de recomendaciones agrícolas por parte de los agricultores. Fuente: (Sharma et al., 2021).	9
2.1	Distribución de estaciones de la red AGROMET utilizadas para modelación agroclimática en Chile central. Adaptado de Fuentes et al. (2018).	13
2.2	Arquitectura de red neuronal utilizada para la predicción de temperatura mínima. Adaptado de Fuentes et al. (2018).	13
2.3	Desempeño espacial del modelo CNN para predicción de temperatura mínima. Adaptado de Bhakare et al. (2025).	17
2.4	Arquitectura del sistema AgAsk para recuperación de conocimiento agrícola mediante modelos de lenguaje. Adaptado de Koopman et al. (2024).	25
3.1	Arquitectura general del sistema propuesto.	30
3.2	Arquitectura multiagente del sistema.	36
3.3	Pipeline de operación del sistema.	37
3.4	Ejemplo de alerta generada por el sistema.	38
3.5	Cobertura espacial del sistema.	39
5.1	Arquitectura general del sistema de alertas agroclimáticas.	48
5.2	Flujo operacional del sistema.	49
5.3	Arquitectura conceptual del módulo de interpretación agroclimática. . .	52
5.4	Modelo lógico simplificado del sistema.	56
6.1	Esquema general de la metodología propuesta.	59

6.2	Diseño experimental de dos niveles utilizado para evaluar el sistema predictivo y la arquitectura Multi-LLM-Agent.	61
6.3	Asignación espacial entre comunas y estaciones meteorológicas mediante proximidad geográfica. Fuente: elaboración propia a partir del sistema desarrollado.	62
6.4	Proceso de fusión jerárquica de fuentes meteorológicas para la construcción de series consolidadas por comuna.	68
6.5	Partición cronológica utilizada para el entrenamiento, calibración y evaluación del sistema predictivo.	72
6.6	Estrategia híbrida de predicción para los horizontes D+1 a D+5.	77
6.7	Protocolo de operación de la arquitectura Multi-LLM-Agent para la generación de alertas agroclimáticas.	78
6.8	Esquema de integración y evaluación de proveedores de modelos de lenguaje dentro de la arquitectura Multi-LLM-Agent.	82
7.1	Cobertura temporal de las fuentes meteorológicas.	90
7.2	Composición de la serie meteorológica consolidada según fuente prioritaria.	91
7.3	Métricas CSI, POD, precisión y F1 para helada en D+1.	93
7.4	CSI de helada por comuna para las veinte comunas con mayor número de registros evaluables.	95
7.5	Comparación entre temperaturas observadas y estimadas por el sistema predictivo para el horizonte D+1. Cada punto representa una predicción evaluada y la línea diagonal roja indica el caso de predicción perfecta.	96
7.6	Variación de CSI, RMSE y cantidad de registros por horizonte D+1 a D+5.	98
7.7	Comparación de CSI, POD, precisión y RMSE de temperatura mínima entre el sistema propuesto y MeteoChile.	100
7.8	Latencia media y distribución de tiempos de respuesta observados para los proveedores LLM utilizados en la arquitectura Multi-LLM-Agent.	102

7.9	Puntajes promedio de consistencia interpretativa, pertinencia agronómica, claridad comunicacional y accionabilidad por proveedor LLM. Escala de 1 a 5.	104
7.10	Ejemplos de alertas agroclimáticas e interacción básica del sistema recibidos mediante WhatsApp durante pruebas de operación.	106
A.1	Arquitectura física del sistema.	
B.1	Modelo físico de base de datos.	
C.1	Pipeline automatizado del sistema.	
D.1	Arquitectura de integración con WhatsApp.	

Índice de tablas

1	Temperaturas críticas de daño por frío (°C) para distintos cultivos en diferentes etapas fenológicas. Fuente Food and Agriculture Organization of the United Nations (2005)	3
2	Temperaturas críticas asociadas a estrés térmico en etapas reproductivas de distintos cultivos. Fuente: Tran et al. (2025)	4
3	Desempeño de modelos para la predicción de temperatura mínima (heladas) en Nuevo México, con diferentes horizontes. Adaptado de Talsma et al. (2023).	15
4	Temperaturas críticas de daño por heladas en cultivos. Adaptado de Food and Agriculture Organization of the United Nations (2005).	20
5	Resumen del estado del arte según componentes abordados.	28
6	Comparación conceptual de enfoques para la generación de alertas agroclimáticas	32
7	Responsabilidades de los agentes que componen la arquitectura Multi-LLM-Agent	36
8	Trazabilidad entre objetivos específicos y requisitos funcionales	45
9	Fuentes de datos meteorológicos utilizadas por el sistema.	55

ÍNDICE DE TABLAS

10	Componentes tecnológicos utilizados por la plataforma.	57
11	Fuentes de datos utilizadas por el sistema y su rol metodológico.	63
12	Variables de entrada utilizadas por el sistema predictivo propuesto.	69
13	Dimensiones de evaluación de los proveedores de modelos de lenguaje.	83
14	Secuencia general del pipeline diario automatizado.	85
15	Matriz de confusión utilizada para evaluar la detección de eventos térmicos.	86
16	Cobertura temporal y territorial de las fuentes meteorológicas utilizadas.	89
17	Composición de la serie meteorológica consolidada después de la fusión jerárquica.	90
18	Porcentaje de valores faltantes por fuente y variable meteorológica.	92
19	Desempeño global del sistema predictivo para D+1.	92
20	Comunas con mayor y menor desempeño CSI para helada en D+1.	94
21	Métricas de regresión para temperatura en D+1.	96
22	Resultados disponibles por horizonte de predicción.	97
23	Comparación operacional entre el sistema predictivo propuesto y Meteochile para pares comunes.	99
24	Resultados operacionales de los proveedores LLM.	101
25	Evaluación cualitativa retrospectiva de mensajes generados por los proveedores LLM. Escala de 1 a 5.	103
26	Casos representativos de predicción y validación.	107
27	Servicios utilizados por la plataforma.	
28	Variables utilizadas por el sistema predictivo.	
29	Modelos incluidos en el ensamble de clasificación.	
30	Entorno de ejecución utilizado para generar los resultados experimentales.	
31	Variables meteorológicas disponibles por fuente.	
32	Rúbrica cualitativa utilizada para evaluar mensajes generados por la arquitectura Multi-LLM-Agent.	
33	Métricas de helada por comuna para D+1.	

ÍNDICE DE TABLAS

34	Distribución de la serie consolidada según fuente ganadora.
35	Robustez de formato de los proveedores de modelos de lenguaje.
36	Resultados de evaluación cualitativa retrospectiva asistida por IA.

1. Introducción

La agricultura enfrenta actualmente un escenario de creciente variabilidad climática, caracterizado por el aumento en la frecuencia e intensidad de eventos térmicos extremos. Diversos informes internacionales han advertido que tanto las heladas como las olas de calor representan riesgos significativos para la productividad agrícola y la seguridad alimentaria (World Bank, 2023; World Meteorological Organization, 2025). Estos eventos afectan directamente el rendimiento de cultivos sensibles y generan pérdidas económicas relevantes, particularmente en regiones donde la agricultura constituye un pilar fundamental del desarrollo productivo.

En este escenario, se ha estimado que el aumento de temperaturas extremas puede provocar reducciones de rendimiento superiores al 10 % en cultivos clave a nivel global, con impactos económicos que se traducen en pérdidas de miles de millones de dólares en el sector agrícola (Kumar et al., 2025; Tran et al., 2025). Como se observa en la Figura 1.1, el aumento de la temperatura media tiene un impacto directo en el rendimiento de cultivos clave, evidenciando reducciones significativas en escenarios de calentamiento sin adaptación.

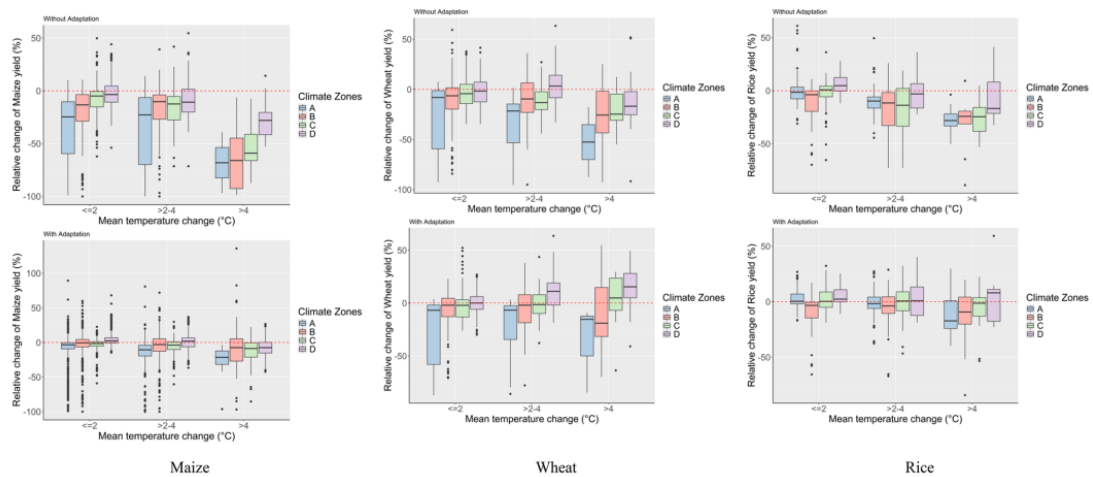


Figura 1.1: Proyección de cambios en el rendimiento de cultivos (maíz, trigo y arroz) en función del aumento de la temperatura media, considerando escenarios con y sin adaptación. Fuente: (Tran et al., 2025).

CAPÍTULO 1 : INTRODUCCIÓN

En el caso de Chile, instituciones como la Fundación para la Innovación Agraria (FIA) han evidenciado que los sistemas actuales de control y manejo de heladas presentan limitaciones en su efectividad, generando pérdidas significativas en el sector agrícola (Fundación para la Innovación Agraria, 2023b). Asimismo, el fortalecimiento de redes de monitoreo agroclimático impulsado por el Instituto de Investigaciones Agropecuarias (INIA) ha incrementado la disponibilidad de datos meteorológicos en tiempo real (Instituto de Investigaciones Agropecuarias, 2025).

No obstante, la disponibilidad de datos provenientes de estaciones meteorológicas, servicios de pronóstico e imágenes satelitales no garantiza por sí sola su integración ni su transformación en información útil para la toma de decisiones en campo. En consecuencia, se configura una brecha tecnológica relevante: aunque los agricultores cuentan con acceso a información climática, no siempre disponen de herramientas accesibles e integradas que permitan anticipar con precisión un evento térmico extremo y evaluar su impacto productivo.

En esta línea, la literatura en Climate-Smart Agriculture (CSA) destaca la necesidad de sistemas digitales que faciliten la gestión del riesgo climático y fortalezcan la resiliencia productiva (Fundación para la Innovación Agraria, 2023a; World Bank, 2023). Por tanto, surge la necesidad de desarrollar soluciones capaces de integrar información climática heterogénea y transformarla en alertas tempranas comprensibles y accionables.

Tradicionalmente, gran parte de los esfuerzos tecnológicos se han concentrado en la predicción de heladas, dada su alta capacidad de generar daños inmediatos en etapas fenológicas sensibles. De acuerdo con la Food and Agriculture Organization of the United Nations (FAO), los umbrales térmicos de daño por frío varían significativamente según el cultivo y su estado de desarrollo, pudiendo observarse afectaciones incluso a temperaturas cercanas a -1 °C en etapas críticas (Food and Agriculture Organization of the United Nations, 2005).

Como se presenta en la Tabla 1, la sensibilidad al frío varía considerablemente entre

CAPÍTULO 1 : INTRODUCCIÓN

cultivos y etapas fenológicas. Mientras algunos cultivos toleran temperaturas bajo -8 °C durante la germinación, otros como tomate o arroz pueden sufrir daños a temperaturas cercanas a 0 °C. Esto evidencia la necesidad de considerar umbrales específicos en la evaluación del riesgo agrícola.

***Tabla 1:** Temperaturas críticas de daño por frío (°C) para distintos cultivos en diferentes etapas fenológicas. Fuente Food and Agriculture Organization of the United Nations (2005)*

Cultivo	Germinación	Floración	Fructificación
Trigo (Spring wheat)	-9 a -10	-1 a -2	-2 a -4
Avena (Oats)	-8 a -9	-1 a -2	-2 a -4
Cebada (Barley)	-7 a -8	-1 a -2	-2 a -4
Maíz (Corn)	-2 a -3	-1 a -2	-2 a -3
Papa (Potato)	-2 a -3	-1 a -2	-1 a -2
Arroz (Rice)	-0.5 a -1	-0.5 a -1	-0.5 a -1
Tomate (Tomato)	0 a -1	0 a -1	0 a -1
Algodón (Cotton)	-1 a -2	-1 a -2	-2 a -3
Repollo (Cabbage)	-5 a -7	-2 a -3	-6 a -9

Sin embargo, el aumento sostenido de eventos de calor extremo ha incorporado un nuevo factor de riesgo productivo. Estudios recientes muestran que temperaturas supe-

CAPÍTULO 1 : INTRODUCCIÓN

rios a 30 °C o 35 °C durante etapas reproductivas pueden reducir significativamente el rendimiento en cultivos como trigo, maíz y soja, con pérdidas que pueden superar el 15 % en escenarios de estrés térmico severo (Kumar et al., 2025; Tran et al., 2025).

Como se observa en la Tabla 2, el impacto del estrés térmico no depende únicamente del valor absoluto de temperatura, sino también de la etapa fenológica del cultivo. En particular, fases como la formación de polen, floración y fecundación presentan una alta sensibilidad térmica, donde temperaturas superiores a 32 °C o 35 °C pueden provocar esterilidad del polen, reducción en la formación de granos y disminución del rendimiento final.

Este fenómeno no solo impacta la producción, sino que también incrementa los costos operacionales y reduce la rentabilidad del sistema agrícola, reforzando la necesidad de sistemas que permitan anticipar este tipo de eventos y su impacto asociado.

Tabla 2: Temperaturas críticas asociadas a estrés térmico en etapas reproductivas de distintos cultivos. Fuente: Tran et al. (2025)

Cultivo	Etapla crítica	Temperatura crítica (°C)	Impacto
Arroz (Rice)	Floración / Polinización	33 – 38	Esterilidad de polen
Trigo (Wheat)	Antesis	30 – 38	Reducción de granos
Maíz (Maize)	Polinización	>35	Esterilidad masculina y femenina
Sorgo (Sorghum)	Floración	33 – 40	Reducción de rendimiento
Cebada (Barley)	Desarrollo floral	~30	Inviabilidad de polen
Tomate (Tomato)	Floración	29 – 32	Disminución de cuaja de frutos
Arveja (Peas)	Floración	28 – 33	Reducción de semillas

En consecuencia, esta sensibilidad en etapas reproductivas evidencia que el impacto del calor extremo no depende únicamente de la magnitud de la temperatura, sino también del momento en que ocurre, reforzando la necesidad de sistemas capaces de contextualizar el riesgo climático en función del desarrollo del cultivo.

CAPÍTULO 1 : INTRODUCCIÓN

A partir de lo anterior, el presente trabajo propone el desarrollo de un sistema basado en inteligencia artificial (IA) y una arquitectura de sistema multiagente (MAS), orientado a anticipar eventos térmicos extremos tanto de bajas como de altas temperaturas. Los sistemas multiagente han sido ampliamente utilizados en aplicaciones complejas donde se requiere la coordinación de múltiples procesos autónomos, permitiendo la integración modular de distintas fuentes de datos y tareas especializadas. En el ámbito agrícola, estos enfoques han demostrado ser efectivos para la gestión distribuida de información y la toma de decisiones basada en múltiples criterios.

Adicionalmente, los avances recientes en los modelos de lenguaje de gran escala (LLM) han permitido desarrollar sistemas capaces de interpretar información especializada y comunicarla de manera comprensible para el usuario final, especialmente cuando se integran con mecanismos de recuperación de conocimiento (RAG) en dominios específicos (Koopman et al., 2024; Xia et al., 2026). En consecuencia, el concepto de sistemas basados en múltiples agentes impulsados por LLM (Multi-LLM-Agent) surge como una alternativa para combinar capacidades de procesamiento, razonamiento y comunicación en una arquitectura unificada.

A diferencia de enfoques tradicionales centrados exclusivamente en la predicción meteorológica, esta propuesta integra de manera conjunta tres componentes fundamentales: la predicción de variables climáticas, la interpretación de dichas variables mediante umbrales térmicos asociados al daño agrícola y la comunicación de resultados a través de interfaces accesibles para el usuario final.

La solución contempla la integración de datos provenientes de estaciones meteorológicas, interfaz de programación de aplicaciones (API) de pronóstico y otras fuentes climáticas relevantes, su procesamiento mediante modelos de aprendizaje automático (ML) entrenados con datos históricos y la generación de alertas automatizadas a través de una plataforma conversacional basada en WhatsApp.

La elección de este canal se fundamenta en su alta adopción y uso cotidiano en el contexto chileno, donde se posiciona como la plataforma digital de comunicación más

CAPÍTULO 1 : INTRODUCCIÓN

utilizada, alcanzando un 80,1 % de uso habitual según la X Encuesta de Acceso y Uso de Internet, y manteniéndose como una de las principales actividades de comunicación en línea con un 76,2 % de usuarios que reportan utilizarlo para mensajería en la XI Encuesta (Subsecretaría de Telecomunicaciones & Cadem, 2024, 2025).

Asimismo, la viabilidad de este enfoque se ve reforzada por la alta penetración de internet y conectividad móvil en el país, especialmente en zonas rurales, donde el 94,5 % de los hogares dispone de acceso a internet y un 51,4 % se conecta exclusivamente mediante redes móviles (Subsecretaría de Telecomunicaciones & Cadem, 2025). En términos aplicados, el uso de WhatsApp permite facilitar el acceso a información relevante en tiempo real mediante un canal ampliamente conocido y utilizado por los usuarios, reduciendo las barreras de adopción tecnológica y favoreciendo la entrega oportuna de alertas agrícolas. Como se observa en la Figura 1.2, WhatsApp presenta una clara predominancia respecto a otras plataformas digitales en Chile.

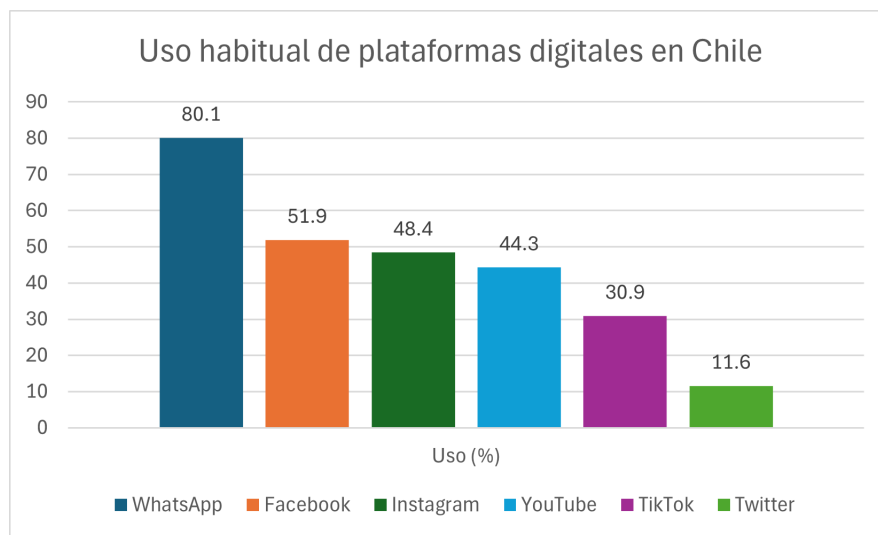


Figura 1.2: *Uso habitual de plataformas digitales en Chile. Fuente: Elaboración propia a partir de Subsecretaría de Telecomunicaciones y Cadem (2024)*

En términos de alcance, este trabajo se enfoca en el diseño e implementación de un sistema agroclimático capaz de integrar múltiples fuentes de información, identificar condiciones de riesgo térmico y entregar alertas oportunas y contextualizadas para su

aplicación en entornos agrícolas, con énfasis en la toma de decisiones operativas a nivel de usuario.

Desde el punto de vista técnico, el problema se aborda como una tarea de detección de eventos raros, donde la superación de umbrales térmicos define la condición de alerta. Este enfoque permite evaluar el desempeño del sistema mediante métricas orientadas a eventos críticos, fundamentales en aplicaciones de gestión de riesgo climático.

En síntesis, esta memoria se posiciona en la intersección entre la predicción climática, la interpretación agronómica y la comunicación efectiva de información, proponiendo una solución tecnológica que transforma datos climáticos dispersos en conocimiento útil y accionable, contribuyendo al desarrollo de una agricultura más resiliente, digitalizada y adaptada a escenarios de variabilidad climática.

A partir de lo anterior, el siguiente apartado presenta la problemática específica que motiva este trabajo, profundizando en las limitaciones actuales de los sistemas agroclimáticos.

1.1. Problemática

A pesar de los avances en monitoreo agroclimático y de la creciente disponibilidad de datos meteorológicos, la agricultura continúa enfrentando una limitación crítica en la transformación de estos datos en información útil para la toma de decisiones. En la práctica, los agricultores acceden a información fragmentada proveniente de estaciones meteorológicas, pronósticos generales y reportes técnicos; sin embargo, carecen de herramientas que integren estos datos y los traduzcan en alertas claras, oportunas y accionables (Selvaraju et al., 2013; World Bank, 2023).

Desde esta perspectiva, uno de los principales problemas radica en la desconexión entre la información meteorológica disponible y su impacto real sobre los cultivos. Los sistemas actuales suelen entregar variables como temperatura mínima o máxima sin considerar los umbrales térmicos específicos asociados al daño agrícola, los cuales

CAPÍTULO 1 : INTRODUCCIÓN

dependen tanto del cultivo como de su estado fenológico. Como consecuencia, la interpretación del riesgo queda sujeta al conocimiento del usuario, dificultando la toma de decisiones oportunas y aumentando la probabilidad de pérdidas productivas (Food and Agriculture Organization of the United Nations, 2005; Richetti et al., 2025).

Adicionalmente, los eventos extremos como las heladas y las olas de calor presentan una naturaleza poco frecuente pero altamente impactante, lo que los convierte en un problema de detección de eventos raros. Bajo este enfoque, los modelos tradicionales tienden a optimizar el desempeño promedio, sin priorizar la correcta identificación de eventos críticos, que son precisamente los que generan mayores pérdidas económicas (Ferro & Stephenson, 2011; Saito & Rehmsmeier, 2015).

De manera complementaria, existe una brecha significativa en la forma en que la información es comunicada al usuario final. La adopción de servicios agroclimáticos depende no solo de la precisión del pronóstico, sino también de su accesibilidad, claridad y pertinencia. No obstante, muchos sistemas requieren conocimientos técnicos para su interpretación o no se encuentran integrados en plataformas de uso cotidiano, lo que limita su utilización en el contexto agrícola (Cegnar et al., 2023; Sharma et al., 2021). Como se muestra en la Figura 1.3, la forma en que se entrega la información agroclimática influye directamente en su adopción por parte de los agricultores. En particular, el uso de plataformas accesibles y alineadas con los hábitos del usuario, como sistemas basados en telefonía móvil, puede mejorar significativamente el cumplimiento de recomendaciones agronómicas.

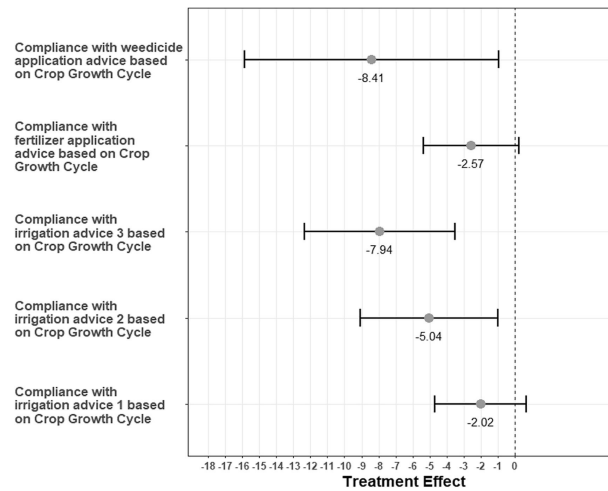


Figura 1.3: Efecto del uso de servicios de mensajería móvil en la adopción de recomendaciones agrícolas por parte de los agricultores. Fuente: (Sharma et al., 2021).

Finalmente, si bien existen avances en modelos de predicción meteorológica basados en inteligencia artificial, estos no abordan directamente la escala de decisión agrícola ni incorporan mecanismos de traducción de variables climáticas a impacto productivo (Kochkov et al., 2024; Lam et al., 2023). En consecuencia, la ausencia de sistemas que integren predicción, interpretación mediante umbrales de daño y comunicación efectiva constituye una limitación crítica para la gestión del riesgo climático en la agricultura.

En síntesis, la problemática identificada no radica únicamente en la disponibilidad de datos o en la capacidad de predicción, sino en la falta de articulación entre la generación de información climática, su interpretación en términos de impacto agrícola y su comunicación efectiva al usuario final. En este sentido, surge la necesidad de desarrollar soluciones que integren múltiples fuentes de datos, identifiquen condiciones de riesgo basadas en umbrales térmicos y entreguen alertas claras, oportunas y accionables mediante interfaces accesibles para los usuarios agrícolas.

A partir de esta problemática, se definen los objetivos del presente trabajo, orientados a abordar las limitaciones identificadas mediante un enfoque integrado.

1.2. Objetivos

En coherencia con la brecha identificada, los objetivos se orientan a integrar estos tres componentes en un sistema unificado.

1.2.1. Objetivo Principal

Desarrollar e implementar un sistema basado en inteligencia artificial y arquitectura multiagente capaz de integrar la predicción de eventos térmicos extremos, su interpretación mediante umbrales de daño agrícola y la comunicación de alertas oportunas y comprensibles para apoyar la toma de decisiones en el ámbito agrícola.

1.2.2. Objetivos Específicos

- Integrar y procesar datos provenientes de estaciones meteorológicas, servicios de pronóstico y otras fuentes climáticas relevantes.
- Diseñar e implementar modelos de aprendizaje automático para la predicción de eventos térmicos extremos, incluyendo heladas y olas de calor.
- Definir y aplicar umbrales térmicos críticos basados en literatura técnica para interpretar el impacto agrícola de eventos térmicos extremos y apoyar la generación de alertas tempranas.
- Desarrollar una arquitectura multi-agente que permita el procesamiento automatizado, modular y escalable de la información climática.
- Implementar un sistema de comunicación mediante una plataforma conversacional (WhatsApp) para la entrega de alertas personalizadas.
- Evaluar el desempeño del sistema en la detección de eventos térmicos extremos mediante métricas orientadas a la evaluación de eventos poco frecuentes y conjuntos de datos desbalanceados.

2. Estado del Arte

El desarrollo de soluciones para la gestión del riesgo agroclimático ha evolucionado significativamente en los últimos años, impulsado por la expansión de las tecnologías digitales aplicadas a la agricultura, incluyendo sensores, IoT, analítica de datos e Inteligencia Artificial (IA), así como por una mayor disponibilidad de datos agrometeorológicos y servicios climáticos digitales (Food and Agriculture Organization of the United Nations & World Meteorological Organization, 2021; Sotomayor et al., 2023).

No obstante, la literatura muestra que estos avances aún presentan brechas importantes en la articulación entre la producción de información climática, su traducción en decisiones agrícolas y su llegada efectiva al usuario final (Born et al., 2021; Cegnár et al., 2023).

2.1. Predicción de eventos de heladas mediante aprendizaje automático

La predicción de heladas ha evolucionado desde métodos empíricos y modelos físicos de escala sinóptica hacia enfoques de aprendizaje automático (ML), capaces de capturar relaciones no lineales entre variables meteorológicas y condiciones microclimáticas.

Un antecedente temprano y directamente relacionado con Chile es el trabajo de Möller-Acuña et al. (2017), quienes desarrollaron un modelo basado en información agrometeorológica histórica (datos de 2010–2016) para anticipar episodios de helada en la Región del Maule, con un horizonte de predicción evaluado principalmente de hasta 12 horas, sin considerar divisiones estandarizadas como 6h, 24h o 48h.

En este estudio, el dataset presentaba un marcado desequilibrio de clases, típico de eventos raros (con una mayor proporción de casos de “no helada” respecto a heladas), lo que fue abordado mediante técnicas como SMOTE para el sobremuestreo sintético de la clase minoritaria. Este enfoque se vincula directamente con el problema de detección de

CAPÍTULO 2 : ESTADO DEL ARTE

eventos raros. Al comparar distintos algoritmos, Random Forest (RF) obtuvo el mejor desempeño general, alcanzando eficiencias superiores al 85–90 % en la mayoría de los escenarios evaluados, superando a métodos como SVM y redes neuronales.

En una línea similar, también en Chile, Fuentes et al. (2018) propusieron una red neuronal artificial (ANN) entrenada mediante el algoritmo de backpropagation para predecir la temperatura mínima del día siguiente y detectar heladas radiativas a partir de variables como temperatura del aire, humedad relativa, radiación, precipitación, dirección y velocidad del viento. Utilizando datos de diez estaciones de la red AGROMET entre 2010 y 2017, los autores obtuvieron un error cuadrático medio de 2,99 °C en la predicción de temperatura mínima. Asimismo, en la detección de heladas, reportaron una exactitud promedio de 98 %, una sensibilidad de 86 % y una tasa de error de 2 %. Este trabajo resulta particularmente relevante, ya que demuestra la factibilidad de construir modelos operativos utilizando datos de estaciones meteorológicas de uso agrícola en la zona central de Chile.

En este contexto, la distribución espacial de estaciones meteorológicas utilizadas en estos estudios constituye un factor clave para la calidad del modelo. Como se muestra en la Figura 2.1, la red AGROMET permite capturar la variabilidad climática presente en la zona central del país.

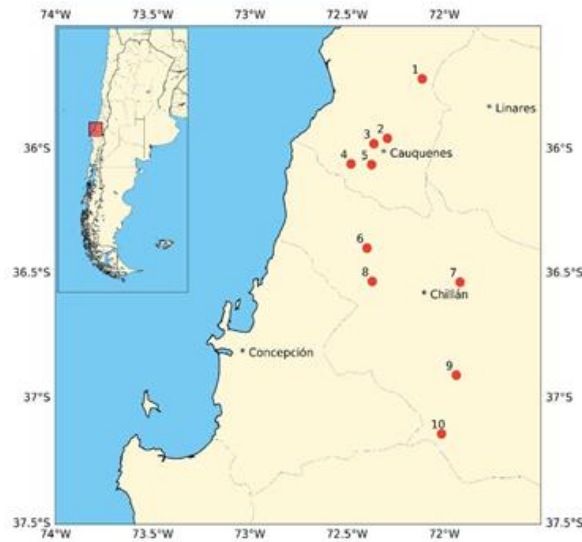


Figura 2.1: Distribución de estaciones de la red AGROMET utilizadas para modelación agroclimática en Chile central. Adaptado de Fuentes et al. (2018).

El modelo propuesto por Fuentes et al. (2018) se basa en una ANN que integra múltiples variables meteorológicas como entradas, tal como se observa en la Figura 2.2.

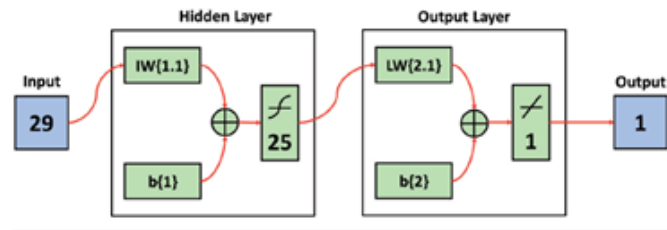


Figura 2.2: Arquitectura de red neuronal utilizada para la predicción de temperatura mínima. Adaptado de Fuentes et al. (2018).

