

UNIVERSIDAD NACIONAL DE JAÉN
CARRERA PROFESIONAL DE INGENIERÍA CIVIL



**ESTIMACIÓN DEL NIVEL DE RIESGO DE INUNDACIÓN FLUVIAL
MEDIANTE EL ALGORITMO RANDOM FOREST EN EL SECTOR
MAGLLANAL, JAÉN, CAJAMARCA, 2022**

**TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE
INGENIERO CIVIL**

**Autores : Bach. Muñoz Bustamante, Yeni Lizbeth
Bach. Yamunaque Tocto, Lady Abigail**
Asesor : Mg. Billy Alexis Cayatopa Calderón

LÍNEA DE INVESTIGACIÓN LI_IC_03 Gestión y Tecnología

JAÉN – PERÚ, JUNIO, 2024

NOMBRE DEL TRABAJO

ESTIMACIÓN DEL NIVEL DE RIESGO DE INUNDACIÓN FLUVIAL MEDIANTE EL ALGORITMO RANDOM FOREST EN EL SECTOR

AUTOR

Yeni Lizbeth Muñoz Bustamante Lady Abigail Yamunaque Tocto

RECuento DE PALABRAS

19376 Words

RECuento DE CARACTERES

111189 Characters

RECuento DE PÁGINAS

107 Pages

TAMAÑO DEL ARCHIVO

4.3MB

FECHA DE ENTREGA

Jun 25, 2024 12:32 PM GMT-5

FECHA DEL INFORME

Jun 25, 2024 12:34 PM GMT-5

● **13% de similitud general**

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

- 12% Base de datos de Internet
- Base de datos de Crossref
- 5% Base de datos de trabajos entregados
- 2% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossref

● **Excluir del Reporte de Similitud**

- Material bibliográfico
- Coincidencia baja (menos de 15 palabras)

 UNIVERSIDAD NACIONAL DE JAÉN
Dr. Alexander Huamán Mera
Responsable de la Unidad de Investigación de la Facultad de Ingeniería



UNIVERSIDAD NACIONAL DE JAÉN

Ley de Creación N° 29304

Universidad Licenciada con Resolución del Consejo Directivo N° 002-2018-SUNEDU/CD

ESCUELA PROFESIONAL DE INGENIERÍA CIVIL



FORMATO 03: ACTA DE SUSTENTACIÓN

En la ciudad de Jaén, el día 24 de junio del año 2024, siendo las 12:00 horas, se reunieron de manera presencial los integrantes del Jurado:

Presidente : Dra. Zadith Nancy Garrido Campaña.

Secretario : Dr. Manuel Emilio Milla Pino.

Vocal : M. Sc. Marcos Antonio Gonzales Santisteban., para evaluar la Sustentación del

Informe Final:

() Trabajo de Investigación

(**X**) Tesis

() Trabajo de Suficiencia Profesional

Titulado: **“NIVEL DE RIESGO ANTE INUNDACIÓN FLUVIAL APLICANDO LA METODOLOGÍA CENEPRED EN EL SECTOR MAGLLANAL- JAÉN - CAJAMARCA - 2022”**, presentado por las tesoreras **Muñoz Bustamante Yeni Libeth y Yamunaque Tocto Lady Abigail** de la Escuela Profesional de Ingeniería Civil de la Universidad Nacional de Jaén.

Después de la sustentación y defensa, el Jurado acuerda:

(**X**) Aprobar () Desaprobar (**X**) Unanimidad () Mayoría

Con la siguiente mención:

- | | | |
|----------------|------------|--------|
| a) Excelente | 18, 19, 20 | () |
| b) Muy bueno | 16, 17 | () |
| c) Bueno | 14, 15 | (15) |
| d) Regular | 13 | () |
| e) Desaprobado | 12 ó menos | () |

Siendo las 13:00 horas del mismo día, el Jurado concluye el acto de sustentación confirmando su participación con la suscripción de la presente.

Dra. Zadith Nancy Garrido Campaña
Presidente

Dr. Manuel Emilio Milla Pino
Secretario

M. Sc. Marcos Antonio Gonzales Santisteban
Vocal

ÍNDICE

ÍNDICE DE FIGURAS.....	iv
ÍNDICE DE TABLAS	vii
ÍNDICE DE ANEXOS.....	viii
RESUMEN	xi
ABSTRACT	xii
I. INTRODUCCIÓN.....	13
1.1. Realidad Problemática.....	13
1.2. Antecedentes de la investigación	16
1.3. Objetivos	29
II. MATERIALES Y MÉTODOS	30
2.1. Objeto de estudio.....	30
2.2. Características generales del área de estudio.....	30
2.3. Población, muestra y muestreo.....	31
2.4. Variables de estudio	31
2.5. Tipo de investigación	33
2.6. Procedimiento.....	34
III. RESULTADOS.....	74
3.1. Identificación del nivel de peligro como consecuencia de inundación fluvial	74
3.2. Análisis del nivel de vulnerabilidad como consecuencia de inundación fluvial	76
3.3. Establecer el nivel de riesgo de inundación fluvial.....	79
3.4. Modelación del comportamiento del nivel de riesgo de inundación fluvial mediante el algoritmo Random Forest en el Sector Magllanal, Jaén.....	80
IV. DISCUSIÓN	95
V. CONCLUSIONES Y RECOMENDACIONES.....	99
5.1. Conclusiones	99
5.2. Recomendaciones.....	101
VI. REFERENCIAS BIBLIOGRÁFICAS.....	103
AGRADECIMIENTO	108

DEDICATORIA	109
ANEXOS	110

ÍNDICE DE FIGURAS

Figura 1. Ubicación geográfica del área estudiada	30
Figura 2. Metodología general para la estimación del nivel de riesgo	34
Figura 3. Coordinación con el presidente de la Junta Vecinal del Sector Magllanal	36
Figura 4. Vista en planta de la zanora Magllanal.....	37
Figura 5. Muros de contención de 2 metros de altura.....	38
Figura 6. Badenes con tirante hidráulico menor a 0.50 metros.....	38
Figura 7. Puente peatonal sin mantenimiento.....	39
Figura 8. Viviendas construidas sobre muros de contención.....	40
Figura 9. Altura del tirante de agua.....	42
Figura 10. Realización de encuestas	46
Figura 11. Importación de la base de datos.....	57
Figura 12. Limpieza de datos.....	57
Figura 13. Datos de las variables seleccionadas	59
Figura 14. Interfaz de Weka.....	63
Figura 15. Interfaz de la opción Classify	63
Figura 16. Modelo I (LY_1)	65
Figura 17. Modelo II (LY_2)	66
Figura 18. Modelo III (LY_3).....	67
Figura 19. Modelo IV (LY_4).....	68
Figura 20. Modelo V (LY_5).....	69
Figura 21. Matriz de confusión de la clasificación	71
Figura 22. Niveles de peligro del área estudiada	76

Figura 23. Nivel de vulnerabilidad social.....	77
Figura 24. Nivel de vulnerabilidad económica.....	77
Figura 25. Nivel de vulnerabilidad ambiental.....	78
Figura 26. Niveles de la vulnerabilidad del área de estudio del Sector Magllanal	78
Figura 27. Niveles de Riesgo.....	79
Figura 28. Histograma de la base de datos.....	82
Figura 29. Gráfico de bigotes de la base de datos.....	83
Figura 30. Variables del modelo I (LY_1).....	85
Figura 31. Variables del modelo II (LY_2)	86
Figura 32. Variables del modelo III (LY_3)	87
Figura 33. Variables del modelo IV (LY_4).....	88
Figura 34. Variables del modelo V (LY_5)	89
Figura 35. Niveles de riesgo	90
Figura 36. Valores de la escala desarrollada por Saaty.....	115
Figura 37. Matriz de riesgo.....	295
Figura 38. Vista en planta de la zanora Magllanal.....	320
Figura 39. Vista de los elementos expuestos	320
Figura 40. Puente peatonal sin mantenimiento.....	321
Figura 41. Puente peatonal con presencia de grietas, barandas oxidadas, maderas rotas, y fierros expuestos.....	321
Figura 42. Levantamiento topográfico con GNSS aplicando la metodología RTK en la zanora Magllanal.....	322

Figura 43. Levantamiento topográfico con GNSS aplicando la metodología RTK dentro de la zanora Magllanal.....	322
Figura 44. Levantamiento topográfico con GNSS aplicando la metodología RTK fuera de la zanora Magllanal.....	323
Figura 45. Realización de encuestas en la calle José Balta.....	324
Figura 46. Realización de encuestas en la calle Dos de Mayo	324
Figura 47. Realización de encuestas en la calle Pedro Ruiz Gallo	325
Figura 48. Realización de encuestas en la calle José Carlos Mariátegui	325
Figura 49. Realización de encuestas en la calle Manuel Gonzáles Prada.....	326
Figura 50. Realización de encuestas en la calle Túpac Amaru.....	326
Figura 51. Realización de encuestas en la calle Lolo Fernández.....	327
Figura 52. Calicata en la zanora Magllanal.....	327
Figura 53. Muestras de suelo por estrato	328
Figura 54. Cuarteo Manual- "Método B".....	328
Figura 55. Ensayo determinación del contenido de humedad MTC-108	329
Figura 56. Ensayo de densidad	329
Figura 57. Ensayo de Análisis granulométrico.....	330
Figura 58. Ensayo de determinación de límite líquido, límite plástico e índice de plasticidad	330
Figura 59. Procesamiento de datos en el software ARGIS en el laboratorio SIRANDA de la UNJ.....	331
Figura 60. Elaboración de mapas de peligro, vulnerabilidad y riesgo en el software ARGIS en el laboratorio SIRANDA de la UNJ	331

ÍNDICE DE TABLAS

Tabla 1. Operacionalización de variables	32
Tabla 2. Simbología y rango de las variables recolectadas del peligro y la vulnerabilidad	51
Tabla 3. Rankin de atributos según el algoritmo Random Forest.....	54
Tabla 4. Variables que aportan mayor información a la variable de salida	56
Tabla 5. Variables seleccionadas para los cinco grupos	60
Tabla 6. Parámetro de evaluación.....	74
Tabla 7. Factores condicionantes.....	75
Tabla 8. Factores desencadenantes	75
Tabla 9. Estadística de las variables.....	81
Tabla 10. Precisión de los modelos según el algoritmo Random Forest	90
Tabla 11. Predicción del modelo LY_1	91
Tabla 12. Predicción del modelo LY_3	93

ÍNDICE DE ANEXOS

Anexo 1. Matriz de consistencia.....	111
Anexo 2. Escala desarrollada por Saaty.....	114
Anexo 3. Estudio hidrológico de la microcuenca de la zanora Magllanal.....	116
Anexo 4. Ponderación del parámetro de evaluación-Proceso de Análisis Jerárquico	127
Anexo 5. Susceptibilidad del área estudiada- Proceso de Análisis Jerárquico	130
Anexo 6. Estudio de suelo para el parámetro condicionante: tipo de suelo.....	143
Anexo 7. Información de la precipitación del SENAMHI de la Estación de Jaén	172
Anexo 8. Componentes de la vulnerabilidad	182
Anexo 9. Formato de encuesta para el análisis de vulnerabilidad	187
Anexo 10. Resumen de encuestas para analizar el nivel de vulnerabilidad.....	193
Anexo 11. Análisis de la vulnerabilidad en la dimensión social-Proceso de Análisis Jerárquico	196
Anexo 12. Análisis de vulnerabilidad de la exposición social-Proceso de Análisis Jerárquico	199
Anexo 13. Análisis de vulnerabilidad de la fragilidad social-Proceso de Análisis Jerárquico	204
Anexo 14. Análisis de la vulnerabilidad de la resiliencia social- Proceso de Análisis Jerárquico	211
Anexo 15. Análisis de la vulnerabilidad en la dimensión económica-Proceso de Análisis Jerárquico	221
Anexo 16. Análisis de vulnerabilidad de la exposición económica- Proceso de Análisis Jerárquico	224

Anexo 17. Análisis de vulnerabilidad de la fragilidad económica- Proceso de Análisis Jerárquico	230
Anexo 18. Análisis de vulnerabilidad de la resiliencia económica- Proceso de Análisis Jerárquico	237
Anexo 19. Vulnerabilidad en la dimensión ambiental- Proceso de Análisis Jerárquico	247
Anexo 20. Análisis de vulnerabilidad de la exposición ambiental- Proceso de Análisis Jerárquico	250
Anexo 21. Análisis de vulnerabilidad de la fragilidad ambiental- Proceso de Análisis Jerárquico	253
Anexo 22. Análisis de vulnerabilidad de la resiliencia ambiental- Proceso de Análisis Jerárquico	258
Anexo 23. Matriz Saaty de dimensiones social, económica y ambiental	263
Anexo 24. Matriz Saaty de los parámetros de fragilidad en la dimensión Social.....	267
Anexo 25. Matriz Saaty de los parámetros de resiliencia en la dimensión Social.....	271
Anexo 26. Cuadro de Matriz Saaty de los parámetros de fragilidad en la dimensión económica	276
Anexo 27. Cuadro de Matriz Saaty de los parámetros de resiliencia en la dimensión económica	280
Anexo 28. Determinación de rangos de nivel de vulnerabilidad	285
Anexo 29. Resultados de nivel de vulnerabilidad por lote	291
Anexo 30. Matriz de riesgo.....	294
Anexo 31. Estratificación de los niveles de peligro en el area de estudio del Sector Magllanal.....	296

Anexo 32. Estratificación de los niveles de vulnerabilidad en el área de estudio del Sector Magllanal.....	298
Anexo 33. Estratificación del nivel de riesgo en el área de estudio del Sector Magllanal	303
Anexo 34. Solicitud de constancia de Co-asesor y evaluador en la ponderación de parámetros en la matriz SAATY	308
Anexo 35. Uso de licencia del software ARGIS en el laboratorio SIRANDA de la UNJ	315
Anexo 36. Panel Fotográfico	319
Anexo 37. Delimitación de la zanora Magllanal.....	332
Anexo 38. Delimitación de la microcuenca de la zanora Magllanal	334
Anexo 39. Mapas	336
Anexo 40. Validación de instrumentos.....	364

RESUMEN

La presente investigación tuvo como finalidad estimar el nivel de riesgo de inundación fluvial mediante el algoritmo Random Forest en el Sector Magllanal, Jaén, Cajamarca. Fue de tipo básica, de diseño descriptivo no experimental y de enfoque cuantitativo; ya que se empleó la metodología CENEPRED para identificar y analizar las variables que influyen en el peligro y la vulnerabilidad del sector; a partir de estas variables, se desarrollaron cinco modelos utilizando el algoritmo Random Forest para estimar el nivel de riesgo de 68 lotes en escenarios futuros. Los resultados indicaron que el nivel riesgo de inundación fluvial es muy alto para el 1.47% de los lotes (1 lote), alto para el 66.18% (46 lotes) y medio para el 32.35% (22 lotes); con el modelo más preciso alcanzando una precisión del 98.53% con un margen de error menor al 5%. Concluyendo que el algoritmo Random Forest estimó que el nivel de riesgo de inundación fluvial en el Sector Magllanal será alto; para lo cual, se propuso medidas estructurales adecuadas y planes efectivos que permitan mitigar el riesgo de inundación fluvial en el área estudiada.

Palabras clave: Riesgo, peligro, vulnerabilidad e inundación fluvial.

ABSTRACT

The purpose of this research was to estimate the level of fluvial flood risk using the Random Forest algorithm in the Magllanal Sector, Jaén, Cajamarca. It was a basic, descriptive non-experimental design with a quantitative approach, since the CENEPRED methodology was used to identify and analyze the variables that influence the hazard and vulnerability of the sector; based on these variables, five models were developed using the Random Forest algorithm to estimate the risk level of 68 lots in future scenarios. The results indicated that the fluvial flood risk level is very high for 1.47% of the lots (1 lot), high for 66.18% (46 lots) and medium for 32.35% (22 lots); with the most accurate model reaching an accuracy of 98.53% with a margin of error of less than 5%. In conclusion, the Random Forest algorithm estimated that the level of fluvial flood risk in the Magllanal Sector will be high; for which, adequate structural measures and effective plans were proposed to mitigate the risk of fluvial flooding in the studied area.

Key words: Risk, hazard, vulnerability and fluvial flooding.

I. INTRODUCCIÓN

1.1. Realidad Problemática

A nivel mundial, se observa un aumento de las inundaciones debido a las lluvias intensas a causa del cambio climático y demográfico la cual representan uno de catástrofes naturales más frecuentes y destructivos que afectan a la población (Organización Meteorológica Mundial [OMM], 2021). En los últimos 23 años Brasil, Colombia y Perú han sido los países más afectados por los desastres ante inundaciones, en la cual se han registrado más de 681 inundaciones graves que han ocasionado 28 billones de dólares de pérdidas económicas en diferentes tipos de sectores, así como también 49 millones de personas afectadas (UNDRR, 2023). Asimismo, en otros países del mundo como en Pakistán el costo de pérdidas económicas por las inundaciones que se produjo a causa del fenómeno climático La Niña es de US\$ 15,000 millones; donde muchas personas perdieron sus pertenencias y 1700 perdieron la vida. También en Florida debido al huracán Ian provocó graves inundaciones generando daños a infraestructuras y daño eléctrico lo que causó US\$ 65.000 millones de pérdidas económicas y 114 pérdidas de vidas humanas (Munich RE, 2022).

A nivel nacional, 16 regiones y 483 distritos en Perú se ha visto en riesgo debido a la presencia del ciclón Yaku, un fenómeno de características tropicales no organizado (Centro de Operaciones de Emergencia Nacional [COEN], 2023), ha provocado intensas lluvias la cual ha generado inundaciones, desbordes de ríos y activaciones de quebradas en diferentes zonas de nuestro país afectando la población, infraestructura de viviendas, pública y vial (Portillo et al., 2023). Asimismo, en el año 2023 debido a las inundaciones se han registrado 12 200 de familias damnificadas, 61 muertos, 57 heridos, 49 111 afectados, adicionalmente se generaron múltiples

pérdidas de bienes materiales y daños en infraestructura, en la cual se contabilizaron afectaciones en 642 aulas, 1 326 viviendas, 94 centros de salud, 188 puentes y 638,4 km de carretera (Instituto Nacional de Defensa Civil [INDECI], 2023) .

En los últimos años, el Sector Magllanal ha enfrentado una vulnerabilidad significativa debido a los recurrentes desbordamientos de la zanora Magllanal; esto a causa de lluvias intensas, lo que supera el cauce principal de la zanora provocando su activación y en efecto generando consecuencias graves con pérdidas humanas y económicas en la población del Sector Magllanal. Además, la zanora del Sector carece de medidas tanto estructurales como no estructurales que permitan mitigar los riesgos de inundación fluvial, lo que aumenta aún más la exposición de la población a estos peligros. En este contexto, resulta imprescindible realizar una estimación del nivel de riesgo en el que se encontrara la población aledaña a la zanora Magllanal mediante el algoritmo Random Forest.

1.1.1. Formulación del Problema

¿Cuál será el nivel de riesgo de inundación fluvial mediante el algoritmo Random Forest en el Sector Magllanal, Jaén, Cajamarca,2022?

1.1.2. Hipótesis

El nivel de riesgo de inundación fluvial mediante el algoritmo Random Forest se encontrará en un nivel alto con un error menor al 5% en el Sector Magllanal, Jaén, Cajamarca.

1.1.3. Justificación

Desde el punto de vista Metodológico se justifica porque en la ciudad de Jaén, Cajamarca no existe investigaciones sobre esta problemática. Esta falta de indagación previa destaca la necesidad de ampliar los antecedentes y la base de conocimientos existentes en relación a la estimación del nivel de riesgo de inundación fluvial mediante el algoritmo Random Forest. La ausencia de estudios recientes subraya la relevancia y oportunidad de la presente investigación, la cual propone contribuir significativamente al corpus existente de conocimientos en este ámbito específico. Al abordar esta laguna de información, se aspira a proporcionar aportes sustanciales que no solo amplíen la comprensión del problema, sino que también sirvan como base para futuras investigaciones y decisiones basadas en el algoritmo Random Forest para la estimación del nivel de riesgo de inundación fluvial en la región.

Desde el punto de vista social la presente investigación se justifica porque se fundamenta en la recurrente afectación que experimenta la población del sector Magllanal durante episodios de intensas lluvias. En tales circunstancias, el desbordamiento de la zanora Magllanal provoca inundaciones, ocasionando perjuicios tanto a nivel personal como en la infraestructura pública y vial. Esta investigación es crucial porque proporcionará información esencial para abordar, predecir y mitigar el riesgo de inundación fluvial. Además, los resultados obtenidos de la investigación serán fundamentales para respaldar la formulación de políticas públicas y la asignación de recursos, facilitando así la implementación de medidas preventivas y de mitigación, que servirá como base para la planificación de inversiones públicas destinadas a reducir el riesgo de inundación en el sector Magllanal.

Desde el punto de vista técnico la investigación permitirá estimar el nivel de riesgo ante inundación fluvial en la zona Magllanal mediante el algoritmo Random Forest. Para ello se realizará el diagnóstico de las áreas vulnerables ante el peligro de inundación fluvial. Además, este enfoque busca salvaguardar la integridad de la población y la infraestructura en escenarios futuros, para contribuir en la implementación de medidas específicas con el fin de garantizar mayor seguridad en la comunidad local, lo que incidirá positivamente en la optimización y eficiencia de las políticas de desarrollo regional, consolidando así un enfoque integral y sostenible.

1.2. Antecedentes de la investigación

1.2.1. A nivel internacional

Hernández (2023) en su investigación titulada “Modelaje de inundaciones en el departamento de Antioquia a partir de redes neuronales” tuvo como objetivo evaluar diferentes modelos de machine learning para predecir las pérdidas por inundación en la producción de agricultores colombianos; con la finalidad de diseñar un seguro indexado. La metodología en la presente investigación es de tipo básica y de diseño descriptivo con enfoque cuantitativo en la cual se obtuvo información de datos históricos de inundaciones, se realizó la construcción de base datos representativa; asimismo se utilizó el método de Cesarini et al. (2021) y Figueiredo et al. (2018) para la implementación, evaluación y selección de modelos. Se obtuvo como resultado 27,000 configuraciones y 15 combinaciones con su respectivo modelo de regresión logística; donde se seleccionó el modelo del algoritmo Hyperband con el fin de reducir el espacio de configuraciones respecto a la función pérdida. Concluyendo que el mejor modelo de redes neuronales es el algoritmo Hyperband para predecir las pérdidas por inundación y analizar el costo promedio de un

seguro indexado por inundaciones; el cual alcanzó una precisión de 0.3150 en validación y 0.2868 para los datos de prueba.

Cárdenas et al. (2022) en su investigación titulada “Sistema de alerta temprana de inundaciones para el río Arauca basado en técnicas de inteligencia artificial” tuvo como objetivo diseñar un modelo matemático a partir de tres clases de tendencias: lineal, potencia y relaciones potenciales para el cauce del río Arauca basado en un sistema inteligente de alerta temprana de inundaciones. La metodología empleada es de tipo básica y de diseño descriptivo con enfoque cuantitativo, en lo cual se utilizó técnicas de Inteligencia Artificial (AI), como: redes neuronales, ANFIS y algoritmo RN a partir de variables seleccionadas que definirán la tendencia del modelo matemático de caudal del río Arauca. Se obtuvo como resultado que los tres modelos matemáticos: lineal, potencia y relaciones potenciales, siendo este último el mejor modelo que permite generar el sistema de alerta temprana más precisa de inundaciones para el río Arauca con un error de 0.9989. Concluyendo que los modelos basados en redes neuronales artificiales son más efectivos que los modelos hidrológicos tradicionales; y el uso de este tipo de soluciones puede ser una buena alternativa para generar sistemas de alerta temprana.

