

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO
FACULTAD DE CIENCIAS FISICAS, QUIMICAS Y MATEMATICAS
ESCUELA PROFESIONAL DE FISICA



TESIS

**PREDICCIÓN DE LA CONCENTRACIÓN DE RADÓN EN EL
SUELO APLICANDO TÉCNICAS DE MACHINE LEARNING**

PRESENTADO POR:

BACH. RAFAEL EDUARDO PONCE AMANCA
PARA OPTAR AL TÍTULO PROFESIONAL DE
FISICO

ASESOR:

MGT. DAVID TEOFILO VERA VICTORIA

CO-ASESOR:

ING. BRIANT GARCÍA FERNÁNDEZ BACA

CUSCO - PERU

2025

INFORME DE ORIGINALIDAD

(Aprobado por Resolución Nro.CU-303-2020-UNSAAC)

El que suscribe, Asesor del trabajo de investigación/tesis titulada: PREDICCIÓN DE LA CONCENTRACIÓN DE RADÓN EN EL SUELO APLICANDO TÉCNICAS DE MACHINE LEARNING presentado por: RAFAEL EDUARDO PONCE AMANCA con DNI Nro.: 75782255 para optar al título profesional/grado académico de FISICO Informo que el trabajo de investigación ha sido sometido a revisión por 01 veces, mediante el Software Antiplagio, conforme al Art. 6° del **Reglamento para Uso de Sistema Antiplagio de la UNSAAC** y de la evaluación de originalidad se tiene un porcentaje de 6 %.

Evaluación y acciones del reporte de coincidencia para trabajos de investigación conducentes a grado académico o título profesional, tesis

Porcentaje	Evaluación y Acciones	Marque con una (X)
Del 1 al 10%	No se considera plagio.	X
Del 11 al 30 %	Devolver al usuario para las correcciones.	
Mayor a 31%	El responsable de la revisión del documento emite un informe al inmediato jerárquico, quien a su vez eleva el informe a la autoridad académica para que tome las acciones correspondientes. Sin perjuicio de las sanciones administrativas que correspondan de acuerdo a Ley.	

Por tanto, en mi condición de asesor, firmo el presente informe en señal de conformidad y **adjunto** la primera página del reporte del Sistema Antiplagio.

Cusco, 03 de diciembre de 2024



Firma

Mgt. David Teofilo Vera Victoria
Nro. de DNI: 23925699
ORCID del Asesor 0000-0001-6296-1227
Nro. de DNI del Co-Asesor: 44435915
ORCID del Co-Asesor 0000-0002-7460-9308

Se adjunta:

1. Reporte generado por el Sistema Antiplagio.
2. Enlace del Reporte Generado por el Sistema Antiplagio: **oid: 27259:412149637**

Rafael Eduardo Ponce Amanca

Predicción de la concentración de radón en el suelo aplicando tecnicas de machine learning

 Universidad Nacional San Antonio Abad del Cusco

Detalles del documento

Identificador de la entrega

trn:oid:::27259:412149637

Fecha de entrega

3 dic 2024, 8:41 a.m. GMT-5

Fecha de descarga

3 dic 2024, 9:00 a.m. GMT-5

Nombre de archivo

TESIS_RAFAEL_PONCE_ML_CUSCO.pdf

Tamaño de archivo

29.7 MB

102 Páginas

28,664 Palabras

150,774 Caracteres

6% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Filtrado desde el informe

- ▶ Bibliografía
- ▶ Texto citado
- ▶ Texto mencionado
- ▶ Coincidencias menores (menos de 10 palabras)

Exclusiones

- ▶ N.º de coincidencias excluidas

Fuentes principales

- 4%  Fuentes de Internet
- 1%  Publicaciones
- 5%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

No se han detectado manipulaciones de texto sospechosas.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

Agradecimientos

A mi asesor de tesis, Mgt. David Teofilo Vera Victoria, docente principal del departamento de Física de la UNSAAC. Al Ing. Briant García Fernández Baca, Responsable del proyecto Act13 "Monitoreo y caracterización de fallas activas", especialista en neotectónica y fallas geológicas activas del INGEMMET, por sus importantes recomendaciones en temas de geología. Al MSc. César Jheferson Guevara Pillaca, por su constante interés, preocupación y apoyo durante el tiempo de realización de esta investigación, por sus valiosos aportes y recomendaciones para esta tesis, basados en su amplia experiencia como investigador en temas de radón, física nuclear y radiaciones ionizantes.

Dedicatoria

Carpe diem, Paob.

Resumen

La concentración de radón en el suelo representa un riesgo radiológico significativo debido a su capacidad para infiltrarse en interiores y sus afecciones a la salud humana. Tradicionalmente, la medición de radón se ha realizado mediante diversos métodos como detectores de trazas nucleares, cámaras de ionización, espectrometría de centelleo líquido y monitores electrónicos de radón. En este estudio, exploramos la aplicación del machine learning para estimar la distribución espacial del radón en el suelo como un complemento a las mediciones experimentales in situ. Esta investigación tiene como objetivo principal desarrollar un modelo basado en redes neuronales artificiales para el mapeo de la concentración de radón en suelo del Valle del Cusco. Para esto se evaluó la distribución espacial de la concentración de radón en el suelo y su relación con factores geológicos y físicos. Se empleó un Sistema de Información Geográfica (SIG) para integrar datos de geología, geofísica y mediciones de radón. Los datos integrados incluyen variables de entrada (la litología, la proximidad a fallas geológicas, el tamaño de grano y la densidad aparente) y una variable de salida (los datos de mediciones activas de concentración de radón en suelo) estos se utilizaron para entrenar y validar el modelo de predicción. Se utilizaron dos métodos de validación, el error medio absoluto (MAE) y el error cuadrático medio (RMSE). Posterior a ellos se generó un mapeo de concentración de radón basado en las predicciones del modelo. Los resultados muestran que la concentración de radón en el suelo varía significativamente en el área de estudio. Por medio de las mediciones activas se encontraron niveles comprendidos entre 0.1 y 22 (kBq/m^3). Por otro lado, los métodos de validación (RMSE: 0.1810 y MAE: 0.0933) muestran que la red neuronal artificial (RNA) funcionó bien para la predicción de la concentración de radón en suelo, presentando valores bajos de error. Finalmente, el mapa generado permite identificar zonas con mayor potencial de concentración de radón, lo cual es de gran relevancia para la evaluación de riesgos asociados a la exposición a este gas carcinógeno. Además, se encontró una relación significativa entre las zonas de alta concentración de radón en el suelo y presencia de fallas geológicas.

Palabras clave: Radón, Machine Learning, Mapa de radón, Fallas Geológicas, Zonas Propensas.

Abstract

The concentration of radon in soil represents a significant radiological risk due to its ability to infiltrate indoors and its effects on human health. Traditionally, radon measurement has been performed using various methods such as nuclear trace detectors, ionization chambers, liquid scintillation spectrometry and electronic radon monitors. In this study, we explore the application of machine learning to estimate the spatial distribution of radon in soil as a complement to in situ experimental measurements. The main objective of this research is to develop a model based on artificial neural networks for mapping the radon concentration in soil of the Cusco Valley. For this purpose, the spatial distribution of radon concentration in soil and its relationship with geological and physical factors were evaluated. A Geographic Information System (GIS) was used to integrate data from geology, geophysics and radon measurements. The integrated data included input variables (lithology, proximity to geological faults, grain size and bulk density) and an output variable (data from active soil radon concentration measurements) which were used to train and validate the prediction model. Two validation methods were used, the mean absolute error (MAE) and the root mean square error (RMSE). A radon concentration mapping was then generated based on the model predictions. The results show that the radon concentration in soil varies significantly in the study area. Active measurements found levels between 0.1 and 22 (kBq/m^3). On the other hand, the validation methods (RMSE: 0.1810 and MAE: 0.0933) show that the artificial neural network (ANN) performed well for the prediction of radon concentration in soil, presenting low error values. Finally, the generated map allows the identification of areas with higher radon concentration potential, which is of great relevance for the assessment of risks associated with exposure to this carcinogenic gas. In addition, a significant relationship was found between areas of high radon concentration in the soil and the presence of geological faults.

Keywords: Radon, Machine Learning, Radon Map, Geological Faults, Prone Areas.

Índice general

1. Introducción	3
1.1. Planteamiento del Problema	4
1.2. Formulación del Problema	5
1.2.1. Problema General	5
1.2.2. Problemas Específicos	5
1.3. Justificación	5
1.4. Objetivos de la Investigación	6
1.4.1. Objetivo General	6
1.4.2. Objetivos Específicos	6
2. Marco Teórico Conceptual	7
2.1. Antecedentes	7
2.2. Física del Radón	10
2.2.1. Radiactividad Natural	10
2.2.2. Decaimiento radioactivo y cadenas de desintegración radiactiva .	12
2.2.2.1. Desintegración Alfa	13
2.2.2.2. Desintegración beta	14
2.2.3. Equilibrio secular y ecuaciones de Bateman	15
2.2.4. Radón en suelo	16
2.2.5. Propiedades del radón	17
2.2.6. Liberación de radón a partir de granos minerales	17
2.2.7. Distribución de Radio en el suelo y Parámetros Granulo-métricos	19
2.2.8. Proceso de migración de radón	19
2.2.8.1. Emanación de radón	20
2.2.8.2. Transporte de radón	20

2.2.8.3. Exhalación de radón	21
2.2.9. Difusión de radón	22
2.2.9.1. Difusión de bulk	22
2.2.9.2. Difusión de knudsen	23
2.2.10. Radón en viviendas	23
2.2.11. Radón en salud	24
2.3. Radón y Geología	26
2.3.1. Radionuclidos	26
2.3.2. Impacto de la geología en los niveles de radiación	26
2.3.3. Litología	28
2.3.4. Permeabilidad del suelo	29
2.3.5. Porosidad del suelo	31
2.3.6. Granulometría del suelo	31
2.3.7. Geología Estructural	33
2.3.8. Cartografía Geológica y de Radón	35
2.4. Machine Learning	37
2.4.1. Introducción al Machine Learning	37
2.4.1.1. Aprendizaje Supervisado - Supervised Learning	37
2.4.1.2. Aprendizaje no Supervisado - Unsupervised Learning	37
2.4.2. Redes Neuronales Artificiales	38
2.4.2.1. Redes Neuronales Recurrentes (RNN)	38
2.4.2.2. Red de memoria a corto-largo plazo - Long short-term memory (LSTM)	40
2.4.3. Hiperparámetros	41
2.4.3.1. Optimización de Hiperparámetros	41
2.4.3.2. Funciones de Activación	41
2.4.3.3. Tasa de Aprendizaje	41
2.4.3.4. Tamaño de Lote	42
2.4.3.5. Número de Épocas	42
3. Hipótesis y Variables	43
3.1. Hipótesis	43
3.1.1. Hipótesis General	43
3.1.2. Hipótesis Específicas	43

4. Métodos y equipos	45
4.1. Área de estudio	45
4.2. Tipo de Investigación	46
4.3. Impacto de la Investigación	46
4.4. Técnicas de recolección de información	46
4.5. Herramientas de Análisis de la Información	47
4.6. Metodología	47
4.6.1. Medición activa de radón en el gas del suelo	48
4.6.2. Generación de mapas insumo	52
4.6.2.1. Elaboración del mapa de Unidades Litoestratigráficas	52
4.6.2.2. Elaboración del mapa de Granulometría	52
4.6.2.3. Elaboración del mapa de Radón promedio	52
4.6.2.4. Elaboración del mapa de Densidad de Bulk	53
4.6.2.5. Elaboración del mapa de Proximidad a fallas	55
4.6.3. Preprocesamiento de los datos	56
4.6.3.1. Limpieza de datos	57
4.6.3.2. Transformación de datos	58
4.6.3.3. Reducción de dimensionalidad	61
4.6.3.4. División del conjunto de datos	61
4.6.4. Desarrollo de la red neuronal	63
4.6.4.1. Diseño del modelo LSTM	63
4.6.4.2. Entrenamiento del modelo LSTM	66
4.6.5. Evaluación de la red neuronal	66
4.6.5.1. Mean absolute error - MAE	66
4.6.5.2. Root mean square error - RMSE	67
4.6.6. Procedimiento de generación de un mapa de concentración de gas radón en suelo por medio de Machine Learning	67
5. Resultados	71
5.1. Radón en el gas del suelo en la ciudad de Cusco	71
5.1.1. Contexto geológico	71
5.1.2. Contexto geológico estructural	72
5.1.3. Contexto litológico	72
5.1.4. Niveles de concentración en el valle del Cusco	74

5.2. Mapas insumo generados	76
5.2.1. Mapa de Unidades Litoestratigraficas	76
5.2.2. Mapa de Granulometría	78
5.2.3. Mapa de Radón promedio	80
5.2.4. Mapa de Densidad de Bulk	81
5.2.5. Mapa de Proximidad de fallas	82
5.3. Modelo de red neuronal artificial	83
5.3.1. Estructura de la red	83
5.3.2. Métricas de evaluación	83
5.3.3. Predicciones y valores observados experimentalmente	84
5.4. Mapa de radón generado con Machine Learning	85
6. Discusiones y Conclusiones	89
6.1. Discusiones	89
6.1.1. Gas radón como herramienta para la detección de fallas geológicas	89
6.1.2. Concentración de gas radón en suelo y en interiores	92
6.2. Conclusiones	92
6.2.1. Limitaciones	93
6.2.2. Recomendaciones	93

Índice de cuadros

2.1. Resumen de la exposición al radón en lugares de trabajo distintos de las minas (UNSCEAR, 2022)	11
2.2. Propiedades del radón, (Baskaran, 2016a).	17
2.3. Coeficiente de emaciación de radón en medios, (Cinelli et al., 2019). . .	18
2.4. Coeficiente de difusión de radon en medios (Cothorn and Smith, 1987) .	22
2.5. Nucleidos primordiales, modificado de (Cinelli et al., 2019)	26
2.6. Nivel de radiación típica en suelo	27
2.7. Formación de rocas y contenido de radionuclidos	28
2.8. Clasificación de la permeabilidad	29
2.9. Coeficiente de permeabilidad de suelos	30
2.10. Rango de porosidad según el tipo de suelo	31
4.1. Estadísticas descriptivas, rango intercuartil (IQR), rango intercuartil superior y rango intercuartil inferior	53
4.2. Estadísticas descriptivas de variables de entrada	56
4.3. Tabla del conjunto de datos posterior a la limpieza de data	58
4.4. Conjunto de datos normalizado	59
4.5. Tabla de codificación por etiquedas de la litología del valle del cusco, (Carlotto et al., 2011)	60
4.6. Conjunto de datos obtenido después del pre procesamiento	62
5.1. Litología departamento del Cusco	73
5.2. Tabla de unidades geológicas de la provincia del Cusco; (Carlotto et al., 2011)	77
5.3. Tabla de clasificación granulométrica de la provincia de Cusco	79
5.4. Estructura de la red neuronal	83
5.5. Resultados de las métricas de evaluación para la RNA	83

Índice de figuras

2.1. Mapa de radiación gamma natural España, (Suarez et al., 2000).	10
2.2. Fuentes de radiación natural (Cinelli et al., 2019).	11
2.3. Cadena de desintegración radiactiva - Uranio 238 (Cinelli et al., 2019) .	12
2.4. Decaimiento alfa Rn-222	13
2.5. Decaimiento alfa, (Cinelli et al., 2019).	13
2.6. Desintegración β , (Brosed and de Física Médica, 2011)	14
2.7. Desintegración beta negativo (Cinelli et al., 2019)	14
2.8. Desintegración beta positivo (Cinelli et al., 2019).	14
2.9. Equilibrio secular U-238 y Th-234 (Gilmore, 2008)	15
2.10. Mapa de exposición potencial de radón España, (García-Talavera San Miguel et al., 2013).	16
2.11. Formación de radón en granos minerales (Otton, 1992).	18
2.12. Comparación de emanación de Rn222 para distribución homogénea y distribución superficial de Ra226. (Baskaran, 2016b)	19
2.13. Migración de radón, (Cinelli et al., 2019).	20
2.14. Modelo del fenómeno de emanación del radón, (Sakoda et al., 2011). . .	21
2.15. Coeficiente de difusión de bulk en medios (Nielson and Rogers, 1982). .	22
2.16. Coeficiente de difusión de Knudsen en medios (Nielson and Rogers, 1982).	23
2.17. Fuentes y vías de entrada del radón en interiores (Cothorn and Smith, 1987)	24
2.18. Niveles de concentración de radón por país (WHO, 2009)	25
2.19. Fuentes de los principales radionucleidos naturales de las series de desintegración 238U, 232Th y 40K; Fuente primaria: lecho rocoso (nivel 1); fuente secundaria: suelo (nivel 2); fuentes adicionales: discontinuidades (por ejemplo, fallas geológicas, nivel 3); fuente antropogénica: materiales de construcción y contaminación ambiental (nivel 4). (Cinelli et al., 2019).	27
2.20. Clasificación de rocas	28

2.21. Nivel de permeabilidad de suelos basado en su estructura, (Cinelli et al., 2019).	29
2.22. Distribución de radón en suelo, por (Otton, 1992)	30
2.23. Análisis mecánicos de J. A. UDDEN con los nombres de los sedimentos adecuados para la clasificación (Wentworth, 1922).	32
2.24. Mapa neotectónico región Cusco, (Benavente et al., 2013).	34
2.25. Mapa geológico de los cuadrángulos 27r, 27s, 28r, 28s correspondiente a la región Cusco, (Carlotto et al., 2011).	35
2.26. Mapa de U en suelo Europa, (Cinelli et al., 2019).	36
2.27. Mapa de Th en suelo Europa, (Cinelli et al., 2019).	36
2.28. Diagrama red neuronal artificial, (Bishop and Nasrabadi, 2006).	38
2.29. Diagrama red neuronal recurrente (RNN), (Zhang et al., 2021).	39
2.30. Diseño de patrones de una RNN(Goodfellow et al., 2016)	39
2.31. Diagrama de celda de memoria, (Goodfellow et al., 2016).	40
2.32. Compuertas modelo LSTM , (Zhang et al., 2021).	40
2.33. Diagrama de flujo de entreno iterativo para un modelo con diferentes hiperparámetros, (Zhang et al., 2021).	41
4.1. Mapa político provincia Cusco.	45
4.2. Esquema de la metodología	48
4.3. Monitores de radón	48
4.4. Zona de estudio valle del Cusco	49
4.5. Esquema para medir gas de suelo con el monitor RTM 2200	50
4.6. Mediciones de radón en gas de suelo - método directo	51
4.7. Diagrama de cajas, concentración de radón en suelo	53
4.8. Raster densidad de bulk, (Poggio et al., 2021)	54
4.9. Relleno de valores sin datos (en gris) de un ráster	55
4.10. Fallas geológicas activas identificadas Benavente et al. (2013)	55
4.11. Código para la eliminación de valores nulos	57
4.12. Código para la eliminación de valores duplicados	57
4.13. Código para la normalización de los datos, rango de 0 a 1	59
4.14. Relleno de valores sin datos (en gris) de un ráster	62
4.15. Código para la división del conjunto de datos	63
4.16. Arquitectura modelo LSTM	63
4.17. Arquitectura LSTM apilada; (Brownlee, 2017)	64