En una línea complementaria, Diedrichs et al. (2018) demostraron que la incorporación de información termodinámica proveniente de sensores cercanos, dentro de una red Internet de las Cosas (IoT), permite mejorar la predicción de temperaturas mínimas y de eventos de helada en escalas locales.

Posteriormente, Talsma et al. (2023) compararon modelos Random Forest (RF), redes neuronales profundas (DNN) y red neuronal convolucional (CNN) para horizontes de predicción entre 6 y 48 horas, utilizando aproximadamente diez años de datos horarios de una estación agrícola en Nuevo México. Además, evaluaron el desempeño de estos modelos en un segundo sitio, con el objetivo de analizar su capacidad de generalización.

Sus resultados muestran que el error de predicción (RMSE) aumenta de forma consistente con el horizonte temporal, pasando de valores cercanos a 1.5–1.7 °C a 6 horas hasta valores superiores a 3 °C a 48 horas. Asimismo, se observa una degradación del desempeño al aplicar los modelos en un entorno distinto al de entrenamiento, lo que evidencia limitaciones en la generalización de estos enfoques. Los resultados comparativos se presentan en la Tabla 3.

Tabla 3: Desempeño de modelos para la predicción de temperatura mínima (heladas) en Nuevo México, con diferentes horizontes. Adaptado de Talsma et al. (2023).

Modelo	Horizonte (h)	Datos (período)	RMSE (°C)	R ²	F1 (0°C)
RF	6	Alcalde, NM (10 años)	1.72	0.98	0.95
	12		2.07	0.97	0.94
	24		2.37	0.96	0.94
	30		2.83	0.94	0.93
	36		3.01	0.94	0.92
	48		3.16	0.93	0.92
DNN	6	Alcalde	1.53	0.98	0.96
	12		1.84	0.98	0.95
	24		2.23	0.96	0.94
	30		2.69	0.95	0.92
	36		2.96	0.95	0.91
	48		2.94	0.94	0.93
CNN	6	Alcalde	1.57	0.98	0.96
	12		1.88	0.98	0.95
	24		2.19	0.97	0.95
	30		2.66	0.95	0.93
	36		2.89	0.95	0.92
	48		3.06	0.94	0.93
CNN	6	Freshies Farm, NM (2018–2020)	1.97	–	–
	12		3.27	–	–
	24		3.45	–	–
	30		3.70	–	–
	36		3.86	–	–
	48		3.94	–	–
RF	6	Freshies Farm	2.00	–	–
	12		3.29	–	–
	24		3.59	–	–
	30		3.63	–	–
	36		4.09	–	–
	48		3.88	–	–

CAPÍTULO 2 : ESTADO DEL ARTE

Como se observa en la Tabla 3, todos los modelos presentan un incremento significativo del error a medida que aumenta el horizonte de predicción, lo que refleja la creciente incertidumbre asociada a estimaciones de mayor plazo. En particular, los modelos basados en redes neuronales (DNN y CNN) presentan un desempeño ligeramente superior en horizontes cortos, mientras que las diferencias entre modelos tienden a reducirse a medida que aumenta el horizonte.

De manera complementaria, otros estudios han ampliado este enfoque hacia la clasificación horaria y la mejora de la representación espacial del riesgo. Noh et al. (2021) desarrollaron un sistema horario de predicción de heladas optimizado para huertos, combinando datos de estaciones meteorológicas automáticas e imágenes de cámara, logrando altos niveles de clasificación correcta tanto en observaciones nacionales como en entornos de huerto.

Más recientemente, Almeida et al. (2025) propusieron un comité de votación basado en Adaptive Boosting (AdaBoost), combinando un índice de heladas derivado de un modelo meteorológico regional con un modelo de ML, evaluando horizontes de 24 y 72 horas para Brasil y zonas cercanas de Sudamérica. De manera complementaria, Bhakare et al. (2025) abordaron el problema desde el downscaling espacial, reduciendo la resolución de pronósticos de temperatura mínima desde 12 km hasta 250 m mediante modelos ANN, RF y CNN, destacando el mejor desempeño de la CNN en terrenos complejos.

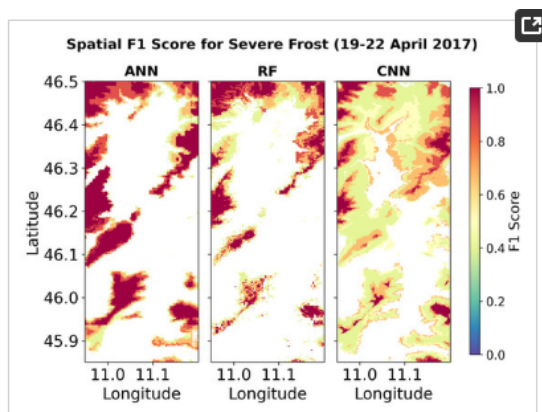


Figura 2.3: Desempeño espacial del modelo CNN para predicción de temperatura mínima. Adaptado de Bhakare et al. (2025).

A pesar de los altos niveles de desempeño reportados en contextos locales, la transferencia de estos modelos a nuevos predios o microclimas sigue siendo un desafío abierto, especialmente cuando varían factores como la topografía, la instrumentación o la densidad de observaciones.

Además, la predicción de heladas constituye un problema inherentemente desbalanceado, dado que los eventos críticos ocurren con menor frecuencia que las condiciones normales. En este sentido, métricas como la exactitud global pueden sobrestimar el desempeño real del sistema. Por ello, para aplicaciones operacionales resulta fundamental complementar el error promedio con métricas como recall, precisión y otras orientadas a la detección de eventos raros (Ferro & Stephenson, 2011; Pérez Tárraga et al., 2025; Saito & Rehmsmeier, 2015; Talsma et al., 2023).

2.2. Modelos avanzados de predicción meteorológica y climática basados en inteligencia artificial

A diferencia de los enfoques locales descritos anteriormente, los avances recientes en IA aplicada al pronóstico meteorológico resultan altamente relevantes, ya que mejoran significativamente la calidad de los datos de entrada que alimentan los sistemas locales de alerta.

En este contexto, Lam et al. (2023) introducen GraphCast, un modelo basado en redes neuronales de grafos entrenado con cuatro décadas de datos de reanálisis ERA5 del ECMWF. Este modelo es capaz de generar pronósticos globales de hasta 10 días con alta resolución espacial (0.25°), superando al sistema determinista de referencia (HRES) en más del 90 % de las variables y horizontes evaluados. Además, destaca por su eficiencia computacional, produciendo predicciones en menos de un minuto, en contraste con los métodos numéricos tradicionales que requieren horas de cálculo en supercomputadores (Lam et al., 2023).

Estos resultados consolidan la idea de que los enfoques basados en datos pueden complementar e incluso superar, en ciertos escenarios, a la predicción numérica tradicional. En esta misma línea, desarrollos posteriores han ampliado las capacidades de estos modelos. En particular, Price, Sanchez-Gonzalez et al. (2024) proponen GenCast, un modelo probabilístico basado en técnicas de difusión que extiende el enfoque determinista de GraphCast hacia la generación de pronósticos por conjuntos (ensemble).

Este modelo genera aproximadamente 50 trayectorias posibles del estado atmosférico, lo que permite capturar la incertidumbre inherente al sistema climático y extender el horizonte de predicción hasta 15 días. Entrenado también con datos ERA5, GenCast ha demostrado superar al sistema por conjuntos del ECMWF (ENS) en el 97.2 % de las combinaciones evaluadas, alcanzando incluso un 99.8 % de superioridad en horizontes mayores a 36 horas (Price, Sanchez-Gonzalez et al., 2024).

En conjunto, estos avances evidencian un cambio de paradigma en el pronóstico meteorológico, donde los modelos basados en ML no solo replican, sino que mejoran el desempeño de los sistemas tradicionales, proporcionando predicciones más rápidas, precisas y con estimaciones explícitas de incertidumbre.

Este progreso tiene implicancias directas en el ámbito agroclimático, ya que permite alimentar sistemas de alerta temprana con información meteorológica de mayor calidad, particularmente en variables críticas como la temperatura, mejorando así la detección oportuna de eventos extremos como heladas y olas de calor.

En la misma dirección, Kochkov et al. (2024) presentan NeuralGCM, un modelo de circulación general que combina un solucionador diferenciable de la dinámica atmosférica con componentes de aprendizaje profundo (DL). Este enfoque demuestra que es posible obtener un desempeño competitivo tanto en pronóstico meteorológico como en simulación climática de largo plazo, con costos computacionales considerablemente menores que los de los modelos físicos convencionales.

Para aplicaciones agroclimáticas, estos avances no reemplazan la modelación local, pero sí refuerzan la viabilidad de construir sistemas de alerta apoyados en productos meteorológicos generados o refinados mediante IA.

Sin embargo, si bien estos avances refuerzan la calidad potencial del pronóstico meteorológico de base, no resuelven por sí mismos la escala de decisión agrícola, la cual depende de condiciones locales y, frecuentemente, microclimáticas. Desde esta perspectiva, el desafío no termina en predecir la temperatura, sino en traducir información atmosférica de mayor escala en riesgo localizado e interpretable para contextos productivos específicos.

Esta brecha es ampliamente reconocida en la literatura de servicios climáticos como el *last-mile problem*, que describe la dificultad de transferir información climática desde su generación a nivel global o regional hasta su uso efectivo en la toma de decisiones a nivel local (Vaughan & Dessai, 2014; World Bank, 2023).

Por ello, el valor de estos modelos para aplicaciones agroclimáticas depende de su articulación con procesos de downscaling espacial, la selección de variables relevantes y la interpretación mediante umbrales térmicos de impacto (Kochkov et al., 2024; Lam et al., 2023; Möller et al., 2022).

2.3. Impacto de las temperaturas extremas y umbrales térmicos en los cultivos

Si bien los avances en predicción meteorológica permiten estimar variables climáticas con mayor precisión, su utilidad en el ámbito agrícola depende de comprender su impacto sobre los cultivos.

El problema de las heladas no puede entenderse únicamente desde la meteorología; también requiere considerar los umbrales de daño biológico y productivo. Desde esta perspectiva, la FAO ha documentado que el daño por helada ocurre cuando la temperatura del tejido vegetal cae por debajo de valores críticos que dependen del cultivo, del órgano afectado y del estado fenológico (Food and Agriculture Organization of the United Nations, 2005). Esta base resulta fundamental para diferenciar entre una baja temperatura meteorológica y un evento potencialmente dañino para la producción.

En esta línea, la FAO define umbrales críticos de daño térmico específicos para distintos cultivos y etapas fenológicas. La Tabla 4 presenta ejemplos representativos de estos valores.

Tabla 4: Temperaturas críticas de daño por heladas en cultivos. Adaptado de Food and Agriculture Organization of the United Nations (2005).

Cultivo	Temperatura crítica (°C)
Vid (brotación)	-1 a -2
Trigo	-2 a -4
Frutales	-1 a -3

En paralelo, la literatura reciente ha mostrado que el calor extremo constituye un riesgo agrícola tan relevante como el frío. Zhao et al. (2017) estimaron que cada aumento de 1 °C en la temperatura media global reduce, en promedio, los rendimientos de trigo, arroz, maíz y soja en 6,0 %, 3,2 %, 7,4 % y 3,1 %, respectivamente.

De manera más reciente, Tran et al. (2025) identificaron efectos umbral en la respuesta de los cultivos al calentamiento. En trigo, las pérdidas aumentan de 6,1 % a 8,2 % por grado cuando el incremento térmico supera 2,38 °C; en arroz, las pérdidas pasan de 1,1 % a 7,1 % por grado al superar 3,13 °C; y en maíz se observa una disminución promedio de 4,03 % por grado sin un umbral definido.

A nivel fisiológico, Kumar et al. (2025) sintetizan que el estrés térmico altera procesos críticos como la absorción de agua, la actividad enzimática, la fotosíntesis y el desarrollo reproductivo. En conjunto, esta evidencia refuerza que un sistema agroclimático no debiera restringirse únicamente a la predicción de heladas, sino también incorporar la detección de episodios de calor capaces de generar pérdidas significativas.

Desde una perspectiva aplicada, no basta con pronosticar temperaturas mínimas o máximas; es necesario vincular estas variables con umbrales térmicos de daño dependientes del cultivo, del órgano afectado y del estado fenológico. Esta distinción entre variable meteorológica y umbral de impacto es central, ya que una misma temperatura puede tener consecuencias productivas muy diferentes según el contexto biológico.

Adicionalmente, la literatura reciente advierte que muchos modelos agronómicos aún presentan dificultades para representar eventos térmicos breves pero severos, precisamente aquellos de mayor interés para un sistema de alerta temprana (Food and Agriculture Organization of the United Nations, 2005; Richetti et al., 2025). En este sentido, se evidencia la necesidad de enfoques que no solo predigan variables climáticas, sino que también permitan interpretar su impacto en función de condiciones específicas del cultivo.

2.4. Servicios agroclimáticos y sistemas de alerta temprana

A partir de la necesidad de interpretar correctamente el impacto de las variables climáticas, surge el desafío de comunicar esta información de manera efectiva a los usuarios finales.

La literatura sobre servicios climáticos subraya que la utilidad práctica de un pronóstico depende de su capacidad para ser traducido en mensajes accionables. Bajo este enfoque, Hansen et al. (2019) explican que los servicios climáticos deben alinearse con las necesidades de los tomadores de decisiones y ser adecuadamente interpretados para su uso. Asimismo, Selvaraju et al. (2013) sostienen que los servicios climáticos locales para la agricultura deben integrar información meteorológica con conocimientos locales, canales de comunicación confiables y mecanismos de retroalimentación bidireccional, de modo que apoyen decisiones concretas de manejo.

En esta misma dirección, reportes recientes del programa CCAFS de CGIAR enfatizan la necesidad de co-diseñar los servicios climáticos junto a los usuarios finales, incorporando sus prácticas, contextos productivos y capacidades de interpretación, con el fin de mejorar su adopción y utilidad en la toma de decisiones agrícolas (CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS), 2019). De manera complementaria, estudios recientes confirman que la adopción de los servicios agrometeorológicos depende no solo de la precisión de los pronósticos, sino también de su accesibilidad, comprensibilidad y pertinencia para el agricultor (Cegnar et al., 2023).

Desde una perspectiva operacional, Shin et al. (2025) describen un sistema de alerta temprana implementado en Corea del Sur, orientado a explotaciones agrícolas a escala metropolitana. Este sistema integra datos meteorológicos a nivel predial, pronóstico de riesgos y entrega de alertas a agricultores y organismos locales mediante plataformas digitales y sistemas de información agroclimática coordinados. Este enfoque destaca por su carácter institucional y por la integración de múltiples actores dentro del proceso de generación y difusión de alertas.

De forma complementaria, Sharma et al. (2021) analizan experiencias en India, donde los avisos agrometeorológicos son distribuidos a agricultores a través de telefonía móvil, principalmente mediante SMS y servicios de mensajería. Sus resultados muestran que este tipo de canales mejora la adopción de recomendaciones agronómicas

y el uso de pronósticos meteorológicos en decisiones de riego y manejo.

En conjunto, estos estudios evidencian que la utilidad de una alerta no depende únicamente del desempeño predictivo del sistema, sino también de su capacidad de llegar al usuario final en un formato comprensible, oportuno y accionable. Asimismo, la adopción efectiva de estos servicios requiere monitoreo y evaluación de uso, ya que la recepción del mensaje no garantiza por sí sola cambios en las prácticas de manejo.

En el contexto chileno, la Red Agroclimática Nacional (AGROMET) constituye un antecedente institucional relevante para la provisión de información meteorológica agrícola (Ministerio de Agricultura de Chile, 2024). No obstante, persiste el desafío de articular el pronóstico meteorológico con umbrales térmicos de impacto y recomendaciones agronómicas en una experiencia de usuario coherente, accesible y orientada a la toma de decisiones (Sharma et al., 2021; World Meteorological Organization, 2023).

2.5. Sistemas conversacionales y recuperación de conocimiento para asesoría agrícola

En este contexto, los sistemas conversacionales emergen como una alternativa para mejorar la accesibilidad y comprensión de la información agroclimática.

Una línea más reciente del estado del arte aborda cómo hacer accesible la información agrícola especializada mediante interfaces conversacionales. Bajo este enfoque, Koopman et al. (2024) desarrollan AgAsk, un agente conversacional orientado a responder consultas de agricultores a partir de documentos científicos, desarrollado en un contexto académico con aplicaciones en el ámbito agrícola australiano.

El sistema utiliza modelos modernos de recuperación de información y una arquitectura de despliegue sobre Telegram, integrando componentes de búsqueda documental y modelos de lenguaje para transformar conocimiento científico en respuestas accesibles para el usuario, como se observa en la Figura 2.4.

El valor de este enfoque radica en reconocer que una parte importante del conoci-

CAPÍTULO 2 : ESTADO DEL ARTE

miento agrícola existe en texto libre y no en bases de datos estructuradas, lo que requiere mecanismos de recuperación y presentación adaptados al lenguaje de los usuarios. Sin embargo, el sistema se centra principalmente en la consulta de información documental y no integra datos meteorológicos en tiempo real ni procesos de interpretación agroclimática, lo que limita su aplicación en sistemas de alerta temprana basados en condiciones dinámicas.

Esta limitación evidencia una brecha relevante, particularmente en contextos como el chileno, donde no existen soluciones equivalentes que combinen acceso conversacional al conocimiento agrícola con monitoreo climático e interpretación de riesgos a nivel local.

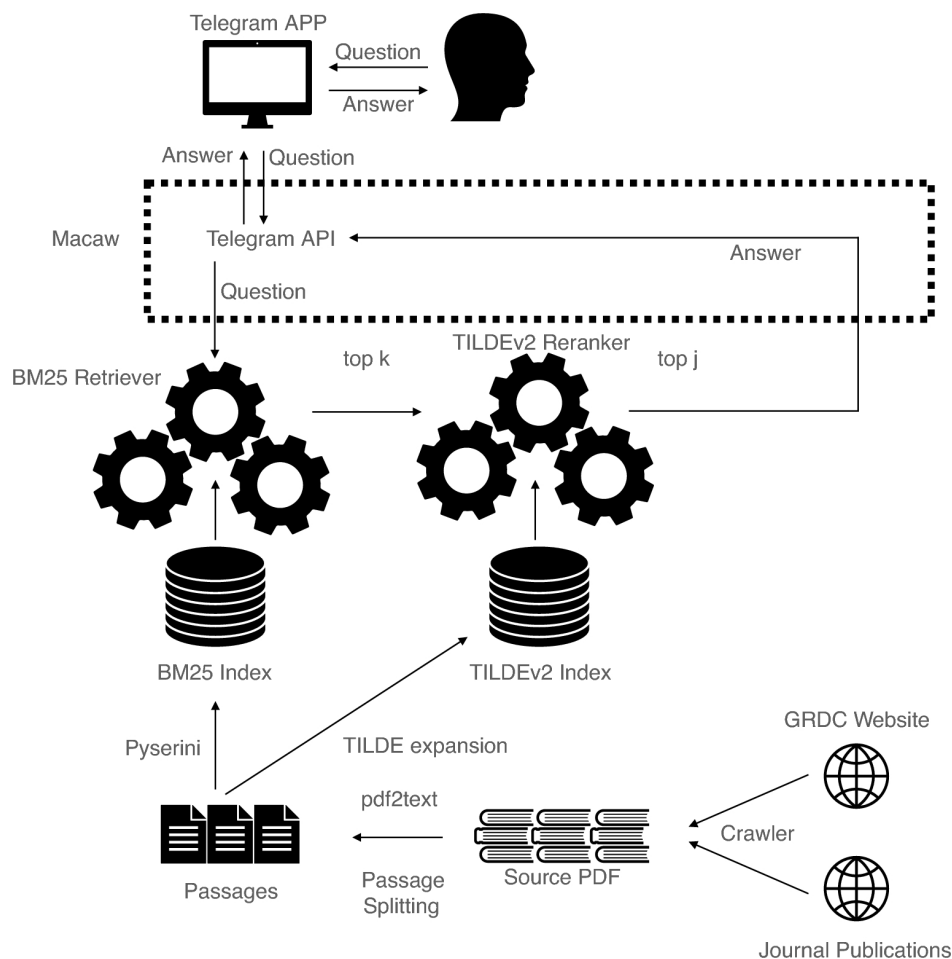


Figura 2.4: Arquitectura del sistema AgAsk para recuperación de conocimiento agrícola mediante modelos de lenguaje. Adaptado de Koopman et al. (2024).

Más recientemente, Xia et al. (2026) proponen ChatCEA, un agente inteligente para agricultura en ambiente controlado que integra LLM con generación aumentada por recuperación (RAG) y una base de conocimiento específica del dominio. Los autores muestran que la incorporación de conocimiento especializado mejora la precisión, la capacidad de responder preguntas cuantitativas y la concisión de las respuestas frente a buscadores y modelos generales.

Si bien estos estudios no se centran en la predicción de heladas o eventos de calor extremo, resultan altamente relevantes para el diseño de interfaces conversacionales capaces de traducir información técnica en recomendaciones comprensibles para el

usuario final.

En este contexto, los agentes conversacionales representan una capa prometedora de acceso, interpretación y explicación de información agroclimática. No obstante, su valor depende de que las respuestas estén ancladas en fuentes verificables y en reglas de decisión explícitas. En otras palabras, la interfaz conversacional no reemplaza al sistema de predicción ni a la lógica de alerta, sino que actúa como una capa de traducción y trazabilidad para el usuario.

En consecuencia, la brecha actual no radica únicamente en la capacidad de responder preguntas agrícolas, sino en la integración de mecanismos de recuperación de conocimiento con datos meteorológicos operacionales y criterios transparentes de activación de alertas (Koopman et al., 2024; Xia et al., 2026).

2.6. Síntesis y brecha de la literatura

La revisión realizada muestra que la literatura dispone de avances significativos en distintas capas del problema agroclimático. Por una parte, existen antecedentes sólidos para la predicción local de heladas mediante modelos de ML, incluyendo aplicaciones sobre datos agrometeorológicos en Chile y en otros contextos agrícolas (Almeida et al., 2025; Fuentes et al., 2018; Möller-Acuña et al., 2017; Noh et al., 2021; Talsma et al., 2023).

Por otra parte, los desarrollos recientes en pronóstico meteorológico basado en IA y técnicas de downscaling espacial han ampliado las posibilidades de generar información atmosférica de mayor calidad para aplicaciones locales (Bhakare et al., 2025; Kochkov et al., 2024; Lam et al., 2023; Möller et al., 2022). A ello se suma evidencia agronómica robusta que demuestra que tanto las bajas temperaturas como el calor extremo pueden producir pérdidas significativas, y que tales impactos dependen de umbrales térmicos específicos del cultivo y del estado fenológico (Food and Agriculture Organization of the United Nations, 2005; Kumar et al., 2025; Richetti et al., 2025; Tran et al., 2025; Zhao et al., 2017).

Asimismo, la literatura sobre servicios climáticos y alertas tempranas ha mostrado que la utilidad de un pronóstico depende no solo de su precisión, sino también de su localización, comprensibilidad y capacidad de llegar al usuario final mediante mecanismos de comunicación adecuados (Cegnar et al., 2023; Selvaraju et al., 2013; Sharma et al., 2021; Shin et al., 2025; World Meteorological Organization, 2023). Finalmente, los avances recientes en agentes conversacionales y sistemas basados en LLM con RAG evidencian el potencial de traducir conocimiento técnico en interacciones más accesibles para el usuario, aunque todavía presentan una integración limitada respecto de flujos meteorológicos operacionales y reglas explícitas de activación de alertas (Koopman et al., 2024; Xia et al., 2026).

En conjunto, estos antecedentes evidencian que la brecha principal no radica en la ausencia de técnicas aisladas, sino en la falta de articulación entre tres componentes que, en la práctica, deberían operar de manera integrada: (i) la predicción de variables térmicas, (ii) la interpretación de dichas variables mediante umbrales térmicos de riesgo, y (iii) la comunicación oportuna de resultados en formatos comprensibles y accionables.

La Tabla 5 presenta un resumen de los trabajos revisados, mapeando cada uno de ellos respecto de estos tres componentes. Como se observa, la mayoría de las propuestas aborda solo una o dos de estas dimensiones, lo que evidencia la ausencia de soluciones integradas capaces de cubrir el ciclo completo desde la generación de datos hasta la toma de decisiones en terreno.

En función de la brecha identificada, el presente trabajo se posiciona en la intersección de estas tres capas, proponiendo una solución orientada a integrar predicción, interpretación y comunicación en un sistema unificado. En particular, se busca convertir información climática heterogénea en alertas agroclimáticas interpretables y accionables para contextos agrícolas específicos, sentando las bases para el desarrollo de la propuesta presentada en el capítulo siguiente.

CAPÍTULO 2 : ESTADO DEL ARTE

Tabla 5: Resumen del estado del arte según componentes abordados.

Trabajo	Predicción	Interpretación	Comunicación	Datos en tiempo real
Möller-Acuña et al. (2017)	✓	×	×	✓
Fuentes et al. (2018)	✓	×	×	✓
Diedrichs et al. (2018)	✓	×	×	✓
Talsma et al. (2023)	✓	×	×	×
Noh et al. (2021)	✓	×	×	✓
Almeida et al. (2025)	✓	×	×	✓
Bhakare et al. (2025)	✓	×	×	×
Lam et al. (2023)	✓	×	×	✓
Price, Sanchez-Gonzalez et al. (2024)	✓	×	×	✓
Kochkov et al. (2024)	✓	×	×	×
Shin et al. (2025)	✓	✓	✓	✓
Sharma et al. (2021)	×	×	✓	✓
Ministerio de Agricultura de Chile (2024)	✓	×	✓	✓
Koopman et al. (2024)	×	×	✓	×
Xia et al. (2026)	×	×	✓	×
Propuesta de este trabajo	✓	✓	✓	✓

Para finalizar, el capítulo siguiente presenta la propuesta de solución, orientada a integrar los distintos componentes analizados en un sistema unificado.

3. Propuesta de Solución

En función de la brecha identificada en el capítulo anterior, se propone el diseño e implementación de un sistema agroclimático basado en una arquitectura Multi-LLM-Agent, orientado a integrar la predicción de eventos térmicos extremos, su interpretación mediante umbrales térmicos de daño agrícola y la comunicación de alertas tempranas comprensibles y accionables para el usuario final.

La propuesta se enfoca inicialmente en la Región Metropolitana de Chile, considerando la disponibilidad de datos meteorológicos, su relevancia agrícola y la factibilidad de implementación en un contexto controlado.

A diferencia de enfoques tradicionales, esta propuesta aborda de manera conjunta las tres dimensiones del problema, cerrando el ciclo completo desde la generación de información climática hasta su utilización en la toma de decisiones agrícolas.

3.1. Visión general de la solución

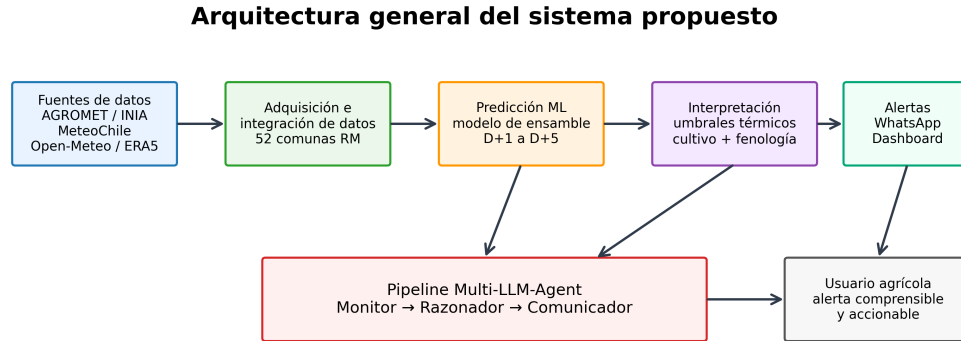
La arquitectura propuesta se fundamenta en un enfoque Multi-LLM-Agent, donde múltiples agentes especializados colaboran para transformar información meteorológica en conocimiento útil para la toma de decisiones agrícolas.

A diferencia de enfoques tradicionales centrados exclusivamente en la predicción de variables climáticas, la propuesta incorpora una capa de razonamiento distribuido basada en modelos de lenguaje de gran escala (LLMs), permitiendo interpretar predicciones, contextualizar riesgos y generar recomendaciones comprensibles para los usuarios finales.

En este contexto, los modelos predictivos constituyen una fuente de información para la arquitectura Multi-LLM-Agent, mientras que el principal aporte de la solución corresponde a la coordinación de agentes especializados capaces de monitorear condiciones meteorológicas, razonar sobre su impacto potencial y comunicar alertas agroclimáticas contextualizadas.

CAPÍTULO 3 : PROPUESTA DE SOLUCIÓN

Para comprender de manera global el funcionamiento del sistema propuesto, la Figura 3.1 presenta su arquitectura general.



El sistema integra predicción, interpretación agronómica y comunicación en un flujo unificado.

Figura 3.1: *Arquitectura general del sistema propuesto.*

Tal como se muestra en la Figura 3.1, el sistema integra múltiples componentes en un flujo continuo que transforma datos climáticos en alertas accionables.

Cabe señalar que la arquitectura presentada corresponde a un nivel conceptual, cuyo diseño técnico detallado será desarrollado en el Capítulo 5.

El sistema opera mediante un pipeline automatizado que se ejecuta diariamente en horario nocturno, iniciando a las 02:00 AM y finalizando aproximadamente antes de las 04:00 AM. Este ciclo nocturno permite recopilar datos recientes, validar información del día anterior, generar predicciones y emitir alertas antes del inicio de la jornada agrícola, de modo que la información esté disponible para los usuarios antes de las 06:00 AM aproximadamente.

3.2. Justificación del Enfoque Multi-LLM-Agent

La agricultura inteligente requiere transformar grandes volúmenes de información meteorológica en conocimiento útil para la toma de decisiones. Si bien los modelos

CAPÍTULO 3 : PROPUESTA DE SOLUCIÓN

predictivos permiten detectar la posible ocurrencia de eventos térmicos extremos, la generación de alertas efectivas exige además interpretar dichas predicciones considerando el contexto agrícola, los umbrales de daño asociados a distintos cultivos y las acciones que pueden ser adoptadas por los usuarios.

Una aproximación basada en un único modelo de lenguaje concentra todas estas responsabilidades dentro de un mismo componente, dificultando la trazabilidad del razonamiento, la reutilización de conocimiento especializado y la evaluación independiente de cada función. Asimismo, limita la capacidad de incorporar nuevos procesos de análisis sin afectar el comportamiento global del sistema.

Por esta razón, la presente investigación propone una arquitectura Multi-LLM-Agent compuesta por agentes especializados que colaboran para transformar información meteorológica y predicciones numéricas en alertas agroclimáticas contextualizadas. Cada agente cumple una responsabilidad específica dentro del proceso de monitoreo, interpretación y comunicación, permitiendo separar funciones, mejorar la modularidad de la solución y facilitar la incorporación de nuevas capacidades en el futuro.

La arquitectura propuesta constituye el elemento central de esta investigación, actuando como una capa inteligente capaz de integrar conocimiento meteorológico, información agrícola y capacidades de razonamiento basadas en modelos de lenguaje para apoyar la toma de decisiones en contextos agroclimáticos.