García et al. (2021) en su investigación titulada “Estimación de la susceptibilidad a inundaciones en la cuenca inferior de Río Negro, Argentina” tuvo como objetivo estimar las zonas o áreas más susceptibles a desbordes fluviales en la cuenca inferior del Río Negro Argentina. La metodología en la presente investigación es de tipo básica y de diseño descriptivo con enfoque cuantitativo, en la cual se utilizó el método de Evaluación Multicriterio junto con el Sistema de Información Geográfica (SIG) para analizar factores físicos como textura del suelo, pendiente

topográfica y la geomorfología de un área de 2828 km². Se obtuvo como resultado zonas con susceptibilidad media de un 42% (1187.4 km²) y con susceptibilidad baja un 53.35% (1506.5 km²) y con susceptibilidad más alta un 4.7% (134.2 km²). Concluyendo que la susceptibilidad a inundaciones en la cuenca inferior de Río Negro de Argentina es de nivel media a alta, en zonas con drenaje del suelo deficiente y en pendientes topográficas de 1%. El aporte de esta investigación en mi tesis es el uso del (SIG) para el análisis y registro de datos.

Fraga et al. (2021) en su investigación titulada “MERLIN: Una nueva herramienta para la predicción del riesgo de inundaciones en la demarcación hidrográfica Galicia-Costa” tuvo como objetivo estimar el riesgo de inundaciones a partir de predicciones de caudales para determinar los calados y extensión de zonas inundadas. La metodología en la presente investigación es de tipo básica y de diseño descriptivo, con enfoque cuantitativo, en la cual se utilizó el modelo hidrológico HEC-HMS para predecir los caudales y el modelo hidráulico 2D Iber+ para predecir la extensión de las zonas inundadas a partir de observaciones y predicciones hidro-meteorológicas en 4 cuencas de Galicia-Costa. Se obtuvo como resultado que el riesgo de inundación muestra una mejor predicción en las cuencas de mayor tamaño siendo estas el río Landro y Verdugo con caudales máximos de 301.5 m³/s y 253.8 m³/s respectivamente; con un intervalo de confianza del 95% de los caudales predichos. Se concluye que el Modelo de Evaluación del Riesgo Local de Inundación (MERLIN) es una herramienta eficaz para estimar el riesgo de inundación utilizando modelos meteorológicos, hidrológicos e hidráulicos; demostrando su precisión y confiabilidad para determinar la extensión de las inundaciones previstas en las zonas costeras de Galicia Costa.

Alarcón et al. (2020) en su investigación titulada “Estimación de áreas vulnerables a inundaciones en zonas urbanas: Morelia, Michoacán, México”, tuvo como objetivo determinar las

áreas vulnerables a inundaciones en la ciudad causadas por el río Grande de Morelia, mediante la aplicación de dos modelos matemáticos. La metodología en la presente investigación es de tipo básica y de diseño descriptivo, con enfoque cuantitativo, en la cual aplicó el modelo hidrológico en la cuenca utilizando el software HEC-HMS y el otro modelo hidráulico sobre el río utilizando el software IBER para identificar las zonas propensas a inundaciones. Se obtuvo como resultado del primer modelo con HEC-HMS la simulación de tormentas con períodos de retorno de 100 a 500 años en 24 horas y del modelo hidráulico en IBER las zonas vulnerables a inundaciones, para avenidas esperadas con períodos de retorno de 100 a 500 años. Se concluye que mediante la modelación hidráulica se puede estimar para períodos de retorno de 100 a 500 años las zonas vulnerables a inundaciones por avenidas máximas.

1.2.2. A nivel nacional

López et al. (2021) en su investigación titulada “Modelo predictivo basado en machine learning para la estimación de vulnerabilidades de riesgo de inundación y deslizamiento. Caso de estudio: instituciones educativas del Perú” tuvo como objetivo proponer un modelo predictivo Usando machine learning, con la finalidad de estimar riesgo de inundación y deslizamiento en instituciones educativas del Perú. La metodología en la presente investigación es de tipo aplicada, de diseño descriptivo-transversal en la cual se usó machine learning entrenando el modelo usando diversos algoritmos para la estimación de vulnerabilidades como árboles de decisión, K vecino más próximo, Multilayer Perceptron y el Random Forest y datos de un dataset con más de 65 000 registros publicados por el Ministerio de Educación del Perú. Se obtuvo como resultado que el mejor modelo con mayor predicción basado en machine learning para la estimación de vulnerabilidades en instituciones es el modelo de Random Forest en la cual presenta un porcentaje de acierto de 92,407 % para el riesgo de deslizamiento y 97,612 % para el riesgo de inundación.

Concluyendo que el método Random Forest es una herramienta de apoyo en la cual brinda mejores resultados para la estimación de vulnerabilidades de riesgo de inundación y deslizamiento con la cual se podrían tomar mejores decisiones para elegir donde construir una institución educativa.

Lopez (2021) en su investigación titulada "Evaluación del Riesgo por Inundaciones en las Viviendas del Asentamiento Humano Sagrado Corazón de Jesús, Distrito de Castilla - Piura" tuvo como objetivo determinar el riesgo ante inundaciones para proponer inversiones a fin de disminuir o mitigar el riesgo. La metodología en la presente investigación es de tipo básica, de diseño descriptivo y de enfoque cuantitativo ya que se identificó y mapeo a través del Software GIS el nivel de peligro y vulnerabilidad de 421 viviendas, asimismo para evaluar el riesgo por inundación fue por medio del Centro Nacional de Estimación, Prevención y Reducción del Riesgo de Desastres (CENEPRED). Los resultados indicaron que el nivel de riesgo por inundación es muy alto en 1.7% de las viviendas (7 viviendas), alto en 7.4% (31 viviendas) y medio en 89.8% (383 viviendas). Concluyendo que el nivel de riesgo por inundación se encuentra entre medio y alto, se propuso desarrollar actividades inmediatas y prioritarias para el manejo del riesgo. El aporte de esta investigación en mi tesis es que se utilizará la misma metodología aplicada por CENEPRED.

Anchelía y Mori (2020) en su investigación titulada "Determinación de zonas susceptibles a inundaciones y análisis comparativo del Proceso de Análisis Jerárquico (AHP) y Random Forest (RF). Caso de estudio: cuenca baja del río Chancay Lambayeque" tuvo como objetivo analizar y comparar el método AHP y RF, con la finalidad de fijar el más eficaz para determinar las zonas susceptibles a inundaciones en la cuenca baja del río Chancay Lambayeque. La metodología en la presente investigación es de tipo aplicada, de diseño descriptivo-transversal; ya que se recolectó

datos históricos de inundaciones y se interpretó imágenes satelitales, asimismo para el análisis comparativo de los modelos se tomó en cuenta 198 muestras (geología, cercanía del río, pendiente y precipitación); donde se utilizó 70% para el entrenamiento del modelo RF y el 30% para la validación de ambos modelos. Se obtuvo como resultado que el método RF tiene un valor de predicción de 0,9941 y el método AHP un valor de 0,9774. Concluyendo que el método más eficaz es el Random Forest para determinar la susceptibilidad frente a inundaciones; además de predecir el fenómeno para la toma de decisiones, así como medidas de control estructurales y no estructurales. El aporte de esta investigación en mi tesis es que se utilizará la metodología Random Forest (RF).

Guillen y Sifuentes (2020) en su investigación titulada “Evaluación de Riesgos Ante Amenazas de Inundación por Desborde del Rio Chorobal, Sector El Inca, Chao, Virú, La Libertad – 2020” tuvo como objetivo proponer medidas estructurales para reducir el riesgo en el tramo del Rio Chorobal límite con el sector El Inca. La metodología en la presente investigación es de tipo aplicada, de alcance explicativo y diseño no experimental en la cual para determinar los niveles de riesgos presentes en el territorio de estudio fue mediante (CENEPRED). Como resultado se obtuvo que el área de estudio predomina el nivel de riesgo medio al peligro de inundación por desborde. Concluyendo que la exposición de riesgos por desborde del río Chorobal radica principalmente debido al poco mantenimiento periódico, para lo cual se optó como medida estructural de solución los trabajos de limpieza y descolmatación del cauce; asimismo la colocación de plantones de forestación en el cauce para reducir el riesgo existente.

Zúñiga (2020) en su investigación titulada “Evaluación de Riesgo por Inundación Fluvial de la Ciudad de Huancavelica, Perú” tuvo como objetivo evaluar y determinar las zonas y niveles

de riesgo bajo, medio, alto y muy alto por inundación fluvial en la ciudad de Huancavelica. La metodología en la presente investigación es de tipo básica, de diseño descriptivo no experimental, en la cual utilizó el CENEPRED para determinar los niveles de peligrosidad, vulnerabilidad y en consecuencia los niveles de riesgo por inundación fluvial. Como resultado se obtuvo que las zonas con mayor riesgo por inundación fluvial son las zonas con cauces activos, zonas de erosión, los valles-quebradas y los acantilados con períodos de retorno de (10 y 50), (50 y 100), (100 y 200) y (200 y 500) años y los niveles de riesgo bajo, medio, alto y muy alto están en rangos de $0.001 \leq R \leq 0.005$, $0.005 \leq R \leq 0.018$, $0.018 \leq R \leq 0.066$ y $0.066 \leq R \leq 0.252$ respectivamente. Concluyendo que, al evaluar los diferentes niveles, la ciudad de Huancavelica tiene un nivel de riesgo bajo y medio por inundación fluvial.

1.2.3. A nivel regional

Ticlla (2022) en su investigación titulada “Evaluación del nivel de riesgo a inundación según la metodología CENEPRED, en el Río San Mateo, tramo puente “san Mateo” al puente “LLoque Yupanqui”, Chota” tuvo como objetivo estimar los niveles de peligrosidad, vulnerabilidad y riesgo en áreas probables a inundaciones fluviales por eventos de máxima intensidad en el río “San Mateo”. La metodología en la presente investigación es de tipo aplicada, de diseño descriptivo y con enfoque mixto en la cual utilizó el Centro Nacional de Estimación, Prevención y Reducción del Riesgo de Desastres (CENEPRED) para determinar los niveles de peligro, vulnerabilidad y riesgo. Como resultado se obtuvo niveles de peligrosidad muy alto en 2.53 ha, alto 0.68 ha, medio 9.51 ha y bajo 5.67 ha de 241 lotes; niveles de vulnerabilidad de 173 familias del barrio San Mateo es muy alta en 119 lotes, alta 68 lotes y media 54 lotes y niveles de riesgo muy alto en 0.69 ha, alto 6.15 ha, medio 2.55 ha y bajo 9.55 ha frente a inundaciones fluviales del río “San Mateo” para tiempos de retorno de 10, 25, 50 y 100 años. Se concluye que

el nivel de riesgo es bajo y las áreas con mayor probabilidad a inundaciones son las que presentan pendientes menores a 5%.

Goicochea (2021) en su investigación titulada “Estimación del Nivel de Riesgo por Inundación de la Quebrada Cruz Blanca para la Zona Urbana-Cajamarca, 2020” tuvo como objetivo estimar el nivel de riesgo por inundación para prevenir pérdidas económicas. La metodología en la presente investigación es de tipo básica, diseño descriptivo con enfoque cuantitativo en la cual utilizó el modelamiento hidrológico e hidráulico para determinar las áreas inundables y un diagnóstico de las condiciones físicas, económicas, sociales y ambientales de 628 viviendas para estimar las variables de vulnerabilidad y peligrosidad. Como resultado se obtuvo que la Zona Urbana-Cajamarca tiene en general un alto nivel de riesgo por inundación afectando un total de 48.4 hectáreas de las cuales 16.1 ha son calles, 31.4 ha son viviendas, 0.45 ha son lotes sin construcción, 0.17 ha son áreas de recreación y 0.297 ha son centros educativos. Concluyendo que el alto nivel de riesgo por inundación afecta directamente proporcional a la población para lo cual se debe elaborar obras de prevención y mitigación, asimismo capacitación en temas de gestión de riesgos.

Alaya Y Riquero (2020) en su investigación titulada “Caracterización de los modelamientos hidráulicos numéricos de inundaciones fluviales, Cajamarca 2020” tuvo como objetivo determinar las principales características y el método más usado para un óptimo modelamiento hidráulico de inundaciones fluviales. La metodología en la presente investigación es de tipo aplicada, de diseño descriptivo y con enfoque mixto en la cual se seleccionó, analizó y caracterizó 20 trabajos de modelamientos hidráulicos de inundaciones fluviales; se utilizó dos

fichas como instrumentos de recolección de datos y se determinó los parámetros más usados en el modelamiento. Se obtuvo como resultado que el 55% de trabajos recopila información topográfica de investigaciones anteriores, el 55% emplea un modelo digital del terreno (MDT), el 52% utiliza parámetros geomorfológicos de investigaciones anteriores, el 45% calcula los caudales máximos con software, el 76% emplea software HEC-RAS para el modelamiento hidráulico, el 61% realiza un modelo unidimensional, entre otros. Se concluye que el método más usado y las principales características para un modelamiento hidráulico se determina con respecto a la topografía, hidrología e hidráulica.

Mantilla (2019) en su investigación titulada “Identificación de áreas inundables mediante modelamiento hidráulico de la quebrada Chinchinmarca - Sector 24 (Villa Huacariz) - Cajamarca” tuvo como objetivo identificar las áreas inundables más vulnerables del sector 24 Villa Huacariz por avenidas extraordinarias mediante modelación hidráulica de la Quebrada Chinchinmarca - Cajamarca. La metodología en la presente investigación es de tipo básica y diseño descriptivo, con enfoque cuantitativo, en la cual se utilizó el modelamiento hidráulico mediante el software IBER para modelar los caudales para diferentes tiempos de retorno; el uso de ArcGIS para generar mapas temáticos donde se muestra la intensidad de inundación basada en criterios del INDECI. Se obtuvo como resultado que las áreas inundables se encuentran sobre la cota del terreno para los caudales $Q = 71.11 \text{ m}^3/\text{s}$; $Q = 66.09 \text{ m}^3/\text{s}$ y $Q = 63.24 \text{ m}^3/\text{s}$ para distintos periodos de retorno de 308.16, 475.06 y 975.29 años respectivamente, los tirantes oscilaron entre 0.39 y 3.50 m como máximo, obtenido un nivel de intensidad baja. Se concluye que la modelación hidráulica es una herramienta valiosa para evitar pérdidas de vidas, minimizar daños materiales e identificar las zonas más vulnerables frente a eventos de inundación.

Díaz (2019) en su investigación titulada “Determinación de zonas de inundación mediante modelamiento hidráulico del río Chonta en el caserío Tartar Chico, distrito de Baños del Inca” tuvo como objetivo estimar los caudales de máximas avenidas para identificar las zonas y niveles de inundación en el caserío de Tartar Chico. La metodología en la presente investigación es de tipo básica y diseño descriptivo, con enfoque cuantitativo, en la cual utilizó el modelamiento hidráulico mediante el software Iber para mitigar posibles daños a zonas agrícolas, ganaderas y vía asfáltica por eventuales inundaciones considerando tres caudales máximos. Como resultado se obtuvo de los caudales de mayor magnitud las zonas de pastos cortos, zonas de pastos altos, zona de asfalto susceptibles a una inundación con áreas de 5.82 ha, 076 ha, 0.17 ha respectivamente, además se determinaron los niveles como zonas de medio (0.38) y bajo (0.19) riesgo del área de estudio. Se concluye que mediante el modelamiento hidráulico permite delimitar y estimar las áreas inundables de cada zona de pastos cortos, pastos altos y asfalto.

1.2.4. A nivel local

Delgado (2023) en su investigación titulada “Riesgo por inundación del río La Envinada en el sector Pueblo Nuevo, distrito Chontalí, provincia Jaén, departamento Cajamarca” tuvo como objetivo determinar el nivel de riesgo por inundación en el sector Pueblo Nuevo del distrito de Chontalí. La metodología en la presente investigación es de tipo aplicada y de diseño no experimental; la cual utilizó el manual del Instituto Nacional de Defensa Civil (INDECI) para determinar los niveles de peligro, vulnerabilidad y riesgo; el método Hec-Ras para la modelación hidráulica con el fin de determinar áreas de inundación en períodos de retorno de 10, 50, 100 y 200 años. Como resultado se obtuvo el nivel de peligro muy alto en un rango de $0.75 < P \leq 1.00$ y

nivel de vulnerabilidad alto en un rango $0.50 < V \leq 0.75$; asimismo las áreas con más riesgo de inundación son 537.406 m², 1332.171 m², 1724.281 m² y 2362.400 m² para los periodos de retorno 10, 50, 100 y 200 respectivamente, encontrándose en las progresivas 0+150 y 0+250 a la altura del sector Pueblo Nuevo. Se concluye que el nivel de riesgo es alto ante inundación del río La Envinada en el Sector Pueblo Nuevo.

Paredes (2023) en su investigación titulada “Evaluación del nivel de riesgo de inundación por la quebrada Tumbillan en la localidad de Bellavista Viejo de la provincia de Jaén, región Cajamarca” tuvo como objetivo evaluar y estimar el nivel de riesgo por inundación de la quebrada Tumbillan en las áreas más vulnerables. La metodología en la presente investigación es descriptiva, según su finalidad es aplicada, de enfoque cuantitativo y diseño no experimental en la cual se evaluó el riesgo mediante la metodología CENEPRED e INDECI, se determinó el nivel de peligro en función de la frecuencia de inundación, la intensidad, y características geológicas de la localidad y se evaluó la vulnerabilidad mediante encuestas realizadas a 30 viviendas. Se obtuvo como resultado un nivel de riesgo de inundación muy alto frente a una avenida máxima; además las áreas con mayor vulnerabilidad a inundaciones se ubican en el lado izquierdo de la quebrada, que incluyen terrenos de cultivo y la propia localidad, donde el flujo del agua alcanza tirantes máximos superiores a 0.90 metros. Se concluye que la metodología del INDECI y CENEPRED son una herramienta fundamental para analizar la vulnerabilidad, peligro y determinar el nivel riesgo por inundación.

Guerrero (2022) en su investigación titulada “Estimación del nivel de riesgo por inundación del río Chinchipe en el centro poblado Puerto Huallape, Jaén-Cajamarca” tuvo como

objetivo estimar el nivel riesgo por inundación para prevenir las consecuencias de los desastres. La metodología aplicada en la investigación fue de tipo descriptiva y cuantitativa, debido a que se realizó la valorización del peligro y la vulnerabilidad mediante 6 tipos encuestas de 8 preguntas relacionadas a las viviendas, estructuras, eventos o fenómenos como las precipitaciones; a un total de 210 personas del centro poblado de Huallape. Se obtuvo como resultado que el nivel de peligro por inundación es medio en un rango de $0.00 < NP \leq 0.25$, el grado de vulnerabilidad es alta en un rango de $0.25 < VT \leq 0.75$. Concluyendo que el nivel de riesgo por inundación se encuentra en un nivel medio debido a que la población se expone a construir sus viviendas en las fajas marginales del río Chinchipe.

Villalobos (2021) en su investigación titulada “Estudio de inundación por efectos del desborde del río Amojú, aplicando los modelos matemáticos HEC-RAS Y HEC-HMS en la ciudad de Jaén, distrito de Jaén provincia de Jaén, departamento de Cajamarca”, tuvo como objetivo realizar un modelamiento hidrológico e hidráulico para calcular los caudales máximos y predecir ante una eventual inundación las partes afectadas y crítica mediante el software HEC-HMS y HEC-RAS. La metodología en la presente investigación es aplicada y diseño descriptivo, con enfoque cuantitativo. Como resultado se obtuvo del modelamiento hidrológico mediante el software HEC-HMS un caudal máximo de 210 m³/s para un periodo de retorno de 500 años y del modelamiento hidráulico a través del software HEC-RAS las partes menos afectadas de la ciudad (Norte y Sur); y la inundación crítica se da para el periodo de retorno de 500 años. Concluyendo que el modelamiento hidráulico a través del software HEC-RAS se obtuvo mejor resultados para predecir una eventual inundación por el desborde del río Amojú.

García (2019) en su investigación titulada “Modelado hidráulico del río Amojú aplicando el Software Iber para zonificar áreas inundables en la ciudad de Jaén” tuvo como objetivo identificar y zonificar áreas de mayor probabilidad de inundación para prevenir desastres y pérdidas económicas o de vidas humanas ante cualquier fenómeno pluviométrico. La metodología es de tipo básica y diseño descriptivo, con enfoque cuantitativo, en la cual realizó el análisis de los parámetros geomorfológicos de la cuenca del río Amojú, la modelación hidrológica de una máxima avenida en la cuenca del río y el modelado hidráulico aplicando el Software Iber. Se obtuvo como resultado que las áreas con mayor inundación se encuentran en la zona 1 con 0.929 hectáreas, la zona 2 con 0.482 hectáreas, la zona 5 con 0.504 hectáreas y la zona 6 con 2.633 hectáreas (sector San Camilo); para un tiempo de retorno de 140 años y un riesgo admisible de 25%. Concluyendo que el modelado hidráulico permite identificar las zonas con mayor probabilidad de inundación frente a lluvias de gran intensidad.

1.3.Objetivos

1.3.1. Objetivo General

- Estimar el nivel de riesgo de inundación fluvial mediante el Algoritmo Random Forest en el Sector Magllanal, Jaén, Cajamarca, 2022.

1.3.2. Objetivos Específicos

- Identificar el nivel del peligro como consecuencia de inundación fluvial en el Sector Magllanal, Jaén.
- Analizar el nivel vulnerabilidad como consecuencia de inundación fluvial en el Sector Magllanal, Jaén.
- Establecer el nivel de riesgo de inundación fluvial en la población del Sector Magllanal, Jaén.
- Modelar el comportamiento del nivel de riesgo de inundación fluvial mediante el algoritmo Random Forest en el Sector Magllanal, Jaén.

II. MATERIALES Y MÉTODOS

2.1. Objeto de estudio

La presente investigación se realizó para estimar el nivel de riesgo de inundación fluvial en el Sector Magllanal. El trabajo de investigación tiene como objetivo hacer estimaciones de la inundación fluvial con fines de tomar predicciones para diagnosticar, modelar y zonificar las áreas que se encuentran vulnerables ante el peligro de inundación fluvial con el fin de proponer medidas tanto estructurales como no estructurales para reducir el riesgo.

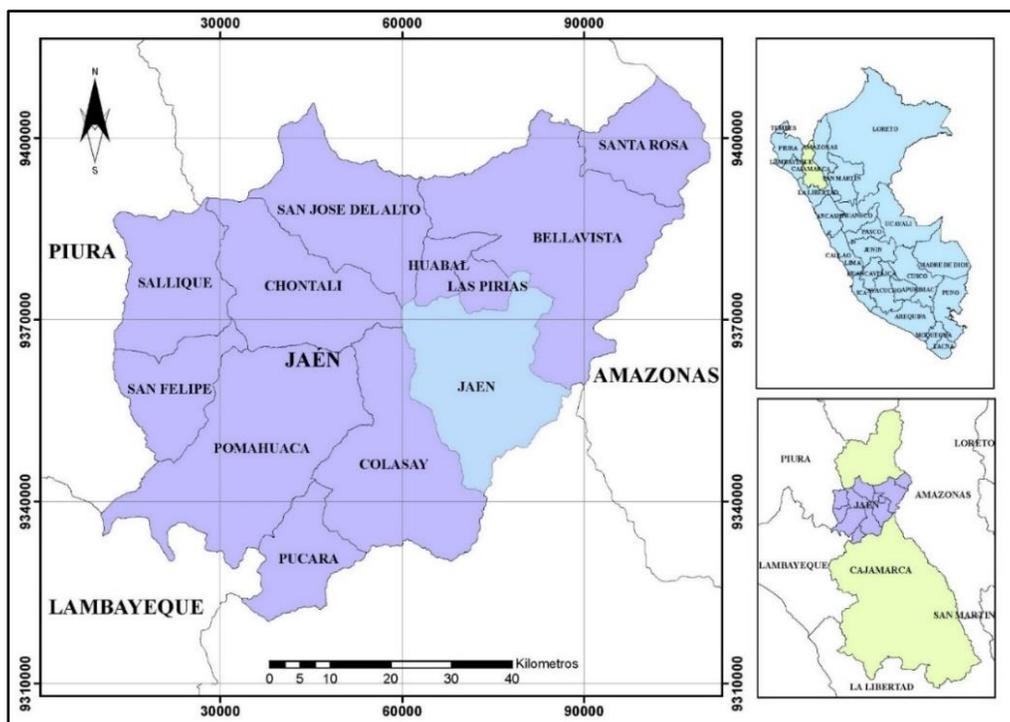
2.2. Características generales del área de estudio

2.2.1. Ubicación geográfica

EL área estudiada está ubicada en la provincia de Jaén, distrito de Jaén, departamento de Cajamarca (ver figura 1).

