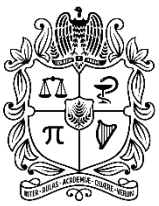


UNIVERSIDAD
NACIONAL
DE COLOMBIA

Predicción temprana de heladas en cultivos de altura, empleando métodos de aprendizaje de máquinas

Evelin Calderón Caro

Universidad Nacional de Colombia
Facultad de Minas, Departamento de Ciencias de la Computación y de la Decisión
Medellín, Colombia
2022

Predicción temprana de heladas en cultivos de altura, empleando métodos de aprendizaje de máquinas

Evelin Calderón Caro

Tesis presentada como requisito parcial para optar al título de:

Magister en Ingeniería - Analítica

Director:

Darío Antonio Castañeda Sánchez, Ph.D

Codirector:

John Willian Branch Bedoya, Ph.D

Línea de Investigación:

Inteligencia artificial

Grupo de Investigación:

Grupo de Investigación y Desarrollo en Inteligencia Artificial GIDIA

Universidad Nacional de Colombia

Facultad de Minas, Departamento de Ciencias de la Computación y de la Decisión

Medellín, Colombia

2022

A mi familia, amigos y compañeros de vida

A Andrés Camilo Peláez Yepes, hasta encontrarte.

“Nuestro universo de víctimas de todos los factores armados, legales e ilegales nos desborda, con el agravante de que la barbarie se ha vuelto rutina y costumbre”.

Fabiola Lalinde

Declaración de obra original

Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.

Evelin Calderón Caro

Nombre

Fecha 14/10/2022

Agradecimientos

Al profesor Darío Antonio Castañeda Sánchez quien dirigió, oriento y acompaño el desarrollo de este trabajo. También, agradezco a todos los que contribuyeron en mi proceso, especialmente, al profesor John William Branch, John Robert Ballesteros y a todos los integrantes del grupo de investigación GIDIA.

A las empresas Soluciones Wiga y Growers Hub Trading Group por proporcionar la subvención de investigación para la realización del proyecto.

A la Universidad Nacional de Colombia por propiciar programas para que los estudiantes puedan continuar sus estudios de posgrado.

Resumen

En Colombia, muchos cultivos están ubicados en los altiplanos de las regiones andinas, en altitudes superiores a 2.500 m s.n.m, donde se concentra la mayor susceptibilidad a la ocurrencia de eventos de heladas. El objetivo de este estudio fue proponer un modelo de predicción temprana de heladas basado en la relación entre estos eventos y variables climáticas, mediante la implementación de algoritmos de aprendizaje de máquinas. Las variables climáticas se obtuvieron a partir de trece estaciones meteorológicas distribuidas en nueve municipios del departamento de Cundinamarca. Las variables registradas fueron la temperatura, humedad relativa, punto de rocío, radiación fotosintéticamente activa y precipitación, estas constituyeron las variables explicativas de los eventos de heladas. Las métricas utilizadas para la evaluación predictiva del rendimiento de los cinco métodos de aprendizaje de máquinas examinados fueron precisión, tasa de verdaderos positivos, tasa de verdaderos negativos, exactitud y puntuación F_1. Se identificó que las horas previas a la ocurrencia de un evento de helada se caracterizan por presentar baja humedad, bajo punto de rocío y alta radiación. Cuatro de los cinco modelos entrenados se desempeñaron satisfactoriamente, con métricas de evaluación superiores al 91 %. La validación cruzada y el análisis estadístico demostraron que el modelo de potenciación del gradiente para la detección de heladas presentó la mayor precisión. Adicionalmente, se evaluaron dos modelos para la predicción de la temperatura mínima y se encontraron métricas de error (error medio absoluto y error cuadrático medio) inferiores a 0,55 °C para una ventana de tiempo de una hora.

Palabras clave: pronóstico, redes neuronales artificiales, temperatura mínima, variables climáticas.

Abstract

Early prediction of Frost events in high altitude crops, using machine learning methods

In Colombia, many crops are located in the highlands of the Andean region, at altitudes above 2,500 m a.s.l., where the greatest susceptibility to the occurrence of frost events is concentrated. The objective of this study was to propose an early frost prediction model based on the relationship between these events and climatic variables, through the implementation of machine learning algorithms. The climatic variables were obtained from thirteen meteorological stations distributed in nine municipalities of the department of Cundinamarca. The variables recorded were temperature, relative humidity, dew point, photosynthetically active radiation, and precipitation, these constituted the explanatory variables of frost events. The metrics used for the predictive evaluation of the performance of the five machine learning methods examined were precision, true positive rate, true negative rate, accuracy, and F_1 score. It was identified that the hours prior to the occurrence of a frost event were characterized by low humidity, low dew point and high radiation. Four of the five trained models performed satisfactorily, with evaluation metrics greater than 91 %. Cross-validation and statistical analysis showed that the gradient boosting model for frost detection had the highest accuracy. Additionally, two models for the prediction of the minimum temperature were evaluated and error metrics (mean absolute error and mean square error) of less than 0.55 °C were found for one hour time window.

Keywords: forecast, artificial neural networks, minimum temperature, climatic variables.

Contenido

Pág.

Resumen ..... **IX**

Lista de figuras ..... **XIII**

Lista de tablas ..... **XV**

Lista de abreviaturas ..... **XVI**

1. Introducción. **1**
1.1 Motivación. 3
1.2 Planteamiento del problema. 3
1.3 Hipótesis. 5
1.4 Objetivos. 5
1.4.1 Objetivo general. 5
1.4.2 Objetivos específicos. 5
1.5 Contribución. 5
1.6 Estructura. 6

2. Descripción del fenómeno de heladas a nivel climático y sus implicaciones en los procesos fisiológicos de las plantas. **7**
2.1 Fenómeno de heladas. 7
2.1.1 Factores climáticos que inciden en la ocurrencia de heladas. 8
2.1.2 Alteración de los procesos fisiológicos de las plantas como efecto de la helada. 10
2.1.3 Estrategias de mitigación de daños por heladas. 13
2.2 Descripción del funcionamiento de los métodos estadísticos y de aprendizaje de máquinas. 15
2.2.1 Modelos estadísticos. 15
2.2.2 Métodos de aprendizaje de máquinas. 16

3. Revisión sistemática de la literatura. **23**
3.1 Introducción. 23
3.2 Metodología. 23
3.3 Resultados y discusión. 26
3.3.1 Algoritmos de aprendizaje de máquinas y otros métodos que han sido utilizados para la predicción de heladas. 26
3.3.2 Características o variables climáticas utilizadas para la predicción de heladas. 35
3.3.3 Métricas de evaluación de los algoritmos implementados en la predicción de heladas. 37

3.3.4 Desafíos relacionados con la predicción de heladas mediante el aprendizaje de máquinas. 38
3.4 Conclusiones. 39

4. Predicción temprana de heladas en cultivos de altura, empleando métodos de aprendizaje de máquinas..... 41
4.1 Caracterización de los datos históricos de las variables climáticas. 41
4.1.1 Área de estudio. 41
4.1.2 Variables meteorológicas. 42
4.1.2 Análisis descriptivo. 46
4.1.3 Preprocesamiento de datos. 46
4.2 Identificación de las variables climáticas de mayor relevancia para la predicción temprana de heladas en cultivos de altura. 47
4.3 Selección del método de aprendizaje de máquinas para la predicción temprana de heladas en cultivos de altura. 48
4.3.1 Método de predicción. 49
4.3.2 Métricas de evaluación de rendimiento y comparación de los métodos de predicción. 52
4.4 Validación del desempeño del método seleccionado para la predicción temprana de heladas en cultivos de altura. 54

5. Experimentos y resultados 56
5.1 Caracterización de los datos históricos de las variables climáticas. 56
5.2 Identificación de las variables climáticas de mayor relevancia para la predicción temprana de heladas en cultivos de altura. 64
5.3 Selección del método de aprendizaje de máquinas para la predicción temprana de heladas en cultivos de altura. 68
5.3.1 Métodos de predicción. 68
5.3.2 Comparación y evaluación de las métricas de rendimiento de los métodos. . 72
5.4 Validación del desempeño del método seleccionado para la predicción temprana de heladas en cultivos de altura. 76

6. Conclusiones y trabajo futuro. 81
6.1 Conclusiones. 81
6.2 Trabajo futuro. 82

7. Referencias. 83

8. ANEXOS..... 95

Lista de figuras

Pág.

Figura 1-1. Daño causado por eventos de helada y escala de daño en cultivo de papa: a) daño muy bajo, b) daño bajo, c) daño moderado, d) daño severo. Nombre de la fuente: (Pino & Chen, 2016). 4
Figura 2-1. Procesos y variables climáticas asociados a la ocurrencia de un evento de helada radiactiva. Nombre de la fuente: Elaboración propia a partir de Arribillaga et al., 2020 & Fundación CK-12, 2021. 9
Figura 2-2. Deshidratación a nivel celular y daño de la estructura de la membrana plasmática como resultado de la formación de cristales de hielo en la región del apoplasto y en la pared celular. Nombre de la fuente: Kochhar & Gujral, 2020. 11
Figura 2-3. Red neuronal artificial cuya arquitectura es el perceptrón multicapa de una capa oculta. Nombre de la fuente: Pedregosa et al. (2011) 17
Figura 3-1. Esquema del protocolo de revisión. 25
Figura 3-2. Distribución de las publicaciones por año. Nombre de la fuente: “Elaboración propia” 26
Figura 4-1. Ubicación de las estaciones meteorológicas en el departamento de Cundinamarca (El tamaño del círculo corresponde a la cantidad de datos). Nombre de la fuente: “Elaboración propia”. 42
Figura 4-2. Cantidad de eventos de heladas por mes. Nombre de la fuente: “Elaboración propia”. 44
Figura 4-3. Histograma de frecuencia para la humedad relativa (a) y la temperatura (b) para la muestra total (gris claro) y el 60% de los datos utilizados para el entrenamiento de los modelos (gris oscuro). Nombre de la fuente: “Elaboración propia”. 49
Figura 4-4. Diagrama para la implementación de la predicción de heladas en la web. Nombre de la fuente: “Elaboración propia”. 55
Figura 5-1. Histograma de frecuencia para la temperatura (a), humedad relativa (b), punto de rocío (c) y radiación fotosintéticamente activa (d) para la temporada de heladas. Nombre de la fuente: “Elaboración propia”. 59
Figura 5-2. Comportamiento de las variables climáticas para las 24 horas del día anterior a eventos de helada (EH) o no asociadas con eventos de heladas (NEH): (a) temperatura, (b) humedad relativa, (c) punto de rocío, (d) precipitación y (e) radiación fotosintéticamente activa. La longitud de las barras de error corresponde al error estándar de la media (SEM). Nombre de la fuente: “Elaboración propia”. 61
Figura 5-3. Frecuencia de eventos de helada por hora. Nombre de la fuente: “Elaboración propia”. 62

Figura 5-4. Comportamiento promedio de las variables climáticas por temporada. (a) temperatura, (b) humedad relativa, (c) punto de rocío y (d) radiación fotosintéticamente activa. La longitud de las barras de error corresponde al error estándar de la media (SEM). Nombre de la fuente: “Elaboración propia”. 63
Figura 5-5. Correlación de las variables climáticas con la ocurrencia de un evento de helada por hora. (a) temperatura, (b) humedad relativa, (c) punto de rocío, (d) precipitación y (e) radiación fotosintéticamente activa. Nombre de la fuente: “Elaboración propia”. 65
Figura 5-6. Importancia para cada variable climática por hora del día. (a) temperatura, (b) humedad relativa, (c) punto de rocío y (d) radiación fotosintéticamente activa a partir de la prueba de Kruskal-Wallis (Kruskal), el método de la mínima redundancia y máxima relevancia (MRMR) y la selección de características mediante bosque aleatorios (RF). Nombre de la fuente: “Elaboración propia”. 67
Figura 5-7. Comparación de los valores reales vs los valores predichos por las redes neuronales (NN) para cuatro ventanas de tiempo diferentes: (a) 10 horas, (b) 7 horas, (c) 3 horas y (d) 1 hora. Nombre de la fuente: “Elaboración propia”. 75
Figura 5-8. Evaluación del rendimiento mediante el área bajo la curva de precisión-sensibilidad o AUC (a) y proporción de verdaderos positivos (TPR) versus la proporción de falsos positivos (FPR) ROC (b), para los algoritmos de regresión logística (LR), árbol de decisión (DT), máquina de soporte vectorial (SVM), árboles de decisión de aumento del gradiente GBDT y red neuronal (NN) con conjunto de prueba. Nombre de la fuente: “Elaboración propia”. 78
Figura 5-9. Visualización de la probabilidad de ocurrencia de un evento de helada para las trece estaciones ubicadas en nueve municipios de Cundinamarca. Nombre de la fuente: “Elaboración propia”. 79

Lista de tablas

Pág.

Tabla 3-1. Descripción de las publicaciones realizados en torno a la predicción de heladas. 28
Tabla 3-2. Variables climáticas utilizadas en las publicaciones para la predicción de heladas y cantidad de inclusión de cada una. 35
Tabla 4-1. Altitud, municipio y número de registros totales por estación meteorológica. 43
Tabla 4-2. Conjunto de datos para la predicción de heladas con variable respuesta binaria (“Helada”). 45
Tabla 4-3. Conjunto de datos para la predicción de heladas con variable respuesta continua (“Tmín”). 45
Tabla 4-4. Valores de ajuste para cada hiperparámetro por método. 50
Tabla 4-5. Descripción de las métricas de evaluación de rendimiento de los métodos de clasificación. 52
Tabla 4-6. Tabla de contingencia para comparación entre los valores pronosticados y observados para las clases binarias. 53
Tabla 4-7. Descripción de las métricas de evaluación de los métodos de regresión. 54
Tabla 5-1. Análisis descriptivo de las variables climáticas a lo largo del año (anual) en comparación con el comportamiento climático en la temporada de helada (Nov – Feb) y la temporada no asociada a heladas (Mar – Oct). 57
Tabla 5-2. Métricas de evaluación del rendimiento de los métodos en los conjuntos de datos de prueba de acuerdo con la estrategia de manejo de datos desbalanceados empleada. 69
Tabla 5-3. Valor óptimo para cada uno de los hiperparámetros determinado mediante validación cruzada por método. 70
Tabla 5-4. Métricas de evaluación del rendimiento de los métodos de clasificación en el conjunto de datos de validación. 72
Tabla 5-5. Métricas de evaluación del rendimiento de los métodos de regresión en el conjunto de datos de validación. 74
Tabla 5-6. Tabla de contingencia para comparación entre los valores pronosticados y observados para una ventana de tiempo de una hora (izquierda) y de diez horas (derecha). 76
Tabla 5-7. Métricas de evaluación del rendimiento de los métodos en el conjunto de datos de prueba. 77
Tabla 5-8. Métricas de evaluación del rendimiento de los métodos de regresión en el conjunto de datos de prueba. 79

Lista de abreviaturas

Abreviatura	Término
RL	Regresión logística
RNA	Redes neuronales artificiales
MLP	Perceptrón multicapa
BA	Bosques aleatorios
m	Variables de entrada
SVM	Máquina de soporte vectorial
DT	Arboles de decisión
GBDT	Potenciación del gradiente con árboles de decisión
ME	Métodos empíricos
SN	Simulación numérica
AA	Aprendizaje de máquinas
ARIMA	Modelos autorregresivos integrados de promedio móvil
IoT	Tecnología de internet de las cosas
MAE	Error medio absoluto
MAD	Desviación media absoluta
MSE	Error cuadrático medio
RMSE	Raíz cuadrada media del error
CC	Coeficiente de correlación
PC	Porcentaje correcto
MAPE	Error porcentual absoluto medio
MaxAPE	Error de porcentaje absoluto máximo
R2	Coeficiente de determinación
POD	Probabilidad de detección
TPR	Sensibilidad o Recall
P	Precisión
Ac	Exactitud
TNR	Especificidad
F_1	Puntuación
FPR	Tasa de falsos positivos
FAR	Tasa de alarmas falsas
CSI	Índice de éxito crítico
HIR	Tasa de aciertos
ISE	Índice de simetría de errores
EH	Evento de helada
NEH	Sin ocurrencia de un evento de helada
CCS	Coeficiente de correlación de Spearman
LR	Regresión logística
DT	Árbol de decisión
GBDT	Árboles de decisión de aumento de gradiente
SVM	Máquina de soporte vectorial
NN	Red neuronal
ROC	Curva característica operativa del receptor
FP	Falsos positivos
FN	Falsos negativos

Abreviatura	Término
VN	Verdaderos negativos
VP	Verdaderos positivos
HR	Humedad relativa
T	Temperatura
RFA	Radiación fotosintéticamente activa
PR	Punto de rocío
PP	Precipitación

1. Introducción.

La temperatura determina los límites latitudinales y altitudinales de la distribución de las especies de plantas en la tierra. En algunas plantas, las condiciones de congelación definen el límite inferior de temperatura, sugerido como el principal regulador de la distribución de especies, cuya diversidad disminuye con la latitud. Asimismo, en las montañas tropicales, el límite superior de elevación de muchas plantas vasculares está determinado por el punto de congelación del agua que se expande en áreas abiertas como los ecosistemas alpinos tropicales (p. ej., páramos). Los tejidos de las plantas tropicales, expuestos a heladas (temperatura < 0°C), pueden dañarse por deshidratación celular y desintegración de las membranas.

Los eventos de helada tienen la mayor probabilidad de ocurrencia en los paisajes montañosos transformados de los ecosistemas montanos tropicales. Estos eventos pueden afectar el desempeño de diversas especies de plantas, incluidas las utilizadas en actividades agrícolas, que cohabitan en estos ecosistemas. En Colombia, muchos cultivos, como flores, papa, caña, café y maíz, se distribuyen en las zonas altas de las cordilleras andinas de los departamentos de Cundinamarca, Boyacá, Norte de Santander y Nariño, y constituyen la principal actividad económica de estas regiones (Ministerio de Agricultura y Desarrollo Rural, 2019). En los Andes, la distribución de estos cultivos puede presentarse a alturas superiores de los 2.500 m s.n.m., constituyendo las áreas con mayor susceptibilidad a la ocurrencia de eventos de heladas (González & Torres, 2012).

Históricamente, los eventos de helada han ocurrido con mayor frecuencia entre diciembre y febrero (Bonilla et al., 2006; González & Torres, 2012; Meza & Gutierréz, 2020). Este período coincide con el ciclo de producción de flores para la temporada de San Valentín, el cual representa alrededor del 15% de las ventas anuales del país. En Colombia, la venta de flores en San Valentín de 2020 fue de 700 millones de tallos por un valor de US$ 1.414 millones (Becerra, 2021). Por lo tanto, comprender las causas de las heladas y

pronosticarlas es fundamental para salvaguardar y mantener la producción de flores y otros productos agrícolas del país.

En este sentido, el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) es un organismo público vinculado al Ministerio de Ambiente y Desarrollo Sostenible de Colombia, cuyo objetivo es generar conocimiento y garantizar el acceso a la información sobre el estado de los recursos naturales y las condiciones hidrometeorológicas en todo el país. Por otra parte, la Mesa Técnica Agroclimática, una iniciativa del gobierno para evitar pérdidas en el sector agropecuario provocadas por la variabilidad climática, ha brindado recomendaciones para la prevención de riesgos por eventos de heladas. Entre las principales se encuentran los planes de siembra que consideren los fenómenos climáticos, la reducción del establecimiento de cultivos susceptibles, la implementación de cercas vivas y la adecuación del sistema de riego (Ministerio de Agricultura y Medio Ambiente, 2020).

Otras estrategias de mitigación incluyen adaptaciones al suelo, suministro de energía a través de fuentes de calor o calentadores de aire, microaspersiones o uso de ventiladores (González & Torres, 2012; Snyder & de Melo-abreu, 2010). Además, se han propuesto otras estrategias basadas en la genética y la ingeniería celular (Colinet et al., 2010; Marmolejo & Ruiz, 2018), tratamientos químicos con oligoelementos, reguladores de crecimiento sintéticos, antioxidantes y acondicionamiento térmico (Joshi et al., 2020; Kochhar & Gujral, 2020; Liu & Sherif, 2019). El desarrollo de algunas de estas estrategias puede llevar tiempo y ser costoso. Además, las estrategias tradicionales requieren métodos de pronóstico precisos que permitan a los productores anticipar, prepararse y mitigar un posible evento de heladas en la producción agrícola.

Entre las metodologías más recientes utilizadas para la predicción de eventos de heladas, se encuentran las técnicas de aprendizaje de máquinas, como las redes neuronales artificiales (Castañeda & Castaño, 2017; Diedrichs et al., 2018; Latif et al., 2020; Mercado et al., 2016), árboles de decisión (Lee et al., 2016), bosques aleatorios (Diedrichs et al., 2018) y máquinas de soporte vectorial (Ding et al., 2019). Estas técnicas se han utilizado para identificar cuándo se va a producir una reducción de la temperatura por debajo de 0°C y para pronosticar el valor mínimo de temperatura que se alcanzará.

1.1 Motivación.

La falta de preparación para eventos de heladas puede causar pérdidas económicas, especialmente en la cantidad y calidad de los cultivos, e incluso puede causar la destrucción ipso facto de las plantas expuestas (Dujovne et al., 2020; Juurakko et al., 2021). Específicamente, en 2020, los eventos de heladas generaron pérdidas en el 2,0 % del área productiva en Cundinamarca (5400 hectáreas) (El Espectador, 2021); además, para este mismo año se reportó que aproximadamente el 30% de los cultivos en el municipio de Subachoque presentaron afectaciones por este fenómeno, mientras que para el municipio de Nemocón se estimó un impacto por heladas en 400 hectáreas (Rodríguez, 2020). De esta manera, mejorar nuestra capacidad de pronosticar eventos de heladas en los altiplanos colombianos, particularmente en Cundinamarca y Boyacá, es fundamental, como apoyo para la implementación de estrategias de mitigación ante eventos de heladas.

1.2 Planteamiento del problema.

En Colombia, el sector floricultor ha presentado un crecimiento constante en las últimas décadas, debido a la calidad y diversidad de sus productos. Este crecimiento le ha permitido al país establecerse como el segundo exportador de flores a nivel mundial con una participación del 15.2% después de Países Bajos (37.1%) (Superintendencia de Sociedades, 2017). De acuerdo con datos tomados entre septiembre de 2020 y agosto de 2021, el 95% de la producción de flores y follajes en Colombia fue exportada a Estados Unidos, Japón, Reino Unido, Países Bajos y Canadá (International Trade Center, 2020), y representó el 18,5% de las exportaciones agropecuarias del país (USD $1634,4 millones de dólares) durante el periodo señalado (Departamento Administrativo Nacional de Estadística DANE, 2021). La producción de estos ornamentales cuyo fin es la comercialización en el extranjero, se concentra en los departamentos de Cundinamarca y Antioquia, los cuales aportan el 66% y 32% respectivamente (Procolombia, 2019), en un área superior a las 7.000 hectáreas, donde, se han generado aproximadamente 140 mil empleos formales (Instituto Colombiano Agropecuario, 2021).

A pesar de la importancia del sector floricultor en la economía del país, la producción de flores es altamente susceptible al comportamiento de fenómenos climáticos, los cuales pueden afectar profundamente los costos de producción, la productividad de los cultivos, los indicadores de rendimiento y, por tanto, pueden producir pérdidas económicas

significativas para los productores nacionales (Kukal & Irmak, 2018; Majeed et al., 2021; Mcelwee et al., 2019). Por esta razón, es fundamental dentro de los programas de innovación tecnológica, identificar los procesos bioclimáticos que impactan la eficiencia en las actividades productivas asociadas al sector, como el fenómeno de helada; que se presentan cuando la temperatura desciende a valores inferiores al del punto de congelación del agua (Pearce, 2001; Shabala, 2017).

Cuando las especies de plantas tropicales se exponen a las bajas temperaturas que caracterizan los eventos de heladas (<0°C), se genera una lesión en los tejidos de las plantas a causa de la deshidratación celular y la desintegración de la membrana, que se produce cuando el agua contenida en los espacios extracelulares se congela y se forma hielo (Hashempour et al., 2014; Kochhar & Gujral, 2020; Pearce, 2001; Rout, 2020). Entre los síntomas asociados a las alteraciones metabólicas en los tejidos vegetales sensibles a estas temperaturas, se encuentra la reducción en el tamaño de la hoja, marchitamiento, clorosis, necrosis y escaso desarrollo reproductivo (**Figura 1-1**) (Kochhar & Gujral, 2020). Además, se disminuyen las tasas de los procesos fotosintéticos, de respiración y síntesis de proteínas (Kochhar & Gujral, 2020; Li et al., 2018). La magnitud de estos síntomas varía de acuerdo con la especie de planta, su edad y el tiempo de duración de la exposición a estas bajas temperaturas.

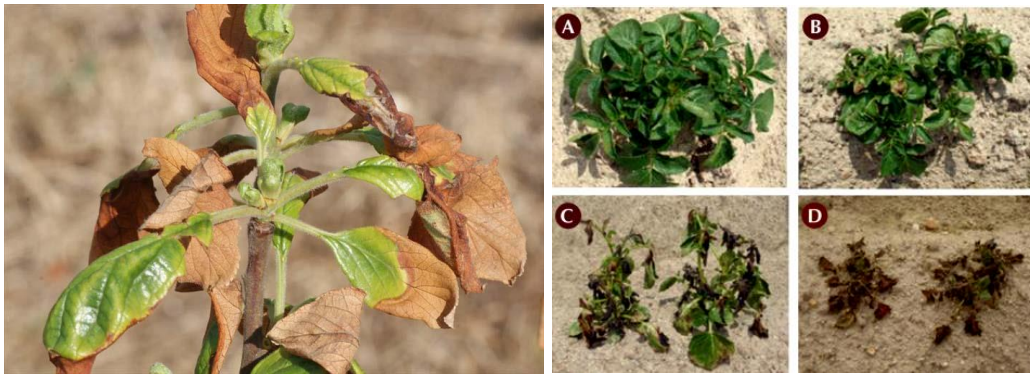


Figura 1-1. Daño causado por eventos de helada y escala de daño en cultivo de papa: a) daño muy bajo, b) daño bajo, c) daño moderado, d) daño severo. Nombre de la fuente: (Pino & Chen, 2016).

En consecuencia, la predicción precisa de eventos de helada es una importante herramienta de alerta temprana para los agricultores en la mitigación de los efectos de este fenómeno sobre los diversos cultivos que se encuentran establecidos en los altiplanos del

país. Por esta razón, pretendemos responder a la pregunta: ¿Es posible predecir la ocurrencia de eventos de helada en función de variables climáticas utilizando estrategias de aprendizaje de máquinas? ¿En caso de ser posible, cuál es el algoritmo con mejor rendimiento para la predicción de los eventos de helada en cultivos de altura?

1.3 Hipótesis.

Es posible predecir la ocurrencia de eventos de helada en cultivos de altura empleando métodos de aprendizaje de máquinas, a partir de la información histórica de cinco variables climáticas.

1.4 Objetivos.

1.4.1 Objetivo general.

Proponer un método para la predicción temprana de heladas en cultivos de altura, empleando métodos de aprendizaje de máquinas.

1.4.2 Objetivos específicos.

- Caracterizar los datos históricos de las variables climáticas para la predicción temprana de heladas en cultivos de altura mediante exploración, descripción y aplicación de técnicas de limpieza de datos.
- Identificar las variables climáticas más relevantes para la predicción temprana de heladas en cultivos de altura, empleando métodos de aprendizaje de máquinas.
- Seleccionar el método de aprendizaje de máquinas para la predicción temprana de heladas en cultivos de altura.
- Validar el desempeño del método seleccionado para la predicción temprana de heladas en cultivos de altura.

1.5 Contribución.

Entre los aportes que se alcanzaron con la realización del presente trabajo se encuentra la revisión sistemática de literatura científica que sintetiza la evolución del pronóstico de heladas en los últimos años, contribuyendo a las investigaciones y trabajos que puedan realizarse en un futuro. Adicionalmente, se caracterizó el comportamiento de cinco

variables climáticas asociadas teóricamente con la ocurrencia de heladas en Colombia, y se identificaron las horas en que estas variables tienen mayor influencia sobre el fenómeno.

La validación del desempeño del método de aprendizaje de máquinas seleccionado para la predicción temprana de heladas lo ratifica como una herramienta útil para la implementación de estrategias de prevención y mitigación, frente a los daños que puede ocasionar este fenómeno en los cultivos de altura en Colombia. Los resultados encontrados en la presente investigación fueron expuestos en el evento internacional "International Conference on Agricultural and Food Science 6th ICAFS" el 30 de septiembre de 2022 y fueron sometidos a través de un artículo científico en la revista Ceres el 26 de mayo de 2022.

1.6 Estructura.

El documento está organizado en cinco secciones además de la introducción. En la primera se hace una explicación detallada de los conceptos clave relacionados con el fenómeno de heladas y los métodos de aprendizaje de máquinas que se utilizarán para realizar la predicción como parte del marco teórico de referencia (capítulo 2). En la segunda sección se presenta la identificación, evaluación y síntesis de los avances y estudios que examinan problemas de investigación similares a través de la revisión sistemática de la literatura (capítulo 3), los cuales orientaron el desarrollo de los objetivos (capítulo 4). Finalmente, se presentan los resultados (capítulo 5), las conclusiones y los trabajos futuros (capítulo 6) relacionados con el problema de investigación.

2. Descripción del fenómeno de heladas a nível climático y sus implicaciones en los procesos fisiológicos de las plantas.

2.1 Fenómeno de heladas.

El comportamiento del clima a nivel regional está influenciado por procesos físicos locales, fenómenos de gran escala como El Niño-Oscilación del Sur (ENOS) y otros modos dominantes de variabilidad climática (Christensen et al., 2013). Asimismo, los sistemas montañosos como la cordillera de los Andes alteran la circulación atmosférica y, como consecuencia, se presentan condiciones climáticas contrastantes (Garreaud, 2009). Particularmente en los sistemas productivos colombianos, identificar y entender estos fenómenos es indispensable para la programación de actividades agronómicas, pues la variabilidad climática puede beneficiar o afectar ampliamente la producción.

Los eventos de heladas se producen cuando la temperatura ambiente alcanza valores iguales o inferiores a cero grados centígrados, y constituyen uno de los fenómenos atmosféricos que pueden afectar los sistemas productivos agrícolas del país (González & Torres, 2012). De acuerdo con la forma en que se origina la helada, esta puede clasificarse en advectiva, radiativa, mixta o por evaporación (Castillo & Castellvi, 2001). Particularmente, en las regiones andinas de Colombia, se presentan heladas radiativas, las cuales se generan como consecuencia de la emisión de energía en forma de calor desde las plantas y el suelo hacia la atmosfera en las noches de baja o nula nubosidad y viento en calma, por lo tanto, la temperatura superficial y de las capas de aire más cercanas a la superficie terrestre desciende (González & Torres, 2012; Hurtado, 1996).

Adicionalmente, las heladas pueden diferenciarse por el aspecto visual que se observa en las plantas como consecuencia de la disminución de la temperatura. Cuando se presenta

una condensación sólida de la humedad del aire que se deposita sobre la superficie foliar de las plantas se forma una capa de hielo, conocida como helada blanca, por el contrario, cuando el contenido de humedad relativa es muy bajo, no hay formación de rocío generándose un mayor daño de los tejidos vegetales, que finalmente se queman, tomado un color negro, por lo que se le denomina helada negra (Arribillaga et al., 2020; González & Torres, 2012; Snyder & de Melo-abreu, 2010).

2.1.1 Factores climáticos que inciden en la ocurrencia de heladas.

La ocurrencia de un evento de helada está influenciada principalmente por la radiación solar que, en conjunto con la circulación de los vientos, el contenido de vapor de agua en la atmósfera, la altitud, latitud y otras características topográficas de la zona, determinan la temperatura del aire. El comportamiento de la radiación incidente sobre la superficie terrestre en zonas cercanas a la línea ecuatorial está estrechamente relacionada con la fracción de cielo cubierto por nubes (González & Torres, 2012).