CAPÍTULO 3 : PROPUESTA DE SOLUCIÓN

Tabla 6: Comparación conceptual de enfoques para la generación de alertas agroclimáticas

Criterio	Sistema tradicional	LLM único	Arquitectura Multi-LLM-Agent propuesta
Trazabilidad del razonamiento	Baja; se basa principalmente en reglas fijas y umbrales predefinidos.	Media; el razonamiento queda concentrado en una única respuesta generada por el modelo.	Alta; el proceso se separa en etapas distinguibles de monitoreo, razonamiento y comunicación.
Modularidad	Baja; las funciones suelen estar acopladas dentro de un mismo flujo de alerta.	Media; concentra varias responsabilidades en un único componente.	Alta; cada agente cumple una responsabilidad específica dentro del sistema.
Escalabilidad	Limitada; incorporar nuevas fuentes o criterios requiere ajustes manuales.	Media; puede adaptarse a nuevos casos, pero aumenta la complejidad del prompt.	Alta; permite incorporar nuevas fuentes, agentes o proveedores de LLM.
Explicación al usuario	Limitada; entrega alertas directas con poca contextualización.	Media-alta; puede generar lenguaje natural, aunque con menor separación de responsabilidades.	Alta; transforma la predicción en mensajes contextualizados y comprensibles para el usuario.

3.3. Enfoque conceptual de la solución

La propuesta se estructura en torno a tres componentes fundamentales que responden directamente a la brecha identificada en el estado del arte: la predicción de eventos térmicos extremos, su interpretación mediante umbrales térmicos y la comunicación de alertas tempranas al usuario final.

Desde una perspectiva sistémica, estos componentes no operan de manera independiente, sino que se integran en un flujo unificado orientado a la toma de decisiones operativas en contextos agrícolas. En particular, la predicción permite estimar la ocurrencia de condiciones climáticas adversas, la interpretación traduce dichas condiciones en niveles de riesgo agrícola específicos y la comunicación transforma estos resultados

en recomendaciones comprensibles y accionables.

Este enfoque permite cerrar el ciclo completo de un sistema agroclimático, abordando no solo la generación de información, sino también su contextualización y uso efectivo en terreno.

3.4. Componentes de la solución

La solución se organiza en tres capas funcionales.

3.4.1. Capa 1 — Adquisición e integración de datos

El sistema integra datos provenientes de múltiples fuentes: observaciones de la red AGROMET (INIA), pronósticos de MeteoChile y datos de reanálisis y forecast de la API Open-Meteo (ERA5).

Los datos son combinados mediante una jerarquía de confiabilidad, priorizando observaciones reales por sobre reanálisis, y estos por sobre bases históricas extendidas.

Para comunas sin estación meteorológica, se aplica interpolación espacial dentro de un radio de 40 km, garantizando cobertura completa en la Región Metropolitana.

3.4.2. Capa 2 — Predicción e interpretación

La predicción de eventos térmicos extremos se realiza mediante un modelo de aprendizaje automático basado en un enfoque de ensamble, diseñado específicamente para la detección de eventos raros como heladas y olas de calor.

El modelo utiliza un conjunto de variables de entrada que incluye mediciones meteorológicas directas (temperatura, humedad, viento), codificación estacional mediante funciones sinusoidales y estadísticas derivadas a partir de ventanas temporales, permitiendo capturar tanto condiciones actuales como tendencias recientes.

El sistema implementa un ensamble de múltiples algoritmos de ML, incluyendo modelos basados en árboles de decisión (Random Forest, Gradient Boosting), métodos

CAPÍTULO 3 : PROPUESTA DE SOLUCIÓN

de vecinos cercanos, máquinas de soporte vectorial y redes neuronales. Las predicciones individuales se combinan mediante un esquema de votación, mejorando la robustez del sistema frente a la variabilidad de los datos.

Dado el carácter altamente desbalanceado de los eventos extremos, se incorporan técnicas de ponderación de clases y calibración de umbrales de decisión, optimizados mediante métricas orientadas a eventos raros, tales como el índice de éxito crítico (CSI). Este enfoque permite mejorar la capacidad del modelo para detectar eventos poco frecuentes sin aumentar significativamente las falsas alarmas.

De manera complementaria, el sistema incorpora modelos de regresión para la estimación de temperatura mínima y máxima, cuya salida es ajustada mediante un mecanismo de corrección de sesgo dinámico basado en el error reciente del modelo por localidad. Este ajuste se calcula a partir del error promedio observado en las últimas predicciones validadas de cada comuna, considerando una ventana máxima de 30 registros recientes y aplicándose solo cuando existen al menos 10 observaciones validadas disponibles. En términos operacionales, la corrección corresponde a una media móvil del sesgo local, definido como la diferencia entre la temperatura predicha y la temperatura real observada, la cual se resta a la predicción futura para reducir desviaciones sistemáticas. Cuando no existe suficiente historial validado para una comuna, el sistema utiliza un sesgo global por defecto estimado a partir del comportamiento agregado del modelo.

La interpretación agronómica se realiza mediante umbrales térmicos, los cuales dependen del cultivo, el órgano afectado y su estado de fenología. Esta capa permite traducir la predicción meteorológica en niveles de riesgo agrícola, diferenciando entre condiciones climáticas adversas y eventos con potencial de generar daño productivo.

El uso de un enfoque de ensamble responde a la necesidad de mejorar la robustez del sistema frente a la variabilidad de los datos meteorológicos y la naturaleza no lineal del problema. En particular, la combinación de modelos permite reducir la varianza y mejorar la capacidad de generalización en comparación con modelos individuales, lo

cual resulta especialmente relevante en contextos con alta incertidumbre y desbalance de clases.

3.4.3. Capa 3 — Pipeline multiagente y comunicación

El sistema utiliza una arquitectura basada en MAS, compuesta por tres agentes:

- **Agente Monitor:** Detecta condiciones de alerta y genera un contexto estructurado.
- **Agente Razonador:** Interpreta el riesgo mediante LLMs utilizando contexto pre-gathering, es decir, información meteorológica, predicciones, umbrales agrícolas y métricas del modelo consultadas previamente desde la base de datos antes de invocar al modelo de lenguaje.
- **Agente Comunicador:** Genera alertas adaptadas a WhatsApp.

En este trabajo, el término contexto pre-gathering se utiliza para describir una estrategia de preparación explícita de información antes de la llamada al LLM. A diferencia de un esquema RAG general, no se realiza búsqueda semántica sobre documentos externos en tiempo real, sino una consulta estructurada a las tablas del sistema, incluyendo predicciones, comunas afectadas, umbrales térmicos, historial reciente y métricas de desempeño.

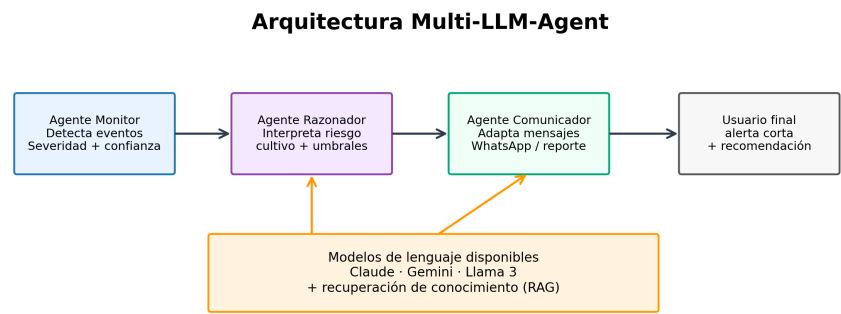
La división de responsabilidades entre agentes especializados permite abordar la complejidad del problema agroclimático de manera modular.

Mientras el Agente Monitor se encarga de recopilar y estructurar información meteorológica relevante, el Agente Razonador concentra las tareas de análisis e interpretación de riesgo utilizando conocimiento agrícola y capacidades de razonamiento basadas en modelos de lenguaje. Finalmente, el Agente Comunicador transforma los resultados obtenidos en mensajes comprensibles para el usuario final.

CAPÍTULO 3 : PROPUESTA DE SOLUCIÓN

Esta separación funcional facilita la incorporación de nuevas fuentes de información, mejora la trazabilidad de las decisiones generadas por el sistema y permite evaluar de manera independiente el desempeño de cada agente dentro de la arquitectura.

La Figura 3.2 muestra esta arquitectura.



La arquitectura separa monitoreo, razonamiento y comunicación para mejorar modularidad y trazabilidad.

Figura 3.2: Arquitectura multiagente del sistema.

Tabla 7: Responsabilidades de los agentes que componen la arquitectura Multi-LLM-Agent

Agente	Entrada	Salida
Monitor	Predicciones meteorológicas y observaciones históricas	Contexto estructurado para análisis
Razonador	Contexto meteorológico y conocimiento agrícola	Evaluación de riesgo y recomendaciones
Comunicador	Resultado del razonamiento	Mensaje final orientado al usuario

El sistema soporta múltiples proveedores de modelos de lenguaje mediante una interfaz abstracta, permitiendo comparar el desempeño de distintos LLM bajo condiciones controladas.

Esta capacidad permite evaluar distintos modelos de lenguaje bajo una arquitectura común, comparando su desempeño en tareas de interpretación agroclimática y genera-

CAPÍTULO 3 : PROPUESTA DE SOLUCIÓN

ción de recomendaciones. De esta forma, la arquitectura propuesta evita depender de un proveedor específico y facilita la incorporación de nuevos modelos a medida que evolucionan las capacidades del estado del arte.

El flujo operativo del sistema se presenta en la Figura 3.3.

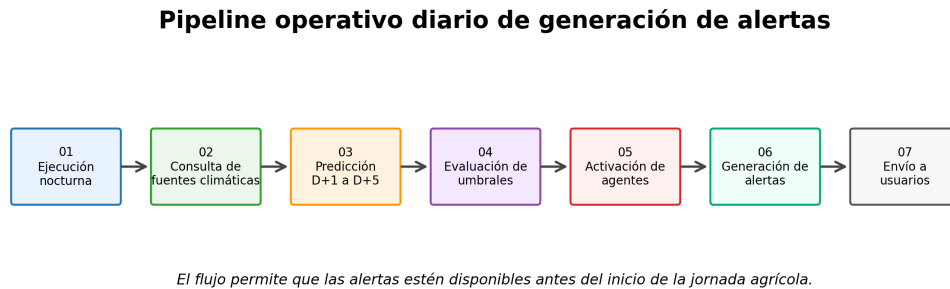


Figura 3.3: Pipeline de operación del sistema.

La adopción de una arquitectura basada en MAS permite desacoplar las distintas funciones del sistema, facilitando su modularidad, escalabilidad y trazabilidad. Este enfoque resulta particularmente adecuado en contextos donde se requiere integrar múltiples fuentes de información y procesos heterogéneos, permitiendo además la incorporación y evaluación de distintos modelos de lenguaje bajo una estructura controlada.

3.5. Pipeline de operación y validación

El sistema opera mediante un ciclo diario automatizado que incluye adquisición de datos, predicción, validación y generación de alertas.

Cada etapa se ejecuta de forma asíncrona, garantizando continuidad operativa incluso ante fallos.

El módulo de validación compara las predicciones generadas por el sistema con observaciones reales provenientes de estaciones meteorológicas, permitiendo evaluar su desempeño en condiciones operacionales.

Dado el carácter desbalanceado de los eventos térmicos extremos, se prioriza el uso de métricas orientadas a la detección de eventos raros, tales como el índice de éxito crítico (CSI), precisión, recall y F1-score. Estas métricas permiten evaluar de manera más adecuada la capacidad del sistema para identificar eventos críticos, evitando sobreestimar el desempeño mediante medidas globales como la exactitud.

De manera complementaria, se utilizan métricas de error continuo, como el RMSE, para evaluar la calidad de las predicciones de temperatura. Este enfoque combinado permite analizar tanto la capacidad predictiva del sistema como su utilidad en la generación de alertas agroclimáticas.

3.6. Ejemplo de salida del sistema

A continuación, se presenta un ejemplo de alerta generada por el sistema.

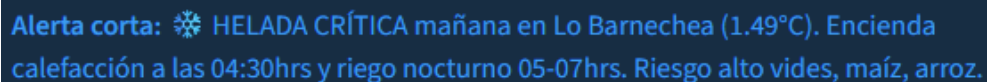
A screenshot of a system alert message. The text is white on a dark blue background. It starts with a snowflake icon, followed by 'HELADA CRÍTICA mañana en Lo Barnechea (1.49°C). Encienda calefacción a las 04:30hrs y riego nocturno 05-07hrs. Riesgo alto vides, maíz, arroz.'

Figura 3.4: Ejemplo de alerta generada por el sistema.

La alerta integra predicción, interpretación y recomendaciones en un único mensaje, facilitando la toma de decisiones.

El ejemplo presentado ilustra la integración de las distintas capas del sistema en una salida unificada, donde se combinan la predicción meteorológica, la interpretación mediante umbrales térmicos y la generación de recomendaciones específicas. Este tipo de salida permite al usuario comprender no solo la ocurrencia del evento, sino también su impacto potencial y las acciones sugeridas para mitigar sus efectos.

3.7. Ámbito de aplicación y alcance

El sistema se implementa en las 52 comunas de la Región Metropolitana de Chile.

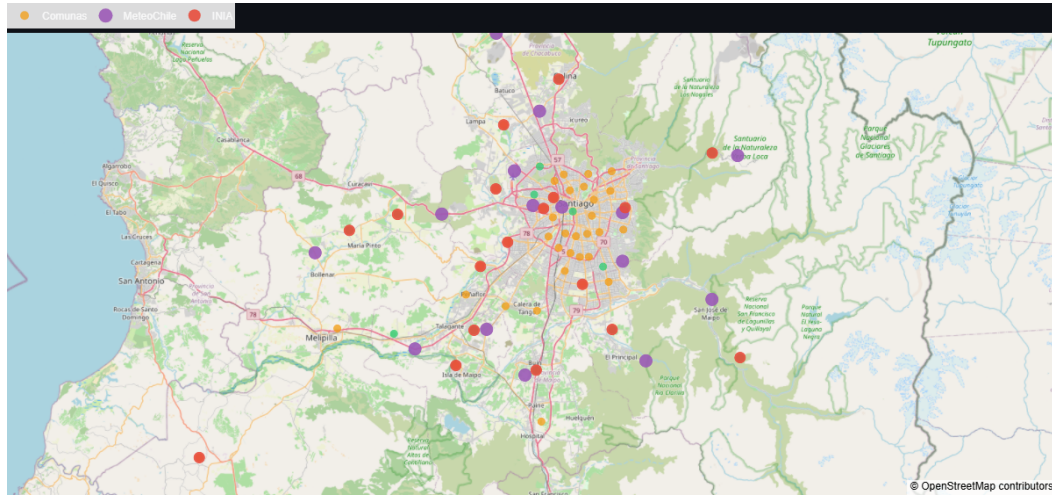


Figura 3.5: Cobertura espacial del sistema.

Esta delimitación permite validar el sistema en un entorno controlado antes de su escalamiento.

3.8. Síntesis de la propuesta

La propuesta desarrollada en este capítulo integra de manera explícita las tres dimensiones identificadas en el estado del arte: la predicción de eventos térmicos extremos, su interpretación mediante umbrales térmicos y la comunicación de alertas tempranas en formatos comprensibles y accionables.

A nivel técnico, el sistema combina modelos de aprendizaje automático basados en ensamble, estrategias de calibración para eventos raros y mecanismos de corrección de sesgo, junto con una arquitectura multiagente que desacopla las tareas de monitoreo, razonamiento e interacción.

La principal contribución de esta investigación corresponde al diseño e implementación de una arquitectura Multi-LLM-Agent orientada a agricultura inteligente. La arquitectura coordina agentes especializados capaces de monitorear condiciones meteorológicas, interpretar riesgos agroclimáticos y generar alertas contextualizadas para los usuarios finales.

CAPÍTULO 3 : PROPUESTA DE SOLUCIÓN

La solución integra modelos predictivos basados en aprendizaje automático, conocimiento agrícola representado mediante umbrales térmicos y mecanismos de comunicación automatizada, permitiendo transformar información meteorológica en recomendaciones accionables para la toma de decisiones.

Asimismo, la capacidad de incorporar distintos modelos de lenguaje bajo una estructura común proporciona un marco flexible para evaluar estrategias de razonamiento asistido por inteligencia artificial en aplicaciones agroclimáticas.

Finalmente, los componentes y decisiones de diseño presentados establecen la base para la definición de requisitos, arquitectura técnica y metodología de evaluación, los cuales serán desarrollados en los capítulos siguientes.

En este sentido, la propuesta no solo constituye una solución técnica, sino también un marco integrador que articula distintas áreas del conocimiento —meteorología, agronomía e inteligencia artificial— en un sistema orientado a la toma de decisiones. Este enfoque permite avanzar hacia sistemas agroclimáticos más efectivos, centrados en el usuario y adaptados a las condiciones reales de operación.

No obstante, la propuesta considera ciertas limitaciones y supuestos de alcance. En esta etapa, la cobertura geográfica del sistema se restringe principalmente a la Región Metropolitana, por lo que su extensión a otras zonas requeriría incorporar nuevas comunas, estaciones meteorológicas y validaciones locales. Asimismo, el funcionamiento de la capa multiagente depende de la disponibilidad de proveedores externos de LLMs, lo que introduce una dependencia operacional asociada a sus API, costos y continuidad del servicio. Finalmente, si bien el sistema genera alertas en lenguaje natural orientadas a usuarios agrícolas, la validación con usuarios reales no forma parte del alcance de esta etapa, quedando como una línea de trabajo futura para evaluar comprensión, utilidad y adopción en condiciones de uso reales.

4. Requisitos del Sistema

A partir de la problemática identificada y de la propuesta de solución presentada en el capítulo anterior, es necesario definir los requisitos que deberá satisfacer el sistema propuesto. Estos requisitos permiten establecer de forma clara las funcionalidades esperadas y las características de calidad que deberá cumplir la solución para apoyar la generación de alertas agroclimáticas basadas en eventos térmicos extremos.

Dado que la propuesta se centra en la integración de datos meteorológicos, la predicción de eventos extremos, la interpretación de su impacto agrícola y la comunicación de alertas mediante WhatsApp, los requisitos se orientan principalmente al flujo de procesamiento de información y generación de conocimiento útil para la toma de decisiones.

4.1. Requisitos Funcionales

Los requisitos funcionales describen las capacidades que deberá proporcionar el sistema para cumplir con los objetivos definidos en esta memoria.

- RF1 Adquisición de datos meteorológicos: El sistema deberá obtener información meteorológica desde fuentes externas confiables, incluyendo datos observados y pronosticados de variables relevantes para la detección de eventos térmicos extremos.

Entre las variables consideradas se incluyen:

- Temperatura mínima.
- Temperatura máxima.
- Humedad relativa.
- Velocidad del viento.
- Radiación solar.

CAPÍTULO 4 : REQUISITOS DEL SISTEMA

- Otras variables disponibles que puedan contribuir al desempeño predictivo.
- RF2 Almacenamiento y preparación de datos: El sistema deberá almacenar la información recopilada y ejecutar procesos de preparación de datos que permitan:
 - Eliminar registros inválidos.
 - Tratar valores faltantes.
 - Normalizar formatos.
 - Generar conjuntos de datos aptos para entrenamiento y predicción.
- RF3 Entrenamiento de modelos predictivos: El sistema deberá permitir el entrenamiento de modelos de aprendizaje automático utilizando datos históricos meteorológicos.

El proceso deberá considerar:

 - Selección de variables predictoras.
 - Entrenamiento del modelo.
 - Validación mediante datos históricos.
 - Almacenamiento del modelo resultante.
- RF4 Generación de predicciones: El sistema deberá generar predicciones diarias asociadas a eventos térmicos extremos, considerando tanto condiciones de helada como episodios de temperaturas elevadas.

Las predicciones deberán estimar la probabilidad de ocurrencia del evento y las variables necesarias para su posterior interpretación.
- RF5 Interpretación mediante umbrales agrícolas: El sistema deberá complementar las predicciones meteorológicas utilizando umbrales térmicos obtenidos desde literatura científica y documentación técnica agrícola.

Esta interpretación deberá permitir determinar el nivel de riesgo potencial para distintos cultivos frente a temperaturas extremas.

- RF6 Generación de alertas agroclimáticas: El sistema deberá generar alertas cuando las condiciones pronosticadas indiquen una probabilidad significativa de ocurrencia de un evento térmico extremo.

Las alertas deberán incluir:

- Tipo de evento.
 - Fecha esperada.
 - Temperatura pronosticada.
 - Nivel de riesgo asociado.
 - Recomendaciones generales cuando corresponda.
- RF7 Comunicación de alertas: El sistema deberá enviar las alertas generadas mediante la plataforma WhatsApp utilizando la API correspondiente.
- Los mensajes deberán ser presentados en lenguaje natural para facilitar su comprensión por parte de los usuarios finales.

4.2. Requisitos No Funcionales

Los requisitos no funcionales describen características de calidad necesarias para el correcto funcionamiento del sistema.

- RNF1 Disponibilidad de datos: El sistema deberá operar utilizando fuentes de datos meteorológicos accesibles de forma periódica y con cobertura suficiente para la zona de estudio.
 - RNF2 Exactitud predictiva: Los modelos desarrollados deberán ser evaluados mediante métricas apropiadas para eventos poco frecuentes, permitiendo cuantificar objetivamente su capacidad predictiva.
- Entre las métricas consideradas se incluyen:

- Precision.

- Recall.
 - F1-Score.
 - Critical Success Index (CSI).
-
- RNF3 Automatización: El proceso de adquisición de datos, generación de predicciones y emisión de alertas deberá ejecutarse de forma automática sin requerir intervención manual continua.
 - RNF4 Escalabilidad: La arquitectura propuesta deberá permitir incorporar nuevas zonas geográficas, fuentes de datos y tipos de eventos climáticos sin requerir modificaciones significativas en su estructura general.
 - RNF5 Interpretabilidad: Los resultados generados por el sistema deberán ser comprensibles para usuarios sin conocimientos especializados en meteorología o aprendizaje automático.

4.3. Interfaces del Sistema

La solución contempla tres interfaces principales asociadas al flujo operacional del sistema, las cuales son:

1. Interfaz de adquisición de datos: Corresponde a la conexión con servicios meteorológicos externos encargados de suministrar observaciones históricas y pronósticos climáticos.
2. Interfaz interna de procesamiento: Corresponde al intercambio de información entre los componentes encargados de la preparación de datos, la predicción, la interpretación mediante umbrales y la generación de alertas.
3. Interfaz de comunicación: Corresponde al mecanismo utilizado para enviar alertas a los usuarios mediante WhatsApp, actuando como canal de difusión de la información generada por el sistema.

4.4. Trazabilidad de Requisitos

La Tabla 8 presenta la relación entre los objetivos específicos definidos y los requisitos funcionales propuestos.

Tabla 8: Trazabilidad entre objetivos específicos y requisitos funcionales

Objetivo específico	Requisitos asociados
Integrar datos meteorológicos	RF1, RF2
Desarrollar modelos predictivos	RF3, RF4
Incorporar umbrales agrícolas	RF5
Generar alertas agroclimáticas	RF6
Implementar comunicación mediante WhatsApp	RF7

4.5. Supuestos del Sistema

El diseño y operación del sistema propuesto se fundamentan en una serie de supuestos que delimitan el alcance de la solución y establecen las condiciones necesarias para su correcto funcionamiento.

- Disponibilidad de fuentes meteorológicas: Se asume que las fuentes de datos meteorológicos utilizadas por el sistema se encuentran disponibles durante la operación normal de la plataforma y proporcionan información con un nivel de calidad suficiente para realizar procesos de predicción e interpretación agroclimática.
- Disponibilidad de observaciones históricas: Se asume la existencia de registros meteorológicos históricos adecuados para el entrenamiento y validación de los

CAPÍTULO 4 : REQUISITOS DEL SISTEMA

modelos predictivos utilizados por la solución.

- Disponibilidad de servicios de pronóstico: Se considera que las plataformas externas utilizadas para la obtención de pronósticos meteorológicos mantienen continuidad operacional y permiten acceder a la información requerida por el sistema.
- Validez de los umbrales agrícolas: Se asume que los umbrales térmicos utilizados para la interpretación de riesgos agroclimáticos corresponden a valores respaldados por literatura científica, documentación técnica especializada o conocimiento experto del dominio agrícola.
- Disponibilidad de servicios de mensajería: Se asume la disponibilidad de la plataforma de comunicación utilizada para el envío de alertas, permitiendo la entrega oportuna de mensajes a los usuarios finales.
- Disponibilidad de modelos de lenguaje: La arquitectura Multi-LLM-Agent propuesta asume la disponibilidad de al menos un modelo de lenguaje capaz de ejecutar las tareas de monitoreo, razonamiento o comunicación definidas dentro de la arquitectura.
- Conectividad de los usuarios finales: Se considera que los usuarios poseen acceso a dispositivos móviles y conectividad suficiente para recibir las alertas generadas por el sistema mediante los canales de comunicación definidos.

Estos supuestos permiten acotar el problema abordado por la investigación, concentrando el esfuerzo en el diseño e implementación de la arquitectura Multi-LLM-Agent y en la generación de alertas agroclimáticas inteligentes, sin depender de factores externos que escapen al control de la solución propuesta.

5. Diseño de Arquitectura del Sistema

El presente capítulo describe la arquitectura propuesta para el sistema de alertas agroclimáticas desarrollado en esta memoria. La solución tiene como objetivo integrar datos meteorológicos provenientes de distintas fuentes, generar predicciones de eventos térmicos extremos mediante técnicas de aprendizaje automático, interpretar dichas predicciones considerando umbrales agrícolas y comunicar oportunamente alertas a los usuarios a través de WhatsApp.

La arquitectura fue diseñada siguiendo un enfoque modular, permitiendo separar las distintas etapas del procesamiento de información y facilitando la escalabilidad, mantenibilidad e integración de nuevas fuentes de datos o modelos predictivos en futuras versiones del sistema.

5.1. Visión General de la Arquitectura

La solución propuesta se organiza en seis componentes principales:

1. Adquisición de datos meteorológicos.
2. Preparación y procesamiento de datos.
3. Predicción de eventos térmicos extremos.
4. Interpretación agroclimática.
5. Generación de alertas.
6. Comunicación de alertas.

El flujo general de procesamiento comienza con la recopilación automática de información meteorológica desde distintas fuentes externas. Posteriormente, los datos son procesados y transformados para ser utilizados por modelos de aprendizaje automático

CAPÍTULO 5 : DISEÑO DE ARQUITECTURA DEL SISTEMA

encargados de generar predicciones. Estas predicciones son evaluadas utilizando umbrales térmicos asociados a distintos cultivos, permitiendo estimar niveles de riesgo. A partir de esta interpretación, el sistema genera alertas estructuradas y posteriormente las comunica a los usuarios mediante un canal accesible, como WhatsApp.

La Figura 5.1 presenta una visión general de la arquitectura propuesta.



Figura 5.1: Arquitectura general del sistema de alertas agroclimáticas.

5.2. Flujo Operacional del Sistema

La Figura 5.2 muestra el flujo operacional implementado por la solución.

CAPÍTULO 5 : DISEÑO DE ARQUITECTURA DEL SISTEMA

El proceso inicia con la obtención de datos meteorológicos históricos y pronosticados. Posteriormente se realizan tareas de validación, limpieza y transformación de datos para generar conjuntos consistentes que permitan alimentar los modelos predictivos.

Las predicciones generadas son evaluadas mediante reglas basadas en umbrales térmicos agrícolas, permitiendo identificar condiciones de riesgo para distintos cultivos. Finalmente, el sistema genera alertas que son distribuidas mediante WhatsApp.

Flujo Operacional del Sistema



Figura 5.2: Flujo operacional del sistema.

5.3. Componentes del Sistema

5.3.1. Módulo de Adquisición de Datos

Este módulo es responsable de recopilar información meteorológica desde diversas fuentes externas. Su función principal consiste en centralizar datos históricos y pronosticados que posteriormente serán utilizados por los modelos predictivos.

La arquitectura contempla la integración de distintas plataformas meteorológicas con el propósito de aumentar la cobertura geográfica, disponibilidad de información y robustez del sistema frente a posibles fallas de una fuente específica.

Los datos obtenidos incluyen variables tales como temperatura mínima, temperatura máxima, humedad relativa, velocidad del viento, radiación solar y otras variables meteorológicas relevantes para la detección de eventos térmicos extremos.

5.3.2. Módulo de Preparación de Datos

Una vez recopilada la información meteorológica, ésta es sometida a procesos de validación, limpieza y transformación.

Las tareas realizadas por este módulo incluyen:

- Eliminación de registros inválidos.
- Tratamiento de valores faltantes.
- Homogeneización de formatos.
- Integración de múltiples fuentes de datos.
- Generación de variables derivadas, incluyendo estadísticas de ventana temporal y codificación estacional.

En particular, el sistema calcula variables como *rolling_max7*, correspondiente al promedio móvil de la temperatura máxima de los días previos en una ventana de siete

CAPÍTULO 5 : DISEÑO DE ARQUITECTURA DEL SISTEMA

días, y *rolling_min3*, correspondiente al mínimo móvil de la temperatura mínima en una ventana de tres días. Asimismo, se codifica el día del año mediante variables sinusoidales, *dia_sin* y *dia_cos*, para capturar patrones estacionales asociados al ciclo anual.

El objetivo de esta etapa es garantizar la calidad de los datos utilizados durante el entrenamiento y operación de los modelos predictivos.

5.3.3. Módulo de Predicción

El módulo de predicción constituye el núcleo analítico del sistema y tiene como objetivo estimar la ocurrencia de eventos térmicos extremos a partir de información meteorológica histórica y pronosticada.

Para ello se utilizan modelos de aprendizaje automático entrenados con datos recopilados desde distintas fuentes meteorológicas integradas por la plataforma.

A partir de las variables procesadas por el módulo de preparación de datos, el sistema genera predicciones asociadas a temperatura mínima, temperatura máxima y probabilidad de ocurrencia de eventos extremos para cada comuna analizada.

Estas predicciones constituyen la entrada principal del módulo de interpretación agroclimática.

5.3.4. Módulo de Interpretación Agroclimática

Este módulo complementa el proceso de predicción mediante la interpretación de los resultados generados por los modelos de aprendizaje automático.

Su objetivo es transformar predicciones meteorológicas en información útil para la toma de decisiones agrícolas, considerando umbrales térmicos de daño obtenidos desde literatura científica y documentación técnica especializada.

La implementación se basa en una arquitectura multiagente compuesta por tres agentes especializados:

CAPÍTULO 5 : DISEÑO DE ARQUITECTURA DEL SISTEMA

- Agente Monitor: Supervisa la información meteorológica y las predicciones generadas por el sistema.
- Agente Razonador: Evalúa los resultados considerando umbrales térmicos específicos y determina el nivel de riesgo asociado a cada evento.
- Agente Comunicador: Transforma dicha información en mensajes comprensibles para los usuarios finales.

La decisión de emitir una alerta se basa en reglas deterministas asociadas a los umbrales definidos por el sistema, garantizando trazabilidad y reproducibilidad en el proceso de análisis.

La Figura 5.3 presenta la arquitectura conceptual de este módulo.

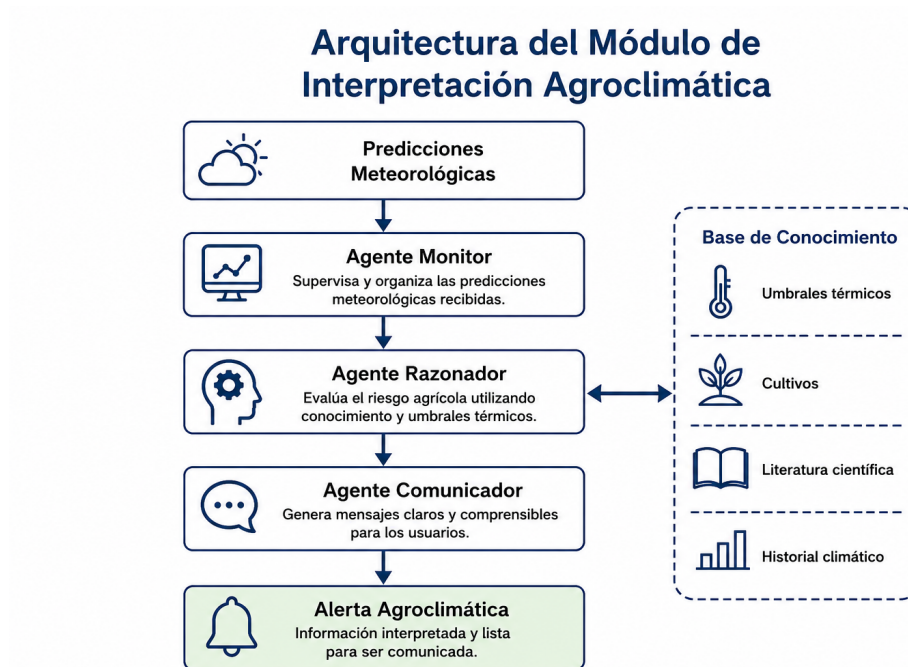


Figura 5.3: Arquitectura conceptual del módulo de interpretación agroclimática.