Figura 1

Ubicación geográfica del área estudiada



Fuente: Adaptado de Carta Geológica Huacal - Ingemmet.

2.3.Población, muestra y muestreo

2.3.1. Población

La población considerada fue el Sector Magllanal la cual está conformada por un área de 31.03 ha. Según coordenadas UTM zona 17 M se encuentra al Este 741526.91 m E y Norte 9368931.37 m S. Este Sector es afectado por inundaciones a causa de la activación de la zanora Magllanal que en consecuencia afecta a la infraestructura de viviendas y servicios públicos en general. Ver en el anexo 39 el mapa satelital del Sector Magllanal -Jaén-Cajamarca.

2.3.2. Muestra.

La muestra que se utilizó estuvo constituida por las áreas próximas a la zanora Magllanal, comprendidas por las calles José Balta, Dos de Mayo, Túpac Amaru, Mariátegui, Lolo Fernández, Manuel Gonzales Prada y Pedro Ruiz; con un área de 3.95 ha, se opta por estas calles porque se ubican dentro del área con mayor riesgo ante inundación por activación de la zanora Magllanal, en la ciudad de Jaén, Cajamarca. El mapa satelital de la delimitación del área en estudio se observa en el anexo 39.

2.3.3. Muestreo

El tipo de muestreo que se llevará a cabo en este trabajo de investigación es no probabilístico por conveniencia.

2.4.Variables de estudio

- Nivel de Riesgo

2.4.1. Operacionalización de variables.

Tabla 1*Operacionalización de variables*

Variable	Dimensiones	Indicadores	Unidad	Técnica de Recolección de datos	Instrumento de Recolección de datos
	Peligro	Parámetros geomorfológicos	m ²		
		Tirante de agua	m	Observación	ficha de registro de datos
Nivel de Riesgo	Vulnerabilidad social	Exposición		Observación	Fichas técnicas
	Vulnerabilidad económica	Fragilidad	Adimensional	Análisis de documentos	Software Excel
	Vulnerabilidad ambiental	Resiliencia			
	Riesgo	Parámetros de evaluación	-	Minería de datos	Ficha de registro de datos

Fuente: Elaboración propia.

2.5.Tipo de investigación

- Por su tipo

El tipo de investigación es básica porque busca recoger, recopilar información sobre la realidad de las cosas para enriquecer el conocimiento científico, además a futuro esta investigación conduce alguna aplicación valiosa (Suárez, 2012). Por ende, la presente investigación es básica, debido a que recolecta datos en campo de forma directa para estimar el nivel de riesgo de inundación fluvial del Sector Magllanal de la ciudad de Jaén mediante el algoritmo Random Forest; asimismo los resultados obtenidos se difundirán para que tanto la población como las autoridades y entidades públicas puedan tomar medidas preventivas para mitigar el riesgo de inundación fluvial en el Sector; con la finalidad de proteger y dar seguridad a la población y sus medios de vida.

- Según su diseño

Es no experimental debido a que “en este diseño no hay estímulos o condiciones experimentales a las que se sometan las variables de estudio, los sujetos del estudio son evaluados en su contexto natural sin alterar ninguna situación” (Gonzales y Covinos,2021, p.82). Por ello, la investigación es de naturaleza no experimental, ya que no se realizó cambios deliberados en la variable estudiada, además se utilizó la información resultante de los parámetros de evaluación para estimar el nivel de riesgo de inundación fluvial en el área estudiada del Sector Magllanal de la ciudad de Jaén.

- Según su alcance

Es transversal -descriptivo debido a que a través de cualquier fenómeno o problema que se someta a un análisis o investigación, miden y recolectan datos para recabar información y así luego poder describir lo que se investiga (Hernández y Mendoza, 2018). Por esta razón, la presente

investigación es descriptiva; debido a que, a través de la observación, análisis de documentos se describió la variable en estudio.

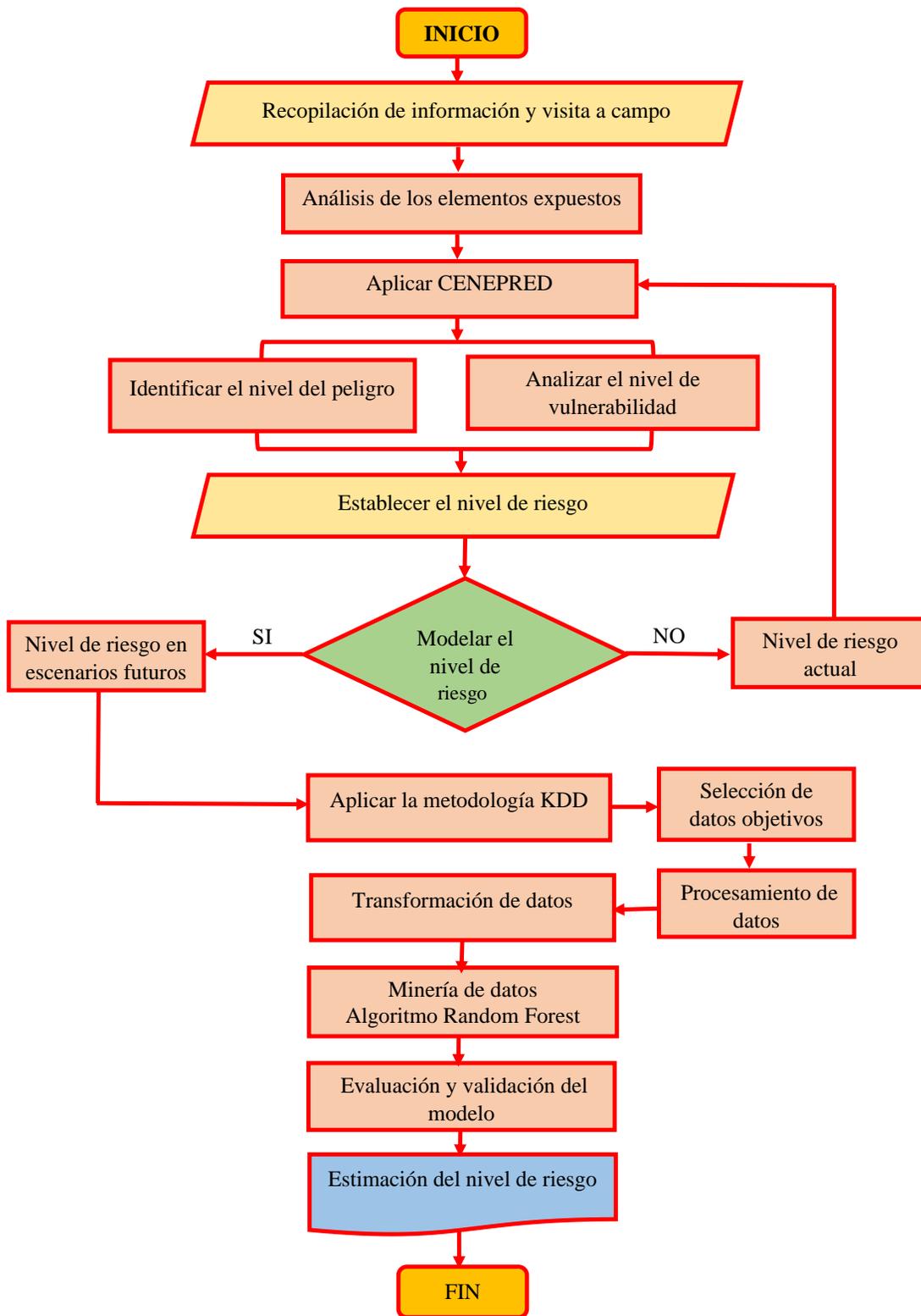
- **Según su enfoque**

Es de enfoque cuantitativo cuando la variable en estudio asume valores o unidades de medida (Gonzales y Covinos,2021). Por ello, la presente investigación según su enfoque es cuantitativa, porque la variable empleada permitió estimar el nivel de riesgo de inundación fluvial mediante el algoritmo Random Forest, a través de técnicas cuantitativas como son la recolección de información mediante fichas técnicas, levantamiento topográfico, estudio de mecánica de suelos, minería de datos y diferentes fuentes de investigación.

2.6.Procedimiento

Figura 2

Metodología general para la estimación del nivel de riesgo



2.6.1. Recopilación de información y visita a campo

2.6.1.1.Reconocimiento del área en estudio.

Se realizó recorridos por las calles del área estudiada, abarcando la calle Dos de Mayo, José Carlos Mariátegui, Pedro Ruiz Gallo, Manuel Gonzales Prada, Túpac Amaru, Lolo Fernández y la calle José Balta con el fin de obtener datos relevantes y observaciones directas del área estudiada. Luego, se hizo coordinaciones con el presidente de la Junta Vecinal del Sector Magllanal, en estas coordinaciones se obtuvo el apoyo para realizar esta investigación (ver figura 3); donde se conoció con más detalle la situación problemática, las familias afectadas y el mal estado de las estructuras, entre más información para la realización de esta investigación.

Figura 3

Coordinación con el presidente de la Junta Vecinal del Sector Magllanal



Nota. En la figura 3, se puede visualizar las coordinaciones con el presidente de la Junta Vecinal del Sector Magllanal, se obtuvo un conocimiento más profundo a partir de la información adicional obtenida durante esta investigación.

2.6.1.2. Análisis de los elementos expuestos.

a) Situación actual del cauce de la zanora

En la situación actual del cauce de la zanora se observó la presencia de vegetación y desechos sólidos que han sido arrojados por la población del sector, además se observó el desagüe de las viviendas aledañas la cual discurren por el cauce de la zanora tanto en su margen izquierda y derecha, contaminando sus aguas y poniendo en riesgo la salud de la población (ver figura 4).

Figura 4

Vista en planta de la zanora Magllanal



Nota. En la figura 4, se observa la situación actual del cauce de la zanora Magllanal, contaminada sus aguas poniendo en riesgo a la salud de la población.

b) Situación actual de muros de contención

La situación actual de los muros de contención de 2m de altura se encontraron agrietados, erosionados, y con desfogues de tuberías de desagüe, exponiendo a la población del Sector Magllanal, visualizar en la (figura 5).

Figura 5

Muros de contención de 2 metros de altura



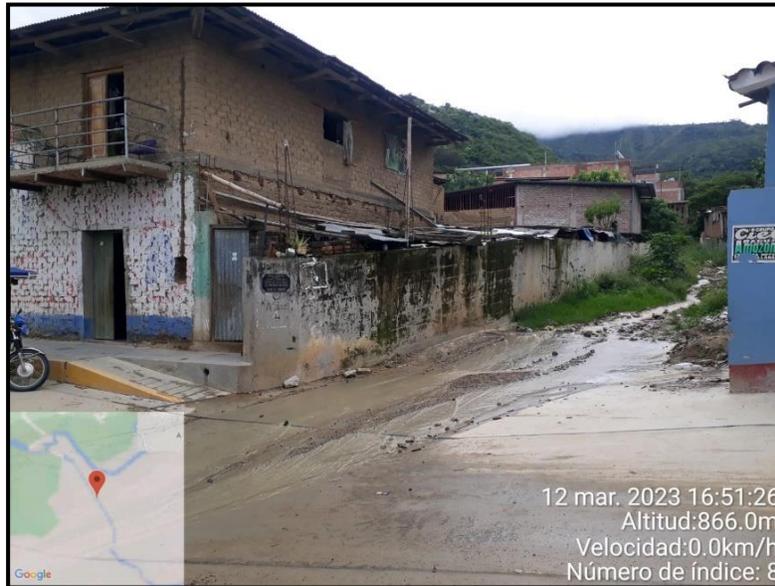
Nota. En la figura 5, se observa los muros de contención en mal estado, exponiendo a la población del Sector Magllanal.

c) Situación actual de badenes

La situación actual de badenes con un tirante hidráulico inferior a 0.50 metros ubicados entre las calles Dos de Mayo, Lolo Fernández y Pedro Ruiz Gallo, los cuales se encontraron obstruidos por desmonte; desviando el flujo del agua por las calles y dificultando el tránsito vehicular como peatonal (ver figura 6).

Figura 6

Badenes con tirante hidráulico menor a 0.50 metros



Nota. En la figura 6, se observa la situación actual de los badenes en mal estado de conservación, dificultando el tránsito vehicular y peatonal.

d) Situación actual del puente peatonal

La situación actual del puente peatonal, se encuentra en mal estado por falta de mantenimiento, presentando grietas, barandas oxidadas, maderas rotas, y fierros expuestos, perjudicando y exponiendo a la población del Sector Magllanal (ver figura 7).

Figura 7

Puente peatonal sin mantenimiento



Nota. En la figura 7, se observa la situación actual del puente peatonal en mal estado por falta de mantenimiento, exponiendo a la población del área estudiada.

e) Construcciones de vivienda en la ribera de la zanora Magllanal

Se observó columnas y muros de las viviendas construidas dentro de la zanora Magllanal y sobre los muros de contención respectivamente, exponiendo a las viviendas y a la población a sufrir daños por inundación fluvial (ver figura 8).

Figura 8

Viviendas construidas sobre muros de contención



Nota. En la figura 8, se observa construcciones de viviendas sobre los muros de contención de la zanora exponiendo a la población a sufrir daños por inundación fluvial.

2.6.2. Identificar el nivel de peligro

Para identificar el nivel de peligro por inundación fluvial en el Sector Magllanal, se aplicó la metodología CENEPRED tomando en consideración el parámetro de evaluación y la susceptibilidad en base a factores condicionantes y desencadenantes en el área de estudio; en los cuales se utilizó el Proceso de Análisis Jerárquico para realizar la ponderación de los parámetros de acuerdo a los criterios de la escala desarrollada por Saaty (ver anexo 2).

2.6.2.1. Parámetro de evaluación.

Para identificar y definir el parámetro de evaluación del peligro, se realizó un reconocimiento detallado del área estudiada. Donde se observó marcas en los muros contención de la zanora, las cuales indicaban la altura máxima alcanzada por el flujo del agua; además se notó la presencia de residuos sólidos dentro del cauce, lo que aumenta el riesgo de obstrucción del cauce, y por consiguiente aumenta la altura del flujo del agua generando desbordamientos.

Basándonos en las evidencias identificadas en campo nos permitieron considerar la altura de tirante de agua como el principal parámetro de evaluación (ver figura 9). Para su ponderación se tomó en cuenta el estudio hidrológico de la microcuenca de la zanora Magllanal (ver anexo 3) y la recopilación de la información en campo del parámetro de evaluación del peligro por tirante de agua.

Figura 9

Altura del tirante de agua



Nota. La figura 9, muestra la altura del tirante de agua menor a 0.50 metros y una altura máxima 2 metros.

De acuerdo a los resultados obtenidos en campo y al estudio hidrológico de la microcuenca (ver anexo 3), se construyó los descriptores del parámetro de evaluación (altura del tirante de agua), considerando los descriptores desde el más crítico (1.40m – 2m) al menos influyente (0 - 0.50m). Para la ponderación del parámetro altura de tirante de agua se realizó el Proceso de Análisis Jerárquico (ver anexo 4).

2.6.2.2. Susceptibilidad del área estudiada.

Para determinar la susceptibilidad del área estudiada, se evaluó los factores condicionantes como la pendiente, geomorfología, litología y tipo de suelo; y teniendo como factor desencadenante la precipitación. Posteriormente, se realizaron los mapas para cada factor (ver anexo 39).

2.6.2.2.1. Factores condicionantes.

Para evaluar el comportamiento de los factores que condicionan el área estudiada ante el riesgo de inundación fluvial provocado por el incremento de la altura del tirante de agua debido a lluvias intensas, se analizaron diversos aspectos de los factores como la pendiente, las unidades geomorfológicas, la litología y el tipo de suelo; aspectos que fueron recopilados del geoportal EVAR (Recursos para Evaluación de Riesgos). Asimismo, se consideró los descriptores de los factores condicionantes, desde lo que más predomina a lo que menos predomina en el área de estudio; así como también desde el más crítico al menos influyente. Para la ponderación del parámetro de factores condicionantes se realizó el Proceso de Análisis Jerárquico (ver anexo 5).

- Pendiente

Se obtuvo los descriptores de la pendiente del área en estudio empleando el software ArcGIS, en el cual se consideró a los descriptores desde el más crítico con una pendiente mayor a 50° y al menos influyente con una pendiente menor a 10°. Para la ponderación del parámetro pendiente se realizó el Proceso de Análisis Jerárquico (ver anexo 5).

- Unidades geomorfológicas

Se obtuvo los descriptores de las unidades geomorfológicas que predominan en el área de estudio, a base de información obtenida como: vertiente o piedemonte aluvial, montaña en roca volcánico-sedimentaria, colina en roca sedimentaria, llanura o planicie inundables y montaña en

roca sedimentaria. Para la ponderación del parámetro unidades geomorfológicas se realizó el Proceso de Análisis Jerárquico (ver anexo 5).

- **Unidades litológicas**

Se obtuvo los descriptores de las unidades litológicas que predominan en el área de estudio, a base de información obtenida del geportal de EVAR como: Rocas de grano fino, Depósito aluvial, fluvial-lacustre, roca caliza, fluvial-aluvial. Para la ponderación del parámetro unidades litológicas se realizó el Proceso de Análisis Jerárquico (ver anexo 5).

- **Tipo de suelo**

Para la evaluación de este parámetro se realizó el estudio de suelo del área estudiada (ver anexo 6); con el fin de determinar el tipo de suelo se realizó una calicata de 2m de profundidad para la obtención de muestras, luego se realizó los ensayos respectivos en el laboratorio; obteniendo como resultado un suelo Arenas arcillosas (SC), Gravas arcillosas (GC) y Gravas mal graduadas (GP); donde dicho resultado sirvió de base para la ponderación de los descriptores.

Los descriptores del tipo de suelo se basó de información recopilada del geportal EVAR, donde se obtuvo un tipo de suelo gravoso, arenoso, limoso y arcilloso (Qh-al), calizas arenosas (Ki-chu), conglomerado (Qp-ta), areniscas cuarzosas (Ki-g) y limolitas (N-be); información que fue corroborada con la carta geológica de Jaén 12-f que brinda el INGEMMET (ver anexo 39) estando dentro de las formaciones geológicas de Jaén, en la cual lo que más predomina en nuestra área de estudio según el estudio de suelos realizado es el tipo de suelo gravoso, arenoso, limoso, arcilloso. Para la ponderación del parámetro tipo de suelo se realizó el Proceso de Análisis Jerárquico (ver anexo 5).

2.6.2.2. Factores desencadenantes.

Debido a antecedentes de las intensas lluvias en el área de estudio se consideró a la precipitación como el causante ante una inundación fluvial.

Para el factor precipitación se recopiló información de la Estación de Jaén desde el año 2013 hasta el mes de Marzo del año 2023 ,considerándose una precipitación máxima diaria de 101 mm/día que ocurrió en el mes de Noviembre del año 2021 (ver anexo 7); comparando con los umbrales de precipitación del SENAMHI de la Estación de Jaén que va desde el año 1964 hasta el año 2019, se determina que la precipitación máxima diaria se encuentra en la categoría de “extremadamente lluvioso” con umbrales de precipitación de $RR > 44.8\text{mm}$ y un percentil de $RR/\text{día} > 99\text{p}$.

En base a ello, se tomó en cuenta los descriptores de la precipitación, desde el más crítico (extremadamente lluvioso) hasta el menos crítico (normal). Para la ponderación del parámetro precipitación se realizó el Proceso de Análisis Jerárquico (ver anexo 5).

2.6.3. Analizar el nivel de vulnerabilidad

Para el análisis de la vulnerabilidad por inundación, se recaudó información mediante encuestas para evaluar los factores de exposición, fragilidad y resiliencia en 68 lotes (ver anexo 9), abarcando las dimensiones social, económica y ambiental.

En el área de estudio se realizó 68 encuestas de acuerdo a la infraestructura de las viviendas, la red de agua potable, la red de desagüe y vías de comunicación. Las encuestas se realizaron en las calles Dos de mayo, José Carlos Mariátegui, Pedro Ruiz Gallo, Manuel Gonzales Prada, Túpac Amaru, Lolo Fernández y la Calle José Balta (ver figura 10). Luego se utilizó el Proceso de Análisis Jerárquico para realizar la ponderación de los parámetros en la matriz Saaty. Finalmente, se realizó el mapa del nivel de vulnerabilidad (ver anexo 39).

Figura 10

Realización de encuestas



Nota. En la figura 10, se observa la realización de encuestas a la población del área estudiada para el análisis de la vulnerabilidad.

2.6.3.1. Vulnerabilidad en la dimensión social.

Se analizó a la población expuesta frente a una inundación en el área de estudio con respecto a los factores de exposición, fragilidad y resiliencia. Para la ponderación del parámetro análisis de la vulnerabilidad en la dimensión social se realizó el Proceso de Análisis Jerárquico (ver anexo 11).

2.6.3.1.1. Análisis de la exposición social.

Se consideró como parámetros de evaluación al número de personas que viven en cada lote, principalmente en los lotes más vulnerables ante el desborde de la zanora Magllanal; así como, el tiempo de exposición de cada lote por inundación. Para la ponderación de los parámetros del análisis de vulnerabilidad de la exposición social se realizó el Proceso de Análisis Jerárquico (ver anexo 12).

2.6.3.1.2. Análisis de la fragilidad social.

Se consideró como parámetros de evaluación al grupo etario, porcentaje de personas afectadas según su edad ante inundación y el acceso a los servicios básicos que tiene cada lote.

Para la ponderación de los parámetros del análisis de la fragilidad social se realizó el Proceso de Análisis Jerárquico (ver anexo 13).

2.6.3.1.3. Análisis de la resiliencia social.

Se consideró como parámetros de evaluación a la capacitación en Gestión de Riesgo de Desastres (GRD), el conocimiento sobre desastres por inundación, la actitud ante una inundación y las campañas de publicidad relacionadas con temas de GRD. Para la ponderación de los parámetros del análisis de la resiliencia social se realizó el Proceso de Análisis Jerárquico (ver anexo14).

2.6.3.2.Vulnerabilidad en la dimensión económica.

Se analizó a las viviendas e infraestructura; así como, los componentes básicos que cuenta cada familia con respecto a los factores de exposición, fragilidad y resiliencia. Para la ponderación del parámetro de vulnerabilidad en la dimensión económica se realizó el Proceso de Análisis Jerárquico (ver anexo 15).

2.6.3.2.1. Análisis de la exposición económica.

Se consideró como parámetros de evaluación a la cercanía de la zanora respecto a los lotes y la exposición de los servicios básicos de agua potable y desagüe frente a inundaciones. Para la ponderación de los parámetros del análisis de la exposición económica se realizó el Proceso de Análisis Jerárquico (ver anexo 16).

2.6.3.2.2. Análisis de la fragilidad económica.

Se consideró como parámetros de evaluación el material principal en la estructura de las viviendas, el estado de conservación de la misma y su antigüedad de construcción. Para la ponderación de los parámetros del análisis de la fragilidad económica se realizó aplicando el Proceso de Análisis Jerárquico (ver anexo 17).

2.6.3.2.3. Análisis de la resiliencia económica.

Se consideró como parámetros de evaluación a la infraestructura con respecto al drenaje, ocupación (acceso a un empleo), ingreso promedio familiar y capacitación en temas económicos. Para la ponderación de los parámetros del análisis de la resiliencia económica se realizó aplicando el Proceso de Análisis Jerárquico (ver anexo 18).

2.6.3.3. Vulnerabilidad en la dimensión ambiental.

Se analizó los servicios básicos en relación con la zanora Magllanal y la disposición de residuos sólidos de cada familia con respecto a los factores de exposición, fragilidad y resiliencia.

Para la ponderación del parámetro de vulnerabilidad en la dimensión ambiental se realizó el Proceso de Análisis Jerárquico (ver anexo 19).

2.6.3.3.1. Análisis de la exposición ambiental.

Se consideró como parámetro de evaluación a la cercanía de los componentes del sistema de alcantarillado sanitario a la zanora. Para la ponderación del parámetro de la vulnerabilidad de la exposición ambiental se realizó el Proceso de Análisis Jerárquico (ver anexo 20).