La nubosidad es una de las principales variables que inciden en la ocurrencia de heladas (Lindkvist et al., 2000) y de otros fenómenos climáticos, debido a que regula la cantidad de radiación de onda corta y larga tanto a nivel regional como global (Alapaty et al., 2012; Stephens, 2005). Cuando se presenta alta nubosidad, la transmisión de energía descendente (de las nubes a la superficie terrestre) se incrementa en función de la temperatura de las nubes y se reduce la pérdida neta de energía (Snyder, 2000). Además, la energía que se transmite por radiación descendente del cielo, se transfiere por conducción desde el suelo hacia el aire, y por convección se propicia el movimiento de masas de aire cálidas ascendentes y masas de aire frías descendentes, que transportan el calor entre regiones con temperaturas diferentes (Snyder, 2000).

Así, el registro del comportamiento de la radiación es indispensable para la predicción de la ocurrencia de heladas en las regiones tropicales, debido a que la radiación térmica emitida del suelo hacia el espacio se presenta cuando la nubosidad es reducida en el día, por lo tanto, no se absorbe la radiación del suelo que es nuevamente emitida hacia la superficie terrestre cuando hay nubes (Arribillaga et al., 2020). Bajo esta condición, aumenta la probabilidad de un descenso gradual severo de la temperatura que inicia desde las horas de la noche hasta las primeras horas de la mañana siguiente, momento en que

alcanza valores inferiores a 0°C (González & Torres, 2012). Sin embargo, cuando se presentan las condiciones de una helada por radiación y la velocidad del viento es superior a 7,2 $km.h^{-1}$, la transferencia de calor del aire compensa la perdida de energía por radiación (Snyder & de Melo-abreu, 2010).

Por lo tanto, se ha asociado las horas de frío en la madrugada con noches de calma (velocidad del viento menor a 1,0 $km.h^{-1}$) (del Angel & Sarmiento, 2011). Esta variable influye tanto en la ocurrencia como en la intensidad de los eventos de helada (Lhomme & Vacher, 2003), debido a que en noches en que se presentan velocidades superiores, el viento mezcla el aire frío cercano a la superficie con el aire más cálido de las capas superiores mediante el transporte de calor por convección, evitando así la inversión térmica y, por consiguiente, disminuyendo la probabilidad de que se presente un evento de helada (Figura 2-1) (González & Torres, 2012; Ribeiro et al., 2006).

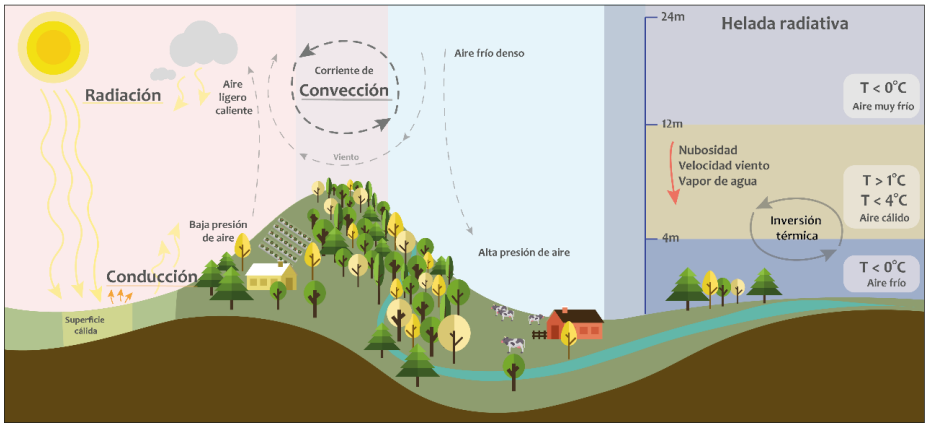


Figura 2-1. Procesos y variables climáticas asociados a la ocurrencia de un evento de helada radiactiva. Nombre de la fuente: Elaboración propia a partir de Arribillaga et al., 2020 & Fundación CK-12, 2021.

El vapor de agua contenido en la atmósfera también influye en la cantidad de energía que se emite en forma de calor por radiación desde la superficie terrestre, debido a que el agua puede absorber esta energía y emitirla nuevamente hacia la superficie. Por lo tanto, a mayor cantidad de vapor de agua, la pérdida de calor disminuye y, como consecuencia, la reducción de la temperatura en una noche de baja nubosidad y en condiciones de calma será más gradual e incluso puede evitarse la ocurrencia de valores inferiores a 0°C (González & Torres, 2012).

Además de las variables climáticas, la topografía también presenta una relación con la intensidad y el daño que provocan las heladas. La curvatura del terreno es una variable importante para evaluar el riesgo de ocurrencia de helada en áreas montañosas, como lo demuestran Lindkvist et al. (2000), quienes encontraron que las heladas se presentan con mayor frecuencia en valles estrechos seguido de zonas cóncavas y planas y, por el contrario, las áreas elevadas y convexas es donde hay una menor ocurrencia de heladas. Adicionalmente, se ha identificado que en las zonas de las llanuras cercanas a las montañas se presentan daños más graves por heladas que en las laderas (Chang et al., 2014).

Teniendo en cuenta lo anterior, en este estudio se evaluó el comportamiento de la temperatura, humedad relativa, punto de rocío, velocidad del viento, precipitación y radiación fotosintéticamente activa a lo largo del día, como variables fundamentales para entender la ocurrencia de eventos de heladas en el altiplano de Cundinamarca, y mediante métodos de aprendizaje de máquinas, predecir a priori la probabilidad de ocurrencia de este fenómeno.

2.1.2 Alteración de los procesos fisiológicos de las plantas como efecto de la helada.

En términos fisiológicos, un evento de helada ocurre cuando se forma hielo en el interior de los tejidos de las plantas, los cuales se afectan e incluso mueren como consecuencia de la solidificación del agua (Kochhar & Gujral, 2020; Snyder & de Melo-abreu, 2010). La congelación del agua en el interior de los tejidos vegetales inicia en el espacio extracelular o apoplasto, a partir del cual se genera un desplazamiento del agua desde el interior de las células, que al llegar al apoplasto se agrega con los cristales de hielo y ocasiona deshidratación, desbalance iónico y daños en la membrana plasmática (Figura 2-2) (Clarkson et al., 1988; Kochhar & Gujral, 2020). La afectación de la integridad de los canales de las membranas ocurre debido a que, con la disminución de la temperatura, los ácidos grasos insaturados que la componen cambian de un estado fluido a semicristalino, generando fugas de solutos (Kochhar & Gujral, 2020).

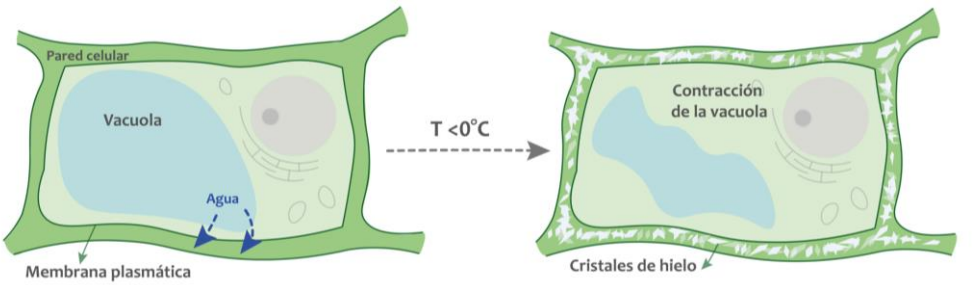


Figura 2-2. Deshidratación a nivel celular y daño de la estructura de la membrana plasmática como resultado de la formación de cristales de hielo en la región del apoplasto y en la pared celular. Nombre de la fuente: Kochhar & Gujral, 2020.

Cuando la disminución de la temperatura se prolonga, la savia que se encuentra en los haces vasculares conductores, principalmente en el xilema, también puede solidificarse. Los gases disueltos en la savia se vuelven menos solubles al agua con la congelación y, finalmente, tienden a formar burbujas que se expanden y colapsan al descongelarse (Kochhar & Gujral, 2020). Esta disfunción fisiológica se conoce como cavitación y está asociada con la pérdida de la conductividad hidráulica en el xilema y la deshidratación de los tejidos de la parte superior de la planta (Taiz & Zeiger, 2002).

Adicionalmente, el marchitamiento de las plantas puede intensificarse durante las primeras horas de descenso de la temperatura, especialmente cuando no se realiza un suministro de agua suficiente, que desencadene una reducción en el potencial hídrico de las hojas y, como consecuencia, las raíces pierden la capacidad de absorber agua y transportarla hacia los tejidos superiores (Nagasuga et al., 2011; Taiz & Zeiger, 2002; Wolfe, 1991). A esta pérdida en la capacidad de absorción de agua se adiciona una disminución en la adsorción de iones y en el transporte de estos entre los diferentes órganos de las plantas. Así, una reducción de la temperatura también genera un déficit en el contenido de nutrientes de la planta (Lukatkin et al., 2012).

Por otro lado, un descenso en la temperatura puede reducir la velocidad de reacciones bioquímicas, alterando así procesos fisiológicos (régimen hídrico, nutrición mineral, fotosíntesis, respiración y metabolismo) y modificando las tasas de transcripción de genes de varias proteínas de bajo peso molecular (Lukatkin et al., 2012). Estás disfunciones

metabólicas se ven reflejadas en la planta a través de síntomas como la reducción de la expansión de las hojas, marchitamiento, clorosis, necrosis y escaso desarrollo productivo (Kochhar & Gujral, 2020). La severidad de los síntomas y el tipo de lesión por frío puede aparecer de diferentes formas, dependiendo de la especie, el estado de desarrollo y la altura sobre el nivel del suelo de las plantas, la temperatura de la hoja, las condiciones nutricionales y sanitarias, la duración de la exposición a la helada y la temperatura mínima alcanzada (Chang et al., 2014; D. A. Gómez, 2014; Kochhar & Gujral, 2020; Lukatkin et al., 2012).

El proceso de fotosíntesis también puede verse afectado, debido a que se detiene el transporte de carbohidratos en el floema desde las hojas hacia otros órganos de las plantas, se reduce la actividad de las enzimas claves en el ciclo de Calvin y se genera un daño oxidativo de los fotosistemas en las membranas de los cloroplastos como resultado del aumento en la peroxidación de lípidos y la degradación de los pigmentos fotosintéticos (clorofila, caroteno y xantofilas) (Fryer et al., 1998; Kingston-Smith & Foyer, 2000 en Lukatkin et al., 2012). Además, el proceso de respiración puede afectarse como consecuencia de la disminución de la fosforilación oxidativa a causa de la hinchazón y degeneración de las mitocondrias (Yin et al., 2009).

La alteración de estos procesos fisiológicos como resultado de la ocurrencia de un evento de helada ha generado pérdidas significativas en actividades productivas agrícolas. Por ejemplo, se ha identificado que, entre los diversos fenómenos climáticos, las heladas son las que generan un mayor impacto económico en los agroecosistemas de Estados Unidos, equivalente a USD $ 1,5 billones anuales (Juurakko et al., 2021). Específicamente para los cultivos del este y medio oeste se han reportado daños equivalentes a USD $ 2 mil millones para los meses de primavera (Gu et al., 2008). Similarmente, los cultivos de vides en Argentina y Chile han registrado afectaciones por heladas en 35 mil y 3 mil hectáreas, respectivamente (Dujovne et al., 2020). Las estimaciones de las pérdidas productivas asociadas a eventos de heladas para Colombia equivalen a USD $ 32,7 millones para los cultivos de rosas, sin embargo, es necesario obtener valores más precisos de la incidencia de este fenómeno tanto en el sector floricultor como en otros sectores productivos de la Sabana de Bogotá.

2.1.3 Estrategias de mitigación de daños por heladas.

Se han propuesto diversas estrategias para la mitigación de daños por heladas basadas en el incremento de la tolerancia al frío. De acuerdo con Lukatkin et al. (2012), estas estrategias pueden agruparse en: uso de ingeniería genética y celular (Colinet et al., 2010; Galiba et al., 2009; Marmolejo & Ruiz, 2018), tratamiento químico con oligoelementos, reguladores de crecimiento sintético y antioxidantes (Joshi et al., 2020; Kochhar & Gujral, 2020; Liu & Sherif, 2019) y, por último, estrategias de acondicionamiento térmico como calentamiento intermedio y estrés por calor. La síntesis de proteínas anticongelantes es otra de las estrategias para la protección de daño contra heladas, debido a que estas tienen la capacidad de permanecer hidratadas (Kochhar & Gujral, 2020).

Adicionalmente, Snyder & de Melo-Abreu (2010) agruparon los métodos de reducción del daño por heladas en pasivos y activos. En el primer grupo ubicaron estrategias enfocadas en la prevención mediante la selección de áreas de siembra, implementación de técnicas para el drenado del aire frío, selección de plantas tolerantes, cubiertas de suelo, la modificación del entorno físico del cultivo con la implementación de barreras naturales, y el uso adecuado de prácticas culturales (González & Torres, 2012; Ribeiro et al., 2006; Simnitt et al., 2017; Snyder & de Melo-abreu, 2010; Striegler et al., 2007). Por otro lado, entre los métodos activos agruparon las actividades enfocadas en evitar la formación de hielo al interior de los tejidos vegetales después de una alerta emitida el día previo o incluso horas antes del descenso de la temperatura (Ribeiro et al., 2006; Snyder & de Melo-abreu, 2010). Entre estos métodos se encuentra la aplicación de riego por aspersión, la calefacción del aire, aislamiento con espumas y el uso de máquinas de viento (Hu et al., 2016, 2018; Lukatkin et al., 2012; Olszewski et al., 2017; Ribeiro et al., 2006; Snyder & de Melo-abreu, 2010).

La implementación de estrategias de mitigación de daños por heladas, mediante métodos pasivos o activos, debe enmarcarse en los objetivos de desarrollo sostenible: promover la agricultura sostenible, garantizar la disponibilidad de agua y su gestión sostenible, y garantizar modalidades de consumo y producción sostenibles (Chavarro et al., 2018). Adicionalmente, es imprescindible conocer con antelación la ocurrencia de un fenómeno de helada y la temperatura mínima que se alcanzará, para así desplegar de manera eficiente las estrategias de mitigación.

La aplicación del Internet (IoT) ha tenido un impacto positivo en la agricultura inteligente, en la automatización de procesos como el suministro preciso de agua y nutrientes en cultivos de invernadero sin suelo, en la transición a la agricultura 4.0, entre otros (Maraveas & Bartzanas, 2021). El almacenamiento de datos de sensores en la nube y la integración con tecnologías inteligentes para el monitoreo remoto de los niveles de agua de las plantas, el contenido nutricional, el pH del suelo, la humedad y la temperatura integrado con la inteligencia artificial, el aprendizaje de máquinas y la automatización de operaciones agrícolas se ha traducido en importantes ahorros de costos y mejoras en los rendimientos (Maraveas & Bartzanas, 2021).

La implementación de modelos de predicción también brinda a los agricultores información en tiempo real sobre las condiciones climáticas, mitigando así los efectos adversos del cambio climático y el calentamiento global. Sin embargo, la adopción de nuevas tecnologías y la agricultura de precisión en los países en desarrollo ha sido baja, contribuyendo a ampliar las brechas en la seguridad alimentaria entre las naciones desarrolladas y emergentes (Maraveas & Bartzanas, 2021).

En este sentido, Castañeda & Castaño (2020) desarrollaron un sistema inteligente para la gestión del riesgo de heladas a partir del monitoreo de variables climáticas, la predicción de la temperatura al interior del invernadero a través de redes neuronales y la implementación de un sistema inteligente de gestión del riego anti-heladas. La interacción de la información climatológica y el sistema de riego se da mediante un sitio web que permite la interconexión, adquisición y monitoreo de información en tiempo real a través de sistemas de telefonía móvil (GSM/GPRS) y servicios de internet (TCP/IP). La red neuronal artificial predice la temperatura al interior de los invernaderos con un coeficiente de determinación R^2 de 84,1% y un error medio absoluto de 0,53°C, a partir del registro de la humedad relativa, la temperatura del aire exterior, la radiación solar, la velocidad del viento y la humedad relativa del aire interior. Un sistema de memoria asociativa difusa controla la activación de una bomba de agua para la aplicación de agua por aspersión sobre el invernadero.

El sistema de riego se activa con cinco niveles de acuerdo con la intensidad de la helada, el primer nivel de activación entre 0 y 20 % asociado a la no ocurrencia de heladas con

temperaturas superiores a 7,5 °C, el segundo nivel entre 20 y 40 % cuando se predice una posible helada (temperaturas entre 2,5 y 7,5 °C), el tercer nivel entre 40 y 60% para heladas leves con temperaturas entre -2,5 y 2,5 °C, el cuarto nivel entre 60 y 80% de activación del sistema de riego para heladas severas (temperatura entre -2,5 y -7,5 °C) y finalmente, para heladas fuertes (temperatura inferiores a -7,5 °C) el sistema de riego se activa entre un 80 y 100 % (Castañeda & Castaño, 2020).

2.2 Descripción del funcionamiento de los métodos estadísticos y de aprendizaje de máquinas.

A pesar de que la predicción de los fenómenos climáticos ha sido compleja debido al comportamiento no lineal de estos fenómenos, a la calidad de la información, la capacidad de procesamiento y a la cantidad de variables que pueden tener influencia, la aplicación de métodos de aprendizaje de máquinas ha sido una alternativa que ha venido en aumento en los últimos años. Los trabajos realizados para la predicción de heladas en los cuales han implementado métodos de aprendizaje de máquinas comenzaron a aparecer a partir del año 2000. Entre los métodos se encuentran las redes neuronales, los bosques aleatorios, los árboles de decisión y las máquinas de soporte vectorial para la predicción de la temperatura mínima y la identificación de la ocurrencia de una helada como variable respuesta binaria. Por lo tanto, esta sección se enfoca en la descripción de los algoritmos implementados en el presente estudio para la predicción de eventos de heladas tanto para abordar el problema desde una perspectiva de clasificación como para la predicción de la temperatura mínima que presupone un problema de regresión.

2.2.1 Modelos estadísticos.

- **Regresión logística (RL).**

El algoritmo de Regresión Logística describe la relación entre una variable respuesta y su predictor o variables independientes a través de la Ecuación 2-1.

Ecuación 2-1

$$logit[\pi(x_i)] = \exp\left(\alpha_0 + \beta' x_i\right)$$

Ecuación 2-2

$$P\ (Y = 1) = \frac{\exp\ (\alpha_0 + \ \beta' x_i)}{1 + \ \exp\ (\alpha_0 + \ \beta' x_i)}$$

Donde α_0 es el intercepto y β' son los coeficientes de las variables climáticas explicativas (x_i). La Ecuación 2-2 muestra la probabilidad de ocurrencia de un evento de helada EH (Y) en función de las variables climáticas (x_i) (Juna et al., 2022; Lee et al., 2016; Zhu et al., 2019). Se asume que la probabilidad con valor superior a 0,5 indica la ocurrencia de un evento de helada, mientras que con un valor inferior se asocia con la no ocurrencia de helada.

2.2.2 Métodos de aprendizaje de máquinas.

- **Redes neuronales artificiales (RNA).**

Las redes neuronales artificiales constituyen un proceso de aprendizaje de máquinas conocido como aprendizaje profundo, las cuales están compuestas por nodos o neuronas interconectadas en una estructura de capas que se asemejan al cerebro animal. Esta estructura permite la creación de un sistema adaptable implementado por las computadoras para aprender de los errores y generar mejoras continuas. Esta configuración matemática de las redes neuronales permite la predicción de variables climáticas mediante la aproximación de funciones lineales y no lineales de una manera más eficiente en comparación con otros métodos convencionales (Fuentes et al., 2018).

Entre los diferentes tipos de RNA, el perceptrón multicapa (MLP por sus siglas en inglés) es el más común y se caracteriza por ser una red jerárquica conectada con retroalimentación proyectada hacia el futuro y compuesta por tres o más capas (capa de entrada, capa de salida y capas intermedias u ocultas) que a su vez están conformadas por una o más neuronas, como se muestra en la Figura 2-3 (Fuentes et al., 2018; Pedregosa et al., 2011).

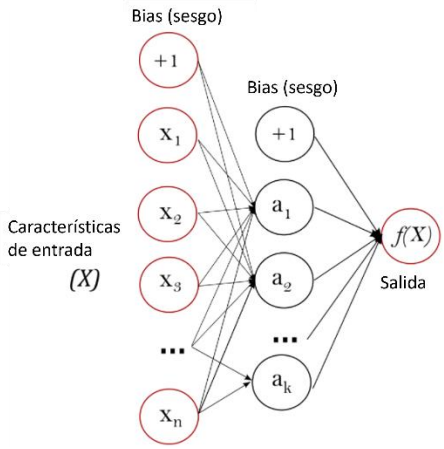


Figura 2-3. Red neuronal artificial cuya arquitectura es el perceptrón multicapa de una capa oculta. Nombre de la fuente: Pedregosa et al. (2011)

La implementación de las RNA se realizará a partir de la metodología propuesta por Fuentes et al., (2018). Estos implementaron la red neuronal multicapa de tres capas (capa de entrada, capas ocultas y capa de salida). La capa de entrada (parte izquierda de la Figura 2-3) corresponde a las variables de entrada. Conectada a esta capa se encuentran las neuronas de la capa oculta, las cuales transforman los valores de la capa anterior con una suma lineal ponderada $w_1x_1 + w_2x_2 + \ldots + w_mx_m$, seguida de una función de activación no lineal, como la función tan hiperbólica $f(o)$: $R^m \rightarrow R^o$ (Fuentes et al., 2018; Pedregosa et al., 2011). Los valores resultantes de la última capa oculta corresponden a la entrada de la capa de salida donde son transformados para generar la predicción.

Las neuronas del tipo MLP pueden combinarse con las neuronas de propagación hacia atrás o retropropagación (del inglés backpropagation) para obtener mejores métricas de rendimiento. La propagación hacia atrás de errores se basa en el método del gradiente descendente para minimizar el error al ajustar los pesos de acuerdo con la función de error mediante una serie de iteraciones. Adicionalmente, este método utiliza la función de entrenamiento de Levenberg-Marquardt (LM), con el fin de minimizar la función de suma de errores al cuadrado y proporcionar una solución numérica al problema de minimizar funciones no lineales mediante la Ecuación 2-3 (Fuentes et al., 2018).

Ecuación 2-3

$$(J^t J + \lambda i)\delta = J^t E$$

Donde:

J es la matriz jacobiana del sistema

$J^t J$ es la matriz hessiana aproximada

λ es el factor de amortiguamiento de Levenberg que se ajusta en cada iteración

i es la matriz identidad

δ es el vector de actualización de peso

E es el vector de error

El vector de error E contiene la salida de errores para cada vector de entrada utilizado en el entrenamiento de la red. Cuando este parámetro disminuye rápidamente indica que se debe utilizar un valor λ inferior, como consecuencia, el algoritmo se aproxima a un Gauss-Newton. Por el contrario, cuando el valor del vector de error no se reduce, λ puede incrementarse para dirigirlo hacia el descenso del gradiente. La determinación del menor valor de la raíz del error cuadrado medio (RMSE) como función de pérdida, al computar la diferencia entre la salida de la red neuronal y el valor observado asociado, genera una interrupción automática del proceso de entrenamiento (Fuentes et al., 2018).

- **Bosques aleatorios (BA).**

La predicción de la ocurrencia de una helada también puede realizarse mediante el entrenamiento del método de bosques aleatorios. Este constituye un método de ensamble que contiene técnicas de perturbación y combinación diseñadas para árboles, y se basa en la creación de un conjunto de clasificadores al implementar la aleatoriedad. La predicción del conjunto se da como la predicción promedio de los clasificadores individuales (Pedregosa et al., 2011).

La implementación de estos métodos de ensamble es el resultado de la aplicación del principio de compensación de sesgo-varianza, debido a que generalmente los métodos estadísticos de predicción con sesgo reducido presentan alta varianza. Al implementar un regresor que combine modelos de bajo sesgo con modelos de alta varianza, se puede generar una disminución general de la varianza, por lo tanto, en promedio esta combinación presenta una mayor precisión que cualquiera de los modelos individuales, cuando los modelos agregados no se encuentran correlacionados (Eccel et al., 2007).

Cada uno de los árboles individuales es construido mediante la selección de muestras al azar del conjunto de entrenamiento las cuales se usan para generar una estimación no sesgada del error de clasificación a medida que se agregan árboles al bosque. Posteriormente, entre el total de variables de entrada, se realiza una selección al azar de un grupo de características o variables de entrada (m). La división optima de las variables se obtiene con el criterio de importancia de Gini. Finalmente, cada árbol se entrena hasta que no se encuentren variables m adicionales para agregar. Este algoritmo continua hasta que se crea un número constante de árboles (Diedrichs et al., 2018; Eccel et al., 2007). Por lo tanto, cada árbol se entrena con datos ligeramente diferentes y la predicción de una nueva observación es el resultado del promedio de las predicciones de todos los árboles individuales que forman el modelo.

- **Máquina de soporte vectorial (SVM).**

Las máquinas de vectores de soporte (SVM por sus siglas en inglés) son un conjunto de métodos de aprendizaje de máquinas utilizados para la clasificación, la regresión y detección de valores atípicos. Las SVM se caracterizan por ser efectivas en espacios de gran dimensión, eficaz cuando el número de dimensiones es superior al número de muestras y es eficiente en el uso de la memoria debido a que funciona con un conjunto de puntos de entrenamiento en la función de decisión conocidos como vectores de soporte (Almansour et al., 2019; Ding et al., 2019; Pedregosa et al., 2011).

El algoritmo construye un hiperplano que alcanza la distancia máxima entre los datos etiquetados, lo que permite separar muestras positivas (eventos de heladas) y negativas (eventos de no heladas) (Almansour et al., 2019; Dinh et al., 2021; Pedregosa et al., 2011). Varias líneas pueden separar muestras positivas y negativas. Sin embargo, la línea óptima se selecciona utilizando vectores de soporte para desarrollar un hiperplano de margen máximo a través de la Ecuación 2-4 y Ecuación 2-5 (Almansour et al., 2019; Pedregosa et al., 2011).

Ecuación 2-4

$$f(x) = sgn\left(\sum_{i=1}^{n} y_i \alpha_i \, K(x, x_i) + b\right)$$

Ecuación 2-5

$$\sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j . y_i y_j . K(x_i, x_j)$$

Donde: y es la variable respuesta, b es el sesgo, α_i se obtiene resolviendo la Ecuación 2-5 en cada entrada y x_j son los vectores de soporte si su correspondiente α_i es mayor a 0.

Los cambios en los parámetros del algoritmo SVM pueden modificar el margen de separación y suavizarlo para permitir que algunos puntos se encuentren en el lado incorrecto del hiperplano sin que se presente una afectación de la precisión porcentual del clasificador. Por otro lado, la función kernel permite separar los puntos de datos en un espacio de dimensión superior cuando estos no pueden separarse en una dimensión más baja, implicando un aumento en el número de muestras de entrenamiento para evitar el sobreajuste (Almansour et al., 2019; Pedregosa et al., 2011).

- **Árboles de decisión (DT).**

Como una estructura de árbol, el algoritmo, a través de un criterio de separación, divide repetidamente el conjunto de datos, formando nodos y ramas, que se corresponden con las características y la categorización de las observaciones, respectivamente (Charbuty & Abdulazeez, 2021). La evaluación de la métrica de ganancia de información para cada variable permite la segmentación (Ecuación 2-6). Esta métrica refleja la reducción de la entropía y conduce a una clasificación eficaz del conjunto de datos. (Charbuty & Abdulazeez, 2021).

Ecuación 2-6

$$Gain\ (S, A) = \sum_{v \in V(A)} \frac{|S_v|}{|S|} Entropy(S_V)$$

Donde el rango de atributo A is $V(A)$, y S_V es un subconjunto del conjunto S igual al valor del atributo v (Charbuty & Abdulazeez, 2021).

- **Potenciación del gradiente con árboles de decisión (GBDT).**

Se basa en un método aditivo por etapas y permite la optimización de funciones de pérdida. En cada etapa se ajustan árboles de regresión de *n* clases sobre el gradiente negativo de la función de pérdida de desviación binomial o multinomial (Pedregosa et al., 2011). El

algoritmo emplea técnicas de impulso y desarrolla una serie de árboles de decisión para procesar conjuntos de modelos de predicción de bajo rendimiento (Danandeh Mehr, 2021). El proceso comienza con un modelo creado a partir de una característica x y una variable dicotómica y generando dos subconjuntos, el de las muestras clasificadas y el de las muestras no clasificadas. En la siguiente etapa, con el subconjunto sin clasificar y otra característica, se genera un nuevo modelo. A través del algoritmo GBDT, el error de todo el sistema disminuye en cada paso de clasificación (Danandeh Mehr, 2021).

3. Revisión sistemática de la literatura.

3.1 Introducción.

Teniendo en cuenta los conceptos expuestos en el marco teórico de referencia y con el fin de orientar la definición del problema de investigación, planificación, gestión y análisis de los datos, en esta sección se realizó una recopilación de las investigaciones previas relacionadas con la predicción de heladas a partir de variables climáticas mediante la revisión sistemática de la literatura, la cual permite recopilar el conocimiento existente y examinar el estado de un campo especifico mediante un proceso reproducible, científico y transparente (Linnenluecke et al., 2020).

3.2 Metodología.

Con el fin de implementar pronósticos de la ocurrencia de eventos de helada, como una de las herramientas de alerta temprana para la mitigación de los efectos que puede ocasionar este fenómeno, se realizó un levantamiento de las investigaciones con alcances relacionados con el aprendizaje de máquinas y la predicción de las heladas y se evaluaron las diferentes metodologías implementadas siguiendo el protocolo de revisión sistemática de literatura propuesto por Kitchenham & Charters (2007). En este sentido, se empezó por definir las preguntas de investigación, las cadenas de búsqueda y los criterios de selección de publicaciones, también se definieron las bases de datos a utilizar, las cuales fueron Science Direct (https://www.sciencedirect.com/ | Physical Sciences and Engineering and Life Sciences journals), Scopus (http://www.scopus.com) y Google Scholar (https://scholar.google.com/). La búsqueda fue realizada el 6 de septiembre de 2022.

Se realizó la revisión de las publicaciones en las tres bases de datos, se seleccionaron las investigaciones relevantes y se realizó un proceso de filtrado y evaluación mediante los criterios de exclusión. A partir de las investigaciones resultantes de la etapa de filtro y evaluación, se extrajeron, sintetizaron y documentaron los datos relevantes (autores, año

de publicación, tipo de publicación y otra información relacionada con las preguntas de investigación).

Las preguntas de investigación que se definieron fueron: PI1 ¿Cuáles algoritmos de aprendizaje de máquinas se han utilizado para la predicción de heladas?, PI2 ¿a partir de cuáles características o variables de entrada se ha realizado la predicción de heladas?, PI3 ¿Cuáles han sido las métricas de evaluación de los algoritmos implementados en la predicción de heladas?, PI4 ¿Cuáles son los desafíos relacionados con la predicción de heladas mediante el aprendizaje de máquinas?

La entrada inicial para la búsqueda básica fue “predicción de la temperatura mínima” Y “predicción de heladas” para la obtención de una visión amplia de las investigaciones. Posteriormente, se utilizaron los criterios de exclusión, se procesaron los resultados y se implementó una cadena de búsqueda de mayor complejidad, tanto en español como en inglés, para evitar la pérdida de investigaciones relevantes ((“evento de helada” OR “temperatura mínima” OR “agricultura de precisión” OR “cultivos”) AND (“aprendizaje de máquinas” OR “pronóstico” OR “clasificación” OR “regresión”)). La cadena de búsqueda fue igual para las tres bases de datos consultadas. Se recopilaron 254 investigaciones.

Los criterios de exclusión fueron CE1: la publicación no está relacionada con la predicción de heladas, CE2: la publicación no está escrita en español o inglés, CE3: la publicación fue realizada antes del 2000, CE4: la publicación no se encontraba en una revista indexada y CE5: la publicación no responde a ninguna de las preguntas de investigación. Con los criterios de exclusión se realizó un primer filtro para la selección de las publicaciones al evaluar los títulos y resúmenes y, posteriormente, se realizó la recopilación, comparación y evaluación de los artículos completos para una selección final de 30 publicaciones (Figura 3-1) distribuidas en el periodo de tiempo evaluado como se muestra en la Figura 3-2).