5.3.5. Módulo de Generación de Alertas

Este componente recibe la información procesada por el módulo de interpretación agroclimática y genera alertas estructuradas para los usuarios.

Las alertas consideran:

- Tipo de evento.
- Fecha esperada de ocurrencia.
- Temperatura pronosticada.
- Nivel de riesgo.
- Recomendaciones generales.

La generación de alertas busca proporcionar información clara, oportuna y comprensible para facilitar la toma de decisiones en contextos agrícolas.

5.3.6. Módulo de Comunicación

El módulo de comunicación es responsable de distribuir las alertas generadas por el sistema utilizando la plataforma WhatsApp.

Este componente permite entregar información directamente a los usuarios mediante mensajes en lenguaje natural, facilitando el acceso oportuno a la información desde dispositivos móviles sin requerir plataformas adicionales.

La comunicación mediante WhatsApp fue seleccionada debido a su amplia adopción y facilidad de uso, especialmente en contextos donde el acceso a herramientas tecnológicas especializadas puede ser limitado.

La arquitectura de comunicación, el proceso de envío, el registro de auditoría y el formato general de los mensajes se describen en el Anexo D.

5.4. Fuentes de Datos Meteorológicos

La calidad de las predicciones depende directamente de la calidad y cobertura de los datos utilizados. Por esta razón, el sistema integra información proveniente de diversas fuentes meteorológicas complementarias.

Las principales fuentes consideradas son:

- Red Agrometeorológica del INIA.
- Open-Meteo.
- Dirección Meteorológica de Chile.
- Atlas Agroclimático de Chile.

Cada una de estas fuentes aporta información relevante para la construcción de conjuntos de datos históricos y pronosticados utilizados por los modelos predictivos.

La Tabla 9 resume las fuentes utilizadas.

CAPÍTULO 5 : DISEÑO DE ARQUITECTURA DEL SISTEMA

Tabla 9: Fuentes de datos meteorológicos utilizadas por el sistema.

Fuente	Tipo de datos	Período / resolución	Uso principal	Referencia
INIA	Observaciones agrometeorológicas históricas y recientes.	Resolución diaria y horaria según disponibilidad de estación.	Entrenamiento, validación y contraste con observaciones reales.	Ministerio de Agricultura de Chile, 2024
Open-Meteo	Reanálisis histórico y pronósticos meteorológicos.	Datos históricos desde 1940; resolución horaria, con grillas ERA5, ERA5-Land y modelos recientes de mayor resolución.	Cobertura espacial, respaldo histórico y predicciones operacionales por comuna.	Open-Meteo, 2026a
MeteoChile	Observaciones y pronósticos meteorológicos oficiales.	Resolución diaria según disponibilidad pública por localidad.	Complementación y comparación externa de predicciones.	Dirección Meteorológica de Chile, 2026
Atlas Agroclimático	Información agroclimática histórica territorial.	Series agroclimáticas históricas agregadas por zona o localidad.	Apoyo al entrenamiento inicial y caracterización agroclimática.	Universidad de Chile, 2026

5.5. Modelo Lógico de Datos

El sistema utiliza un modelo de datos orientado a almacenar observaciones meteorológicas, predicciones y alertas generadas.

Las principales entidades consideradas son:

- Comunas.
- Estaciones meteorológicas.
- Observaciones meteorológicas.
- Predicciones.
- Umbrales térmicos.

- Alertas.
- Agricultores o usuarios suscritos.

La Figura 5.4 presenta una representación simplificada del modelo lógico utilizado.

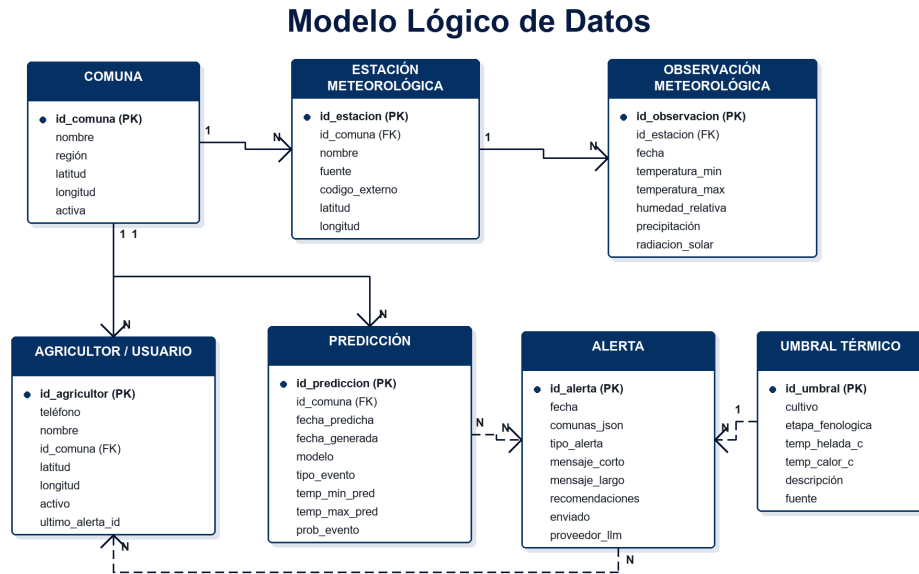


Figura 5.4: Modelo lógico simplificado del sistema.

5.6. Infraestructura Tecnológica

La solución fue implementada mediante una arquitectura basada en servicios distribuidos, organizada en componentes especializados para procesamiento, almacenamiento, visualización, automatización y comunicación. La Tabla 10 resume las principales tecnologías utilizadas y el rol que cumplen dentro del sistema.

CAPÍTULO 5 : DISEÑO DE ARQUITECTURA DEL SISTEMA

Tabla 10: Componentes tecnológicos utilizados por la plataforma.

Componente	Tecnología	Justificación
Backend	Python / FastAPI	Permite implementar servicios REST para la consulta de predicciones, métricas, alertas y ejecución del pipeline. FastAPI fue seleccionado por su soporte nativo para aplicaciones asincrónicas, documentación automática de endpoints y facilidad de integración con servicios externos.
Base de datos	MySQL	Almacena observaciones meteorológicas, predicciones, métricas, umbrales, alertas y usuarios suscritos. Se utiliza una base relacional debido a la estructura tabular de los datos y a la necesidad de mantener relaciones entre comunas, estaciones, observaciones y predicciones.
Motor predictivo	Python / Scikit-learn / XGBoost / LightGBM / CatBoost	Permite entrenar y ejecutar modelos de aprendizaje automático para la predicción de eventos térmicos extremos y temperatura mínima/máxima. Estas librerías fueron seleccionadas por su madurez, disponibilidad de modelos supervisados y buen desempeño en datos tabulares.
Comunicación	WhatsApp / API de mensajería	Permite entregar alertas en lenguaje natural a usuarios finales mediante un canal ampliamente utilizado. La comunicación se implementa mediante una interfaz programática que permite enviar mensajes y gestionar interacciones básicas con los usuarios.
Automatización	APScheduler	Ejecuta tareas programadas para scraping, validación, predicción, entrenamiento y generación de alertas. Se utiliza para mantener una operación diaria autónoma sin intervención manual continua.
Despliegue	Docker / Docker Compose	Permite ejecutar los servicios de forma reproducible y desacoplada, incluyendo backend, base de datos y componentes auxiliares. Esto facilita la portabilidad entre ambientes de desarrollo y ejecución.

CAPÍTULO 5 : DISEÑO DE ARQUITECTURA DEL SISTEMA

La elección de estas tecnologías responde a la necesidad de construir una plataforma modular y reproducible. FastAPI centraliza la lógica del sistema y expone los servicios principales; MySQL permite mantener la persistencia estructurada de los datos; el motor predictivo en Python facilita la integración con bibliotecas de aprendizaje automático; y APScheduler permite automatizar el ciclo operacional nocturno del sistema. Finalmente, Docker Compose permite desplegar los componentes como servicios independientes, mientras que WhatsApp actúa como canal de comunicación hacia los usuarios finales.

Los detalles de implementación, despliegue, configuración de infraestructura y componentes tecnológicos específicos se presentan en el Anexo A.

6. Metodología

En este capítulo se describe la metodología utilizada para diseñar, construir, operar y evaluar el sistema propuesto. La estrategia metodológica integra dos líneas de trabajo complementarias. La primera corresponde al desarrollo de un sistema predictivo para la detección de eventos térmicos extremos a escala comunal. La segunda corresponde a la implementación y evaluación de una arquitectura Multi-LLM-Agent, orientada a interpretar las predicciones generadas y transformarlas en alertas agroclimáticas contextualizadas para usuarios del sector agrícola.

La metodología no se limita a la construcción de modelos de aprendizaje automático, sino que considera un flujo completo de adquisición de datos, preparación de series temporales, entrenamiento, inferencia, validación, interpretación de riesgo y comunicación de resultados. Esta integración permite evaluar tanto la calidad de las predicciones meteorológicas como el aporte de la arquitectura Multi-LLM-Agent en tareas de razonamiento, recomendación y comunicación dentro del dominio de la agricultura inteligente.

Esquema general de la metodología propuesta



Figura 6.1: Esquema general de la metodología propuesta.

La Figura 6.1 resume el flujo metodológico seguido en la investigación, desde la integración de fuentes meteorológicas hasta la generación y evaluación de alertas agroclimáticas mediante la arquitectura Multi-LLM-Agent.

6.1. Enfoque Metodológico y Diseño Experimental

La investigación corresponde a un estudio aplicado con enfoque mixto de predominio cuantitativo. Su propósito es diseñar e implementar un artefacto tecnológico funcional que permita detectar eventos térmicos extremos y generar alertas agroclimáticas orientadas a la toma de decisiones.

El componente cuantitativo se concentra en la evaluación del sistema predictivo propuesto. Para ello, se construyen modelos de clasificación y regresión a partir de datos meteorológicos históricos y observaciones recientes, evaluando su desempeño mediante métricas de verificación adecuadas para eventos poco frecuentes. Los resultados se comparan con fuentes externas de referencia, específicamente Open-Meteo y MeteoChile.

El componente cualitativo y descriptivo se concentra en la evaluación de la arquitectura Multi-LLM-Agent. Esta arquitectura utiliza modelos de lenguaje para interpretar el contexto meteorológico y agrícola asociado a una alerta, generar recomendaciones y producir mensajes comprensibles para usuarios finales. La comparación se realiza bajo condiciones controladas, proporcionando a cada proveedor de modelo de lenguaje exactamente el mismo contexto de entrada, instrucciones y esquema de salida esperado.

El diseño experimental se organiza en dos niveles independientes:

1. Evaluación del sistema predictivo: se analiza la capacidad del sistema para detectar eventos de helada y ola de calor, así como para estimar temperaturas mínimas y máximas.
2. Evaluación de la arquitectura Multi-LLM-Agent: se compara el comportamiento de distintos modelos de lenguaje en tareas de interpretación agroclimática, reco-

mendación y comunicación de alertas.



Figura 6.2: Diseño experimental de dos niveles utilizado para evaluar el sistema predictivo y la arquitectura Multi-LLM-Agent.

La Figura 6.2 resume esta separación metodológica, permitiendo diferenciar la evaluación del desempeño predictivo de la evaluación de la interpretación, recomendación y comunicación generada por los modelos de lenguaje.

6.2. Área de Estudio y Alcance Territorial

El sistema opera sobre las 52 comunas activas de la Región Metropolitana de Chile. Cada comuna se considera una unidad de predicción independiente, con sus propias series de observaciones, modelos entrenados, predicciones almacenadas y métricas de evaluación.

La elección de la Región Metropolitana responde a su diversidad territorial y climática. La zona incorpora sectores urbanos, rurales, precordilleranos y agrícolas, con condiciones térmicas heterogéneas que pueden afectar de manera distinta la ocurrencia de heladas y olas de calor. Esta variabilidad hace necesario considerar cada comuna

CAPÍTULO 6 : METODOLOGÍA

como una unidad territorial diferenciada, evitando suponer que una misma condición meteorológica afecta de forma equivalente a toda la región.

La asignación de estaciones meteorológicas a cada comuna se realiza mediante la fórmula de Haversine, utilizada para calcular la distancia geodésica entre la ubicación de una estación y el centroide de la comuna correspondiente. Cuando una comuna no dispone de estaciones propias con información suficiente, el sistema amplía la búsqueda a estaciones de la Red AGROMET del Instituto de Investigaciones Agropecuarias ubicadas en un radio máximo de 40 km.

Esta estrategia permite ampliar la cobertura territorial del sistema sin requerir infraestructura de medición específica en cada comuna. Sin embargo, la proximidad espacial se considera una aproximación operativa y no implica que las estaciones cercanas representen de forma idéntica las condiciones microclimáticas de cada territorio.

La Figura 6.3 presenta la distribución territorial de las comunas consideradas, las estaciones meteorológicas disponibles y las conexiones espaciales utilizadas durante la asignación de datos. Estas conexiones representan la relación de proximidad empleada por el sistema para asociar una comuna con estaciones cercanas, especialmente cuando no dispone de una estación propia con cobertura suficiente.

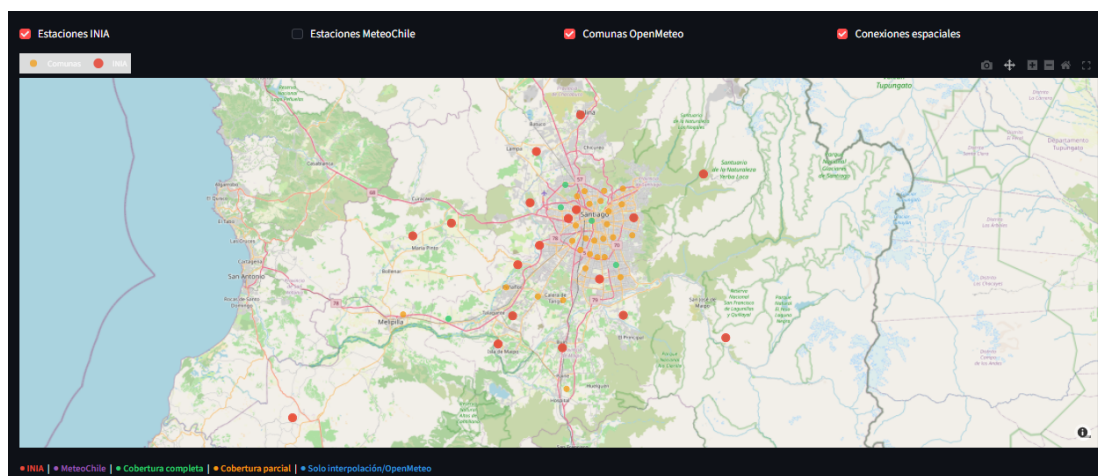


Figura 6.3: Asignación espacial entre comunas y estaciones meteorológicas mediante proximidad geográfica. Fuente: elaboración propia a partir del sistema desarrollado.

6.3. Fuentes de Datos

El sistema integra fuentes de datos con distintos períodos de cobertura, niveles de prioridad y propósitos metodológicos. Esta combinación permite disponer de información histórica para el entrenamiento, observaciones recientes para la validación y pronósticos operacionales para la generación de alertas.

Tabla 11: Fuentes de datos utilizadas por el sistema y su rol metodológico.

Fuente	Cobertura temporal	Información principal	Rol metodológico
Atlas Agroclimático de Chile	2004–2016	Temperatura mínima y máxima, humedad relativa, precipitación y radiación solar.	Base histórica de largo plazo para el entrenamiento inicial.
Red AGROMET del INIA	Observaciones recientes	Mediciones directas de estaciones agrometeorológicas.	Fuente de mayor prioridad para validación, actualización y reentrenamiento.
Open-Meteo histórico	2019–presente	Reanálisis meteorológico diario para todas las comunas.	Complemento temporal y espacial de las observaciones disponibles.
Open-Meteo Forecast	D+1 a D+5	Pronóstico de temperatura, humedad, precipitación y radiación solar.	Variables auxiliares para predicción multi-horizonte y referencia externa.
MeteoChile	Pronóstico diario	Predicción meteorológica oficial.	Referencia operacional para comparar el sistema predictivo propuesto.

6.3.1. Atlas Agroclimático de Chile

La base histórica utilizada por el sistema corresponde a un conjunto de datos procesado localmente a partir de información del Repositorio Climático del Atlas de Riesgos Climáticos de Chile (ARClím) Ministerio del Medio Ambiente de Chile, 2026. El ar-

chivo utilizado en la implementación, denominado `datos_atlas_limpios_completos.csv`, contiene registros diarios preparados previamente para su incorporación al sistema.

Durante la etapa de importación, el conjunto de datos se filtra para conservar únicamente las comunas pertenecientes a la Región Metropolitana, identificadas mediante códigos territoriales asociados a la región. Además, se restringe el período temporal comprendido entre los años 2004 y 2016.

Esta fuente se almacena en la tabla de observaciones con la etiqueta **ATLAS** y constituye la capa histórica base utilizada durante el entrenamiento inicial del sistema predictivo. Sus variables incluyen temperatura mínima diaria, temperatura máxima diaria, precipitación, radiación solar y variables de humedad disponibles en el conjunto procesado.

6.3.2. Red AGROMET del Instituto de Investigaciones Agropecuarias

Las observaciones recientes se obtienen desde la plataforma Agrometeorología, disponible en Instituto de Investigaciones Agropecuarias, 2026. Esta plataforma concentra información proveniente de estaciones agrometeorológicas de distintas instituciones; sin embargo, para esta investigación se utilizan específicamente las estaciones identificadas como pertenecientes al Instituto de Investigaciones Agropecuarias.

La incorporación de estos datos se realiza mediante procesos automatizados de extracción diaria y horaria. El proceso diario descarga registros consolidados de estaciones, mientras que el proceso horario permite complementar la información térmica reciente cuando resulta necesario para la operación del sistema.

Las observaciones obtenidas desde estas estaciones se almacenan con la etiqueta **INIA** y representan la fuente de mayor prioridad dentro de la estrategia de fusión de datos. Su función principal es aportar mediciones recientes de terreno para validar las predicciones emitidas, actualizar las series meteorológicas y alimentar los procesos periódicos de reentrenamiento.

6.3.3. Open-Meteo

Open-Meteo se utiliza mediante dos servicios complementarios: la API histórica y la API de pronóstico meteorológico Open-Meteo, 2026a, 2026b.

La API histórica se consulta mediante el servicio archive-api.open-meteo.com/v1/archive. Esta fuente permite recuperar datos meteorológicos previos para las comunas de estudio y se utiliza como complemento temporal y espacial entre la base histórica proveniente de ARClím y las observaciones recientes obtenidas desde estaciones INIA. En el sistema, estos registros se almacenan bajo una etiqueta asociada a Open-Meteo.

La API de pronóstico se consulta mediante el servicio api.open-meteo.com/v1/forecast. Esta fuente entrega variables meteorológicas proyectadas, tales como temperatura mínima, temperatura máxima, humedad relativa, precipitación y radiación solar, para los horizontes futuros considerados por el sistema.

El pronóstico de Open-Meteo cumple dos funciones. Primero, aporta variables auxiliares para construir predicciones en los horizontes D+2 a D+5. Segundo, actúa como fuente de referencia externa para contrastar el desempeño del sistema predictivo propuesto.

6.3.4. Dirección Meteorológica de Chile

Las predicciones meteorológicas de referencia se obtienen desde el portal de Servicios Climáticos de la Dirección Meteorológica de Chile Dirección Meteorológica de Chile, 2026. La incorporación de esta información se realiza mediante un proceso automatizado de extracción desde el portal institucional.

Los registros obtenidos se almacenan con la etiqueta METEOCHILE. Esta fuente no interviene en el entrenamiento de los modelos ni en la generación interna de predicciones. Su función metodológica es actuar como una referencia operacional externa para comparar el desempeño del sistema predictivo propuesto frente a un producto meteorológico institucional.

La comparación con MeteoChile permite evaluar el comportamiento del sistema respecto de una fuente disponible públicamente para usuarios y organizaciones que requieren información meteorológica en la Región Metropolitana.

El detalle de las variables disponibles por fuente, su cobertura temporal y su rol dentro de la serie consolidada se presenta en el Anexo F. La política de fusión jerárquica y la trazabilidad de los registros se describen específicamente en el Anexo F.

6.4. Gestión y Trazabilidad de la Información

La información utilizada por el sistema se almacena en una base de datos relacional implementada en MySQL. El objetivo de esta capa no es únicamente persistir datos, sino mantener trazabilidad entre las observaciones de entrada, los modelos utilizados, las predicciones generadas, las alertas emitidas y las métricas obtenidas posteriormente.

Las entidades centrales almacenan información territorial, estaciones meteorológicas, observaciones diarias, predicciones, métricas, alertas, umbrales agrícolas y registros de ejecución del pipeline. Esta organización permite reconstruir el ciclo de vida de una predicción, desde la fuente meteorológica utilizada hasta la alerta entregada al usuario.

La trazabilidad resulta especialmente relevante en este trabajo, ya que permite diferenciar entre errores originados en la calidad de los datos, en el sistema predictivo o en la etapa de interpretación realizada por la arquitectura Multi-LLM-Agent. El detalle del modelo físico de datos y sus relaciones se presenta en los anexos correspondientes.

El modelo físico de datos, las entidades involucradas y sus relaciones se presentan en el Anexo B.

6.5. Preparación de Datos y Construcción de Características

6.5.1. Fusión jerárquica de fuentes

Las fuentes meteorológicas disponibles se combinan bajo una política de prioridad jerárquica:

$$\text{INIA} > \text{Open-Meteo histórico} > \text{Atlas Agroclimático} \quad (1)$$

Esta política busca privilegiar observaciones directas de estaciones meteorológicas cuando están disponibles. El procedimiento comienza utilizando el Atlas Agroclimático como base histórica de largo plazo. Posteriormente, las fechas coincidentes son actualizadas con datos históricos de Open-Meteo. Finalmente, cuando existen observaciones de la Red AGROMET, estas reemplazan a las fuentes anteriores.

La Figura 6.4 ilustra el procedimiento de integración aplicado a las fuentes meteorológicas. El proceso utiliza ARClím como base histórica, actualiza los registros coincidentes con datos históricos de Open-Meteo y finalmente prioriza las observaciones directas obtenidas desde estaciones INIA.

Fusión jerárquica de fuentes meteorológicas

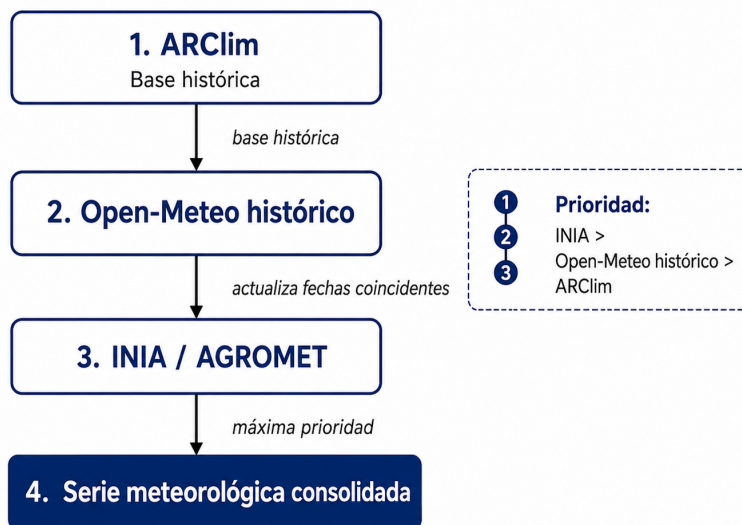


Figura 6.4: Proceso de fusión jerárquica de fuentes meteorológicas para la construcción de series consolidadas por comuna.

La fusión no busca generar una serie artificialmente uniforme. Cuando no existe información suficiente para una fecha o variable crítica, la observación se excluye del conjunto de entrenamiento. De este modo, se evita introducir valores sintéticos en variables que podrían afectar la calidad de los modelos predictivos.

El catálogo comparativo de variables por fuente y el detalle complementario de la política de prioridad se presentan en el Anexo F.

6.5.2. Tratamiento de valores faltantes

Las fuentes de datos meteorológicos pueden contener registros incompletos debido a interrupciones de estaciones, ausencia de mediciones o diferencias en la cobertura temporal. Por esta razón, se aplica un tratamiento diferenciado según el tipo de variable.

Los valores faltantes de precipitación se imputan con cero como una decisión operacional adoptada para mantener la continuidad de las series utilizadas por el sistema.

La radiación solar se trata mediante interpolación lineal entre observaciones dis-

ponibles. Para los extremos de las series temporales se aplica relleno hacia adelante o hacia atrás, según corresponda.

Las variables consideradas críticas para el entrenamiento, tales como temperatura mínima, temperatura máxima y humedad relativa, no se imputan de forma artificial. Las observaciones incompletas en estas variables son excluidas del conjunto de entrenamiento para evitar introducir patrones no observados.

6.5.3. Variables de entrada

A partir de las series meteorológicas fusionadas se construye un vector de nueve características para cada observación diaria. Estas variables incluyen información meteorológica directa, componentes de estacionalidad y estadísticas móviles de temperatura.

Tabla 12: Variables de entrada utilizadas por el sistema predictivo propuesto.

Variable	Descripción	Origen
temp_min	Temperatura mínima diaria registrada en grados Celsius.	Observada
temp_max	Temperatura máxima diaria registrada en grados Celsius.	Observada
humedad_rel	Humedad relativa diaria expresada como porcentaje.	Observada
precipitacion	Precipitación acumulada diaria en milímetros.	Observada
radiacion_solar	Radiación solar global diaria.	Observada
dia_sin	Componente seno de la posición anual del día.	Derivada
dia_cos	Componente coseno de la posición anual del día.	Derivada
rolling_max7	Promedio móvil de siete días de temperatura máxima.	Derivada
rolling_min3	Mínimo móvil de tres días de temperatura mínima.	Derivada

Las variables `dia_sin` y `dia_cos` representan la estacionalidad anual mediante

una codificación cíclica:

$$\text{dia_sin} = \sin\left(2\pi \cdot \frac{\text{DOY}}{365,25}\right) \quad (2)$$

$$\text{dia_cos} = \cos\left(2\pi \cdot \frac{\text{DOY}}{365,25}\right) \quad (3)$$

donde *DOY* representa el día del año. Esta transformación permite modelar la continuidad estacional entre diciembre y enero, evitando una discontinuidad artificial entre los extremos del calendario.

Las variables de ventana deslizante permiten sintetizar la tendencia térmica reciente. Para prevenir fuga de información temporal, estas características se calculan utilizando solamente datos disponibles hasta el día anterior a la predicción. De esta forma, se evita la fuga de información temporal (data leakage), es decir, el uso de datos futuros durante el entrenamiento o la inferencia, práctica que puede generar estimaciones de desempeño artificialmente optimistas Kapoor y Narayanan, 2023.

6.6. Definición de Eventos Térmicos

El sistema aborda dos condiciones térmicas de riesgo relevantes para la agricultura: helada y ola de calor. Ambas se modelan como tareas de clasificación binaria independientes, orientadas a anticipar condiciones meteorológicas que podrían afectar a cultivos sensibles.

Para fines de detección temprana, se definen los siguientes umbrales operacionales:

$$\text{HELADA : } T_{\text{mín},t+1} < 3,0,^{\circ}\text{C} \quad (4)$$

$$\text{OLA CALOR : } T_{\text{máx},t+1} \geq 30,0,^{\circ}\text{C} \quad (5)$$

donde $T_{\min,t+1}$ y $T_{\max,t+1}$ corresponden, respectivamente, a la temperatura mínima y máxima estimadas para el día siguiente.

El evento denominado helada se activa cuando la temperatura mínima prevista es inferior a 3,0°C. Este valor se utiliza como un umbral operacional de alerta temprana ante condiciones potencialmente riesgosas para cultivos sensibles a bajas temperaturas, sin asumir que representa una temperatura crítica universal de daño para todas las especies agrícolas Food and Agriculture Organization of the United Nations, 2005; Fundación para la Innovación Agraria, 2023b.

De manera análoga, el evento ola de calor se activa cuando la temperatura máxima prevista alcanza o supera los 30,0°C. Este umbral permite identificar escenarios de posible estrés térmico y se utiliza como criterio inicial de detección, considerando que la susceptibilidad efectiva depende del cultivo, la etapa fenológica y las condiciones locales Kumar et al., 2025; Richetti et al., 2025.

Posteriormente, la arquitectura Multi-LLM-Agent complementa estos criterios generales mediante umbrales específicos por cultivo y etapa fenológica. De este modo, el sistema predictivo identifica una condición meteorológica de riesgo a escala comunal, mientras que la arquitectura multiagente contextualiza su posible impacto sobre los cultivos afectados.

6.7. Entrenamiento del Sistema Predictivo Propuesto

6.7.1. Partición temporal de los datos

Para cada comuna y tipo de evento, el conjunto de datos se divide cronológicamente en un 80 % destinado al entrenamiento y un 20 % reservado para la evaluación final. La partición temporal evita que observaciones futuras sean utilizadas durante el ajuste de los modelos.

El conjunto de entrenamiento se subdivide también de forma cronológica en dos partes. El 75 % se utiliza para ajustar los modelos base, mientras que el 25 % restante

se emplea para calibrar los umbrales de decisión. En términos del conjunto completo, esta división corresponde a un 60 % para ajuste, un 20 % para calibración interna y un 20 % para evaluación final.

La Figura 6.5 resume la partición cronológica utilizada para el ajuste, la calibración de umbrales y la evaluación final del sistema predictivo. La organización temporal de los conjuntos evita utilizar observaciones futuras durante el entrenamiento, reduciendo el riesgo de fuga de información y de estimaciones de desempeño artificialmente optimistas Kapoor y Narayanan, 2023.

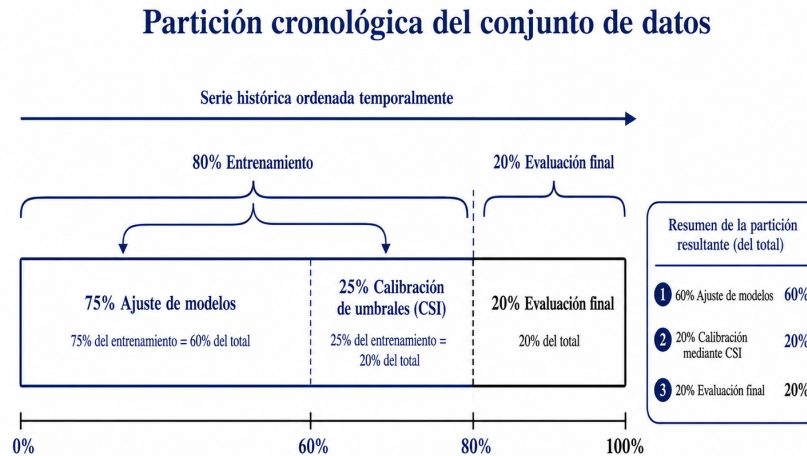


Figura 6.5: Partición cronológica utilizada para el entrenamiento, calibración y evaluación del sistema predictivo.