2.6.3.3.2. Análisis de la fragilidad ambiental.

Se consideró como parámetros de evaluación a la disposición de residuos sólidos y a la disposición de excretas. Para la ponderación del parámetro de la vulnerabilidad de la fragilidad ambiental se realizó el Proceso de Análisis Jerárquico (ver anexo 21).

2.6.3.3.3. *Análisis de la resiliencia ambiental.*

Se consideró como parámetros de evaluación al manejo de residuos sólidos y al conocimiento en temas ambientales. Para la ponderación del parámetro de la vulnerabilidad de la resiliencia ambiental se realizó aplicando el Proceso de Análisis Jerárquico (ver anexo 22).

2.6.4. *Establecer el nivel del riesgo*

Para establecer el nivel de riesgo primero se identificó los elementos expuestos a sufrir daños y pérdidas ante el fenómeno natural de inundación fluvial; luego se identificó el peligro al que está expuesto el Sector Magllanal mediante el parámetro de evaluación altura de tirante de agua y la susceptibilidad del área estudiada; seguidamente se analizó los componentes de la vulnerabilidad en sus dimensiones social, económica y ambiental, a base de los factores de exposición, fragilidad y resiliencia que tiene el área de estudio a ser afectada; y finalmente se precedió a calcular el nivel de riesgo por lote mediante la metodología CENEPRED, la cual resultó de conjugar el peligro con la vulnerabilidad. Asimismo, se realizó la matriz de riesgo para establecer los rangos para cada uno de los niveles de riesgo, como se visualiza en el (anexo 30).

2.6.5. *Modelar el nivel de riesgo*

Para modelar el nivel de riesgo de inundación fluvial en el Sector Magllanal, se siguió un enfoque metodológico que involucró la identificación del nivel de peligro, el análisis de la vulnerabilidad y la determinación del nivel de riesgo utilizando la metodología CENEPRED. Esta metodología incorporó el Proceso de Análisis Jerárquico para ponderar las variables relevantes. Estas variables fueron procesadas mediante el modelo matemático Random Forest, permitiendo estimar el nivel de riesgo de inundación fluvial para escenarios futuros.

2.6.5.1. Base de datos a partir de las variables que influyen en el nivel de riesgo.

Tomando en consideración la información recopilada del área de estudio y de diversas tesis relacionadas con el nivel de riesgo de inundación fluvial recopiladas de múltiples repositorios del Perú desde el año 2017 hasta el 2023, se ha construido una base de datos compuesta por 505 instancias. Esta base de datos incluye 23 variables de vulnerabilidad y 6 variables de peligro. La finalidad fue de generar una base de datos que permitan validar a través de un modelo matemático como lo es el algoritmo Random Forest, para estimar el nivel de riesgo de inundación fluvial del área estudiada del Sector Magllanal.

Las variables contempladas del peligro fueron: Altura de tirante de agua, pendiente, unidades geomorfológicas, litología, tipo de suelo y precipitación.

Las variables de la vulnerabilidad fueron: Número de personas que viven en cada lote, el tiempo de exposición de cada lote por inundación, grupo etario, % de población afectada, acceso a servicios básicos, capacitación en temas de GRD, el conocimiento sobre desastres por inundación, la actitud ante una inundación, campañas de difusión relacionadas con temas de GRD, cercanía de la zanora respecto a los lotes, servicio básico de agua potable y desagüe expuesto ante una inundación, material principal en la estructura de las viviendas, el estado de conservación de las viviendas, la antigüedad de construcción de las viviendas, la infraestructura presenta drenaje, ocupación (acceso a un empleo), ingreso promedio familiar, capacitación en temas económicos, la cercanía de los componentes del sistema de alcantarillado sanitario a la zanora, disposición de residuos sólidos, disposición de excretas, manejo de residuos sólidos y conocimiento en temas ambientales. Las variables según su simbología y rango se muestran en la Tabla 2.

Tabla 2*Simbología y rango de las variables recolectadas del peligro y la vulnerabilidad*

GRUPO	CONCEPTO	VARIABLES	SIMBOLOGÍA	R
Peligro	Parámetro de evaluación	Altura de tirante de agua	ATA	[0.032-0.415]
	Factores condicionantes	Pendiente	P	[0.062-0.416]
		Unidades geomorfológicas	UG	[0.050-0.395]
		Litología	L	[0.037-0.513]
		Tipo de suelo	TP	[0.045-0.439]
	Factores desencadenantes	Precipitación	PP	[0.033-0.430]
Vulnerabilidad social	Exposición	Número de personas que viven en cada lote	NPL	[0.052-0.433]
		Tiempo de exposición de su vivienda por inundación	TEV	[0.049-0.441]
	Fragilidad	Grupo etario	GE	[0.043-0.424]
		Porcentaje de población afectada	PPA	[0.043-0.424]
		Acceso a servicios básicos	ASB	[0.043-0.424]
	Resiliencia	Capacitación en temas de GRD	CTGRD	[0.035-0.379]
		Conocimiento sobre desastres por inundación	CDI	[0.036-0.474]
		Actitud frente a una inundación	AFI	[0.036-0.474]
	Campañas de publicidad	CPGRD	[0.036-0.474]	

		relacionadas con temas de GRD		
	Exposición	Cercanía de la zanora a nivel de lotes	CZNL	[0.048-0.462]
		Servicio básico de agua potable y desagüe expuesto ante una inundación	SBAD	[0.048-0.462]
		Material principal en la estructura de las viviendas	MPEV	[0.037-0.468]
Vulnerabilidad económica	Fragilidad	Estado de conservación de las viviendas	ECV	[0.044-0.391]
		Antigüedad de construcción de las viviendas	ACV	[0.044-0.391]
		La infraestructura cuenta con drenaje	ICD	[0.048-0.462]
	Resiliencia	Ocupación	O	[0.048-0.462]
		Ingreso promedio familiar	IPF	[0.048-0.462]
		Capacitación en temas económicos	CTE	[0.048-0.462]
	Exposición	Cercanía de los componentes del sistema de alcantarillado sanitario a la zanora	CCAZ	[0.048-0.462]
Vulnerabilidad ambiental		Disposición de residuos sólidos	DRS	[0.048-0.462]
	Fragilidad	Disposición de excretas	DE	[0.048-0.462]
	Resiliencia	Manejo de residuos sólidos	MRS	[0.048-0.462]

	Conocimiento en temas ambientales	CTA	[0.048-0.462]
Riesgo	Nivel	N	BAJO, MEDIO, ALTO, MUY ALTO

Nota. La tabla 2, muestra las variables que forman parte del peligro y la vulnerabilidad en la base de datos, su simbología y rango.

Luego se realizó la estadística descriptiva de la base de datos empleando el lenguaje de programación Python, mediante la librería de pandas con el comando `df.describe()`, obteniéndose el promedio, desviación estándar, el valor mínimo, 25%, 50%, 75% y el valor máximo de cada variable. Estos datos son fundamentales para comprender la distribución y variabilidad de cada una de las variables analizadas (ver tabla 9).

Seguidamente se realizó, el gráfico de bigotes, la cual proporciona una representación visual de la distribución de los datos de las variables seleccionadas (ver figura 29).

2.6.5.2. Modelación del nivel de riesgo mediante la metodología KDD.

Este proceso de modelación conocido como “descubrimiento de conocimiento en bases de datos” (Knowledge Discovery Databases (KDD)), incluye la preparación de datos, el análisis estadístico, la aplicación de algoritmos de minería de datos y la evaluación e interpretación de los resultados. Como resultado de este procedimiento, se logra el descubrimiento de conocimiento el nivel de riesgo de inundación fluvial en el Sector Magllanal, Jaén.

2.6.5.2.1. Etapa de selección.

Identificada la necesidad de determinar el nivel de riesgo de inundación fluvial en el Sector Magllanal, Jaén, mediante el uso de minería de datos utilizando el algoritmo Random Forest, se procedió a la creación de un conjunto de datos objetivo sobre la cual se llevó a cabo el proceso de descubrimiento. En esta fase, se empleó el software Weka para la selección de atributos utilizando

el algoritmo "Correlation Ranking" como se observa en la tabla 3, el cual asigna ponderaciones a cada variable en función de su relevancia con respecto a la variable de salida (nivel de riesgo).

Tabla 3

Rankin de atributos según el algoritmo Random Forest

VARIABLES	SIMBOLOGÍA	RANKING
Pendiente	P	0.53297
Precipitación	PP	0.52254
Altura del tirante de agua	ATA	0.51701
Litología	L	0.26414
Cercanía de los componentes del sistema de alcantarillado sanitario a la zanora	CCAZ	0.19736
Tipo de suelo	TP	0.17886
Tiempo de exposición de su vivienda por inundación	TEV	0.17108
Cercanía de la zanora a su vivienda	CZNL	0.17027
Campañas de publicidad sobre temas de GRD	CPGRD	0.16041
Capacitación en temas de GRD	CTGRD	0.1425
La infraestructura cuenta con drenaje	ICD	0.14051
Capacitación en temas de conservación ambiental	CTA	0.13731
Capacitación en temas económicos	CTE	0.11549
Geomorfología	UG	0.11106
Grupo etario	GE	0.10797
Manejo de residuos sólidos	MRS	0.10704
Disposición de residuos sólidos	DRS	0.08685

Conocimiento sobre algún desastre por inundación	CDI	0.07731
Tipo de disposición de excretas	TDE	0.07454
Porcentaje de población afectada	PPA	0.07031
Número de personas que viven en cada lote	NPL	0.06313
Servicio básico de agua potable y desagüe expuesto ante una inundación	SBAD	0.06087
Material de construcción que se utilizó en la vivienda	MPEV	0.05892
Cuál es su ingreso promedio familiar	IPF	0.05666
Con que tipo de empleo cuenta	O	0.05625
Actitud frente a una inundación	AFI	0.05237
Antigüedad de construcción de la vivienda	ACV	0.02964
Acceso a los servicios básicos	ASB	0.02873
Estado de conservación de la vivienda	ECV	0.00928
Nivel	N	BAJO, MEDIO, ALTO, MUY ALTO

Nota. En la presente tabla 3, se muestra la información de las variables de entrada con respecto de la variable de salida (Nivel de riesgo).

Después de llevar a cabo el proceso de clasificación de atributos, se optó por seleccionar aquellas variables que presentaban un valor superior a 0.14. De esta manera, se trabajó con un

conjunto de 10 variables de entrada (Ver Tabla 4), las cuales proporcionan mayor información para la variable de salida (nivel de riesgo).

Tabla 4

Variables que aportan mayor información a la variable de salida

VARIABLES	SIMBOLOGÍA	RANKING
Pendiente	P	0.53297
Precipitación	PP	0.52254
Altura del tirante de agua	ATA	0.51701
Litología	L	0.26414
Cercanía de los componentes del sistema de alcantarillado sanitario a la zanora	CCAZ	0.19736
Tipo de suelo	TP	0.17886
Tiempo de exposición de su vivienda por inundación	TEV	0.17108
Cercanía de la zanora a su vivienda	CZNL	0.17027
Campañas de publicidad sobre temas de GRD	CPGRD	0.16041
Capacitación en temas de GRD	CTGRD	0.1425

Nota. La tabla 4, muestra las 10 variables seleccionadas que proporcionan mayor información a la variable de salida.

2.6.5.2.2. Etapa de preprocesamiento.

Se procedió a realizar la limpieza de los datos con el propósito de mejorar su calidad, eliminando aquellos datos anómalos, abordando dos problemáticas comunes: la presencia de valores faltantes en algunos registros y la existencia de datos duplicados. Para realizar esta tarea, se utilizó Python a través de la interfaz de Jupyter, empleando las librerías de análisis de datos como Pandas y NumPy (ver figuras 11 y 12).

Figura 11

Importación de la base de datos

```
df=pd.read_csv('Base de datos.csv')
df.sample(15)
```

	Pendiente	Litología	Tipo de suelo	Precipitación	Altura del tirante	Tiempo de exposición de su vivienda por inundación	Capacitación en temas de GRD	Campañas de publicidad sobre temas de GRD	Cercanía de la zanora a su vivienda	Cercanía de los componentes del sistema de alcantarillado sanitario a la zanora	Nivel de riesgo
54	0.416	0.513	0.150	0.430	0.415	0.098770	0.105075	0.128735	0.101798	0.101798	ALTO
15	0.416	0.513	0.281	0.430	0.415	0.259648	0.105075	0.079582	0.160352	0.101798	ALTO
24	0.161	0.244	0.150	0.179	0.173	0.440802	0.379267	0.473565	0.462063	0.462063	MEDIO
42	0.161	0.513	0.281	0.179	0.173	0.259648	0.379267	0.473565	0.462063	0.228105	MEDIO
9	0.416	0.513	0.281	0.430	0.415	0.151569	0.035359	0.282092	0.101800	0.101800	ALTO
56	0.416	0.513	0.150	0.430	0.415	0.440802	0.264740	0.079582	0.101798	0.101798	ALTO
26	0.416	0.513	0.150	0.430	0.415	0.098770	0.264740	0.282092	0.101798	0.101798	ALTO
13	0.416	0.513	0.281	0.430	0.415	0.259648	0.215559	0.282092	0.101798	0.160352	ALTO
36	0.262	0.513	0.085	0.288	0.307	0.440802	0.379267	0.473565	0.462063	0.228105	MEDIO
1	0.161	0.244	0.085	0.430	0.415	0.151569	0.105075	0.036027	0.101798	0.228105	ALTO
32	0.262	0.513	0.281	0.288	0.307	0.440802	0.379267	0.473565	0.462063	0.462063	MEDIO

Nota. La figura 11, muestra el proceso de importación de la base de datos en las librerías de Python (Pandas y Numpy).

Figura 12

Limpieza de datos

```
LIMPIEZA DE LOS DATOS

b=df.describe()
def plt_boxplot(df, ft):
    df.boxplot(column=[ft])
    plt.show()
plt_boxplot(df, 'P')

def outliers(df, ft):
    Q1=df[ft].quantile(0.25)
    Q3=df[ft].quantile(0.75)
    IQR=Q3-Q1

    lower_bound=Q1-1.5*IQR
    upper_bound=Q3+1.5*IQR

    ls=df.index[(df[ft]<lower_bound)|(df[ft]>upper_bound) ]
    return ls

index_list=[]
for L in ['Pendiente','Litología','Tipo de suelo', 'Precipitación', 'Altura del tirante', 'Tiempo de exposición de su vivienda por inundación',
          'Capacitación en temas de GRD', 'Campañas de publicidad sobre temas de GRD', 'Cercanía de la zanora a su vivienda',
          'Cercanía de los componentes del sistema de alcantarillado sanitario a la zanora'
]:
    index_list.extend(outliers(df,L))

index_list

def remove(df,ls):
    ls=sorted(set(ls))
    df=df.drop(ls)
    return df
df_cleaned=remove(df, index_list)
df_cleaned.shape
```

Nota. La figura 12, muestra el proceso de eliminación y limpieza de los datos duplicados.

Después de ejecutar el comando "datos.drop_duplicates()" de la biblioteca Pandas, se obtuvieron un total de 505 instancias de los datos recopilados, los cuales fueron guardados en un archivo denominado 'Datos-ok' utilizando el comando df.reset_index().to_csv('Datos-ok.CSV', header=True, index=False). El archivo resultante 'Datos-ok' se puede observar en la Figura 13.

Figura 13

Datos de las variables seleccionadas

	A	B	C	D	E	F	G	H	I	J	K
1	Pendie	Litologi	Tipo de	Precipi	Altura	Tiempc	Capacit	Campa	Cercan	Cercan	Nivel de riesgo
2	0.416	0.513	0.281	0.43	0.415	0.25965	0.10507	0.12873	0.2281	0.2281	ALTO
3	0.416	0.513	0.085	0.43	0.415	0.15157	0.10507	0.03603	0.1018	0.2281	ALTO
4	0.161	0.513	0.281	0.179	0.173	0.25965	0.37927	0.47356	0.46206	0.16035	MEDIO
5	0.262	0.513	0.085	0.288	0.307	0.25965	0.26474	0.47356	0.46206	0.1018	MEDIO
6	0.262	0.513	0.281	0.288	0.307	0.25965	0.37927	0.47356	0.46206	0.2281	MEDIO
7	0.161	0.513	0.281	0.179	0.173	0.4408	0.10507	0.47356	0.46206	0.1018	MEDIO
8	0.262	0.513	0.085	0.288	0.307	0.4408	0.26474	0.47356	0.46206	0.46206	MEDIO
9	0.262	0.513	0.281	0.288	0.307	0.4408	0.37927	0.47356	0.46206	0.2281	MEDIO
10	0.416	0.513	0.15	0.43	0.415	0.15157	0.26474	0.12873	0.1018	0.2281	ALTO
11	0.416	0.513	0.281	0.43	0.415	0.15157	0.03536	0.28209	0.46206	0.2281	ALTO
12	0.416	0.513	0.085	0.43	0.415	0.15157	0.10507	0.07958	0.46206	0.1018	ALTO
13	0.416	0.513	0.085	0.43	0.415	0.15157	0.21556	0.28209	0.2281	0.1018	ALTO
14	0.416	0.513	0.15	0.43	0.415	0.15157	0.21556	0.07958	0.1018	0.2281	ALTO
15	0.416	0.513	0.281	0.43	0.415	0.25965	0.21556	0.28209	0.1018	0.16035	ALTO
16	0.416	0.513	0.085	0.43	0.415	0.15157	0.10507	0.12873	0.16035	0.16035	ALTO
17	0.416	0.513	0.281	0.43	0.415	0.25965	0.10507	0.07958	0.16035	0.1018	ALTO
18	0.161	0.513	0.085	0.179	0.173	0.4408	0.37927	0.47356	0.46206	0.2281	MEDIO
19	0.416	0.513	0.15	0.43	0.415	0.25965	0.10507	0.07958	0.16035	0.16035	ALTO
20	0.416	0.513	0.15	0.43	0.415	0.15157	0.37927	0.07958	0.46206	0.1018	ALTO

Nota. La figura 13, muestra los datos limpiados tras haber sido procesados con Python.

2.6.5.3.3. Etapa de transformación.

En esta etapa se buscó identificar características relevantes para representar los datos en relación al nivel de riesgo de inundación fluvial en el Sector Magllanal, Jaén, en la cual, se utilizaron métodos de reducción de dimensiones o transformación. Estos métodos se emplearon con el propósito de disminuir el número efectivo de variables consideradas o encontrar representaciones invariables de los datos, con el objetivo de crear cinco grupos o modelos distintos.

Tabla 5*Variables seleccionadas para los cinco grupos*

GRUPO	VARIABLES
MODELO I (LY_1)	-Pendiente
	-Precipitación
	-Altura del tirante
	-Litología
	-Cercanía de los componentes del sistema de alcantarillado sanitario a la zanora
	-Tipo de suelo
	-Tiempo de exposición de su vivienda por inundación
	-Cercanía de la zanora a su vivienda
	-Campañas de publicidad sobre temas de GRD
	-Capacitación en temas de GRD
MODELO II (LY_2)	-Pendiente
	-Precipitación
	-Altura del tirante
	-Litología
	-Cercanía de los componentes del sistema de alcantarillado sanitario a la zanora
	-Tipo de suelo
	-Tiempo de exposición de su vivienda por inundación
MODELO III (LY_3)	-Cercanía de la zanora a su vivienda
	-Campañas de publicidad sobre temas de GRD
	-Capacitación en temas de GRD
	-Precipitación
	-Altura del tirante

	-Altura del tirante
	-Litología
MODELO IV	-Cercanía de la zanora a su vivienda
(LY_4)	-Tipo de suelo
	-Campañas de publicidad sobre temas de GRD
	-Capacitación en temas de GRD
	-Cercanía de los componentes del sistema de alcantarillado sanitario a la zanora
	-Tipo de suelo
	-Tiempo de exposición de su vivienda por inundación
MODELO V	-Cercanía de la zanora a su vivienda
(LY_5)	-Pendiente
	-Precipitación

Nota. La tabla 5, muestra las variables seleccionadas para los cinco grupos creados (modelo I (LY1), modelo II (LY_2), modelo III (LY_3), modelo IV (LY_4) y modelo V (LY_5)).

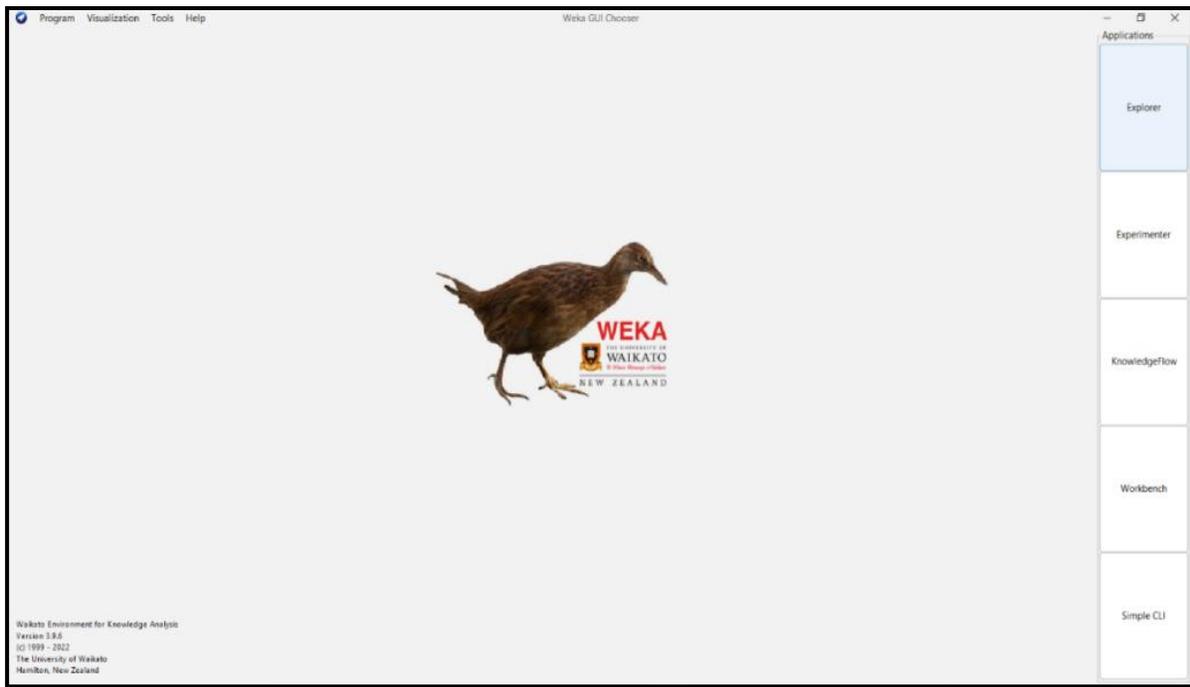
2.6.5.3.4. Etapa minería de datos.

En esta etapa se empleó el software WEKA versión 3.9.6 (ver figura 14). Para la ejecución del estudio se utilizó la interfaz gráfica del programa, ingresando a la pestaña "Explorer" y "Classify". Además, se empleó la técnica de validación cruzada k-folds con un valor k igual a 10; esta técnica ha sido seleccionada debido a que proporciona una estimación más concreta de la precisión del modelo, al reducir la varianza a través de la división del conjunto de datos en múltiples pliegues. Este enfoque mejora las particiones aleatorias de los datos, lo que contribuye a obtener resultados finales más robustos y confiables.

- **Preprocess:** En esta pestaña se encuentran disponibles una variedad de filtros que se aplican al conjunto de datos. Estas herramientas fueron fundamentales para preparar y organizar los datos antes de realiza los modelos.
- **Classify:** Esta función nos permitió aplicar algoritmos de clasificación y regresión al conjunto de datos (ver figura 15)
- **Cluster:** Esta función nos permitió aplicar algoritmos de agrupamiento al conjunto de datos.
- **Associate:** Esta función nos permitió aplicar el algoritmo para obtener reglas de asociación.
- **Select Attributes:** Esta función nos permitió seleccionar las variables o atributos para cada modelo.
- **Visualize:** Esta función nos permitió visualizar pares de atributos, lo cual es fundamental para comprender las relaciones y patrones entre diferentes variables en los datos.
- **Experimenter:** Este proceso fue fundamental para aplicar diversos algoritmos de minería de datos en diferentes conjuntos de datos.
- **Knowledge Flow:** Esta funcionalidad nos permitió realizar las mismas acciones que la herramienta "Explorer", pero de manera visual y gráfica.