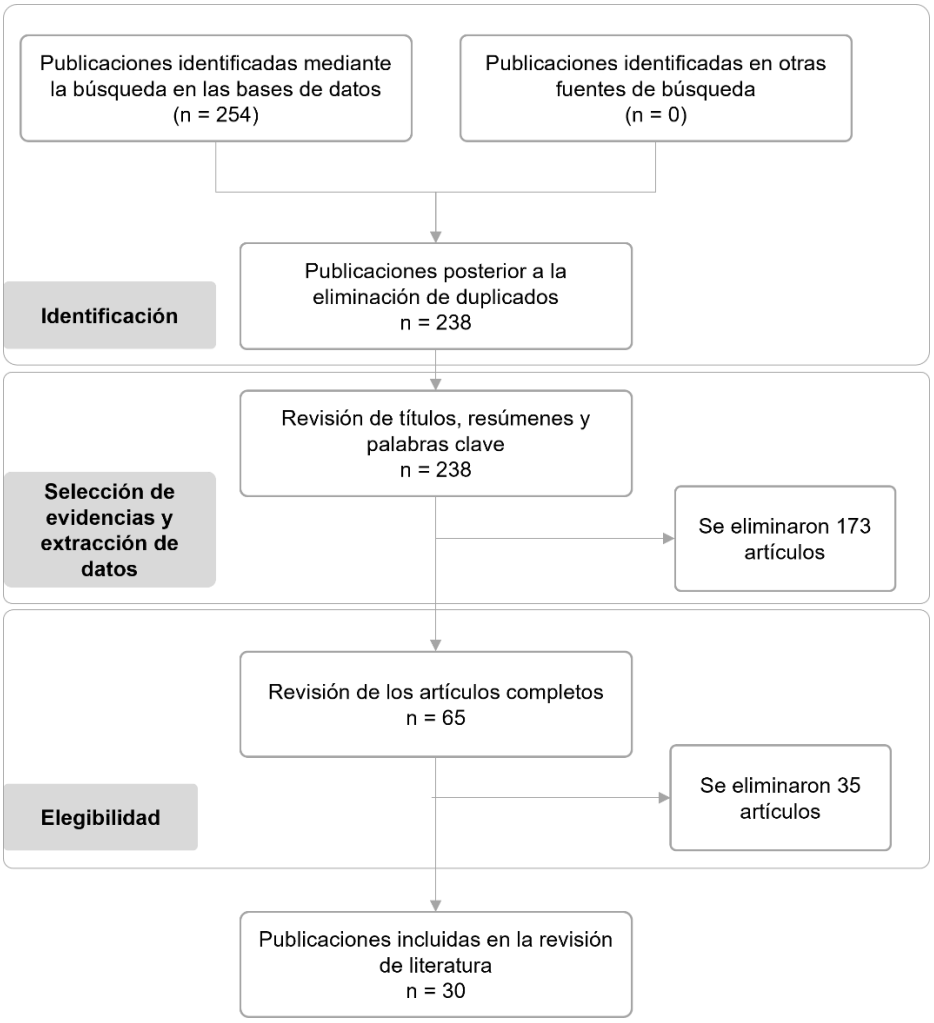


Figura 3-1. Esquema del protocolo de revisión.

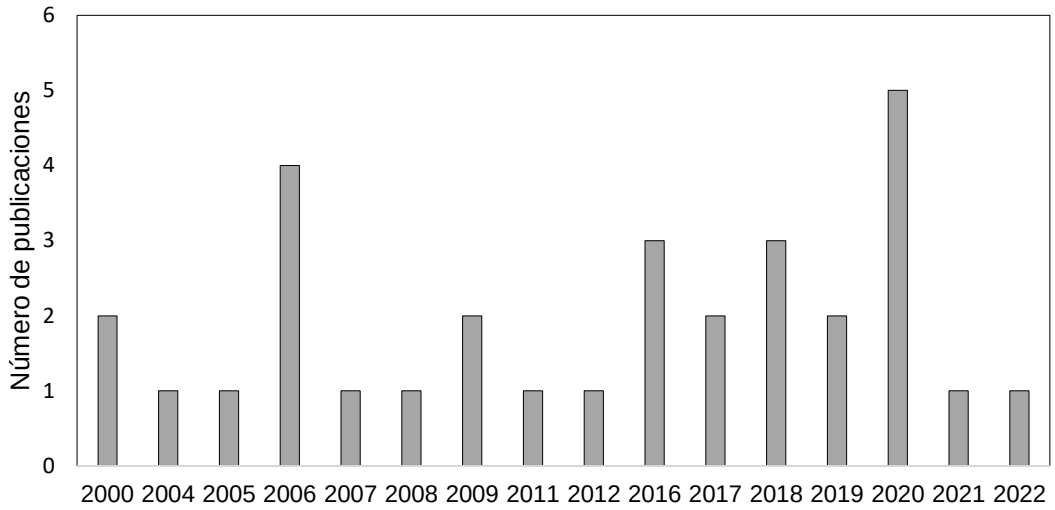


Figura 3-2. Distribución de las publicaciones por año. Nombre de la fuente: “Elaboración propia”

3.3 Resultados y discusión.

3.3.1 Algoritmos de aprendizaje de máquinas y otros métodos que han sido utilizados para la predicción de heladas.

El modelamiento y la predicción de las variables meteorológicas se ha realizado mediante métodos estadísticos y numéricos (Bonilla et al., 2006). Sin embargo, los primeros identifican tendencias en el comportamiento de las variables a pesar de que las condiciones atmosféricas no se repiten con exactitud en el mismo espacio y los segundos requieren computación de alto rendimiento (Bonilla et al., 2006). Asimismo, la predicción de los fenómenos climáticos ha sido compleja debido a la cantidad de variables relacionadas, a la calidad de la información y al comportamiento de tipo no lineal que presentan estos fenómenos (Garay y Cicero, 2001 en Ovando et al., 2005); por lo tanto, en los últimos años ha aumentado la aplicación de la inteligencia artificial en la predicción del clima.

Particularmente para la predicción de heladas, Diedrichs et al. (2018) agruparon los trabajos de investigación que se han realizado en métodos empíricos (ME), de simulación numérica (SN) y de aprendizaje de máquinas (AA), de acuerdo con la metodología empleada. Los métodos empíricos son los que se desarrollan mediante la construcción de fórmulas algebraicas a partir del análisis estadístico gráfico de series de parámetros,

mientras que los métodos de SN se desarrollan a partir del estudio de los fenómenos y la integración de herramientas matemáticas para modelar, simular y predecir el comportamiento del clima como el modelo de investigación y pronóstico del tiempo (WRF por sus siglas en inglés), diseñado para la investigación atmosférica y predicción operativa (Diedrichs et al., 2018).

El aprendizaje de máquinas constituye una de las ramas de la inteligencia artificial, mediante la cual se busca que las computadoras puedan aprender e inducir conocimiento a través del aprendizaje supervisado y no supervisado (Mercado et al., 2016). El aprendizaje supervisado es un método del aprendizaje de máquinas en el cual los modelos son entrenados utilizando datos etiquetados, por lo tanto, el valor del atributo objetivo para el conjunto de datos es conocido (Mercado et al., 2016). A su vez, el aprendizaje supervisado puede implementarse para resolver dos tipos de problema de acuerdo con la clase de atributo objetivo a predecir (Mercado et al., 2016; Ozgur, 2004). Si el atributo objetivo es un conjunto finito de resultados es un problema de clasificación, mientras que si el atributo objetivo es una variable continua, constituye un problema de regresión (Mercado et al., 2016; Ozgur, 2004).

Para la predicción de las heladas se han empleado técnicas de aprendizaje de máquinas supervisado, tanto para identificar cuando se va a presentar una reducción de la temperatura inferior a 0°C (problema de clasificación) (Cadenas et al., 2020; Ding et al., 2019; Meza & Gutierréz, 2020), como para pronosticar el comportamiento de la temperatura o el valor mínimo que alcanzará esta variable (problema de regresión) (Ding et al., 2020; Guillén et al., 2020; Latif et al., 2020; Zhou et al., 2022).

Entre los métodos de AA se han implementado algoritmos como el k-vecinos más cercanos, los bosques aleatorios, las máquinas de soporte vectorial y los árboles de decisión que, en conjunto con las redes neuronales, corresponden al 73% de los trabajos de investigación seleccionados, mientras que los métodos de ME y SN están representados en un 10 y 13 %, respectivamente. El 6,7% restante se clasificaron como modelos causales. Entre los métodos de ME y SN que se han implementado se encuentran la regresión logística, regresión lineal múltiple, modelos lineales generalizados, modelos autorregresivos integrados de promedio móvil (ARIMA) y la parametrización de las

ecuaciones de Brunt, Swinbank y Chudnosky para la predicción de eventos de helada (Tabla 3-1).

Tabla 3-1. Descripción de las publicaciones realizados en torno a la predicción de heladas.

Referencia	Área de estudio	Rango de tiempo	N° muestras	Variables climáticas	Metodología	Métrica	Tipo de método
(Snyder, 2000)	ND[1]	1990	ND	• Temperatura dos horas después de la puesta de sol	Predicción de la temperatura media horaria del aire hasta el amanecer de la mañana siguiente a partir de ecuaciones propuestas.	• Ac	ME
(Verdes et al., 2000)	Zavalla, Argentina	1973-1989	2.601 (1 EM[2])	• Temperatura mínima, de bulbo seco y húmedo • Nubosidad • Intensidad del viento	Predicción de la temperatura mínima (regresión) y ocurrencia de helada (clasificación) • Redes neuronales artificiales prealimentada • Clasificador Bayesiano • K-Vecinos más cercanos	• TPR • Ac • FPR • FAR • CSI	AA
(Maqsood et al., 2004)	Southern Saskat-chewan, Canadá	2001	ND	• Temperatura • Humedad relativa • Velocidad del viento	• Redes neuronales artificiales (perceptrón multicapa, recurrentes y base radial) y métodos de ensamble	• MAD • RMSE • CC	AA
(Ovando et al., 2005)	Córdoba, Argentina	24 años	3.650 (1 EM)	• Temperatura • Humedad relativa • Punto de rocío • Dirección y velocidad del viento • Nubosidad	Predicción de la temperatura mínima • Redes neuronales prealimentadas de propagación hacia atrás	• MSE	AA
(Bonilla et al., 2006)	Cundinamarca, Colombia	1990-2003	720	• Temperatura mínima de termómetro seco	Predicción de temperatura mínima • Modelo univariado de red neuronal	• MSE	AA
(Ghielmi & Eccel, 2006)	Trentino, Italia	2001-2004 1983-2004 1999-2004	ND (6 EM)	• Temperatura del atardecer • Humedad • Velocidad del viento • Temperatura del suelo • Duración de la noche	• Redes neuronales de perceptrón multicapa • Ecuaciones de Brunt, Swinbank, Chudnosky	• Ac • CC • HIR • FAR • CSI • ISE • PC	AA - ME

Referencia	Área de estudio	Rango de tiempo	N° muestras	Variables climáticas	Metodología	Métrica	Tipo de método
						• POD	
(Jain et al., 2006)	Georgia, Estados Unidos	1998-2002	ND	• Temperatura • Humedad relativa • Velocidad del viento • Radiación solar • Precipitación	Predicción de la temperatura • Redes neuronales artificiales prealimentadas de propagación hacia atrás	• MAE	AA
(Smith et al., 2006)	Georgia, Estados Unidos	1998-2005	32.000 19 EM	• Temperatura • Humedad relativa • Velocidad del viento • Radiación solar • Precipitación	Predicción de la temperatura mínima • Redes neuronales artificiales prealimentadas de propagación hacia atrás	• MAE	AA
(Eccel et al., 2007)	Alpes italianos, Trentino, Italia	2001-2005	ND	• Precipitación • Nubosidad • Presión atmosférica • Velocidad del viento • Punto de rocío • Temperatura • Humedad específica • Altura geopotencial	Predicción de la temperatura mínima (regresión) • Regresión multilineal • Bosques aleatorios • Redes neuronales artificiales de perceptrón multicapa	• MAE	AA
(Prabha & Hoogenboom, 2008)	Estados Unidos	1993–2006	(8 EM)	• Temperatura • Humedad relativa • Temperatura del suelo • Humedad volumétrica del suelo • Velocidad del viento • Presión superficial • Radiación solar	Modelo de investigación y pronóstico del tiempo (WRF) • Pronósticos meteorológicos a escala mesolocal de investigación avanzada y pronósticos meteorológicos de última generación (WRF-ARW)	• MAE • RMSE • CC	SN
(Smith et al., 2009)	Georgia, Estados Unidos	1997-2005	3,3 millones	• Humedad relativa • Velocidad del viento • Temperatura • Precipitación • Radiación solar	Predicción de la temperatura mínima • Redes neuronales artificiales prealimentadas de retropropagación	• MAE	AA
(Xier, 2009)	Estocolmo, Suecia	1756-2007	3.024 meses	• Temperatura promedio mensual	Predicción de la temperatura promedio mensual • Modelo lineal generalizado	Evaluación de residuales y prueba	SN

Referencia	Área de estudio	Rango de tiempo	N° muestras	Variables climáticas	Metodología	Métrica	Tipo de método
					• Modelo autorregresivo integrado de media móvil (ARIMA)	de hipótesis	
(Shamsnia et al., 2011)	Irán	1990-2009	ND	• Precipitación • Temperatura promedio • Humedad relativa	Predicción de la precipitación, temperatura y humedad • ARIMA	Criterio de información de Akaike	SN
(Wen et al., 2012)	Gansu, China	ND	ND	Modificación de los campos iniciales de WRF con: • Temperatura del suelo y del aire • Humedad del suelo y del aire • Clasificación de la vegetación	• Modelo de investigación y pronóstico del clima (WRF)	• RMSE	SN
(Lee et al., 2016)	Corea del Sur	1973-2014	4.325 (6 EM)	• Temperatura mínima del aire y del pasto • Punto de rocío • Velocidad del viento • Humedad relativa • Nubosidad	Predicción del evento de helada con algoritmos de clasificación • Árboles de decisión • Regresión logística	• POD • FAR • HIR	AA
(Mercado et al., 2016)	Cochabamba, Bolivia	2011-2015	178.450 (6 EM)	• Temperatura • Radiación solar • Velocidad del viento • Humedad • Precipitación	Predicción de la temperatura • Red neuronal prealimentada de una sola capa	CC	AA
(Yu et al., 2016)	Shandong Province, China	2013-2014	1000	• Temperatura y humedad del aire dentro del invernadero • Temperatura y humedad del suelo en el invernadero • Radiación solar • Velocidad del viento • Temperatura del aire al exterior del invernadero	Predicción de la temperatura mínima • Máquina de soporte vectorial de mínimos cuadrados	• MAE • MAPE • MSE	AA
(Castañeda & Castaño, 2017)	Región central México	1 año	52.560	• Temperatura del aire • Humedad relativa • Velocidad del viento • Flujo de radiación solar global	Predicción de la temperatura • Red neuronal artificial (perceptrón multicapa)	• R^2	AA

Referencia	Área de estudio	Rango de tiempo	N° muestras	Variables climáticas	Metodología	Métrica	Tipo de método
(Diedrichs et al., 2018)	Mendoza, Argentina	2001-2016	5 EM	Promedio, mínimo y máximo diario: • Temperatura • Humedad	Predicción de la temperatura mínima (regresión) y ocurrencia de evento de helada (clasificación) • Redes Bayesianas • Regresión logística • Bosques aleatorios	• MAE • RMSE • TPR • F1	AA
(Fuentes et al., 2018)	Regiones de Maule y Biobío, Chile	2010-2017	24.572 10 EM	• Temperatura • Humedad relativa • Radiación • Velocidad y dirección del viento • Precipitación • Punto de rocío	Predicción de la temperatura mínima • Redes neuronales artificiales basada en en retropropagación	• MSE • Ac • TPR	AA
(Guillen et al., 2018)	Murcia, España	2012-2018		• Humedad relativa promedio, mínima y máxima, • Radiación promedio, máxima y acumulada • Velocidad y dirección del viento • Precipitación • Punto de rocío • Déficit de presión de vapor • Temperatura mínima	Predicción de la temperatura mínima (regresión) y la ocurrencia de eventos de helada (clasificación) • Árboles de decisión	• MAE • RMSE • CC • R^2 • Ac • TPR • p	AA
(Ding et al., 2019)	Japón	2016-2017	ND	• Temperatura del aire • Humedad relativa • Radiación neta • Velocidad del viento	Predicción de eventos de heladas con algoritmos de clasificación • Máquina de soporte vectorial	• TPR • FPR • CC • R^2	AA
(Guillén et al., 2019)	Murcia, España	2018-2019	17.212	• Temperatura	Redes neuronales recurrentes (memoria a corto y largo plazo)	• RMSE • MAE • CC • R^2	AA
(Cadenas et al., 2020)	Murcia, España	2012–2018	5 EM	• Humedad relativa promedio, máxima, mínima • Radiación promedio, máxima y acumulada	Predicción de la temperatura mínima (regresión) y la ocurrencia de eventos de helada (clasificación) • k-vecinos más cercanos	• MSE • CC • R^2 • Ac • TPR • TNR	AA

Referencia	Área de estudio	Rango de tiempo	N° muestras	Variables climáticas	Metodología	Métrica	Tipo de método
				• Velocidad (máxima y promedio) y dirección del viento • Precipitación • Punto de rocío • Déficit de presión de vapor • Temperatura mínima		• F1	
(Guillén et al., 2020)	Murcia, España	ND	144 registros por día	• Temperatura • Humedad • Velocidad del viento	Predicción de la temperatura mínima • Redes neuronales recurrentes	• RMSE • MAE	AA
(Latif et al., 2020)	Washington y California, Estados Unidos	2014-2019	ND	• Temperatura • Datos del suelo	Predicción de la temperatura mínima • Redes neuronales convolucionales	• Ac	AA
(Meza & Gutierréz, 2020)	Cundinamarca, Colombia	2005-2019	5277 (1 EM)	• Temperatura mínima y máxima del aire • Humedad relativa • Precipitación • Día pluviométrico • Temperatura de bulbo seco mínima y máxima	Predicción ocurrencia de eventos de helada (clasificación) • Regresión logística • Redes Bayesianas	• TPR • F1 • P	ME
(Ding et al., 2020)	Ikuta, Japón	2017	9800	• Temperatura del aire • Humedad • Radiación	Predicción ocurrencia de eventos de helada • Modelos causales	• TPR • P	MC
(Ding et al., 2021)	Hokkaido, Japón	2019	ND	• Temperatura del aire • Radiación • Presión de vapor	Predicción ocurrencia de eventos de helada • Modelos causales	• TPR • P	MC
(Zhou et al., 2022)	New South Wales, Australia	2016 - 2017	30 EM	• Temperatura • Punto de rocío • Humedad relativa • Velocidad del viento	Predicción de la temperatura mínima • Redes neuronales recurrentes	• MSE	AA

[1]ND= No definido. [2]EM= Estación meteorológica.

Uno de los primeros trabajos enfocados en el pronóstico de heladas mediante algoritmos de aprendizaje de máquinas fue el de Verdes et al. (2000), quienes predijeron la ocurrencia de un evento de helada por el pronóstico de la temperatura mínima, la cual fue función del registro histórico de la temperatura de bulbo seco y húmedo, la nubosidad y la intensidad del viento. Nueve años más tarde, Smith et al. (2009) predijeron la temperatura del aire

durante todo el año mediante redes neuronales artificiales a partir del monitoreo de la humedad relativa, la velocidad del viento, la temperatura, la precipitación y la radiación solar. Además, estos autores incluyeron atributos temporales agrupados en cuatro niveles (medianoche, mañana, mediodía y tarde), los cuales fueron asignados mediante funciones de membresía triangular en el rango de 0 a 1. Para el entrenamiento de la red neuronal, estos autores utilizaron un conjunto de datos con 258 atributos y obtuvieron un desempeño de la predicción para el conjunto de datos de prueba de 0,516 °C de error absoluto medio en el horizonte de una hora y de 1,873 °C de error absoluto medio en el horizonte de doce horas.

Por otra parte, Yu et al. (2016) desarrollaron un algoritmo de predicción de temperatura basado en máquinas de soporte vectorial de mínimos cuadrados y encontraron pronósticos de valores de temperatura máxima y mínima en invernadero más precisos que las realizadas mediante máquinas de soporte vectorial estándar y redes neuronales de propagación hacia atrás. Estos autores realizaron la predicción de la temperatura con una ventana de tiempo de una, dos, cuatro y seis horas a partir del registro histórico de la humedad del aire, radiación solar, temperatura y humedad del suelo, velocidad del viento y temperatura externa, con una frecuencia horaria por un periodo de dos meses. Los resultados mostraron que la precisión de la predicción se disminuye a medida que se aumenta el intervalo de tiempo, en este sentido, el MSE fue de 0,008 para la predicción de temperatura una hora al futuro, mientras que para una ventana de tiempo de seis horas al futuro fue de 1,35, por lo tanto, los autores concluyeron que el método propuesto fue preciso y efectivo para predecir la temperatura mínima y máxima dentro del invernadero.

Adicionalmente, Fuentes et al. (2018) emplearon un concepto similar al de Verdes et al. (2000) para la predicción de la temperatura mínima del aire del día siguiente mediante redes neuronales, con una mayor cantidad de datos y una mayor cantidad de variables climáticas (temperatura, humedad relativa, radiación, velocidad y dirección del viento, precipitación y punto de rocío). En la validación del método se determinó que el error cuadrático medio en la predicción de la temperatura mínima fue de 2.99 °C; y la precisión media total en la detección de heladas mediante métodos de clasificación fue del 98% (86% de sensibilidad). Los autores atribuyen los errores en la detección de heladas a la precisión de las estaciones meteorológicas, las condiciones climáticas y geográficas locales, y el número de parámetros en la construcción de los métodos.

En este mismo año, Diedrichs et al. (2018) utilizaron registros de los valores promedio, mínimo y máximo diario de la temperatura y la humedad relativa, obtenidos a partir de sensores conectados desde diversas ubicaciones mediante tecnología de internet de las cosas (IoT). Esta tecnología les permitió mejorar el rendimiento de la predicción de heladas con bosques aleatorios, regresión logística y redes bayesianas. Asimismo, Guillén et al. (2019, 2020) diseñaron un algoritmo para predecir las temperaturas bajas a partir de información obtenida bajo una infraestructura IoT desplegada en dos parcelas en Murcia (sureste de España), con una frecuencia de un registro cada diez minutos. Los resultados obtenidos fueron satisfactorios de acuerdo con las métricas de evaluación (MSE inferior a un grado centígrado y R^2 superior a 0,95) (Guillén et al., 2020).

La tecnología IoT permitió obtener información de mayor calidad en periodos de muestreo menores, por lo tanto, las métricas de desempeño de los métodos también aumentaron. En este sentido, Cadenas et al. (2020) utilizaron información recolectada mediante esta tecnología (1 registro cada cinco minutos) para la predicción de heladas. Estos autores agruparon cada variable climática en un valor por hora (mínimo, promedio y máximo) e identificaron los atributos que hacen referencia a una misma medida en un único atributo expresado en un valor difuso. Mediante la técnica de regresión basada en k-vecinos más cercanos encontraron un MSE y MAE de 0,49 y 0,51 °C, respectivamente.

Estas investigaciones en conjunto con los avances propuestos por Ding et al. (2019) y Latif et al. (2020) han demostrado el potencial de las técnicas de aprendizaje de máquinas para la predicción de heladas. Sin embargo, la ocurrencia de los eventos de helada están influenciado por factores ambientales con un impacto acumulado, por lo tanto Ding et al. (2019, 2020, 2021) propusieron la inclusión del modelo causal (MC), mediante el cual se puede capturar la relación causal imprecisa entre las lecturas pasadas de los factores ambientales y la ocurrencia futura de heladas. Estos autores realizaron experimentos con una muestra de información recopilada por un mes (entre diciembre de 2016 y enero de 2017), con una resolución de un registro por minuto para las variables temperatura, humedad relativa, radiación neta y velocidad del viento. A partir de este conjunto de datos de series continuas y el evento de helada como variable respuesta binara, Ding et al. (2020, 2021) predijeron la ocurrencia de este fenómeno mediante modelos causales, con un TPR

de 89%. Sin embargo, estos métodos dependen de la disponibilidad constante de los registros de cada variable climática.

3.3.2 Características o variables climáticas utilizadas para la predicción de heladas.

La mitad de los trabajos de investigación realizados en torno a la predicción de heladas a partir del año 2000 (53%), han implementado algoritmos de redes neuronales artificiales entrenadas con diversas variables climáticas de estaciones meteorológicas, sensores y otros instrumentos de medición. Como se muestra en la Tabla 3-2, las variables climáticas de mayor uso para la predicción de heladas fueron la temperatura, humedad relativa, velocidad del viento, radiación solar, precipitación y punto de rocío, las cuales presentaron una participación del 100 %, 73 %, 60 %, 43 %, 33 % y 23 % respectivamente, en el total de publicaciones revisadas.

Tabla 3-2. Variables climáticas utilizadas en las publicaciones para la predicción de heladas y cantidad de inclusión de cada una.

Variable climática	Número de inclusión en las publicaciones	Referencia
Temperatura	30	Ver referencias recopiladas en Tabla 3-1.
Humedad relativa	22	(Maqsood et al., 2004; Ovando et al., 2005; Ghielmi & Eccel, 2006; Jain et al., 2006; Smith et al., 2006; Prabha & Hoogenboom, 2008; Smith et al., 2009; Shamsnia et al., 2011; Wen et al., 2012; Lee et al., 2016; Mercado et al., 2016; Yu et al., 2016; Castañeda & Castaño, 2017; Diedrichs et al., 2018; Fuentes et al., 2018; Guillen et al., 2018; Ding et al., 2019; Cadenas et al., 2020; Guillén et al., 2020; Meza & Gutierréz, 2020; Ding et al., 2020; Zhou et al., 2022).
Velocidad del viento	18	(Maqsood et al., 2004; Ovando et al., 2005; Ghielmi & Eccel, 2006; Jain et al., 2006; Smith et al., 2006; Eccel et al., 2007; Prabha & Hoogenboom, 2008; Smith et al., 2009; Lee et al., 2016; Mercado et al., 2016; Yu et al., 2016; Castañeda & Castaño, 2017; Fuentes et al., 2018; Guillen et al., 2018; Ding et al., 2019;

Variable climática	Número de inclusión en las publicaciones	Referencia
		Cadenas et al., 2020; Guillén et al., 2020; Zhou et al., 2022).
Radiación solar	13	(Jain et al., 2006; Smith et al., 2006; Prabha & Hoogenboom, 2008; Smith et al., 2009; Mercado et al., 2016; Yu et al., 2016; Castañeda & Castaño, 2017; Fuentes et al., 2018; Guillen et al., 2018; Ding et al., 2019; Cadenas et al., 2020; Ding et al., 2020; Ding et al., 2021)
Precipitación	10	(Jain et al., 2006; Smith et al., 2006; Eccel et al., 2007; Smith et al., 2009; Shamsnia et al., 2011; Mercado et al., 2016; Fuentes et al., 2018; Guillen et al., 2018; Cadenas et al., 2020; Meza & Gutierréz, 2020).
Punto de rocío	7	(Ovando et al., 2005; Eccel et al., 2007; Lee et al., 2016; Fuentes et al., 2018; Guillen et al., 2018; Cadenas et al., 2020; Zhou et al., 2022).
Temperatura del suelo	5	(Ghielmi & Eccel, 2006; Prabha & Hoogenboom, 2008; Wen et al., 2012; Yu et al., 2016; Latif et al., 2020).
Nubosidad	4	(Verdes et al., 2000; Ovando et al., 2005; Eccel et al., 2007; Lee et al., 2016).
Dirección del viento	4	(Ovando et al., 2005; Fuentes et al., 2018; Guillen et al., 2018; Cadenas et al., 2020).
Humedad del suelo	4	(Prabha & Hoogenboom, 2008; Wen et al., 2012; Yu et al., 2016; Latif et al., 2020).
Déficit de presión de vapor	3	(Guillen et al., 2018; Cadenas et al., 2020; Ding et al., 2021).
Presión atmosférica	1	(Eccel et al., 2007).
Duración de la noche	1	(Ghielmi & Eccel, 2006).
Humedad específica	1	(Eccel et al., 2007).
Altura geopotencial	1	(Eccel et al., 2007).
Día pluviométrico	1	(Meza & Gutierréz, 2020).

Por otra parte, el 27 % de los trabajos de investigación publicados, que presentan relación con el problema en cuestión, corresponden a países ubicados en el centro y sur del continente americano (Argentina, Bolivia, Chile, Colombia, México). Particularmente en Colombia (7 %), Bonilla et al. (2006) y Meza & Gutierréz (2020) han realizado el pronóstico de heladas a partir del registro histórico de las variables de temperatura mínima y máxima del aire, temperatura de bulbo seco, precipitación y humedad relativa obtenidas de la

estación meteorológica del IDEAM “Tibaitatá” ubicada en el municipio de Madrid (Cundinamarca).

En la publicación de Bonilla et al. (2006) se utilizaron datos diarios de la temperatura de termómetro seco (°C) medida a las 7:00 a.m., 1:00 p.m. y 7:00 p.m., con un registro histórico comprendido entre los años 1990 y 2003 para la predicción de la temperatura mínima de la madrugada del día siguiente generada a las siete de la noche. De manera similar, Meza & Gutierréz (2020) realizaron el entrenamiento de modelos de regresión logística y redes bayesianas para la predicción de la temperatura mínima ($T_{mín}$) del día siguiente, con el fin de identificar la ocurrencia de un evento de helada ($T_{mín} \leq 0$ °C). La temperatura máxima del aire registrada por día (°C), la humedad relativa promedio (%), la precipitación ($mm.h^{-1}$) y la temperatura de bulbo seco mínima y máxima del día (°C) constituyeron las variables de entrada utilizadas por estos autores, las cuales fueron reportadas entre 2005 y 2019.

En ambos trabajos se identificó que debe realizarse un preprocesamiento de los datos para identificar valores perdidos, inconsistentes, duplicados, ruido y *outliers* debido a que las estaciones se descalibran cuando no se realizan con frecuencia procesos de mantenimiento. También, estos autores determinaron que el fenómeno de helada ocurre en los meses secos del año (diciembre, enero y febrero) y puede predecirse a partir de variables climáticas con un promedio de error de 2,16 °C y una efectividad del 73% al implementar redes neuronales sencillas (Bonilla et al., 2006) y superiores al 93% (TPR) con modelos de regresión logística y redes bayesianas (Meza & Gutierréz, 2020).

3.3.3 Métricas de evaluación de los algoritmos implementados en la predicción de heladas.

Las métricas empleadas para la evaluación del desempeño de los diferentes tipos de métodos enfocados en la predicción del valor de temperatura mínima (74 % de participación en las investigaciones revisadas), son el error medio absoluto (MAE), la desviación media absoluta (MAD), el error cuadrático medio (MSE), la raíz cuadrada media del error (RMSE), el coeficiente de correlación (CC), el porcentaje correcto (PC), el error porcentual absoluto medio (MAPE), el error de porcentaje absoluto máximo (MaxAPE) y el coeficiente de determinación (R^2). Asimismo, para los métodos de clasificación (26 %), cuyo fin es la detección de un evento de helada, se han empleado métricas de desempeño

como la probabilidad de detección (POD), la sensibilidad o Recall (TPR), precisión (P), exactitud (Ac), la tasa de falsos positivos (FPR), la tasa de alarmas falsas (FAR), la puntuación F1, el índice de éxito crítico (CSI), la tasa de aciertos (HIR) y el índice de simetría de errores (ISE). Adicionalmente, en la mayoría de las publicaciones se ha utilizado la validación cruzada como método de evaluación. Como se muestra en la Tabla 3-1, la métrica de evaluación MAE ha sido la más utilizada en las publicaciones evaluadas, seguida por TPR, Ac, CC, RMSE y MSE.

3.3.4 Desafíos relacionados con la predicción de heladas mediante el aprendizaje de máquinas.

Uno de los desafíos relacionados con la predicción de heladas identificado fue la insuficiente cantidad de datos, los cuales pueden no representar la varianza real de comportamiento de las variables climáticas, por lo tanto, se recomienda utilizar datos de diferentes circunstancias climáticas y vegetación, integrar diversas fuentes de datos y mayores periodos de tiempo, para lograr mejores predicciones y generar información de calidad disponible para los agricultores.