4.18. Arquitectura LSTM bidireccional; (Brownlee, 2017)	64
4.19. Estructura de red neuronal profunda	65
4.20. Implementacion del codigo para generar la red	65
4.21. Diseño red LSTM	66
4.22. Esquema para la obtención de las predicciones	68
4.23. Esquema para la elaboración del mapa de radón empleando las predicciones	68
4.24. Provincia del Cusco con grilla de 1Km x 1Km	68
5.1. Mapa geológico de los cuadrangulos 27r, 27s, 28r, 28s correspondiente a la región Cusco, (Carlotto et al., 2011).	71
5.2. Mapa Neotectonico de la provincia del Cusco	72
5.3. Mapa litológico del Departamento del Cusco (Vílchez et al., 2020) . . .	73
5.4. Diagrama de cajas de la concentración de radón en el gas del suelo en el Valle del Cusco	74
5.5. Gráfico de cuantil-cuantil de las concentraciones de radon en suelo del valle del Cusco	75
5.6. Mapa Litoestratigrafico de la provincia de Cusco	76
5.7. Mapa Granulometrico Cusco	78
5.8. Mapa radón base Cusco	80
5.9. Mapa Densidad de bulk de suelo Cusco	81
5.10. Mapa proximidad fallas Cusco	82
5.11. Mapa generado por machine learning y valores experimentales	84
5.12. Esquema de grillado de la provincia del Cusco	85
5.13. a) Mapa de concentración de radón distrito de Ccorca, b) Mapa de concentración de radón distrito de Poroy	86
5.14. a) Mapa de concentración de radón distrito de Cusco, b) Mapa de concentración de radón distrito de Wanchaq	86
5.15. a) Mapa de concentración de radón distrito de Santiago, b) Mapa de concentración de radón distrito de San Sebastián	87
5.16. a) Mapa de concentración de radón distrito de San Jerónimo, b) Mapa de concentración de radón distrito de Saylla	87
5.17. Mapa de concentración de radón en suelo de la provincia del Cusco por medio de Machine Learning	88
6.1. Histograma de la distribución de concentración de radón	89
6.2. Mapa de fallas y puntos de muestreo mediciones activas	90
6.3. Mapa de fallas y áreas con niveles anómalos generados por ML	91

Capítulo 1

Introducción

El radón es un gas radiactivo que se encuentra naturalmente en el medio ambiente, se origina principalmente en el subsuelo, como producto la cadena de desintegración radiactiva del uranio (U-238), principalmente del isótopo de radio (Ra-226) presentes en rocas de la corteza terrestre (Cinelli et al., 2019).

Diversos factores influyen en la concentración del radón en el suelo, como la geología, la permeabilidad del suelo, la porosidad del suelo, la distribución átomos de radio entre otros (Tareen et al., 2019). Por lo tanto, conocer el potencial de radón geogenico es de gran importancia en la evaluación de riesgos radiológicos. Estos factores influyen directamente en la liberación y migración del radón hacia la superficie del suelo y su transporte a la atmósfera. Por otro lado, la variabilidad de los parámetros meteorológicas afectan en el transporte y la exhalación del radón hacia la atmósfera (Elío et al., 2023). Comprender y considerar la influencia de los factores geofísicos, geoquímicos y meteorológicos es crucial para desarrollar modelos robustos para la estimación de concentración de radón en suelo (Rezaie et al., 2023).

Por otro lado, en los últimos años el uso de técnicas de Machine Learning como las redes neuronales artificiales, ha demostrado ventajas significativas frente a los métodos tradicionales de análisis de datos (Giustini et al., 2019). Estas técnicas destacan por su capacidad para aprender patrones complejos y no lineales, manejar grandes volúmenes de información y adaptarse a nuevos datos. Esto las convierte en herramientas particularmente útiles en campos donde múltiples variables se encuentran interrelacionadas, como es el caso de la predicción de la concentración de radón en el suelo. Dada la gran variedad de factores que influyen en el comportamiento del radón, las redes neuronales permiten integrar de manera eficaz parámetros geológicos, geoquímicos, físicos y meteorológicos.

Esta investigación se llevó a cabo en la región del Cusco, específicamente en la provincia del Cusco. Se realizó el apoyo en las mediciones puntuales de la concentración de radón en el suelo mediante el método activo ¹ en los distritos que integran la provincia.

¹Método activo: Involucra uso de dispositivos electrónicos de monitoreo continuo que proporcionan mediciones en tiempo real de la concentración de gas radón. A diferencia de los métodos pasivos, que requieren un período de tiempo más largo para obtener resultados.

La información de estas mediciones corresponde a la actividad 13 (ACT13) del Instituto Geológico, Minero y Metalúrgico (INGEMMET) “ESTUDIO Y MONITOREO DE FALLAS ACTIVAS” en colaboración con la Pontificia Universidad Católica del Perú (PUCP). Posteriormente se estimó la concentración de radón en el suelo por medio del desarrollo de un modelo de redes neuronales artificiales basado en parámetros geológicos y físicos del suelo.

1.1. Planteamiento del Problema

El radón es un gas radiactivo natural que se origina en el suelo, se transporta hacia la atmósfera y se puede acumular en ambientes cerrados. La exposición crónica a este gas representa un riesgo significativo para la salud humana ya que puede causar daños en el ADN incrementando el riesgo de cáncer. Esto genera la necesidad de evaluar su concentración de radón en el gas del suelo para identificar zonas potencialmente peligrosas en el Valle del Cusco. Pero la imposibilidad de realizar mediciones de concentración de radón en un área extensa siguiendo una grilla de alta resolución (100x100 m) constituye una limitación significativa para obtener un mapa de concentración preciso. Para superar esta barrera y estimar los valores de radón en las zonas sin datos, se propone el empleo de técnicas de Machine Learning, las cuales permitirán modelar la distribución espacial del radón.

La concentración de radón en el suelo varía según factores geológicos, geofísicos y geoquímicos, junto con parámetros meteorológicos. Debido a la gran cantidad de factores se complica la precisión con la que los modelos pueden actuar en diferentes contextos. A pesar de los avances en la comprensión de estos factores, existen limitaciones significativas en las metodologías tradicionales de modelado. La mayoría de los enfoques convencionales no logran capturar de manera efectiva las interacciones no lineales entre las variables. Los métodos basados en geoestadística tradicional, utilizados para predecir las concentraciones de radón, presentan limitada precisión y capacidad de generalización, especialmente en áreas donde la variabilidad de los factores es alta. En contraste, las técnicas de Machine Learning han demostrado un mejor desempeño en estos estudios, ofreciendo mejor precisión al analizar grandes conjuntos de datos geoespaciales y ambientales.

En este contexto, es fundamental investigar cómo las técnicas de Machine Learning pueden ser aplicadas para desarrollar modelos robustos que estimen la concentración de radón en el suelo a partir de datos geoespaciales, contribuyendo así a una mejor comprensión y gestión del riesgo radiológico asociado a este gas carcinógeno .

1.2. Formulación del Problema

1.2.1. Problema General

¿Como se puede utilizar un modelo de predicción basado en redes neuronales artificiales para realizar un mapeo de concentración de radón en el suelo en el Valle del Cusco?

1.2.2. Problemas Específicos

- ¿Cuáles son niveles de concentración de radón en el suelo en el Valle del Cusco?
- ¿Qué variables geológicas y físicas son fundamentales para modelar la concentración de radón en el suelo en el Valle del Cusco?
- ¿Cómo un modelo de redes neuronales artificiales puede ayudarnos a estimar los valores de concentración de radón en el suelo en zonas sin mediciones directas?

1.3. Justificación

La provincia del Cusco se asienta sobre diversas formaciones geológicas que se encuentran constituidas por rocas sedimentarias, metamórficas e ígneas, con presencia de fallas geológicas activas como la falla de Tambomachay, la falla Cusco, el sistema de fallas Pachatusan entre otras. Por otro lado, en esta zona existen antiguas formaciones volcánicas, abundancia de galerías filtrantes y aguas termales. Dado este contexto resulta importante realizar un estudio en la provincia del Cusco ya que presenta las condiciones que favorecen la presencia de niveles elevados de concentración radón en el suelo.

Dada la acelerada expansión urbana, la elevada densidad poblacional y la ausencia de una planificación territorial integral en la zona, resulta crucial llevar a cabo un estudio detallado que permita comprender y abordar de manera efectiva y aplicable los riesgos asociados a la concentración de radón en suelo. El desarrollo de un modelo de redes neuronales artificiales permitirá un mapeo mas amplio de la concentración de radón en el suelo a partir de mas de 500 mediciones in situ. Esto facilitara identificar zonas de alto potencial de riesgo lo que proporciona una herramienta para la implementación de estrategias de protección radiológica ambiental, la planificación urbana y la toma de decisiones en la gestión de riesgo de desastres promoviendo entornos seguros y saludables.

1.4. Objetivos de la Investigación

1.4.1. Objetivo General

Desarrollar un modelo de predicciones de machine learning basado en redes neuronales artificiales para el mapeo de la concentración de radón en suelo del Valle del Cusco.

1.4.2. Objetivos Específicos

- Realizar mediciones in situ de las concentraciones de radón en suelo en el Valle del Cusco, siguiendo un protocolo estandarizado.
- Elaborar mapas de los principales parámetros geológicos y físicos del Valle del Cusco, utilizando sistemas de información geográfica (SIG).
- Generar un mapa de concentración de radón en suelo a partir de las estimaciones obtenidas por el modelo de redes neuronales artificiales

Capítulo 2

Marco Teórico Conceptual

2.1. Antecedentes

El radón es un elemento químico descubierto a finales del siglo XIX en Alemania de naturaleza incolora, inodora e insípida (USEPA, 2002), su número atómico es el 86 (Partington, 1957). El isótopo más común del radón es el radón-222, tiene un periodo de semidesintegración de aproximadamente 3,823 días (Cothorn and Smith, 1987) y continuamente puede ser encontrado en el suelo o en las rocas, por lo que resulta adecuado para el estudio geológico (Feng et al., 2022).

El radón (Rn-222) forma parte de la cadena de desintegración del Uranio (U-238). En este contexto existe un segundo isótopo más importante, el (Rn-220), conocido como torón que posee un periodo de semidesintegración de 55 segundos (Cothorn and Smith, 1987), formando parte de la cadena de desintegración del Torio (Th-232). El resto de estos isótopos pasan a ser irrelevantes para estudios medioambientales.

Se conoce que la principal vía de exposición humana al radón es la inhalación en el entorno cotidiano (Saeed and Hassan, 2015). Sin embargo, también es posible estar expuesto a este gas mediante el contacto directo o la ingesta de su progenie (WHO, 2016). El impacto de la exposición a una dosis causada por la progenie del radón-222 en contacto directo con tejido humano conduce a severos problemas de salud (Baldwin and Grantham, 2015; Doss, 2018; Hughes et al., 2022). De hecho, el radón es la principal causa de cáncer de pulmón entre los no fumadores según la Agencia de Protección del Medio Ambiente de los Estados Unidos (U.S. Environmental Protection Agency (USEPA), por sus siglas en inglés) (USEPA, 2002), y catalogado como un carcinógeno humano por la Agencia Internacional para la Investigación del Cáncer (International Agency for Research on Cancer (IARC), por sus siglas en inglés) (IARC et al., 1988). Un ejemplo notable es Estados Unidos, donde aproximadamente el 10 % de las muertes anuales por cáncer de pulmón están relacionadas con la exposición al radón (US Cancer Statistics et al., 2016).

Por otro lado, al tratarse de un gas radioactivo, el radón representa una amenaza biológica para las células debido a las radiaciones emitidas que resulta de su desintegración a lo largo del tiempo. Durante este proceso, se emiten partículas alfa, beta y

gamma (Valcarce et al., 2022). Como consecuencia, una exposición prolongada a este gas es claramente perjudicial para la salud humana.

Dado el riesgo que supone la exposición al radón para la población, existen diversas normativas enfocadas en la medición de su concentración, la mitigación de sus efectos y las acciones de reparación cuando son necesarias. Estas normativas están establecidas en las Normas Básicas de Seguridad de la Agencia Internacional de Energía Atómica (International Atomic Energy Agency (IAEA), por sus siglas en inglés) (IAEA, 2015, 2014) y la Comunidad Europea de Energía Atómica (European Atomic Energy Community (EURATOM), por sus siglas en inglés) (ESR, 2015). Así mismo a nivel nacional, la entidad encargada de regular estos estándares es el Instituto Peruano de Energía Nuclear (IPEN), que ha establecido políticas de seguridad basadas en los estándares recomendados por la Comisión Internacional de Protección Radiológica (International Commission on Radiological Protection (ICRP), por sus siglas en inglés). Estas políticas están plasmadas en el reglamento de seguridad radiológica de 1997 (IPEN, 1997).

En los estudios medioambientales, el objetivo es identificar patrones espacio-temporales (como tendencia, periodicidad, auto correlación, anomalías) y comprender su relación con el fenómeno estudiado. A lo largo de los años, el radón ha demostrado ser una herramienta valiosa en una amplia gama de investigaciones, por ejemplo: estudios atmosféricos (Gregorič et al., 2020), cálculo de emisiones de gases de efecto invernadero (Levin et al., 2021), caracterización de contaminación en el suelo (De Miguel et al., 2018), entre otras aplicaciones científicas.

La concentración de radón en el suelo ha sido objeto de creciente interés en los últimos años, debido a su relación con la geología y las condiciones meteorológicas. Estudios previos han demostrado que la geología de una región desempeña un papel crucial en la liberación de este gas (Mose and Mushrush, 1997; Elío et al., 2023). Las propiedades del suelo, como la granulometría, la porosidad, la humedad, la densidad aparente y la permeabilidad, así como las características litológicas y geoquímicas de las formaciones subyacentes, influyen significativamente en la concentración de radón en el suelo (Thumvijit et al., 2020). Además, la presencia de altos niveles de uranio y radio en ciertos tipos de suelos y rocas, elementos progenitores del radón, que intensifican este fenómeno (Baskaran, 2016a). Por otro lado, las condiciones meteorológicas también ejercen una influencia notable en la concentración y dispersión del radón en el suelo (Zafrir et al., 2013; Garavaglia et al., 2000). Factores como la temperatura, la humedad y la presión atmosférica pueden modificar la tasa de emanación y el transporte del gas a través de los poros del suelo.

En investigaciones previas, se han utilizado diferentes enfoques para evaluar la concentración de radón en el suelo, como detectores de trazas nucleares (Liza Necio-sup, 2017; Pereyra et al., 2015), cámaras de difusión con detectores de trazas nucleares (Guevara, 2019), mediciones directas utilizando monitores de radón (Oni et al., 2022; Hosoda et al., 2020), análisis de isótopos de radón (Baskaran, 2016a), modelos de simulación (Zheng et al., 2012), trazador para la exploración geoquímica (Baskaran, 2016a), estudios sísmicos (Tareen et al., 2019; Chen et al., 2018), peligro para la salud humana en el medio ambiente (Elío et al., 2017, 2023). Estos estudios han proporcionado información valiosa sobre la distribución espacial y temporal del radón en el suelo en

diferentes regiones geográficas.

Paralelamente, el campo de la inteligencia artificial ha aportado avances significativos en el estudio de radón en el suelo. En particular las redes neuronales artificiales han demostrado ser herramientas poderosas para predecir y modelar la concentración de gas radón (Tareen et al., 2019; Zheng et al., 2012). Las redes neuronales sobresalen en el manejo de grandes conjuntos de información, la detección de patrones y la predicción de fenómenos complejos lo que las convierte en una herramienta prometedora en la investigación de radón en suelo. (Zhang et al., 1998; Hosoda et al., 2020; Duong et al., 2021).

Sin embargo, se requiere un enfoque más integrado que considere tanto la influencia geológica como las condiciones meteorológicas para comprender mejor la distribución y los factores que afectan la concentración de radón en el suelo. Además, la aplicación de modelos de redes neuronales en este contexto específico aún necesita ser explorada y validada.

2.2. Física del Radón

2.2.1. Radiactividad Natural

La radiactividad en general son las transformaciones nucleares espontáneas de átomos inestables que dan lugar a la formación de nuevos elementos (Cember, 2009). Se refiere a la radiación ionizante emitida de forma espontánea por ciertos elementos presentes en el medio ambiente. Esta radiación es el resultado de la desintegración de átomos inestables en elementos más estables. Hay tres tipos principales de radiación ionizante emitida en procesos de radiactividad natural: radiación alfa, beta y gamma. Los niveles de estas radiaciones varían según las características geológicas de cada región y son representados en mapas específicos, como el mapa de radiación gamma natural de España (Fig. 2.1).

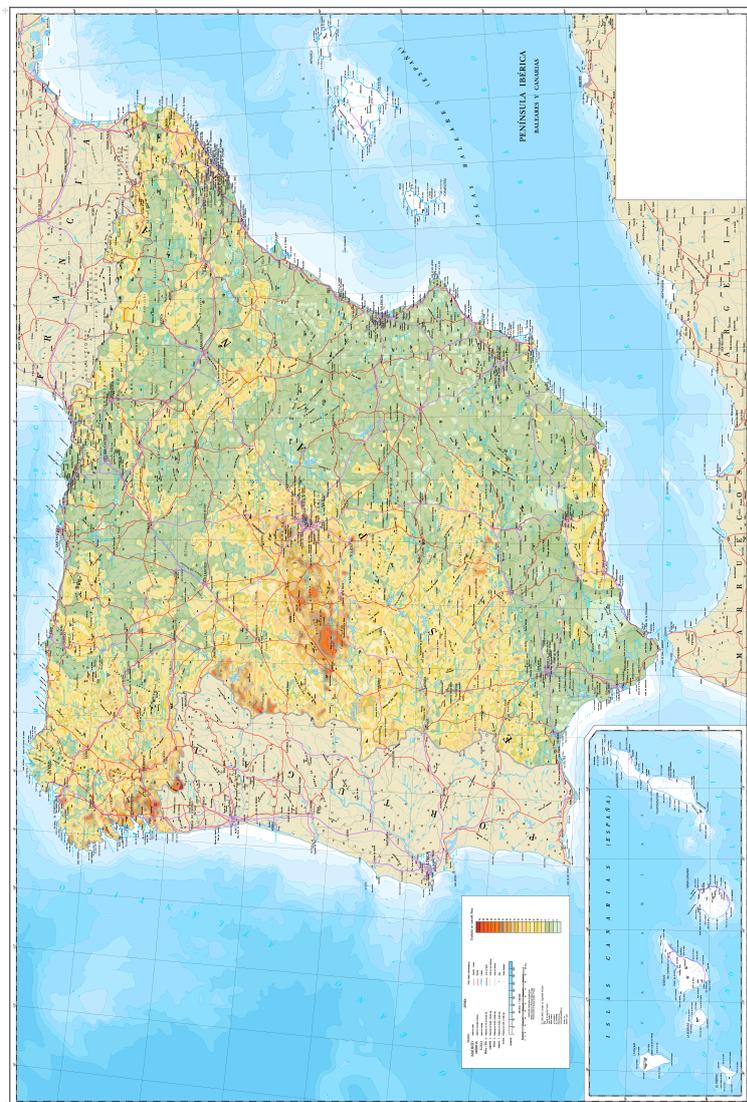


Figura 2.1: Mapa de radiación gamma natural España, (Suarez et al., 2000).