Figura 14

Interfaz de Weka

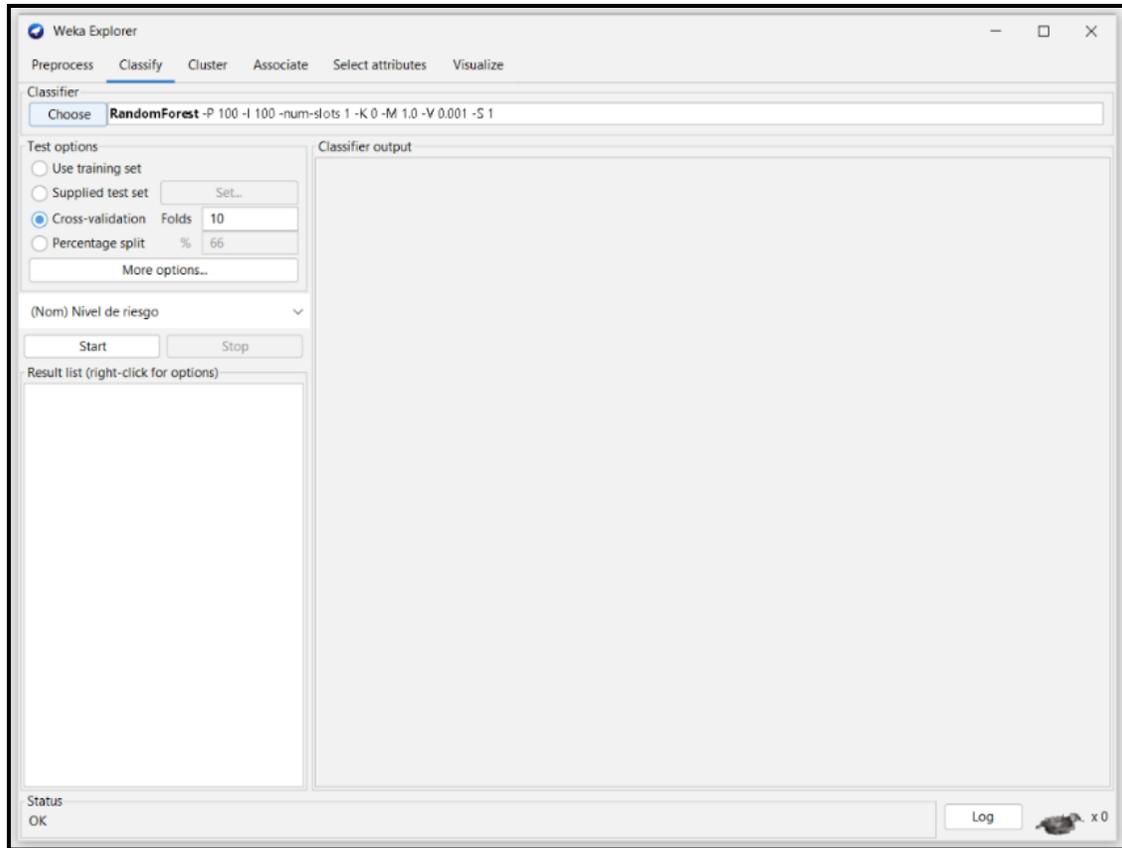


Nota. En la figura 14, se muestra la interfaz gráfica de Weka 3.9.6.

De todas las funcionalidades mencionadas, se empleó la función "Explorer", seleccionando la herramienta "Classify", como se muestra en la figura 15.

Figura 15

Interfaz de la opción Classify



Nota. La figura 15, muestra la interfaz de Classify de la herramienta de Explorer.

- **Método de evaluación**

En esta fase, se empleó la técnica de Validación Cruzada (Cross-Validation) junto con un conjunto de datos de prueba proporcionado (Supplied test set) para llevar a cabo el entrenamiento del modelo y la predicción del nivel de riesgo.

- **Supplied test set:** Nos permitió cargar un conjunto de datos diferentes a los de entrenamiento, con los cuales se realizó la estimación del nivel de riesgo de inundación.

- **Cross-validation:** La evaluación con validación cruzada consistió en dividir las instancias en varias carpetas según el parámetro Folds. En cada evaluación, se usaron las instancias de una carpeta como datos de prueba, mientras que el resto se utilizaron como datos de entrenamiento para construir el modelo. Los errores calculados representarán el promedio de todas las ejecuciones.

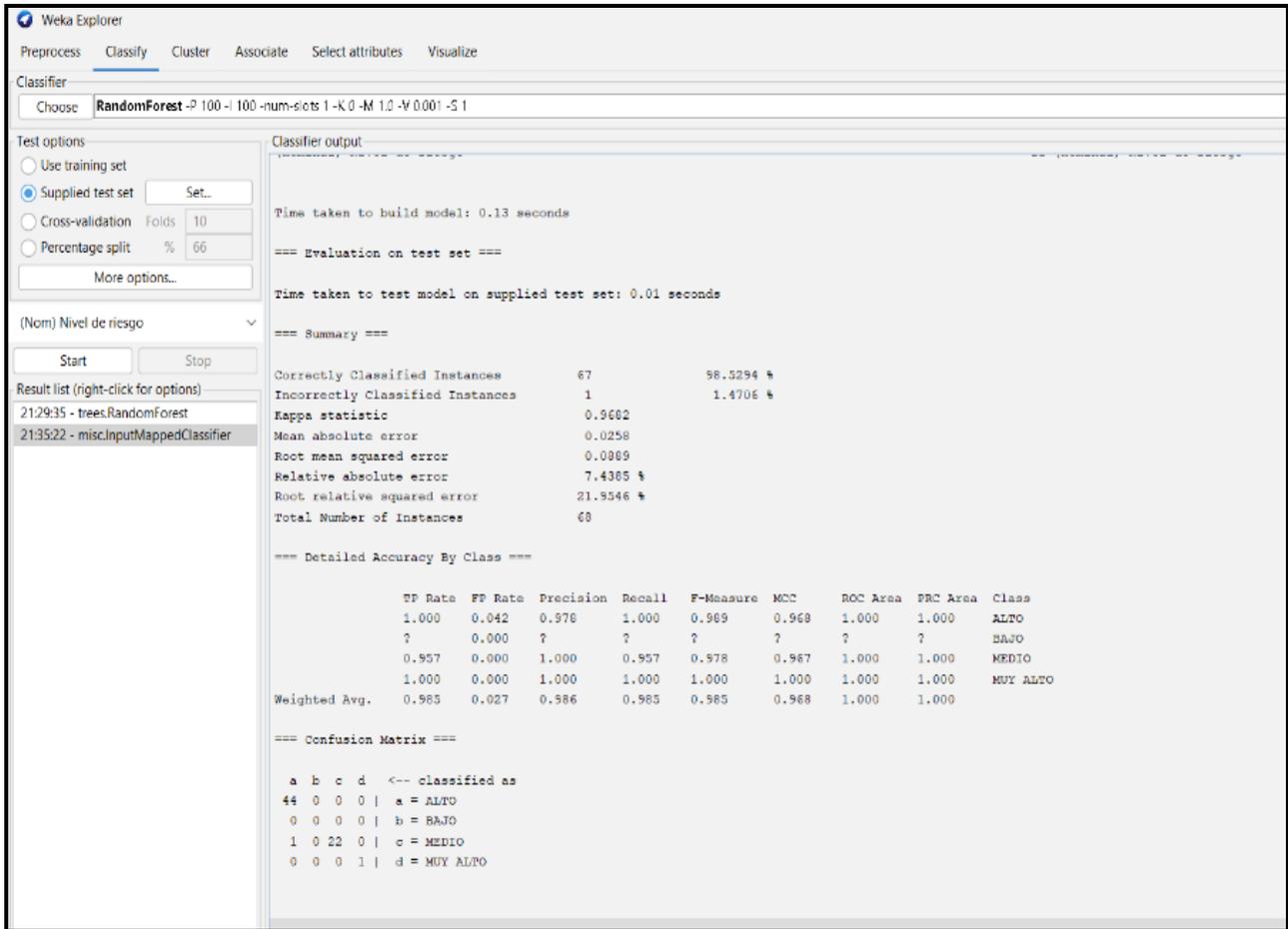
El algoritmo Random Forest fue aplicado a los diferentes modelos.

- **Random Forest**

Es un algoritmo de aprendizaje robusto que utiliza múltiples árboles de decisión aleatorios y combina sus predicciones mediante promedios. Su funcionamiento consta de dos fases: la fase de construcción, en la que se construyen múltiples árboles de decisión utilizando diferentes instancias de entrenamiento para garantizar un modelo preciso y menos sobre ajustado, y la fase de predicción, en la que se promedian los resultados de cada árbol para realizar una predicción final. El rendimiento se evalúa en función de varios criterios, variaciones en los valores de los parámetros del clasificador, sensibilidad al ruido y cambios en el tamaño del conjunto de entrenamiento. Para medir la eficacia, se comparan con los árboles de clasificación analizando el rendimiento en relación con éstos. Las predicciones, clasificación y validación cruzada de cada modelo con el algoritmo Random Forest se puede observar en las figuras 16, 17, 18, 19 y 20 respectivamente.

Figura 16

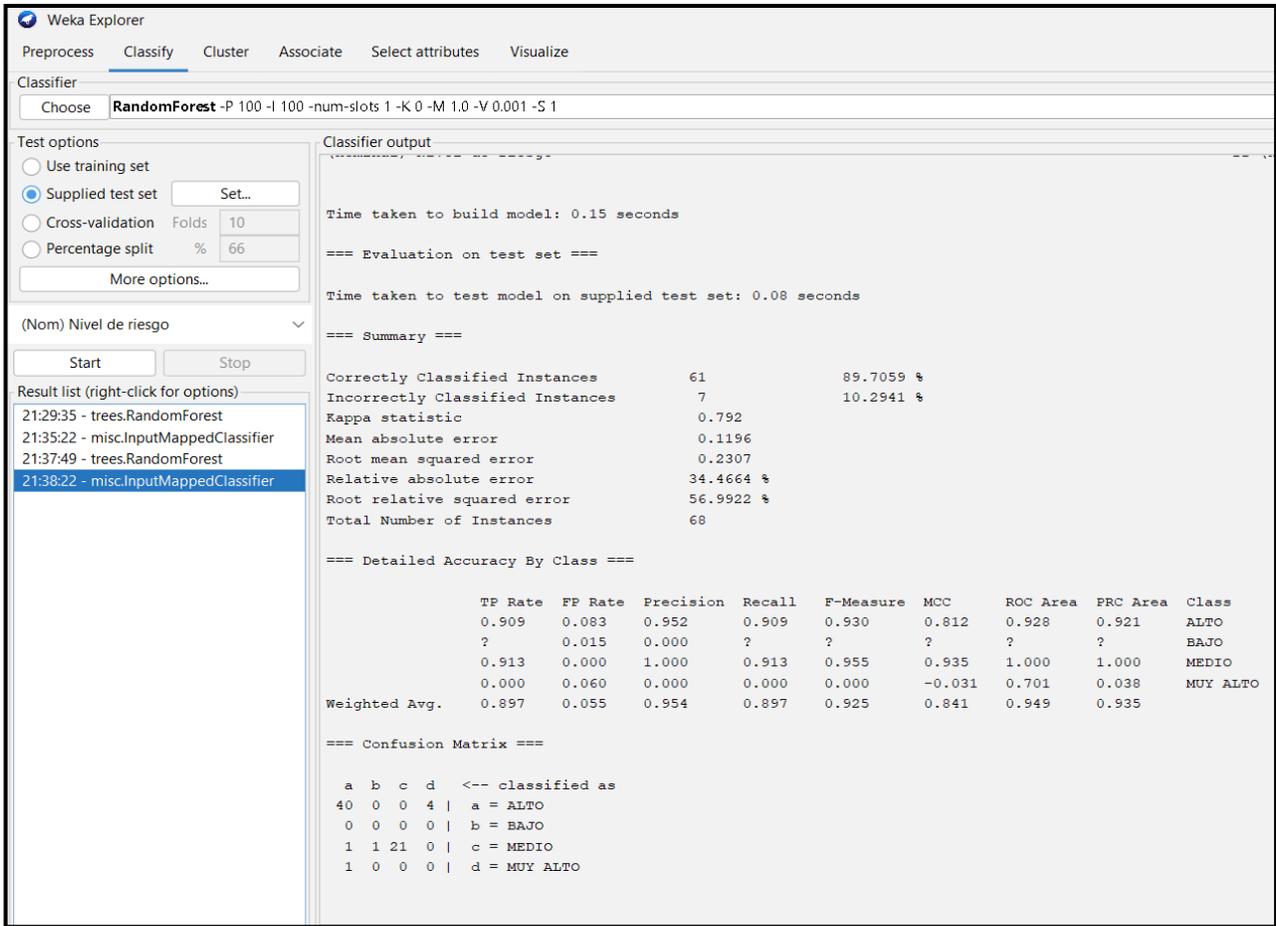
Modelo I (LY_1)



Nota. La figura 16, muestra la clasificación del modelo I (LY_1) con el algoritmo Random Forest y validación cruzada.

Figura 17

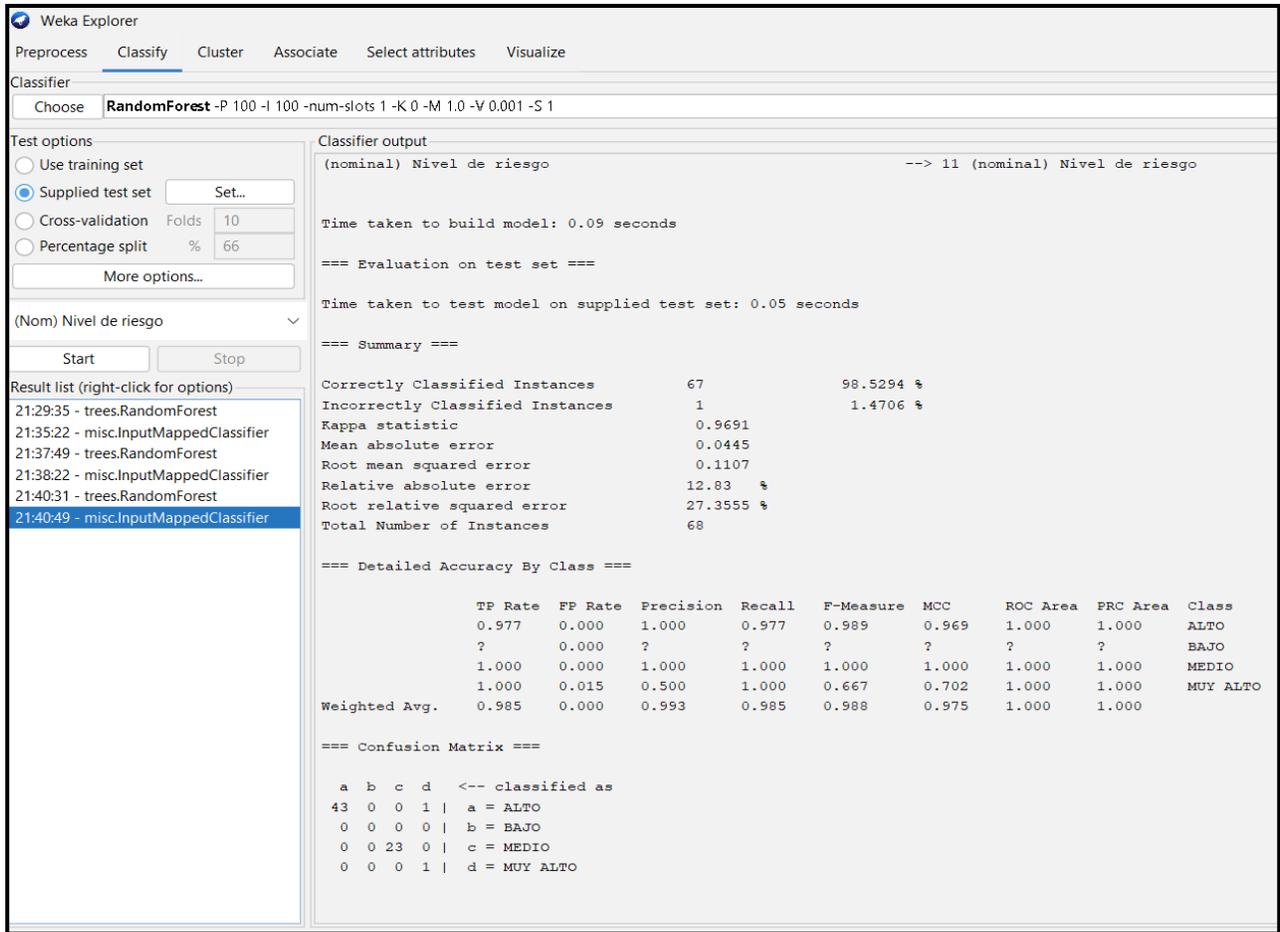
Modelo II (LY_2)



Nota. La figura 17, muestra la clasificación del modelo II (LY_2) con el algoritmo Random Forest y validación cruzada.

Figura 18

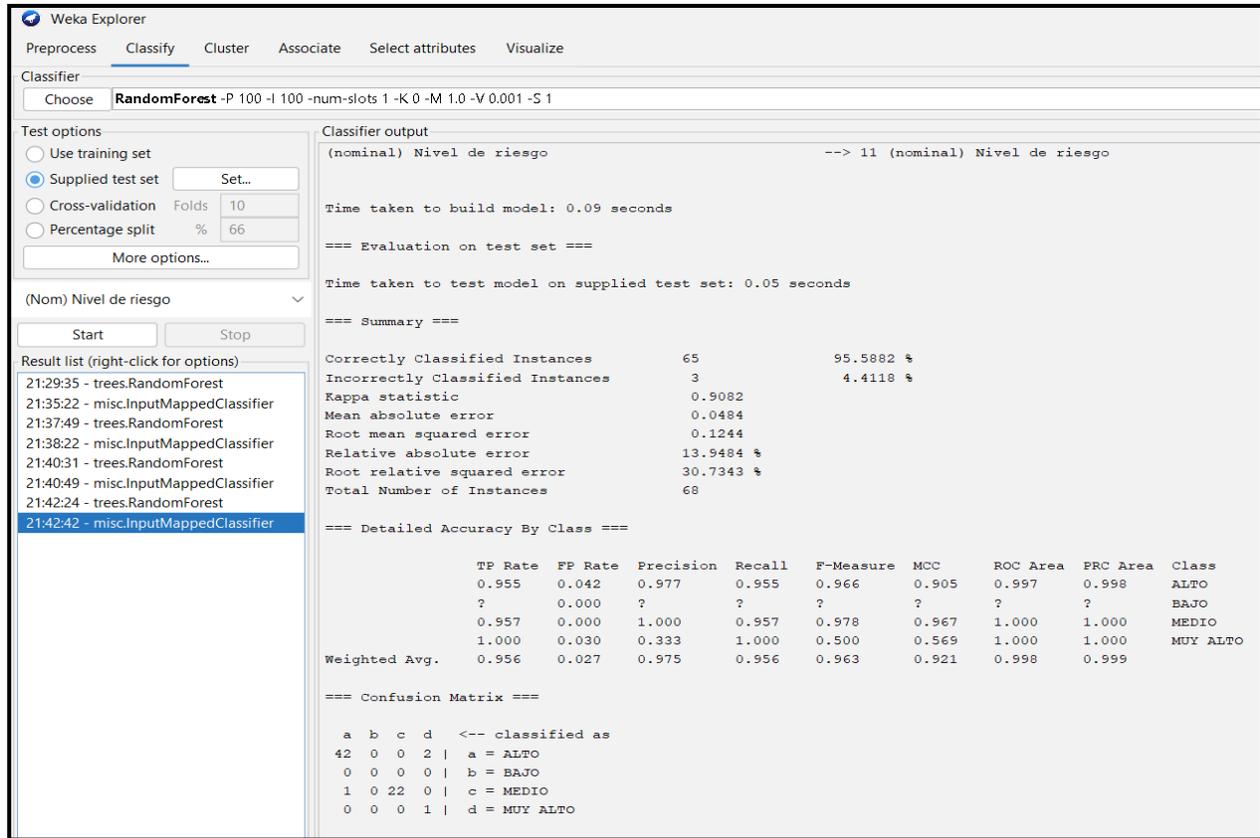
Modelo III (LY_3)



Nota. La figura 18, muestra la clasificación del modelo III (LY_3) con el algoritmo Random Forest y validación cruzada.

Figura 19

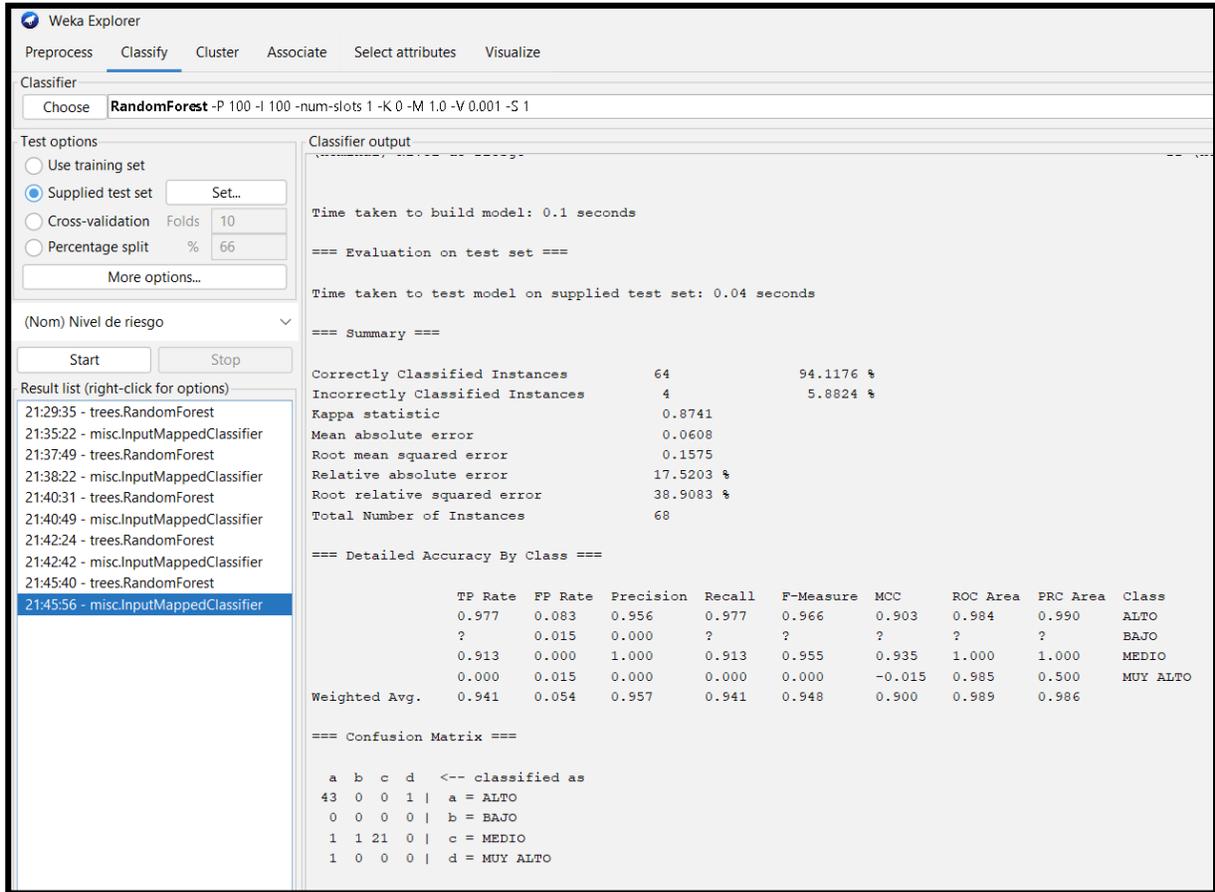
Modelo IV (LY_4)



Nota. La figura 19, muestra la clasificación del modelo IV (LY_4) con el algoritmo Random Forest y validación cruzada.

Figura 20

Modelo V (LY_5)



Nota. La figura 20, muestra la clasificación del modelo V (LY_5) con el algoritmo Random Forest y validación cruzada.