Otro de los desafíos identificados en los algoritmos de aprendizaje de máquinas para la clasificación en la predicción de la ocurrencia de los eventos de helada, donde la variable respuesta fue manejada como una variable binaria, hace referencia al desbalanceo de los datos. El desbalanceo se refiere a la distribución desigual de los datos en clases, por lo tanto, el rendimiento de los clasificadores se afecta debido a que las predicciones se direccionan hacia la clase que tiene una mayor cantidad de datos, generando clasificadores con rendimientos sesgados. Las clases que presentan una mayor cantidad de datos son conocidas como clases mayoritarias (Patel et al., 2020). Por el contrario, las clases minoritarias son las que tiene pocas instancias o registros de datos. Generalmente, las clases minoritarias son las que se necesitan pronosticar, por lo tanto, es necesario implementar métodos para incrementar su reconocimiento (Patel et al., 2020).

A partir de la revisión de literatura sobre la clasificación de datos desbalanceados para redes de sensores inalámbricos, Patel et al. (2020), identificaron que las estrategias para tratar el desequilibrio de los datos se agrupan en re-muestreo (re-sampling), preprocesamiento de los datos y aplicación de clasificadores tradicionales o modificación

del clasificador. En este sentido, García et al., 2012 en Patel et al. (2020), al comparar métodos de re-muestreo, encontraron que para una relación de desbalanceo baja, es indiferente utilizar estrategias de sobremuestreo o submuestreo. Sin embargo, para obtener una mejor clasificación con datos que presentan un desbalanceo alto, se recomienda el uso de estrategias de sobremuestreo como la técnica de sobremuestreo sintético de minorías (SMOTE del inglés Synthetic Minority Oversampling Technique), la cual genera información sintética entre ejemplos minoritarios a partir de la similitud basada en características y, como consecuencia, el clasificador tradicional mejora el límite de decisión cerca de los ejemplos minoritarios. Varios investigadores han encontrado que el sobremuestreo influye significativamente en el rendimiento de los clasificadores, tanto al implementar la estrategia SMOTE como otros enfoques que han sido desarrollados tal como el algoritmo de sobremuestreo sintético adaptativo (ADASYN del inglés oversample using adaptive synthetic) (Patel et al., 2020).

Finalmente, Patel et al. (2020) recomiendan utilizar estrategias de sub o sobremuestreo, disminuir el peso de los atributos con mayor presencia y aumentar el peso de los atributos con menor cantidad de instancias, el aprendizaje sensible a los costos para evitar la preselección de hiperparámetros mediante un ajuste dinámico y el uso de métodos combinados para el manejo de datos desequilibrados.

3.4 Conclusiones.

La revisión de la literatura permitió identificar que es posible pronosticar la ocurrencia de eventos de helada mediante algoritmos de aprendizaje de máquinas a partir del registro histórico de diversas variables climáticas en diferentes periodos de tiempo. Entre las variables climáticas que han sido utilizadas para la predicción de este fenómeno se tiene que, la temperatura, humedad relativa, velocidad del viento, radiación solar, precipitación y punto de rocío corresponden al 79 % del total de variables utilizadas. Sin embargo, la selección de estas variables depende de la disponibilidad del conjunto de datos.

Los métodos de mayor uso en la predicción de heladas fueron las redes neuronales artificiales para la predicción de la temperatura mínima y métodos de clasificación como los árboles de decisión, la regresión logística, los bosques aleatorios y las máquinas de soporte vectorial.

Las publicaciones relacionadas con la predicción de heladas desde el año 2021 son relativamente bajas y únicamente el 27 % del total de las investigaciones se han realizado en países latinoamericanos. Para Colombia, se han efectuado algunas aproximaciones, empero, estos métodos de pronóstico dependen de la disponibilidad de información de las estaciones meteorológicas del IDEAM, las cuales sólo tienen registros de un valor por variable climática al día y ha sido necesario implementar procesos para la limpieza y obtención de datos de calidad. Por lo tanto, el monitoreo de las variables meteorológicas en tiempo real a lo largo del día puede aportar información tanto para analizar el comportamiento de las heladas en países tropicales, como para la generación de predicciones y su posterior inclusión como una de las estrategias de prevención y mitigación de heladas.

4. Predicción temprana de heladas en cultivos de altura, empleando métodos de aprendizaje de máquinas.

En este capítulo se aborda la propuesta metodológica para el desarrollo, implementación y evaluación de diferentes algoritmos de clasificación y regresión, seleccionados a partir de la información recopilada en la revisión de la literatura, para el pronóstico de la ocurrencia de eventos de heladas a partir de cinco variables meteorológicas monitoreadas con tecnología IoT, en trece estaciones meteorológicas ubicadas en el altiplano cundiboyacense. La primera sección hace referencia a la caracterización de los datos históricos de las variables climáticas para la predicción temprana de heladas en cultivos de altura mediante exploración, descripción y aplicación de técnicas de limpieza de datos. Posteriormente, se exponen los métodos para la identificación de las variables climáticas más relevantes para la predicción temprana de heladas en cultivos de altura y para la selección del método de aprendizaje de máquinas de acuerdo con las métricas de rendimiento. Este capítulo finaliza con la validación del desempeño del método seleccionado para la predicción temprana de heladas en cultivos de altura.

4.1 Caracterización de los datos históricos de las variables climáticas.

4.1.1 Área de estudio.

El altiplano Cundiboyacense presenta un rango altitudinal entre 460 y 4.240 msnm., por lo tanto, hay un amplio rango de unidades biogeográficas (Rivera et al., 2004 in Gómez et al., 2021). La región presenta un régimen de lluvia bimodal con una precipitación anual entre 580 y 1.000 mm año^{-1}, los valores más altos se presentan desde abril a mayo y de octubre a noviembre (Aguilar & Torres, 2010; Guhl, 2013). Además, esta región tiene una

temperatura media anual de 13,5 °C y mantiene un amplio rango de oscilaciones diarias entre 3 °C y 28 °C. Temperaturas por debajo de 2 °C generalmente ocurren en las primeras horas de la mañana de los meses secos (Aguilar & Torres, 2010; Guhl, 2013).

La obtención de la información climática se realizó a partir de trece estaciones meteorológicas ubicadas al interior de cultivos de flores, distribuidos en diez municipios del departamento de Cundinamarca (Chía, Facatativá, Gachancipá, Madrid, Nemocón, Sesquilé, Suesca, Tocancipá, Ubaté y Zipaquirá) (Figura 4-1). Wiga SAS y Growers Hub Trading (GHT) son las empresas que gestionan las estaciones.

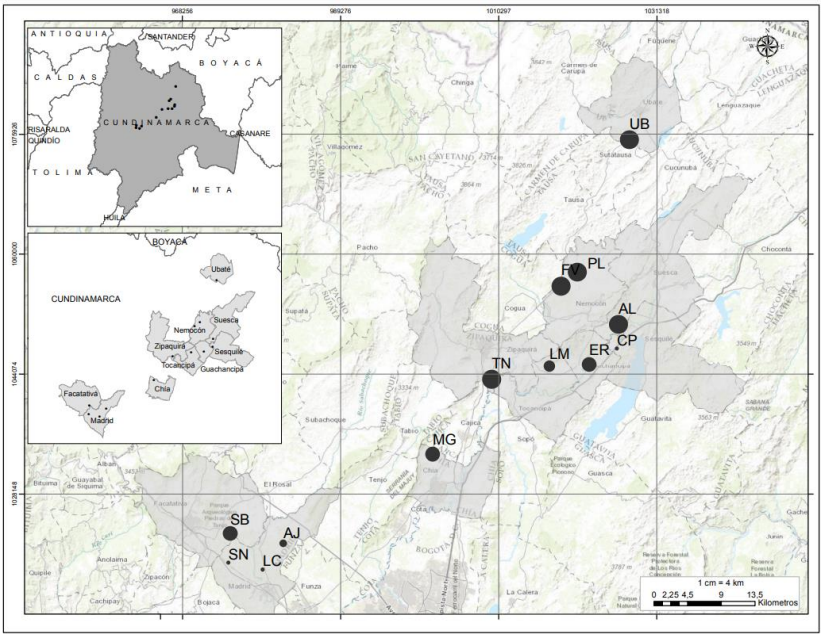


Figura 4-1. Ubicación de las estaciones meteorológicas en el departamento de Cundinamarca (El tamaño del círculo corresponde a la cantidad de datos). Nombre de la fuente: “Elaboración propia”.

4.1.2 Variables meteorológicas.

Los registros adquiridos entre 2017 y 2022 tuvieron una resolución de una hora (24 registros día^{-1}). El número total de registros se detalla en la Tabla 4-1. La fecha de instalación entre estaciones fue diferente, lo que implica discrepancias en el número total de registros. Dado el carácter periódico del fenómeno de las heladas, este estudio se centró en la temporada seca del año, que corresponde a los meses de noviembre, diciembre, enero y febrero (Figura 4-2).

Tabla 4-1. Altitud, municipio y número de registros totales por estación meteorológica.

Estación	Altitud (m s.n.m.)	Municipio	Fecha inicial	Fecha final	Días registrados	Registros totales
MG consultores (MG)	2.564	Chía	Nov 2018	Feb 2022	465	11.160
Finca Santa Barbara (SB)	2.582	Facatativá	Nov 2018	Feb 2022	477	11.448
Flores El Rosal (ER)	2.566	Guachancipá	Nov 2018	Feb 2022	480	11.520
Flores Las Acacias (LC)	2.562	Madrid	Nov 2020	Feb 2022	224	5.376
Santana Flowers (SN)	2.583	Madrid	Nov 2020	Feb 2022	235	5.640
Flores Alejandra (AL)	2.567	Madrid	Dec 2019	Feb 2022	301	7.224
Cultivos La Planicie (PL)	2.584	Nemocón	Nov 2017	Feb 2022	580	13.920
Florval (FV)	2.575	Nemocón	Nov 2017	Feb 2022	583	13.992
Flores El Ciprés (CP)	2.568	Sesquilé	Nov 2020	Feb 2022	232	5.568
Flores El Aljibe (AJ)	2.568	Suesca	Nov 2017	Feb 2022	573	13.752
Flores La Mana (LM)	2.567	Tocancipá	Jan 2019	Feb 2022	416	9.984
Flores Ubaté (UB)	2.575	Ubaté	Nov 2017	Feb 2022	585	14.040
Flores El Tandil (TN)	2.569	Zipaquirá	Nov 2017	Feb 2022	578	13.872

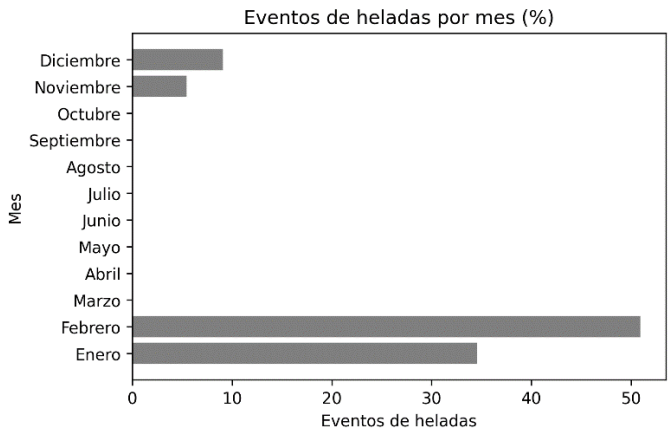


Figura 4-2. Cantidad de eventos de heladas por mes. Nombre de la fuente: “Elaboración propia”.

A partir de la información de las trece estaciones se identificaron los eventos de helada (registros de temperatura inferiores a 0 °C) como variable respuesta y cinco variables climáticas como variables explicativas. Las variables climáticas fueron la temperatura (T, °C), radiación fotosintéticamente activa (RFA, µmol m^{-2} s^{-1}), humedad relativa (HR, %), punto de rocío (PR, °C) y precipitación (PP, mm $m^{-2}h^{-1}$). La variable respuesta (ocurrencia de un evento de helada), se manejó como una variable binaria, tomando el valor de uno (1) las 24 horas anteriores del día en que la temperatura fue menor a 0 °C. Es decir, al momento de ocurrir un evento de helada, la magnitud de las variables climáticas explicativas 24 horas del día anterior se asoció con un valor de uno en la variable respuesta. En cualquier otro caso, la variable respuesta tomó el valor de cero, asociado a las magnitudes de las variables climáticas correspondientes. Cada variable explicativa se refiere a una combinación de las variables climáticas y la hora del día; por lo tanto, el conjunto de datos resultante para la exploración de los datos estuvo compuesto por 123 columnas (5 variables climáticas x 24 horas del día), incluyendo el día, mes y año de cada registro Tabla 4-2.

Tabla 4-2. Conjunto de datos para la predicción de heladas con variable respuesta binaria (“Helada”).

HR0	…	HR23	PP0	…	PP23	PR0	…	PR23	RFA0	RFA6	…	RFA23	T0	…	T23	Año	Mes	Día	Helada
97.5	…	95.5	0	…	0	8.7	…	10.2	0	874.1	…	0	9.1	…	10.9	2018	1	2	0
96.7	…	95.1	0	…	0	9	…	9.2	0	712.9	…	0	9.5	…	10	2018	1	2	0
97.6	…	96.5	0	…	0	8.4	…	9.8	0	796.7	…	0	8.7	…	10.3	2018	1	2	1
96.9	…	93.3	1.1	…	0	10	…	10.7	0	700	…	0	10.4	…	11.7	2018	1	2	0
95.8	…	92.3	0	…	0	8.2	…	10.3	0	1383.8	…	0	8.8	…	11.5	2018	1	2	0
95.4	…	96.7	0	…	0	9.3	…	7.8	0	1667.7	…	0	10	…	8.3	2018	1	3	0
93.1	…	96.1	0	…	0	7.8	…	7.8	0	1193.5	…	0	8.8	…	8.4	2018	1	3	0
96.8	…	96.4	0	…	0	9	…	6.9	0	1606.4	…	0	9.5	…	7.4	2018	1	3	1
94.2	…	95.6	0	…	0	10	…	8.5	0	1503.2	…	0	10.9	…	9.2	2018	1	3	0

Donde HR: humedad relativa, PR: Punto de rocío, PP: Precipitación, RFA: Radiación fotosintéticamente activa, T: Temperatura, Tmín: Temperatura mínima. Cada número representa la hora del día, donde 0 hace referencia a las 12:00 a.m. y 23 corresponde a las 11:00 p.m.

De manera similar, se construyó el conjunto de datos para la predicción de la temperatura mínima del día siguiente. Las variables de entrada o explicativas estaban constituidas por la combinación de las variables climáticas y la hora del día; por lo tanto, el conjunto de datos resultante para el entrenamiento de los modelos de regresión estuvo compuesto por 123 columnas (5 variables climáticas x 24 horas del día), incluyendo el día, mes y año de cada registro. Sin embargo, la variable respuesta fue la temperatura mínima de la madrugada del día siguiente Tabla 4-3.

Tabla 4-3. Conjunto de datos para la predicción de heladas con variable respuesta continua (“Tmín”).

HR0	…	HR23	PP0	…	PP23	PR0	…	PR23	RFA0	RFA6	…	RFA23	T0	…	T23	Año	Mes	Día	Tmín
97.5	…	95.5	0	…	0	8.7	…	10.2	0	874.1	…	0	9.1	…	10.9	2018	1	2	10.8
96.7	…	95.1	0	…	0	9	…	9.2	0	712.9	…	0	9.5	…	10	2018	1	2	11.5
97.6	…	96.5	0	…	0	8.4	…	9.8	0	796.7	…	0	8.7	…	10.3	2018	1	2	-0.6
96.9	…	93.3	1.1	…	0	10	…	10.7	0	700	…	0	10.4	…	11.7	2018	1	2	8.9
95.8	…	92.3	0	…	0	8.2	…	10.3	0	1383.8	…	0	8.8	…	11.5	2018	1	2	12.4
95.4	…	96.7	0	…	0	9.3	…	7.8	0	1667.7	…	0	10	…	8.3	2018	1	3	10.9
93.1	…	96.1	0	…	0	7.8	…	7.8	0	1193.5	…	0	8.8	…	8.4	2018	1	3	12.3
96.8	…	96.4	0	…	0	9	…	6.9	0	1606.4	…	0	9.5	…	7.4	2018	1	3	-1.9
94.2	…	95.6	0	…	0	10	…	8.5	0	1503.2	…	0	10.9	…	9.2	2018	1	3	9.1

4.1.2 Análisis descriptivo.

Se implementaron las estadísticas más comunes (mínimo, 1er, 2do y 3er cuartil, máximo, curtosis y asimetría) para describir el comportamiento anual de cada variable climática y compararlo con el comportamiento de estas en la temporada de heladas (entre noviembre y febrero). Adicionalmente, se realizó una descripción detallada de las variables climáticas para diferenciar el comportamiento promedio horario de cada variable, durante las 24 horas previas a la ocurrencia de un evento de helada (EH) vs las horas no asociadas a la ocurrencia de un evento de helada (NEH). Las horas se etiquetaron utilizando el sistema horario de 24 horas, donde 0 corresponde a las 12 a.m. y 23 indica las 11 p.m. Finalmente, se evidenció el comportamiento de cada variable climática mediante histogramas de frecuencia, evaluación de la normalidad mediante la prueba de Kolmogorov-Smirnov y la comparación del promedio de cada variable por temporada de helada.

4.1.3 Preprocesamiento de datos.

El proceso de limpieza incluyo la identificación de valores perdidos, inconsistentes, duplicados y atípicos. Los datos faltantes se imputaron por medio de la aproximación del vecino más cercano KNNImputer (Juna et al., 2022), de la biblioteca “Sklearn Impute” de python, y se estandarizaron con la función Sklearn StandarScale (Ecuación 4-1) (Pedregosa et al., 2011) para evitar el efecto de su magnitud. Tanto para la identificación de las variables climáticas de mayor relevancia en la predicción de heladas como para el entrenamiento de los modelos, únicamente se tuvo en cuenta los días con los veinticuatro registros completos.

Ecuación 4-1

$$z = \frac{(x-u)}{s}$$

Donde x corresponde a cada observación, u y s corresponden a la media y desviación estándar de las muestras de entrenamiento, respectivamente.

Teniendo en cuenta que los algoritmos de clasificación de aprendizaje de máquinas son sensibles al desbalanceo de los datos y que los eventos positivos (ocurrencia de heladas) correspondían únicamente al 2% de los datos, se implementaron y compararon tres estrategias para el manejo de datos desbalanceados, las cuales fueron el ajuste de

parámetros del modelo para la penalización de la clase mayoritaria, el método de sobremuestreo SMOTE (del inglés Synthetic Minority Oversampling Technique) y submuestreo en la clase mayoritaria.

Mediante la estrategia de submuestreo se seleccionó un subconjunto de datos de la clase mayoritaria mediante un algoritmo de dos pasos; en el primero para cada muestra de la clase minoritaria, se mantienen los *M* vecinos más cercanos, posteriormente, se seleccionan las muestras de la clase mayoritaria para las cuales la distancia promedio a los *N* vecinos más cercanos es mayor. Este algoritmo no es tan susceptible a valores atípicos debido a la selección de la muestra del primer paso (Zhang & Mani, 2003).

El balanceo mediante los métodos de sobre y submuestreo se realizó con las funciones “SMOTETomek” y “NearMiss” respectivamente, de la librería Imbalanced learn de Python. La penalización en los algoritmos de regresión logística, arboles de decisión y la máquina de soporte vectorial se realizó incluyendo el parámetro de peso de las clases o “class weight” equivalente a equilibrado o ‘balanced’. Mediante la penalización, se asignó una ponderación para cada clase de datos; los pesos se ajustaron para ser inversamente proporcionales a la frecuencia de cada clase de los datos de entrada (Pedregosa et al., 2011). Dado que las redes neuronales de la librería de Sklearn no presentan el parámetro para la penalización de las clases, se utilizó la librería “Keras”, mediante la cual se asignó la función de perdida de entropía cruzada binaria o “binary cross entropy” (Ho & Wookey, 2020).

4.2 Identificación de las variables climáticas de mayor relevancia para la predicción temprana de heladas en cultivos de altura.

Posterior al preprocesamiento de los datos, se evaluó el tipo y el grado de relación entre los eventos de heladas y las variables climáticas según la hora del día con el coeficiente de correlación de Spearman (CCS). También se identificó la existencia de diferencias significativas entre los valores de cada variable climática por hora relacionados y no relacionados con eventos de helada a través de la prueba de Kruskal Wallis. Esta es una

prueba no paramétrica que determina si las medianas de dos o más grupos son diferentes al contrastar la hipótesis nula (todas las muestras provienen de una misma población) con la hipótesis alternativa (al menos una muestra proviene de una población con una distribución diferente) (Hecke, 2012).

La selección de las variables climáticas más relacionadas con los eventos de heladas y el grado de relación según la hora del día se evaluó mediante el análisis de componentes principales y la selección de características con los puntajes más altos mediante árboles aleatorios y la función SelectKBest de Sklearn. Está ultima metodología permite seleccionar k variables independientes, mediante una medición directa de la relación entre cada característica y la variable respuesta sin tener en cuenta las interacciones entre características (Pedregosa et al., 2011). La selección de características mediante los árboles aleatorios se realizó mediante la obtención del valor de la importancia de cada variable basado en el índice de impureza de Gini (Dimitriadis et al., 2018). Adicionalmente, se empleó la metodología de la mínima redundancia y máxima relevancia como método rápido para determinar el conjunto de características relevantes y complementarias a través de la librería “mrmr” de Python (Bugata & Drotar, 2020).

Las variables meteorológicas con las horas especificas seleccionadas como las de mayor influencia en los eventos de heladas se usaron como variables de entrada para los modelos de predicción. No se tuvieron en cuenta las variables de entrada altamente correlacionadas entre sí (CCS > 80%).

4.3 Selección del método de aprendizaje de máquinas para la predicción temprana de heladas en cultivos de altura.

Con el objetivo de seleccionar un método de aprendizaje de máquinas para la predicción temprana de heladas en cultivos de altura se evaluaron cinco métodos para la identificación de un evento de helada como variable respuesta categórica binaria (regresión logística, árbol de decisión, árboles de decisión de aumento del gradiente, máquina de soporte vectorial y red neuronal). Adicionalmente, para identificar la severidad de la helada es imprescindible conocer el valor al que descenderá la temperatura para ajustar la cantidad de agua a aplicar mediante el sistema de riego por aspersión, por lo tanto, se evaluaron

dos algoritmos para la predicción de la temperatura mínima que se alcanzará a partir de la construcción de redes neuronales y bosques aleatorios.

4.3.1 Método de predicción.

- **Predicción de heladas mediante métodos de clasificación.**

Se dividió el conjunto de datos en tres subconjuntos, el primero con un 60 % para el entrenamiento del modelo por medio de validación cruzada de cinco iteraciones, el segundo (20 %) para pruebas y el tercero (20 %) para validación. El subconjunto para el entrenamiento de los modelos represento el comportamiento del total de registros como se muestra para la humedad y la temperatura en la Figura 4-3. Las métricas de rendimiento (precisión, sensibilidad, especificidad, exactitud y puntuación F1) se calcularon en función de los datos de prueba y validación. Los métodos predictivos utilizados fueron la regresión logística (LR), árbol de decisión (DT), árboles de decisión de aumento de gradiente (GBDT), máquina de soporte vectorial (SVM) y red neuronal (NN). La optimización del aprendizaje y ajuste de los algoritmos se realizó con la función pipeline y el método gridsearcCV de la librería Sklearn para el ajuste de hiperparámetros (Tabla 4-4). En el entrenamiento de los modelos, únicamente se tuvo en cuenta la información de cada variable climática hasta las 16:00 horas, para tener una ventana de tiempo de 14 horas para el despliegue de las estrategias de mitigación.

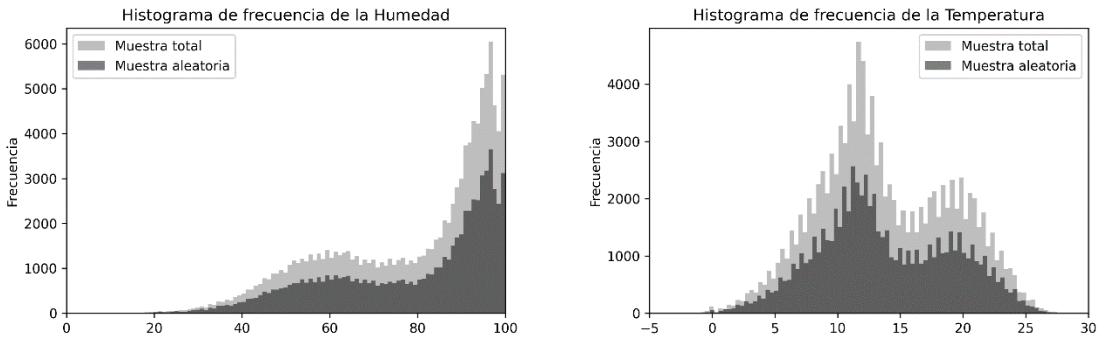


Figura 4-3. Histograma de frecuencia para la humedad relativa (a) y la temperatura (b) para la muestra total (gris claro) y el 60% de los datos utilizados para el entrenamiento de los modelos (gris oscuro). Nombre de la fuente: “Elaboración propia”.

Tabla 4-4. Valores de ajuste para cada hiperparámetro por método.

Método	Librería	Hiperparámetros	Rango de valores para cada hiperparámetro
LR[1]	LogisticRegression (Sklearn)	C (Inverso de la fuerza de regularización)	{0,001 – 50}
		Regularización	{l1, l2, elasticnet}
		Solucionador (solver)	{newton-cg, lbfgs, liblinear, sag, saga}
DT[2]	DecisionTreeClassifier (Sklearn)	Criterio	{gini, entropy}
		Divisor (splitter)	{best, random}
		Profundidad máxima del árbol	{1 - 50}
		Número mínimo de muestras	{1 - 20}
		Máximo número de nodos de hoja	{1 - 50}
GBDT[3]	GradientBoostingClassifier (Sklearn)	Tasa de aprendizaje	{0,0001 – 0,3}
		Número mínimo de muestras	{1 - 50}
		Profundidad máxima	{1 - 50}
		Número máximo de características	{1 - 20}
SVM[4]	SVC (Sklearn)	Regularización	{0,5 – 30}
		Tipo de Kernel	{linear, poly, rbf, sigmoid, precomputed}
		Grado de la función del Kernel polinomial	{1 – 5}
		Gamma	{auto, scale}
NN[5]	MLPRegressor (Sklearn)	Número de neuronas en la i-ésima capa oculta	{(2,), (5,), (10,), (20,), (50,)}
		Función de activación para la capa oculta	{identity, logistic, tanh, relu}
		Solucionador (solver)	{lbfgs, sgd, adam}
		Regularización L2	{0,001 – 1}

Método	Librería	Hiperparámetros	Rango de valores para cada hiperparámetro
		Tasa de aprendizaje	{0,005; 0,01; 0,05; 0,075; 0,1; 0,5}
		Tamaño del lote o Batch size	{50 – 400}

[1]Regresión logística. [2]Árbol de decisión. [3]Arboles de decisión de aumento del gradiente. [4]Máquina de soporte vectorial. [5]Red neuronal.

- **Predicción de la temperatura mínima mediante métodos de regresión.**

Para modelar la relación entre la temperatura mínima y las variables climáticas registradas según la hora del día con mayor relación con las heladas, se entrenó un algoritmo de bosques aleatorios. En el entrenamiento se estableció la cantidad de árboles y el número de variables de entrada por nodo. Los valores de estos parámetros se optimizaron mediante el conjunto de datos de entrenamiento y la obtención del RMSE. Se probaron valores para la cantidad de árboles entre 100 y 1000 con intervalos de 100 y se probó el número de variables de entrada por nodo entre 3 y 50; se seleccionaron los valores que produjeron un RMSE inferior. Para cada ventana de tiempo se estimó la cantidad de árboles y el número de variables de entrada por nodo. El entrenamiento se realizó mediante la función RandomForestRegressor de la librería Sklearn con el conjunto de entrenamiento y se probó en los conjuntos de validación y prueba.

Por otro lado, se desarrolló una red neuronal artificial multicapa basado en el método de descenso del gradiente, usando el algoritmo de retropropagación para el cálculo del gradiente descendente. Se configuraron tres capas (capa de entrada, ocultas y de salida). La capa de entrada contenía las variables climáticas para cada hora del día seleccionadas en la sección anterior, en la capa oculta se evaluaron 1, 5 10, 15, 20, 25 y 30 neuronas y se implementó una función de activación de unidad lineal rectificada (ReLu). Finalmente, se utilizó una capa de salida que hace referencia a la variable a predecir. En la capa de salida se utilizó una función de transferencia lineal.

El proceso de entrenamiento se detuvo cuando se presentó un aumento en el MSE, indicando que no se presentarán mejoras en la generalización. Las derivadas parciales del MSE se usaron en este método para ajustar los pesos mediante el algoritmo del gradiente

descendente (Fuentes et al., 2018). El entrenamiento se realizó mediante la librería Keras con el conjunto de entrenamiento y se probó en los conjuntos de validación y prueba.

En el entrenamiento de ambos métodos, se compararon cuatro ventanas de tiempo diferentes (una hora, tres horas, siete horas y diez horas) para la predicción de la temperatura mínima, debido a que la magnitud de la temperatura permite realizar un ajuste de la cantidad de agua a aplicar (Castañeda & Castaño, 2017) y podrá usarse en invernaderos automatizados, los cuales necesitan una ventana de tiempo menor para la activación del sistema de aspersión o ventilación.

4.3.2 Métricas de evaluación de rendimiento y comparación de los métodos de predicción.

- **Predicción de heladas mediante métodos de clasificación.**

Las métricas para la evaluación del rendimiento y la comparación de los métodos de predicción de clasificación binaria fueron la precisión (proporción de casos clasificados correctamente de todos los positivos reales), sensibilidad (proporción de casos positivos clasificados correctamente), especificidad (proporción de casos negativos clasificados correctamente), exactitud (proporción de casos clasificados correctamente de todos los puntos de datos), y puntuación F1 (promedio ponderado entre precisión y sensibilidad) (Fuentes et al., 2018) (Tabla 4-5). Estas métricas de evaluación del rendimiento se calcularon para todos los métodos tanto en los conjuntos de datos de prueba como en los de validación.

Tabla 4-5. Descripción de las métricas de evaluación de rendimiento de los métodos de clasificación.

Métricas	Ecuación
Precisión (p)	$p = \frac{VP}{VP + FP}$
Sensibilidad (TPR)	$TPR = \frac{VP}{VP + FN}$
Especificidad (TNR)	$TNR = \frac{VN}{VN + FP}$
Exactitud (Ac)	$Ac = \frac{VP + VN}{VP + VN + FP + FN}$

Puntuación F_1	$F1 = 2\frac{precision\,.recall}{precision + recall}$

VP es el número de valores predichos de manera acertada; en otras palabras, son verdaderos positivos o eventos de heladas correctamente pronosticados; FP son los falsos positivos o eventos no asociados con heladas pronosticados como heladas; FN son los falsos negativos o eventos de heladas que no se pronosticaron y VN corresponde a verdaderos negativos o días sin ocurrencia de heladas correctamente pronosticados. Estas categorías se especificaron utilizando la tabla de contingencia (Tabla 4-6) y la matriz de confusión (Diedrichs et al., 2018). Se utilizaron dos parámetros de rendimiento adicionales que fueron la relación entre la tasa de verdaderos positivos (sensibilidad) y la tasa de falsos positivos (Cho et al., 2021), denominada curva característica operativa del receptor (ROC), y el área bajo la curva de precisión-sensibilidad (AUC-PR) (DeVries et al., 2021).

Tabla 4-6. Tabla de contingencia para comparación entre los valores pronosticados y observados para las clases binarias.

Observación	**Predicción**	
	Ocurrencia de heladas (PP)	**No ocurrencia de heladas (PN)**
Ocurrencia de heladas (P)	VP[1]	FN[2]
No ocurrencia de heladas (N)	FP[3]	VN[4]

[1]Verdaderos positivos. [2]Falsos negativos. [3]Falsos positivos. [4]Verdaderos negativos.