La radiactividad natural está presente en numerosos materiales, como el uranio (U-238), el torio (Th-232), el potasio (K-40) y sus productos de desintegración (DOE, United states, 1993). Estos materiales se encuentran en rocas, suelos, agua e incluso en el cuerpo humano. La exposición a la radiactividad puede tener dos orígenes principales como se observa en la Figura 2.2a, de origen artificial, como los materiales radiactivos utilizados en la medicina e industria que representan un 20 % y de origen natural, como la radiación cósmica proveniente del espacio exterior, por ingestión y radón que representa un 80 %. Dentro de estas ultimas el radón figura como la mayor fuente de radiación natural con un 52 %, la tabla 2.1 muestra un resumen de la exposición a radón en diferentes medios.

Fuente	Practica	Promedio Rn ((Bq/m ³)), (rango)	Promedio anual dosis efectiva (mSv), (rango)
Bajo tierra	Cueva-espectáculo/ Mina-turística	3090 (20 ->30000)	5.1 (0.01 - 50)
	Balneario termal	250 (30 -2810)	3.4 (0.2 - 18)
	Metro/túnel/aparcamiento	28 (5- 85)	0.2 (0.04 - 0.6)
	Laboratorio/almacén/ bodega/oficina	400 (20 - 8660)	2.7 (0.1 - 8.4)
	En superficie	Escuela/universidad/ hospital	180 (10 - 1360)
En superficie	Edificio de oficinas/ empresas	74 (23 - 273)	0.5 (0.2 - 2.0)
	Fábrica	66 (10 - 220)	0.43 (0.1 - 1.3)
	Almacén/bodega	190 (10 - 860)	1.4 (0.1 - 6.2)
	Planta acuática/ criadero de peces	793 (20 - 2720)	1.4 (0.01 - 4.7)

Tabla 2.1: Resumen de la exposición al radón en lugares de trabajo distintos de las minas (UNSCEAR, 2022)

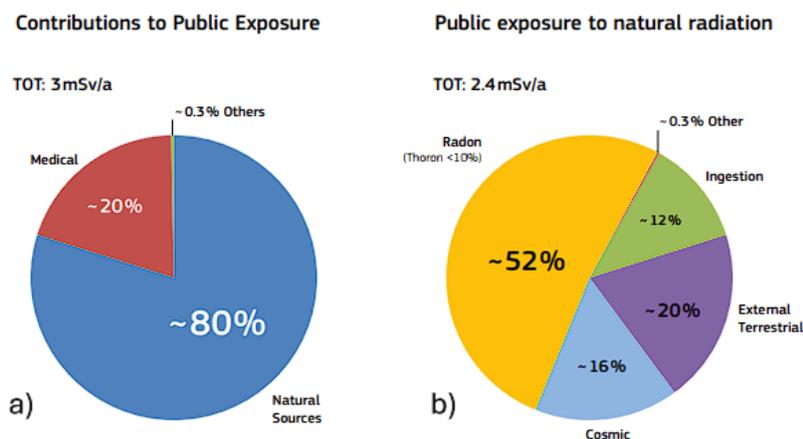


Figura 2.2: Fuentes de radiación natural (Cinelli et al., 2019).

La exposición a la radiactividad natural se encuentra ligada a diversos factores, como la geolocalización, geología, altitud, estilo de vida entre otros. La dosis promedio anual mundial es de aproximadamente 2.4 milisieverts (mSv) (UNSCEAR, 2000).

2.2.2. Decaimiento radioactivo y cadenas de desintegración radiactiva

El decaimiento radioactivo es un proceso natural. Por medio de este los núcleos atómicos inestables emiten partículas o radiación electromagnética para alcanzar un nuevo estado mas estable. Esta transformación se origina como consecuencia de la inestabilidad de ciertos núcleos atómicos. Estos tienen una proporción desequilibrada de protones y neutrones dentro del núcleo (Steiner, 2016). Para lograr una configuración mas estable, estos núcleos atómicos inestables emiten diversas partículas subatómicas, como partículas alfa (equivalentes a núcleos de helio), partículas beta (electrones o positrones) o radiación gamma (ondas electromagnéticas).

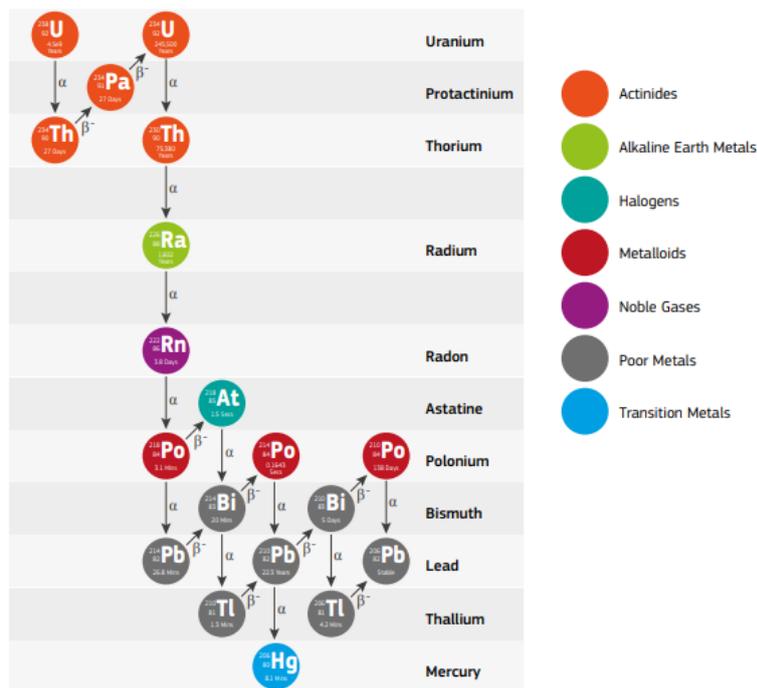


Figura 2.3: Cadena de desintegración radiactiva - Uranio 238 (Cinelli et al., 2019)

El decaimiento radioactivo esta representado por un modelo matemático. La ley de decaimiento exponencial. Esta describe la disminución gradual de la cantidad de átomos radiactivos en una muestra con el paso del tiempo. En este proceso cada núcleo posee una probabilidad constante de desintegrarse, esto implica que el decaimiento no es lineal, sino exponencial. La ecuación 2.1 expresa la cantidad de núcleos radiactivos en un instante de tiempo (t).

$$N(t) = N_0 e^{-\lambda t} \quad (2.1)$$

La ley de decaimiento exponencial relaciona el número de núcleos radiactivos en un instante dado ($N(t)$), el número inicial de núcleos (N_0), la constante de desintegración (λ) y el tiempo transcurrido (t). Dos conceptos clave asociados a esta ley son el periodo de semidesintegración ($t_{1/2}$) y la constante de desintegración (λ). La constante de desintegración λ indica la probabilidad de que un núcleo radiactivo se desintegre por unidad de tiempo y el periodo de semidesintegración el tiempo necesario para que la

mitad de los núcleos radiactivos iniciales se desintegren. Ambas se relacionan mediante la ecuación 2.2.

$$t_{1/2} = \frac{\ln(2)}{\lambda} \quad (2.2)$$

En la naturaleza podemos encontrar tres cadenas de desintegración radiactiva: la del Uranio 238, Torio 232 y Actinio 235. En la Figura 2.3 se observa la cadena del Uranio-238: Comienza con el Uranio-238 y pasa por una serie de elementos como Torio-234, Protactinio-234, Uranio-234, Torio-230, Radio-226 y, finalmente, genera el Radón-222. El proceso continúa hasta llegar a un núcleo estable de Plomo-206. La cadena del Uranio-238 es la que representa un problema para salud humana dado que en su decaimiento se encuentra el gas Radón-222.

2.2.2.1. Desintegración Alfa

El decaimiento alfa es un fenómeno característico de elementos muy pesados frecuentemente con un $Z > 90$ como el uranio o radio. Estos elementos, dada su composición nuclear inestable, experimentan un proceso de desintegración en el que emiten partículas alfa.

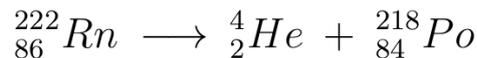


Figura 2.4: Decaimiento alfa Rn-222

A medida que la partícula alfa es liberada, el núcleo original experimenta una disminución en su número de protones y neutrones. Como se observa en la Figura 2.5 en esta forma de desintegración, el núcleo “padre” libera una partícula alfa (núcleo de helio altamente energético), que está formada por dos protones y dos neutrones. Como resultado de esta desintegración, el núcleo ‘hijo’ o descendiente se forma en un estado excitado o directamente en su estado fundamental (Brosed and de Física Médica, 2011).

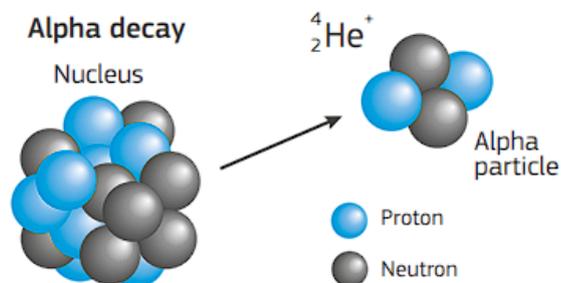


Figura 2.5: Decaimiento alfa, (Cinelli et al., 2019).

Un ejemplo de esto en nuestro caso (Fig.2.4). En esta reacción, el Rn-222, que tiene 86 protones y 136 neutrones, se desintegra para formar el Po-218, con 84 protones

y 134 neutrones. La pérdida de dos protones y dos neutrones explica la disminución de 4 unidades en el número de masa (de 222 a 218) y de 2 unidades en el número atómico (de 86 a 84).

2.2.2.2. Desintegración beta

La desintegración beta es un proceso de decaimiento radiactivo en el cual un núcleo atómico inestable emite una partícula beta (un electrón o un positrón) con el fin de alcanzar un estado más estable, estos procesos están descritos por 2.9.

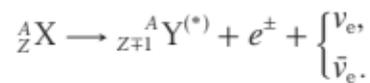


Figura 2.6: Desintegración β , (Brosed and de Física Médica, 2011)

En el caso de una desintegración beta negativa (β^-) (Fig.2.7), se da la emisión de una partícula beta negativa (un electrón) y un antineutrino durante la transformación de un neutrón en un protón. Como resultado, el número de protones del núcleo aumenta en uno, transformando el elemento en otro con un número atómico mayor.

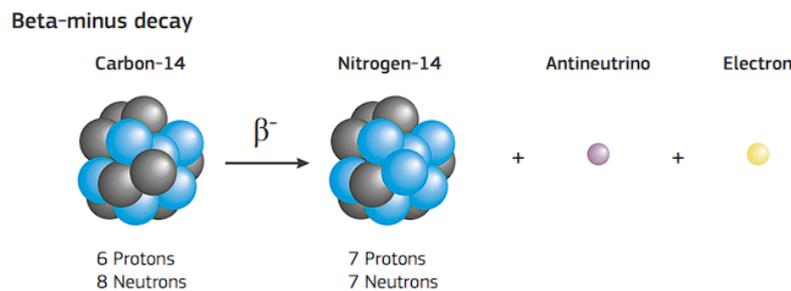


Figura 2.7: Desintegración beta negativo (Cinelli et al., 2019)

En la desintegración beta positiva (β^+) (Fig.2.8), se da la emisión de un positrón y un neutrino durante la transformación de un protón en un neutrón.

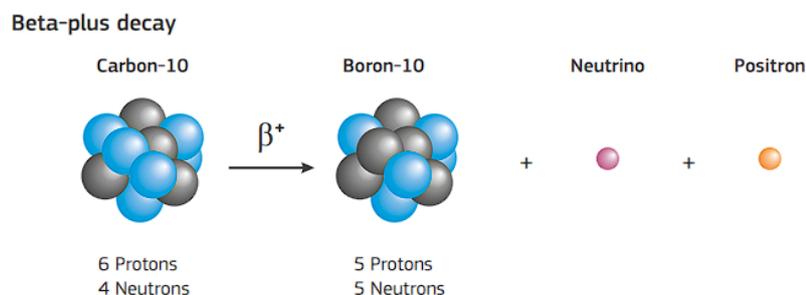


Figura 2.8: Desintegración beta positivo (Cinelli et al., 2019).

2.2.3. Equilibrio secular y ecuaciones de Bateman

El equilibrio secular y las ecuaciones de Bateman, son fundamentales para entender las cadenas de desintegración radiactiva. El equilibrio secular es un caso especial que ocurre en las cadenas de desintegración cuando un isótopo padre tiene una vida media mucho mayor que la de sus isótopos hijos (Soria, 2020).

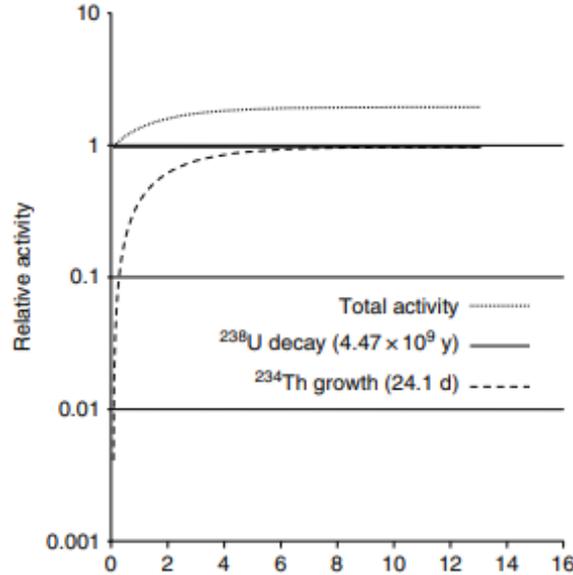


Figura 2.9: Equilibrio secular U-238 y Th-234 (Gilmore, 2008)

Para esto las ecuaciones de Bateman muestran que, con el tiempo, el número de núcleos hijos en la cadena tiende a estabilizarse en una proporción constante con respecto al padre. Así para un equilibrio secular en una cadena de dos miembros (un isótopo padre A_1 y un hijo A_2), con $\lambda_1 \ll \lambda_2$ (es decir, A_1 se desintegra muy lentamente en comparación con A_2), las ecuaciones de Bateman se aproximan a:

$$N_2(t) \cong \frac{\lambda_1}{\lambda_2} N_1(t) \quad (2.3)$$

Las ecuaciones de Bateman están directamente relacionadas con el concepto de equilibrio secular en las cadenas de desintegración radiactiva. Las ecuaciones de Bateman describen cómo evoluciona la cantidad de cada isótopo en una cadena de desintegración con el tiempo. Para esto una cadena de desintegración radiactiva simple donde un isótopo A_1 se desintegra en A_2 , que a su vez en A_3 sucesivamente hasta A_n con una constante de desintegración λ_i para cada A_i se puede escribir como ecuación 2.4 para el isótopo padre.

$$N_1(t) = N_1(0)e^{-\lambda_1 t} \quad (2.4)$$

De acuerdo con lo desarrollado en (Soria, 2020), para cualquier elemento A_i en la cadena después de cierto tiempo (t) una expresión generalizada se representa mediante la ecuación 2.5.

$$N_i(t) = N_1(0) \sum_{j=1}^i \left(\prod_{k=1}^i \frac{\lambda_k}{\lambda_k - \lambda_j} \right) e^{-\lambda_j t} \quad (2.5)$$

Un ejemplo común es el uranio-238, al desintegrarse, origina una cadena de desintegraciones sucesivas que involucra diversos isótopos de torio y radio. Estos isótopos, a su vez, también se desintegran en otros isótopos radiactivos hasta llegar a isótopos estables. Durante este proceso, las concentraciones de los isótopos intermedios permanecen constantes. Dado que su tasa de desintegración es igual a la tasa de producción a pesar de que los átomos individuales continúan desintegrándose.

2.2.4. Radón en suelo

El radón en suelo se genera de manera natural a partir de la desintegración del uranio presente en el suelo y las rocas (Otton, 1992). El radón está catalogado como una de las principales fuentes de exposición a la radiación ionizante para los seres humanos.

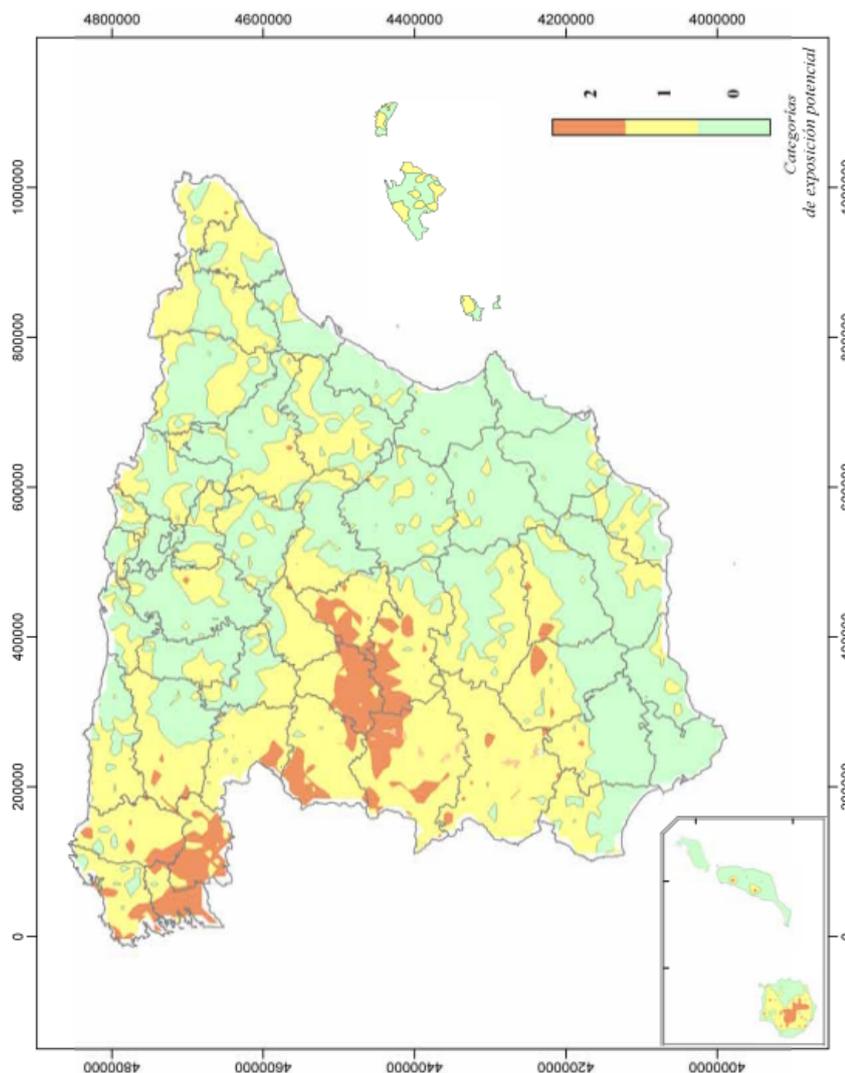


Figura 2.10: Mapa de exposición potencial de radón España, (García-Talavera San Miguel et al., 2013).