6.7.2. Normalización de variables

Las variables de entrada presentan escalas heterogéneas. Mientras la temperatura se expresa en grados Celsius, la precipitación se mide en milímetros y la humedad relativa se representa como porcentaje. Por esta razón, los modelos sensibles a escala son entrenados sobre variables normalizadas.

La normalización se aplica mediante estandarización a media cero y desviación estándar unitaria. Este procedimiento se utiliza para K-Nearest Neighbors, Máquinas de Vectores de Soporte y Perceptrón Multicapa.

Los modelos basados en árboles de decisión no requieren normalización, debido a que su mecanismo de partición no depende de la escala original de las variables. Los parámetros de normalización se ajustan exclusivamente sobre el conjunto de entrenamiento y se aplican posteriormente de manera consistente en validación, prueba e inferencia.

6.7.3. Ensamble heterogéneo de clasificadores

El sistema predictivo propuesto utiliza un ensamble heterogéneo de siete clasificadores. Los modelos son entrenados de manera independiente para cada combinación de comuna y evento térmico.

Los clasificadores considerados corresponden a XGBoost, Random Forest, LightGBM, CatBoost, K-Nearest Neighbors, Máquina de Vectores de Soporte con kernel radial y Perceptrón Multicapa.

La utilización de modelos heterogéneos responde a que cada familia de algoritmos posee distintos sesgos inductivos y estrategias para representar relaciones entre variables. Al combinar sus resultados, se busca reducir la dependencia de un único enfoque de aprendizaje y obtener una predicción final más robusta frente a variaciones en los datos.

Los hiperparámetros específicos, las versiones de bibliotecas y los artefactos utilizados para el entrenamiento se presentan en el Anexo E. El código fuente, los scripts de reproducción y los archivos de configuración no sensibles se encuentran disponibles en el repositorio técnico del proyecto Ahumada Avendaño, 2026.

6.7.4. Manejo del desbalance de clases

Los eventos de helada y ola de calor presentan una frecuencia menor respecto de los días sin eventos extremos. Esta condición genera desbalance de clases, lo que puede inducir a los modelos a predecir mayoritariamente la clase normal Saito y Rehmsmeier,

2015.

Para mitigar este problema, se utilizan mecanismos de ponderación de clases en aquellos modelos que lo permiten. Las ponderaciones se calculan considerando la proporción entre eventos positivos y negativos dentro del conjunto de entrenamiento de cada comuna.

Este enfoque permite aumentar la sensibilidad del sistema frente a eventos poco frecuentes, evitando que una alta cantidad de días normales produzca una falsa impresión de buen desempeño global.

6.7.5. Calibración de umbrales mediante CSI

Cada clasificador genera una probabilidad estimada de ocurrencia del evento. En lugar de utilizar el umbral convencional de 0,5, se selecciona un umbral específico para cada modelo a partir de la maximización del *Critical Success Index* sobre el conjunto de validación Ferro y Stephenson, 2011.

$$CSI = \frac{TP}{TP + FP + FN} \quad (6)$$

donde TP representa verdaderos positivos, FP falsas alarmas y FN eventos reales no detectados.

El umbral se busca dentro del intervalo $[0,05; 0,95]$, utilizando incrementos de 0,01. Esta estrategia permite ajustar la sensibilidad de los modelos a la naturaleza poco frecuente de los eventos térmicos extremos, penalizando tanto las falsas alarmas como la omisión de eventos reales.

6.7.6. Votación mayoritaria

La predicción final del ensamble se obtiene mediante votación mayoritaria o *hard voting*. Cada clasificador emite un voto positivo cuando su probabilidad estimada supera el umbral calibrado individualmente.

Cuando al menos cuatro de los siete clasificadores emiten un voto positivo, el sistema predice la ocurrencia del evento. El número de votos favorables y su proporción se almacenan junto a cada predicción, permitiendo disponer de un indicador de consenso interno entre los modelos del ensamble.

6.7.7. Regresión de temperatura

Además de los clasificadores de eventos, el sistema entrena regresores orientados a estimar la temperatura mínima y máxima del día siguiente. Estos modelos utilizan el mismo vector de características empleado por los clasificadores.

Las estimaciones numéricas de temperatura complementan la predicción binaria de eventos. Mientras el ensamble permite identificar la posible ocurrencia de helada u ola de calor, los regresores aportan valores concretos de temperatura que posteriormente son utilizados por el Agente Monitor, el Agente Razonador y el Agente Comunicador.

6.8. Corrección Adaptativa de Sesgo

Durante la operación del sistema se monitorean las diferencias entre las temperaturas estimadas y las temperaturas observadas posteriormente. Este seguimiento permite identificar la existencia de sesgos sistemáticos en las predicciones de temperatura mínima y máxima.

La corrección adaptativa se calcula utilizando el error promedio de las últimas 30 predicciones validadas de una comuna:

$$\varepsilon = T_{\text{pred}} - T_{\text{obs}} \quad (7)$$

Cuando existen al menos diez observaciones validadas para la comuna, se utiliza el sesgo local promedio. Cuando no existe suficiente historial local, se utiliza un sesgo global calculado a partir de las predicciones disponibles en el sistema.

Para evitar ajustes excesivos ante condiciones atípicas, el valor de corrección se aco-

ta dentro de un intervalo definido durante la configuración del sistema. La temperatura corregida se obtiene restando el sesgo estimado a la predicción original.

Adicionalmente, se aplica una regla de consistencia entre la estimación numérica y la clasificación del evento. Cuando la temperatura mínima corregida se encuentra bajo el umbral de helada, pero el ensamble clasificador indica una condición normal, la predicción se ajusta hacia la categoría de helada. Esta regla busca evitar que una temperatura estimada bajo umbral quede sin la alerta correspondiente.

6.9. Predicción Multi-Horizonte

El sistema genera predicciones para un horizonte de cinco días. El horizonte D+1 corresponde al día siguiente a la ejecución del pipeline, mientras que D+2 a D+5 representan los días posteriores.

Para D+1, el sistema predictivo utiliza observaciones reales recientes y variables derivadas disponibles al momento de la inferencia. El ensamble de clasificadores identifica la posible ocurrencia de eventos térmicos extremos y los regresores estiman temperaturas mínimas y máximas.

Para los horizontes D+2 a D+5, se utiliza una estrategia híbrida. Las variables temporales se actualizan mediante una cascada iterativa, incorporando las predicciones generadas en pasos anteriores. Sin embargo, las temperaturas futuras y variables meteorológicas auxiliares se obtienen desde el pronóstico numérico de Open-Meteo.

Esta decisión busca evitar la acumulación progresiva de error que puede ocurrir cuando un modelo autoregresivo utiliza sus propias salidas como entrada durante varios días consecutivos. De este modo, el sistema predictivo conserva su capacidad de clasificación contextual, mientras que las variables meteorológicas futuras se apoyan en un pronóstico numérico especializado.

La Figura 6.6 resume la estrategia utilizada para generar predicciones en los distintos horizontes temporales. Mientras el horizonte D+1 se construye directamente me-

diante el sistema predictivo, los horizontes D+2 a D+5 combinan la actualización iterativa de variables temporales con las variables meteorológicas futuras obtenidas desde Open-Meteo Open-Meteo, 2026b.

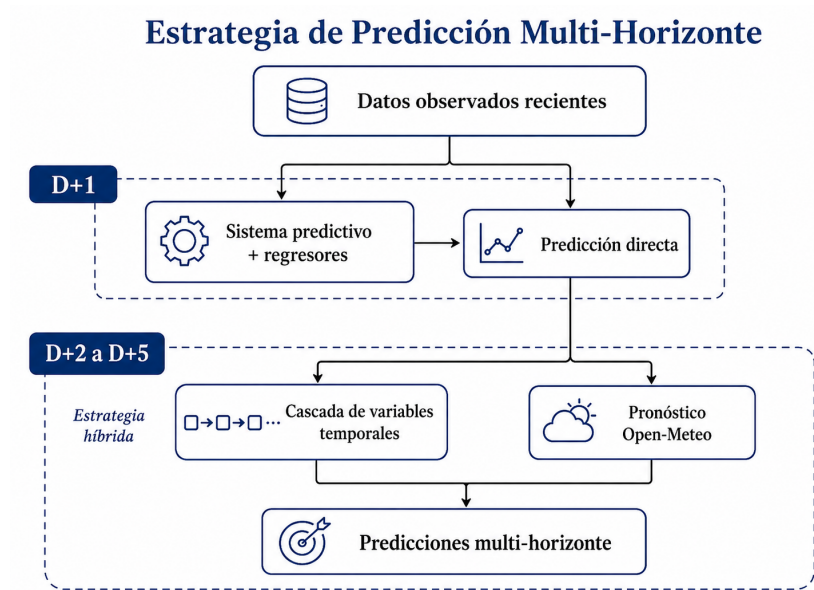


Figura 6.6: Estrategia híbrida de predicción para los horizontes D+1 a D+5.

Para los horizontes D+2 a D+5 se aplica también una regla de consistencia inversa. Cuando la temperatura mínima pronosticada por Open-Meteo supera claramente el umbral de helada, una clasificación positiva inducida únicamente por tendencias térmicas previas puede ser revertida a condición normal. Esta medida reduce la probabilidad de falsas alarmas durante períodos de aumento térmico posterior a jornadas frías.

6.10. Protocolo de la Arquitectura Multi-LLM-Agent

La arquitectura Multi-LLM-Agent se activa una vez que el sistema predictivo identifica una condición potencialmente riesgosa. Su función es transformar predicciones numéricas y reglas meteorológicas en alertas agroclimáticas interpretadas para usuarios agrícolas.

La arquitectura está compuesta por tres agentes funcionales: Monitor, Razonador y Comunicador. El diseño arquitectónico detallado de estos componentes se presenta en el capítulo anterior. En esta sección se describe el protocolo utilizado para operar y evaluar la arquitectura.

La Figura 6.7 presenta el protocolo de operación utilizado para ejecutar la arquitectura Multi-LLM-Agent. A diferencia del diagrama arquitectónico presentado en el capítulo anterior, la figura enfatiza las decisiones metodológicas que permiten mantener trazabilidad y comparabilidad entre ejecuciones: activación determinista de alertas, recopilación previa del contexto y generación estructurada del análisis.

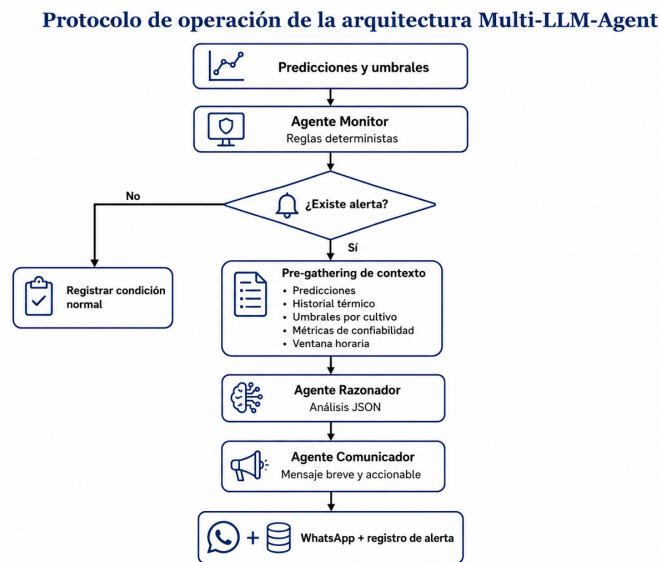


Figura 6.7: Protocolo de operación de la arquitectura Multi-LLM-Agent para la generación de alertas agroclimáticas.

6.10.1. Agente Monitor

El Agente Monitor constituye el punto de entrada del pipeline multiagente. Su función es determinar si existe una condición de alerta que justifique activar las etapas posteriores.

La decisión de activación se realiza mediante reglas deterministas implementadas en Python. Esto garantiza que la detección de eventos sea reproducible, trazable y no

dependa de la variabilidad de un modelo de lenguaje.

Se activa una alerta de helada cuando existe una predicción de dicho evento y se cumple al menos una de las siguientes condiciones:

- La temperatura mínima estimada es menor o igual a 3,0°C.
- El ensamble registra al menos cuatro votos favorables de los siete clasificadores.

Se activa una alerta de ola de calor cuando existe una predicción de dicho evento y se cumple al menos una de las siguientes condiciones:

- La temperatura máxima estimada alcanza o supera los 30,0°C.
- El ensamble registra al menos cuatro votos favorables de los siete clasificadores.

Cuando ambos tipos de evento se presentan simultáneamente en distintas comunas, el sistema clasifica la situación como una alerta mixta. Cuando no se superan los criterios de activación, el pipeline registra una condición normal y no ejecuta las etapas de razonamiento ni comunicación.

El uso de modelos de lenguaje dentro de este agente es opcional y limitado a la generación de una explicación breve en lenguaje natural. Si la respuesta no cumple el formato esperado o no puede ser interpretada correctamente, el sistema conserva la decisión determinista y continúa su operación sin depender del proveedor LLM.

6.10.2. Agente Razonador

El Agente Razonador recibe la alerta activada por el Monitor y genera una interpretación agroclimática enriquecida. Para ello, el sistema recopila previamente la información necesaria desde la base de datos y la entrega al modelo de lenguaje mediante un contexto estructurado.

Este procedimiento se denomina *pre-gathering*. En lugar de permitir que el modelo solicite datos de forma iterativa mediante herramientas externas, toda la información relevante se incorpora antes de invocar al LLM.

El contexto entregado al Agente Razonador incluye:

- Predicciones de temperatura mínima y máxima para las comunas afectadas.
- Tipo de evento detectado y nivel de consenso del ensamble de clasificadores.
- Historial térmico de los últimos siete días.
- Umbrales críticos por cultivo y etapa fenológica.
- Métricas recientes de confiabilidad del sistema predictivo.
- Ventana horaria estimada del evento cuando corresponde a una alerta de helada.

A partir de esta información, el modelo de lenguaje debe producir un análisis estructurado en formato JSON. El resultado incluye un resumen del riesgo, cultivos potencialmente afectados, tendencia térmica reciente, nivel de confiabilidad y recomendaciones priorizadas para los agricultores.

La decisión de utilizar *pre-gathering* responde a una razón metodológica fundamental: todos los modelos de lenguaje evaluados reciben exactamente el mismo contexto. Por lo tanto, las diferencias observadas en sus respuestas pueden atribuirse al comportamiento del modelo y no a diferencias en la información disponible.

6.10.3. Agente Comunicador

El Agente Comunicador transforma el análisis estructurado generado por el Razonador en un mensaje de alerta orientado a usuarios agrícolas.

El mensaje debe ser breve, claro y accionable. La comunicación incluye el tipo de evento detectado, las comunas afectadas, temperaturas relevantes, nivel de riesgo y recomendaciones prioritarias.

El canal principal de comunicación implementado es WhatsApp, seleccionado para entregar alertas de manera oportuna a dispositivos móviles. La utilización de servicios de mensajería móvil para distribuir recomendaciones agroclimáticas ha mostrado potencial para apoyar la adopción de consejos agrícolas por parte de los usuarios Sharma et al., 2021.

6.11. Evaluación Comparativa de Modelos de Lenguaje

La evaluación de los modelos de lenguaje se estructuró en dos dimensiones complementarias: una evaluación operacional automatizada y una evaluación cualitativa retrospectiva de los mensajes generados. Los proveedores considerados corresponden a Claude, Gemini y Llama 3, integrados mediante una interfaz común dentro de la arquitectura Multi-LLM-Agent.

La evaluación operacional analiza el comportamiento técnico de cada proveedor durante la ejecución del pipeline de alertas. Para cada ejecución se almacenan el proveedor utilizado, el modelo invocado, la respuesta estructurada generada por el Agente Razonador, el mensaje final generado por el Agente Comunicador, el tiempo total de procesamiento y el resultado del parseo del formato JSON.

Las métricas operacionales consideradas corresponden a la latencia de respuesta, la proporción de respuestas estructuradas válidas y la necesidad de aplicar mecanismos de recuperación o *fallback*. Estas métricas permiten evaluar si los modelos pueden integrarse de manera confiable dentro de un flujo automatizado de generación de alertas agroclimáticas.

La Figura 6.8 presenta el esquema general utilizado para integrar los proveedores de modelos de lenguaje dentro de la arquitectura. Cada proveedor recibe un contexto estructurado generado previamente por el Agente Monitor, el cual contiene información meteorológica, predicciones, umbrales agrícolas, historial reciente y antecedentes relevantes de la alerta. Posteriormente, el Agente Razonador genera una interpretación estructurada y el Agente Comunicador transforma dicho resultado en un mensaje orien-

tado al usuario final.

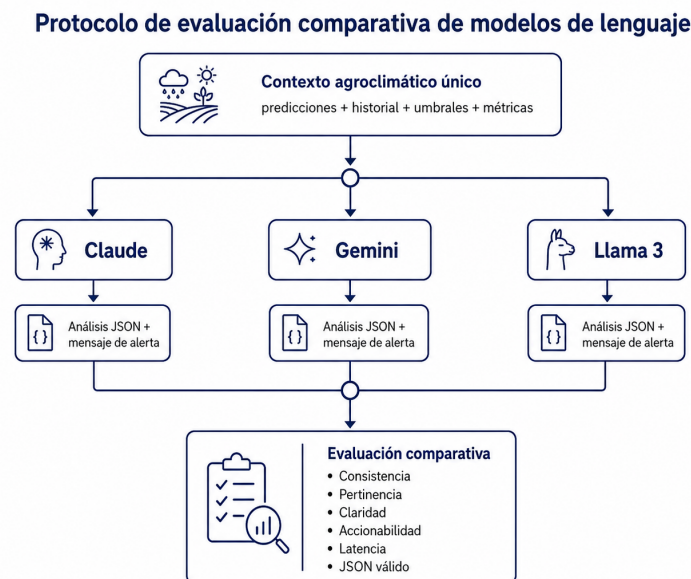


Figura 6.8: Esquema de integración y evaluación de proveedores de modelos de lenguaje dentro de la arquitectura Multi-LLM-Agent.

Cuando existen ejecuciones correspondientes a un mismo contexto agroclimático, los proveedores pueden ser comparados bajo condiciones equivalentes de entrada. Sin embargo, debido a la naturaleza operacional del sistema y a la disponibilidad desigual de registros almacenados por proveedor, no todas las ejecuciones analizadas corresponden necesariamente a los mismos eventos o contextos. Por esta razón, las comparaciones de latencia, robustez de formato y comportamiento general se interpretan como resultados operacionales descriptivos y no como una evaluación experimental pareada concluyente.

Además de la evaluación automatizada, se incorporó una evaluación cualitativa retrospectiva de los mensajes generados y almacenados por el sistema. Esta evaluación considera cada mensaje final junto con el contexto meteorológico y agrícola disponible al momento de su generación. El objetivo es analizar si las respuestas producidas por los modelos son coherentes con la alerta detectada y si entregan recomendaciones comprensibles y potencialmente útiles para usuarios agrícolas.

CAPÍTULO 6 : METODOLOGÍA

La evaluación cualitativa considera las dimensiones descritas en la Tabla 13. Las dimensiones de latencia y robustez de formato se obtienen automáticamente desde los registros del sistema. En cambio, consistencia interpretativa, pertinencia agronómica, claridad comunicacional y accionabilidad se evalúan mediante una rúbrica ordinal de cinco niveles, donde 1 representa un desempeño deficiente y 5 representa un desempeño alto.

Tabla 13: Dimensiones de evaluación de los proveedores de modelos de lenguaje.

Dimensión	Criterio de evaluación
Consistencia interpretativa	Coherencia entre el análisis generado y el contexto meteorológico, agrícola y predictivo entregado.
Pertinencia agronómica	Adecuación de las recomendaciones al evento detectado, cultivos afectados y nivel de riesgo informado.
Claridad comunicacional	Comprensibilidad del mensaje para usuarios agrícolas no especializados.
Accionabilidad	Presencia de recomendaciones concretas que puedan ser aplicadas preventivamente.
Latencia	Tiempo total requerido para completar el razonamiento y generar el mensaje final.
Robustez de formato	Capacidad de generar una respuesta JSON válida y ajustada al esquema definido.

La rúbrica cualitativa se aplica sobre el contenido generado por los modelos, considerando exclusivamente su correspondencia con el contexto de entrada y la calidad comunicacional de la alerta. La pauta de puntuación detallada, los criterios de cada nivel y el procedimiento de evaluación se presentan en el Anexo G.

La evaluación cualitativa tiene carácter exploratorio. Su propósito no es reemplazar una validación con agricultores, asesores agrícolas o especialistas agronómicos, sino

proporcionar una primera aproximación estructurada a la calidad de las recomendaciones generadas por los modelos de lenguaje. Una validación posterior con usuarios reales permitiría complementar estos resultados mediante criterios de utilidad práctica, comprensión y aplicabilidad en terreno.

En consecuencia, la evaluación no busca demostrar equivalencia estadística ni superioridad definitiva entre proveedores. Su objetivo es caracterizar el comportamiento técnico y comunicacional de los modelos de lenguaje dentro de una tarea especializada: interpretar eventos agroclimáticos y transformarlos en alertas comprensibles y accionables para usuarios agrícolas.

6.12. Operación Automatizada y Reentrenamiento

El sistema opera mediante un planificador de tareas configurado en la zona horaria America/Santiago. El flujo diario respeta el orden de dependencia entre adquisición de datos, validación de predicciones, generación de nuevos pronósticos y activación del pipeline Multi-LLM-Agent.

CAPÍTULO 6 : METODOLOGÍA

Tabla 14: *Secuencia general del pipeline diario automatizado.*

Hora	Etapas	Descripción
02:00	Ingesta INIA	Obtención de observaciones recientes desde estaciones agrometeorológicas.
02:15	Actualización horaria	Incorporación de datos horarios recientes para estimar condiciones térmicas.
02:30	Validación	Comparación entre predicciones previas y observaciones reales.
03:00	Ingesta MeteoChile	Obtención del pronóstico oficial utilizado como referencia externa.
03:15	Pronóstico Open-Meteo	Actualización de variables meteorológicas para horizontes futuros.
03:30	Predicción D+1	Generación de predicciones para el día siguiente.
03:35	Predicción D+2 a D+5	Generación de predicciones para horizontes posteriores.
03:45	Pipeline Multi-LLM-Agent	Activación de Monitor, Razonador y Comunicador cuando existe alerta.

El sistema incorpora además un ciclo de reentrenamiento periódico. Este proceso combina la base histórica disponible con observaciones recientes obtenidas durante la operación. El objetivo es reducir el envejecimiento de los modelos y permitir que incorporen progresivamente patrones térmicos locales observados.

El reentrenamiento se ejecuta semanalmente y requiere un mínimo de 200 observaciones diarias válidas por comuna. Cuando una comuna no alcanza este umbral, se conserva el modelo de la versión anterior. Esta decisión evita generar modelos entrenados con información insuficiente.

El flujo automatizado, las etapas operacionales y la programación de tareas del pipeline se detallan en el Anexo C.

6.13. Métricas de Evaluación del Sistema Predictivo

La evaluación del sistema predictivo se realiza diariamente durante la etapa de validación, comparando las predicciones emitidas para una fecha determinada con las observaciones reales disponibles posteriormente.

Las métricas se calculan para el sistema predictivo propuesto, Open-Meteo y MeteoChile, permitiendo comparar el comportamiento de cada fuente frente a los eventos observados.

Tabla 15: Matriz de confusión utilizada para evaluar la detección de eventos térmicos.

	Evento observado	Sin evento observado
Evento predicho	Verdadero positivo (TP)	Falsa alarma (FP)
Sin evento predicho	Evento no detectado (FN)	Verdadero negativo (TN)

La Tabla 15 resume los resultados posibles al contrastar una predicción con la observación real de un evento térmico. Estos valores se utilizan para calcular las métricas de clasificación presentadas a continuación.

6.13.1. Critical Success Index

El *Critical Success Index* corresponde a la métrica principal para evaluar eventos poco frecuentes. A diferencia de la exactitud global, no considera los verdaderos negativos, evitando que una gran cantidad de días normales oculte deficiencias en la detección de eventos extremos Ferro y Stephenson, 2011.

$$CSI = \frac{TP}{TP + FP + FN} \quad (8)$$

Un valor de CSI cercano a 1 representa una detección altamente precisa, mientras

que un valor cercano a 0 indica bajo desempeño en la identificación de eventos.

6.13.2. Probabilidad de Detección

La probabilidad de detección, también denominada *Probability of Detection* o POD, mide la proporción de eventos reales que fueron identificados correctamente por el modelo.

$$POD = \frac{TP}{TP + FN} \quad (9)$$

Esta métrica resulta relevante para agricultura inteligente, ya que los eventos no detectados pueden impedir que los usuarios adopten medidas preventivas oportunas.

6.13.3. Raíz del Error Cuadrático Medio

La raíz del error cuadrático medio permite evaluar el desempeño de las estimaciones numéricas de temperatura mínima y máxima.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{T}_i - T_i)^2} \quad (10)$$

donde $\hat{T}_i$ representa la temperatura estimada y T_i corresponde a la temperatura observada.

El RMSE se expresa en grados Celsius y complementa las métricas de clasificación, permitiendo analizar la precisión de las estimaciones continuas de temperatura.

6.14. Síntesis Metodológica

La metodología desarrollada integra la construcción de un sistema predictivo de eventos térmicos extremos y la evaluación de una arquitectura Multi-LLM-Agent orientada a agricultura inteligente.

CAPÍTULO 6 : METODOLOGÍA

La primera línea metodológica aborda la preparación de datos meteorológicos, el entrenamiento de un ensamble de clasificadores, la estimación de temperaturas, la corrección de sesgo y la evaluación comparativa frente a fuentes externas de referencia.

La segunda línea metodológica aborda la transformación de predicciones numéricas en alertas agroclimáticas contextualizadas. Para ello, la arquitectura Multi-LLM-Agent combina reglas deterministas de activación, conocimiento agrícola estructurado, modelos de lenguaje y mecanismos de comunicación orientados a usuarios finales.

La separación entre predicción e interpretación permite evaluar de manera independiente la calidad del sistema predictivo y el comportamiento de los proveedores de modelos de lenguaje. En consecuencia, esta metodología establece las condiciones necesarias para analizar, en el capítulo siguiente, los resultados obtenidos tanto en la detección de eventos térmicos extremos como en la generación de alertas agroclimáticas mediante una arquitectura Multi-LLM-Agent.

7. Resultados

Este capítulo presenta los resultados obtenidos a partir del corte de datos disponible al 3 de julio de 2026. La evaluación considera la cobertura y calidad de las series meteorológicas, el desempeño del sistema predictivo para eventos de helada y ola de calor, la estimación de temperatura, la evaluación multi-horizonte, la comparación operacional con MeteoChile y el funcionamiento de la arquitectura Multi-LLM-Agent.

Los resultados deben interpretarse según el período efectivo de operación disponible. En particular, la evaluación de heladas se concentra en un período otoño-invierno, mientras que la evidencia disponible para ola de calor y para los horizontes D+3 a D+5 es reducida. Por ello, estos últimos resultados se presentan como evidencia operacional preliminar y no como una validación concluyente.

7.1. Cobertura, integración y calidad de las series meteorológicas

La Tabla 16 resume la cobertura disponible para las fuentes empleadas por el sistema. El Atlas Agroclimático de Chile constituye la base histórica entre 2004 y 2016. Open-Meteo histórico aporta información desde 2020 y las observaciones de la red INIA/AGROMET complementan el período reciente con mediciones directas de estación.

Tabla 16: Cobertura temporal y territorial de las fuentes meteorológicas utilizadas.

Fuente	Inicio	Fin	Registros	Comunas	Rol metodológico
Atlas Agroclimático (ARClím)	01-01-2004	31-12-2016	246.875	52	Base histórica de largo plazo
Open-Meteo histórico	01-01-2020	23-04-2026	119.860	52	Complemento temporal y espacial
INIA/AGROMET	01-01-2020	01-07-2026	68.899	20	Observaciones de estación; prioridad máxima

La cobertura disponible evidencia una brecha temporal entre 2017 y 2019. Este intervalo no fue completado mediante imputación, ya que la metodología privilegia mantener únicamente observaciones disponibles y evitar la incorporación de valores

CAPÍTULO 7 : RESULTADOS

sintéticos que puedan alterar el entrenamiento. Esta limitación se considera al interpretar la continuidad histórica de las series.

La estrategia de integración aplica la prioridad INIA/AGROMET > Open-Meteo histórico > Atlas Agroclimático. Después de resolver los solapamientos por comuna y fecha, la serie consolidada contiene 368.082 pares únicos comuna × fecha. La composición final se presenta en la Tabla 17

Tabla 17: Composición de la serie meteorológica consolidada después de la fusión jerárquica.

Fuente ganadora	Pares consolidados	Participación	Descripción
Atlas Agroclimático	246.875	67,07 %	Base histórica 2004–2016
Open-Meteo histórico	82.349	22,37 %	Reanálisis 2020–2026
INIA/AGROMET	38.858	10,56 %	Observaciones directas de mayor prioridad
Total	368.082	100,00 %	Pares únicos comuna × fecha

La mayor parte de la serie consolidada corresponde al Atlas Agroclimático, debido a que esta fuente cubre el intervalo histórico más extenso. Open-Meteo histórico aporta continuidad durante el período reciente y las observaciones INIA representan la fracción de mayor prioridad cuando existen registros de estación disponibles. La Figura 7.1 presenta la cobertura temporal por fuente, mientras que la Figura 7.2 resume la composición porcentual de la serie consolidada.

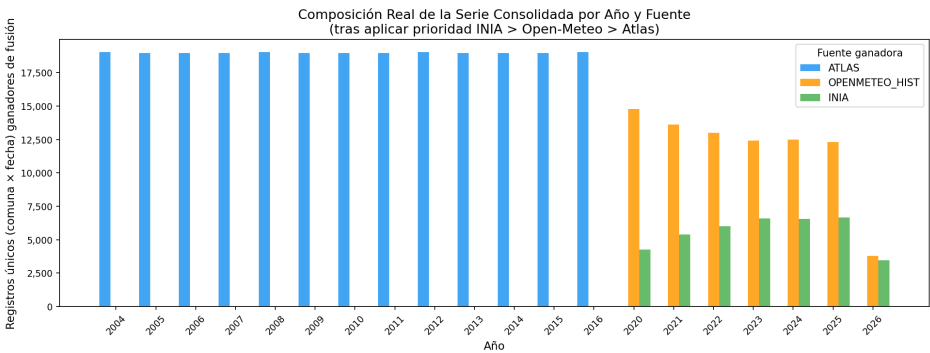
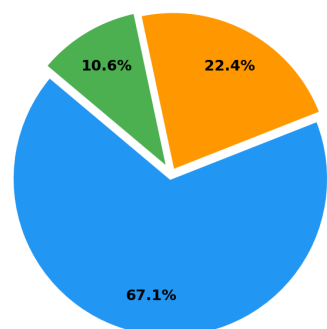


Figura 7.1: Cobertura temporal de las fuentes meteorológicas.

CAPÍTULO 7 : RESULTADOS

Composición Real de la Serie Consolidada tras Fusión Jerárquica
Total: 368,082 pares únicos (comuna × fecha)
Prioridad: INIA > Open-Meteo > Atlas



ATLAS (67.1% 246,875 registros)
Base histórica 2004-2016
Open-Meteo Histórico (22.4% 82,349 registros)
Reanálisis ERA5 2020-2026
INIA/AGROMET (10.6% 38,858 registros)
Estaciones reales — prioridad máxima

Figura 7.2: Composición de la serie meteorológica consolidada según fuente prioritaria.

La Tabla 18 presenta la proporción de registros que no disponían de una medición para cada variable meteorológica antes de la etapa de preparación de datos. Atlas Agroclimático y Open-Meteo histórico presentan una cobertura prácticamente completa en las variables críticas del sistema, particularmente temperatura mínima, temperatura máxima y humedad relativa. En contraste, la fuente INIA/AGROMET presenta mayores porcentajes de ausencia, especialmente en radiación solar, situación esperable en una red operacional de estaciones físicas, donde pueden ocurrir interrupciones de sensores, transmisión o disponibilidad de registros.