- **Predicción de la temperatura mínima mediante métodos de regresión.**

Para la evaluación de los métodos de predicción de la temperatura mínima (redes neuronales y bosques aleatorios) se utilizó el error medio absoluto (MAE), el error cuadrático medio (MSE) y la raíz cuadrada media del error (RMSE), las cuales se calcularon mediante las ecuaciones de la Tabla 4-7.

Tabla 4-7. Descripción de las métricas de evaluación de los métodos de regresión.

Métricas	Ecuación[1]
Error medio absoluto	$MAE = \frac{1}{n}\sum_{i=1}^{n} \lvert y_i - \hat{y}_i \rvert^2$
Error cuadrático medio	$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$
Raíz cuadrada media del error	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$

[1]Donde y_i es el valor predicho y $\hat{y}_i$ es el valor observado.

Adicionalmente, para la evaluación de los métodos de predicción de la temperatura mínima, se realizó una clasificación de los valores predichos con el conjunto de prueba para determinar la cantidad de falsos positivos y falsos negativos (Guillén et al., 2020). Los falsos positivos hacen referencia al error tipo 1 y los falsos negativos están relacionados con el error tipo 2. Para la clasificación se consideraron eventos de helada cuando la temperatura predicha fue inferior o igual a 0 °C, en caso contrario, se consideró como NEH. A partir de esta clasificación se identificaron los verdaderos positivos (VP), verdaderos negativos (VN), falsos positivos (FP) y falsos negativos (FN) y se calcularon las métricas descritas para los métodos de clasificación (Tabla 4-5).

4.4 Validación del desempeño del método seleccionado para la predicción temprana de heladas en cultivos de altura.

La validación del desempeño de los métodos de clasificación binaria y regresión se realizó mediante el cálculo de las métricas expuestas en la sección anterior (Tabla 4-5 y Tabla 4-7) para el conjunto de datos de validación. Estos resultados se compararon con las métricas obtenidas en trabajos similares reportados en la literatura.

El mejor algoritmo de clasificación binaria se implementó en la web, para que, a partir de la información de temperatura, humedad relativa, punto de rocío y radiación registrada diariamente en las trece estaciones, los usuarios puedan obtener información acerca de la

probabilidad de ocurrencia de un evento de helada en la madrugada del día siguiente para cada uno de los diez municipios del área de estudio.

La implementación de los métodos en el sitio web se ejecutó por medio de una interfaz de programación de aplicaciones (API). La creación del API se realizó en Azure Databricks, el cual proporciona una versión integrada de MLflow, que permite la ejecución, seguimiento y registro de modelos con el código utilizado, los parámetros, datos de entrada, métricas y archivos de salida. Además, permite la generación de una API de tipo transferencia de estado representacional (REST). El API está alojada en la ruta en el servidor de seguimiento de MLflow y se puede insertar en cualquier lugar donde los usuarios ejecuten código (Zaharia et al., 2018).

Posterior a la generación del API, esta se implementó en un sitio web mediante el lenguaje de programación C Sharp. En este se realizó el desarrollo del back end y front end. Los datos de las variables climáticas en tiempo real para la generación de las nuevas predicciones, a partir de la información de las estaciones meteorológicas, se consultaron desde SQL. Los valores predichos también se almacenaron en SQL (Figura 4-4).

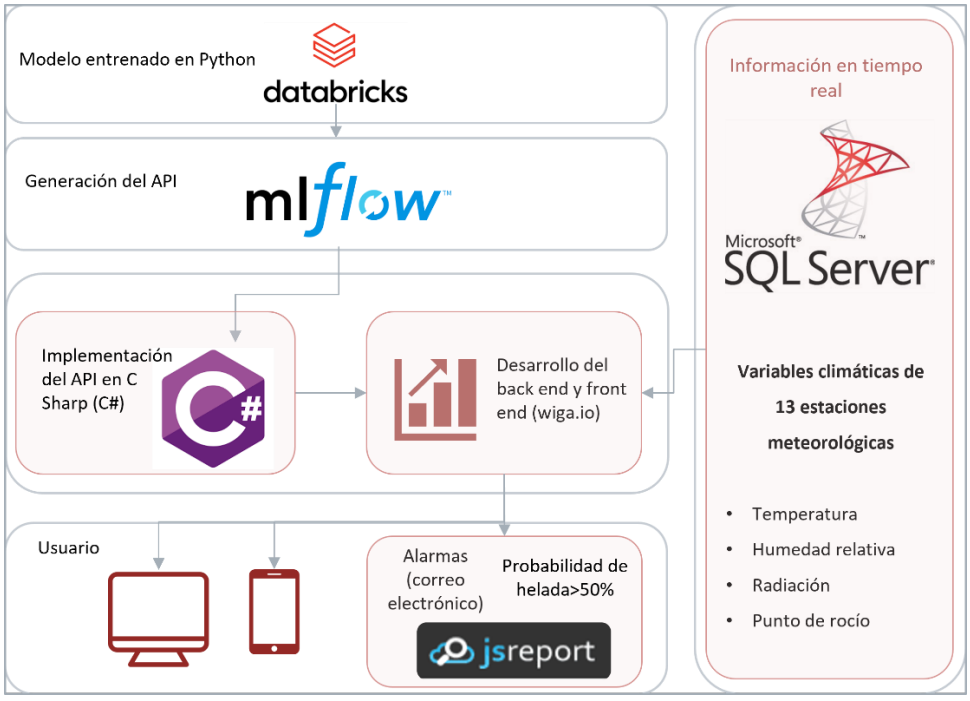


Figura 4-4. Diagrama para la implementación de la predicción de heladas en la web. Nombre de la fuente: “Elaboración propia”.

5. Experimentos y resultados

En la primera sección de este capítulo se recopilaron los resultados obtenidos relacionados con la descripción del comportamiento de las cinco variables climáticas que se utilizaron para la predicción de heladas y se evidenciaron las diferencias en el comportamiento de estas variables por hora para los días no asociados con eventos de heladas en contraste con los días previos a la ocurrencia de una helada. Posteriormente, se identificaron las variables climáticas y las respectivas horas del día en que estas presentan una mayor influencia en la ocurrencia de este fenómeno. Finalmente, se presenta el desempeño de los algoritmos de aprendizaje de máquinas tanto para la identificación de la ocurrencia de heladas como una variable respuesta categórica binaria, como para la predicción de la temperatura. El desempeño de cada método fue evaluado con el conjunto de datos de prueba y de validación.

5.1 Caracterización de los datos históricos de las variables climáticas.

La región en estudio se caracterizó por presentar una temperatura promedio anual de 13,6 °C con valores entre -3,1 y 28,7 °C, estos valores son similares a los reportados por Aguilar & Torres (2010); Guhl (2013) & Mayorga et al. (2020); además, para el periodo evaluado el 75 % de los valores de temperatura fueron inferiores a 17,1 °C. Con relación a la HR, el valor promedio anual fue de 79,8%, con una mediana de 86,5 % y valores entre 8,3 y 100 %; este valor promedio anual estuvo dentro del rango promedio especificado por Luengas et al. (2021) para el departamento de Cundinamarca (60 a 80 %). Adicionalmente, el promedio anual para la RFA, PR y PP fueron 364,9 (0 a 3.300) µmol m^{-2} s^{-1}, 9,7 (-10,6 a 19,9) °C, y 0,02 (0 a 60,4) mm m^{-2} h^{-1} respectivamente, como se muestra en la Tabla 5-1.

El promedio anual de la RFA estuvo entre los valores reportados por Mayorga et al. (2020) para el municipio de Pasca (Cundinamarca), localizado a una altitud de 2.498 m s.n.m. Estos autores reportaron valores promedio mensuales entre 300 y 500 µmol m^{-2} s^{-1}.

Tabla 5-1. Análisis descriptivo de las variables climáticas a lo largo del año (anual) en comparación con el comportamiento climático en la temporada de helada (Nov – Feb) y la temporada no asociada a heladas (Mar – Oct).

Variable climática*	Estadísticas									
	Promedio	**DS[1]**	**Min[2]**	**Mediana**	**Q1[3]**	**Q2[4]**	**Q3[5]**	**Max[6]**	**Curt[7]**	**As[8]**
Anual										
T (°C)[9]	13,6	4,7	-3,1	12,9	10,5	12,9	17,1	28,7	-0,3	0,1
HR (%)[10]	79,8	17,8	8,3	86,5	66,3	86,5	95	100	-0,3	-0,8
RFA (µmol m^{-2} s^{-1})[11]	364,9	587,1	0	9,7	0	9,7	564,5	3300	2,9	1,8
PR (°C)[12]	9,7	2,5	-10,6	10,2	8,8	10,2	11,3	19,9	8,3	-1,9
PP (mm m^{-2} h^{-1})[13]	0,02	0,5	0	0	0	0	0	60,4	8626,3	78,47
Temporada de heladas										
T (°C)	13,6	5,3	-3,1	12,8	9,9	12,8	18,0	28,7	-0,7	0,1
HR (%)	78,7	18,6	10,5	86	63,6	86,0	94,6	100	-0,5	-0,8
RFA (µmol m^{-2} s^{-1})	367,7	580	0	3,4	0	3,4	568,3	3300	2,2	1,7
PR (°C)	9,4	2,7	-14,4	10	8,1	10	11,3	19,9	1,4	-1
PP (mm m^{-2} h^{-1})	0,01	0,7	0	0	0	0,2	0,4	0,9	7237,3	118,6
Temporada no asociada a heladas										
T (°C)	13,7	4	0	13	11	13	16,7	28,1	-0,1	0,3
HR (%)	81,1	15,7	10	86,6	69,3	86,6	94,6	100	-0,4	-0,8
RFA (µmol m^{-2} s^{-1})	362,6	572,8	0	12,9	0	12,9	551,6	3300	3,5	1,9
PR (°C)	10,3	1,8	0,4	10,5	9,4	10,5	11,4	15,3	3,3	-1,3
PP (mm m^{-2} h^{-1})	0,03	0,6	0	0	0	0	0	60,4	6571	72

[1]Desviación estándar. [2]Valor mínimo. [3]1er cuartil. [4]2do cuartil. [5]3er cuartil. [6]Valor máximo. [7]Curtosis. [8]Asimetría. [9]Temperatura. [10]Humedad relativa. [11]Radiación fotosintéticamente activa. [12]Punto de rocío. [13]Precipitación.

De acuerdo con el coeficiente de curtosis para la evaluación anual (Tabla 5-1), la T y HR presentaron una baja concentración de valores alrededor de su media y, por lo tanto, su distribución puede asociarse con una curva platicúrtica (Curt < 0). Sin embargo, este coeficiente puede variar para una misma variable climática monitoreada en un lugar específico por diferentes sensores, como lo evidencian Trilles et al. (2021). Estos autores determinaron valores de curtosis entre -0,7 y 11,98 para la temperatura y entre -1,2 y -0,8 para la humedad relativa en la localidad de Castellón, España.

Por otro lado, RFA y PR presentaron una distribución leptocúrtica debido a la alta concentración de valores alrededor de la media (Curt > 0). El alto valor del coeficiente de curtosis para PP se debió a que el 75% de los valores corresponden a cero; es decir, los eventos de PP corresponden a menos del 25% para el período evaluado. El comportamiento de la curtosis para estas dos variables en los períodos evaluados (anual, temporada de heladas y temporada no asociada a heladas) fue similar. Sin embargo, algunos autores han informado diferencias en la distribución de datos al comparar varias estaciones (Brito et al., 2019) o diferentes intervalos de tiempo durante varios años (Moumen et al., 2021). El coeficiente de curtosis para cada variable climática fue diferente de cero, por lo que su distribución no fue normal. Adicionalmente, mediante el estadístico de prueba de Kolmogorov-Smirnov para cada variable climática por hora del día y su respectivo nivel de significancia, se encontró que ninguna variable presenta una distribución normal (valor $p < 0,05$) (Anexo 1).

El coeficiente de asimetría mostró una distribución de los valores de temperatura cercana a la simetría (As = 0,1) como se muestra en la Figura 5-1 (a), mientras que HR y PR presentaron un sesgo negativo; con alargamiento de valores inferiores a la media en la distribución (As < 0) (Figura 5-1 b y c); por el contrario, Brito et al. (2019) encontraron coeficientes de asimetría negativos para la temperatura y valores positivos para la humedad relativa. La RFA y PP presentaron una asimetría positiva (sesgo > 0) como se muestra en la Figura 5-1 (d). Adicionalmente, el comportamiento de la asimetría no varió para los tres periodos evaluados.

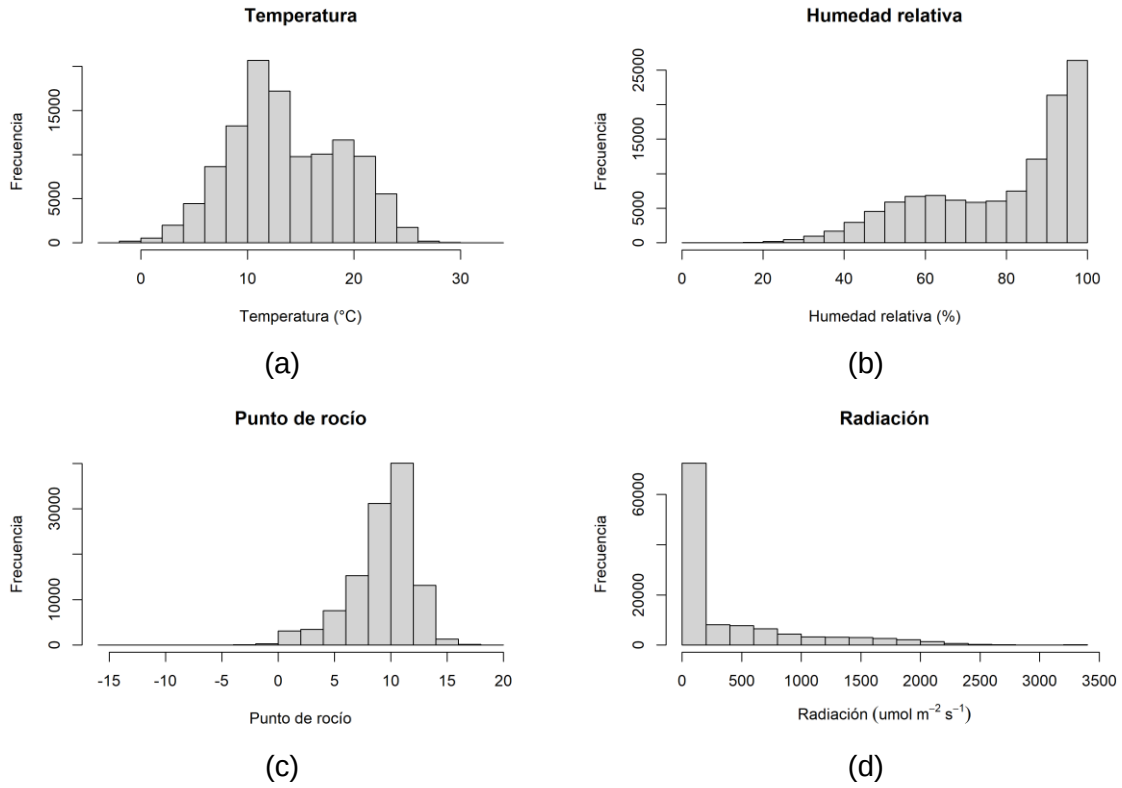


Figura 5-1. Histograma de frecuencia para la temperatura (a), humedad relativa (b), punto de rocío (c) y radiación fotosintéticamente activa (d) para la temporada de heladas. Nombre de la fuente: “Elaboración propia”.

La temperatura promedio en la época de heladas fue igual a la media anual (13,6 °C); sin embargo, la T mínima para los meses no asociados a heladas fue de 3,1 °C superior a la mínima registrada en época de heladas. Los valores promedio de HR, PR y PP fueron superiores en 3,7, 8,7 y 66,7 %, respectivamente, para los meses comprendidos entre marzo y octubre en comparación con la época de heladas (Tabla 5-1). La RFA fue la única variable climática cuyo valor promedio fue mayor en la temporada de heladas (367,7 µmol m^{-2} s^{-1}) en comparación con el valor promedio anual (364,9 µmol m^{-2} s^{-1}) y el valor promedio de la temporada sin heladas (362,6 µmol m^{-2} s^{-1}). Por otro lado, las estaciones que registraron el mayor número de eventos de heladas fueron PL (25,4%), UB (12,3%), FV (10,9%) y ER (10,9%), ubicadas en el noroeste del departamento de Cundinamarca.

Adicionalmente, el comportamiento promedio de las variables climáticas durante las 24 horas previas a un EH en comparación con un NEH evidenciaron diferencias, como se muestra en la Figura 4-3. Los días asociados a EH presentaron una temperatura media por hora que aumentó entre las 6:00 y las 14:00 de 1,7 a y 23,0 °C, mientras que a partir de las 14:00 hasta las 6:00 disminuyó. El comportamiento de la curva de temperatura del día NEH en comparación con un día asociado a heladas se mantuvo inferior, entre las 10:00 y las 16:00, y por encima, entre las 16:00 y las 10:00 (Figura 5-2a). Las curvas de HR presentaron un descenso convexo a partir de las 7:00 con un punto de inflexión mínimo (33,8 %, 53,1 % para EH y NEH) a las 123:00 h; a partir de este momento se inicia un crecimiento convexo con un punto máximo de inflexión a las 7:00 (96,8 %, 94,2 % para EH y NEH).

La curva de HR de un día NEH mostró un comportamiento similar a un día EH, entre las 00:00 y las 9:00 horas, y estuvo por debajo entre las 9:00 y las 12:00 horas (Figura 5-2b). Adicionalmente, el comportamiento del PR para los días no asociados a fenómenos de helada se mantuvo siempre inferior a la curva de las 24 horas del día previo a la ocurrencia de la helada, y ambas tienen etapas distintivas a lo largo del día. A las 6:00, el comportamiento promedio del PR en ambas curvas presento los valores más bajos (1,3 °C EH, 7,5 °C día), y tres horas después, a las 9:00 a.m., se evidenciaron los valores promedio por hora más altos (8,6 °C EH, 10,9 °C NEH), exhibiendo una alta tasa de cambio. A partir de estos valores máximos, ambas curvas decrecieron hasta las 6:00 a. m. del día siguiente; sin embargo, los valores promedio por hora en los días asociados a heladas (EH) presentaron la mayor tasa de descenso (Figura 5-2c).

En cuanto al comportamiento de la precipitación, los días EH fueron secos, mientras que los días NEH presentaron una mayor humedad, con eventos de lluvia entre las 11:00 y las 19:00 horas principalmente (Figura 5-2d). El comportamiento promedio de la radiación para los días EH y NEH fue similar; sin embargo, entre las 7:30 y las 17:00 horas, la radiación para los días asociados a heladas estuvo por encima de la curva NEH, con los valores máximos registrados a las 11:00 horas (1548,7 µmol m^{-2} s^{-1} EH días, y 1283,5 µmol m^{-2} s^{-1} NEH días) (Figura 5-2e). Las horas del día con mayor frecuencia de eventos de heladas fueron las 5:00 y 6:00 (Figura 5-3).

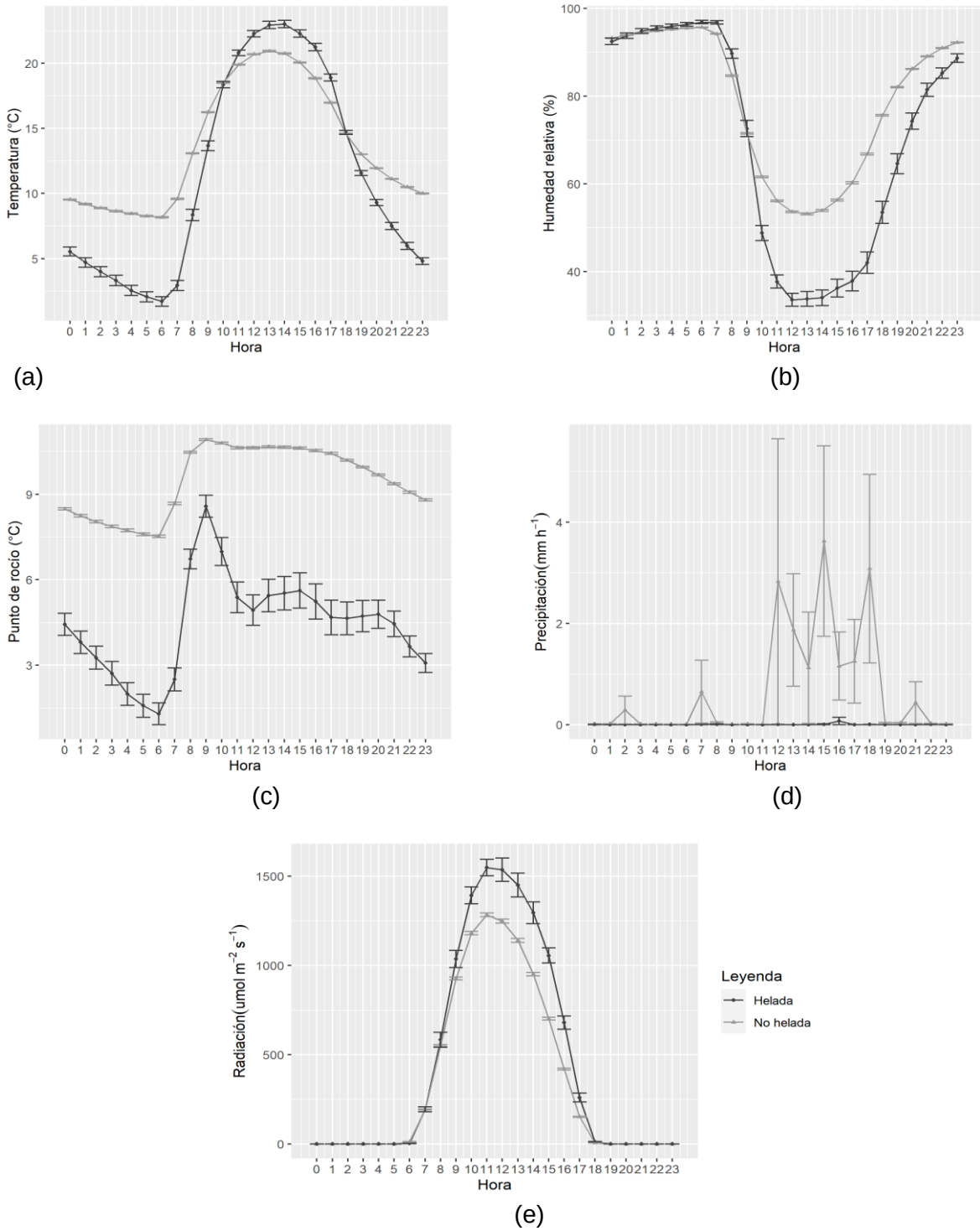


Figura 5-2. Comportamiento de las variables climáticas para las 24 horas del día anterior a eventos de helada (EH) o no asociadas con eventos de heladas (NEH): (a) temperatura, (b) humedad relativa, (c) punto de rocío, (d) precipitación y (e) radiación fotosintéticamente

activa. La longitud de las barras de error corresponde al error estándar de la media (SEM). Nombre de la fuente: “Elaboración propia”.

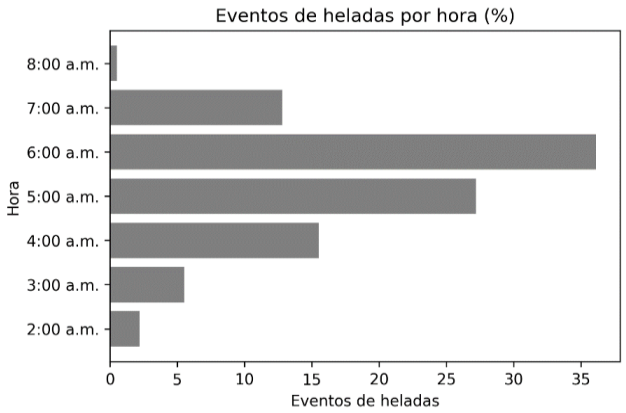


Figura 5-3. Frecuencia de eventos de helada por hora. Nombre de la fuente: “Elaboración propia”.

En resumen, el comportamiento de las variables climáticas de las horas previas a un EH se caracterizan por ser secas, con baja HR, marcadas por una reducción de la nubosidad entre las 11:00 y las 14:00 horas, por lo tanto, aumenta la radiación térmica emitida desde el suelo hacia el espacio (Arribillaga et al., 2020). En las condiciones anteriores, con vientos de baja velocidad (<7,2 km h^{-1}), no se compensa la pérdida de energía por radiación, provocando un descenso paulatino de la temperatura desde las primeras horas de la noche hasta las primeras horas de la madrugada, alcanzando valores por debajo de 0°C (González & Torres, 2012; Snyder & de Melo-abreu, 2010).

Otra de las variables que influye en la cantidad de energía emitida como calor de radiación desde la superficie terrestre es el vapor de agua contenido en la atmósfera, debido a que el agua puede absorber esta energía y devolverla a la superficie. Por lo tanto, una mayor cantidad de vapor de agua genera menos pérdida de calor. En esta etapa, el descenso de temperatura en una noche con poca nubosidad y condiciones de calma será más paulatino e incluso evitará alcanzar valores por debajo de 0°C (González & Torres, 2012; Snyder & de Melo-abreu, 2010).

Por otra parte, se evaluó el comportamiento promedio de las variables climáticas por temporada de helada (Figura 5-4). Cada temporada hace referencia al periodo entre noviembre y febrero del siguiente año para cada uno de los cinco años entre el 2017 y 2022. Las diferencias en los valores promedio para cada variable por temporada de helada pueden deberse a la ocurrencia de fenómenos globales como El Niño-Oscilación del Sur u otras condiciones a escala regional o local. Estas diferencias indican que la partición de la base de datos en los subconjuntos para el entrenamiento y la validación debería realizarse de manera aleatoria, con el fin de evitar sesgos asociados a cada temporada y considerar la varianza para todas las temporadas.

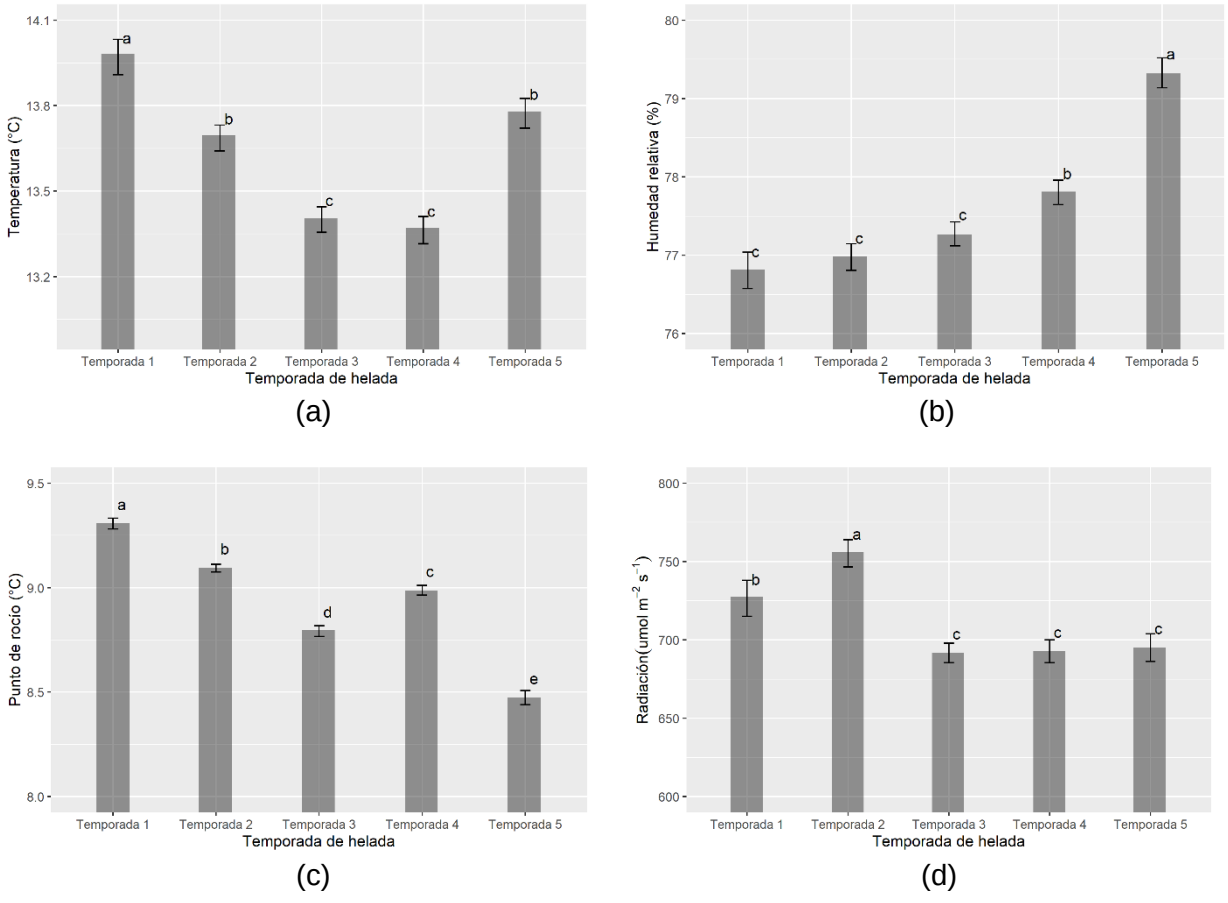


Figura 5-4. Comportamiento promedio de las variables climáticas por temporada. (a) temperatura, (b) humedad relativa, (c) punto de rocío y (d) radiación fotosintéticamente activa. La longitud de las barras de error corresponde al error estándar de la media (SEM). Nombre de la fuente: “Elaboración propia”.

5.2 Identificación de las variables climáticas de mayor relevancia para la predicción temprana de heladas en cultivos de altura.

El coeficiente de correlación de Spearman y su nivel de significancia (Anexo 2) evidenció que la temperatura presenta una relación directamente proporcional con los EH entre las 11:00 y 17:00 horas (valor $p < 0,05$) y de crecimiento inverso entre las 19:00 y 7:00 horas (valor $p < 0,05$), a partir de esta ultima hora, la relación comienza a presentar un debilitamiento hasta las 11:00 horas donde se vuelve positiva (Figura 5-5a). De igual forma, HR y EH presentaron una correlación de crecimiento positiva entre las 2:00 y las 7:00 (valor $p < 0,05$); a partir de este momento se produce un debilitamiento, para tornarse negativo entre las 10:00 a 01:00 del día siguiente (Figura 5-5b). Además, PR presentó una correlación negativa con los eventos de helada a lo largo del día, los valores más altos estuvieron entre -0,33 y -0,15 (Figura 5-5c) (Anexo 2). Se calculó la correlación mediante un método no paramétrico (coeficiente de correlación de Spearman) debido a que los datos no obedecen a una distribución normal.

La precipitación y la RFA tuvieron un impacto débil en la ocurrencia de eventos de heladas. La correlación entre PP y eventos de heladas fue negativa y presentó valores entre -0,002 y -0,026, a excepción de los valores obtenidos entre las 8:00 y las 12:00 que fueron positivos y cercanos a cero (0,004 a 0,005) (Figura 5-5d). La correlación entre RFA y la ocurrencia de eventos de heladas fue positiva entre las 7 a.m. y 6 p. m. (valor $p < 0,05$), con valores máximos entre las 15:00 y las 16:00 horas; y negativo de 19:00 a 06:00 (-0,013 a -0,002) como se muestra en la Figura 5-5e. Las variables climáticas con la menor correlación con la ocurrencia de eventos de helada (valor $p > 0,05$) fueron la humedad de las 0:00, 8:00 y 23:00 horas, la precipitación de las 10:00 y 11:00 horas, la radiación entre las 19:00 y 23:00 horas y la temperatura de las 10:00 y 18:00 horas (Anexo 2).