Los niveles presentes de concentración de radón en el suelo generalmente son muy variables. Evidenciando áreas de mayor o menor concentración como se observa en la Figura 2.10 correspondiente al mapa de radón de España. Esta concentración se mide en Becquerelios por metro cubico (Bq/m^3). Existen diversos factores que influyen la concentración de este gas en el suelo como por ejemplo la composición geológica de la zona, la permeabilidad del suelo, las condiciones hidrológicas, la presencia de fallas geológicas entre otros. La mineralogía de las rocas y minerales presentes en la zona juega un papel crucial, ya que la geometría cristalina determina la facilidad con la que los átomos de radón pueden escapar de la estructura cristalina (Cinelli et al., 2019). Algunas formaciones geológicas, como las rocas graníticas, tienen una mayor concentración de uranio y, por lo tanto, pueden liberar mas radón al suelo

2.2.5. Propiedades del radón

El radón es un gas radiactivo incoloro e inodoro (USEPA, 2002), tiene una densidad mayor que la del aire y es altamente soluble en líquidos, como el agua. Al ser un gas noble tiene una configuración electrónica estable en su capa de electrones más externa, lo que relativamente lo vuelve inerte y poco propenso a reaccionar químicamente con otras sustancias. En comparación con los demás gases nobles, el radón es el más pesado y tiene el punto de fusión, ebullición, la temperatura crítica y la presión crítica más elevados (Cothorn and Smith, 1987). La tabla 2.2 presenta las propiedades atómicas, físicas y químicas del radón.

Propiedad	Valor
Número atómico	86
Peso atómico estándar	222
Configuración electrónica	6s ² 6p ⁶
Densidad	9,73 kg/m ³
Punto de fusión (°K)	202
Punto de ebullición normal (°K)	208,2
Calor de fusión (kJ mol ⁻¹)	3.247
Calor de vaporización (kJ mol ⁻¹)	18.0
Entalpía de primera ionización (kJ mol ⁻¹)	1037
Estados de oxidación	0, 2, 6
Electronegatividad	2,2 (escala de Pauling)
Radio covalente (nm)	0,150
Radio de van der Waals (nm)	0,220

Tabla 2.2: Propiedades del radón, (Baskaran, 2016a).

2.2.6. Liberación de radón a partir de granos minerales

La liberación de radón a partir de los granos minerales representa un proceso clave ya que este contribuye a los niveles de concentración de radón que se pueden encontrar en suelo. Minerales como la uraninita y la torita que contienen uranio y

radio, se desintegran y liberan radón como producto de su decaimiento radioactivo. En el proceso de liberación (Fig. 2.3), la ubicación del átomo de radio dentro del grano mineral (lo cerca que este de la superficie del grano) y la dirección del retroceso del átomo de radón (si es hacia la superficie o hacia el interior del grano) determinan si el átomo de radón recién formado entra o no en el espacio poroso entre los granos minerales (Otton, 1992).

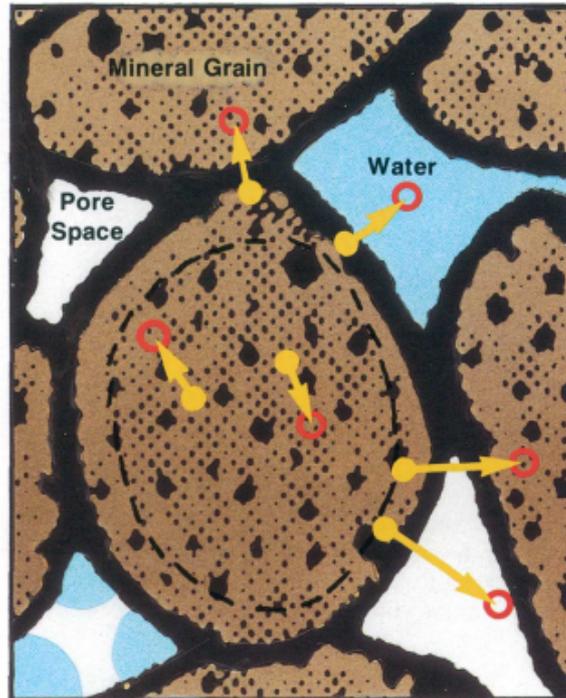


Figura 2.11: Formación de radón en granos minerales (Otton, 1992).

Otro factor que es necesario tomar en cuenta es la humedad del suelo, esta puede afectar la desintegración radiactiva de los minerales y, por lo tanto, la liberación de radón. Por otro lado, condiciones ambientales, como la temperatura y la presión, también pueden influir en la liberación de radón. En la tabla 2.3 se observa los coeficientes de emanación de radón. Estos indican la fracción de radón que realmente es liberada al espacio poroso o a la atmósfera dependiendo del medio en el que se ubica.

	Rocas	Suelos	Materiales de construcción
Coefficiente de emanación	2.1 - 32	0.17 - 80	0.1 - 58

Tabla 2.3: Coeficiente de emaciación de radón en medios, (Cinelli et al., 2019).

2.2.7. Distribución de Radio en el suelo y Parámetros Granulo-métricos

La distribución de radio en el suelo está estrechamente relacionada con los parámetros granulo-métricos del suelo. Estos se refieren a la distribución de tamaños de partículas presentes en el suelo. Los parámetros granulo-métricos, como el tamaño de las partículas, la porosidad y la textura del suelo, pueden influir en la movilidad y retención del radio en el suelo. Tomando en cuenta esto, la ubicación, tamaño y forma son las principales características que intervienen en la cantidad de radón que puede escapar de poros intersticiales (Baskaran, 2016b). Entonces existirá una diferencia notable en función a la distribución de los minerales en el grano, ya sea una distribución homogénea o por lo contrario una distribución concentrada en la superficie de este, podemos evidenciar esta diferencia en la Figura 2.32. La tasa de liberación de radón

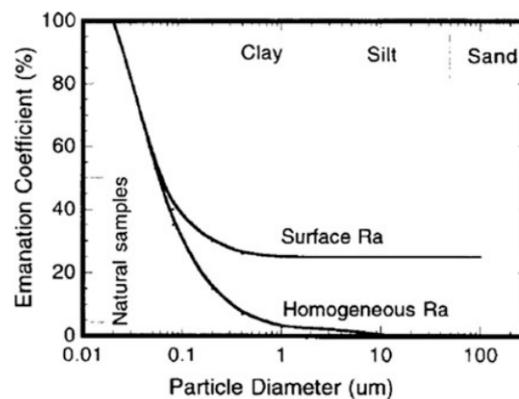


Figura 2.12: Comparación de emanación de Rn222 para distribución homogénea y distribución superficial de Ra226. (Baskaran, 2016b)

puede variar dependiendo de la concentración y distribución de los minerales radiactivos en el suelo. Por otro lado la permeabilidad del suelo afecta la difusión y transporte del radón a través de los poros del suelo.

2.2.8. Proceso de migración de radón

La migración del radón es un proceso clave. Este describe el movimiento del radón partiendo desde su fuente de emanación en el suelo o rocas hasta terminar en su exhalación a la atmosfera. Diversos factores influyen en como se dará el desarrollo de este proceso, por ejemplo la permeabilidad del suelo y de las rocas, la presencia de grietas o fracturas, la actividad del agua subterránea y otros factores locales (Rezaie et al., 2023). En la Figura 2.13 se puede observar el proceso de migración del radón descrito en 3 etapas.

- Emanación: El radón se mueve a través de los poros y espacios interconectados presentes en el suelo y las rocas mediante un proceso llamado difusión.

- Transporte: El radón puede ser transportado por el agua subterránea, especialmente en áreas con acuíferos contaminados con radón.
- Exhalación: El radón puede ser liberado al aire desde el suelo y las rocas a través de procesos de exhalación.

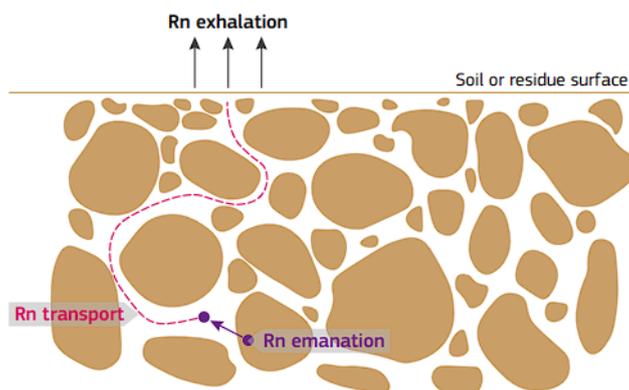


Figura 2.13: Migración de radón, (Cinelli et al., 2019).

2.2.8.1. Emanación de radón

En este proceso los átomos de radón formados a partir de la desintegración del radio escapan de los granos al espacio intersticial, el gas se desplaza desde áreas de mayor concentración a áreas de menor concentración hasta alcanzar un equilibrio. La emanación está relacionada con el retroceso alfa y la difusión de los átomos de radón (Sakoda et al., 2011), siendo el retroceso alfa el factor más importante. Sin embargo, algunos átomos de radón pueden quedar atrapados en los espacios circundantes (Fig.2.14). Los átomos de radón con mayor probabilidad de alcanzar el espacio poroso en el suelo son los producidos por la desintegración de átomos de radio ubicados muy cerca a la superficie del grano (Sakoda et al., 2011). Adicional a esto otro factor importante corresponde a el contenido de humedad que exista en el suelo, esta posee un efecto sobre el coeficiente de emanación. De esta manera el agua producto de la humedad de suelo posee un poder de detención por mucho superior al que genera el aire. Generando una reducción de la pérdida de emanación de los átomos de radón (Cinelli et al., 2019).

2.2.8.2. Transporte de radón

El radón se transporta a través del suelo por difusión y convección. Este se encuentra influenciado por la humedad del suelo, la temperatura y la presencia de materia orgánica (Cinelli et al., 2019; Petermann et al., 2021). La actividad humana, como la construcción de edificios y la excavación del suelo, también afecta el transporte de radón en el suelo (Wiegand, 2001).

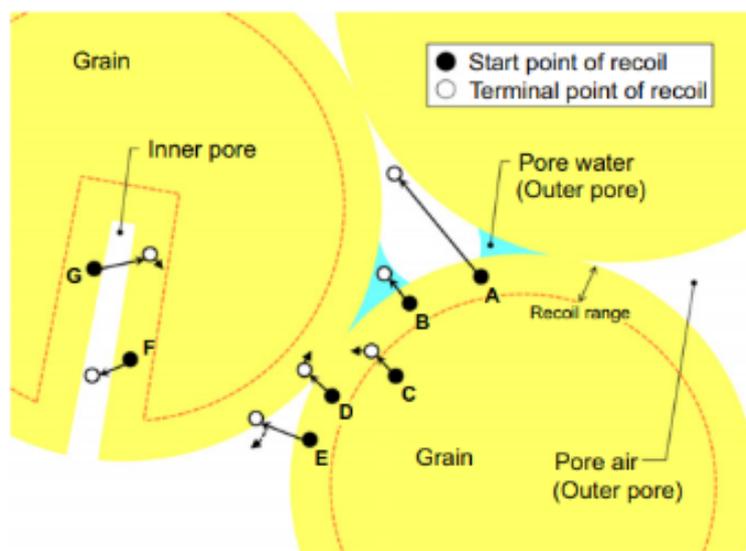


Figura 2.14: Modelo del fenómeno de emanación del radón, (Sakoda et al., 2011).

El radón se transporta a través de los poros y grietas del suelo debido a gradientes de concentración, ya que al ser un gas obedece las leyes de difusión (Cothorn and Smith, 1987). Es decir, va desde zonas de alta concentración en el subsuelo hacia zonas de baja concentración en la atmósfera. La difusión del radón varía en diferentes tipos de medios, en la Tabla 2.4 se presentan los coeficientes de difusión del radón y la longitud de difusión en algunos medios. Por otro lado, la velocidad del viento, la diferencia de presión y la temperatura generan corrientes convectivas que arrastran el gas radón presente en el suelo hacia la superficie. El radón se puede mover con facilidad a través de medios porosos como el suelo o la roca fracturada, suelos como la arena gruesa y grava, presentan alta permeabilidad; mientras que, los suelos arcillosos son menos permeables (Otton, 1992).

2.2.8.3. Exhalación de radón

La tasa de exhalación de radón a la atmósfera es un proceso complejo que depende de varios factores como la geología de la zona de estudio, la porosidad, textura del suelo, la humedad, la temperatura del suelo y las propiedades físicas (Cinelli et al., 2019). Un factor a tener en cuenta es la altitud, zonas de gran altitud se caracterizan por poseer suelos menos compactos, compuestos generalmente por elementos gruesos. Por otro lado, zonas de baja altitud con pendientes más suaves, se caracterizan en su mayoría por poseer suelos compuestos de fracciones más finas que inhiben el movimiento del radón del suelo al aire (Rezaie et al., 2023). Consecuentemente esto disminuye así la tasa de exhalación de radón.

2.2.9. Difusión de radón

La difusión del radón es el proceso mediante el cual este gas se transporta a través de los materiales porosos presentes en el suelo, rocas y otros. Cuando el radón se genera este tiende a migrar hacia la atmósfera, debido a la diferencia de presión entre el suelo y la esta. En la tabla 2.4 se presenta los diferentes coeficiente de difusión de radón. Los coeficientes mencionados describen la capacidad del radón para desplazarse a través de estos diferentes medios debido a gradientes de concentración. El coeficiente es un parámetro fundamental para comprender la migración de radón desde suelos y rocas hasta el aire en el interior de edificios.

Medio	Coefficiente de difusión (cm ² s ⁻¹)	Longitud de difusión (m)
Aire	10 ⁻¹	2,4
Agua	10 ⁻⁵	
Arena	3x10 ⁻²	1,5
Argilita	8x10 ⁻⁵	
Hormigón	2x10 ⁻⁵	0,04 - 0,26
Minerales cristalinos	10 ⁻⁹ - 10 ⁻²⁰	

Tabla 2.4: Coeficiente de difusion de radon en medios (Cothorn and Smith, 1987)

2.2.9.1. Difusión de bulk

La difusión consiste principalmente en el flujo neto de átomos, en el contexto de la teoría de difusión, los coeficientes de difusión están definidos por la ley de Fick 2.6 (Fick, 1855)

$$J = -D \frac{\partial \phi}{\partial x} \quad (2.6)$$

La difusión de bulk, también conocida como difusión en masa, es otro mecanismo importante en el transporte de radón en el suelo. A diferencia de la difusión a través de los poros del suelo, la difusión de bulk implica el movimiento del radón a través de la matriz sólida del suelo en lugar de los espacios porosos (Nielson and Rogers, 1982), definido en 2.7.

$$D_a = TD_A \quad (2.7)$$

Este proceso es especialmente importante en suelos con baja porosidad y baja permeabilidad, donde la difusión a través de los poros es limitada.

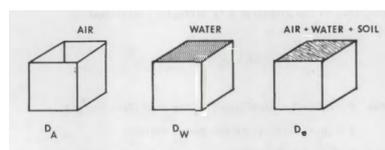


Figura 2.15: Coeficiente de difusión de bulk en medios (Nielson and Rogers, 1982).

2.2.9.2. Difusión de Knudsen

La difusión de Knudsen es un proceso de difusión molecular. Ocurre cuando las moléculas de radón se mueven en línea recta a través de los poros sin interactuar significativamente con las paredes de los poros. La difusión se rige por la teoría de la difusión de Knudsen. Definido por (Nielson and Rogers, 1982) en 2.8

$$\frac{1}{D_a} = \frac{1}{\tau D_k} + \frac{1}{\tau D_o} \quad (2.8)$$

La comprensión de la difusión de Knudsen es fundamental para evaluar la movilidad del radón en el suelo, especialmente en suelos con características particulares, como arcillas compactadas o sedimentos con estructuras micro-porosas.

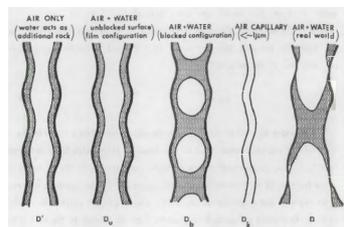


Figura 2.16: Coeficiente de difusión de Knudsen en medios (Nielson and Rogers, 1982).

2.2.10. Radón en viviendas

Como se ha mencionado, el radón es un gas liberado al ambiente por la desintegración de radionúclidos como el uranio y el radio, los cuales se encuentran distribuidos en el suelo, las rocas o los materiales de construcción. No obstante, en presencia de edificaciones, este tiende a migrar hacia el interior de las edificaciones. Este flujo de radón se ve favorecido por tres factores principales: las diferencias de presión de aire entre el suelo y la atmósfera dentro de la vivienda, la presencia de aberturas en suelo de la vivienda y el aumento de la permeabilidad alrededor de los cimientos (Otton, 1992). Estas condiciones facilitan la entrada del gas radón hacia el interior de los edificios, representando un riesgo para la salud de sus ocupantes

Es importante tener en cuenta que la permeabilidad del suelo y la naturaleza de las formaciones geológicas subyacentes pueden afectar significativamente las concentraciones de radón en el interior de las viviendas. La Figura 2.17 proporciona una representación visual donde se observa las vías de migración. Grietas en el suelo, uniones entre suelos y paredes, huecos alrededor de tuberías, pequeños poros en paredes de bloques huecos, paredes huecas y sumideros facilitan el transporte del radón hacia las edificaciones. Por otro lado, un factor geológico influyente es la proximidad de la zona residencial a estructuras geológicas como fallas y lineamientos. Estas estructuras actúan como vías para la migración del radón desde el lecho rocoso hacia la superficie (Rezaie et al., 2023).