- **Resumen de modelos obtenidos**

Después de construir los modelos utilizando el algoritmo Random Forest en Weka, obtenemos un resumen de cada modelo con la siguiente estructura:

- Número de instancias clasificadas correctamente.
- Numero de instancias clasificadas incorrectamente.
- El estadístico Kappa es un índice que mide el grado de acuerdo entre varios

expertos, comparándolo con el nivel de acuerdo que podría ocurrir por azar.

- Error cuadrático medio: Es un estimador que mide la media de los errores al cuadrado, que es la diferencia entre el estimador y lo estimado al cuadrado.
- Error absoluto: La diferencia entre el valor obtenido y su valor verdadero.
- Error absoluto relativo: La fórmula $[(Xi - Xt)/Xt] \times 100$, donde $Xi - Xt$ es el error absoluto y Xt es el valor real.
- Número total de instancias.

2.6.5.3.5. Etapa de evaluación y validación.

- Predicción de los modelos

Después de haber desarrollado los cinco modelos utilizando la base de datos, se seleccionaron tres modelos para la validación debido a su mayor porcentaje de predicción durante el entrenamiento, siendo el modelo LY_1, LY_3 y LY_4. En el proceso de validación, se emplearon los datos calculados del nivel de riesgo de inundación fluvial de 68 viviendas ubicadas en la población del Sector Magllanal, Jaén, mediante la aplicación de la Metodología Cenepred.

- Evaluación de modelos matriz de confusión

Luego de haber estimado el nivel de riesgo utilizando el algoritmo Random Forest a través de los cinco modelos, se procedió a evaluarlos mediante la matriz de confusión. Esta evaluación permitió identificar cuál de los modelos presentó el menor error en la predicción. A continuación, se exponen los principales criterios de predicción como se muestra en la figura 21.

Figura 21

Matriz de confusión de la clasificación

		CLASE PREDICHA	
		Positivo	Negativo
Actual	Positivo	Positivo Verdadero (PV)	Negativo Falso (NF)
	Negativo	Positivo Falso (PV)	Negativo Verdadero (NV)

- **Instancias Correctamente Clasificadas (ICC):** Se refiere a la proporción de instancias correctamente clasificadas, obtenida al dividir el número de instancias correctamente clasificadas entre el total de instancias.

$$ICC = \frac{PV + NV}{PV + NP + PF + NF}$$

- **Instancias Incorrectamente Clasificadas (IIC):** Se define como la proporción de instancias clasificadas de manera incorrecta, calculada al dividir el número de instancias incorrectamente clasificadas entre el total de instancias.

$$IIC = \frac{PF + NF}{PV + NV + PF + NF}$$

- **Tasa TP:** Es la proporción de casos positivos que fueron correctamente identificados por el modelo en relación con todos los casos positivos.

$$Tasa TP = \frac{PV}{PV + NF}$$

- **Tasa FP:** Esta métrica representa la proporción de casos negativos que fueron incorrectamente clasificados como positivos por el modelo.

$$Tasa FP = \frac{PF}{PF + NV}$$

- **Precisión:** Representa la proporción de predicciones positivas correctas sobre el total de instancias que el modelo ha clasificado como positivas.

$$\textit{Precisión} = \frac{PV}{PV + PF}$$

- **F-Mesure:** Se calcula como la media armónica de la precisión y el recall, proporcionando una evaluación más completa del rendimiento del modelo

$$\textit{FMesure} = 2 * \frac{\textit{Precisión} * \textit{Recall}}{\textit{Precisión} + \textit{Recall}}$$

- **Accuracy:** Representa la proporción de predicciones correctas realizadas por el modelo sobre el total de instancias evaluadas.

$$\textit{Accuracy} = \frac{PV + NV}{PV + NV + PF + NF}$$

- **Sensibilidad:** Representa la proporción de predicciones positivas correctas sobre el total de instancias que realmente pertenecen a esa clase.

$$\textit{Sensibilidad} = \frac{PV}{PV + NF}$$

- **Especificidad:** Representa la proporción de predicciones negativas correctas sobre el total de instancias que realmente pertenecen a la clase negativa.

$$\textit{Especificidad} = \frac{NV}{NV + PF}$$

III. RESULTADOS

3.1. Identificación del nivel de peligro como consecuencia de inundación fluvial

Para identificar el nivel de peligro de inundación fluvial en el área de estudio del Sector Magllanal, se aplicó la metodología CENEPRED, considerando el parámetro de evaluación y la susceptibilidad basada en factores condicionantes y desencadenantes en el área de estudio. Se utilizó el Proceso de Análisis Jerárquico para ponderar los parámetros en la matriz Saaty.

En la tabla 6, se presenta el parámetro de evaluación y sus descriptores del área estudiada en el Sector Magllanal. Asimismo; en las tablas 7 y 8, se muestra los factores condicionantes y desencadenantes respectivamente.

Tabla 6

Parámetro de evaluación

PARÁMETRO	DESCRIPTORES				
Altura del tirante de agua (m)	1.40 – 2 m	1.10 - 1.4 m	0.80 - 1.10m	0.5 - 0.80 m	0 - 0.50 m

Nota. La tabla 6, muestra el parámetro de evaluación (altura del tirante de agua en (m)) y sus descriptores desde el más crítico (1.40m – 2m) al menos influyente (0 - 0.50m).

Tabla 7*Factores condicionantes*

PARÁMETRO		DESCRIPTORES			
Pendiente	>50°	31° - 50°	21°-30°	11°-20°	0°-10°
Geomorfología	Vertiente o piedemonte aluvial	Montaña en roca volcano-sedimentaria	Colina en roca sedimentaria	Llanura o planicie inundable	Montaña en roca sedimentaria
litología	Rocas de grano fino	Depósito aluvial	Fluvial-lacustre	Roca caliza	Fluvial-aluvial
Tipo de suelo	Gravoso, arenoso, limoso y arcilloso	Calizas arenosas	Conglomerado	Areniscas cuarzosas	Limolitas

Nota. La tabla 7, presenta los parámetros de los factores condicionantes, ordenados según su importancia en función de las características del área estudiada, desde el más crítico hasta el menos influyente.

Tabla 8*Factores desencadenantes*

PARÁMETRO	DESCRIPTORES				
Precipitación	RR/día>99p Extremadamente lluvioso	95p<RR/día≤99p Muy lluvioso	90p<RR/día≤95p (Lluvioso)	75p<RR/día≤90p Moderadamente lluvioso	RR/día≤75p (Normal)

Nota. La tabla 8, muestra los parámetros de los factores desencadenantes (precipitación), y sus descriptores desde lo más crítico (extremadamente lluvioso) a lo menos crítico (normal).

3.1.1. Niveles de peligro

En la siguiente figura 22, se muestra como resultado los niveles de peligro y sus respectivos rangos obtenidos a través del proceso de Análisis Jerárquico.

Figura 22

Niveles de peligro del área estudiada

NIVEL DE PELIGRO	RANGO
MUY ALTO	$0.294 \leq P \leq 0.422$
ALTO	$0.172 \leq P < 0.294$
MEDIO	$0.075 \leq P < 0.172$
BAJO	$0.038 \leq P < 0.075$

Nota. La figura 22, muestra los niveles de peligro muy alto, alto medio y bajo con sus respectivos rangos del área estudiada.

3.1.2. Estratificación de los niveles de peligro

En el anexo 31, se muestra la estratificación del peligro del área de estudio de Sector Magllanal, con sus respectivos rangos obtenidos a través de utilizar el proceso de Análisis Jerárquico.

3.1.3. Mapa de zonificación del nivel de peligro del área estudiada

Mapa de zonificación encontrado a partir del procesamiento de la información en el software ArcGIS. El formato del mapa se observa en el anexo 39.

3.2. Análisis del nivel de vulnerabilidad como consecuencia de inundación fluvial

Para el análisis de la vulnerabilidad a inundaciones, se recopiló información sobre los elementos expuestos (infraestructura de viviendas, red de agua potable, red de desagüe y vías de comunicación) mediante encuestas que abarcaban factores de exposición, fragilidad y resiliencia en 68 lotes, considerando las dimensiones social, económica y ambiental. Asimismo, se utilizó el Proceso de Análisis Jerárquico para realizar la ponderación de los parámetros en la matriz Saaty.

En las figuras 23, 24 y 25, se muestra el nivel de vulnerabilidad para la dimensión social, económica y ambiental respectivamente.

Figura 23

Nivel de vulnerabilidad social

NIVEL DE VULNERABILIDAD DIMENSIÓN SOCIAL	RANGO
MUY ALTO	$0.459 \leq V \leq 0.656$
ALTO	$0.325 \leq V < 0.459$
MEDIO	$0.255 \leq V < 0.325$
BAJO	$0.197 \leq V < 0.255$

Nota. La figura 23, muestra el nivel de vulnerabilidad muy alto, alto, medio y bajo para la dimensión social con sus respectivos rangos.

Figura 24

Nivel de vulnerabilidad económica

NIVEL DE VULNERABILIDAD DIMENSIÓN ECONÓMICA	RANGO
MUY ALTO	$0.317 \leq V \leq 0.552$
ALTO	$0.242 \leq V < 0.337$
MEDIO	$0.166 \leq V < 0.242$
BAJO	$0.106 \leq V < 0.166$

Nota. La figura 24, muestra el nivel de vulnerabilidad muy alto, alto, medio y bajo para la dimensión económica con sus respectivos rangos.

Figura 25

Nivel de vulnerabilidad ambiental

NIVEL DE VULNERABILIDAD DIMENSIÓN AMBIENTAL	RANGO
MUY ALTO	$0.228 \leq V \leq 0.462$
ALTO	$0.160 \leq V < 0.228$
MEDIO	$0.102 \leq V < 0.160$
BAJO	$0.048 \leq V < 0.102$

Nota. La figura 25, muestra el nivel de vulnerabilidad muy alto, alto, medio y bajo para la dimensión ambiental con sus respectivos rangos.

3.2.1. Niveles de la vulnerabilidad

En la siguiente figura 26, se muestra los niveles de la vulnerabilidad del área de estudio del Sector Magllanal; así como, sus respectivos rangos obtenidos a través del Proceso de Análisis Jerárquico.

Figura 26

Niveles de la vulnerabilidad del área de estudio del Sector Magllanal

NIVEL DE LA VULNERABILIDAD	RANGO
MUY ALTO	$0.335 \leq V \leq 0.557$
ALTO	$0.242 \leq V < 0.335$
MEDIO	$0.174 \leq V < 0.242$
BAJO	$0.117 \leq V < 0.174$

Nota. La figura 26, muestra los niveles de vulnerabilidad muy alto, alto, medio y bajo con sus respectivos rangos.

3.2.2. Estratificación de los niveles de vulnerabilidad

En el anexo 32, se muestra la estratificación de la vulnerabilidad del área de estudio del Sector Magllanal; con sus respectivos rangos obtenidos a través de utilizar el Proceso de Análisis Jerárquico.

3.2.3. Mapa de zonificación del nivel de la vulnerabilidad

En el anexo 39, se presenta el mapa de vulnerabilidad por inundación fluvial en el área de estudio del Sector Magllanal. Este mapa se elaboró utilizando el software ArcGIS, a partir de la información recopilada sobre los elementos expuestos.

3.3. Establecer el nivel de riesgo de inundación fluvial

Para establecer el nivel de riesgo de inundación fluvial en el área de estudio del Sector Magllanal, se aplicó la metodología CENEPRED. Este proceso incluyó el uso del Análisis Jerárquico para identificar los rangos de niveles de peligro y vulnerabilidad. Posteriormente, se combinó la información de peligro y vulnerabilidad para calcular el nivel de riesgo de inundación para cada lote.

3.3.1. Niveles de riesgo

En la siguiente figura 27, se muestra los niveles de riesgo y sus respectivos rangos obtenidos a través del proceso de Análisis Jerárquico.

Figura 27

Niveles de Riesgo

NIVEL DE RIESGO	RANGO
MUY ALTO	$0.098 \leq R \leq 0.235$
ALTO	$0.042 \leq R < 0.098$
MEDIO	$0.013 \leq R < 0.042$
BAJO	$0.004 \leq R < 0.013$

Nota. La figura 27, muestra los niveles de riesgo muy alto, alto, medio y bajo con sus respectivos rangos

3.3.2. Estratificación del nivel de riesgo

En el anexo 33, se muestra la estratificación del nivel de riesgo, del área de estudio del Sector Magllanal; con sus respectivos rangos obtenidos a través de utilizar el Proceso de Análisis Jerárquico.

3.3.3. Mapa de zonificación del nivel de riesgo

Utilizando el software ArcGIS, se elaboró un mapa que representa el nivel de riesgo por inundación fluvial en el Sector Magllanal, basado en los elementos expuestos. Este mapa se encuentra detallado en el anexo 39.

3.4. Modelación del comportamiento del nivel de riesgo de inundación fluvial mediante el algoritmo Random Forest en el Sector Magllanal, Jaén

Para la modelación del comportamiento del nivel de riesgo, se identificó y analizó mediante la metodología CENEPRED, 23 variables en la vulnerabilidad y 6 variables en el peligro que influyen en el nivel riesgo del Sector, a base de estas variables se realizó la matriz de datos que estuvo conformada por 505 instancias.

En la tabla 9, se presenta la estadística descriptiva de las variables que influyen en el nivel de riesgo del Sector. En las figuras 28 y 29, se muestra el histograma y la caja de bigotes de la base de datos respectivamente, los cuales proporcionan una representación visual de la distribución de los datos de las variables seleccionadas.

Tabla 9*Estadística de las variables*

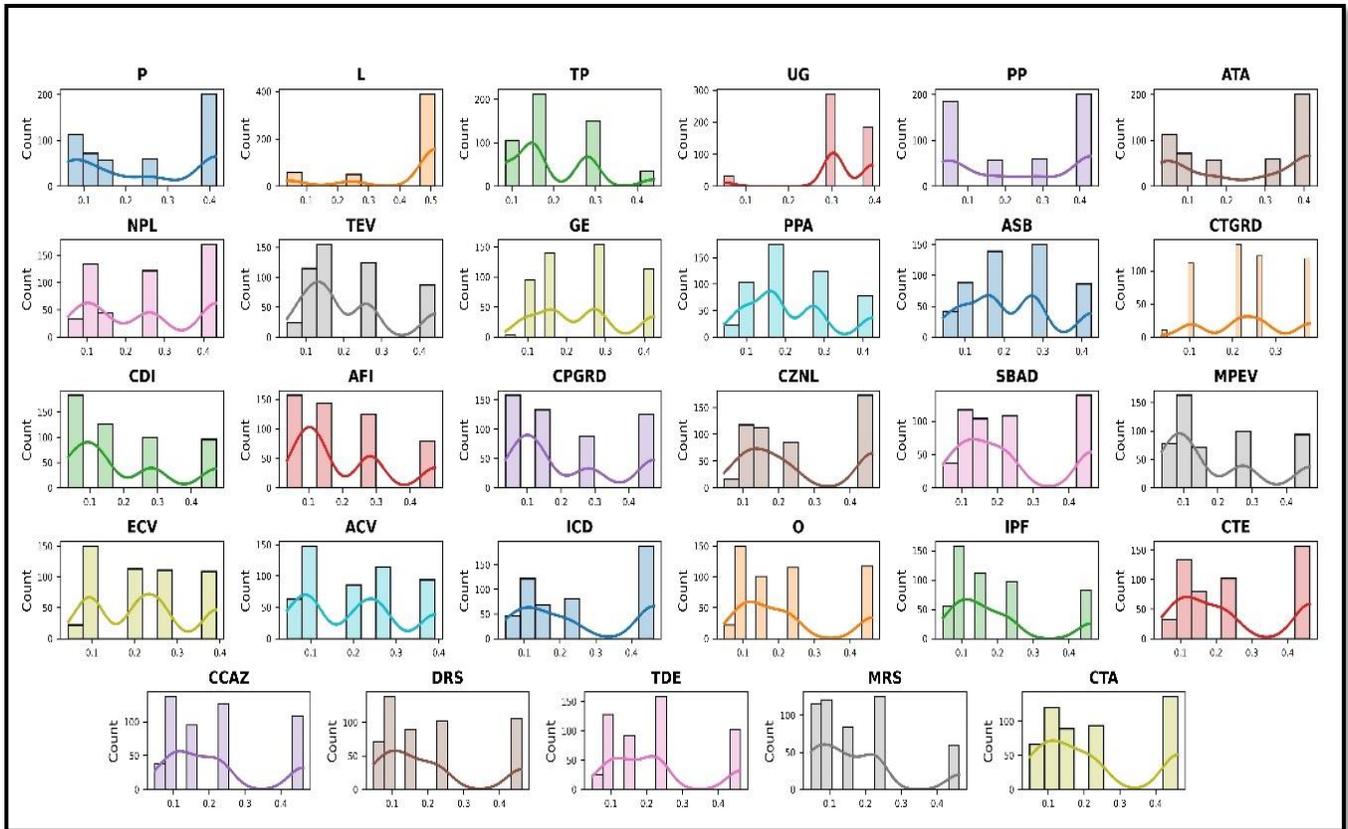
Variables	cantidad	Desviación						
		promedio	estándar	Mínimo	25%	50%	75%	máximo
P	505	0.24400792	0.15204807	0.062	0.099	0.262	0.416	0.416
L	505	0.42784752	0.16472304	0.037	0.513	0.513	0.513	0.513
TP	505	0.19555644	0.09874535	0.085	0.15	0.15	0.281	0.439
UG	505	0.32106139	0.08263055	0.05	0.304	0.304	0.395	0.395
PP	505	0.24414059	0.16996928	0.033	0.07	0.288	0.43	0.43
ATA	505	0.23991287	0.16434567	0.032	0.073	0.307	0.415	0.415
NPL	505	0.25274729	0.14402662	0.05181843	0.10032528	0.26207464	0.43278948	0.43278948
TEV	505	0.2110474	0.12208504	0.04921123	0.09876982	0.15156932	0.25964798	0.44080166
GE	505	0.24192003	0.11666666	0.04328056	0.16605214	0.27203298	0.27203298	0.42384088
PPA	505	0.21183526	0.11272002	0.04328056	0.09479344	0.16605214	0.27203298	0.42384088
ASB	505	0.21880419	0.11847683	0.04328056	0.09479344	0.16605214	0.27203298	0.42384088
CTGRD	505	0.23768592	0.09924147	0.03535919	0.21555934	0.21555934	0.26473966	0.37926682
CDI	505	0.20080512	0.15442264	0.0360274	0.07958174	0.12873465	0.28209159	0.47356462
AFI	505	0.20371811	0.1407945	0.0360274	0.07958174	0.12873465	0.28209159	0.47356462
CPGRD	505	0.22387408	0.16074817	0.0360274	0.07958174	0.12873465	0.28209159	0.47356462
CZNL	505	0.25786276	0.15394442	0.04768184	0.10179833	0.22810453	0.46206282	0.46206282
SBAD	505	0.23606596	0.1485694	0.04768184	0.10179833	0.16035249	0.46206282	0.46206282
MPEV	505	0.19382005	0.15199799	0.03655579	0.08934584	0.13330218	0.2730206	0.46777559
ECV	505	0.21866434	0.11224343	0.04447106	0.09605067	0.20985026	0.25849891	0.3911291

ACV	505	0.20059269	0.117474	0.04447106	0.09605067	0.20985026	0.25849891	0.3911291
ICD	505	0.25913071	0.16402781	0.04768184	0.10179833	0.22810453	0.46206282	0.46206282
O	505	0.22338163	0.14061704	0.04768184	0.10179833	0.16035249	0.22810453	0.46206282
IPF	505	0.19214019	0.13170899	0.04768184	0.10179833	0.16035249	0.22810453	0.46206282
CTE	505	0.24515944	0.15426076	0.04768184	0.10179833	0.22810453	0.46206282	0.46206282
CCAZ	505	0.21801542	0.13925496	0.04768184	0.10179833	0.16035249	0.22810453	0.46206282
DRS	505	0.20492715	0.14363034	0.04768184	0.10179833	0.16035249	0.22810453	0.46206282
TDE	505	0.22266791	0.13285986	0.04768184	0.10179833	0.22810453	0.22810453	0.46206282
MRS	505	0.17328208	0.12460942	0.04768184	0.10179833	0.16035249	0.22810453	0.46206282
CTA	505	0.22557731	0.15373861	0.04768184	0.10179833	0.16035249	0.46206282	0.46206282

Nota. La tabla 9, muestra la estadística descriptiva de las variables, la distribución y variabilidad de cada variable en la base de datos, para identificar tendencias, patrones y posibles anomalías en los datos.

Figura 28

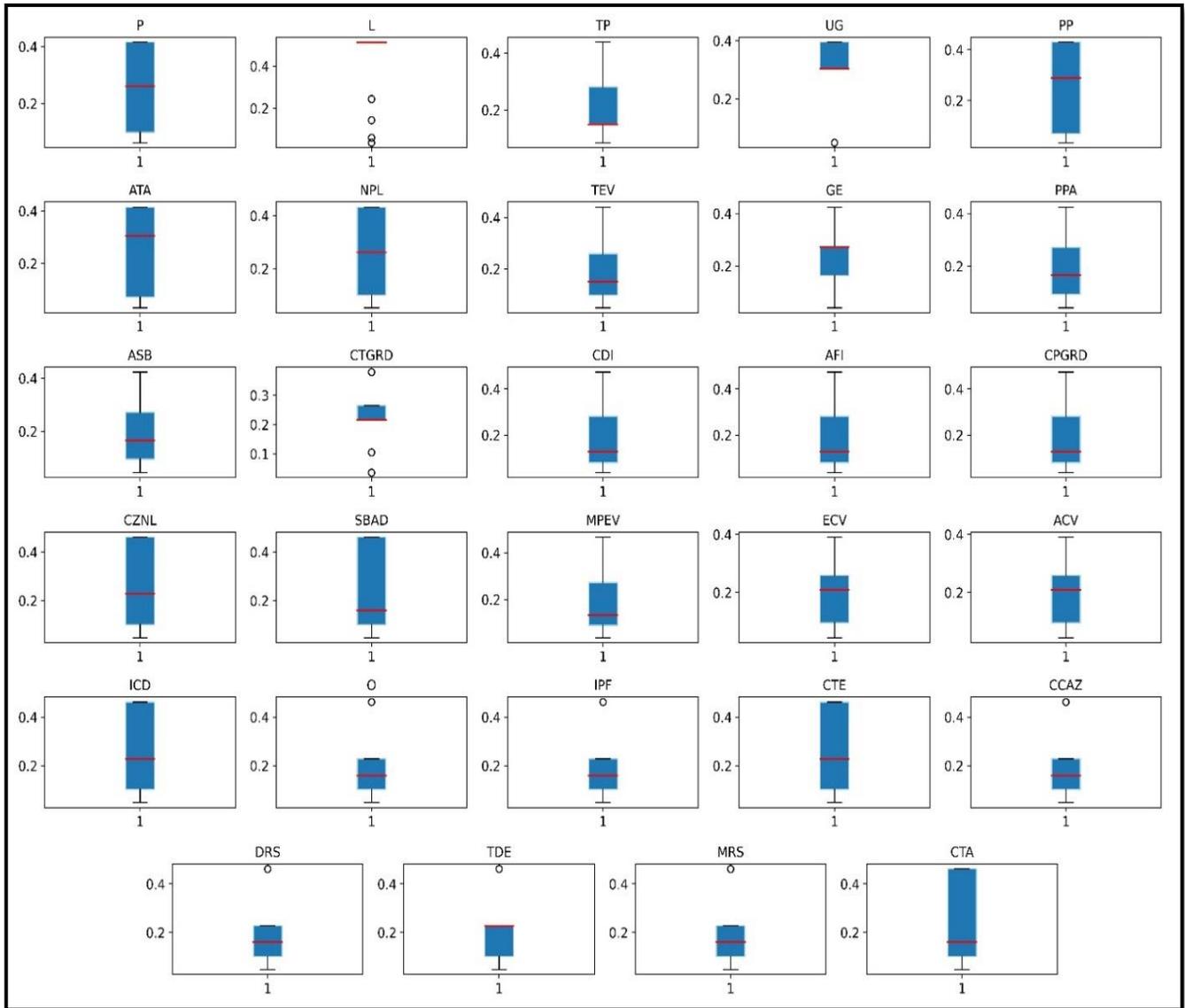
Histograma de la base de datos



Nota. La figura 28, se muestra el histograma de la distribución de las variables, donde se observa que cada barra representa un intervalo de valores y la altura de la barra indica cuántas observaciones caen dentro de ese intervalo.