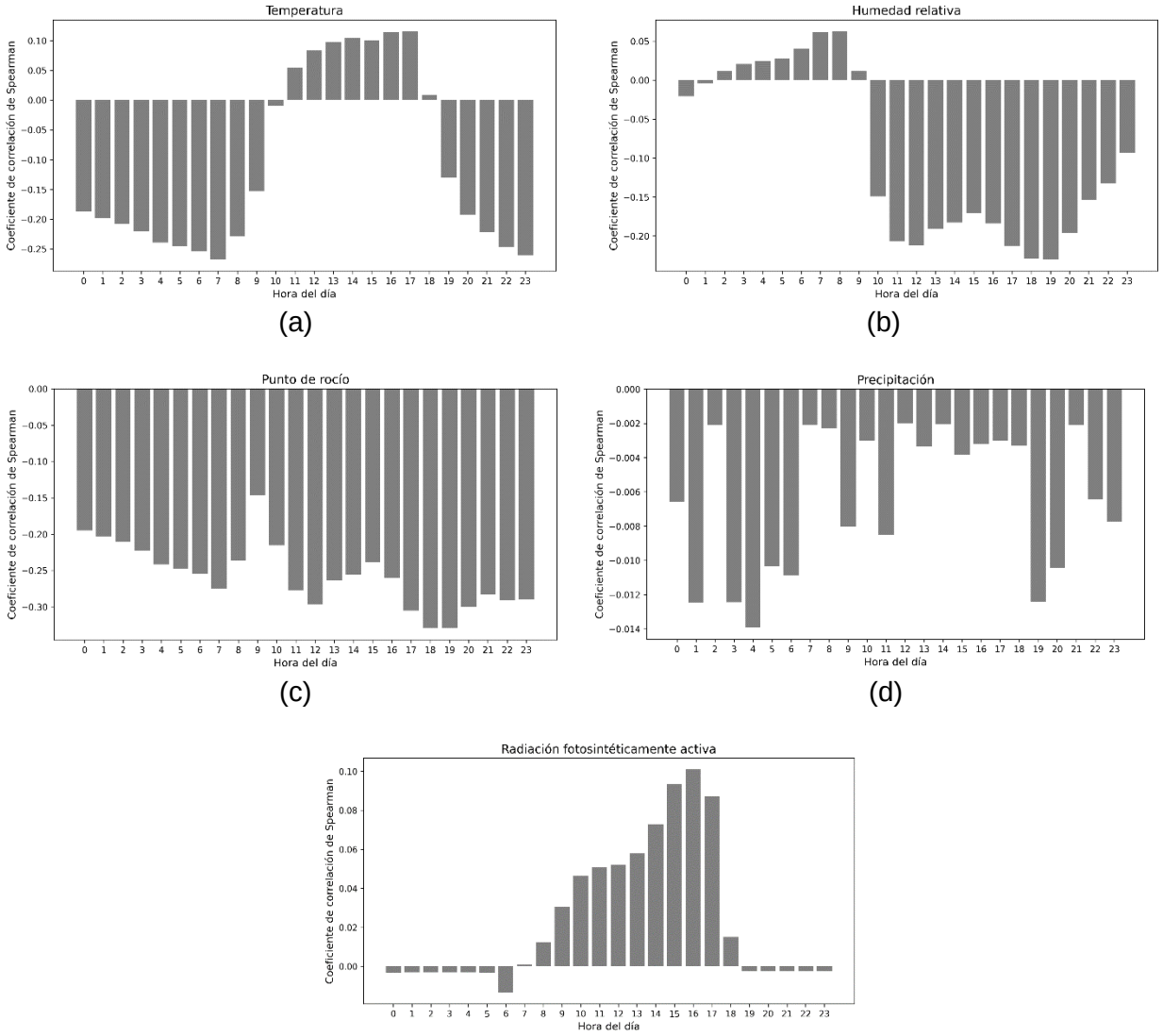


Figura 5-5. Correlación de las variables climáticas con la ocurrencia de un evento de helada por hora. (a) temperatura, (b) humedad relativa, (c) punto de rocío, (d) precipitación y (e) radiación fotosintéticamente activa. Nombre de la fuente: “Elaboración propia”.

Los resultados encontrados en el presente estudio difieren de los reportados por Lee et al. (2016), quienes establecieron una relación positiva entre PR y HR con la probabilidad de ocurrencia de heladas. Además, Ding et al. (2019) al evaluar un dataset con una frecuencia de 60 datos por hora encontraron que la temperatura del aire, la radiación neta y la velocidad del viento tenían una correlación negativa con las heladas (-0,37, -0,2 y -0,28, respectivamente) y una correlación positiva con la humedad relativa (0,35).

Otra de las variables climáticas que han sido utilizadas para la predicción de heladas es la velocidad del viento, debido a que cuando la velocidad supera un valor umbral específico, impide la formación de la capa de inversión térmica cerca del suelo (Castañeda & Castaño, 2017; Ding et al., 2019; Fuentes et al., 2018; Ghielmi & Eccel, 2006; Lee et al., 2016; Mercado et al., 2016; Ovando et al., 2005). Sin embargo, Ding et al. (2019) encontraron valores de correlación inferiores entre la velocidad del viento y la ocurrencia de heladas y Ovando et al. (2005) determinaron que los mejores pronósticos eran aquellos que no incluían variables relacionadas con el viento o la humedad relativa.

Por otro lado, a partir de la prueba de Kruskal-Wallis se encontraron diferencias estadísticamente significativas al comparar la humedad relativa de los días asociados a eventos de helada, con los días no relacionados con este fenómeno, para todas las horas del día a excepción de las 12 a.m. y las 9 a.m. (Anexo 3). Asimismo, se encontró que el punto de rocío presenta valores significativamente diferentes a lo largo del día, exceptuando los valores entre las 11 a.m. y las 4 p.m. La temperatura no fue significativamente diferente entre las 11 p.m. y la 1 a.m., tampoco se presentaron diferencias a las 10 a.m y a las 6 p.m. Por el contrario, la precipitación no presentó diferencias significativas al comparar los valores registrados por hora para ningún momento del día. La radiación presento diferencias significativas entre las 9 a.m. y las 6 p.m. (Anexo 3).

De manera similar, mediante el método de la mínima redundancia y máxima relevancia se encontró que las variables más relevantes fueron el punto de rocío entre la 11 a.m. y las 23:00 p.m., la temperatura entre las 4 a.m. y 11 a.m. y en la noche a partir de las 21:00 p.m. hasta las 23:00 p.m. Las mayores puntuaciones fueron para el punto de rocío a las 18:00 p.m., 19:00 p.m. y 12 m. y la temperatura de las 8 am y 11 a.m. Las horas del día en las que la función SelectKBest detectó la puntuación más alta fue para la temperatura entre las 02:00 y las 10:00, HR entre las 10:00 y las 16:00, PR entre la 01:00 y las 04:00 y RFA entre las 9:00 y 16:00.

Finalmente, las mayores puntuaciones obtenidas mediante el algoritmo de árboles aleatorios fueron para la humedad entre las 16:00 p.m. y 21:00 p.m., la temperatura entre las 21:00 p.m. y 23:00 p.m., seguidos del punto de rocío entre las 19:00 p.m. y las 23:00 p.m. con puntuaciones entre 0,02 y 0,06 (Anexo 4). La precipitación para todas sus horas

obtuvo los puntajes más bajos (<0,00014). A partir de la identificación de las variables más importante mediante cada metodología, se asignó a cada variable un valor de importancia (de cero a tres), donde se asignó el valor de cero a las variables que no fueron importantes mediante ninguna de las tres metodologías utilizada; por el contrario, el valor de tres indica que la variable fue seleccionada en todas las metodologías. Se seleccionaron las variables con importancia mayor a uno para el entrenamiento de los modelos (Figura 5-6) y la radiación entre las 9 a.m. y 18:00 p.m. Tal como se evidenció en el análisis de correlación, el punto de rocío es la variable que presenta una mayor incidencia en la ocurrencia de heladas a lo largo del día.

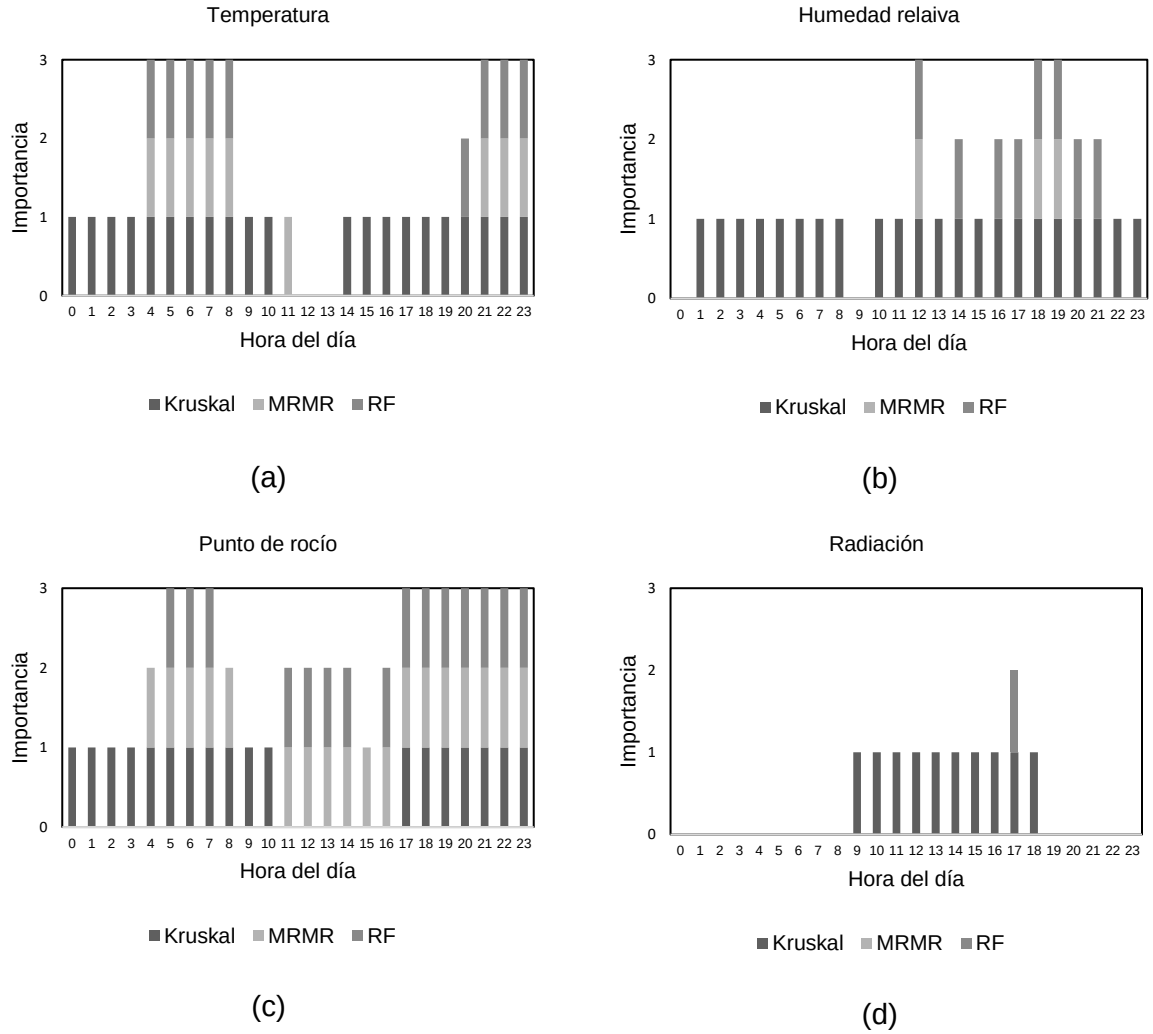


Figura 5-6. Importancia para cada variable climática por hora del día. (a) temperatura, (b) humedad relativa, (c) punto de rocío y (d) radiación fotosintéticamente activa a partir de la prueba de Kruskal-Wallis (Kruskal), el método de la mínima redundancia y máxima

relevancia (MRMR) y la selección de características mediante bosque aleatorios (RF). Nombre de la fuente: “Elaboración propia”.

En síntesis, la ocurrencia de eventos de helada está asociada con temperaturas en la noche de 9,3 ± 1,7 °C, 7,5 ± 2,0 °C, 5,9 ± 2,0 °C y 4,8 ± 1,9 °C para las 20:00, 21:00, 22:00 y 23:00 horas; humedad de 33,6 ± 11,0 % a las 12:00 horas y entre 53,5 y 64,6 ± 18,1 % para el intervalo de horas entre las 18:00 y 19:00. Adicionalmente, el punto de rocío con valores entre 1,5 y 2,5 ± 2,9 °C entre las 5:00 y las 7:00 horas e inferiores a 5,0 ± 3,6 °C entre las 17:00 y 23:00 horas indican la ocurrencia de una posible helada.

Tabla 5-2. Comportamiento de la temperatura, humedad relativa y punto de rocío por hora para el día previo a la ocurrencia de una helada.

Hora del día	Temperatura (°C)	Humedad relativa (%)	Punto de rocío (°C)
5:00			1,6
6:00			1,3
7:00			2,5
12:00		33,6	
17:00			4,7
18:00		53,5	4,6
19:00		64,6	4,7
20:00	9,3		4,8
21:00	7,5		4,5
22:00	5,9		3,6
23:00	4,8		3,1

5.3 Selección del método de aprendizaje de máquinas para la predicción temprana de heladas en cultivos de altura.

5.3.1 Métodos de predicción.

- **Predicción de heladas mediante métodos de clasificación.**

En el conjunto de datos de entrenamiento se incluyeron las variables más fuertemente relacionadas con EH hasta las 16:00 del día anterior, con el fin de tener una ventana de

tiempo disponible para implementar las estrategias de prevención y mitigación de heladas. Teniendo en cuenta que el conjunto de datos se encontraba altamente desbalanceado, se compararon tres estrategias para el manejo de datos desbalanceados y se encontraron las mejores métricas al implementar la técnica de sobremuestreo de minorías sintéticas (SMOTE) para los cinco algoritmos (Tabla 5-3). Además, el método de red neuronal desarrollado con la librería keras cuya función de activación fue ReLu, el optimizador Adam, la función de perdida de entropía cruzada binaria para la penalización de clases y entrenado con 20 épocas, también presento altas métricas de rendimiento (p = 0,72, TPR = 0,67, Ac = 0,98 y F_1 = 0,69).

Tabla 5-3. Métricas de evaluación del rendimiento de los métodos en los conjuntos de datos de prueba de acuerdo con la estrategia de manejo de datos desbalanceados empleada.

Estrategia de balanceo	Métrica	LR[1]	DT[2]	SVM[3]	GBDT[4]	NN[5]
Sobremuestreo SMOTE	p[6]	0,91	0,91	1,00	0,94	0,92
	TPR[7]	0,94	0,91	0,64	0,96	0,93
	TNR[8]	0,91	0,91	1,00	0,93	0,92
	Ac[9]	0,92	0,91	0,82	0,95	0,92
	F_1 score[10]	0,92	0,91	0,78	0,95	0,93
Subsampling	p	0,45	1,00	0,36	0,45	0,55
	TPR	0,99	0,20	0,10	0,10	0,30
	TNR	0,40	1,00	0,96	0,99	0,99
	Ac	0,71	0,41	0,25	0,6	0,76
	F_1 score	0,62	0,33	0,16	0,16	0,39
Penalización	p	0,45	0,27	0,18	NA	0,72
	TPR	0,14	0,07	1,00	NA	0,67
	TNR	0,99	0,99	0,99	NA	0,82
	Ac	0,95	0,93	0,99	NA	0,98
	F_1 score	0,21	0,11	0,31	NA	0,69
Sin balancear	p	0,27	0,18	0,18	0,18	0,27
	TPR	0,75	1,00	1,00	0,33	0,75
	TNR	0,99	0,99	0,99	0,99	0,99
	Ac	0,99	0,99	0,99	0,98	0,99

Estrategia de balanceo	Métrica	LR[1]	DT[2]	SVM[3]	GBDT[4]	NN[5]
	F_1 score	0,4	0,31	0,31	0,24	0,4

[1]Regresión logística. [2]Árbol de decisión. [3]Máquina de soporte vectorial. [4]Árboles de decisión de aumento del gradiente. [5]Red neuronal. [6]Precisión. [7]Tasa de verdaderos positivos. [8]Tasa de verdaderos negativos. [9]Exactitud. [10]Puntuación F_1.

Con la estrategia para el manejo de datos desbalanceados SMOTE, la mejor puntuación para el método LR (0,948) se obtuvo con C = 5,0 (inverso de la fuerza de regularización), regularización = L1 y solucionador liblineal. De igual forma, para DT (0,917), GBDT (0,995), SVM (0,993) y NN (0,927) se encontraron los hiperparámetros óptimos especificados en la Tabla 5-4. Además, en el entrenamiento de los cinco métodos se determinó mediante el análisis de componentes principales que 30 es el número de componentes principales que explican la mayor parte de variabilidad de los datos.

Tabla 5-4. Valor óptimo para cada uno de los hiperparámetros determinado mediante validación cruzada por método.

Método	Hiperparámetros	Valor óptimo	Tiempo de entrenamiento (segundos)[1]
LR[2]	C	5,0	120,2
	Regularización	L1	
	Solucionador (solver)	Liblinear	
DT[3]	Criterio	Gini	583,4
	Divisor (splitter)	Best	
	Profundidad máxima del árbol	10,0	
	Número mínimo de muestras	10,0	
	Máximo número de nodos de hoja	20,0	
GBDT[4]	Tasa de aprendizaje	0,075	4.206,5
	Número mínimo de muestras	20,0	
	Profundidad máxima	39,0	
	Número máximo de características	2,0	
SVM[5]	Regularización	20,0	1.816,6
	Tipo de Kernel	Rbf	

Método	Hiperparámetros	Valor óptimo	Tiempo de entrenamiento (segundos)[1]
NN[6]	Grado de la función del Kernel polinomial	0,5	24.300
	Gamma	Auto	
	Número de neuronas en la i-ésima capa oculta	2,0	
	Función de activación para la capa oculta	Identity	
	Solucionador (solver)	Lbfgs	
	Regularización	3,0	
	Tasa de aprendizaje	0,01	
	Criterio	2,0	

[1]Para el entrenamiento de los modelos se utilizó la máquina “Python 3 Google Compute Engine backend” (GPU). [2]Regresión logística. [3]Árbol de decisión. [4]Máquina de soporte vectorial. [5]Árboles de decisión de aumento del gradiente. [6]Red neuronal.

En algunos experimentos, las comparaciones del rendimiento predictivo entre los sistemas de predicción tradicionales y el ajuste de hiperparámetros muestran que la búsqueda de parámetros óptimos a través del ajuste de hiperparámetros proporciona mejores resultados y predicciones de mayor calidad (Ramasamy et al., 2021; Shahhosseini et al., 2022). Sin embargo, los parámetros óptimos dependen del conjunto de datos y los valores predeterminados arbitrarios del algoritmo respectivo, y a través de una búsqueda aleatoria, se puede generar un aumento en los costos computacionales (Schratz et al., 2019; Shahhosseini et al., 2022) y tiempos de ejecución, los cuales son mayores a medida que aumenta el número de hiperparámetros. Para este estudio se encontró que, entre los cinco métodos desarrollados, el mayor tiempo de ejecución fue para el algoritmo de NN (24300 segundos), seguido de GBDT (4206,5) y SVM (1816,6). Los tiempos de ejecución del entrenamiento en el caso del LR y DT fueron inferiores a 600 segundos.

- **Predicción de la temperatura mínima mediante métodos de regresión.**

Para la predicción de la temperatura mínima se desarrolló un método de bosques aleatorios y una red neuronal multicapa mediante el algoritmo de retropropagación, a partir de un conjunto de series temporales meteorológicas. Se encontraron las menores métricas

de error (MSE, MAE y RMSE) tanto en los datos de entrenamiento, validación y prueba para la red neuronal con 20 neuronas en la capa oculta, función de activación “ReLu”, optimizador “Adam”, taza de aprendizaje de 0,001, función de pérdida error cuadrático medio, 50 épocas y un tamaño del lote de 120 para la ventana de tiempo de una hora. El número de capas ocultas y la cantidad de neuronas en cada capa podría cambiar para cada configuración de la red y tipo de datos (Fuentes et al., 2018). Sin embargo, la adición de una capa oculta permite que la red aprenda patrones más complejos.

Asimismo, para el método de bosque aleatorios, se determinó las menores métricas de error con un número de 100 árboles para la ventana de tiempo de una hora. Para ambos métodos se evidenció que el error disminuye al reducir la ventana de tiempo de diez horas a una hora, en este sentido, el RMSE, MAE y MSE se reducen en 1,3 °C, 1,1 °C y 3,4 °C en promedio, respectivamente (Tabla 5-6). A nivel general, los resultados de los dos métodos son satisfactorios y se puede afirmar que mediante ambos algoritmos se puede predecir la temperatura mínima.

5.3.2 Comparación y evaluación de las métricas de rendimiento de los métodos.

- **Predicción de heladas mediante métodos de clasificación.**

Con la estrategia para el manejo de datos desbalanceados SMOTE, los cinco métodos entrenados presentaron un buen desempeño debido a sus métricas de evaluación de clasificación (exactitud, precisión, sensibilidad, especificidad y puntaje F1), superiores al 91 %, excepto la SVM, cuyas métricas fluctuaron entre 64 y 100 %. Todos los métodos probados fueron efectivos y eficientes en la detección de EH. La validación cruzada y el análisis estadístico demostraron la mayor precisión del método NN en la detección de EH con p = 0,95, TPR = 0,92, TNR = 0,95, Ac = 0,93 y puntuación F_1 = 0,94 en el conjunto de datos de validación (Tabla 5-5). Por el contrario, el método SVM presentó las métricas de evaluación de desempeño más bajas con p = 1,00, TPR = 1,00, TNR = 0,71, Ac = 0,80 y puntuación F_1 = 0,75. Diedrichs et al. (2018) también encontraron mejores rendimientos al evaluar la sensibilidad para los algoritmos de regresión logística y bosques aleatorios al implementar la estrategia SMOTE.

Tabla 5-5. Métricas de evaluación del rendimiento de los métodos de clasificación en el conjunto de datos de validación.

Métrica	LR[1]	DT[2]	SVM[3]	GBDT[4]	NN[5]
p[6]	0,95	0,93	1,00	0,95	0,95
TPR[7]	0,92	0,94	1,00	0,92	0,92
TNR[8]	0,95	0,93	0,71	0,94	0,95
Ac[9]	0,93	0,93	0,80	0,93	0,93
F_1 score[10]	0,93	0,93	0,75	0,93	0,94

[1]Regresión logística. [2]Árbol de decisión. [3]Máquina de soporte vectorial. [4]Árboles de decisión de aumento del gradiente. [5]Red neuronal. [6]Precisión. [7]Tasa de verdaderos positivos. [8]Tasa de verdaderos negativos. [9]Exactitud. [10]Puntuación F_1.

El método LR también presentó métricas de alto rendimiento en el conjunto de datos de validación (Ac = 0,93, p = 0,95, TPR = 0,92, TNR = 0,95 y puntuación F1 = 0,93) y un tiempo de ejecución más corto en comparación con GBDT y NN; lo que indica que los métodos GBDT, NN y LR se pueden utilizar para predecir EH con un error de clasificación bajo. Por otra parte, la SVM presentó un rendimiento más bajo en comparación con los métodos desarrollados por Ding et al. (2019). Estos autores encontraron métricas de evaluación de clasificación de 0,92 (Ac), 0,9 (p) y 0,99 (TPR) para el pronóstico de heladas con una ventana de tiempo de tres horas, con el algoritmo de máquina de vector de soporte y la temperatura, humedad relativa y radiación como variables de entrada, monitoreadas con una frecuencia de un registro por minuto.

Para la predicción de la ocurrencia de heladas también se han implementado otros algoritmos como el K-vecinos más cercanos (Cadenas et al., 2020), el cual fue entrenado con siete variables climáticas (humedad relativa, radiación solar, velocidad del viento, precipitación, punto de rocío, déficit de presión de vapor y temperatura) monitoreadas con una frecuencia de un registro cada cinco minutos, para la predicción de este fenómeno con una ventana de tiempo de una hora. El método desarrollado por estos autores presentó una exactitud y una tasa de verdaderos negativos (Ac = 0,96 y TNR = 0,98) superior a los encontrados con los diferentes métodos en la presente investigación; sin embargo, la tasa de verdaderos positivos y la puntuación F_1 fue inferior (TPR = 0,88 y F_1 = 0,88).

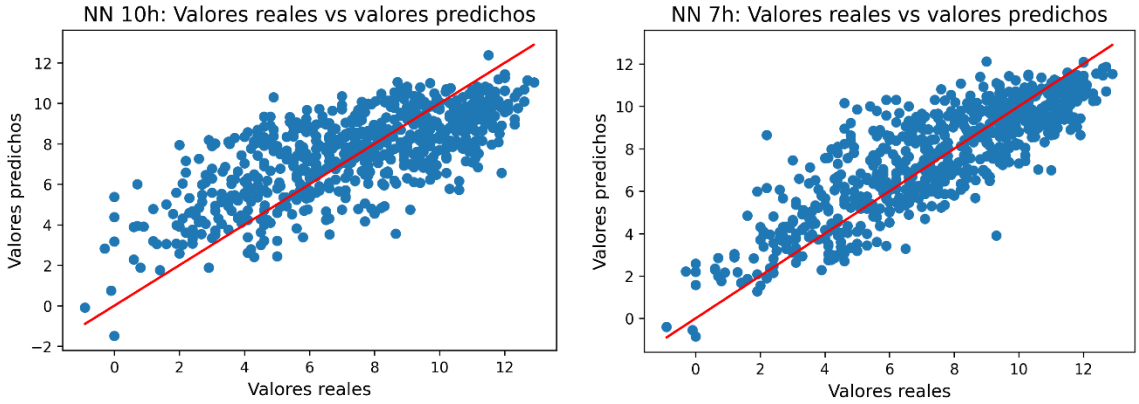
- **Predicción de la temperatura mínima mediante métodos de regresión.**

Al analizar las métricas de error para los métodos de predicción de la temperatura mínima, se puede concluir que los resultados son satisfactorios. Los menores errores se obtuvieron para la predicción con una hora de antelación a la ocurrencia de las heladas (inferiores a 0,71 °C) y fueron similares para el método basado en bosques aleatorios RF (RMSE = 0,71, MAE = 0,53, MSE = 0,51) y la red neuronal NN (RMSE = 0,71, MAE = 0,54, MSE = 0,49). Por el contrario, los errores más altos (superiores a 1,5 °C) se obtuvieron con la mayor ventana de tiempo (diez horas) y fueron RMSE = 2,04, MAE = 1,65 y MSE = 4,17 para el método de bosque aleatorios y RMSE = 1,91, MAE = 1,55 y MSE = 3,65 para la red neuronal (Tabla 5-6). Asimismo, al comparar la distribución de los valores predichos con los valores reales se puede observar que, al reducir la ventana de tiempo, los valores presentaron una menor dispersión y se acercaron a la línea roja que indica que los valores predichos son iguales a los reales u observados (Figura 5-7).

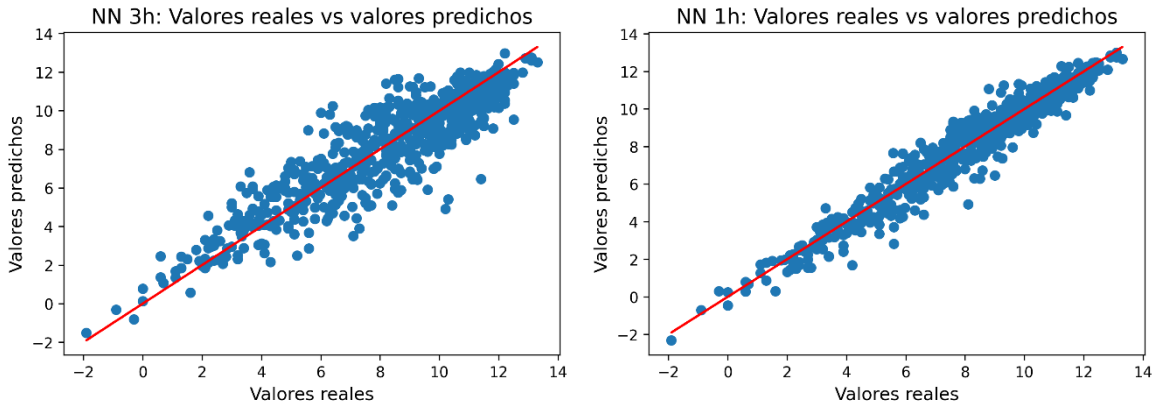
Tabla 5-6. Métricas de evaluación del rendimiento de los métodos de regresión en el conjunto de datos de validación.

Método	Ventana de tiempo (horas)	RMSE[1]	MAE[2]	MSE[3]
NN[4]	10	1,91	1,55	3,65
	7	1,49	1,18	2,24
	3	1,27	0,99	1,60
	1	0,71	0,54	0,49
RF[5]	10	2,04	1,65	4,17
	7	1,59	1,28	2,55
	3	1,28	0,99	1,63
	1	0,71	0,53	0,51

[1]Raíz del error cuadrático medio. [2]Error medio absoluto. [3]Error cuadrático medio. [4]Red neuronal. [5]Bosques aleatorios.



(a) (b)



(c) (d)

Figura 5-7. Comparación de los valores reales vs los valores predichos por las redes neuronales (NN) para cuatro ventanas de tiempo diferentes: (a) 10 horas, (b) 7 horas, (c) 3 horas y (d) 1 hora. Nombre de la fuente: “Elaboración propia”.

Teniendo en cuenta que, desde el punto de vista del problema de las heladas, un error del método al no predecir un evento de helada y que efectivamente se presente una reducción de la temperatura inferior a 0 °C es un error muy grave, ya que el agricultor puede perder toda su cosecha. Por el contrario, un error en el método indicando que se presenta una helada y en realidad no se presenta, es menos grave debido a que no se pierde la cosecha; sin embargo, se pueden perder los recursos en la implementación de estrategias de mitigación de heladas, los cuales son inferiores a la pérdida del material vegetal (Guillén et al., 2020).

Para la identificación de este tipo de errores, se detectaron los falsos positivos y falsos negativos con los valores de temperatura mínima predichos por la red neuronal, donde T mínima ≤ 0 °C indico un EH y valores superiores a cero se clasificaron como NEH. Con esta clasificación, se calcularon las métricas descritas en la Tabla 4-5. Se encontraron las mejores métricas para el conjunto de datos de prueba con una ventana de tiempo de una hora (TNR = 1,00, TPR = 0,80, p = 1,00, F_1 = 0,89 y Ac = 1,00). Por el contrario, en la ventana de tiempo de diez horas se identificó un mayor número de falsos positivos (TNR = 0,99, TPR = 1,00, p = 0,44, F_1 = 0,61 y Ac = 0,99). Sin embargo, en ambos periodos de tiempo se predijeron correctamente la totalidad de los eventos de helada como se muestra en las tablas de contingencia (Tabla 5-7).

Tabla 5-7. Tabla de contingencia para comparación entre los valores pronosticados y observados para una ventana de tiempo de una hora (izquierda) y de diez horas (derecha).

Observación	Predicción	
	PP[3]	PN[4]
P[1]	4	0
N[2]	2	674

Observación	Predicción	
	PP	PN
P	4	0
N	5	671

[1]Eventos de heladas reales. [2]Eventos no asociados con heladas reales. [3]Eventos de heladas predichos. [4]Eventos no asociados con heladas predichos.

Estas métricas fueron similares a las reportadas por Fuentes et al. (2018) & Guillén et al. (2020), quienes obtuvieron un Ac promedio de 0,98. Sin embargo, deben realizarse pruebas adicionales para evaluar el desempeño de los métodos con una mayor cantidad de eventos de helada.