En este contexto, estudios demuestran que la concentración de gas radón en el suelo contribuye en los niveles de concentración de radón en el aire interior (Vaupotič

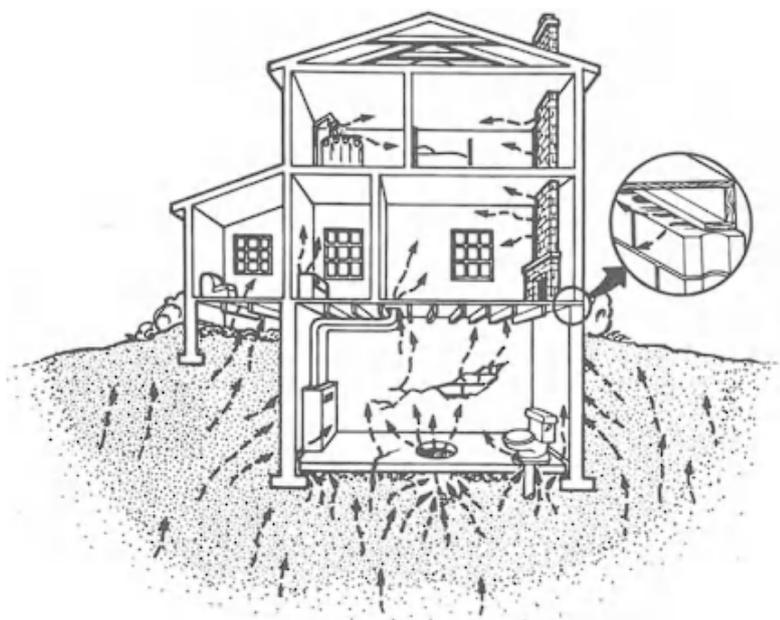


Figura 2.17: Fuentes y vías de entrada del radón en interiores (Cothorn and Smith, 1987)

et al., 2002). Presentando correlaciones positivas y lineales en diferentes contextos (Al-Khateeb et al., 2017; Kumar and Chauhan, 2014; Farid, 1997). Por lo tanto, este gas se filtra a través del suelo, ingresa a las viviendas y puede acumularse en el interior alcanzando altos niveles de concentración en el aire, representando un riesgo para la salud de los habitantes.

2.2.11. Radón en salud

La exposición constante de la población a diversas fuentes de radiación tanto naturales como artificiales dependiendo de su origen, representa un problema de salud pública. Una de las principales fuentes naturales de radiación ionizante es el radón. Este se destaca ya que representa un riesgo significativo para la salud, siendo una amenaza biológica especialmente para las células del cuerpo (Valcarce et al., 2022). La Organización Mundial de la Salud (OMS) ha catalogado al gas radón como la segunda causa principal de cáncer de pulmón después del tabaco. Se estima que es responsable de aproximadamente el 15% de los casos de esta enfermedad a nivel mundial (WHO, 2009).

El riesgo que representa la exposición al gas radón para la salud pública ha generado gran preocupación a nivel global, especialmente en zonas donde se registran altos niveles de radón en interiores. Los habitantes de estas zonas se encuentran más propensos a desarrollar alguna enfermedad pulmonar. En el contexto nacional, el Reglamento de Seguridad Radiológica (D.S. N° 009-97-EM) del Instituto Peruano de Energía Nuclear (IPEN), establece los niveles de referencia de radón en interiores en 200 y 1000 (Bq/m^3) para viviendas y puestos de trabajo respectivamente. Aunque el promedio de

concentración de radón en el Perú es bajo (Fig.2.18), investigaciones relacionadas con la medición de concentraciones de radón en el aire interior de viviendas de Lima (Guevara, 2019; Pereyra et al., 2015), han evidenciado valores superiores a los recomendados.

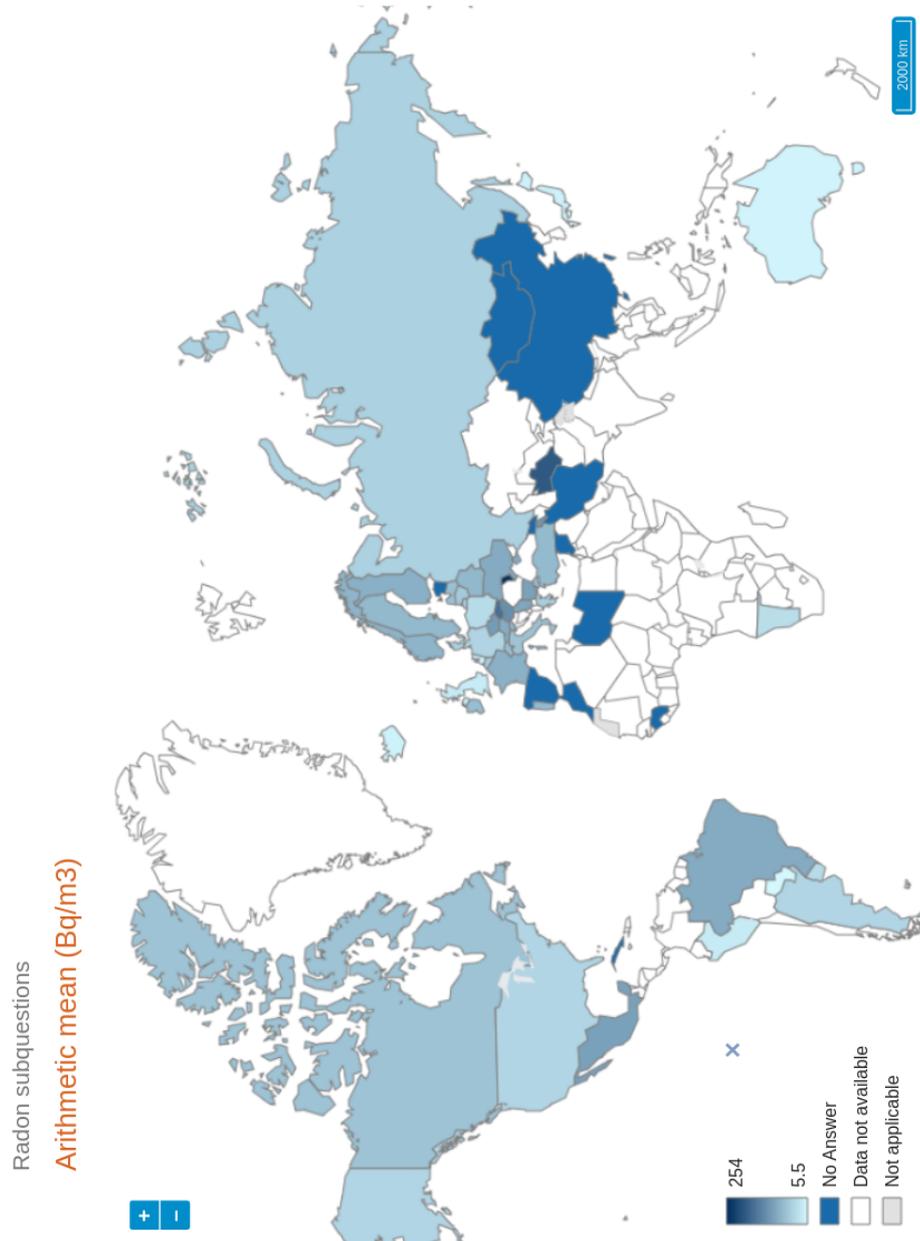


Figura 2.18: Niveles de concentración de radón por país (WHO, 2009)

La principal preocupación en temas de salud relacionados con el radón es su progenie. Esta se refiere a los elementos radiactivos que se generan cuando el radón-222 se desintegra. Los cuales al ser sólidos se adhieren al material particulado en el aire, pudiendo ser inhalados y alcanzando tejidos blandos del tracto respiratorio. Al permanecer depositadas pueden desintegrarse mediante decaimiento alfa afectando la salud de las personas.

2.3. Radón y Geología

2.3.1. Radionuclidos

Los radionúclidos son isótopos radiactivos, es decir, versiones inestables de los elementos químicos que emiten radiación ionizante. Estos radionúclidos tabla 2.5 se encuentran naturalmente en el medio ambiente y pueden ser producidos por procesos naturales, como la desintegración radiactiva de elementos como el uranio, el torio y el potasio-40. Algunos ejemplos de radionúclidos comunes incluyen el uranio-238, el uranio-235, el torio-232, el potasio-40 y el radón-222.

Radionúclido	Nombre	Semivida (años)	Radiaciones Principales	Concentración típica en la corteza terrestre (Bq/kg)
^{40}K	Potasio	$1,26 \times 10^9$	β, γ	630
^{50}V	Vanadio	6×10^{15}	γ	2×10^{-5}
^{87}Rb	Rubidio	$4,8 \times 10^{10}$	β	70
^{113}Cd	Cadmio	$8,04 \times 10^{15}$	β	$< 2 \times 10^{-6}$
^{115}In	Indio	6×10^{14}	β	2×10^{-5}
^{123}Te	Telurio	$1,2 \times 10^{13}$	Rayos X	2×10^{-7}
^{138}La	Lantano	$1,12 \times 10^{11}$	β, γ	2×10^{-2}
^{142}Ce	Cerio	$> 5 \times 10^{16}$	No reportado	$< 1 \times 10^{-5}$
^{144}Nd	Neodimio	$2,4 \times 10^{15}$	α	3×10^{-4}
^{147}Sm	Samario	$1,05 \times 10^{11}$	α	0.7
^{152}Gd	Gadolinio	$1,1 \times 10^{14}$	α	7×10^{-6}
^{174}Hf	Hafnio	$2,0 \times 10^{15}$	α	2×10^{-7}
^{176}Lu	Lutecio	$2,2 \times 10^{10}$	e^{-}, γ	0.04
^{187}Re	Renio	$4,3 \times 10^{10}$	β	1×10^{-3}
^{190}Pt	Platino	$6,9 \times 10^{11}$	α	7×10^{-8}
^{192}Pt	Platino	1×10^{15}	α	3×10^{-6}
^{209}Bi	Bismuto	2×10^{19}	α	$< 4 \times 10^{-9}$
^{128}Te	Telurio	$7,7 \times 10^{24}$	Doble β	-

Tabla 2.5: Nucleidos primordiales, modificado de (Cinelli et al., 2019)

2.3.2. Impacto de la geología en los niveles de radiación

La geología se encuentra intrínsecamente ligada a la radiactividad natural. Esto debido a que las distintas rocas y minerales contienen diferentes cantidades de elementos radiactivos naturales. Elementos como el uranio, el torio y el potasio, se descomponen con el tiempo liberando radiación al medio ambiente (Cinelli et al., 2019). El nivel de radiación del suelo se puede ver afectado dependiendo del tipo de roca, de la cantidad de elementos radiactivos que contiene y su grado de erosión. Por otro lado, la composición y la estructura del subsuelo se toman como algunos de los principales influyentes en los niveles de radiación natural o de fondo (Cinelli et al., 2019). La tabla 2.6 muestra los niveles de radiación en mSv/año de las diferentes capas de suelo. En este contexto, la generación de la radiación de fondo esta relacionada con procesos ocurridos en el

ámbito geológico como la movilización, precipitación, concentración y lixiviación de los radionucleidos naturales (Fig.2.19, nivel 1) (Cinelli et al., 2019).

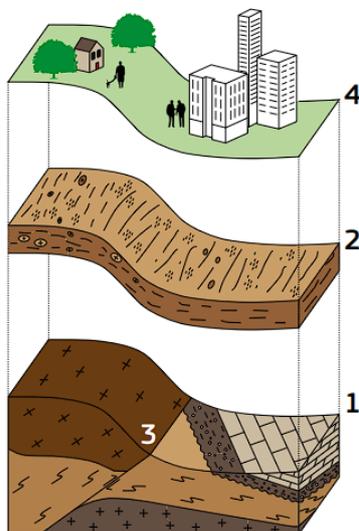


Figura 2.19: Fuentes de los principales radionucleidos naturales de las series de desintegración ^{238}U , ^{232}Th y ^{40}K ; Fuente primaria: lecho rocoso (nivel 1); fuente secundaria: suelo (nivel 2); fuentes adicionales: discontinuidades (por ejemplo, fallas geológicas, nivel 3); fuente antropogénica: materiales de construcción y contaminación ambiental (nivel 4). (Cinelli et al., 2019).

La presencia y concentración de estos radionúclidos depende de la geología de la zona, rocas como los granitos, esquistos y rocas sedimentarias tienden a presentar un mayor contenido de uranio y radio, por ende, una mayor emanación de radón. Por otro lado, factores como la permeabilidad, la composición del suelo, la dinámica de la capa freática y la tectónica (Fig. 2.19, nivel 3) influyen en la migración del gas radón.

Capa de suelo	Nivel de radiación (mSv/año)
Tierra vegetal	0.1-0.5
Sub suelo	0.05-0.1
Lecho rocoso	0.01-0.05

Tabla 2.6: Nivel de radiación típica en suelo

Otro aspecto importante a tener en cuenta es la relación entre el radón y las fallas geológicas. Debido a que estas últimas al ser zonas fracturadas en la corteza terrestre actúan como rutas que favorecen la migración del radón (Chen et al., 2018). Debido al mayor grado de permeabilidad en las zonas de fallas, se generan anomalías en la concentración de radón y geogases (Fu et al., 2017). La proximidad al plano de falla y la litología del lecho rocoso son los principales factores que controlan la velocidad de migración del radón y su concentración en el suelo poco profundo (Cinelli et al., 2019; Rezaie et al., 2023; Hosoda et al., 2020).

2.3.3. Litología

Se conoce que la radiactividad natural presente en las rocas esta asociada con el tipo de roca. Las rocas pueden clasificarse en ígneas (Fig.2.20(a)), sedimentarias (Fig.2.20(c)) o metamórficas (Fig.2.20(b)) según su proceso de formación (Cinelli et al., 2019).



Figura 2.20: Clasificación de rocas

La litología representa un parámetro importante a tener en consideración para el estudio de concentración de radón en suelo, se ha demostrado su relevancia en diferentes estudios (Rezaie et al., 2023; Petermann et al., 2021; Oni et al., 2022). Ya que esta describe las características de las rocas que constituyen una determinada formación geológica basado en columnas estratigráficas. Considerando que estas pueden contener radionúclidos debido a su formación y a la presencia de elementos radiactivos en los materiales de los que se componen. El contenido de uranio varía significativamente entre distintas litologías. Como se observa en la tabla 2.7 las rocas ígneas, especialmente el granito, tienen comúnmente niveles más altos de uranio en comparación con las rocas sedimentarias y metamórficas.

Tipo de Roca	Formación	Contenido de Radionúclidos
Rocas Ígneas	Enfriamiento y solidificación del magma fundido	Altas concentraciones de uranio y torio, desintegración en radón y otros productos de desintegración
Rocas Sedimentarias	Acumulación y cementación de sedimentos	Contenido variable dependiendo de la composición del sedimento original. Puede contener trazas de uranio y otros radionúclidos
Rocas Metamórficas	Transformación de rocas preexistentes por altas temperaturas y presiones	Redistribución y concentración variable de radionúclidos en minerales específicos o bandas de minerales

Tabla 2.7: Formación de rocas y contenido de radionuclidos

2.3.4. Permeabilidad del suelo

La permeabilidad, es la capacidad del suelo para permitir el flujo de agua, aire y otros fluidos a través de él. El radón al ser un gas tiende a moverse desde áreas de alta concentración hacia áreas de baja concentración, lo que conduce a su dispersión en el suelo y su eventual liberación a la atmósfera. En diversos estudios realizados en países como Irlanda y Alemania se observó como la velocidad de difusión del radón en el suelo depende bastante de la permeabilidad del suelo (Ciotoli et al., 2017; Hughes et al., 2022; Petermann et al., 2021), que está determinada por la estructura del suelo y la presencia de porosidad interconectada.

La estructura del suelo se refiere a cómo se agrupan las partículas de suelo y cómo se organizan entre sí (Fig. 2.21). Esta puede tener un impacto significativo en la permeabilidad de los suelos. Depende de factores como el material parental, la química y la estructura de los minerales, la actividad biológica y las condiciones ambientales (como la presencia de arcillas hinchables o la congelación y descongelación) (Cinelli et al., 2019).

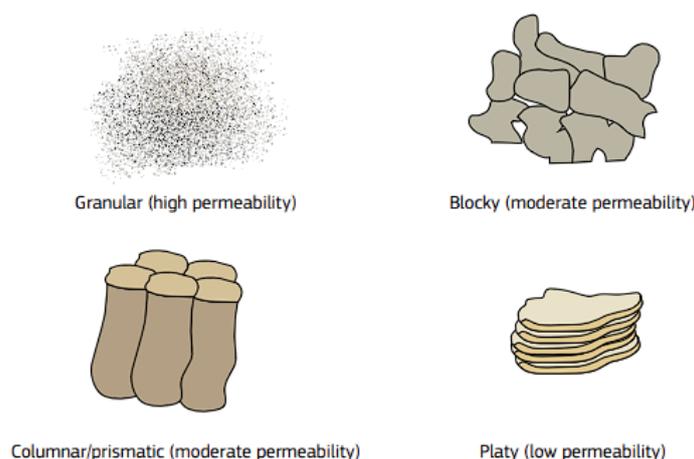


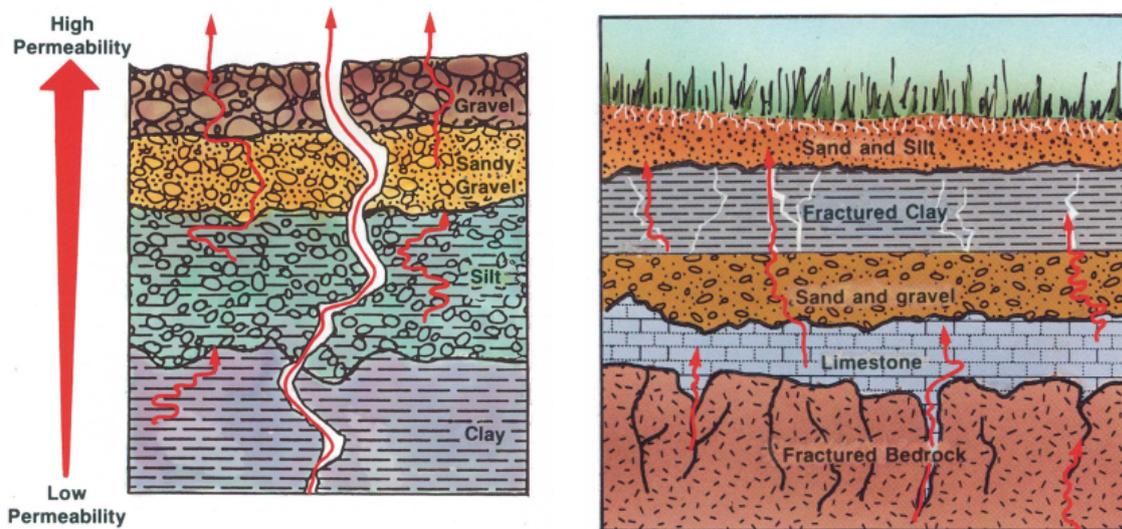
Figura 2.21: Nivel de permeabilidad de suelos basado en su estructura, (Cinelli et al., 2019).