Los valores faltantes no se tratan de manera uniforme. La precipitación se imputa con cero, la radiación solar se completa mediante interpolación temporal y las observaciones con ausencia en temperatura mínima, temperatura máxima o humedad relativa se excluyen del conjunto de entrenamiento, con el objetivo de evitar la incorporación de valores artificiales en variables críticas.

CAPÍTULO 7 : RESULTADOS

Tabla 18: Porcentaje de valores faltantes por fuente y variable meteorológica.

Fuente	T. mínima	T. máxima	Humedad	Precipitación	Radiación
Atlas Agroclimático	0,00 %	0,00 %	0,00 %	0,00 %	0,08 %
Open-Meteo histórico	0,00 %	0,00 %	0,00 %	0,00 %	0,00 %
INIA/AGROMET	12,78 %	12,78 %	12,91 %	15,25 %	22,56 %

7.2. Desempeño del sistema predictivo para el horizonte D+1

La evaluación D+1 considera el corte operativo disponible de 3.206 predicciones validadas. La Tabla 19 presenta las métricas de clasificación para heladas y olas de calor, calculadas a partir de las matrices de confusión correspondientes.

Tabla 19: Desempeño global del sistema predictivo para D+1.

Evento	CSI	POD	Precisión	F1	TP	FP	FN	TN	N
Helada	0,395	0,773	0,447	0,566	625	773	184	1.624	3.206
Ola de calor	0,000	0,000	0,000	0,000	0	15	4	3.187	3.206

Para heladas, el sistema alcanzó un CSI de 0,395, una probabilidad de detección de 0,773, una precisión de 0,447 y un F1-score de 0,566. En términos prácticos, el ensamble identificó 625 de 809 heladas observadas durante el período evaluado. El resultado refleja una estrategia orientada a privilegiar la detección de eventos potencialmente dañinos, aunque con una cantidad relevante de falsas alarmas: 773 de las 1.398 alertas positivas de helada no fueron confirmadas posteriormente.

En el caso de ola de calor, el período de evaluación contiene solo cuatro eventos reales y no se registraron verdaderos positivos. Por ello, las métricas obtenidas para este evento no permiten caracterizar de forma concluyente la capacidad predictiva del

CAPÍTULO 7 : RESULTADOS

sistema. La evaluación de ola de calor deberá ampliarse con registros de primavera y verano antes de establecer conclusiones sobre su rendimiento.

La Figura 7.3 permite visualizar las métricas principales de D+1 para helada. Debido a la baja cantidad de eventos de ola de calor, no se recomienda interpretar gráficamente su comparación directa con heladas como evidencia de superioridad o inferioridad del modelo.

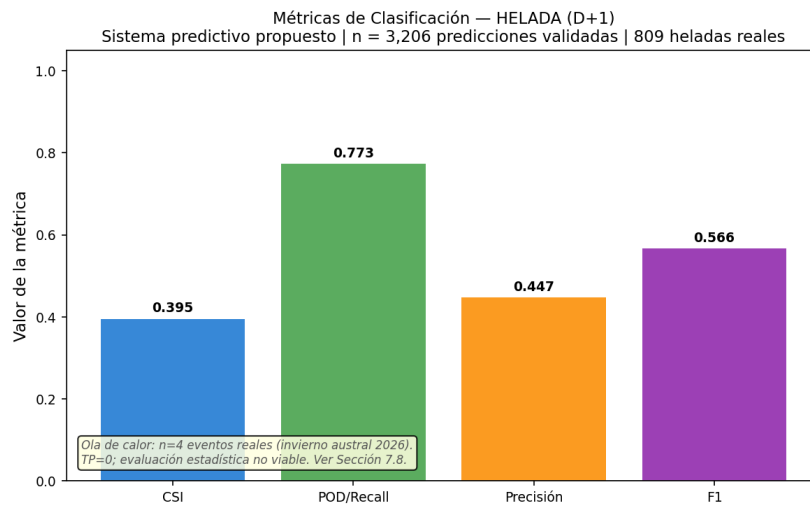


Figura 7.3: Métricas CSI, POD, precisión y F1 para helada en D+1.

7.2.1. Variación del desempeño entre comunas

Para analizar la variabilidad territorial, se calcularon métricas desde TP, FP, FN y TN por comuna, evitando promediar métricas previamente agregadas. La Tabla 20 muestra las cinco comunas con mayor y menor CSI, restringiendo el ranking a comunas con al menos cinco heladas reales durante el período evaluado.

CAPÍTULO 7 : RESULTADOS

Tabla 20: *Comunas con mayor y menor desempeño CSI para helada en D+1.*

Grupo	Comuna	CSI	POD	Precisión	F1	TP	FP	FN	Heladas reales
Mayor CSI	Maipú	0,597	0,895	0,641	0,747	34	19	4	38
Mayor CSI	Lampa	0,582	0,821	0,667	0,736	32	16	7	39
Mayor CSI	Peñalolén	0,538	1,000	0,538	0,700	14	12	0	14
Mayor CSI	Macul	0,520	0,929	0,542	0,684	13	11	1	14
Mayor CSI	Quinta Normal	0,520	0,929	0,542	0,684	13	11	1	14
Menor CSI	La Reina	0,000	0,000	0,000	0,000	0	5	6	6
Menor CSI	Pirque	0,000	0,000	0,000	0,000	0	4	10	10
Menor CSI	Padre Hurtado	0,088	0,120	0,250	0,162	3	9	22	25
Menor CSI	Tiltil	0,111	0,200	0,200	0,200	1	4	4	5
Menor CSI	Cerro Navia	0,176	0,429	0,231	0,300	3	10	4	7

Maipú y Lampa presentan los mayores valores de CSI dentro del corte analizado, con 0,597 y 0,582, respectivamente. En contraste, La Reina y Pirque no registran verdaderos positivos en el período disponible. Estas diferencias deben interpretarse con cautela, puesto que dependen de la frecuencia de heladas, de la cobertura efectiva de observaciones y de características locales no modeladas explícitamente. La información actual no permite atribuir causalmente el desempeño a un único factor.

Con el propósito de analizar la variabilidad territorial del desempeño del sistema, se calculó el *Critical Success Index* (CSI) de detección de heladas para cada comuna evaluada. Esta métrica resume la capacidad del sistema para identificar eventos de helada, penalizando tanto las falsas alarmas como los eventos no detectados y sin considerar los verdaderos negativos. Por ello, un valor de CSI más alto representa una mejor correspondencia entre las alertas emitidas y los eventos realmente observados.

La Figura 7.4 presenta los resultados obtenidos para las veinte comunas con mayor cantidad de registros evaluables en el horizonte D+1. Cada barra corresponde al

CAPÍTULO 7 : RESULTADOS

CSI calculado a partir de las predicciones y observaciones disponibles para una comuna específica durante el período analizado. La selección de las veinte comunas busca evitar comparaciones basadas en muestras demasiado pequeñas, dado que la frecuencia de heladas y la disponibilidad de observaciones no fueron homogéneas en todo el territorio.

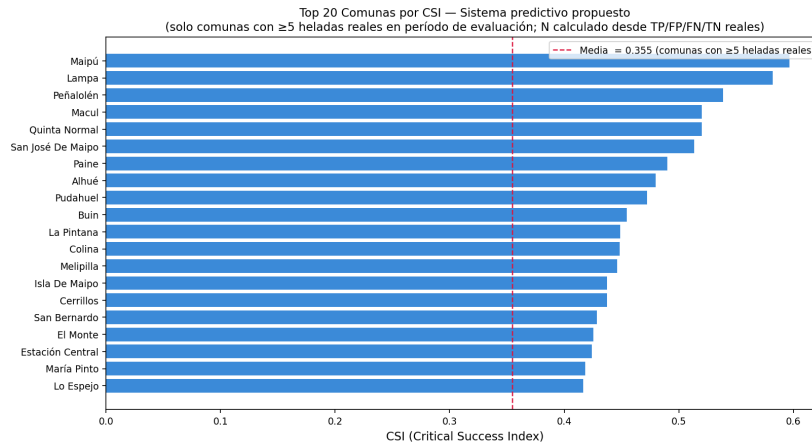


Figura 7.4: CSI de helada por comuna para las veinte comunas con mayor número de registros evaluables.

El ensamble aplica una regla de activación cuando al menos cuatro de siete clasificadores emiten un voto positivo. La distribución de votos puede incorporarse como figura complementaria para mostrar el nivel de consenso interno de los modelos y la concentración de las decisiones en torno a los extremos de 0 y 7 votos.

7.3. Desempeño de regresión de temperatura

La Tabla 21 presenta el error de las estimaciones de temperatura mínima y máxima para D+1. La temperatura mínima obtuvo un RMSE de 3,211 °C, mientras que la temperatura máxima alcanzó un RMSE de 5,031 °C.

CAPÍTULO 7 : RESULTADOS

Tabla 21: Métricas de regresión para temperatura en D+1.

Variable	RMSE (°C)	MAE (°C)	Sesgo medio (°C)	N
Temperatura mínima	3,211	2,551	+0,084	3.194
Temperatura máxima	5,031	4,073	-0,237	3.206

Con el fin de evaluar la precisión de las estimaciones continuas de temperatura para el horizonte D+1, se compararon las temperaturas observadas posteriormente con las temperaturas predichas por el sistema. La Figura 7.5 presenta diagramas de dispersión para temperatura mínima y máxima, donde cada punto representa una predicción evaluada. La línea diagonal roja corresponde al escenario de predicción perfecta, es decir, aquellos casos en que la temperatura estimada coincide exactamente con la temperatura observada.

Los puntos ubicados próximos a esta línea representan predicciones con menor error, mientras que una mayor distancia respecto de la diagonal indica una discrepancia más alta entre el valor estimado y el observado. La figura permite observar que las estimaciones de temperatura mínima presentan una concentración relativamente más cercana a la diagonal de referencia que las estimaciones de temperatura máxima, lo que resulta coherente con los valores de RMSE calculados para ambas variables.

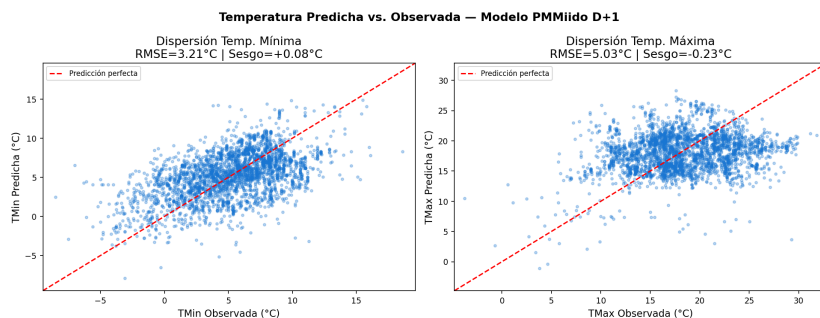


Figura 7.5: Comparación entre temperaturas observadas y estimadas por el sistema predictivo para el horizonte D+1. Cada punto representa una predicción evaluada y la línea diagonal roja indica el caso de predicción perfecta.

CAPÍTULO 7 : RESULTADOS

El sesgo promedio de temperatura mínima fue de $+0,084\text{ }^{\circ}\text{C}$, mientras que para temperatura máxima fue de $-0,237\text{ }^{\circ}\text{C}$. En el agregado del período evaluado, estos valores indican que no existe un desplazamiento sistemático de gran magnitud en las estimaciones almacenadas. Sin embargo, los valores de RMSE muestran que persiste una diferencia relevante entre las temperaturas previstas y observadas, especialmente para las máximas diarias.

El sistema incorpora una corrección adaptativa de sesgo basada en las últimas predicciones validadas por comuna. En el corte actual, el RMSE de temperatura mínima cambió de $3,2106\text{ }^{\circ}\text{C}$ a $3,2095\text{ }^{\circ}\text{C}$ tras aplicar la corrección, equivalente a una variación de $-0,03\%$. Por tanto, en este período la corrección debe entenderse principalmente como un mecanismo operativo de estabilización local y no como una mejora concluyente del desempeño global.

7.4. Evaluación multi-horizonte D+1 a D+5

La Tabla 22 resume los resultados disponibles para los cinco horizontes de predicción. El horizonte D+1 concentra la mayor cantidad de observaciones validadas. Los horizontes posteriores disponen de una muestra considerablemente menor, debido a que el período de operación y validación aún es acotado.

Tabla 22: Resultados disponibles por horizonte de predicción.

Horizonte	N	Heladas reales	CSI	POD	RMSE Tmin	RMSE Tmax	Fuente de temperatura
D+1	3.206	809	0,395	0,773	3,211	5,031	Regresor propio (RF)
D+2	167	37	0,027	0,027	3,875	2,923	Open-Meteo Forecast
D+3	58	18	0,000	0,000	2,243	2,601	Open-Meteo Forecast
D+4	6	4	0,000	0,000	3,504	4,219	Open-Meteo Forecast
D+5	6	3	0,250	0,333	4,442	6,378	Open-Meteo Forecast

La Figura 7.6 complementa la información presentada en la Tabla 22, mostrando de forma visual la variación del desempeño y del tamaño muestral para los horizontes D+1

CAPÍTULO 7 : RESULTADOS

a D+5. La figura integra tres dimensiones: el CSI obtenido en la detección de heladas, el RMSE asociado a las estimaciones de temperatura mínima y máxima, y la cantidad de registros validados en cada horizonte.

La interpretación de los indicadores debe realizarse considerando simultáneamente el número de registros disponibles. Mientras D+1 cuenta con una base amplia de observaciones y eventos reales, los horizontes posteriores presentan una disminución pronunciada de casos evaluables. Por esta razón, los valores observados para D+3, D+4 y D+5 no permiten establecer comparaciones concluyentes respecto de D+1, sino que deben entenderse como resultados exploratorios del período de operación disponible.

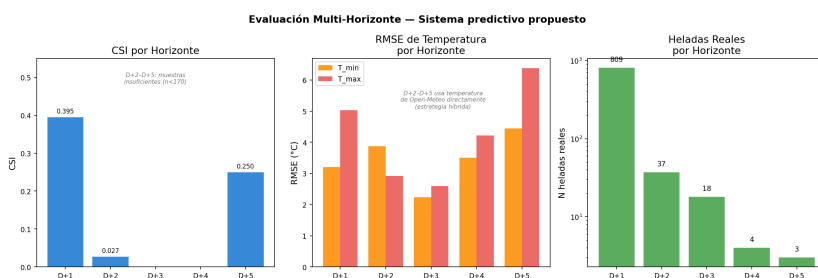


Figura 7.6: Variación de CSI, RMSE y cantidad de registros por horizonte D+1 a D+5.

Como se observa en la Figura 7.6, D+1 concentra la mayor cantidad de registros evaluables y constituye el único horizonte con una base suficiente para una evaluación cuantitativa más representativa dentro del período analizado. D+2 dispone de 167 registros y presenta un CSI de 0,027, asociado a un único verdadero positivo. En cambio, D+3 a D+5 contienen entre 6 y 58 registros, por lo que sus métricas deben interpretarse únicamente de manera descriptiva.

La disminución del tamaño muestral también limita la interpretación de los valores de RMSE para los horizontes posteriores. Por ejemplo, un RMSE aparentemente menor en un horizonte específico no implica necesariamente un mejor comportamiento predictivo, ya que puede estar calculado sobre un conjunto reducido de fechas y condiciones meteorológicas.

Para los horizontes D+2 a D+5, la temperatura futura utilizada por el sistema pro-

viene directamente de Open-Meteo Forecast, mientras que los clasificadores aplican una estrategia híbrida basada en variables meteorológicas futuras y actualización iterativa de características temporales. En consecuencia, las métricas de estos horizontes representan el comportamiento del flujo híbrido completo y no exclusivamente del modelo de aprendizaje automático propio.

7.5. Comparación operacional con fuentes externas

La comparación con Open-Meteo no se incorpora como resultado principal, ya que la base de datos contiene solo 104 predicciones D+1 validadas para esta fuente, correspondientes a cinco días y nueve heladas reales. Esta cantidad es insuficiente para sostener una comparación robusta.

En cambio, la comparación con MeteoChile se realizó sobre una intersección estricta de 3.034 pares comuna-fecha. Ambos sistemas fueron evaluados con la misma definición de helada observada: temperatura mínima real inferior a 3,0 °C. Los resultados se presentan en la Tabla 23

Tabla 23: Comparación operacional entre el sistema predictivo propuesto y MeteoChile para pares comunes.

Sistema	CSI	POD	Precisión	F1	RMSE Tmin	TP	FP	FN	TN	N
Sistema predictivo propuesto	0,394	0,764	0,449	0,566	3,147	583	716	180	1.555	3.034
MeteoChile	0,084	0,085	0,833	0,155	4,431	65	13	698	2.258	3.034

Sobre el conjunto común, el sistema predictivo propuesto alcanzó CSI de 0,394 y POD de 0,764, mientras que MeteoChile obtuvo CSI de 0,084 y POD de 0,085. MeteoChile presentó mayor precisión, debido a que generó menos alertas positivas y concentró sus aciertos en un número reducido de eventos. El sistema propuesto, en cambio, detectó una mayor proporción de heladas observadas, a costa de una mayor tasa de falsas alarmas.

El sistema predictivo propuesto presentó además un menor RMSE para temperatura

CAPÍTULO 7 : RESULTADOS

mínima, con 3,147 °C frente a 4,431 °C de MeteoChile. Sin embargo, esta comparación debe interpretarse como un análisis operacional descriptivo y no como una demostración definitiva de superioridad, dado que no se aplicaron pruebas formales de significancia estadística y ambas fuentes pueden diferir en su resolución espacial, cobertura territorial y objetivo de pronóstico.

La Figura 7.7 permite visualizar de manera comparativa los indicadores presentados en la Tabla 23. En particular, se contrastan métricas orientadas a la detección de eventos, tales como CSI y POD, junto con la precisión y el RMSE de temperatura mínima. Esta representación permite distinguir que los sistemas no presentan el mismo comportamiento operacional: mientras un sistema puede generar menos alertas y alcanzar una mayor precisión, también puede dejar sin detectar una proporción relevante de los eventos observados.

En este contexto, el CSI resume el equilibrio entre eventos correctamente detectados, falsas alarmas y eventos no detectados, mientras que el POD refleja la proporción de heladas reales que fueron identificadas. Por su parte, la precisión indica qué proporción de las alertas emitidas correspondió efectivamente a una helada observada. Por ello, los indicadores deben interpretarse de manera conjunta y no de forma aislada.

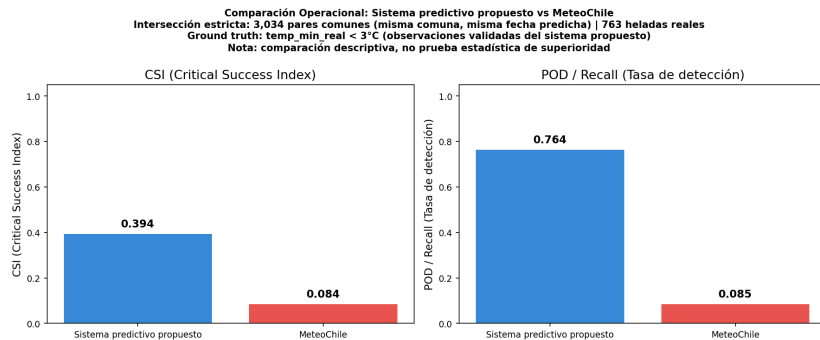


Figura 7.7: Comparación de CSI, POD, precisión y RMSE de temperatura mínima entre el sistema propuesto y MeteoChile.

Como se observa en la Figura 7.7, el sistema predictivo propuesto alcanzó valores superiores de CSI y POD respecto de MeteoChile, lo que indica una mayor capacidad para detectar las heladas observadas dentro del conjunto común de evaluación. En

contraste, MeteoChile presentó una precisión superior, asociada a una estrategia más conservadora de emisión de alertas, con menor cantidad de falsos positivos, pero también con un número considerablemente mayor de eventos no detectados.

7.6. Evaluación de la arquitectura Multi-LLM-Agent

Durante el período analizado, la arquitectura Multi-LLM-Agent procesó 199 alertas, de las cuales 195 correspondieron a eventos de helada y 4 a alertas mixtas. La Tabla 24 resume el comportamiento operacional observado para los tres proveedores de modelos de lenguaje integrados en el sistema.

Tabla 24: Resultados operacionales de los proveedores LLM.

Proveedor	Modelo	Alertas	JSON sin fallback	JSON final válido	Latencia media (ms)	Latencia mediana (ms)
Claude	claude-haiku-4-5-20251001	73	100,0 %	100,0 %	18.660	17.818
Gemini	gemini-2.5-flash	55	90,9 %	100,0 %	8.300	7.276
Llama 3	llama-3.3-70b-versatile (Groq)	71	100,0 %	100,0 %	2.987	2.658

Claude y Llama 3 generaron respuestas JSON válidas sin requerir mecanismos de recuperación en el 100 % de sus ejecuciones registradas. Gemini alcanzó un 90,9 % de respuestas válidas antes de aplicar correcciones, debido a cinco casos donde el campo destinado al mensaje corto contenía un bloque JSON en lugar de texto plano. No obstante, el mecanismo de parseo y *fallback* permitió que las respuestas finales fueran válidas en el 100 % de los casos almacenados.

En términos de latencia, Llama 3 presentó el menor tiempo promedio observado, seguido por Gemini y Claude. Sin embargo, estos resultados deben interpretarse como evidencia operacional descriptiva, dado que los proveedores no procesaron exactamente el mismo conjunto total de alertas ni fueron evaluados exclusivamente bajo un diseño pareado con contextos idénticos.

La Figura 7.8 complementa los resultados de la Tabla 24, mostrando la latencia media y la dispersión de los tiempos de respuesta registrados para cada proveedor. La

CAPÍTULO 7 : RESULTADOS

latencia corresponde al tiempo transcurrido entre el envío de la solicitud al proveedor y la recepción de una respuesta procesable por la arquitectura, por lo que constituye un indicador relevante para un sistema orientado a la emisión oportuna de alertas.

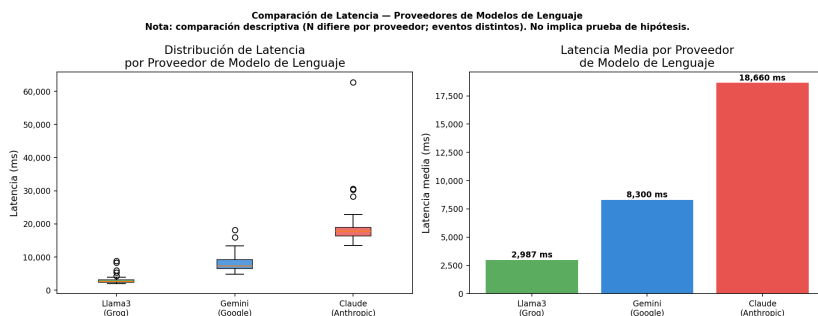


Figura 7.8: Latencia media y distribución de tiempos de respuesta observados para los proveedores LLM utilizados en la arquitectura Multi-LLM-Agent.

Como se observa en la Figura 7.8, Llama 3 presentó la menor latencia media dentro del período analizado, seguido por Gemini y Claude. Claude presentó tanto el mayor promedio como una mayor dispersión relativa en sus tiempos de respuesta. Estas diferencias sugieren comportamientos operacionales distintos entre proveedores, aunque no permiten establecer una comparación experimental concluyente de rendimiento.

7.6.1. Evaluación cualitativa retrospectiva

Con el propósito de complementar la evaluación operacional, se realizó una evaluación cualitativa retrospectiva de los mensajes de alerta almacenados por el sistema. La evaluación se realizó utilizando una rúbrica estructurada de cuatro dimensiones: consistencia interpretativa, pertinencia agronómica, claridad comunicacional y accionabilidad.

La evaluación no corresponde a una validación agronómica o con agricultores. Por tanto, los puntajes obtenidos deben interpretarse como indicadores exploratorios de calidad textual y coherencia contextual, y no como una medida definitiva de utilidad práctica en terreno.

CAPÍTULO 7 : RESULTADOS

A partir de los 199 registros operacionales, se obtuvieron 160 mensajes únicos luego de eliminar ejecuciones duplicadas correspondientes a una misma fecha y proveedor. Del total, 126 mensajes correspondieron a 42 fechas en las cuales los tres proveedores generaron alertas para un mismo contexto predictivo. Durante la asignación de puntajes, la identidad del proveedor se ocultó para reducir sesgos asociados al nombre del modelo; posteriormente, los resultados se vincularon nuevamente con cada proveedor para su análisis agregado.

La Tabla 25 presenta los resultados promedio obtenidos por proveedor. Cada dimensión se evaluó en una escala de 1 a 5, donde 1 representa un desempeño deficiente y 5 un desempeño alto.

Tabla 25: Evaluación cualitativa retrospectiva de mensajes generados por los proveedores LLM. Escala de 1 a 5.

Proveedor	N	Consistencia	Pertinencia	Claridad	Accionabilidad	Global	Rev. manual
Claude	59	4,61	4,17	4,92	4,25	4,49	14
Gemini	44	4,18	4,73	4,43	4,50	4,46	6
Llama 3	57	3,28	4,00	3,04	3,25	3,39	11

En el conjunto evaluado, Claude obtuvo el mayor puntaje global promedio, con 4,49, seguido estrechamente por Gemini, con 4,46. Llama 3 obtuvo un puntaje global promedio de 3,39. Claude presentó el valor más alto en consistencia interpretativa y claridad comunicacional, mientras que Gemini obtuvo los mayores puntajes en pertinencia agronómica y accionabilidad.

Los resultados sugieren que Claude y Gemini generaron mensajes generalmente coherentes con el contexto disponible, comprensibles y con recomendaciones aplicables en términos generales. En cambio, los mensajes de Llama 3 tendieron a ser más breves y menos detallados, lo que se reflejó principalmente en menores puntajes de claridad y accionabilidad.

La Figura 7.9 presenta visualmente los puntajes promedio obtenidos en las cuatro

CAPÍTULO 7 : RESULTADOS

dimensiones cualitativas. La figura permite observar que las diferencias más marcadas se concentran en claridad comunicacional y accionabilidad, mientras que los puntajes de pertinencia agronómica se mantienen relativamente altos para los tres proveedores.

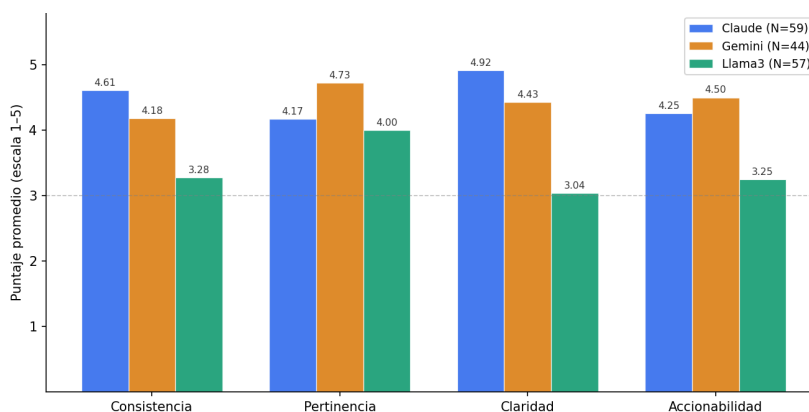


Figura 7.9: Puntajes promedio de consistencia interpretativa, pertinencia agronómica, claridad comunicacional y accionabilidad por proveedor LLM. Escala de 1 a 5.

Se identificaron 31 mensajes con al menos una bandera de revisión manual. Los principales motivos correspondieron a contexto predictivo insuficiente durante las primeras semanas de operación, inconsistencias entre temperaturas mencionadas y contexto disponible, recomendaciones vacías o poco específicas y presencia de JSON crudo en algunos mensajes almacenados. En particular, se detectaron tres registros donde el mensaje corto contenía directamente una estructura JSON, lo que afecta negativamente su claridad y utilidad para un usuario final.

Estos casos muestran que la calidad de salida no depende únicamente del proveedor LLM, sino también de la calidad y completitud del contexto disponible, de la correcta aplicación del parser y de la presencia de recomendaciones estructuradas en el pipeline. Los ejemplos de mayor y menor puntaje, junto con los casos marcados para revisión, se presentan en el Anexo H.

7.6.2. Evidencia de entrega de alertas mediante WhatsApp

Como verificación complementaria de la operación extremo a extremo, se realizaron pruebas de entrega de alertas mediante el canal de comunicación implementado sobre WhatsApp. La Figura 7.10 presenta ejemplos de mensajes recibidos en un dispositivo móvil durante pruebas de funcionamiento del sistema.

Los mensajes incluyen alertas de helada, alertas mixtas y mensajes informativos asociados al servicio. En particular, las alertas comunican el tipo de evento detectado, las comunas potencialmente afectadas, temperaturas relevantes, nivel de confiabilidad y recomendaciones preventivas generales. Asimismo, se observa la disponibilidad de comandos básicos para consultar predicciones, alertas activas y gestionar la suscripción a notificaciones.

Esta evidencia permite verificar que la arquitectura no solo genera resultados estructurados internamente, sino que también logra transformarlos en mensajes legibles y entregarlos mediante un canal accesible para el usuario final. Sin embargo, estas pruebas corresponden a una validación técnica de entrega y formato; no constituyen una evaluación formal de comprensión, utilidad práctica o adopción por parte de agricultores.

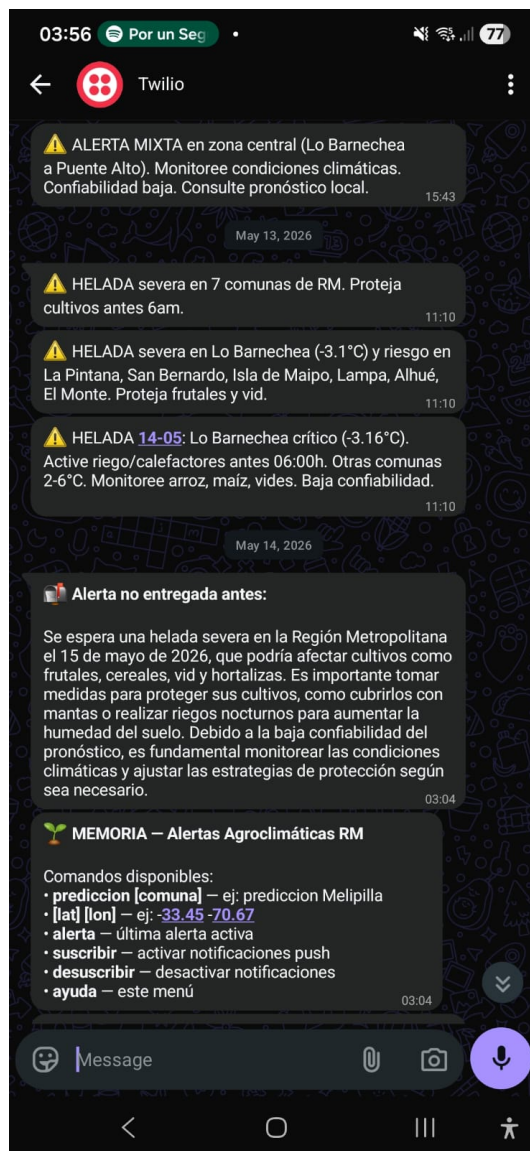


Figura 7.10: Ejemplos de alertas agroclimáticas e interacción básica del sistema recibidos mediante WhatsApp durante pruebas de operación.

7.7. Casos representativos del pipeline

La Tabla 26 presenta tres ejemplos reales extraídos de las predicciones almacenadas por el sistema. Los casos incluyen una helada correctamente detectada, una condición normal correctamente clasificada y una alerta de ola de calor que posteriormente no fue confirmada.