Figura 29

Gráfico de bigotes de la base de datos

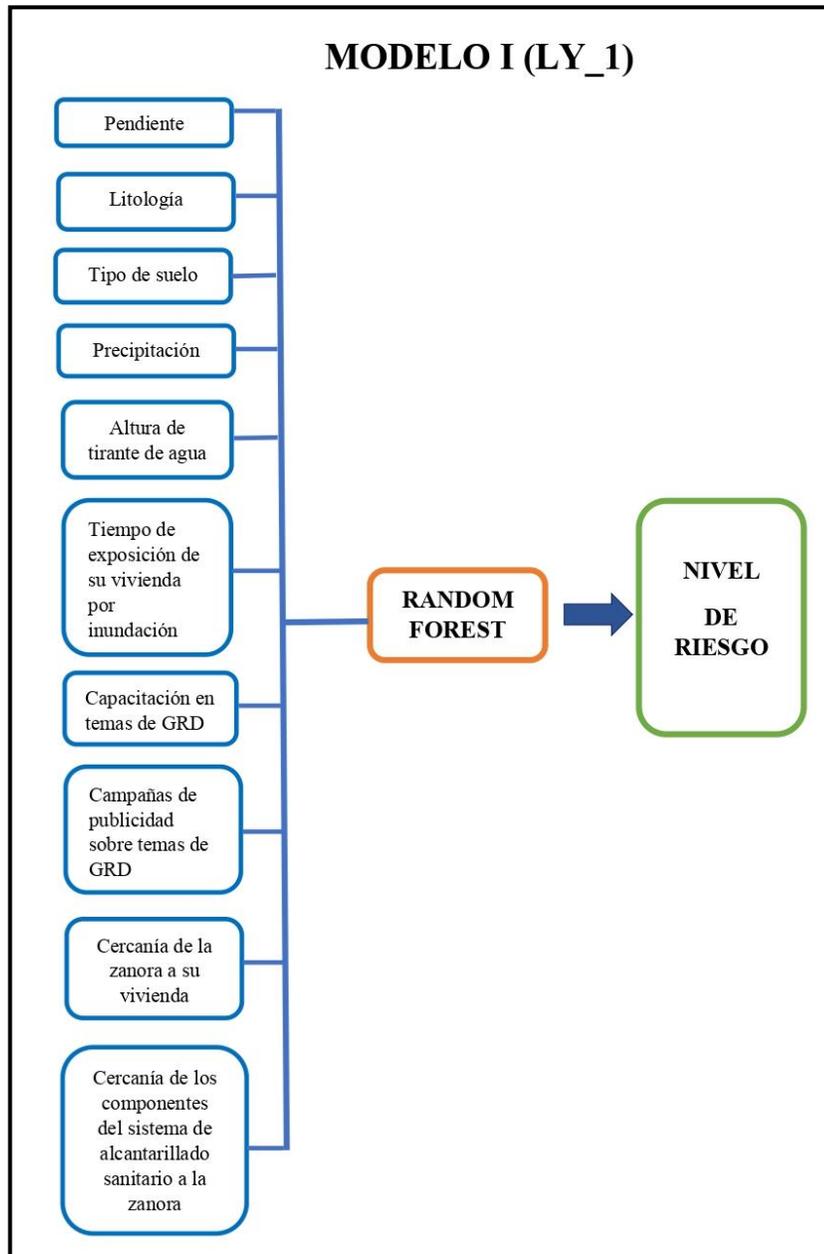


Nota. La figura 29, muestra la distribución, la tendencia central, la dispersión y la presencia de valores atípicos de las variables seleccionadas.

Con base a estas variables se desarrollaron 5 modelos (LY_1, LY_2, LY_3, LY_4 y LY_5) mediante el algoritmo Random Forest para estimar el nivel de riesgo de 68 lotes en escenarios futuros. En las figuras 30, 31, 32, 33 y 34 se muestra los modelos con sus respectivas variables LY_1, LY_2, LY_3, LY_4 y LY_5 respectivamente.

Figura 30

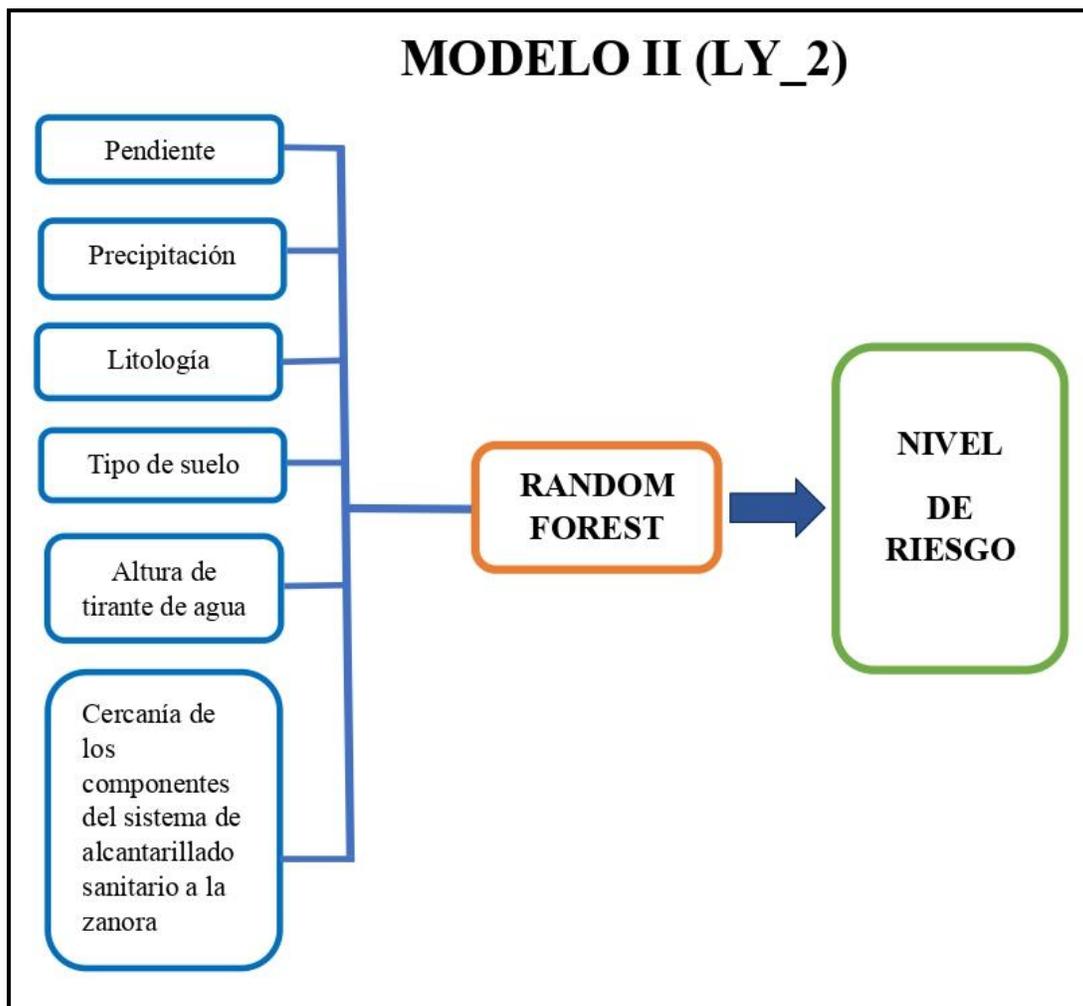
Variables del modelo I (LY_1)



Nota. La figura 30, muestra las 10 variables de entrada que conforman el modelo I y la variable de salida (nivel de riesgo).

Figura 31

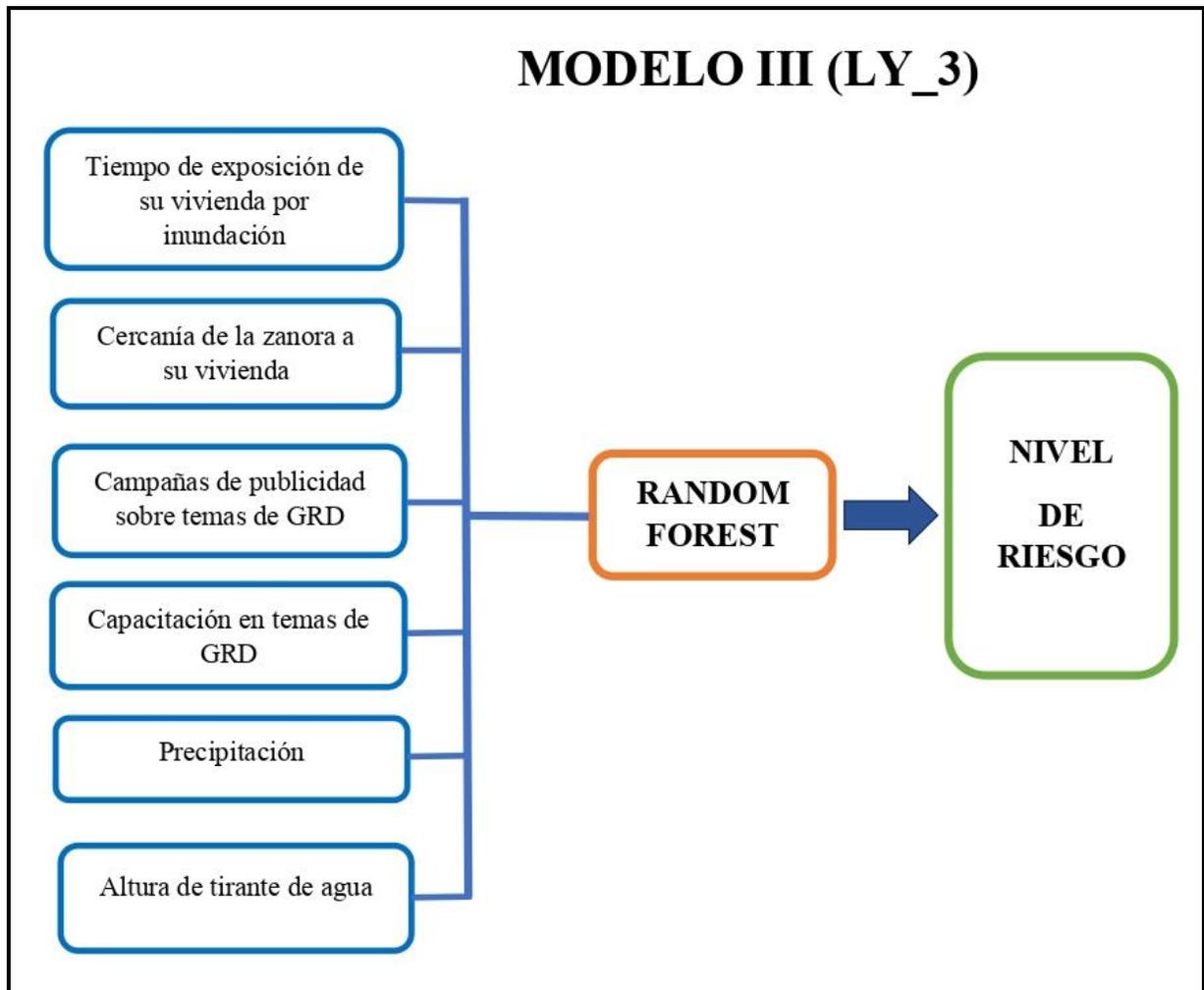
Variables del modelo II (LY_2)



Nota. La figura 31, muestra las 6 variables de entrada que conforman el modelo II y la variable de salida (nivel de riesgo).

Figura 32

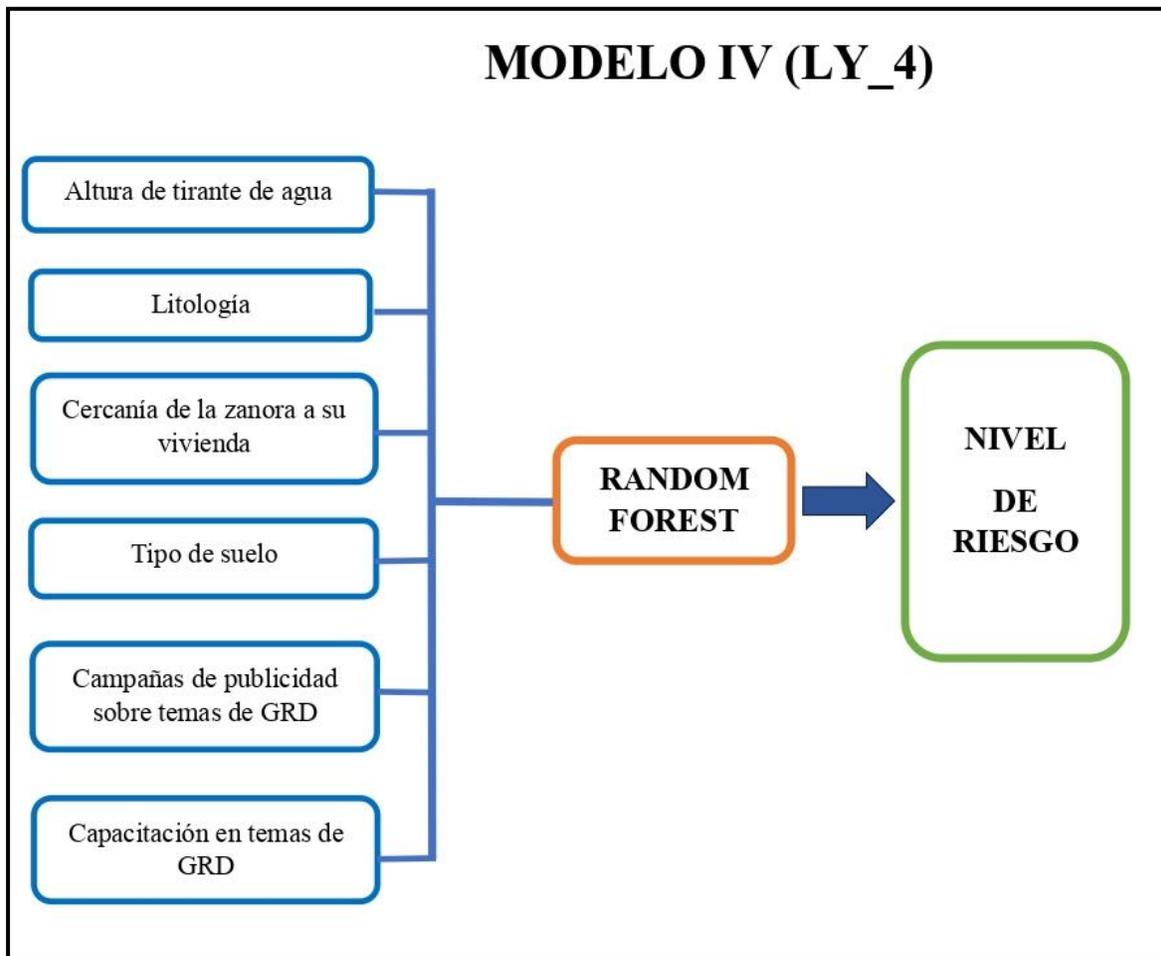
Variables del modelo III (LY_3)



Nota. La figura 32, muestra las 6 variables de entrada que conforman el modelo III y la variable de salida (nivel de riesgo).

Figura 33

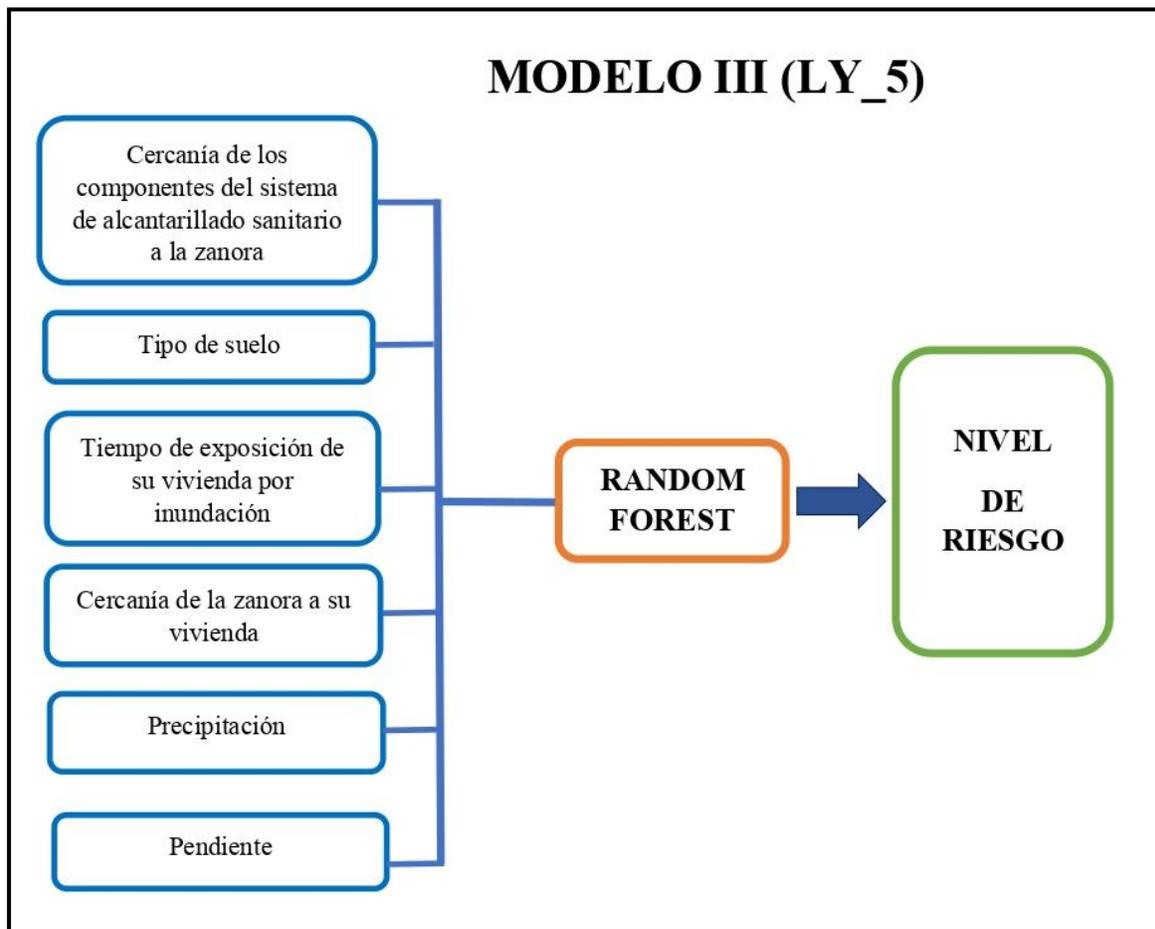
Variables del modelo IV (LY_4)



Nota. La figura 33, muestra las 6 variables de entrada que conforman el modelo IV y la variable de salida (nivel de riesgo).

Figura 34

Variables del modelo V (LY_5)



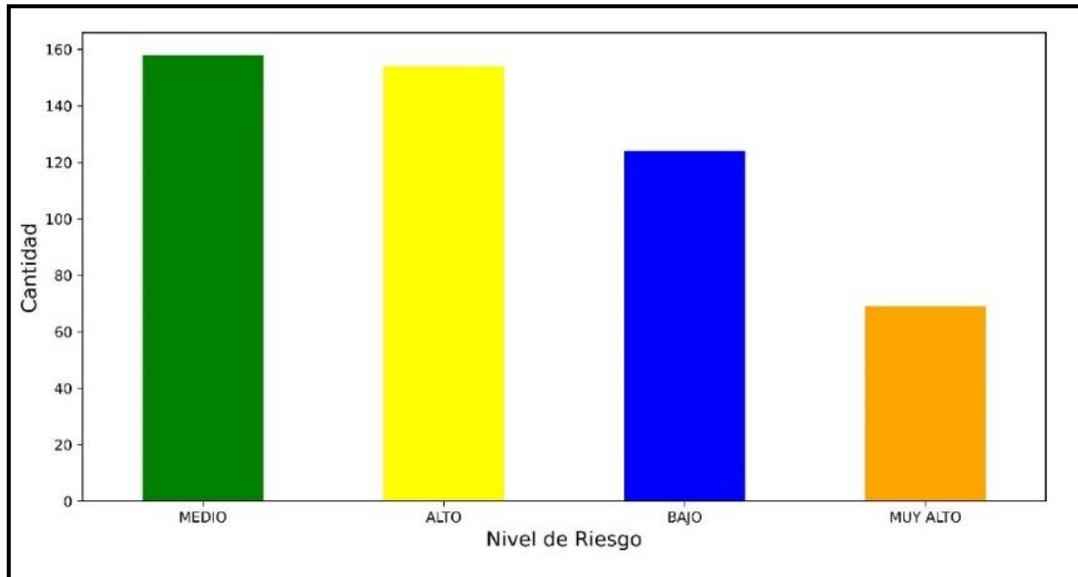
Nota. La figura 34, muestra las 6 variables de entrada que conforman el modelo V y la variable de salida (nivel de riesgo).

Después de asociar cada modelo con sus respectivas variables, se llevó a cabo la modelación del nivel de riesgo utilizando cada uno de ellos. Como resultado de esta modelación, se obtuvieron clasificaciones del nivel de riesgo que incluyeron categorías como "muy alto", "alto", "medio" y "bajo" de la base de datos.

En la figura 35, se muestra la modelación del riesgo con sus diferentes niveles.

Figura 35

Niveles de riesgo



Nota. La Figura 35, presenta los niveles de riesgo de las 505 instancias, mediante 23 variables de peligro y 6 variables de vulnerabilidad, predominando los niveles medio y alto.

Después de haber entrenado y validado los modelos mediante el test de validación cruzada de los modelos LY_1, LY_2, LY_3, LY_4, LY_5 con el algoritmo Random Forest, se obtuvo la precisión de cada modelo.

Tabla 10

Precisión de los modelos según el algoritmo Random Forest

MODELO	PRECISIÓN
LY_1	98.53%
LY_2	89.71%
LY_3	98.53%
LY_4	95.59%
LY_5	94.12%

Nota. La presente tabla 10, detalla la precisión de los 5 modelos según el algoritmo Random Forest, en base a las instancias correctamente e incorrectamente clasificadas, donde el modelo LY_1 y LY_3 presentan mayor precisión con un 98.53%.

Resultados obtenidos de la predicción del nivel de riesgo con el algoritmo Random Forest se presenta en la tabla 11 y 12, donde los modelos LY_1 y LY_3 presentan mayor precisión en la predicción del nivel de riesgo de los 68 lotes.

Tabla 11

Predicción del modelo LY_1

LOTES	REAL	PREDICCIÓN
1	ALTO	ALTO
2	ALTO	ALTO
3	MEDIO	MEDIO
4	MEDIO	ALTO
5	MEDIO	MEDIO
6	MEDIO	MEDIO
7	MEDIO	MEDIO
8	MEDIO	MEDIO
9	ALTO	ALTO
10	ALTO	ALTO
11	ALTO	ALTO
12	ALTO	ALTO
13	ALTO	ALTO
14	ALTO	ALTO
15	ALTO	ALTO
16	ALTO	ALTO
17	MEDIO	MEDIO
18	ALTO	ALTO
19	ALTO	ALTO
20	ALTO	ALTO
21	MEDIO	MEDIO
22	MEDIO	MEDIO
23	ALTO	ALTO
24	MEDIO	MEDIO
25	MEDIO	MEDIO
26	ALTO	ALTO
27	ALTO	ALTO
28	ALTO	ALTO

29	MEDIO	MEDIO
30	MEDIO	MEDIO
31	ALTO	ALTO
32	MEDIO	MEDIO
33	MEDIO	MEDIO
34	ALTO	ALTO
35	MEDIO	MEDIO
36	ALTO	ALTO
37	MEDIO	MEDIO
38	ALTO	ALTO
39	ALTO	ALTO
40	ALTO	ALTO
41	ALTO	ALTO
42	ALTO	ALTO
43	MEDIO	MEDIO
44	ALTO	ALTO
45	ALTO	ALTO
46	ALTO	ALTO
47	ALTO	ALTO
48	MEDIO	MEDIO
49	ALTO	ALTO
50	MEDIO	MEDIO
51	MEDIO	MEDIO
52	ALTO	ALTO
53	MEDIO	MEDIO
54	ALTO	ALTO
55	ALTO	ALTO
56	ALTO	ALTO
57	ALTO	ALTO
58	ALTO	ALTO
59	ALTO	ALTO
60	ALTO	ALTO
61	MEDIO	MEDIO
62	ALTO	ALTO
63	ALTO	ALTO
64	ALTO	ALTO
65	ALTO	ALTO
66	ALTO	ALTO
67	'MUY ALTO'	MUY ALTO
68	ALTO	ALTO

Nota. La Tabla 11, muestra el modelo de predicción LY_1, el cual destaca por una particularidad en el lote 4; mientras que los otros lotes exhiben un nivel de riesgo constante, el lote 4 se distingue al mostrar un nivel de riesgo medio en la actualidad y para escenarios futuros predice un nivel de riesgo alto.