El coeficiente de permeabilidad se expresa en unidades de velocidad, generalmente en metros por segundo (m/s) o centímetros por segundo (cm/s). Los valores del coeficiente de permeabilidad pueden variar ampliamente según las características del suelo y su composición (tabla 2.9).

Clases de permeabilidad del suelo	Coeficiente de permeabilidad (K en m/s)	
	Limite inferior	Limite superior
Permeable	2×10^{-7}	2×10^{-1}
Semi-permeable	1×10^{-11}	1×10^{-5}
Impermeable	1×10^{-11}	5×10^{-7}

Tabla 2.8: Clasificación de la permeabilidad

Los suelos con mayor permeabilidad como los compuestos por arena gruesa y la grava permiten un transporte más rápido del radón a través de sus poros. Por otro lado, suelos impermeables como los compuestos por arcillas dificultan el transporte del gas (Fig.2.22(a)). La permeabilidad puede variar significativamente según el tipo de suelo (Fig.2.22(b)), y se mide mediante el coeficiente de permeabilidad (k) tabla 2.8, que es un valor que cuantifica la facilidad con la que el agua se mueve a través del suelo. Los suelos se pueden clasificar como permeable, semi-permeable e impermeable esto según su eficiente de permeabilidad.



(a) Distribución en función de la permeabilidad del suelo

(b) Distribución por tipos de suelo

Figura 2.22: Distribución de radón en suelo, por (Otton, 1992)

Tipo de suelo	Coefficiente de permeabilidad (m/s)
Arena gruesa	$1 \times 10^{-2} - 1 \times 10^{-1}$
Arena media	$1 \times 10^{-3} - 1 \times 10^{-2}$
Arena fina	$1 \times 10^{-4} - 1 \times 10^{-3}$
Limo	$1 \times 10^{-5} - 1 \times 10^{-4}$
Arcilla	$1 \times 10^{-6} - 1 \times 10^{-5}$
Suelo orgánico	$1 \times 10^{-7} - 1 \times 10^{-6}$
Roca sólida	Muy baja permeabilidad

Tabla 2.9: Coeficiente de permeabilidad de suelos

2.3.5. Porosidad del suelo

La porosidad del suelo se refiere al volumen de espacios vacíos o poros presentes en relación al volumen total del suelo. Estos poros pueden contener aire, agua o una combinación de ambos. El transporte difusivo del radón en los poros del suelo es otro proceso relevante en la dinámica de este gas radiactivo en el medio ambiente. El radón se desplaza a través de los poros del suelo debido a las diferencias de concentración y presión, siguiendo un proceso de difusión gobernado por las leyes de difusión molecular.

La porosidad del suelo puede variar según diversos factores como el tipo de suelo, la estructura del suelo, el contenido de materia orgánica y la compactación del suelo. Los tipos de poros presentes en el suelo incluyen porosidad primaria (espacios creados durante la formación del suelo) y porosidad secundaria (espacios creados posteriormente debido a procesos como la bioturbación o la contracción y expansión del suelo). A continuación, se presenta un rango aproximado de porosidad para los principales tipos de suelo tabla 2.10.

Tipo de suelo	Rango de porosidad (%)
Arena	25 - 40
Limo	40 - 50
Arcilla	40 - 60
Suelo orgánico	50 - 70

Tabla 2.10: Rango de porosidad según el tipo de suelo

2.3.6. Granulometría del suelo

La granulometría del suelo es el estudio y la determinación de tamaño de las partículas que componen el suelo. clasificándolas en tres grupos principales arenas, gravas, limos y arcillas, esta clasificación usualmente se da empleando una escala cronométrica. La escala de (Wentworth, 1922) es la más ampliamente adoptada para la clasificación granulométrica de partículas sedimentarias (Fig.2.23).

Este parámetro es fundamental para entender las propiedades físicas del suelo, como la permeabilidad, la capacidad de retención de agua entre otros. De igual manera existe una relación entre el tamaño de grano y la concentración de radón en el gas del suelo que se encuentra influenciada por la distribución de radio en los granos del suelo. Se ha observado que la concentración de radio aumenta a medida que disminuye el tamaño del grano (Cinelli et al., 2019) esto representa un aumento en el coeficiente de emanación ya que el radio se encuentra distribuido en mayor proporción en las capas superficiales de los granos.

La clasificación de estos suelos es importante, ya que en suelos arenosos y gravosos se encuentran partículas de mayor tamaño y porosidad, permitiendo una mayor difusión y movilidad del radón. Por otro lado suelos limosos y arcillosos con partículas más pequeñas y de menor porosidad pueden restringir el movimiento del radón y provocar acumulaciones subterráneas (Rezaie et al., 2023).

Diameter in Millimeters	No. 75† Gravel †	No. 9 Sandy Gravel	No. 18 Gravelly Sand	No. 16 Sand	No. 68 Silty Sand
64-32.....		.2			
32-16.....		14.0			
16-8.....	3.3	13.2			
8-4.....	65.9	20.0	6.1		
4-2.....	26.0	14.4	7.4	1.7	.2
2-1.....	3.0	11.4	3.6	24.4	.3
1-1/2.....	.3	18.2	7.2	49.2	7.2
1/2-1/4.....	.2	5.6	36.8	8.3	17.9
1/4-1/8.....	.7	2.6	37.4	14.8	48.7
1/8-1/16.....	.3	.2	1.2	1.0	8.4
1/16-1/32.....			.2	.4	13.2
1/32-1/64.....				.f	3.2
1/64-1/128.....					.7
1/128-1/256.....					.1

* J. A. Udden, *loc. cit.*

† The numbers are those used by Udden in describing these sediments. The names are applied by the author of this paper according to the terminology here presented.

Diameter in Millimeters	No. 70 Sandy Silt	No. 129 Silt	No. 131 Clayey Silt	No. 19 Silty Clay
2-1.....				
1-1/2.....	.2	tr.	tr.	
1/2-1/4.....	.9	tr.	tr.	
1/4-1/8.....	11.9	tr.	.4	2.3
1/8-1/16.....	9.8	3.1	5.3	2.0
1/16-1/32.....	28.7	12.4	13.2	5.6
1/32-1/64.....	40.6	37.2	28.2	13.2
1/64-1/128.....	5.9	24.6	31.1	26.4
1/128-1/256.....	1.5	9.6	11.7	19.8
1/256-1/512.....	.2	3.5	6.4	17.8
1/512-1/1024.....		.9	2.4	10.0
1/1024-1/2048.....		.2	.8	2.3

Figura 2.23: Análisis mecánicos de J. A. UDDEN con los nombres de los sedimentos adecuados para la clasificación (Wentworth, 1922).

2.3.7. Geología Estructural

La geología estructural se ocupa del estudio de las estructuras geológicas y las fuerzas que las han formado. Se centra en el análisis de las deformaciones, fracturas y pliegues en las rocas que describen el relieve de una región. El origen y las características de su relieve se encuentra asociado a diversos episodios de modelamiento tectónico y climático, que han sido los motores para la exhumación y conservación de la cadena de los Andes (Benavente et al., 2013).

Es importante destacar que la geología estructural puede ayudar en la identificación y mapeo de áreas con mayor potencial de concentración de radón. Estudios geológicos detallados, incluyendo el mapeo de estructuras geológicas y la caracterización de las rocas y el suelo, pueden proporcionar información geoespacial valiosa que permita la identificación de áreas preferentes de transporte de radón.

En el contexto de geología estructural puntos claves son la tectónica de placas, lineamientos, pliegues y fallas. La relación de las fallas geológicas (Fig.2.24) con la concentración de radón radica en la migración y transporte del gas radón a través de las estructuras geológicas. Cuando existe una falla activa, es decir, una falla que ha experimentado desplazamiento reciente o continua actividad tectónica, puede haber una mayor probabilidad de que el radón se movilice a lo largo de esa estructura.

Las gran cantidad de fracturas y fisuras en la zona de falla pueden facilitar la migración del radón desde las capas más profundas hacia la superficie, debido a su mayor permeabilidad y porosidad en relación con las rocas circundantes, además (Baskaran, 2016a) resalta tres factores que afectan al flujo de Rn-222 en el interior de la tierra, estos son: (1) Convección inducida geotérmicamente, (2) Tensiones presísmicas en la tierra, (3) Emanación subterránea de gas.

En zonas con presencia de fallas geológicas y zonas de fractura el flujo de radón también puede variar en función de factores como la geometría de la falla, fallas no cicatrizadas y cicatrizadas, el volumen de roca fracturada, permeabilidad del núcleo de la falla entre otros (Benà et al., 2022). Estos factores favorecen el transporte de radón a distancia, teniendo como resultado una contribución adicional denominada como radón técnicamente incrementado (TER) (Benà et al., 2022). Por ende se presentan anomalías de radón en estas zonas observando valores significativamente superiores en comparación a los valores de fondo representativos de la zona u unidad geológica (Cinelli et al., 2019).

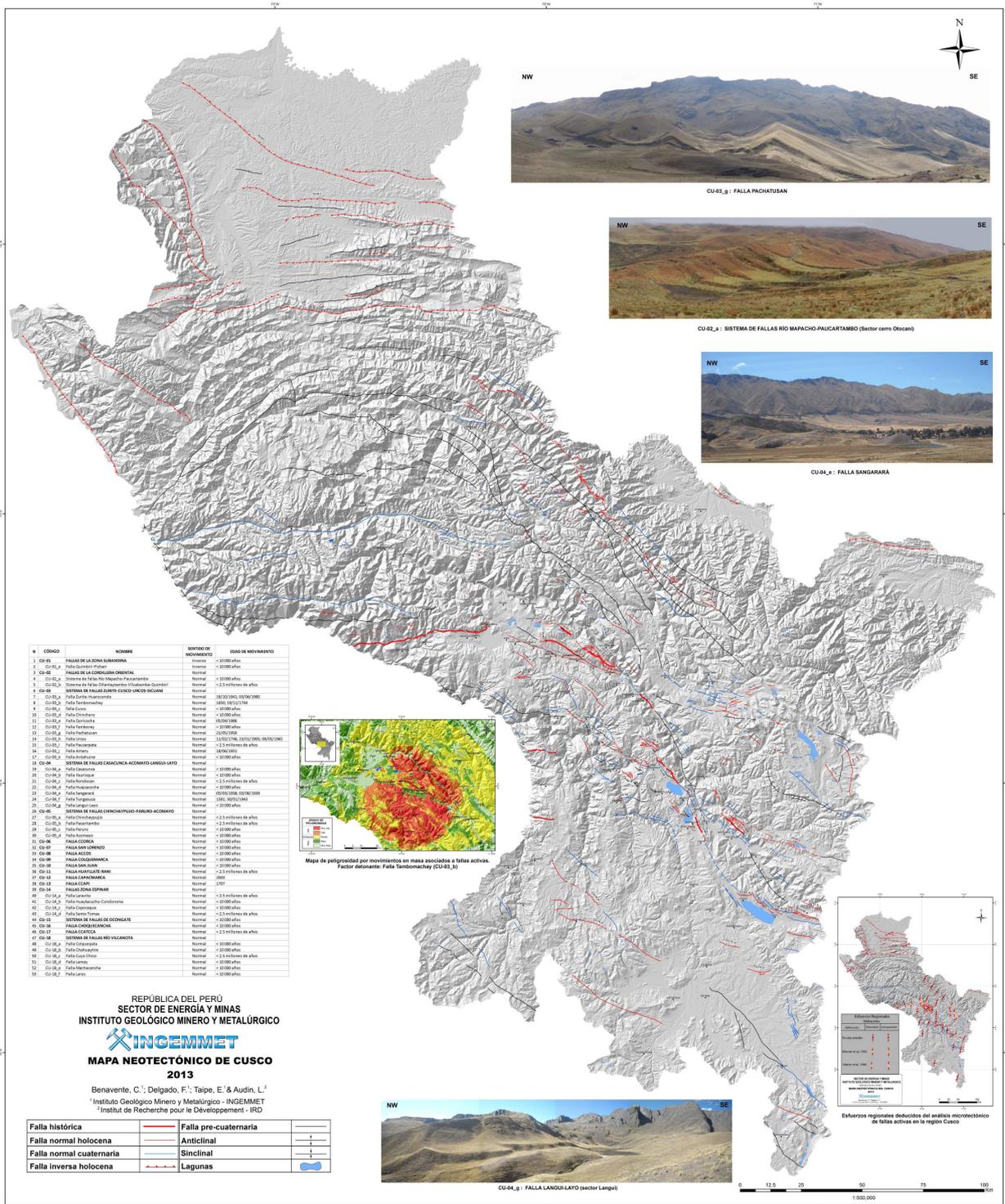


Figura 2.24: Mapa neotectónico región Cusco, (Benavente et al., 2013).

2.3.8. Cartografía Geológica y de Radón

La cartografía geológica y de radiación desempeña un papel fundamental en la comprensión de la geología de una región y la evaluación de factores que contribuyen a la radiación ambiental. Estos mapas son herramientas que permiten visualizar y analizar la distribución espacial de las características geológicas y la radiación en un área determinada (Otton, 1992).

Los mapas geológicos son representaciones gráficas que muestran la distribución y la naturaleza de las unidades geológicas, como rocas, sedimentos, fallas y estructuras geológicas en una región específica. La geología del cusco (Fig.2.25) fue resumido por (Carlotto et al., 2011).

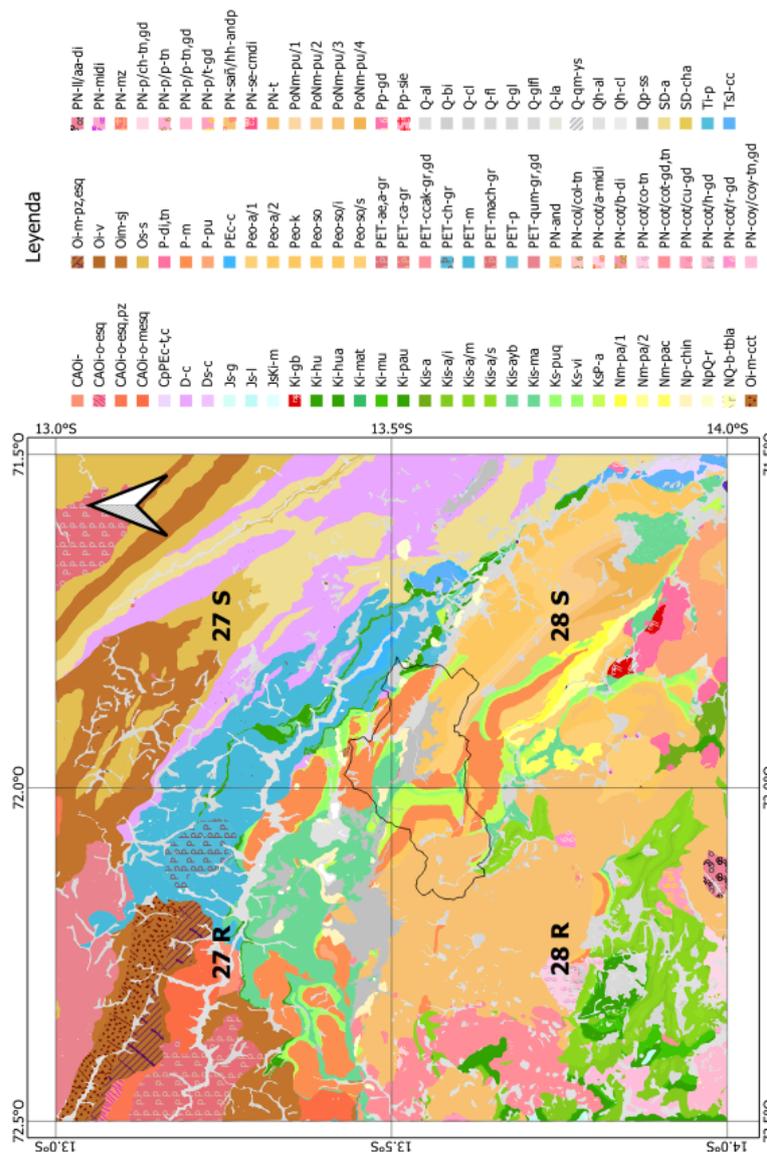


Figura 2.25: Mapa geológico de los cuadrangulos 27r, 27s, 28r, 28s correspondiente a la región Cusco, (Carlotto et al., 2011).

Por otro lado, los mapas de radiación son herramientas utilizadas para evaluar y representar la radiación ambiental en una región determinada, indican los niveles de uranio de los materiales de la superficie (Otton, 1992). Además estos mapas por ejemplo pueden mostrar la concentración de radionúclidos como Uranio (Fig. 2.26), Torio (Fig. 2.27), radón en suelo, aire o agua. También pueden mostrar la radiación ionizante natural, como los rayos gamma, que se produce debido a la presencia de elementos radiactivos en la corteza terrestre.

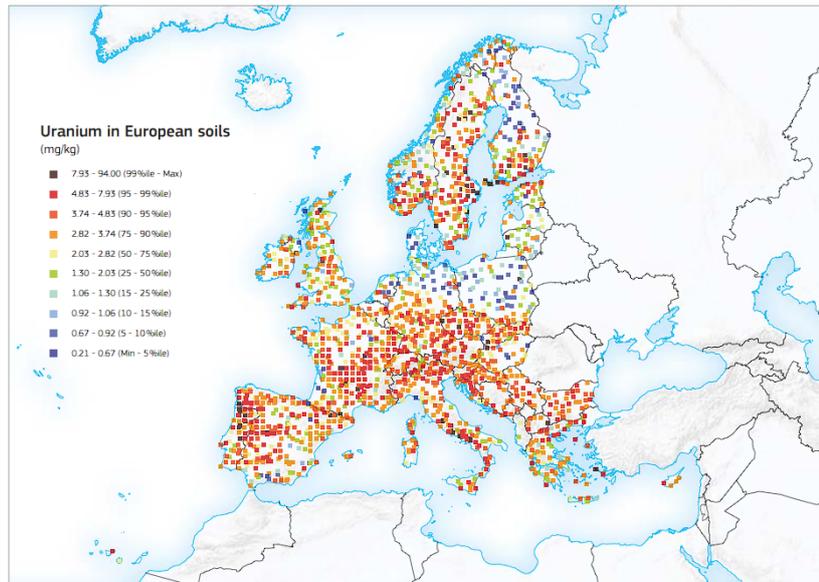


Figura 2.26: Mapa de U en suelo Europa, (Cinelli et al., 2019).