CAPÍTULO 7 : RESULTADOS

Tabla 26: Casos representativos de predicción y validación.

Caso	Comuna	Fecha	Variable	Temp. pred. (°C)	Temp. real (°C)	Prob.	Salida	Resultado
Helada confirmada	Lo Barnechea	23-06-2026	Tmin	-7,90	-3,10	1,000	Helada	Verdadero positivo
Condición normal	Tiltil	06-05-2026	Tmin	8,34	5,80	0,000	Normal	Verdadero negativo
Ola de calor	La Florida	04-05-2026	Tmax	21,41	21,67	0,714	Ola de calor	Falsa alarma

El primer caso corresponde a Lo Barnechea el 23 de junio de 2026. El sistema estimó una temperatura mínima de -7,90 °C y clasificó el evento como helada con probabilidad 1,000. La observación posterior registró una temperatura mínima de -3,10 °C, por lo que el evento fue correctamente identificado. Sin embargo, la diferencia entre ambas temperaturas evidencia que una clasificación correcta del evento no implica necesariamente una estimación numérica exacta de su magnitud.

El segundo caso corresponde a Tiltil el 6 de mayo de 2026. La predicción indicó una condición normal, con temperatura mínima estimada de 8,34 °C. La observación posterior fue de 5,80 °C, manteniéndose por sobre el umbral operacional de helada. En consecuencia, el sistema clasificó correctamente una condición sin evento térmico extremo.

El tercer caso corresponde a una alerta de ola de calor emitida para La Florida el 4 de mayo de 2026. El sistema clasificó el evento con probabilidad 0,714, pero la temperatura máxima observada posteriormente fue de 21,67 °C, valor inferior al umbral operacional de 30,0 °C. Por tanto, el caso fue clasificado como falsa alarma. Este resultado refuerza la necesidad de ampliar la evaluación de olas de calor durante meses estivales, donde exista una mayor frecuencia de temperaturas máximas elevadas.

De forma complementaria, el Anexo A.7 presenta ejemplos anonimizados de mensajes generados por los proveedores, incluyendo casos de alta calidad comunicacional, salidas con recomendaciones insuficientes y registros que requirieron revisión manual. Estos ejemplos permiten ilustrar cómo los resultados cualitativos se manifiestan en mensajes concretos generados por la arquitectura Multi-LLM-Agent.

7.8. Síntesis de resultados y limitaciones observadas

Los resultados obtenidos entregan evidencia inicial del funcionamiento integrado del sistema propuesto, desde la consolidación de datos meteorológicos hasta la generación automatizada de alertas agroclimáticas. En el horizonte D+1, el sistema predictivo alcanzó un CSI de 0,395 y una probabilidad de detección de 0,773 para eventos de helada dentro del corte operativo evaluado. Estos resultados indican que el ensamble logró identificar una proporción relevante de las heladas observadas, aunque manteniendo una cantidad importante de falsas alarmas.

La comparación operacional con MeteoChile, realizada sobre 3.034 pares comuna-fecha comunes, mostró mayores valores de CSI y POD para el sistema predictivo propuesto. Sin embargo, MeteoChile presentó una mayor precisión, asociada a una emisión más conservadora de alertas. En consecuencia, los resultados reflejan un compromiso entre maximizar la detección de eventos potencialmente riesgosos y reducir la cantidad de falsas alarmas. Esta comparación debe interpretarse como descriptiva, debido a que no se aplicaron pruebas formales de significancia estadística.

Respecto de la arquitectura Multi-LLM-Agent, el sistema procesó 199 alertas y logró generar respuestas JSON finales válidas para los tres proveedores después de aplicar mecanismos de validación, parseo y *fallback*. Llama 3 presentó la menor latencia promedio observada, seguido por Gemini y Claude. Claude y Llama 3 generaron respuestas estructuradas válidas sin requerir mecanismos de recuperación en la totalidad de las ejecuciones registradas, mientras que Gemini requirió corrección en una proporción reducida de casos.

La evaluación cualitativa retrospectiva permitió complementar estos resultados operacionales mediante el análisis de consistencia interpretativa, pertinencia agronómica, claridad comunicacional y accionabilidad de los mensajes generados. En el conjunto evaluado, Claude obtuvo el mayor puntaje global promedio, con 4,49, seguido estrechamente por Gemini, con 4,46. Llama 3 alcanzó un puntaje global promedio de 3,39. Claude destacó principalmente en consistencia interpretativa y claridad comunicacio-

CAPÍTULO 7 : RESULTADOS

nal, mientras que Gemini presentó los mayores puntajes en pertinencia agronómica y accionabilidad.

No obstante, estos resultados cualitativos deben interpretarse con cautela. La evaluación fue no realizada por agricultores, asesores agrícolas ni especialistas agrónomos independientes. Por esta razón, los puntajes obtenidos representan una aproximación exploratoria de calidad textual y coherencia contextual, y no una validación definitiva de utilidad práctica en terreno.

Las principales limitaciones identificadas son las siguientes:

1. El período de validación disponible se concentra principalmente en otoño e invierno, por lo que no permite evaluar de manera suficiente el desempeño del sistema frente a olas de calor.
2. Los horizontes D+3 a D+5 contienen una cantidad reducida de registros validados, por lo que sus indicadores deben interpretarse únicamente de manera descriptiva.
3. Solo una parte de las 52 comunas analizadas dispone de observaciones directas de estaciones INIA. Las comunas restantes dependen de la estrategia de asignación espacial y de fuentes meteorológicas complementarias.
4. La cobertura histórica consolidada presenta una brecha temporal entre los años 2017 y 2019, debido a la diferencia entre el período cubierto por Atlas Agroclimático y el inicio de los registros históricos utilizados desde Open-Meteo.
5. Los horizontes D+2 a D+5 utilizan una estrategia híbrida que incorpora variables futuras provenientes de Open-Meteo Forecast. Por ello, sus métricas no representan exclusivamente el desempeño del modelo predictivo propio.
6. La comparación entre proveedores LLM no corresponde completamente a una evaluación experimental pareada, debido a que no todos los proveedores procesaron exactamente el mismo conjunto total de alertas.

7. La calidad de los mensajes generados depende no solo del proveedor LLM, sino también de la completitud del contexto meteorológico y agrícola disponible, de la calidad de las recomendaciones estructuradas y del correcto funcionamiento de los mecanismos de parseo y recuperación.
8. La investigación no incorpora todavía retroalimentación de usuarios finales respecto de comprensión, utilidad práctica, adopción o impacto de las alertas en decisiones agrícolas reales.

En conjunto, los resultados muestran que el sistema es capaz de integrar fuentes meteorológicas heterogéneas, generar predicciones de eventos térmicos, validarlas frente a observaciones disponibles y activar una arquitectura Multi-LLM-Agent para transformar dichas predicciones en mensajes estructurados y comprensibles. La ampliación del período de operación, la incorporación de nuevas observaciones estacionales y una validación futura con usuarios agrícolas permitirán fortalecer la evaluación del sistema en distintos contextos climáticos, territoriales y productivos.

8. Conclusiones

El presente trabajo abordó la brecha existente entre la disponibilidad de información meteorológica y su utilización efectiva para la toma de decisiones agrícolas. Para ello, se desarrolló e implementó un sistema agroclimático basado en inteligencia artificial y una arquitectura Multi-LLM-Agent, orientado a integrar datos meteorológicos heterogéneos, anticipar eventos térmicos extremos, interpretar su posible impacto mediante umbrales agrícolas y comunicar alertas mediante WhatsApp.

El principal aporte del trabajo no corresponde únicamente al uso de modelos predictivos o de modelos de lenguaje de manera aislada, sino a la integración de ambos componentes dentro de un flujo operacional trazable. La solución articula fuentes históricas, observaciones recientes, pronósticos meteorológicos, reglas deterministas de activación, modelos de aprendizaje automático y agentes de lenguaje especializados. De esta manera, el sistema permite recorrer el ciclo completo desde la adquisición de datos hasta la entrega de una alerta comprensible para el usuario final.

8.1. Conclusiones generales

La integración jerárquica de Atlas Agroclimático, Open-Meteo histórico e INIA/AGROMET permitió construir una serie meteorológica consolidada de 368.082 pares únicos comuna–fecha para las 52 comunas de la Región Metropolitana. La política de prioridad aplicada privilegia observaciones directas de INIA/AGROMET cuando están disponibles, seguidas por Open-Meteo histórico y Atlas Agroclimático. Esta estrategia permitió combinar cobertura histórica, continuidad temporal y observaciones recientes de estación, manteniendo explícita la brecha histórica entre 2017 y 2019 y evitando la imputación artificial de variables críticas.

En el horizonte D+1, el sistema predictivo obtuvo un CSI de 0,395 y una probabilidad de detección de 0,773 para eventos de helada. En términos absolutos, el ensamble identificó 625 de las 809 heladas observadas durante el período evaluado. Estos resulta-

CAPÍTULO 8 : CONCLUSIONES

dos muestran que la solución posee capacidad para detectar una proporción importante de eventos potencialmente riesgosos, aspecto especialmente relevante en un sistema de alerta temprana. Sin embargo, la precisión de 0,447 y la presencia de 773 falsas alarmas evidencian que existe un compromiso entre privilegiar la detección de heladas y reducir alertas no confirmadas.

La evaluación de las estimaciones continuas de temperatura mostró un RMSE de 3,211 °C para temperatura mínima y 5,031 °C para temperatura máxima en D+1. Los resultados indican que la predicción de temperatura mínima presentó un desempeño relativamente más estable que la predicción de temperatura máxima. Asimismo, la corrección adaptativa de sesgo produjo una variación marginal en el RMSE de temperatura mínima dentro del corte evaluado. Por tanto, su principal valor actual corresponde a estabilizar operacionalmente desviaciones locales recientes y no a demostrar una mejora global concluyente.

La comparación operacional con MeteoChile se realizó sobre 3.034 pares comuna-fecha comunes. En dicho conjunto, el sistema propuesto obtuvo valores superiores de CSI y POD, mientras que MeteoChile presentó una precisión mayor. Esta diferencia refleja dos comportamientos operacionales distintos: el sistema propuesto prioriza la detección de una mayor proporción de heladas, mientras que MeteoChile emite una menor cantidad de alertas positivas y, por tanto, concentra una mayor proporción de aciertos entre las alertas emitidas. Debido a que no se aplicaron pruebas de significancia estadística, esta comparación debe entenderse como evidencia descriptiva y no como una demostración definitiva de superioridad.

La evaluación multi-horizonte mostró que D+1 es el único horizonte con una cantidad suficiente de observaciones validadas para sostener una interpretación cuantitativa más representativa. Los resultados de D+2 a D+5 deben considerarse preliminares, tanto por el menor tamaño muestral como por la naturaleza híbrida del flujo implementado. En dichos horizontes, las variables meteorológicas futuras provienen de Open-Meteo Forecast, mientras que el sistema propio aporta la clasificación contextual y la actua-

CAPÍTULO 8 : CONCLUSIONES

lización de características temporales. Por consiguiente, sus métricas no representan exclusivamente el desempeño del modelo de aprendizaje automático desarrollado.

La arquitectura Multi-LLM-Agent procesó 199 alertas durante el período analizado. El uso de reglas deterministas en el Agente Monitor permitió mantener la activación de alertas separada de la variabilidad propia de los modelos de lenguaje. Posteriormente, los agentes Razonador y Comunicador transformaron el contexto meteorológico y agrícola disponible en análisis estructurados y mensajes orientados a WhatsApp. Los tres proveedores evaluados lograron respuestas JSON finales válidas después de la aplicación de los mecanismos de validación, parseo y *fallback*. Llama 3 presentó la menor latencia promedio observada, seguido por Gemini y Claude.

La evaluación cualitativa retrospectiva de mensajes permitió complementar la evidencia operacional mediante las dimensiones de consistencia interpretativa, pertinencia agronómica, claridad comunicacional y accionabilidad. En la muestra evaluada, Claude alcanzó el mayor puntaje global promedio, con 4,49, seguido estrechamente por Gemini, con 4,46; Llama 3 obtuvo un puntaje global de 3,39. Claude destacó especialmente en consistencia interpretativa y claridad comunicacional, mientras que Gemini obtuvo los mayores puntajes de pertinencia agronómica y accionabilidad. No obstante, esta evaluación tiene carácter exploratorio, pues no constituye una validación realizada por agricultores, asesores agrícolas ni especialistas agronómicos independientes.

Finalmente, las pruebas de recepción de alertas mediante WhatsApp confirmaron el funcionamiento técnico extremo a extremo del sistema: una predicción almacenada puede activar una alerta, ser interpretada mediante la arquitectura Multi-LLM-Agent, transformarse en un mensaje legible y ser entregada a un dispositivo móvil. Este resultado no demuestra todavía adopción ni utilidad práctica en terreno, pero valida la factibilidad de utilizar un canal de comunicación cotidiano para acercar información agroclimática al usuario final.

8.2. Cumplimiento de objetivos

El objetivo principal de desarrollar e implementar un sistema basado en inteligencia artificial y arquitectura multiagente para integrar predicción, interpretación agrícola y comunicación de alertas fue cumplido dentro del alcance definido para esta investigación. La solución desarrollada integra los componentes necesarios para operar el flujo completo de información agroclimática y genera evidencia experimental respecto de su comportamiento predictivo, operacional y comunicacional.

Respecto de los objetivos específicos, se concluye lo siguiente:

1. Se integraron y procesaron datos provenientes de Atlas Agroclimático, INIA/AGROMET, Open-Meteo y MeteoChile, aplicando reglas de trazabilidad, prioridad de fuentes y tratamiento diferenciado de valores faltantes.
2. Se diseñó e implementó un sistema predictivo basado en un ensamble heterogéneo de siete clasificadores y regresores de temperatura. La evaluación disponible permite caracterizar el desempeño para heladas en D+1; la evaluación de olas de calor permanece como una línea en desarrollo debido a la baja disponibilidad de eventos durante el período analizado.
3. Se incorporaron umbrales térmicos operacionales para helada y ola de calor, complementados por umbrales agrícolas por cultivo y etapa fenológica en la capa de interpretación. Esta separación permitió distinguir entre la detección meteorológica de un evento y su contextualización como posible riesgo agrícola.
4. Se implementó una arquitectura Multi-LLM-Agent compuesta por los agentes Monitor, Razonador y Comunicador. La división de responsabilidades permitió desacoplar la activación determinista de alertas, la interpretación contextual y la construcción del mensaje final.
5. Se implementó la integración con WhatsApp como canal de comunicación y se

verificó técnicamente la recepción de alertas y comandos básicos en un dispositivo móvil durante pruebas de operación.

6. Se evaluó el sistema mediante métricas adecuadas para eventos poco frecuentes, incluyendo CSI, POD, precisión, F1-score, RMSE y matrices de confusión. Asimismo, se incorporó una evaluación operacional y cualitativa exploratoria de los proveedores de modelos de lenguaje.

8.3. Limitaciones del estudio

Los resultados obtenidos deben interpretarse considerando las limitaciones del período de operación y del alcance experimental. La evaluación disponible se concentra principalmente en otoño e invierno, por lo que la evidencia relativa a olas de calor es insuficiente para caracterizar el desempeño del sistema frente a temperaturas máximas extremas. Del mismo modo, los horizontes D+3 a D+5 contienen una cantidad reducida de registros validados, lo que limita la posibilidad de establecer conclusiones robustas respecto de esos plazos.

La cobertura territorial se construyó para las 52 comunas de la Región Metropolitana, pero solo una parte de ellas dispone de observaciones directas de estaciones INIA/AGROMET. En las comunas restantes, la estrategia depende de la asignación espacial de estaciones cercanas y de fuentes complementarias. Esta aproximación permite ampliar la cobertura operacional, pero no reemplaza mediciones locales ni representa necesariamente la variabilidad microclimática de cada predio agrícola.

La serie consolidada presenta una brecha temporal entre 2017 y 2019. Además, la comparación con MeteoChile es descriptiva y no incorporó pruebas estadísticas de hipótesis. Por otra parte, los horizontes D+2 a D+5 utilizan variables futuras provenientes de Open-Meteo Forecast, por lo que su desempeño representa el comportamiento del flujo híbrido y no exclusivamente el del modelo propio.

En la capa Multi-LLM-Agent, los proveedores no procesaron exactamente el mismo

conjunto total de alertas, por lo que las comparaciones de latencia y comportamiento general no corresponden plenamente a un experimento pareado. La evaluación cualitativa de mensajes, aunque estructurada y realizada sin visibilizar el proveedor durante la calificación, no puede reemplazar una validación con especialistas agrícolas o usuarios finales.

8.4. Trabajo futuro

A partir de los resultados y limitaciones identificadas, se proponen las siguientes líneas de trabajo futuro:

1. Ampliar el período de operación y validación para incluir primavera y verano, con el propósito de evaluar adecuadamente la detección de olas de calor y disponer de una base más amplia para los horizontes D+2 a D+5.
2. Incorporar nuevas estaciones meteorológicas, sensores prediales o información topográfica y de uso de suelo, con el objetivo de mejorar la representación de microclimas y reducir la dependencia de asignaciones espaciales por proximidad.
3. Refinar la estrategia de reducción de falsas alarmas mediante calibración dinámica de umbrales, análisis de costos de decisión y mecanismos diferenciados según comuna, cultivo, etapa fenológica y nivel de riesgo.
4. Ejecutar una evaluación experimental pareada de proveedores LLM, utilizando exactamente los mismos contextos de entrada y un identificador común por alerta, para comparar de forma más controlada latencia, robustez y calidad de salida.
5. Realizar una validación cualitativa con agricultores, asesores técnicos y especialistas agronómicos. Esta evaluación permitiría medir comprensión, pertinencia, utilidad práctica, confianza y adopción de las alertas en condiciones reales de uso.

CAPÍTULO 8 : CONCLUSIONES

6. Incorporar mecanismos de retroalimentación del usuario final, tales como confirmación de recepción, reporte de ocurrencia observada y valoración de utilidad, con el fin de enriquecer la trazabilidad y apoyar la mejora continua del sistema.
7. Extender la solución a otras regiones agrícolas de Chile, incorporando nuevas fuentes locales, umbrales específicos y procedimientos de validación territorial antes de su despliegue operativo.

En síntesis, esta investigación demuestra la factibilidad de integrar datos meteorológicos, aprendizaje automático, conocimiento agrícola estructurado y modelos de lenguaje en una plataforma única de alertas agroclimáticas. La solución desarrollada no pretende reemplazar el criterio de los agricultores ni la asesoría agronómica especializada, sino aportar una herramienta de apoyo que permita transformar información climática dispersa en alertas oportunas, trazables y comprensibles. La continuidad de la operación, la ampliación de la validación estacional y la incorporación de usuarios reales permitirán avanzar desde un prototipo funcional hacia un sistema de apoyo a la decisión con mayor madurez operacional y territorial.

Referencias

- Ahumada Avendaño, T. A. (2026). Repositorio técnico de *Multi-LLM-Agent para la Agricultura Inteligente*.
- Almeida, V. A., Anochi, J. A., Rozante, J. R., & de Campos Velho, H. F. (2025). Machine Learning with Voting Committee for Frost Prediction. *Meteorology*, 4(1), 6. <https://doi.org/10.3390/meteorology4010006>
- Bhakare, S., Matiu, M., Crespi, A., & Zardi, D. (2025). Spatial Downscaling of Daily Temperature Minima Using Machine Learning Methods and Application to Frost Forecasting in Two Alpine Valleys. *Atmosphere*, 16(1), 38. <https://doi.org/10.3390/atmos16010038>
- Born, L., Prager, S., Ramirez-Villegas, J., & Imbach, P. (2021). A global meta-analysis of climate services and decision-making in agriculture. *Climate Services*, 22, 100231. <https://doi.org/10.1016/j.cliser.2021.100231>
- Cegnar, T., Boogaard, H., Finklele, K., Lalic, B., Raymond, J., Lifka, S., Schultz, D. M., & Tarchiani, V. (2023). Toward effective communication of agrometeorological services. *Advances in Science and Research*, 20, 9-16. <https://doi.org/10.5194/asr-20-9-2023>
- CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS). (2019). *Climate Services and Farmer Decision-Making*. <https://cgspace.cgiar.org/items/84cc82f7-b48a-475c-8079-2a8db88c14ee>
- Diedrichs, A. L., Bromberg, F., Dujovne, D., Brun-Laguna, K., & Watteyne, T. (2018). Prediction of Frost Events Using Machine Learning and IoT Sensing Devices. *IEEE Internet of Things Journal*, 5(6), 4589-4597. <https://doi.org/10.1109/JIOT.2018.2867333>
- Dirección Meteorológica de Chile. (2026). Portal de Servicios Climáticos. <https://climatologia.meteochile.gob.cl/>
- Ferro, C. A. T., & Stephenson, D. B. (2011). Extremal Dependence Indices: Improved Verification Measures for Deterministic Forecasts of Rare Binary Events.

Weather and Forecasting, 26, 699-713. <https://doi.org/10.1175/WAF-D-10-05030.1>

Food and Agriculture Organization of the United Nations. (2005). Frost damage: Physiology and critical temperatures. <https://www.fao.org/4/y7223e/y7223e0a.htm>

Food and Agriculture Organization of the United Nations & World Meteorological Organization. (2021). *Global outlook on climate services in agriculture: Investment opportunities to reach the last mile*. Food, Agriculture Organization of the United Nations y World Meteorological Organization. <https://www.fao.org/3/cb6941en/cb6941en.pdf>

Fuentes, M., Campos, C., & García-Loyola, S. (2018). Application of Artificial Neural Networks to Frost Detection in Central Chile Using the Next Day Minimum Air Temperature Forecast. *Chilean Journal of Agricultural Research*, 78(3), 327-338. <https://doi.org/10.4067/S0718-58392018000300327>

Fundación para la Innovación Agraria. (2023a). Chile hacia una agricultura sustentable y 4.0. <https://opia.fia.cl/601/w3-article-123471.html>

Fundación para la Innovación Agraria. (2023b). El reporte del vigilante: Sistemas de control y manejo de heladas. <https://opia.fia.cl/601/w3-article-78093.html>

Hansen, J., Furlow, J., Goddard, L., Nissan, H., Vaughan, C., et al. (2019). *Scaling Climate Services to Enable Effective Adaptation Action*. Global Commission on Adaptation. <https://gca.org/reports/scaling-climate-services/>

Instituto de Investigaciones Agropecuarias. (2025). Nueva estación meteorológica entregará datos climáticos clave desde el corazón de Chillán. <https://www.inia.cl/2025/11/24/nueva-estacion-meteorologica-de-inia-entregara-datos-climaticos-clave-desde-el-corazon-de-chillan/>

Instituto de Investigaciones Agropecuarias. (2026). Agrometeorología. <https://agrometeorologia.cl/>

Kapoor, S., & Narayanan, A. (2023). Leakage and the Reproducibility Crisis in Machine-Learning-Based Science. *Patterns*, 4(9), 100804. <https://doi.org/10.1016/j.patter.2023.100804>

- Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Brenner, M. P., Hoyer, S., & Dueben, P. D. (2024). Neural general circulation models for weather and climate. *Nature*, 632, 1060-1066. <https://doi.org/10.1038/s41586-024-07744-y>
- Koopman, B., Mourad, A., Li, H., van der Vegt, A., Zhuang, S., Gibson, S., Dang, Y., Lawrence, D., & Zuccon, G. (2024). AgAsk: an Agent to Help Answer Farmer's Questions from Scientific Documents. *International Journal on Digital Libraries*, 25(4), 569-584. <https://doi.org/10.1007/s00799-023-00369-y>
- Kumar, A., Muthuramalingam, P., Kumar, R., Tiwari, S., Verma, L., Park, S., & Shin, H. (2025). Adapting Crops to Rising Temperatures: Understanding Heat Stress and Plant Resilience Mechanisms. *International Journal of Molecular Sciences*, 26(21), 10426. <https://doi.org/10.3390/ijms262110426>
- Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., & Battaglia, P. (2023). Learning Skillful Medium-Range Global Weather Forecasting. *Science*, 382(6677), 1416-1421. <https://doi.org/10.1126/science.adi2336>
- Ministerio de Agricultura de Chile. (2024). *Red Agroclimática Nacional (AGROMET)*. <https://agromet.cl>
- Ministerio del Medio Ambiente de Chile. (2026). Repositorio Climático: Comunas. https://arclim.mma.gob.cl/features/repositorio_package/comunas/
- Möller, V., van Diemen, R., Matthews, J. B. R., Méndez, C., Semenov, S., Fuglestad, J. S., & Reisinger, A. (2022). Annex II: Glossary. En H.-O. Pörtner, D. C. Roberts, M. Tignor, E. S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem & B. Rama (Eds.), *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 2897-2930). Cambridge University Press. <https://doi.org/10.1017/9781009325844.029>

- Möller-Acuña, P., Ahumada-García, R., & Reyes-Suárez, J. A. (2017). Machine Learning for Prediction of Frost Episodes in the Maule Region of Chile. En S. Ochoa, P. Singh & J. Bravo (Eds.), *Ubiquitous Computing and Ambient Intelligence* (pp. 715-720, Vol. 10586). Springer, Cham. https://doi.org/10.1007/978-3-319-67585-5_69
- Noh, I., Doh, H.-W., Kim, S.-O., Kim, S.-H., Shin, S., & Lee, S.-J. (2021). Machine Learning-Based Hourly Frost-Prediction System Optimized for Orchards Using Automatic Weather Station and Digital Camera Image Data. *Atmosphere*, *12*(7), 846. <https://doi.org/10.3390/atmos12070846>
- Open-Meteo. (2026a). Historical Weather API. <https://open-meteo.com/en/docs/historical-weather-api>
- Open-Meteo. (2026b). Weather Forecast API. <https://open-meteo.com/en/docs>
- Pérez Tárraga, J., Castillo-Cara, M., Arias-Antúnez, E., et al. (2025). Frost forecasting through machine learning algorithms. *Earth Science Informatics*, *18*, 183. <https://doi.org/10.1007/s12145-025-01710-6>
- Price, I., Sanchez-Gonzalez, A., et al. (2024). Accurate medium-range global weather forecasting with generative diffusion models. *Nature*. <https://doi.org/10.1038/s41586-024-08252-9>
- Richetti, J., Sadras, V. O., He, D., Leske, B., Hu, P., Beletse, Y., Cossani, C. M., Nguyen, H., Zheng, B., Deery, D. M., Dreccer, M. F., Whish, J., & Lilley, J. (2025). Challenges in modelling the impact of frost and heat stress on the yield of cool-season annual grain crops. *Frontiers in Plant Science*, *16*, 1613432. <https://doi.org/10.3389/fpls.2025.1613432>
- Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE*, *10*(3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- Selvaraju, R., Khanh, N. D., Dalida Jr., L. U., Alvina, P., Chung, P., Tejada, E., Mendoza Luzcúber, O. J., Mori, R., Madima, T., & Gautam, B. R. (2013). Localizing Climate Information Services for Agriculture [Accessed 2026-04-19]. *WMO*

Bulletin, 62(Special Issue), 12-15. <https://wmo.int/media/magazine-article/localizing-climate-information-services-agriculture>

Sharma, U., Chetri, P., Minocha, S., Roy, A., Holker, T., Patt, A., & Joerin, J. (2021). Do Phone-Based Short Message Services Improve the Uptake of Agri-met Advice by Farmers? A Case Study in Haryana, India. *Climate Risk Management*, 33, 100321. <https://doi.org/10.1016/j.crm.2021.100321>

Shin, Y.-S., Lee, H.-A., Park, S.-H., Han, Y.-K., Shim, K.-M., & Han, S.-J. (2025). Establishment and Operation of an Early Warning Service for Agrometeorological Disasters Customized for Farmers and Extension Workers at Metropolitan Scale. *Atmosphere*, 16(3), 291. <https://doi.org/10.3390/atmos16030291>

Sotomayor, O., Nunes, S., & Rodrigues, M. (2023). *La digitalización al servicio de la transición agroecológica*. Comisión Económica para América Latina y el Caribe. Santiago. Consultado el 21 de abril de 2026, desde https://www.cepal.org/sites/default/files/news/files/23-00546-folleto-la_digitalizacion_web_0.pdf

Subsecretaría de Telecomunicaciones & Cadem. (2024). *Informe final X Encuesta sobre acceso, uso y usuarios de internet 2023*. https://www.subtel.gob.cl/wp-content/uploads/2024/03/Informe_Final_Acceso_y_uso_Internet_2023_VF.pdf

Subsecretaría de Telecomunicaciones & Cadem. (2025). *Informe final XI Encuesta sobre acceso, uso y usuarios de internet 2024*. <https://www.subtel.gob.cl/wp-content/uploads/2025/02/Informe-Final-Subtel-Acceso-y-Uso-Internet-2024.pdf>

Talsma, C. J., Solander, K. C., Mudunuru, M. K., Crawford, B., & Powell, M. R. (2023). Frost Prediction Using Machine Learning and Deep Neural Network Models. *Frontiers in Artificial Intelligence*, 5, 963781. <https://doi.org/10.3389/frai.2022.963781>

Tran, B. L., Tseng, W. C., & Chen, C. C. (2025). Climate change impacts on crop yields across temperature rise thresholds and climate zones. *Scientific Reports*, 15, 23424. <https://doi.org/10.1038/s41598-025-07405-8>

- Universidad de Chile. (2026). Atlas Agroclimático de Chile. <https://agronomia.uchile.cl/publicaciones/139988/atlas-agroclimatico-de-chile>
- Vaughan, C., & Dessai, S. (2014). Climate services for society: origins, institutional arrangements, and design elements for an evaluation framework. *WIREs Climate Change*, 5(5), 587-603. <https://doi.org/10.1002/wcc.290>
- World Bank. (2023). Climate-Smart Agriculture. <https://www.worldbank.org/en/topic/climate-smart-agriculture>
- World Meteorological Organization. (2023). What are Climate Services? <https://gfcs.wmo.int/site/global-framework-climate-services-gfcs/what-are-climate-services>
- World Meteorological Organization. (2025). FAO and WMO report highlights extreme heat risks for agriculture. <https://wmo.int/media/news/fao-and-wmo-report-highlights-extreme-heat-risks-agriculture>
- Xia, F., Pan, J., Zhong, R., Liu, W., Chen, T., Zhu, L., Yang, Y., & Lin, T. (2026). ChatCEA: a Knowledge-Driven Intelligent Service Agent for Controlled Environment Agriculture. *Computers and Electronics in Agriculture*, 248, 111733. <https://doi.org/10.1016/j.compag.2026.111733>
- Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao, Y., Bassu, S., Ciais, P., Durand, J.-L., Elliott, J., Ewert, F., Janssens, I. A., Li, T., Lin, E., Liu, Q., Martre, P., Müller, C., ... Asseng, S. (2017). Temperature Increase Reduces Global Yields of Major Crops in Four Independent Estimates. *Proceedings of the National Academy of Sciences*, 114(35), 9326-9331. <https://doi.org/10.1073/pnas.1701762114>

Anexos

A. Infraestructura Tecnológica

Este anexo presenta los componentes tecnológicos, configuraciones experimentales y resultados complementarios utilizados en la implementación y evaluación del sistema propuesto. Su objetivo es complementar la información presentada en los Capítulos 5, 6 y 7, proporcionando antecedentes técnicos que contribuyen a la trazabilidad y reproducibilidad de la solución.

A.1. Arquitectura física

La solución fue implementada mediante una arquitectura basada en servicios distribuidos desplegados sobre contenedores. Esta organización permite desacoplar los componentes de adquisición de datos, almacenamiento, predicción, interpretación, automatización y comunicación, facilitando el mantenimiento, la escalabilidad y la evolución futura del sistema.

La Figura A.1 presenta la arquitectura física implementada.