5.4 Validación del desempeño del método seleccionado para la predicción temprana de heladas en cultivos de altura.

- **Predicción de heladas mediante métodos de clasificación.**

Los resultados de las métricas de evaluación del desempeño de los algoritmos generados, con el conjunto de datos durante el proceso de prueba, presentaron valores similares a los obtenidos con el conjunto de datos usados en el proceso de validación. Por lo tanto, se podría afirmar que no se presentó un sobreajuste. De manera similar a los resultados expuestos en la sección anterior, con el conjunto de datos de prueba, los métodos LR, DT, GBDT y NN, presentaron una alta precisión de predicción, con valores superiores al 0.92 en todas las métricas consideradas. El método GBDT fue particularmente más preciso en la detección de EH con TPR = 0,96 y Ac = 0,95. El GBDT también fue superior en otras métricas como p, TNR y puntuación F1 tanto en el conjunto de validación como en el conjunto de datos de prueba (Tabla 5-8). Más específicamente, GBDT presento p = 0,94, TNR = 0,93 y puntuación F_1 = 0,95, mientras que la SVM presentó métricas de evaluación de desempeño más bajas con Ac = 0,82, p = 1,00, TPR = 0,64, TNR = 1,00 y puntuación F1 = 0,78, tendencia similar a la obtenida en el proceso de validación. De manera similar, Lee et al. (2016) evaluaron dos algoritmos de clasificación para la predicción de heladas

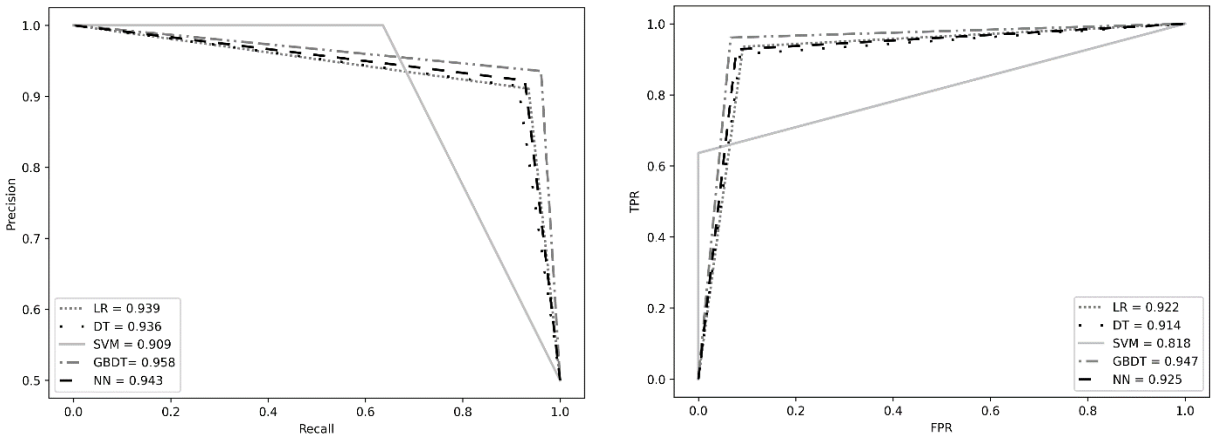
(regresión logística y arboles de decisión) y encontraron métricas de desempeño superiores al 0,90 para la exactitud y valores entre 0,75 y 0,87 para la sensibilidad.

Tabla 5-8. Métricas de evaluación del rendimiento de los métodos en el conjunto de datos de prueba.

Métrica	LR[1]	DT[2]	SVM[3]	GBDT[4]	NN[5]
p[6]	0,91	0,91	1,00	0,94	0,92
TPR[7]	0,94	0,91	0,64	0,96	0,93
TNR[8]	0,91	0,91	1,00	0,93	0,92
Ac[9]	0,92	0,91	0,82	0,95	0,92
F_1 score[10]	0,92	0,91	0,78	0,95	0,93

[1]Regresión logística. [2]Árbol de decisión. [3]Máquina de soporte vectorial. [4]Árboles de decisión de aumento del gradiente. [5]Red neuronal. [6]Precisión. [7]Tasa de verdaderos positivos. [8]Tasa de verdaderos negativos. [9]Exactitud. [10]Puntuación F_1.

La curva ROC evalúa el equilibrio entre la tasa de verdaderos positivos y falsos positivos; las curvas cercanas a la esquina superior izquierda indican el mejor desempeño (TPR = 1 y FPR = 0) (Chipindu et al., 2020). Por lo tanto, todos los métodos de clasificación binaria entrenados presentaron un alto rendimiento. Adicionalmente, como se muestra en la Figura 5-8a, el valor del área bajo la curva AUC del algoritmo GBDT (0,96) fue mayor que los de NN (0,94), LR (0,94), DT (0,94) y SVM (0,91). De manera similar, el valor del área bajo la curva ROC para el método GBDT fue mayor (0,95), seguido de NN (0,93), LR (0,92), DT (0,91) y SVM (0,82), como se muestra en la Figura 5-8b.



(a) AUC-PR (b) ROC

Figura 5-8. Evaluación del rendimiento mediante el área bajo la curva de precisión-sensibilidad o AUC (a) y proporción de verdaderos positivos (TPR) versus la proporción de falsos positivos (FPR) ROC (b), para los algoritmos de regresión logística (LR), árbol de decisión (DT), máquina de soporte vectorial (SVM), árboles de decisión de aumento del gradiente GBDT y red neuronal (NN) con conjunto de prueba. Nombre de la fuente: "Elaboración propia".

Los valores de TPR o sensibilidad constituyen la métrica más apropiada para la evaluación del desempeño de los métodos, ya que el objetivo es predecir la ocurrencia de eventos de heladas y así generar una alerta oportuna para implementar mecanismos de prevención de daños por heladas. En este sentido, es más grave no predecir una helada que generar una falsa alarma. Por lo tanto, se utilizó el método GBDT como una de las principales herramientas para generar alarmas de prevención de daños por eventos de heladas en los cultivos de altura en el país.

Este algoritmo de clasificación binaria se implementó en la web y a partir de la información de la temperatura, humedad relativa, punto de rocío y radiación registrada en las trece estaciones, los usuarios diariamente a partir de las 16:00 horas pueden consultar la probabilidad de ocurrencia de un evento de helada en la madrugada del día siguiente (Figura 5-9). Sin embargo, actualmente la visualización de esta información está restringida a los cultivos de flores que hacen parte del grupo Growers Hub Trading (GHT).

Figura 5-9. Visualización de la probabilidad de ocurrencia de un evento de helada para las trece estaciones ubicadas en nueve municipios de Cundinamarca. Nombre de la fuente: "Elaboración propia".

- **Predicción de la temperatura mínima mediante métodos de regresión.**

Similar a los resultados encontrados con el conjunto de datos de validación, al analizar las métricas de error para los métodos de predicción de la temperatura mínima, con el conjunto de datos de pruebas, se obtuvieron los menores errores para la predicción con una hora de antelación a la ocurrencia de las heladas (inferiores a 0,75 °C) y fueron similares para la red neuronal NN (RMSE = 0,74, MAE = 0,55 y MSE = 0,54) y el método de bosques aleatorios RF (RMSE = 0,71, MAE = 0,54 y MSE = 0,50). Por el contrario, los errores más altos (entre 1,45 y 3,65 °C) se obtuvieron con la mayor ventana de tiempo y fueron de un RMSE = 1,81, MAE = 1,45 y MSE = 3,27 para la red neuronal y RMSE = 1,91, MAE = 1,55 y MSE = 3,65 para el método de bosques aleatorios (Tabla 5-9).

Tabla 5-9. Métricas de evaluación del rendimiento de los métodos de regresión en el conjunto de datos de prueba.

Métodos	Ventana de tiempo (horas)	RMSE[1]	MAE[2]	MSE[3]
NN[4]	10	1,81	1,45	3,27
	7	1,45	1,17	2,10
	3	1,29	1,00	1,67
	1	0,74	0,55	0,54
RF[5]	10	1,91	1,54	3,65

7	1,59	1,27	2,52
3	1,29	1,03	1,67
1	0,71	0,54	0,50

[1]Raíz del error cuadrático medio. [2]Error medio absoluto. [3]Error cuadrático medio. [4]Red neuronal. [5]Bosques aleatorios.

Las métricas obtenidas fueron similares a las reportadas por Smith et al. (2009), estos autores encontraron un MAE de 0,52 °C y 1,87 °C para la predicción de la temperatura mínima con una antelación de una y doce horas, respectivamente, al desarrollar algoritmos de redes neuronales. Asimismo, Cadenas et al. (2020) determinaron la temperatura mínima a partir de información climática y el desarrollo de un algoritmo de k-vecinos más cercanos y encontraron un MSE de 0,49 °C y un MAE de 0,51 °C para una ventana de tiempo de una hora.

Por otra parte, Yu et al. (2016) desarrollaron un algoritmo para la predicción de la temperatura mínima al interior del invernadero basado en máquinas de soporte vectorial de mínimos cuadrados a partir de variables climáticas y obtuvieron un MSE de 0,008 °C para una ventana de tiempo de una hora y de 2,35 °C para seis horas. Finalmente, el algoritmo desarrollado por Guillén et al. (2020) para la predicción de la temperatura mínima a partir de un modelo de memoria a largo plazo (LSTM) presento un RMSE promedio de 0,99 °C y un MAE de 0,57 °C. Estos autores realizaron la predicción de la temperatura de las siguientes 24 horas a partir de información climática obtenida bajo una infraestructura IoT en dos parcelas en Murcia (España), con una frecuencia de un registro cada diez minutos**.**

Como trabajos futuros relacionados con la predicción de la temperatura se encuentra ampliar la cantidad de variables climáticas involucradas en el proceso de entrenamiento y un período de tiempo mayor. Adicionalmente, algunos autores han determinado que se puede obtener una mejora en los pronósticos al entrenar un modelo por estación, debido a que cada área asociada a las estaciones tiene diferentes características geográficas (orografía y morfología del terreno) y condiciones climáticas locales, que pueden ser más importantes que una condición generalizada para el área total en formación de heladas como elevación, inclinación, bosques o colinas cercanas (Fuentes et al., 2018; Ghielmi & Eccel, 2006).

6. Conclusiones y trabajo futuro.

6.1 Conclusiones.

La predicción de heladas es un problema latente para los agricultores, debido a que el cambio climático está provocando cambios bruscos de temperatura incluso en zonas cálidas. Por lo tanto, es imperante conocer con una ventana de tiempo suficiente la ocurrencia de eventos de heladas para gestionar, prever los recursos y desplegar las estrategias de mitigación. En este trabajo se desarrolló un modelo para la detección temprana de heladas a partir de la implementación de cuatro métodos de clasificación de aprendizaje de máquinas y regresión logística basados en variables climáticas. Los resultados se compararon con diferentes métricas de evaluación del rendimiento. Este experimento se llevó a cabo utilizando datos reales para caracterizar y comprender el problema.

Los resultados muestran que la tasa de verdaderos positivos, la exactitud y la precisión de GBDT en el conjunto de datos de validación y prueba fueron superiores entre los algoritmos evaluados (> 92 %). Sin embargo, todos los métodos evaluados con sus hiperparámetros óptimos tuvieron un alto rendimiento, a excepción de SVM.

Adicionalmente, se desarrollaron dos modelos para la detección de heladas a partir de la predicción de la temperatura mínima en cuatro ventanas de tiempo diferentes, mediante los cuales se detectaron el 100% de los eventos de heladas en el conjunto de datos de prueba. Sin embargo, al aumentar la ventana de tiempo también se aumenta la cantidad de falsos positivos y aumentan las métricas de error.

El punto de rocío fue la variable climática con mayor relación con la ocurrencia de eventos de heladas. Además, se identificó que la ocurrencia de este fenómeno está asociada con

temperaturas inferiores a 9,5 °C entre las 20:00 y 23:00 horas, humedad relativa inferior a 64,6 % entre las 18:00 y las 19:00 horas y punto de rocío menor a 5.0 °C entre las 17:00 y 23:00 horas.

A partir de los resultados encontrados, es posible generar una predicción temprana de heladas, como apoyo, en la anticipación de la implementación de estrategias de mitigación y adaptación, frente a los daños que ocasiona este factor en los cultivos de altura colombianos.

El método GBDT, dada su mejor eficiencia en todas las métricas consideradas, se implementó en la web como suporte para los productores de flores, en la estimación de la probabilidad de ocurrencia de una helada, a partir la temperatura, humedad relativa, punto de rocío y radiación.

6.2 Trabajo futuro.

Aunque los métodos presentaron métricas de alto rendimiento cuando se evaluaron con los conjuntos de datos de validación y prueba, es recomendable monitorearlos en tiempo real ya que el sobreajuste es una de las amenazas para los algoritmos de aprendizaje de máquinas.

Desde las perspectivas de este trabajo, planeamos aplicar nuestros hallazgos para prevenir la ocurrencia de heladas en cultivos de altura ubicados en otras regiones tropicales y obtener predicciones de la duración de los eventos de heladas para la implementación óptima de estrategias de mitigación. También se espera incluir otras variables climáticas y obtener predicciones en ventanas de tiempo más amplias. Además, se recomienda desarrollar un modelo por estación ya que habría una alta posibilidad de mejorar los pronósticos, teniendo en cuenta las condiciones particulares de cada lugar.

7. Referencias.

Aguilar, M., & Torres, S. B. (2010). Protocolo de uso y aprovechamiento de la uva de anís en matorrales andinos del Altiplano Cundiboyacense. In A. E. y G. Editores (Ed.), Instituto de Investigación de Recursos Biológicos Alexander von Humboldt. http://repository.humboldt.org.co/handle/20.500.11761/31447

Alapaty, K., Herwehe, J. A., Otte, T. L., Nolte, C. G., Bullock, O. R., Mallard, M. S., Kain, J. S., & Dudhia, J. (2012). Introducing subgrid-scale cloud feedbacks to radiation for regional meteorological and climate modeling. Geophysical Research Letters, 39(24), 1–5. https://doi.org/10.1029/2012GL054031

Almansour, N. A., Syed, H. F., Khayat, N. R., Altheeb, R. K., Juri, R. E., Alhiyafi, J., Alrashed, S., & Olatunji, S. O. (2019). Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study. Computers in Biology and Medicine, 109, 101–111. https://doi.org/10.1016/j.compbiomed.2019.04.017

Arribillaga, D., Bravo, R., Campos, C., Fuentes, M., Gatica, J., Luchabeche, P., Quintana, J., Reyes, M., Salacar, C., Salvo del Pedregal, J., & Vidal, M. (2020). Heladas. Factores, tendencias y efectos en frutales y vides. In R. Bravo, J. Quintana, & M. Reyes (Eds.), Instituto de Investigaciones Agropecuarias: Vol. Boletín IN (N° 417). https://biblioteca.inia.cl/bitstream/handle/123456789/6847/Boletín INIA N° 417?sequence=1&isAllowed=y

Becerra, L. L. (2021). San Valentín, el desquite de los floricultores en pandemia. Portafolio. https://www.portafolio.co/economia/san-valentin-el-desquite-de-los-floricultores-en-pandemia-con-las-exportaciones-de-flores-548989

Bonilla, J. E., Ramirez, J., & Ramirez, O. (2006). Metodología para el diseño de un modelo univariado de Red Neuronal Para El Pronóstico De La Temperatura Mínima En La Zona De Mosquera (Cundinamarca, Colombia). Meteorología Colombiana., 10, 111–120.

Brito, A., Araújo, H. A., & Zebende, G. F. (2019). Detrended Multiple Cross-Correlation Coefficient applied to solar radiation, air temperature and relative humidity. Scientific Reports, 9(1), 1–10. https://doi.org/10.1038/s41598-019-56114-6

Bugata, P., & Drotar, P. (2020). On some aspects of minimum redundancy maximum relevance feature selection. Science China Information Sciences, 63(1), 1–15. https://doi.org/10.1007/s11432-019-2633-y

Cadenas, J. M., Garrido, M. C., Martínez, R., & Guillén, M. A. (2020). Making decisions for frost prediction in agricultural crops in a soft computing framework. Computers and

Electronics in Agriculture, 175(May), 105587. https://doi.org/10.1016/j.compag.2020.105587

Castañeda, A., & Castaño, V. M. (2017). Smart frost control in greenhouses by neural networks models. Computers and Electronics in Agriculture, 137, 102–114. https://doi.org/10.1016/j.compag.2017.03.024

Castañeda, A., & Castaño, V. M. (2020). Internet of things for smart farming and frost intelligent control in greenhouses. Computers and Electronics in Agriculture, 176(June), 105614. https://doi.org/10.1016/j.compag.2020.105614

Castillo, F. E., & Castellvi, F. (2001). Agrometeorología (Mundi-Prensa (ed.); 2a ed.). https://docplayer.es/78435550-Agrometeorologia-2a-edicion-corregida-coordinadores-francisco-elias-castillo-instituto-nacional-de-investigaciones-agrarias.html

Chang, D. C., Sohn, H. B., Cho, J. H., Im, J. S., Jin, Y. I., Do, G. R., Kim, S. J., Cho, H. M., & Lee, Y. B. (2014). Freezing and Frost Damage of Potato Plants: a Case Study on Growth Recovery, Yield Response, and Quality Changes. Potato Research, 57(2), 99–110. https://doi.org/10.1007/s11540-014-9253-5

Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. Journal of Applied Science and Technology Trends, 2(01), 20–28. https://doi.org/10.38094/jastt20165

Chavarro, D., Vélez, M., Montenegro, I., Hernández, A., & Olaya, A. (2018). Objetivos de Desarrollo Sostenible en Colombia y el aporte de la ciencia, la tecnologia y la innovación. “Patrimonio”: Economía Cultural Y Educación Para La Paz (Mec-Edupaz), 2(14), 100–117.

Chipindu, L., Mupangwa, W., Mtsilizah, J., Nyagumbo, I., & Zaman-Allah, M. (2020). Maize Kernel Abortion Recognition and Classification Using Binary Classification Machine Learning Algorithms and Deep Convolutional Neural Networks. Ai, 1(3), 361–375. https://doi.org/10.3390/ai1030024

Cho, S., Kim, Y. J., Lee, M., Woo, J. H., & Lee, H. J. (2021). Correction to: Cut-off points between pain intensities of the postoperative pain using receiver operating characteristic (ROC) curves (BMC Anesthesiology, (2021), 21, 1, (29). https://doi.org/10.1186/s12871-021-01410-w

Christensen, J. H., Krishna Kumar, K., Aldrian, E., An, S.-I., Cavalcanti, I. F. A., de Castro, M., Dong, W., Goswami, P., Hall, A., Kanyanga, J. K., Kitoh, A., Kossin, J., Lau, N.-C., Renwick, J., Stephenson, D. B., Xie, S.-P., & Zhou, T. (2013). Climate Phenomena and their Relevance for Future Regional Climate Change. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge Univ. Press, 1217–1308.

http://www.climatechange2013.org/images/report/WG1AR5_FOD_Ch14_All_Final.pdf

Clarkson, D. T., Earnshaw, M. J., White, P. J., & Cooper, H. D. (1988). Temperature dependent factors influencing nutrient uptake: an analysis of responses at different levels of organization. In Symposia of the Society for Experimental Biology, 42, 281–309.

Colinet, H., Lee, S. F., & Hoffmann, A. (2010). Functional characterization of the Frost gene in Drosophila melanogaster: Importance for recovery from chill coma. PLoS ONE, 5(6), 1–7. https://doi.org/10.1371/journal.pone.0010925

Danandeh Mehr, A. (2021). Drought classification using gradient boosting decision tree. Acta Geophysica, 69(3), 909–918. https://doi.org/10.1007/s11600-021-00584-8

del Angel, J. A., & Sarmiento, A. (2011). Utilización De La Escala Beaufort En La Determinación Del Potencial Eólico. Revista Científica de Ingeniería Energética, 25(1), 13–17. https://rie.cujae.edu.cu/index.php/RIE/article/download/171/169

Departamento Administrativo Nacional de Estadística DANE. (2021). Boletín técnico Exportaciones (EXPO). DANE. https://www.dane.gov.co/index.php/estadisticas-por-tema/comercio-internacional/exportaciones/exportaciones-historicos

DeVries, Z., Locke, E., Hoda, M., Moravek, D., Phan, K., Stratton, A., Kingwell, S., Wai, E. K., & Phan, P. (2021). Using a national surgical database to predict complications following posterior lumbar surgery and comparing the area under the curve and F1-score for the assessment of prognostic capability. Spine Journal, 21(7), 1135–1142. https://doi.org/10.1016/j.spinee.2021.02.007

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

Dimitriadis, S. I., Liparas, D., & Tsolaki, M. N. (2018). Random forest feature selection, fusion and ensemble strategy: Combining multiple morphological MRI measures to discriminate among healhy elderly, MCI, cMCI and alzheimer's disease patients: From the alzheimer's disease neuroimaging initiative (ADNI) data. Journal of Neuroscience Methods, 302, 14–23. https://doi.org/10.1016/j.jneumeth.2017.12.010

Ding, L., Noborio, K., & Shibuya, K. (2019). Frost forecast using machine learning - From association to causality. Procedia Computer Science, 159, 1001–1010. https://doi.org/10.1016/j.procs.2019.09.267

Ding, L., Noborio, K., & Shibuya, K. (2020). Modelling and learning cause-effect - application in frost forecast. Procedia Computer Science, 176, 2264–2273. https://doi.org/10.1016/j.procs.2020.09.285

Ding, L., Tamura, Y., Yoshida, S., Owada, K., Toyoda, T., Morishita, Y., Noborio, K., & Shibuya, K. (2021). Ensemble causal modelling for frost forecast in vineyard. Procedia Computer Science, 192, 3194–3203. https://doi.org/10.1016/j.procs.2021.09.092

Dinh, T. V., Nguyen, H., Tran, X. L., & Hoang, N. D. (2021). Predicting Rainfall-Induced Soil Erosion Based on a Hybridization of Adaptive Differential Evolution and Support Vector Machine Classification. Mathematical Problems in Engineering, 2021. https://doi.org/10.1155/2021/6647829

Dujovne, D., Watteyne, T., Mercado, G., & Diedrichs, A Taffernaberry, J C Perez Peña, J. E. (2020). Wireless Wine: Estimación de rendimiento y ubicación de sensores para la predicción de heladas en los viñedos. Universidad Nacional de La Patagonia Austral.

Eccel, E., Ghielmi, L., Granitto, P., Barbiero, R., Grazzini, F., & Cesari, D. (2007). Prediction of minimum temperatures in an alpine region by linear and non-linear post-processing of meteorological models. Nonlinear Processes in Geophysics, 14(3), 211–222. https://doi.org/10.5194/npg-14-211-2007

El Espectador. (2021). Pérdidas de más de 10.000 hectáreas de cultivos por época de helada y sequía en Cundinamarca. https://www.elespectador.com/bogota/perdidas-mas-de-10000-hectareas-de-cultivos-por-epoca-de-heladas-y-s¿quia-en-cundinamarca-article/

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 CK-12. (2021). Conceptos de Ciencias de la Tierra: La circulación en la atmósfera. California. https://flexbooks.ck12.org/cbook/ck-12-conceptos-de-ciencias-de-la-tierra-grados-6-8-en-espanol/section/7.14/primary/lesson/la-circulación-en-la-atmósfera/

Galiba, G., Vágújfalvi, A., Li, C., Soltész, A., & Dubcovsky, J. (2009). Regulatory genes involved in the determination of frost tolerance in temperate cereals. Plant Science, 176(1), 12–19. https://doi.org/10.1016/j.plantsci.2008.09.016

Garreaud, R. D. (2009). The Andes climate and weather. Advances in Geosciences, 22, 3–11. https://doi.org/10.5194/adgeo-22-3-2009

Ghielmi, L., & Eccel, E. (2006). Descriptive models and artificial neural networks for spring frost prediction in an agricultural mountain area. Computers and Electronics in Agriculture, 54(2), 101–114. https://doi.org/10.1016/j.compag.2006.09.001

Gómez, D. A. (2014). Caracterización, pronóstico y alternativas de manejo de las heladas en el sistema de producción lechero del Valle de Ubaté y Chiquinquirá (Colombia). Universidad Nacional de Colombia.

Gómez, D., Araujo, G., Martínez, F. E., Rodríguez, A. O., Estupiñan, J. M., & Deantonio, L. Y. (2021). Análisis de eventos climáticos extremos asociados a excesos de lluvia y heladas meteorológicas en el Altiplano Cundiboyacense de Colombia. Revista de Climatología, 21, 112–126. https://rclimatol.eu/2021/09/18/analisis-de-eventos-climaticos-extremos-asociados-a-excesos-de-lluvia-y-heladas-meteorologicas-en-el-altiplano-cundiboyacense-de-colombia/

González, O. C., & Torres, C. F. (2012). Actualización nota técnica heladas. Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM). http://www.ideam.gov.co/documents/21021/21147/Documento+FINAL+actualizacion+nota+tecnica+heladas.pdf/e10a0183-62e6-410a-8e96-7e0739f6f06b

Gu, L., Hanson, P. J., Post, W. Mac, Kaiser, D. P., Yang, B., Nemani, R., Pallardy, S. G., & Meyers, T. (2008). The 2007 eastern US spring freeze: Increased cold damage in a warming world? BioScience, 58(3), 253–262. https://doi.org/10.1641/B580311

Guhl, E. (2013). La región hídrica de bogotá - Capítulo Marco conceptual. Revista de La Academia Colombiana de Ciencias Exactas, Físicas y Naturales, 37(144), 327–341.

Guillen, M. A., Cadenas, J. M., Garrido, M. C., Ayuso, B., & Martinez, R. (2018). A Preliminary Study to Solve Crop Frost Prediction Using an Intelligent Data Analysis Process. Intelligent Environments 2018, 23, 97–106. https://doi.org/10.3233/978-1-61499-874-7-97

Guillén, M. A., Martínes, R., Bueno, A., Ayuso, B., Moren, J. L., & Cecilia, J. M. (2019). An LSTM Deep Learning Scheme for Prediction of Low Temperatures in Agriculture. 130–138. https://doi.org/10.3233/AISE190032

Guillén, M. A., Martínez, R., Llanes, A., Bueno, A., & Cecilia, J. M. (2020). A deep learning model to predict lower temperatures in agriculture. Journal of Ambient Intelligence and Smart Environments, 12(1), 21–34. https://doi.org/10.3233/AIS-200546

Hashempour, A., Ghasemnezhad, M., Ghazvini, R. F., & Sohani, M. M. (2014). The Physiological and Biochemical Responses to Freezing Stress of Olive Plants Treated with Salicylic Acid 1. Russian Journal of Plant Physiology, 61(4), 443–450. https://doi.org/10.1134/S1021443714040098

Hashi, E. K., & Md. Shahid Uz Zaman. (2020). Developing a Hyperparameter Tuning Based Machine Learning Approach of Heart Disease Prediction. Journal of Applied Science & Process Engineering, 7(2), 631–647. https://doi.org/10.33736/jaspe.2639.2020

Hecke, T. Van. (2012). Power study of anova versus Kruskal-Wallis test. Journal of Statistics and Management Systems, 15(2–3), 241–247. https://doi.org/10.1080/09720510.2012.10701623

Ho, Y., & Wookey, S. (2020). The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling. IEEE Access, 8, 4806–4813. https://doi.org/10.1109/ACCESS.2019.2962617

Hu, Y. G., Asante, E. A., Lu, Y. Z., Mahmood, A., Buttar, N. A., & Yuan, S. Q. (2018). Review of air disturbance technology for plant frost protection. International Journal of Agricultural and Biological Engineering, 11(3), 21–28. https://doi.org/10.25165/j.ijabe.20181103.3172

Hu, Y. G., Zhao, C., Liu, P. F., Asante, E. A., & Li, P. P. (2016). Sprinkler irrigation system for tea frost protection and the application effect. International Journal of Agricultural and Biological Engineering, 9(5), 17–23. https://doi.org/10.3965/j.ijabe.20160905.1315

Hurtado, G. (1996). Estadísticas de la Helada Meteorológica en Colombia (IDEAM (ed.); METEO/007-).

Instituto Colombiano Agropecuario. (2021). El ICA, soporte para la exportación de flores y ornamentales al mundo para San Valentín. ICA. https://www.ica.gov.co/noticias/ica-san-valentin-flores-colombia-llegan-100-paises

International Trade Center. (2020). List of importers for the selected product: 0603 Flowers and buds, cut for bouquets or decorations. https://www.trademap.org/Country_SelProduct_TS.aspx?nvpm=1%7C%7C%7C%7C%7C0603%7C%7C%7C4%7C1%7C1%7C1%7C2%7C1%7C2%7C1%7C1%7C1

Jain, A., Mcclendon, R. W., & Hoogenboom, G. (2006). Freeze prediction for specific locations using artificial neural networks. 49(6), 1955–1962.

Joshi, N. C., Yadav, D., Ratner, K., Kamara, I., Aviv-Sharon, E., Irihimovitch, V., & Charuvi, D. (2020). Sodium hydrosulfide priming improves the response of photosynthesis to overnight frost and day high light in avocado (Persea americana Mill, cv. ‘Hass’). Physiologia Plantarum, 168(2), 394–405. https://doi.org/10.1111/ppl.13023

Juna, A., Umer, M., Sadiq, S., Karamti, H., Eshmawi, A. A., Mohamed, A., & Ashraf, I. (2022). Water Quality Prediction Using KNN Imputer and Multilayer Perceptron. Water, 14, 2592, 1–19. https://doi.org/10.3390/w14172592

Juurakko, C. L., diCenzo, G. C., & Walker, V. K. (2021). Cold acclimation and prospects for cold-resilient crops. Plant Stress, 2(August), 100028. https://doi.org/10.1016/j.stress.2021.100028

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report, Ver. 2.3 EBSE Technical Report. EBSE.

Kochhar, S. L., & Gujral, S. K. (2020). Plant Physiology. In Cambridge University Press (2nd ed., Issue April).

Kukal, M. S., & Irmak, S. (2018). U. S. Agro-Climate in 20th Century: Growing Degree Days , First and Last Frost , Growing Season Length , and Impacts on Crop Yields. Scientific Reports, January, 1–14. https://doi.org/10.1038/s41598-018-25212-2

Latif, R. M. A., Belhaouari, S. B., Saeed, S., Imran, L. B., Sadiq, M., & Farha, M. (2020). Integration of Google Play Content and Frost Prediction Using CNN: Scalable IoT Framework for Big Data. IEEE Access, 8, 6890–6900. https://doi.org/10.1109/ACCESS.2019.2963590

Lee, H., Chun, J. A., Han, H. H., & Kim, S. (2016). Prediction of Frost Occurrences Using Statistical Modeling Approaches. Advances in Meteorology, 2016. https://doi.org/10.1155/2016/2075186

Lhomme, J.-P., & Vacher, J.-J. (2003). La Mitigación de heladas en los camellones del altiplano andino. Bulletin de l'Institut Français d'études Andines, 32(32 (2)), 377–399. https://doi.org/10.4000/bifea.6556

Li, X., Ahammed, J. G., Li, Z., Zhang, L., Wei, J., Yan, P., Zhang, L.-P., & Han, W.-Y. (2018). Scientia Horticulturae Freezing stress deteriorates tea quality of new fl ush by inducing photosynthetic inhibition and oxidative stress in mature leaves. Scientia Horticulturae, 230(December 2017), 155–160. https://doi.org/10.1016/j.scienta.2017.12.001

Lindkvist, L., Gustavsson, T., & Bogren, J. (2000). A frost assessment method for mountainous areas. Agricultural and Forest Meteorology, 102(1), 51–67. https://doi.org/10.1016/S0168-1923(99)00087-8

Linnenluecke, M. K., Marrone, M., & Singh, A. K. (2020). Conducting systematic literature reviews and bibliometric analyses. Australian Journal of Management, 45(2), 175–194. https://doi.org/10.1177/0312896219877678

Liu, J., & Sherif, S. M. (2019). Combating Spring Frost With Ethylene. Frontiers in Plant Science, 10(October), 1–6. https://doi.org/10.3389/fpls.2019.01408

Luengas, E., Guhl, A., Castro, J. C., González, L. N., & Restrepo, S. (2021). Modeling the correlation between potato disease spread and climate variables to guide fungicide applications in Cundinamarca, Colombia. Naturaleza y Sociedad. Desafíos Medioambientales, 1, 7–42. https://revistas.uniandes.edu.co/doi/full/10.53010

Lukatkin, A., Brazaityte, A., Bobinas, C., & Duchovskis, P. (2012). Chilling injury in chilling-sensitive plants: a review. Agriculture, 99(2), 111–124. https://doi.org/https://doi.org/10.1016/j.postharvbio.2005.04.012

Majeed, M., Bhatti, K. H., & Amjad, M. S. (2021). Impact of climatic variations on the flowering phenology of plant species in Jhelum district, Punjab, Pakistan. Applied Ecology and Environmental Research, 19(August), 3343–3376. https://doi.org/10.15666/aeer/1905

Maqsood, I., Khan, M. R., & Abraham, A. (2004). An ensemble of neural networks for weather forecasting. Neural Computing and Applications, 13(2), 112–122. https://doi.org/10.1007/s00521-004-0413-4

Maraveas, C., & Bartzanas, T. (2021). Application of Internet of Things (IoT) for Optimized Greenhouse Environments. AgriEngineering, 3(4), 954–970. https://doi.org/10.3390/agriengineering3040060

Marmolejo, D., & Ruiz, J. E. (2018). Tolerance of native potatoes (Solanum spp.) to ice creams in the context of climate change. Scientia Agropecuaria, 9(3), 393–400. https://doi.org/10.17268/sci.agropecu.2018.03.10

Mayorga, M., Fischer, G., Melgarejo, L. M., & Parra-Coronado, A. (2020). Growth, development and quality of Passiflora tripartita var. Mollissima fruits under two environmental tropical conditions. Journal of Applied Botany and Food Quality, 93, 66–75. https://doi.org/10.5073/JABFQ.2020.093.009

Mcelwee, P., Castro, P., Marisa, A., Walter, A., & Filho, L. (2019). Climate Change-Resilient Agriculture and Agroforestry. Springer International Publishing. https://doi.org/10.1007/978-3-319-75004-0

Mercado, F. R., García Fernández, W., & Acebey, J. A. H. (2016). Sistema de inteligencia artificial para la predicción temprana de heladas meteorológicas Artificial intelligence system for early prediction of weather frost. Acta Nova, 7 (December), 1683–0768.