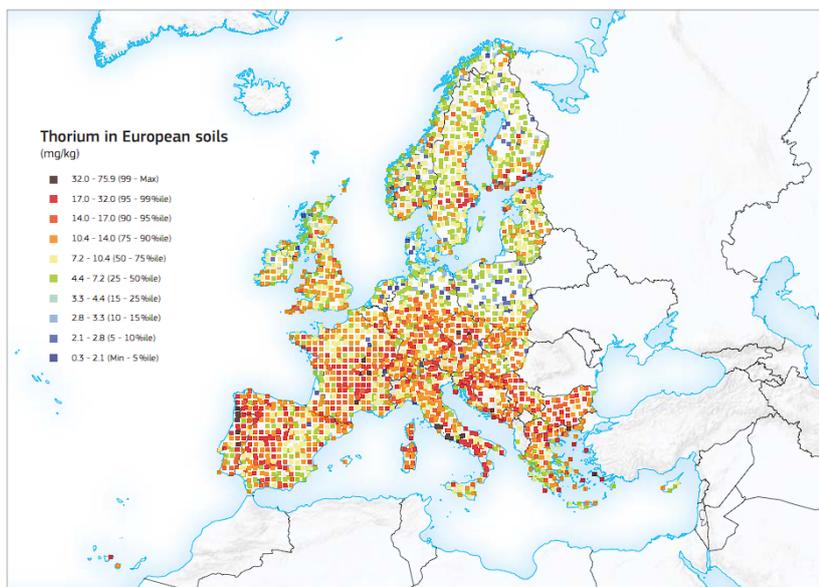


Figura 2.27: Mapa de Th en suelo Europa, (Cinelli et al., 2019).

2.4. Machine Learning

2.4.1. Introduccion al Machine Learning

Machine Learning, o aprendizaje automático, es un campo de estudio de la inteligencia artificial, su principal enfoque es desarrollar algoritmos y modelos capaces de aprender y tomar decisiones a partir de datos sin ser programados explícitamente (Elío et al., 2023). Estos algoritmos se entrenan con conjuntos de datos para reconocer patrones, hacer predicciones y tomar decisiones basadas en la información que se les proporciona. Debemos resaltar el enfoque buscado en este trabajo y el porque de optar por estas técnicas, ya que se ha demostrado que son una potente herramienta para modelar el comportamiento de sistemas complejos (Oni et al., 2022; Janik et al., 2018), donde se dependen de controles o predictores asumidos de forma complicada.

En el contexto del machine learning, se utilizan diferentes enfoques y técnicas, como redes neuronales artificiales (ANN), algoritmos de aprendizaje supervisado y no supervisado, y técnicas de procesamiento de datos. Estos modelos pueden ser aplicados en una amplia variedad de campos y tareas, como reconocimiento de imágenes, procesamiento del lenguaje natural, análisis de datos, predicción y clasificación. el poder predictivo de los métodos de machine learning puede ser sustancialmente superior al de los enfoques tradicionales (Elío et al., 2023), sobre todo en áreas como medicina, finanzas, industria, marketing, entre otros.

2.4.1.1. Aprendizaje Supervisado - Supervised Learning

El aprendizaje supervisado es una subcategoría de machine learning en la que se utilizan algoritmos que aprenden iterativamente de los datos para realizar tareas específicas, como la clasificación y la predicción. (Hastie et al., 2009) plantea la interpretación de estos datos suponiendo que estos proceden de un modelo estadístico que por conveniencia viene representado por la ecuación 2.9.

$$Y = f(X) + \varepsilon \quad (2.9)$$

El aprendizaje supervisado se basa en el uso de un conjunto de datos etiquetados para entrenar un modelo que pueda predecir o clasificar datos nuevos. El conjunto de datos etiquetados consta de pares de datos de entrada y datos de salida deseados, estos generan un conjunto de observaciones de entrenamiento que puede ser representado por la ecuación 2.10 donde x_i y y_i son valores observados (Hastie et al., 2009). Finalmente el objetivo es que el modelo aprende a identificar los patrones que se encuentran en los datos de entrada y a utilizar estos patrones para hacer predicciones precisas en datos no vistos.

$$\tau = (x_i, y_i); i = 1, \dots, N \quad (2.10)$$

2.4.1.2. Aprendizaje no Supervisado - Unsupervised Learning

El aprendizaje no supervisado es una subcategoría de machine learning donde se utilizan algoritmos que aprenden de los datos sin la necesidad de etiquetas o infor-

mación adicional sobre la respuesta deseada, de esta manera estos experimentan con un conjunto de datos que contiene muchas características y, a continuación, aprenden propiedades útiles de la estructura de este conjunto de datos (Goodfellow et al., 2016) siendo esta la principal diferencia con el aprendizaje supervisado ya que en lugar de tener datos con pares de entrada y salida, el aprendizaje no supervisado se basa en identificar patrones, estructuras o relaciones intrínsecas en los datos sin la guía de etiquetas preexistentes.

2.4.2. Redes Neuronales Artificiales

Una red neuronal es un modelo computacional inspirado en encontrar representaciones matemáticas del procesamiento de la información en los sistemas biológicos (Bishop and Nasrabadi, 2006). Consiste en un conjunto de unidades con elementos de computación interconectadas llamadas neuronas (Duong et al., 2021) artificiales que trabajan en conjunto para procesar información y realizar tareas específicas.

Cada neurona artificial recibe entradas, las procesa a través de una función de activación y genera una salida. Estas neuronas están organizadas en capas, donde las entradas se conectan a la capa de entrada, las salidas se obtienen en la capa de salida y existen capas intermedias conocidas como capas ocultas. (Duong et al., 2021)

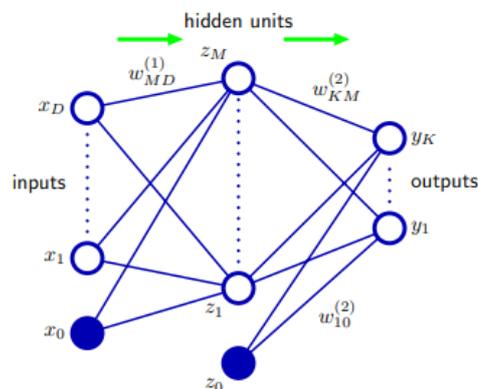


Figura 2.28: Diagrama red neuronal artificial, (Bishop and Nasrabadi, 2006).

Algunas de las ventajas de las redes neuronales artificiales incluyen su capacidad para manejar datos complejos y no lineales, su capacidad para aprender y adaptarse a nuevos datos sin un conocimiento previo de los atributos del objeto (Duong et al., 2021), y su capacidad para encontrar patrones y relaciones en los datos.

2.4.2.1. Redes Neuronales Recurrentes (RNN)

Las redes neuronales recurrentes (RNN) son un tipo de arquitectura de redes neuronales artificiales que están diseñadas para trabajar con datos secuenciales o de naturaleza temporal, son modelos de aprendizaje profundo que capturan la dinámica de secuencias mediante conexiones recurrentes (Zhang et al., 2021). A diferencia de las

redes neuronales tradicionales, las RNN tienen conexiones de retroalimentación, lo que les permite mantener y utilizar información de estados anteriores.

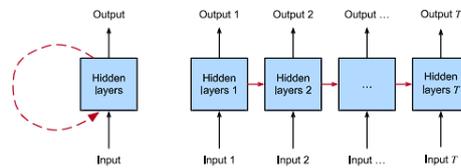


Figura 2.29: Diagrama red neuronal recurrente (RNN), (Zhang et al., 2021).

las RNN pueden considerarse como redes neuronales feedforward en las que los parámetros de cada capa (tanto los convencionales como los recurrentes) se aplican de forma sincronizada (Zhang et al., 2021), La característica principal de las RNN es su capacidad para procesar secuencias de longitud variable, donde cada elemento de la secuencia se considera en un orden específico y se toma en cuenta la información contextual de los elementos anteriores.

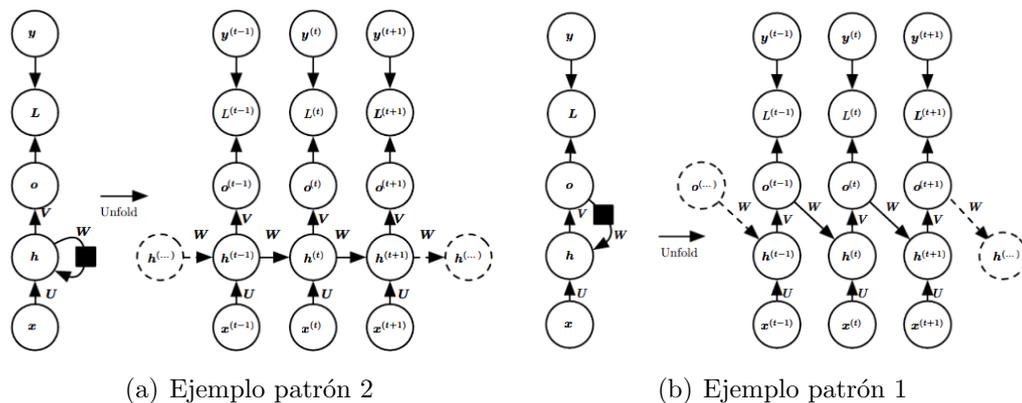


Figura 2.30: Diseño de patrones de una RNN(Goodfellow et al., 2016)

Los patrones de diseño mas resaltantes en las RNN, son por ejemplo, una RNN producen una salida en cada paso temporal y tienen conexiones recurrentes entre las unidades ocultas 2.30(a) y una RNN producen una salida en cada paso de tiempo y tienen conexiones recurrentes sólo desde la salida en un paso de tiempo a las unidades ocultas en el siguiente paso de tiempo 2.30(b) (Goodfellow et al., 2016)

Las RNN son capaces de modelar dependencias a largo plazo en los datos secuenciales gracias a la retroalimentación de las conexiones recurrentes. Sin embargo, pueden enfrentar dificultades al tratar con secuencias largas debido a los problemas de desvanecimiento o explosión del gradiente. Para abordar este problema, se han desarrollado variantes de RNN, como las redes LSTM (Long Short-Term Memory) y las GRU (Gated Recurrent Unit), que incorporan mecanismos de memoria y control de flujo de información para mejorar el manejo de secuencias largas.

2.4.2.2. Red de memoria a corto-largo plazo - Long short-term memory (LSTM)

LSTM (Long Short-Term Memory) es una variante de las redes neuronales recurrentes (RNN) que fue desarrollada para abordar el problema del desvanecimiento del gradiente en el entrenamiento de redes neuronales de largo plazo. Las LSTM tienen conexiones recurrentes de modo que el estado de las activaciones anteriores de la neurona del paso de tiempo anterior se utiliza como contexto para formular una salida (Brownlee, 2017). A diferencia de las RNN tradicionales, las LSTM incorporan una estructura de celdas de memoria que les permite capturar y almacenar información a largo plazo de manera más efectiva.

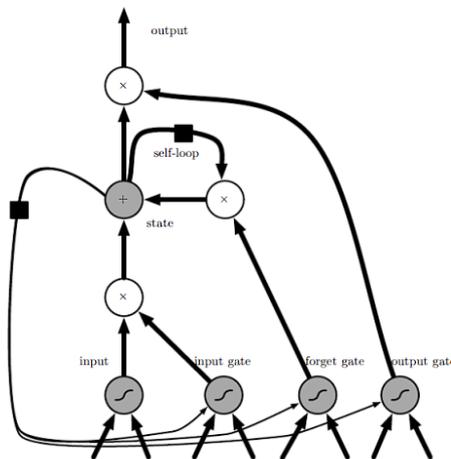


Figura 2.31: Diagrama de celda de memoria, (Goodfellow et al., 2016).

La arquitectura de una LSTM se compone de unidades de memoria llamadas “celdas LSTM”, que están conectadas entre sí a través de compuertas, cada celda de memoria contiene un estado interno, es decir, un nodo con una arista recurrente auto-conectada de peso fijo 1, lo que garantiza que el gradiente pueda pasar por muchos pasos temporales sin desvanecerse (Zhang et al., 2021).

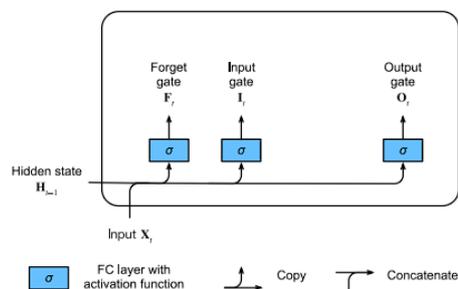


Figura 2.32: Compuertas modelo LSTM, (Zhang et al., 2021).

Cada celda LSTM tiene tres componentes principales: la compuerta de entrada, la compuerta de olvido y la compuerta de salida.

2.4.3. Hiperparámetros

2.4.3.1. Optimización de Hiperparámetros

Los hiperparámetros son parámetros cuyo valor se establece antes del proceso de entrenamiento y no se aprenden directamente de los datos. El rendimiento de todo modelo de aprendizaje automático depende de sus hiperparámetros. Estos controlan la arquitectura, el comportamiento o la estructura del modelo estadístico subyacente (Zhang et al., 2021). Algunos ejemplos comunes de hiperparámetros en redes neuronales incluyen: Taza de aprendizaje, número de capas ocultas y neuronas, función de activación, tamaño del lote (batch size), número de épocas, regularización (L1, L2, dropout). Para la selección de estos existen varios métodos y técnicas (Zhang et al., 2021; Goodfellow et al., 2016), Sin embargo, en la práctica no existe una forma general de elegir los hiperparámetros. En su lugar, los hiperparámetros se establecen a menudo por ensayo y error (Fig.2.33).

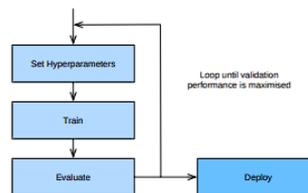


Figura 2.33: Diagrama de flujo de entrenamiento iterativo para un modelo con diferentes hiperparámetros, (Zhang et al., 2021).

2.4.3.2. Funciones de Activación

En el contexto de las redes neuronales artificiales una función de activación está descrita como una función matemática que transforma la salida de una neurona antes de que se pase a la siguiente capa. Esta función transfiere la activación ponderada de la neurona (Brownlee, 2017). La transformación originada por la función introduce no linealidades en el modelo, lo que permite que la red aprenda patrones no lineales, lo cual es crucial para tareas como clasificación, regresión entre otras.

Sin una función de activación, una red neuronal solo podría realizar transformaciones lineales, lo cual sería equivalente a un modelo lineal independientemente de su profundidad. Es decir la combinación de múltiples capas sería equivalente a una sola capa, esto limita su capacidad de la red para resolver problemas complejos.

2.4.3.3. Tasa de Aprendizaje

La tasa de aprendizaje (learning rate) en el contexto de redes neuronales determina la magnitud de los ajustes realizados a los pesos de la red en cada iteración del entrenamiento. Controla cuánto se actualizan los pesos en cada iteración basándose en

el gradiente de la función de pérdida calculado al final de cada lote (Brownlee, 2017). Es decir, define qué tan rápido o lento la red aprende de sus errores.

La tasa de aprendizaje influye directamente en la velocidad y calidad del aprendizaje. Cuando la tasa de aprendizaje es demasiado pequeña, el entrenamiento no sólo es más lento, sino que puede mostrar una disminución muy lenta en la pérdida (Goodfellow et al., 2016). Por otro lado cuando la tasa de aprendizaje alta puede mostrar una disminución abrupta en la pérdida seguida de oscilaciones o un aumento inadvertidamente en el error de entrenamiento en lugar de reducirlo (Goodfellow et al., 2016).

2.4.3.4. Tamaño de Lote

El tamaño de lote (batch size) en el contexto de redes neuronales es clave para el entrenamiento, este define cuántas muestras del conjunto de datos de entrenamiento se procesan antes de que el modelo actualice sus pesos (Brownlee, 2017) basado en el cálculo del gradiente de la función de pérdida del lote. El tipo de entrenamiento varía según el tamaño de lote, en un lote completo (Batch Gradient Descent) se emplea todo el conjunto de datos en cada iteración para calcular el gradiente, en un Mini-Lote (Mini-Batch Gradient Descent) el conjunto de datos se divide en pequeños lotes (e.g., de 32, 64, 128 muestras), y cada iteración del entrenamiento se basa en uno de estos lotes, este es el enfoque más común y el Descenso de Gradiente Estocástico (SGD) que emplea una única muestra del conjunto de datos para cada actualización de los pesos. Este concepto es fundamental para entender cómo se optimizan los modelos de aprendizaje profundo y cómo se equilibra la eficiencia del entrenamiento con la calidad del modelo.

2.4.3.5. Numero de Épocas

El número de épocas es un hiperparámetro clave en el entrenamiento de redes neuronales, una época representa un ciclo completo, osea una completa pasada de todo el conjunto de datos de entrenamiento a través del modelo. Durante cada época, el modelo procesa todos los lotes (o batches) del conjunto de datos, ajustando los pesos en función del gradiente calculado.

El número de épocas es un hiperparámetro crucial que determina la duración del entrenamiento. Un número de épocas insuficiente puede llevar a un modelo subentrenado donde la red no ha capturado los patrones subyacentes en los datos por ende no generaliza bien a nuevos datos, por otro lado un número excesivo puede provocar un sobreentrenamiento produciendo que la red se ajusta demasiado a los datos de entrenamiento, capturando incluso el ruido aleatorio, y disminuye la capacidad de generalizar a nuevos datos.

Capítulo 3

Hipótesis y Variables

3.1. Hipótesis

3.1.1. Hipótesis General

El uso de un modelo de predicción basado en redes neuronales artificiales, que integre variables geológicas, físicas y meteorológicas del Valle del Cusco, permite realizar un mapeo predictivo de la concentración de radón en el suelo.

3.1.2. Hipótesis Específicas

- La implementación de un protocolo estandarizado para realizar mediciones in situ de la concentración de radón en el suelo del Valle del Cusco permitirá obtener datos que reflejan los niveles reales de concentración de radón en el suelo.
- La recopilación de información y el uso de sistemas de información geográfica (SIG) para desarrollar mapas detallados de los principales parámetros geológicos y físicos del Valle del Cusco, permitirá identificar las áreas con mayor concentración potencial de radón y mejorar la precisión de los modelos predictivos.
- Un modelo de redes neuronales artificiales entrenado con datos de mediciones in situ de concentración de radón y variables geológicas y físicas puede predecir los niveles de concentración de radón en zonas del Valle del Cusco donde no se han realizado mediciones directas, mediante la identificación de patrones y correlaciones entre las variables.