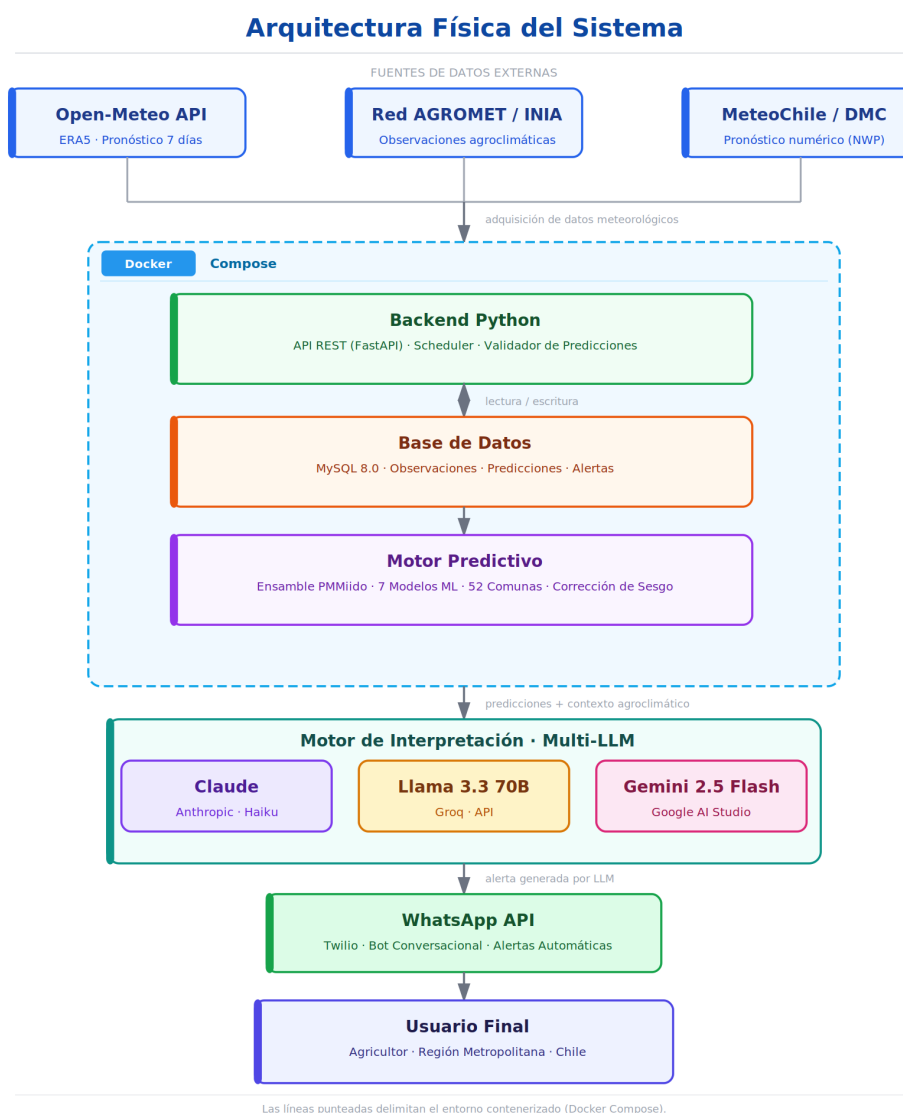


Figura A.1: Arquitectura física del sistema.

A.2. Servicios implementados

La Tabla 27 resume los principales servicios utilizados por la plataforma.

Tabla 27: *Servicios utilizados por la plataforma.*

Servicio	Tecnología	Propósito
Backend	Python / FastAPI	Procesamiento, consulta de predicciones, ejecución del pipeline y exposición de servicios internos.
Base de datos	MySQL	Almacenamiento de observaciones, predicciones, métricas, umbrales, alertas y registros de ejecución.
Motor predictivo	Python / bibliotecas ML	Entrenamiento y ejecución de modelos de clasificación y regresión.
Arquitectura Multi-LLM-Agent	APIs de proveedores LLM	Interpretación de alertas, generación de recomendaciones y construcción de mensajes.
WhatsApp Gateway	API de mensajería / Twilio	Distribución de alertas e interacción básica con usuarios.
Scheduler	APScheduler	Automatización de ingesta, validación, predicción, alertas y reentrenamiento.
Despliegue	Docker / Docker Compose	Ejecución reproducible y desacoplada de servicios.

A.3. Configuración de despliegue

La solución fue desplegada utilizando contenedores Docker y Docker Compose para garantizar consistencia entre los ambientes de desarrollo y operación. La separación de servicios permite mantener aislados el backend, la base de datos, el motor predictivo y los componentes auxiliares de comunicación.

La automatización operacional se configura en la zona horaria **America/Santiago**. Los archivos de configuración, variables de entorno no sensibles y scripts de despliegue se mantienen en el repositorio del proyecto. Las claves de API, credenciales de base de datos y otros secretos no se incluyen en este documento.

B. Modelo Físico de Datos

Este apartado describe el modelo físico de datos utilizado para soportar el almacenamiento de observaciones meteorológicas, predicciones, alertas y trazabilidad de las ejecuciones del sistema.

B.1. Diagrama entidad–relación

La Figura B.1 presenta el modelo físico de datos implementado.



Figura B.1: Modelo físico de base de datos.

B.2. Entidades principales

- **Comunas:** almacena la información territorial de las comunas consideradas dentro de la zona de estudio.
- **Estaciones meteorológicas:** contiene identificadores, ubicación y fuente asocia-

da a las estaciones disponibles.

- **Observaciones meteorológicas:** registra las variables históricas y recientes obtenidas desde Atlas Agroclimático, INIA/AGROMET y Open-Meteo.
- **Predicciones:** almacena las estimaciones de temperatura, probabilidades, votos del ensamble, horizontes de predicción y resultados de validación posterior.
- **Umbrales térmicos:** contiene valores de referencia para interpretar el riesgo agrícola por cultivo y etapa fenológica.
- **Alertas:** registra las alertas activadas, su estado de envío, mensajes asociados y resultados operacionales.
- **Ejecuciones LLM:** conserva proveedor, modelo, contexto estructurado, salida del razonador, mensaje final, latencia y resultado del parseo.

Las definiciones completas de atributos, tipos de datos, relaciones y restricciones se encuentran en el esquema físico de base de datos incluido en el repositorio técnico del proyecto.

C. Automatización del Pipeline

El sistema fue diseñado para operar de forma automática, permitiendo actualizar información meteorológica, validar predicciones previas, generar nuevos pronósticos y emitir alertas sin intervención manual continua.

C.1. Flujo de automatización

La Figura C.1 presenta el flujo automatizado implementado.

Pipeline Automatizado del Sistema

Ejecución programada, secuencial y sin intervención manual

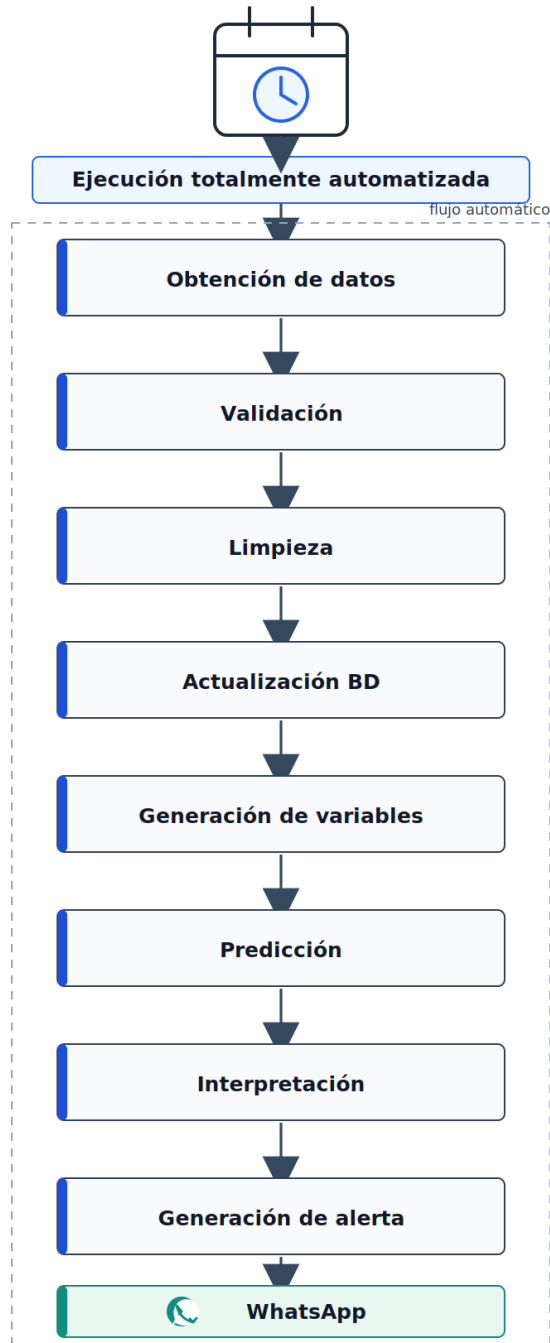


Figura C.1: Pipeline automatizado del sistema.

C.2. Etapas del pipeline

1. Obtención de observaciones meteorológicas recientes.
2. Validación de predicciones emitidas previamente.
3. Actualización de registros meteorológicos y métricas de desempeño.
4. Preparación de datos y generación de variables derivadas.
5. Actualización de variables de pronóstico desde fuentes externas.
6. Ejecución de modelos predictivos para D+1 a D+5.
7. Activación determinista de alertas según umbrales y votos del ensamble.
8. Interpretación agroclimática mediante la arquitectura Multi-LLM-Agent.
9. Generación y envío de mensajes mediante WhatsApp.

C.3. Programación de tareas

La ejecución periódica de los procesos se implementó mediante APScheduler. El pipeline respeta el orden de dependencia entre ingesta, validación, actualización de variables futuras, predicción y comunicación. El proceso de reentrenamiento se ejecuta de manera periódica cuando existen suficientes observaciones válidas por comuna.

D. Integración con WhatsApp

La comunicación con los usuarios se realiza mediante WhatsApp, seleccionado por su amplia adopción, facilidad de uso y disponibilidad en dispositivos móviles. La integración permite distribuir alertas y atender comandos básicos de consulta.

D.1. Arquitectura de comunicación

La Figura D.1 presenta la arquitectura utilizada para el envío de alertas.

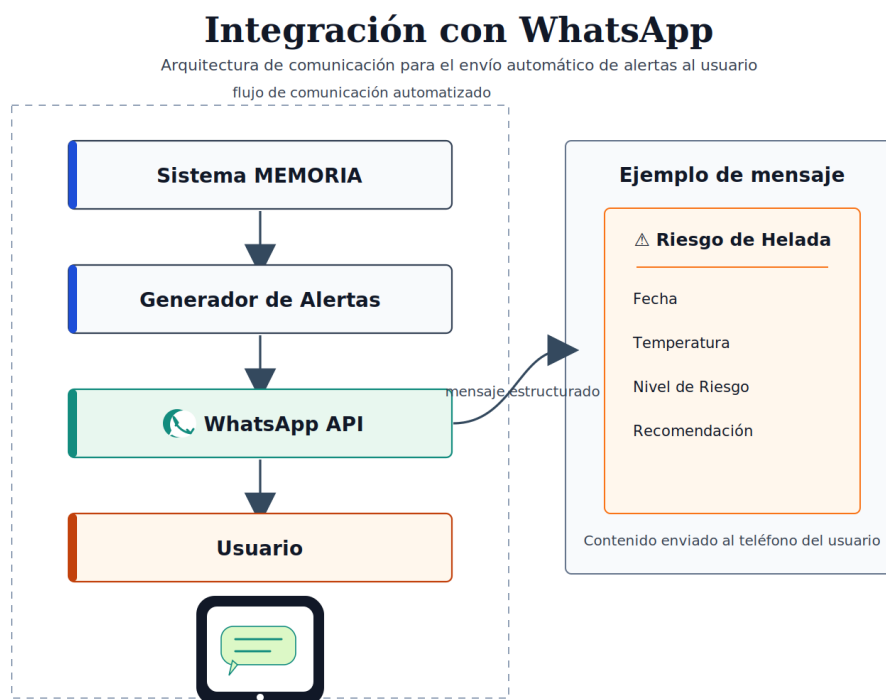


Figura D.1: Arquitectura de integración con WhatsApp.

D.2. Proceso de envío

El envío de alertas considera las siguientes etapas:

1. Activación de una alerta a partir de reglas deterministas.
2. Preparación del contexto meteorológico y agrícola.
3. Generación de análisis estructurado mediante el Agente Razonador.
4. Construcción de mensaje corto y mensaje explicativo mediante el Agente Comunicador.

5. Validación del formato de salida.
6. Envío mediante la API de mensajería.
7. Registro de auditoría, estado de entrega y respuesta del proveedor.

D.3. Formato de mensajes

Los mensajes generados pueden incluir tipo de evento, comuna o zona afectada, fecha esperada de ocurrencia, temperatura relevante, nivel de riesgo, nivel de confiabilidad y recomendaciones preventivas generales. La evidencia de entrega de mensajes mediante WhatsApp se presenta en el Capítulo 7 como parte de la validación técnica extremo a extremo.

E. Configuración Experimental de los Modelos

Este apartado presenta las variables, modelos, reglas de entrenamiento y condiciones de ejecución empleadas para generar los resultados experimentales.

E.1. Variables empleadas

La Tabla 28 presenta las variables utilizadas por el sistema predictivo. Las variables meteorológicas críticas corresponden a temperatura mínima, temperatura máxima y humedad relativa. Las variables derivadas se construyen exclusivamente con información disponible antes de la fecha de predicción.

Tabla 28: Variables utilizadas por el sistema predictivo.

Variable	Tipo	Unidad	Origen	Tratamiento de faltantes	Uso
temp_min	Númerica continua	°C	Atlas, INIA y Open-Meteo histórico	Se excluyen registros incompletos en entrenamiento	Predictor
temp_max	Númerica continua	°C	Atlas, INIA y Open-Meteo histórico	Se excluyen registros incompletos en entrenamiento	Predictor
humedad_rel	Númerica continua	%	Atlas, INIA y Open-Meteo histórico	Se excluyen registros incompletos en entrenamiento; fallback operativo en inferencia	Predictor
precipitacion	Númerica continua	mm	Atlas, INIA y Open-Meteo histórico	Imputación con 0 cuando no existe registro	Predictor
radiacion_solar	Númerica continua	MJ m ⁻²	Atlas, INIA y Open-Meteo histórico	Interpolación lineal y relleno en extremos	Predictor
dia_sin	Derivada	Adimensional	Día del año	Sin nulos	Predictor
dia_cos	Derivada	Adimensional	Día del año	Sin nulos	Predictor
rolling_max7	Derivada	°C	Temperatura máxima histórica	Calculada con datos previos	Predictor
rolling_min3	Derivada	°C	Temperatura mínima histórica	Calculada con datos previos	Predictor
target_helada	Objetivo binario	{0,1}	temp_min del día siguiente	No aplica	Entrenamiento
target_ola_calor	Objetivo binario	{0,1}	temp_max del día siguiente	No aplica	Entrenamiento

Las variables de estacionalidad se calcularon mediante:

$$\text{dia_sin} = \sin\left(\frac{2\pi \cdot \text{DOY}}{365,25}\right)$$

$$\text{dia_cos} = \cos\left(\frac{2\pi \cdot \text{DOY}}{365,25}\right)$$

donde DOY corresponde al día del año. Las variables de ventana temporal se calcularon utilizando exclusivamente registros anteriores al día evaluado:

$$\text{rolling_max7}_t = \frac{1}{7} \sum_{i=1}^7 \text{temp_max}_{t-i}$$

$$\text{rolling_min3}_t = \min(\text{temp_min}_{t-1}, \text{temp_min}_{t-2}, \text{temp_min}_{t-3})$$

El uso de desplazamiento temporal evita incorporar información futura durante el entrenamiento o la inferencia.

E.2. Modelos y configuración de entrenamiento

La Tabla 29 presenta los siete clasificadores incorporados al ensamble heterogéneo. Los modelos se entrenan de manera independiente para cada combinación de comuna

y tipo de evento térmico.

Tabla 29: Modelos incluidos en el ensamble de clasificación.

Modelo	Biblioteca	Configuración principal	Manejo de desbalance	Semilla
XGBoost	xgboost	n_estimators=100, max_depth=5, eval_metric=logloss	scale_pos_weight	42
Random Forest	scikit-learn	n_estimators=100, n_jobs=-1	class_weight=balanced	42
LightGBM	lightgbm	verbosity=-1, n_jobs=-1	class_weight=balanced	42
CatBoost	catboost	verbose=0; configuración persistida en el artefacto del modelo	auto_class_weights=Balanced	42
KNN	scikit-learn	n_neighbors=5	Sin ponderación explícita	No aplica
SVM	scikit-learn	Kernel RBF, probability=True	class_weight=balanced	42
MLP	scikit-learn	Capas (64, 32), max_iter=500, early_stopping=True	Sin ponderación explícita	42

Para la regresión de temperatura mínima y máxima se utilizó RandomForestRegressor con n_estimators=200 y n_jobs=-1.

La partición temporal se realizó sin mezcla aleatoria: 60 % de la serie se empleó para ajuste, 20 % para calibración de umbrales y 20 % para evaluación final. Para cada clasificador se seleccionó el umbral que maximiza CSI dentro del intervalo 0,05 a 0,95, con incrementos de 0,01. La predicción final del ensamble se activa cuando al menos cuatro de los siete clasificadores emiten un voto positivo.

E.3. Entorno de ejecución y reproducibilidad

Tabla 30: Entorno de ejecución utilizado para generar los resultados experimentales.

Elemento	Configuración utilizada
Lenguaje	Python 3.12
Sistema operativo del servidor	Ubuntu 24.04.4 LTS
Base de datos	MySQL desplegado en contenedor Docker
Zona horaria operacional	America/Santiago
Bibliotecas de ML	scikit-learn, XGBoost, LightGBM y CatBoost
Despliegue	Docker / Docker Compose
Semilla global	42
Generación de resultados	Script versionado de generación de resultados experimentales

Para favorecer la reproducibilidad del trabajo, el código fuente, los scripts de generación de resultados, las versiones de dependencias y los archivos de configuración no sensibles se mantienen versionados en el repositorio técnico del proyecto:

<https://github.com/Tabby2109/memoria-pmm>

Las credenciales, claves de API, archivos `.env` y otros elementos sensibles no se incluyen en dicho repositorio.

F. Catálogo de Variables Meteorológicas y Trazabilidad

Esta sección distingue las variables disponibles por fuente de aquellas que fueron efectivamente seleccionadas para el modelo predictivo. Variables como viento, dirección de viento, presión atmosférica o punto de rocío pueden estar disponibles en fuentes

puntuales, pero no forman parte del vector final de nueve predictores utilizado para los resultados reportados en esta memoria.

F.1. Variables disponibles por fuente

Tabla 31: Variables meteorológicas disponibles por fuente.

Fuente	Variables principales	Cobertura	Uso en el sistema
Atlas Agroclimático	Temperatura mínima, temperatura máxima, humedad relativa, precipitación y radiación solar	2004–2016	Base histórica de entrenamiento
Open-Meteo histórico	Temperatura mínima, temperatura máxima, humedad relativa, precipitación y radiación solar	2020–2026	Complemento temporal y espacial
INIA/AGROMET	Temperatura mínima, temperatura máxima, humedad relativa, precipitación, radiación y viento	2020–2026; 20 comunas con observaciones directas	Validación, actualización y reentrenamiento
Open-Meteo Forecast	Temperatura, humedad, precipitación y radiación futuras	D+1 a D+5	Variables auxiliares para predicción multi-horizonte
MeteoChile	Pronóstico oficial de temperatura	Según disponibilidad pública	Referencia operacional externa

F.2. Política de fusión

La construcción de la serie consolidada sigue la prioridad:

INIA > Open-Meteo histórico > Atlas Agroclimático

Para cada par comuna–fecha se conserva el registro disponible de mayor prioridad. Atlas Agroclimático se utiliza como base histórica, Open-Meteo histórico complementa el período reciente e INIA reemplaza las fuentes previas cuando existe observación directa de estación. La fusión no incorpora imputación artificial de temperatura mínima, temperatura máxima ni humedad relativa cuando dichas variables críticas no se encuentran disponibles.

La cobertura consolidada presenta una brecha temporal entre 2017 y 2019, debido a que la base Atlas Agroclimático utilizada finaliza en 2016 y los registros históricos seleccionados desde Open-Meteo comienzan en 2020. Esta brecha se mantiene explícita y no fue completada mediante imputación artificial de variables críticas.

G. Protocolo Reproducible de Evaluación Multi-LLM-Agent

G.1. Contexto entregado al Agente Razonador

Cada alerta se procesa a partir de un contexto estructurado que puede incluir:

1. Tipo de alerta detectada.
2. Comunas afectadas.
3. Temperatura mínima y máxima previstas.
4. Probabilidad o nivel de consenso del ensamble.
5. Historial térmico reciente.
6. Umbrales críticos por cultivo y etapa fenológica.
7. Métricas recientes de confiabilidad del sistema.
8. Ventana horaria estimada cuando corresponde a helada.

El Agente Monitor activa el flujo mediante reglas deterministas. El Agente Razonador recibe el contexto ya preparado y devuelve un análisis estructurado. Posteriormente, el Agente Comunicador transforma el resultado en mensajes cortos y largos adecuados para WhatsApp.

G.2. Esquema de salida esperado

Agente Razonador:

```
{  
  "resumen_riesgo": "texto breve",
```

```
"cultivos_riesgo": ["cultivo 1", "cultivo 2"],
"tendencia_temp": "descripción",
"confiabilidad": "alta|media|baja",
"recomendaciones": ["acción 1", "acción 2", "acción 3"],
"analisis_completo": "párrafo técnico"
}
```

Agente Comunicador:

```
{
  "mensaje_corto": "máximo 160 caracteres",
  "mensaje_largo": "mensaje explicativo de 3 a 5 oraciones",
  "emoji_tipo": "símbolo correspondiente al evento"
}
```

G.3. Rúbrica y procedimiento de evaluación cualitativa retrospectiva

La calidad textual y contextual de los mensajes generados fue evaluada mediante una revisión retrospectiva, utilizando una rúbrica ordinal de cinco niveles. La revisión se efectuó sobre el mensaje final y el contexto meteorológico y agrícola disponible al momento de su generación. Durante la asignación de puntajes, la identidad del proveedor se ocultó para reducir sesgos asociados al nombre del modelo.

Tabla 32: *Rúbrica cualitativa utilizada para evaluar mensajes generados por la arquitectura Multi-LLM-Agent.*

Dimensión	1 — Deficiente	3 — Adecuado	5 — Alto desempeño
Consistencia interpretativa	Contradice datos o inventa información crítica	Refleja los datos principales con omisiones menores	Integra correctamente evento, temperaturas, riesgo y contexto disponible
Pertinencia agronómica	Recomendaciones incorrectas, irrelevantes o poco prudentes	Recomendaciones generales razonables	Recomendaciones prudentes, pertinentes y vinculadas al evento y contexto agrícola
Claridad comunicacional	Mensaje confuso, ambiguo o excesivamente técnico	Mensaje comprensible, aunque mejorable	Mensaje claro, breve, priorizado y adecuado para usuarios no especializados
Accionabilidad	No propone acciones concretas o propone acciones poco aplicables	Incluye acciones generales	Propone acciones concretas, prudentes y aplicables en un contexto agrícola general

La robustez de formato JSON se evaluó automáticamente mediante el parser del sistema y no forma parte de la rúbrica cualitativa. Los puntajes de consistencia, pertinen-

cia, claridad y accionabilidad representan una evaluación exploratoria; no sustituyen una validación realizada por agricultores, asesores agrícolas o especialistas agrónomos independientes.

H. Resultados Complementarios

H.1. Métricas de helada por comuna

La siguiente tabla presenta el detalle de métricas de detección de helada por comuna para el horizonte D+1. Las métricas fueron calculadas desde los conteos individuales de TP, FP, FN y TN, evitando promediar indicadores agregados.

Tabla 33: Métricas de helada por comuna para D+1.

Comuna	CSI	POD	Precisión	F1	TP	FP	FN	TN	N	Heladas reales
Maipú	0,5965	0,8947	0,6415	0,7472	34	19	4	6	63	38
Lampa	0,5818	0,8205	0,6667	0,7356	32	16	7	8	63	39
Peñalolén	0,5385	1,0000	0,5385	0,7000	14	12	0	37	63	14
Macul	0,5200	0,9286	0,5417	0,6842	13	11	1	38	63	14
Quinta Normal	0,5200	0,9286	0,5417	0,6842	13	11	1	38	63	14
San José de Maipo	0,5135	0,8261	0,5758	0,6786	19	14	4	26	63	23
Paine	0,4902	0,8621	0,5319	0,6579	25	22	4	11	62	29
Alhué	0,4800	0,8571	0,5217	0,6486	24	22	4	6	56	28
Pudahuel	0,4722	0,7727	0,5484	0,6415	17	14	5	27	63	22
Buin	0,4545	0,8065	0,5102	0,6250	25	24	6	8	63	31
La Pintana	0,4490	0,8462	0,4889	0,6197	22	23	4	16	65	26
Colina	0,4483	0,7222	0,5417	0,6191	13	11	5	35	64	18
Melipilla	0,4464	0,8333	0,4902	0,6173	25	26	5	6	62	30
Isla de Maipo	0,4375	0,9130	0,4565	0,6087	21	25	2	2	50	23
Cerrillos	0,4375	1,0000	0,4375	0,6087	14	18	0	31	63	14

Tabla 33 – continuación

Comuna	CSI	POD	Precisión	F1	TP	FP	FN	TN	N	Heladas reales
San Bernardo	0,4286	0,8077	0,4773	0,6000	21	23	5	14	63	26
El Monte	0,4255	0,8696	0,4545	0,5970	20	24	3	3	50	23
Estación Central	0,4242	1,0000	0,4242	0,5957	14	19	0	30	63	14
María Pinto	0,4182	0,7419	0,4894	0,5898	23	24	8	7	62	31
Lo Espejo	0,4167	1,0000	0,4167	0,5883	10	14	0	39	63	10
San Miguel	0,4167	1,0000	0,4167	0,5883	10	14	0	39	63	10
Talagante	0,4054	0,8333	0,4412	0,5769	15	19	3	25	62	18
La Florida	0,4000	0,7692	0,4545	0,5714	10	12	3	38	63	13
San Pedro	0,3902	0,8421	0,4211	0,5614	16	22	3	9	50	19
Calera de Tango	0,3793	0,6471	0,4783	0,5500	11	12	6	34	63	17
Lo Barnechea	0,3774	0,8000	0,4167	0,5480	20	28	5	10	63	25
Puente Alto	0,3704	0,7692	0,4167	0,5406	10	14	3	36	63	13
Curacaví	0,3571	0,6522	0,4412	0,5263	15	19	8	14	56	23
El Bosque	0,3500	1,0000	0,3500	0,5185	14	26	0	23	63	14
Pedro Aguirre Cerda	0,3462	0,9000	0,3600	0,5143	9	16	1	37	63	10
La Cisterna	0,3448	1,0000	0,3448	0,5128	10	19	0	34	63	10
San Ramón	0,3448	1,0000	0,3448	0,5128	10	19	0	34	63	10
La Granja	0,3333	1,0000	0,3333	0,5000	7	14	0	42	63	7
Peñaflor	0,3191	0,6818	0,3750	0,4839	15	25	7	11	58	22
Recoleta	0,3077	0,6667	0,3636	0,4706	4	7	2	50	63	6
Huechuraba	0,2667	0,6667	0,3077	0,4211	4	9	2	48	63	6
San Joaquín	0,2667	0,6667	0,3077	0,4211	4	9	2	48	63	6
Providencia	0,2667	0,6667	0,3077	0,4211	4	9	2	48	63	6
Vitacura	0,2667	0,6667	0,3077	0,4211	4	9	2	48	63	6
Independencia	0,2667	0,6667	0,3077	0,4211	4	9	2	48	63	6
Las Condes	0,2500	0,6667	0,2857	0,4000	4	10	2	47	63	6

Tabla 33 – continuación

Comuna	CSI	POD	Precisión	F1	TP	FP	FN	TN	N	Heladas reales
Santiago	0,2500	0,6667	0,2857	0,4000	4	10	2	47	63	6
Ñuñoa	0,2500	0,5000	0,3333	0,4000	3	6	3	51	63	6
Quilicura	0,2353	0,5714	0,2857	0,3809	4	10	3	46	63	7
Renca	0,2353	0,5714	0,2857	0,3809	4	10	3	46	63	7
Conchalí	0,2222	0,5714	0,2667	0,3637	4	11	3	45	63	7
Cerro Navia	0,1765	0,4286	0,2308	0,3000	3	10	4	46	63	7
Tiltil	0,1111	0,2000	0,2000	0,2000	1	4	4	54	63	5
Padre Hurtado	0,0882	0,1200	0,2500	0,1622	3	9	22	16	50	25
La Reina	0,0000	0,0000	0,0000	0,0000	0	5	6	54	65	6
Lo Prado	0,0000	0,0000	0,0000	0,0000	0	1	3	59	63	3
Pirque	0,0000	0,0000	0,0000	0,0000	0	4	10	49	63	10

H.2. Composición de la serie consolidada

Tabla 34: Distribución de la serie consolidada según fuente ganadora.

Fuente ganadora	Pares consolidados	Participación	Descripción
Atlas Agroclimático	246.875	67,07 %	Base histórica 2004–2016
Open-Meteo histórico	82.349	22,37 %	Reanálisis 2020–2026
INIA/AGROMET	38.858	10,56 %	Observaciones directas de mayor prioridad
Total	368.082	100,00 %	Pares únicos comuna × fecha

H.3. Cumplimiento de formato JSON por proveedor

Tabla 35: Robustez de formato de los proveedores de modelos de lenguaje.

Proveedor	Alertas procesadas	JSON correcto sin fallback	Porcentaje sin fallback	JSON final válido
Claude	73	73	100,0 %	100,0 %
Gemini	55	50	90,9 %	100,0 %
Llama 3	71	71	100,0 %	100,0 %

Los cinco casos de Gemini que requirieron *fallback* correspondieron a respuestas donde el campo de mensaje corto contenía un bloque JSON en lugar de texto plano. El mecanismo de validación permitió corregir estos casos antes de construir el mensaje final.

H.4. Evaluación cualitativa retrospectiva

La Tabla 36 presenta los resultados desagregados de la evaluación cualitativa retrospectiva. Se evaluaron 160 mensajes únicos y se identificaron 126 mensajes asociados a 42 fechas con ejecución de los tres proveedores sobre un contexto comparable. Los puntajes se presentan en una escala de 1 a 5.

Tabla 36: Resultados de evaluación cualitativa retrospectiva asistida por IA.

Proveedor	N	Consistencia	Pertinencia	Claridad	Accionabilidad	Global	Rev. manual
Claude	59	4,61	4,17	4,92	4,25	4,49	14
Gemini	44	4,18	4,73	4,43	4,50	4,46	6
Llama 3	57	3,28	4,00	3,04	3,25	3,39	11

Los puntajes corresponden a una evaluación exploratoria y no sustituyen una validación realizada por agricultores, asesores agrícolas o especialistas agronómicos independientes. Los casos con revisión manual corresponden principalmente a contexto

insuficiente, mensajes con recomendaciones poco específicas, discrepancias entre texto y contexto disponible o problemas de formato en la salida original.

H.5. Ejemplos anonimizados de alertas generadas

Los mensajes entregados por la arquitectura pueden incluir el tipo de evento detectado, la comuna o zona potencialmente afectada, temperaturas relevantes, nivel de confiabilidad y recomendaciones preventivas generales. Los ejemplos visuales de recepción de alertas mediante WhatsApp se presentan en el Capítulo 7.

Para proteger la privacidad de los usuarios, no se incluyen números telefónicos, identificadores de cuentas, credenciales ni datos personales. Los registros completos de auditoría y las salidas estructuradas se conservan en la base de datos y en los artefactos técnicos del proyecto.