Tabla 12

Predicción del modelo LY_3

LOTES	REAL	PREDICCIÓN
1	ALTO	ALTO
2	ALTO	ALTO
3	MEDIO	MEDIO
4	MEDIO	MEDIO
5	MEDIO	MEDIO
6	MEDIO	MEDIO
7	MEDIO	MEDIO
8	MEDIO	MEDIO
9	ALTO	ALTO
10	ALTO	ALTO
11	ALTO	ALTO
12	ALTO	ALTO
13	ALTO	ALTO
14	ALTO	ALTO
15	ALTO	ALTO
16	ALTO	ALTO
17	MEDIO	MEDIO
18	ALTO	ALTO
19	ALTO	ALTO
20	ALTO	ALTO
21	MEDIO	MEDIO
22	MEDIO	MEDIO
23	ALTO	ALTO
24	MEDIO	MEDIO
25	MEDIO	MEDIO
26	ALTO	ALTO
27	ALTO	ALTO
28	ALTO	ALTO
29	MEDIO	MEDIO
30	MEDIO	MEDIO
31	ALTO	ALTO
32	MEDIO	MEDIO
33	MEDIO	MEDIO

34	ALTO	ALTO
35	MEDIO	MEDIO
36	ALTO	ALTO
37	MEDIO	MEDIO
38	ALTO	ALTO
39	ALTO	ALTO
40	ALTO	ALTO
41	ALTO	ALTO
42	ALTO	ALTO
43	MEDIO	MEDIO
44	ALTO	ALTO
45	ALTO	MUY ALTO
46	ALTO	ALTO
47	ALTO	ALTO
48	MEDIO	MEDIO
49	ALTO	ALTO
50	MEDIO	MEDIO
51	MEDIO	MEDIO
52	ALTO	ALTO
53	MEDIO	MEDIO
54	ALTO	ALTO
55	ALTO	ALTO
56	ALTO	ALTO
57	ALTO	ALTO
58	ALTO	ALTO
59	ALTO	ALTO
60	ALTO	ALTO
61	MEDIO	MEDIO
62	ALTO	ALTO
63	ALTO	ALTO
64	ALTO	ALTO
65	ALTO	ALTO
66	ALTO	ALTO
67	'MUY ALTO'	MUY ALTO
68	ALTO	ALTO

Nota. En la Tabla 12, detalla el modelo de predicción LY_3, revelando una diferencia notable en el lote 45 en comparación con los demás lotes; mientras que los otros lotes mantienen un nivel de riesgo consistente, el lote 45 se destaca por tener un nivel de riesgo alto en el presente, y se anticipa que este riesgo aumentará significativamente en futuros escenarios, alcanzando un nivel muy alto.

IV. DISCUSIÓN

Basándonos en los resultados obtenidos, confirmamos la hipótesis general que establece que el nivel de riesgo de inundación fluvial, evaluado mediante el algoritmo Random Forest, es alto en el Sector Magllanal, Jaén, Cajamarca, con un error menor al 5%. Se registró un error del 1.47%, lo que indica una estimación de riesgo alto. Estos valores guardan relación con lo que sostiene Anchelía y Mori (2020) donde determina que el nivel de riesgo frente a inundaciones se encuentra en un nivel de riesgo alto con un valor de predicción del 99.41%. con un error de 0.49%. Al igual que López et al. (2021), donde obtuvo una precisión de 97.61% con un error de 2.39% mediante el algoritmo Random Forest. En esta investigación, es importante destacar que la estimación del nivel de riesgo mediante el algoritmo Random Forest fue alta con una precisión de 98.53% de variabilidad observada de inundación fluvial, esto se debe al conjunto de variables que se midieron en esta investigación, y el margen de error del 1.47% se debe a la limitación en la consideración de otras variables, ya que nos ajustamos a la metodología establecida por CENEPRED, que especifica un conjunto definido de variables a evaluar. Además, es relevante señalar que, mientras que la evaluación del nivel de riesgo de inundación fluvial a través de la metodología CENEPRED es descriptiva y produce resultados estáticos, la aplicación del algoritmo Random Forest es cuantitativa y dinamiza el procesamiento de información al describir y predecir el comportamiento del riesgo de manera más eficiente.

Para determinar el peligro en el área de estudio a sufrir inundaciones por el desborde de la zanora Magllanal a causa del fenómeno de lluvias intensas se utilizó el Sistema de información Geográfica (SIG), para el análisis y registro de datos de factores condicionantes como pendiente, geomorfología, geología y tipo de suelo; obteniendo un nivel de peligro medio y alto por tener pendientes entre 11° - 30° , la cual es semejante a la investigación de García et al.(2021) que utilizó la información geográfica SIG para el análisis de factores condicionantes como tipo de suelo, pendiente topográfica y geomorfología en la cual el peligro a sufrir inundaciones en la cuenca inferior de Rio Negro Argentina es media y alta por tener un suelo deficiente y pendientes topográficas de 1%. En comparación con la investigación de López (2021) que identificó y mapeo a través del software GIS a 421 viviendas, obteniendo un nivel de peligro medio porque sólo 9.1% de viviendas se encuentra entre un nivel alto y muy alto; y Guerrero (2022) realizó la valorización del peligro mediante 6 tipos encuestas de 8 preguntas relacionadas a las viviendas, estructuras y eventos o fenómenos de precipitaciones; obteniendo un nivel medio por la construcción de viviendas en las fajas marginales del río Chinchipe.

Para el análisis de la vulnerabilidad por inundación se utilizó la metodología Cenedred, en la cual se recaudó información de los elementos expuestos (infraestructura de las viviendas, la red de agua potable, la red de desagüe y vías de comunicación, entre otros) a base de encuestas realizadas a 68 lotes del área de estudio, obteniéndose un nivel de Vulnerabilidad alto con una 72.06% que representan 49 lotes ; puesto que la exposición a nivel de lotes son vulnerables ante una inundación fluvial por lo que las viviendas están construidas cerca de 0.2km - 1km de la Zonora Magllanal; la cual es semejante a la investigación Guerrero (2022) también determinó el nivel de vulnerabilidad en base encuestas relacionadas a viviendas, estructuras, etc, realizadas a

210 personas, obteniendo un grado de vulnerabilidad alto dado que la población se expone a construir sus viviendas en las fajas marginales del río Chinchipe. La presente investigación es semejante en el mismo nivel de vulnerabilidad, puesto que la investigación de Guerrero los lotes están cerca de un río y en nuestra investigación están cerca una zanora que se activa ante lluvias intensas generando desbordamiento de esta misma.

Al aplicar la metodología Cenepred utilizando el proceso de Análisis Jerárquico al realizar la combinación entre peligro y vulnerabilidad se determinó el nivel de riesgo alto como más relevante por inundación fluvial por desborde de la zanora Magllanal causado por la activación de la zanora Magllanal con umbrales de categoría lluvioso, además por presentar desechos de residuos en el cauce de la zanora y por presentar una susceptibilidad física muy alta en la cual se encuentra ubicada en pendientes de 21° - 30 entre otros aspectos, la cual es semejante a la investigación Guillen y Sifuentes (2020) determinó los niveles de riesgo mediante la metodología Cenepred en el cual predominó el nivel de riesgo medio al peligro de inundación por desborde del río Chorobal debido a su poco mantenimiento periódico. En comparación con la investigación de Zúñiga (2020) utilizó la misma metodología Cenepred en la cual obtuvo como resultado que las zonas con mayor riesgo por inundación fluvial son las zonas con cauces activos, zonas de erosión, los valles-quebradas y los acantilados y Ticlla (2022) determinó que el nivel de riesgo es bajo ya que presenta áreas con mayor probabilidad a inundaciones con pendientes menores a 5%.

Para la modelación del comportamiento del nivel de riesgo, se identificó y analizó mediante la metodología CENEPRED las variables que influyen en el peligro y vulnerabilidad del Sector; a base de estas variables se desarrolló cinco modelos mediante el algoritmo Random Forest para

estimar el nivel de riesgo de 68 lotes en escenarios futuros obteniendo como resultado un nivel de riesgo alto de 66.18% lotes, con una precisión de 98.53%. Los resultados obtenidos son similares a Anchelía y Mori (2020) que comparó y analizó el método de Proceso de Análisis Jerárquico y Random Forest a base de datos históricos, para obtener el método más eficaz en la determinación de zonas susceptibles a inundaciones; concluyendo que el algoritmo Random Forest es efectivo para determinar la susceptibilidad a inundaciones, con un valor de predicción del 99.41%, lo que respalda la elección del algoritmo. Además, López et al. (2021), estimó el riesgo de inundación en instituciones educativas del Perú a través de varios modelos predictivos a base de 65 000 registros publicados por el Ministerio de Educación del Perú, obteniendo como resultado que el mejor modelo es el algoritmo Random Forest con una precisión de 97.612%. Hernández (2023), utilizó diferentes modelos de machine learning a base de datos históricos de inundaciones para predecir las pérdidas por inundación, obteniendo que el mejor modelo de redes neuronales es el algoritmo Hyperband para predecir las pérdidas por inundación con una precisión de 0.3150 en validación y 0.2868 para los datos de prueba; concluyendo que los algoritmos son eficaces para estimar el riesgo de inundación.

V. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

- La estimación del nivel de riesgo ante inundación fluvial mediante el Algoritmo Random Forest en el Sector Magllanal-Jaén-Cajamarca, se encontrará en un nivel de riesgo alto con una precisión de 98.53% y un error menor a 5%, aceptando la hipótesis establecida.
- La identificación del nivel de peligro como consecuencia de inundación fluvial a través de la evaluación de los parámetros de los factores condicionantes y desencadenantes de la población del área de estudio del Sector Magllanal, los resultados muestran que el 19.12 % (13 lotes) se encuentra en un rango de peligro alto $0.172 \leq P < 0.294$ y un 80.88% (55 lotes) está en un rango de peligro medio $0.075 \leq P < 0.172$. En conclusión, el área de estudio presenta un nivel de peligro medio de inundación fluvial debido a la influencia de la zanora Magllanal.
- Se analizó el nivel de vulnerabilidad a través de los parámetros de exposición, fragilidad y resiliencia de las dimensiones social, económica y ambiental de la población estudiada, la cual estratifican la zona en niveles de vulnerabilidad alto ($0.242 \leq V \leq 0.335$) y medio ($0.174 \leq V < 0.242$) de inundación fluvial por la zanora Magllanal. De los 68 lotes tenemos que el 72.06% (49 lotes) se encuentra en un nivel de vulnerabilidad alto y 26.47% (18 lotes) en un nivel de vulnerabilidad medio.
- Se estableció el nivel de riesgo de los 68 lotes de inundación fluvial en el área de estudio del Sector Magllanal, combinando los niveles de peligro y vulnerabilidad. Los resultados mostraron niveles de riesgo alto ($0.042 \leq R \leq 0.098$) y medio ($0.013 \leq R < 0.042$) como más relevantes; en el cual el 69.12% (47 lotes) tienen un riesgo alto y

29.4% (20 lotes) tienen un riesgo medio. Se concluye que los resultados aplicando la Metodología CENEPRED revelaron un nivel de riesgo alto en esta área.

- Se modeló el nivel de riesgo a través del algoritmo Random Forest planteándose cinco modelos LY_1, LY_2, LY_3, LY_4 y LY_5, obteniéndose una precisión del 98.53% con el LY_1, este modelo predice que en el Sector Magllanal, la mayoría de las viviendas 66.18% (46 lotes) experimentarán un nivel alto de riesgo de inundación fluvial, mientras que un 32.35% (22 lotes) tendrán un nivel medio. En general, los resultados del algoritmo Random Forest estimaron un nivel de riesgo alto en esta área.

5.2.Recomendaciones

- Aplicar otros métodos de inteligencia artificial para determinar el nivel de riesgo ante inundación fluvial de un área vulnerable, que sirva de apoyo para la toma de decisiones e implementación de las medidas de mitigación, mediante la extracción de patrones y tendencias.
- Para realizar la identificación del nivel de peligro se debe utilizar el software ARCGIS para recopilar información geográfica en la que se encuentran los elementos expuestos. Además, permite organizar e identificar a través de mapas los factores condicionantes como la pendiente, geomorfología, litología y tipo de suelo del área de estudio.
- Para el análisis del nivel de vulnerabilidad, el orden de los descriptores de cada parámetro, debe considerarse desde el más crítico al menos influyente para obtener la ponderación de parámetros de acuerdo al proceso de análisis jerárquico de la matriz Saaty según el Manual de CENEPRED, y por consiguiente determinar el nivel de riesgo en base a la metodología mencionada.
- Al establecer si este nivel de riesgo afecta a la población y sus medios de vida; proponer medidas de mitigación a fin de evitar el desbordamiento de zanoras, quebradas o ríos; estas medidas pueden ser implementar obras de drenaje pluvial en el área de estudio para evitar que la población sufra daños en sus viviendas por inundación ante lluvias extremadamente intensas, muros contención con refuerzo, puentes peatonales y tipo alcantarilla para facilitar el tránsito peatonal y vehicular; asimismo limpieza del cauce y capacitaciones de riesgos de desastres a la población, rutas de evacuación y zonas seguras ante inundación fluvial como medidas de seguridad alerta temprana.

- Se recomienda utilizar el algoritmo Random Forest y explorar otros algoritmos de machine learning con otras variables de entrada para determinar el nivel de riesgo de inundación en hospitales, colegios y viviendas.

VI. REFERENCIAS BIBLIOGRÁFICAS

- Alacón, A., Chávez, J., Palacios, Ó., & Ibáñez, L. (2020). Estimating areas vulnerable to flooding in urban zones: Morelia, Michoacán, Mexico. *Tecnología y Ciencias del Agua*, 11(3), 01-26. <https://doi.org/10.24850/j-tyca-2020-03-01>
- Alaya García, E., & Riquero Miranda, W. (2020). *Caracterización de los modelamientos hidráulicos numéricos de inundaciones fluviales, Cajamarca 2020 [Tesis de licenciatura, Universidad Privada del Norte]*. Repositorio Institucional UPN. <https://hdl.handle.net/11537/28238>
- Anchelía Carhuaricra, D., & Mori Sáenz, X. (2020). *Determinación de zonas susceptibles a inundaciones y análisis comparativo del Proceso de Análisis Jerárquico (AHP) y Random Forest (RF). Caso estudio: cuenca baja del río Chancay Lambayeque [Tesis, Universidad Nacional Mayor de San Marcos]*. Repositorio de Tesis Digitales. <https://hdl.handle.net/20.500.12672/15868>
- Barrientos, E. P. (s.f.). *Tecnologías de la Información geoespacial al servicio del país*. <https://sites.google.com/view/eportuguez/p%C3%A1gina-principal?authuser=0>
- Cárdenas Rodríguez, S., Vides Herrera, C., & Pardo García, A. (2022). Sistema de alerta temprana de inundaciones para el río Arauca basado en técnicas de inteligencia artificial. *Revista De Investigación, Desarrollo E Innovación*, 12(2), 315-326. <https://doi.org/https://doi.org/10.19053/20278306.v12.n2.2022.15274>
- Centro de Operaciones de Emergencia Nacional (COEN). (19 de Marzo de 2023). *Boletín Informativo* COEN.

<https://cdn.www.gob.pe/uploads/document/file/4281702/BOLETIN%20COEN%20N%20012.pdf.pdf?v=1679244153>

Centro Nacional de Estimación, Prevención y Reducción del Riesgo de Desastres (CENEPRED).

(2014). *Manual para la Evaluación de Riesgos originados por Fenómenos Naturales – 2da Versión*. Neva Studio SAC.

Delgado Ramos, F. (2023). *Riesgo por inundación del Río La Envinada en el Sector Pueblo Nuevo, Distrito Chontali, Provincia Jaén, Departamento Cajamarca [Tesis Titulo Profesional de Ingeniería Civil, Universidad Nacional de Cajamarca]*. Repositorio Institucional. <http://hdl.handle.net/20.500.14074/6172>

Díaz Briones, O. E. (2019). *Determinación de zonas de inundación mediante modelamiento hidráulico del río Chonta en el caserío Tartar Chico, distrito de Baños del Inca [Tesis, Universidad Nacional de Cajamarca]*. Repositorio Universidad Nacional de Cajamarca (UNC). <http://hdl.handle.net/20.500.14074/3450>

Fraga, I., Cea, L., Puertas, J., Mosqueira, G., Quinteiro, B., Botana, S., . . . Taboada, J. (2021). MERLIN: A new tool for flood hazard forecasting at the Galicia-Costa. *Ingeniería del Agua*, 25(3), 215-227. <https://doi.org/https://doi.org/10.4995/ia.2021.15565>

García Tocto, D. Y. (2019). *Modelado Hidráulico del Río Amojú Aplicando el Software Iber para Zonificar Áreas Inundables en la Ciudad de Jaén, Perú [Tesis de pregrado, Universidad Nacional de Jaén]*. Repositorio Institucional UNJ. <http://repositorio.unj.edu.pe/handle/UNJ/228>

García, G., Piccolo, M., & Bohn, V. (2021). Estimación de la susceptibilidad a inundaciones en la cuenca inferior del Río Negro, Argentina. *Finisterra*, 56(118), 51-70. <https://doi.org/10.18055/finis21647>

- Goicochea Cueva, A. R. (2021). *Estimación del Nivel de Riesgo por Inundación de la Quebrada Cruz Blanca para la Zona Urbana-Cajamarca, 2020 [Tesis, Universidad Nacional de Cajamarca]*. Repositorio Universidad Nacional de Cajamarca (UNC). <http://hdl.handle.net/20.500.14074/4453>
- Gonzales, J. A., & Covinos Gallardo, M. (2021). *Diseño y Metodología de la Investigación* (1 ed.). Enfoques consulting EIRL. <https://www.researchgate.net/publication/352157132>
- Guerrero Diaz, R. (2022). *Estimación del nivel de riesgo por inundación del río Chinchipe en el Centro Poblado Puerto Huallape, Jaén-Cajamarca [Tesis titulo profesional de Ingeniero Civil, Universidad Nacional de Cajamarca]*. Repositorio Institucional. <http://hdl.handle.net/20.500.14074/5004>
- Guillen, C., & Sifuentes, H. (2020). *Evaluación de Riesgos Ante Amenazas de Inundación por Desborde del Rio Chorobal, Sector El Inca, Chao, Virú, La Libertad-2020 [Tesis, Universidad César Vallejo]*. Repositorio Institucional. <https://hdl.handle.net/20.500.12692/63771>
- Hernández Rojas, L. (2023). *Modelaje de inundaciones en el departamento de Antioquia a partir de redes neuronales [Proyecto de grado-pregrado, Universidad de los Andes – Colombia]*. Repositorio Institucional Séneca. <http://hdl.handle.net/1992/64076>
- Hernández Sampieri, R., & Mendoza Torres, C. (2018). *Metodología de la investigación Las rutas cuantitativa, cualitativa y mixta* (1 ed.). McGraw-Hill Interamericana.
- Instituto Nacional de Defensa Civil (INDECI). (13 de Marzo de 2023). *Resumen Ejecutivo N°323-2023. Temporada de Lluvias 2022-2023 (N° 165)*: <https://drive.google.com/file/d/1TXLtyd6ne7oRen3-m3hrBYGD9rhL6Vzh/view>

- Lopez Ordinola, L. C. (2021). *Evaluación del Riesgo por Inundaciones en las Viviendas del Asentamiento Humano Sagrado Corazón de Jesús, Distrito de Castilla-Piura [Tesis título profesional, Universidad César Vallejo]*. Repositorio Digital Institucional. <https://hdl.handle.net/20.500.12692/87190>
- López Vega, J., Torres Lázaro, J., & Herrera Quispe, J. (2021). Modelo predictivo basado en machine learning para la estimación de vulnerabilidades de riesgo de inundación y deslizamiento. Caso de estudio: instituciones educativas del Perú. *Actas Del Congreso Internacional De Ingeniería De Sistemas*, 206-207. <https://doi.org/https://doi.org/10.26439/ciis2021.5637>
- Mantilla Briones, F. (2019). *Identificación de áreas inundables mediante modelamiento hidráulico de la quebrada Chinchinmarca - sector 24 (Villa Huacariz) - Cajamarca [Tesis título profesional, Universidad Nacional de Cajamarca]*. Repositorio Institucional. <http://hdl.handle.net/20.500.14074/3693>
- Munich RE. (10 de Enero de 2022). *Hurricanes, cold waves, tornadoes: Weather disasters in USA dominate natural disaster losses in 2021*. <https://www.munichre.com/en/company/media-relations/media-information-and-corporate-news/media-information/2022/natural-disaster-losses-2021.html>
- Organización Meteorológica Mundial (OMM). (5 de Octubre de 2021). *Sequías, tormentas e inundaciones: el agua y el cambio climático dominan la lista de desastres*. <https://onuhabitat.org.mx/index.php/sequias-tormentas-e-inundaciones-el-agua-y-el-cambio-climatico-dominan-la-lista-de-desastres>
- Paredes Delgado, R. (2023). *Evaluación del nivel de riesgo de inundación por la quebrada Tumbillan en la localidad de Bellavista Viejo de la provincia de Jaén, región Cajamarca*

- [Tesis Título Profesional, Universidad Nacional de Cajamarca]. Repositorio Institucional.
<http://hdl.handle.net/20.500.14074/6012>
- Portillo, D., Beytia, V., Laínez, V., Quevedo, P., & Loyola, A. (21 de Marzo de 2023). *El ciclón Yaku, las inundaciones y el impacto en la vida de las niñas y niños de las regiones afectadas*. <https://coperainfanciaperu.com/2023/03/21/el-ciclon-yaku-las-inundaciones-y-el-impacto-en-la-vida-de-las-ninas-y-ninos-de-las-regiones-afectadas-breves-recomendaciones-para-el-desarrollo-infantil/>
- Suárez, M. B. (2012). *Metodología de la investigación científica para ingenieros* (1 ed.). Universidad Nacional Pedro Ruiz Gallo.
- Ticlla Sanchez, Y. (2022). *Evaluación del nivel de riesgo a inundación según metodología CENEPRED, en el Río San Mateo, tramo puente "San Mateo" al puente "Lloque Yupanqui"* [Tesis de licenciatura, Universidad Nacional Autónoma de Chota]. Repositorio institucional UNACH. <http://repositorio.unach.edu.pe/handle/UNACH/205>
- UNDRR. (2023). *PANORAMA DE LOS DESASTRES EN AMÉRICA LATINA Y EL CARIBE 2000 - 2022*. <https://www.unocha.org/publications/report/world/panorama-de-los-desastres-en-america-latina-y-el-caribe-2000-2022>
- Villalobos Granadino, P. O. (2021). *Estudio de inundación por efectos del desborde del río Amojú, aplicando los modelos matemáticos HEC-RAS Y HEC-HMS en la ciudad de Jaén, distrito de Jaén provincia de Jaén, departamento de Cajamarca* [Tesis de grado, USAT]. USAT-Tesis. <http://hdl.handle.net/20.500.12423/4312>
- Zúñiga Ramos, J. A. (2020). *Evaluación de Riesgo por Inundación Fluvial de la Ciudad de Huancavelica, Perú* [Tesis, Universidad Nacional de San Cristóbal de Huamanga]. Repositorio Institucional. <http://repositorio.unsch.edu.pe/handle/UNSCH/3908>