Meza, C. C., & Gutierréz, S. A. (2020). Evaluación de modelos de clasificación para la predicción de heladas en el sector agricultor de Mosquera Cundinamarca, Colombia [Universidad de La Salle]. https://ciencia.lasalle.edu.co/cgi/viewcontent.cgi?article=1778&context=ing_automatizacion

Ministerio de Agricultura y Desarrollo Rural. (2019). Evaluaciones Agropecuarias Municipales EVA. Red de Información y Comunicación Del Sector Agropecuario

Colombiano [Agronet]. https://www.datos.gov.co/Agricultura-y-Desarrollo-Rural/Evaluaciones-Agropecuarias-Municipales-EVA/2pnw-mmge

Ministerio de Agricultura y Medio Ambiente. (2020). “Debemos mantener la guardia con medidas preventivas frente a bajas temperaturas y fenómeno de La Niña”: ministro Rodolfo Zea. https://www.minagricultura.gov.co/noticias/Paginas/“Debemos-mantener-la-guardia-con-medidas-preventivas-frente-a-bajas-temperaturas-y-fenómeno-de-La-Niña”-ministro-Rodolfo-Ze.aspx

Moumen, Z., Elhassnaoui, I., & Daid, F. (2021). Statistical descriptive analysis of three climate variables; Precipitation, temperature and relative humidity . Study cases (Innaouene watershed; Morocco). 03004, 1–7.

Nagasuga, K., Murai-Hatano, M., & Kuwagata, T. (2011). Effects of low root temperature on dry matter production and root water uptake in rice plants. Plant Production Science, 14(1), 22–29. https://doi.org/10.1626/pps.14.22

Olszewski, F., Jeranyama, P., Kennedy, C. D., & DeMoranville, C. J. (2017). Automated cycled sprinkler irrigation for spring frost protection of cranberries. Agricultural Water Management, 189, 19–26. https://doi.org/10.1016/j.agwat.2017.04.014

Ovando, G., Bocco, M., & Sayago, S. (2005). Redes Neuronales Para Modelar Predicción De Heladas. Agricultura Técnica, 65(1). https://doi.org/10.4067/s0365-28072005000100007

Ozgur, A. (2004). Supervised and unsupervised machine learning techniques for text document categorization [Bogazi¸ci University]. In CWL Publishing Enterprises, Inc., Madison (Vol. 2004). http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract

Patel, H., Singh Rajput, D., Thippa Reddy, G., Iwendi, C., Kashif Bashir, A., & Jo, O. (2020). A review on classification of imbalanced data for wireless sensor networks. International Journal of Distributed Sensor Networks, 16(4). https://doi.org/10.1177/1550147720916404

Pearce, R. S. (2001). Plant Freezing and Damage. Annals of Botany, 87(4), 417–424. https://doi.org/10.1006/anbo.2000.1352

Pedregosa, F., Varoquaux, G., Gramfort, Alexandre Michel, Vincent Thirion, Bertrand Grisel, Olivier Blondel, Mathieu Prettenhofer, Peter Weiss, R., & Dubourg, Vincent Vanderplas, Jake Passos, Alexandre Cournapeau, David Brucher, Matthieu Perrot, Matthieu Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research JMLR 12. https://scikit-learn.org/stable/modules/neural_networks_supervised.html#neural-networks-supervised

Pino, M. & Chen, T. (2016). Efectos de las heladas en el cultivo de papa, y desafíos del mejoramiento genético. Boletín Inia. Vol. Nº 331. Pág. 130-141.

Prabha, T., & Hoogenboom, G. (2008). Evaluation of the Weather Research and Forecasting model for two frost events. Computers and Electronics in Agriculture, 64(2), 234–247. https://doi.org/10.1016/j.compag.2008.05.019

Procolombia. (2019). ¿Cómo funciona el sector floricultor en Colombia? https://www.colombiatrade.com.co/noticias/como-funciona-el-sector-floricultor-en-colombia

Ramasamy, L. K., Kadry, S., & Lim, S. (2021). Selection of optimal hyper-parameter values of support vector machine for sentiment analysis tasks using nature-inspired optimization methods. Bulletin of Electrical Engineering and Informatics, 10(1), 290–298. https://doi.org/10.11591/eei.v10i1.2098

Ribeiro, A. C., De Melo-Abreu, J. P., & Snyder, R. L. (2006). Apple orchard frost protection with wind machine operation. Agricultural and Forest Meteorology, 141(2–4), 71–81. https://doi.org/10.1016/j.agrformet.2006.08.019

Rodríguez, K. (2020). Así enfrentan las heladas los municipios más afectados de Cundinamarca. El Espectador. https://www.elespectador.com/bogota/asi-enfrentan-las-heladas-los-municipios-mas-afectados-de-cundinamarca-article-899460/

Rout, B. M. (2020). Advances in Freez- ing Stress Resis- tance in Vegetable Crops. Biotica Research Today, 2(5 Spl.), 261–263. https://bioticainternational.com/ojs/index.php/biorestoday/article/view/150

Sallis, P., Jarur, M., & Trujillo, M. (2008). Frost Prediction Characteristics and Classification Frost Prediction Characteristics and Classification. November. https://doi.org/10.1007/978-3-642-02490-0

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., & Brenning, A. (2019). Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. Ecological Modelling, 406(June), 109–120. https://doi.org/10.1016/j.ecolmodel.2019.06.002

Shabala, S. (2017). Plant stress physiology. In CABI (2ND ed., Vol. 3). https://books.google.com.co/books

Shahhosseini, M., Hu, G., & Pham, H. (2022). Optimizing ensemble weights and hyperparameters of machine learning models for regression problems. Machine Learning with Applications, 7(December 2021), 100251. https://doi.org/10.1016/j.mlwa.2022.100251

Shamsnia, S. A., Shahidi, N., Liaghat, A., Sarraf, A., & Vahdat, S. F. (2011). Modeling of weather parameters using stochastic methods (ARIMA model)(case study: Abadeh

Region, Iran). International Conference on Environment and Industrial Innovation. IPCBEE, 12, 282–285. http://www.ipcbee.com/vol12/55-C30028.pdf

Simnitt, S., Borisova, T., Chavez, D., & Olmstead, M. (2017). Frost protection for Georgia peach varieties: Current practices and information needs. HortTechnology, 27(3), 344–353. https://doi.org/10.21273/HORTTECH03590-16

Smith, B. A., Hoogenboom, G., & McClendon, R. W. (2009). Artificial neural networks for automated year-round temperature prediction. Computers and Electronics in Agriculture, 68(1), 52–61. https://doi.org/10.1016/j.compag.2009.04.003

Smith, B. A., Mcclendon, R. W., & Hoogenboom, G. (2006). Improving Air Temperature Prediction with Artificial Neural Networks. International Journal of Computer and Information Engineering, 1(10), 3159. http://waset.org/publications/10353/improving-air-temperature-prediction-with-artificial-neural-networks

Snyder, R. L. (2000). Principles of Frost Protection. University of California, 1((Long version – Quick Answer FP005)). https://d1wqtxts1xzle7.cloudfront.net

Snyder, R. L., & de Melo-abreu, J. P. (2010). Protección contra las heladas: fundamentos, práctica y economía. In Organización de las Naciones Unidas para la Agricultura y la Alimentación (Ed.), Organización de las Naciones Unidas para la Agricultura y la Alimentación FAO (Volumen 1). https://www.fao.org/3/y7223s/y7223s.pdf

Stephens, G. L. (2005). Cloud feedbacks in the climate system: A critical review. Journal of Climate, 18(2), 237–273. https://doi.org/10.1175/JCLI-3243.1

Striegler, K., Allen, A., Bergmeier, E., & Caple, H. (2007). Understanding and Preventing Freeze Damage in Vineyards. Institute for Continental Climate Viticulture and Enology, University of Missouri-Columbia, 108. https://site.extension.uga.edu/viticulture/files/2018/03/Missouri-Freeze-Conference-Proceedings.pdf#page=39

Superintendencia de Sociedades. (2017). Desempeño del sector floricultor en Bogotá, Colombia. https://www.supersociedades.gov.co/SiteCollectionDocuments/2017/EEEstudio sector Flores-2017 09 28.pdf

Taiz, L., & Zeiger, E. (2002). Plant physiology (S. Associantes (ed.); Third).

Trilles, S., Juan, P., Chaudhuri, S., & Fortea, A. B. V. (2021). Data on CO2, temperature and air humidity records in Spanish classrooms during the reopening of schools in the COVID-19 pandemic. Data in Brief, 39, 107489. https://doi.org/10.1016/j.dib.2021.107489

Verdes, P. F., Granitto, P. M., Navone, H. D., & Ceccatto, H. A. (2000). Frost prediction with machine learning techniques. In I. de F. R. (CONICET-UNR) (Ed.), VI Congreso Argentino de Ciencias de la Computación. (p. 11). http://sedici.unlp.edu.ar/bitstream/handle/10915/23444/SI-026.pdf?sequence=1

Wen, X., Lu, S., & Jin, J. (2012). Integrating remote sensing data with WRF for improved simulations of oasis effects on local weather processes over an Arid Region in Northwestern China. Journal of Hydrometeorology, 13(2), 573–587. https://doi.org/10.1175/JHM-D-10-05001.1

Wolfe, D. W. (1991). Low temperature effects on early vegetative growth, leaf gas exchange and water potential of chilling-sensitive and chilling-tolerant crop species. Annals of Botany, 67(3), 205–212. https://doi.org/10.1093/oxfordjournals.aob.a088124

Xier, L. (2009). Analysis Of Monthly Temperature of Stockholm. Level Essay in Statistics.

Yin, G., Sun, H., Xin, X., Qin, G., Liang, Z., & Jing, X. (2009). Mitochondrial damage in the soybean seed axis during imbibition at chilling temperatures. Plant and Cell Physiology, 50(7), 1305–1318. https://doi.org/10.1093/pcp/pcp074

Yu, H., Chen, Y., Hassan, S. G., & Li, D. (2016). Prediction of the temperature in a Chinese solar greenhouse based on LSSVM optimized by improved PSO. Computers and Electronics in Agriculture, 122, 94–102. https://doi.org/10.1016/j.compag.2016.01.019

Zaharia, M., Chen, A., Davidson, A., Ghodsi, A., Hong, S. A., Konwinski, A., Murching, S., Nykodym, T., Ogilvie, P., Parkhe, M., Xie, F., & Zumar, C. (2018). Accelerating the Machine Learning Lifecycle with MLflow. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 39–45.

Zhang, J. P., & Mani, I. (2003). KNN Approach to Unbalanced Data Distributions: A Case Study Involving Information Extraction. Proceeding of International Conference on Machine Learning (ICML 2003), Washington DC, Workshop o.

Zhou, I., Lipman, J., Abolhasan, M., & Shariati, N. (2022). Minute-wise frost prediction: An approach of recurrent neural networks. Array, 14(October 2021), 100158. https://doi.org/10.1016/j.array.2022.100158

Zhu, C., Idemudia, C. U., & Feng, W. (2019). Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques. Informatics in Medicine Unlocked, 17(March), 100179. https://doi.org/10.1016/j.imu.2019.100179

8. ANEXOS

Anexo 1: Evaluación de la distribución de las variables climáticas por hora mediante la prueba de normalidade de Kolmogorov-Smirnov.

Variable	Prueba de Kolmogorov-Smirnov	Nivel de significancia (Valor p)	Variable	Prueba de Kolmogorov-Smirnov	Nivel de significancia (Valor p)
TMin	0.936	0.000	PuntoRocio12	0.975	0.000
Humedad0	1.000	0.000	PuntoRocio13	0.973	0.000
Humedad1	1.000	0.000	PuntoRocio14	0.972	0.000
Humedad2	1.000	0.000	PuntoRocio15	0.968	0.000
Humedad3	1.000	0.000	PuntoRocio16	0.968	0.000
Humedad4	1.000	0.000	PuntoRocio17	0.973	0.000
Humedad5	1.000	0.000	PuntoRocio18	0.976	0.000
Humedad6	1.000	0.000	PuntoRocio19	0.979	0.000
Humedad7	1.000	0.000	PuntoRocio20	0.980	0.000
Humedad8	1.000	0.000	PuntoRocio21	0.979	0.000
Humedad9	1.000	0.000	PuntoRocio22	0.977	0.000
Humedad10	1.000	0.000	PuntoRocio23	0.973	0.000
Humedad11	1.000	0.000	Radiacion0	0.500	0.000
Humedad12	1.000	0.000	Radiacion1	0.500	0.000
Humedad13	1.000	0.000	Radiacion2	0.500	0.000
Humedad14	1.000	0.000	Radiacion3	0.500	0.000
Humedad15	1.000	0.000	Radiacion4	0.500	0.000
Humedad16	1.000	0.000	Radiacion5	0.500	0.000
Humedad17	1.000	0.000	Radiacion6	0.508	0.000
Humedad18	1.000	0.000	Radiacion7	0.987	0.000
Humedad19	1.000	0.000	Radiacion8	0.990	0.000
Humedad20	1.000	0.000	Radiacion9	0.992	0.000
Humedad21	1.000	0.000	Radiacion10	0.993	0.000
Humedad22	1.000	0.000	Radiacion11	0.990	0.000
Humedad23	1.000	0.000	Radiacion12	0.988	0.000
Precipitacion0	0.500	0.000	Radiacion13	0.986	0.000
Precipitacion1	0.500	0.000	Radiacion14	0.985	0.000
Precipitacion2	0.500	0.000	Radiacion15	0.989	0.000
Precipitacion3	0.500	0.000	Radiacion16	0.988	0.000
Precipitacion4	0.500	0.000	Radiacion17	0.984	0.000
Precipitacion5	0.500	0.000	Radiacion18	0.500	0.000
Precipitacion6	0.500	0.000	Radiacion19	0.500	0.000
Precipitacion7	0.500	0.000	Radiacion20	0.500	0.000
Precipitacion8	0.500	0.000	Radiacion21	0.500	0.000

Precipitacion9	0.500	0.000	Radiacion22	0.500	0.000
Precipitacion10	0.500	0.000	Radiacion23	0.500	0.000
Precipitacion11	0.500	0.000	Temperatura0	0.993	0.000
Precipitacion12	0.500	0.000	Temperatura1	0.988	0.000
Precipitacion13	0.500	0.000	Temperatura2	0.980	0.000
Precipitacion14	0.500	0.000	Temperatura3	0.971	0.000
Precipitacion15	0.500	0.000	Temperatura4	0.958	0.000
Precipitacion16	0.500	0.000	Temperatura5	0.948	0.000
Precipitacion17	0.500	0.000	Temperatura6	0.944	0.000
Precipitacion18	0.500	0.000	Temperatura7	0.973	0.000
Precipitacion19	0.500	0.000	Temperatura8	0.999	0.000
Precipitacion20	0.500	0.000	Temperatura9	1.000	0.000
Precipitacion21	0.500	0.000	Temperatura10	1.000	0.000
Precipitacion22	0.500	0.000	Temperatura11	1.000	0.000
Precipitacion23	0.500	0.000	Temperatura12	1.000	0.000
PuntoRocio0	0.968	0.000	Temperatura13	1.000	0.000
PuntoRocio1	0.960	0.000	Temperatura14	1.000	0.000
PuntoRocio2	0.949	0.000	Temperatura15	1.000	0.000
PuntoRocio3	0.936	0.000	Temperatura16	1.000	0.000
PuntoRocio4	0.924	0.000	Temperatura17	1.000	0.000
PuntoRocio5	0.915	0.000	Temperatura18	1.000	0.000
PuntoRocio6	0.911	0.000	Temperatura19	1.000	0.000
PuntoRocio7	0.949	0.000	Temperatura20	1.000	0.000
PuntoRocio8	0.984	0.000	Temperatura21	1.000	0.000
PuntoRocio9	0.983	0.000	Temperatura22	0.999	0.000
PuntoRocio10	0.978	0.000	Temperatura23	0.997	0.000
PuntoRocio11	0.974	0.000			

Anexo 2: Coeficiente de Correlación de Spearman por cada variable climática para las 24 horas del día y la temperatura mínima de la madrugada siguiente.

Variable	CCS	Nivel de significancia (Valor p[1])	Variable	CCS	Nivel de significancia (Valor p[1])
Humedad0	-0,022	0,138	PuntoRocio12	0,380	**4,47E-155**
Humedad1	-0,044	**0,003**	PuntoRocio13	0,345	**1,68E-126**
Humedad2	-0,063	**2,19E-05**	PuntoRocio14	0,339	**5,00E-122**
Humedad3	-0,078	**1,64E-07**	PuntoRocio15	0,354	**1,43E-133**
Humedad4	-0,087	**5,19E-09**	PuntoRocio16	0,371	**7,61E-148**
Humedad5	-0,093	**3,10E-10**	PuntoRocio17	0,425	**2,44E-197**
Humedad6	-0,103	**4,25E-12**	PuntoRocio18	0,508	**2,01E-295**
Humedad7	-0,108	**3,06E-13**	PuntoRocio19	0,585	**0,00E+00**
Humedad8	-0,025	0,097	PuntoRocio20	0,627	**0,00E+00**

Humedad9	0,153	**4,79E-25**	PuntoRocio21	0,672	**0,00E+00**
Humedad10	0,303	**6,50E-97**	PuntoRocio22	0,721	**0,00E+00**
Humedad11	0,349	**1,69E-129**	PuntoRocio23	0,773	**0**
Humedad12	0,335	**3,87E-119**	Radiacion0	0,041	**0,006**
Humedad13	0,305	**7,82E-98**	Radiacion1	0,035	**0,020**
Humedad14	0,316	**3,39E-105**	Radiacion2	0,013	0,385
Humedad15	0,347	**4,12E-128**	Radiacion3	0,003	0,857
Humedad16	0,394	**1,55E-167**	Radiacion4	0,033	**0,026**
Humedad17	0,447	**1,63E-220**	Radiacion5	0,037	**0,012**
Humedad18	0,430	**2,70E-203**	Radiacion6	0,076	**2,86E-07**
Humedad19	0,360	**2,65E-138**	Radiacion7	-0,071	**1,74E-06**
Humedad20	0,261	**1,81E-71**	Radiacion8	-0,097	**5,28E-11**
Humedad21	0,151	**2,57E-24**	Radiacion9	-0,164	**1,33E-28**
Humedad22	0,054	**0,000**	Radiacion10	-0,169	**2,58E-30**
Humedad23	-0,022	0,149	Radiacion11	-0,161	**1,70E-27**
Precipitacion0	0,097	**7,38E-11**	Radiacion12	-0,134	**1,51E-19**
Precipitacion1	0,112	**4,25E-14**	Radiacion13	-0,153	**4,23E-25**
Precipitacion2	0,111	**7,34E-14**	Radiacion14	-0,204	**1,55E-43**
Precipitacion3	0,099	**3,09E-11**	Radiacion15	-0,245	**6,31E-63**
Precipitacion4	0,089	**2,41E-09**	Radiacion16	-0,292	**2,33E-89**
Precipitacion5	0,077	**2,08E-07**	Radiacion17	-0,292	**1,03E-89**
Precipitacion6	0,047	**0,001**	Radiacion18	-0,205	**2,77E-44**
Precipitacion7	0,037	**0,013**	Radiacion19	0,027	0,073
Precipitacion8	0,043	**0,004**	Radiacion20	0,011	0,457
Precipitacion9	0,038	**0,010**	Radiacion21	0,024	0,104
Precipitacion10	0,024	0,106	Radiacion22	0,009	0,557
Precipitacion11	0,026	0,080	Radiacion23	0,015	3,08E-01
Precipitacion12	0,036	**0,015**	Temperatura0	0,438	**2,97E-211**
Precipitacion13	0,055	**0,000**	Temperatura1	0,458	**4,22E-233**
Precipitacion14	0,072	**1,45E-06**	Temperatura2	0,471	**2,11E-248**
Precipitacion15	0,097	**7,26E-11**	Temperatura3	0,484	**8,26E-264**
Precipitacion16	0,114	**1,35E-14**	Temperatura4	0,487	**3,00E-268**
Precipitacion17	0,127	**1,11E-17**	Temperatura5	0,495	**1,73E-278**
Precipitacion18	0,139	**5,53E-21**	Temperatura6	0,497	**1,50E-280**
Precipitacion19	0,163	**2,55E-28**	Temperatura7	0,467	**1,26E-243**
Precipitacion20	0,173	**1,46E-31**	Temperatura8	0,317	**7,86E-106**
Precipitacion21	0,155	**1,30E-25**	Temperatura9	0,133	**2,24E-19**
Precipitacion22	0,130	**1,95E-18**	Temperatura10	0,000	0,977
Precipitacion23	0,134	**1,68E-19**	Temperatura11	-0,071	**2,07E-06**
PuntoRocio0	0,437	**2,24E-210**	Temperatura12	-0,115	**1,05E-14**
PuntoRocio1	0,453	**3,79E-228**	Temperatura13	-0,124	**5,55E-17**

PuntoRocio2	0,463	**3,25E-239**	Temperatura14	-0,163	**2,38E-28**
PuntoRocio3	0,476	**4,94E-254**	Temperatura15	-0,203	**4,16E-43**
PuntoRocio4	0,481	**4,96E-260**	Temperatura16	-0,255	**8,72E-68**
PuntoRocio5	0,491	**9,85E-273**	Temperatura17	-0,275	**5,16E-79**
PuntoRocio6	0,494	**1,15E-276**	Temperatura18	-0,023	1,30E-01
PuntoRocio7	0,486	**9,92E-267**	Temperatura19	0,344	**4,51E-126**
PuntoRocio8	0,427	**1,64E-199**	Temperatura20	0,530	**0,00E+00**
PuntoRocio9	0,348	**5,43E-129**	Temperatura21	0,637	**0,00E+00**
PuntoRocio10	0,359	**8,42E-138**	Temperatura22	0,718	**0,00E+00**
PuntoRocio11	0,389	**2,16E-163**	Temperatura23	0,782	**0**
Mes	0,126	**1,90E-17**	Ano	-0,025	0,088

[1]El valor p menor a 0,05 indica que la relación entre la variable climática para su correspondiente hora del día es significativa.

Anexo 3: Resultado de la prueba no paramétrica Kruskal Wallis para determinar por cada variable climática para las 24 horas del día, si la mediana es significativamente diferente para los días associados y no asociados a eventos de heladas.

Variable	H	Valor p[1]	Variable	H	Valor p[1]	Variable	H	Valor p[1]
Humedad0	0,029	0,865	Temperatura16	51,808	0,000	Radiacion8	0,673	0,412
Humedad1	0,492	0,483	Temperatura17	46,121	0,000	Radiacion9	3,928	0,047
Humedad2	1,912	0,167	Temperatura18	0,545	0,460	Radiacion10	24,638	0,000
Humedad3	4,660	0,031	Temperatura19	8,712	0,003	Radiacion11	33,059	0,000
Humedad4	8,129	0,004	Temperatura20	10,595	0,001	Radiacion12	45,507	0,000
Humedad5	6,067	0,014	Temperatura21	13,108	0,000	Radiacion13	41,159	0,000
Humedad6	10,389	0,001	Temperatura22	8,531	0,003	Radiacion14	55,414	0,000
Humedad7	14,365	0,000	Temperatura23	2,875	0,090	Radiacion15	11,132	0,001
Humedad8	19,133	0,000	Precipitacion0	0,014	0,904	Radiacion16	14,867	0,000
Humedad9	0,353	0,553	Precipitacion1	1,419	0,234	Radiacion17	4,094	0,043
Humedad10	56,159	0,000	Precipitacion2	1,279	0,258	Radiacion18	35,179	0,000
Humedad11	96,414	0,000	Precipitacion3	1,294	0,255	Radiacion19	1,232	0,267
Humedad12	103,150	0,000	Precipitacion4	1,247	0,264	Radiacion20	1,015	0,314
Humedad13	87,486	0,000	Precipitacion5	0,907	0,341	Radiacion21	1,278	0,258
Humedad14	81,667	0,000	Precipitacion6	0,754	0,385	Radiacion22	1,216	0,270
Humedad15	71,632	0,000	Precipitacion7	0,049	0,825	Radiacion23	1,092	0,296
Humedad16	77,543	0,000	Precipitacion8	0,072	0,789	PuntoRocio0	7,021	0,008
Humedad17	77,796	0,000	Precipitacion9	0,586	0,444	PuntoRocio1	17,516	0,000
Humedad18	66,582	0,000	Precipitacion10	0,526	0,468	PuntoRocio2	34,933	0,000
Humedad19	56,666	0,000	Precipitacion11	0,556	0,456	PuntoRocio3	60,089	0,000
Humedad20	44,237	0,000	Precipitacion12	0,112	0,738	PuntoRocio4	78,614	0,000
Humedad21	24,606	0,000	Precipitacion13	0,923	0,337	PuntoRocio5	84,366	0,000
Humedad22	25,476	0,000	Precipitacion14	1,529	0,216	PuntoRocio6	89,710	0,000

Humedad23	11,925	0,001	Precipitacion15	0,390	0,532	PuntoRocio7	42,313	0,000
Temperatura0	0,832	0,362	Precipitacion16	0,825	0,364	PuntoRocio8	17,634	0,000
Temperatura1	0,209	0,648	Precipitacion17	2,730	0,099	PuntoRocio9	24,358	0,000
Temperatura2	11,616	0,001	Precipitacion18	1,067	0,302	PuntoRocio10	12,495	0,000
Temperatura3	29,258	0,000	Precipitacion19	2,195	0,138	PuntoRocio11	1,228	0,268
Temperatura4	49,768	0,000	Precipitacion20	1,781	0,182	PuntoRocio12	0,455	0,500
Temperatura5	67,241	0,000	Precipitacion21	1,560	0,212	PuntoRocio13	0,814	0,367
Temperatura6	75,464	0,000	Precipitacion22	0,018	0,895	PuntoRocio14	0,387	0,534
Temperatura7	26,520	0,000	Precipitacion23	1,139	0,286	PuntoRocio15	0,176	0,675
Temperatura8	26,481	0,000	Radiacion0	1,341	0,247	PuntoRocio16	0,847	0,357
Temperatura9	12,200	0,000	Radiacion1	1,108	0,293	PuntoRocio17	8,215	0,004
Temperatura10	0,153	0,696	Radiacion2	1,185	0,276	PuntoRocio18	3,098	0,078
Temperatura11	12,322	0,000	Radiacion3	1,294	0,255	PuntoRocio19	5,764	0,016
Temperatura12	30,386	0,000	Radiacion4	1,185	0,276	PuntoRocio20	6,871	0,009
Temperatura13	38,421	0,000	Radiacion5	2,227	0,136	PuntoRocio21	15,875	0,000
Temperatura14	42,751	0,000	Radiacion6	0,075	0,784	PuntoRocio22	18,288	0,000
Temperatura15	40,759	0,000	Radiacion7	3,526	0,060	PuntoRocio23	33,809	0,000

[1]El valor p inferior a 0,05 indica que la mediana para los días no asociados a heladas es diferente a la mediana de la población para los días asociados a eventos de heladas.

Anexo 4: Valor de la importancia de cada variable mediante el índice de impureza de Gini (Bosques aleatorios para la selección de características).

Variable	RF Puntuación	Variable	RF Puntuación	Variable	RF Puntuación
Humedad0	0,006	Precipitacion16	0,001	Radiacion8	0,006
Humedad1	0,003	Precipitacion17	0,000	Radiacion9	0,006
Humedad2	0,004	Precipitacion18	0,000	Radiacion10	0,006
Humedad3	0,007	Precipitacion19	0,000	Radiacion11	0,003
Humedad4	0,004	Precipitacion20	0,000	Radiacion12	0,003
Humedad5	0,006	Precipitacion21	0,000	Radiacion13	0,004
Humedad6	0,002	Precipitacion22	0,000	Radiacion14	0,005
Humedad7	0,005	Precipitacion23	0,000	Radiacion15	0,005
Humedad8	0,008	PuntoRocio0	0,009	Radiacion16	0,005
Humedad9	0,008	PuntoRocio1	0,005	Radiacion17	0,010
Humedad10	0,003	PuntoRocio2	0,004	Radiacion18	0,002
Humedad11	0,007	PuntoRocio3	0,009	Radiacion19	NA
Humedad12	0,013	PuntoRocio4	0,007	Radiacion20	NA
Humedad13	0,008	PuntoRocio5	0,025	Radiacion21	NA
Humedad14	0,010	PuntoRocio6	0,012	Radiacion22	NA
Humedad15	0,004	PuntoRocio7	0,023	Radiacion23	NA

Humedad16	0,009	PuntoRocio8	0,007	Temperatura0	0,002
Humedad17	0,017	PuntoRocio9	0,007	Temperatura1	0,004
Humedad18	0,018	PuntoRocio10	0,004	Temperatura2	0,006
Humedad19	0,021	PuntoRocio11	0,009	Temperatura3	0,008
Humedad20	0,014	PuntoRocio12	0,013	Temperatura4	0,013
Humedad21	0,010	PuntoRocio13	0,010	Temperatura5	0,026
Humedad22	0,005	PuntoRocio14	0,018	Temperatura6	0,013
Humedad23	0,003	PuntoRocio15	0,007	Temperatura7	0,028
Precipitacion0	0,000	PuntoRocio16	0,013	Temperatura8	0,021
Precipitacion1	0,000	PuntoRocio17	0,023	Temperatura9	0,006
Precipitacion2	0,000	PuntoRocio18	0,021	Temperatura10	0,007
Precipitacion3	0,000	PuntoRocio19	0,037	Temperatura11	0,005
Precipitacion4	0,000	PuntoRocio20	0,031	Temperatura12	0,002
Precipitacion5	0,000	PuntoRocio21	0,021	Temperatura13	0,005
Precipitacion6	0,000	PuntoRocio22	0,037	Temperatura14	0,007
Precipitacion7	0,000	PuntoRocio23	0,025	Temperatura15	0,003
Precipitacion8	0,001	Radiacion0	NA	Temperatura16	0,003
Precipitacion9	0,000	Radiacion1	NA	Temperatura17	0,007
Precipitacion10	0,000	Radiacion2	NA	Temperatura18	0,005
Precipitacion11	0,000	Radiacion3	NA	Temperatura19	0,005
Precipitacion12	0,000	Radiacion4	NA	Temperatura20	0,011
Precipitacion13	0,000	Radiacion5	NA	Temperatura21	0,027
Precipitacion14	0,000	Radiacion6	0,004	Temperatura22	0,057
Precipitacion15	0,002	Radiacion7	0,005	Temperatura23	0,064