Capítulo 4

Métodos y equipos

4.1. Área de estudio

La provincia del Cusco (Fig. 4.1) se encuentra ubicada en el corazón de la Cordillera de los Andes en la región sur-oriental del Perú, en la zona central y sur-oriental de la región del Cusco y es una de las trece provincias que conforman la región del Cusco. La provincia del Cusco comprende 8 distritos, abarca un área de 617 Km^2 aproximadamente y posee una población de 447,588 habitantes aproximadamente (INEI, 2017). Limita por el norte con las provincias de Calca y Urubamba, por el sur con la provincia de Paruro, por el este con la provincia de Quispicanchi, y por el oeste con la provincia de Anta. La provincia se ubica entre los 3138 m.s.n.m. (distrito Saylla) y 3635 m.s.n.m. (distrito Ccorcca) y la ciudad de Cusco se ubica en un valle interandino, entre los 3,244 m.s.n.m (distrito San Jerónimo) y los 3,700 m.s.n.m. (sector del Arco de Tica-Tica).

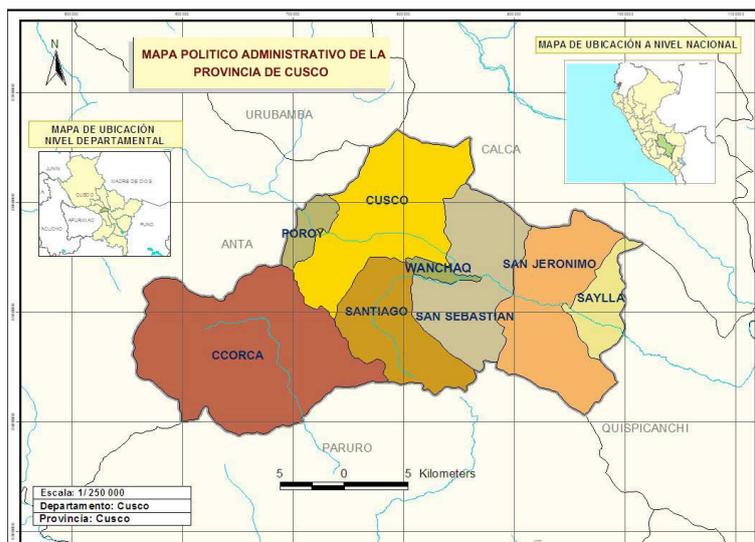


Figura 4.1: Mapa político provincia Cusco.

4.2. Tipo de Investigación

Este estudio se enmarca en una investigación de tipo descriptivo-correlacional, buscando describir las características tanto de la concentración de radón en el suelo (variable a predecir) como de los factores potencialmente asociados (variables predictoras: geológicas, edáficas, ambientales, etc.) en el área de estudio. Posteriormente, se enfocaría en el aspecto correlacional entre las variables predictoras y los niveles de radón, utilizando las técnicas de machine learning como herramienta fundamental para modelar estas relaciones no lineales complejas y desarrollar un modelo capaz de predecir la concentración de radón en suelo.

4.3. Impacto de la Investigación

Este estudio multidisciplinario aportará avances en la comprensión de la física del radón, contribuirá al estudio de la radiación y los procesos de decaimiento del radón, proporcionando datos cuantitativos sobre su concentración en el suelo y su distribución espacial. Esto permitirá un mejor entendimiento de los mecanismos de liberación y dispersión del radón. Además, ayudará a comprender la relación entre la geología de la región y la concentración de radón en el suelo.

4.4. Técnicas de recolección de información

La recolección de datos precisa y relevante es un paso crítico en el desarrollo de modelos de machine learning. Invertir tiempo y esfuerzo en esta etapa garantiza que el modelo sea confiable, preciso y útil.

- Monitoreo insitu de radón: Por medio de instrumentación de campo para medir las concentraciones de radón en el suelo, como cámaras de ionización, que capturan datos en tiempo real de las concentraciones de radón en ubicaciones específicas.
- Bases de datos geológicas y mapas de suelos: Fuentes que proveen información sobre la composición y estructura geológica del área de estudio.
- Análisis documental: Recopilar información de fuentes documentales que pueden dar contexto al comportamiento del radón en el suelo o estudios previos.
- Mediciones meteorológicas y ambientales: Herramientas para medir variables como temperatura, humedad, presión, y condiciones del suelo, ya que estas afectan la migración y concentración de radón.
- Estudios de mapas de riesgos geológicos: Documentos técnicos que exploran zonas con altos niveles de radón, influencias sísmicas, y otros factores geológicos que podrían impactar en la concentración de radón en la zona de estudio.

4.5. Herramientas de Análisis de la Información

- Herramientas para el análisis de datos: Los datos relacionados serán procesados empleando librerías especializadas de Python como Pandas para manipulación de datos y análisis exploratorios, Numpy para cálculos numéricos avanzados y manejo eficiente de matrices, Scipy para realizar análisis estadístico, pruebas de hipótesis y procesamiento avanzado de datos. Scikit-learn, TensorFlow y Keras se emplearán para construir redes neuronales que modelen los patrones no lineales en las concentraciones de radón en función de factores geológicos y meteorológicos. Por otro lado la visualización e interpretación de resultados se dará por medio de un mapeo geoespacial empleando herramientas como QGIS o ArcGIS para visualizar la distribución espacial de las concentraciones de radon, proporcionando una interpretación geográfica del fenómeno en estudio.

4.6. Metodología

La metodología que se propone esta basada en estudios previos realizados en otros contextos (Elío et al., 2023; Al-Khateeb et al., 2017; Rezaie et al., 2023; Oni et al., 2022) para el desarrollo de un modelo de machine learning basado en redes neuronales LSTM para la estimación de la concentración de radón en el gas del suelo a partir de datos de parámetros físicos y geológicos de la zona, obtenidos por las mediciones insitu de concentración de radón y los mapas de los predictores elegidos basados en datos geoespaciales de la zona de estudio, esta descrita por las siguientes fases como se observa en la Figura 4.2.

- Recolección de Datos: Mediciones de radón y generación de mapas insumos geoespaciales.
- Preparación de datos: Limpieza, preprocesamiento y división de la data de manera que quede debidamente estructurada para los algoritmos.
- Desarrollo del Modelo de Red Neuronal: Selección de modelo, arquitectura e implementación utilizando bibliotecas especializadas.
- Entrenamiento del Modelo y evaluación del modelo: Alimentar la red neuronal con la data recolectada, optimización de hiperparametros y evaluación de rendimiento.

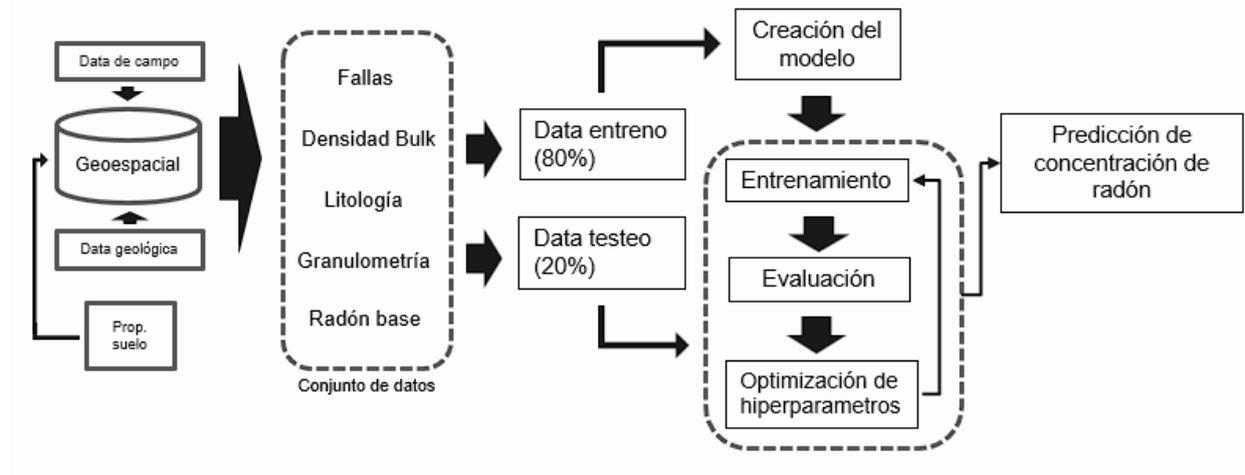


Figura 4.2: Esquema de la metodología

4.6.1. Medición activa de radón en el gas del suelo

En el valle del Cusco se realizó la medición activa de radón en el periodo de octubre y noviembre del año 2022 y los meses de abril y agosto del año 2023. Los puntos de medición están distribuidos en 14 transectos y en grilla, cubriendo diferentes zonas relacionadas con la presencia de fallas activas. La medición activa se realizó mediante los monitores de radón SARAD RTM 2200 (Fig.4.3(b)) y RAD7 (Fig.4.3(a)). Ambos sistemas se basan en el análisis del espectro de energía de partículas alfa. El monitor de radón/toron RTM 2200 (fabricado por SARAD Ltd.), es una estación de medición multiparametro escalable, para monitoreo ambiental y análisis geofísicos. Posee un rango de detección de entre 1 - 10 000 000 (Bq/m^3), el límite de detección y el error de medición del SARAD RTM 2200 es de 500 (Bq/m^3) y $\pm 5\%$ respectivamente Chen et al. (2018). Por otro lado el monitor de radón RAD 7 (fabricado por Durrige Company, Inc) posee un rango de detección de entre 4 - 750,000 (Bq/m^3).



(a) Durrige - RAD 7



(b) SARAD - RTM 2200

Figura 4.3: Monitores de radón

El procedimiento de medición se basa en la inserción de una sonda metálica en el suelo a una profundidad promedio de 60cm, el equipo extrae el gas del suelo por medio de una bomba instalada en su interior con un caudal promedio de 0.3 L/min, este pasa por una cámara con un material desecante en su interior para evitar problemas ocasionados con la humedad, a continuación pasa por un filtro de membrana tipo disco de $0.45\mu\text{m}$ para finalmente ingresar al detector, donde este asocia el número de cuentas de iones de Po-218 detectados a la concentración de radón y análogamente el número de cuentas de iones de Po-216 a la concentración de torón, esta configuración se puede evidenciar en la Figura 4.5.

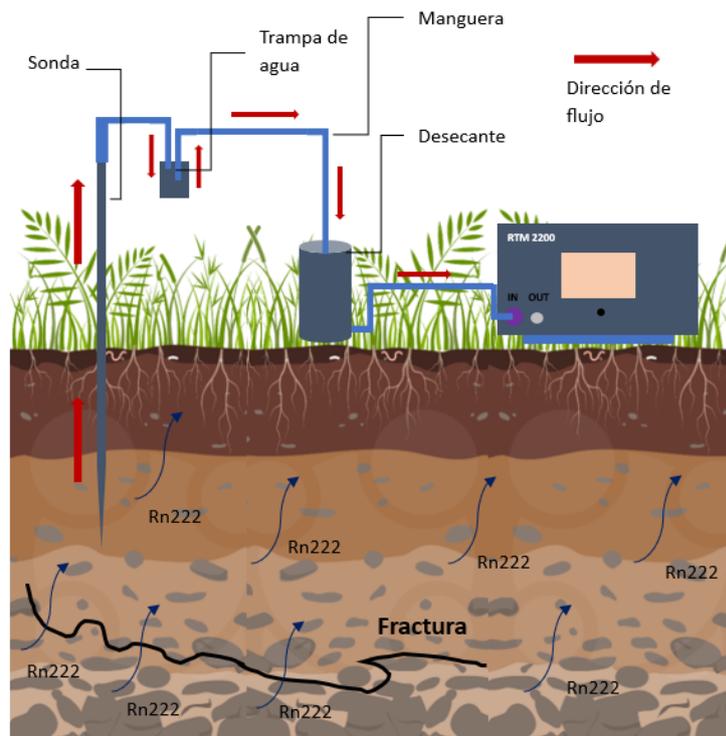


Figura 4.5: Esquema para medir gas de suelo con el monitor RTM 2200

Se realizarán múltiples mediciones de concentración de radón en suelo en distintos puntos de la ciudad del Cusco (Fig. 4.6), la configuración de medición del monitor fue establecida en ciclos de 5 minutos durante un periodo de 45 minutos. Cabe resaltar que se debe tener en cuenta el tiempo necesario para alcanzar el equilibrio entre los átomos de polonio (Po-218) y radón (Rn-222) el cual ocurre después de aproximadamente cuatro o cinco semividas, dado que el tiempo de semivida del Po-218 es de 3.10min aproximadamente. Por otro lado en cambio el estado de equilibrio átomos de polonio (Po-216) y torón (Rn-220) se alcanza rápidamente debido a su corta semivida de 0,145s.

Podemos apreciar como en las diversas mediciones de radón en el gas de suelo se evidencian la variabilidad significativa de concentración de radón (Fig. 4.6), muchos de estos puntos de muestreo ubicados en zonas urbanas de la ciudad del Cusco.

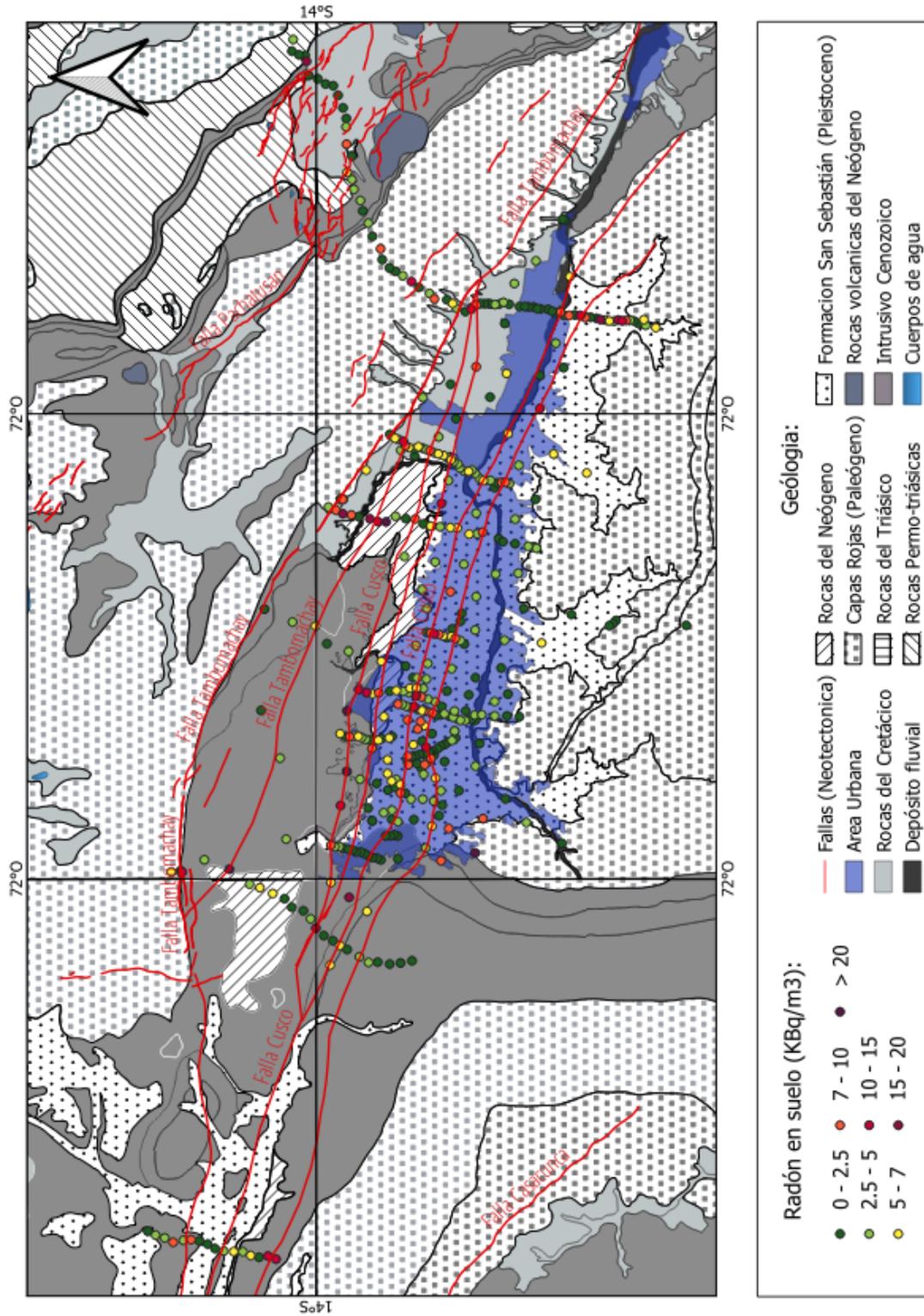


Figura 4.6: Mediciones de radón en gas de suelo - método directo

4.6.2. Generación de mapas insumo

Para este estudio es necesario comprender la variabilidad del gas radiactivo radón en el suelo, ya que esta se encuentra directamente influenciada por factores geológicos y meteorológicos (Elío et al., 2023; Petermann et al., 2021). La generación de estos mapas implica un enfoque técnico que involucra la recopilación de datos radio-métricos, geológicos y ambientales. Para este fin se utilizó las cartas geológicas del cuadrángulo de Cusco a escala 1:500000 del INGEMMET (Carlotto et al., 2011). De manera análoga se procederá a la obtención de datos de parámetros físicos y ambientales adicionales relacionados por medio de bases de datos.

4.6.2.1. Elaboración del mapa de Unidades Litoestratigráficas

Para la elaboración de este mapa se realizó una recopilación de información cartográfica geológica existente. Se utilizaron los mapas a escala 1:50000 (Carlotto et al., 2011) a partir de estos se realizó un análisis de la meta data disponible para la zona de estudio. Se identificó una clasificación ya existente basada en un código único que hace referencia a cada unidad. Partiendo de esto se empleó una categorización por medio de una escala de colores para cada código único asociado a la unidad litoestratigráfica y a las propiedades litológicas y físicas observables.

4.6.2.2. Elaboración del mapa de Granulometría

La granulometría abarca el estudio y la medición del tamaño de las partículas en un suelo, sedimento, roca entre otros. Este predictor posee una gran relevancia en la emanación y migración de radón en el gas de suelo (Rezaie et al., 2023; Petermann et al., 2021; Elío et al., 2017). La elaboración de este mapa se dividió en dos procesos. Primero se realizó la identificación de todas las geologías presentes en la zona de estudio esto por medio de recopilación de información de la zona resumida por (Carlotto et al., 2011), posterior a esto se realizó una clasificación más amplia de estas en unidades. Segundo, teniendo esta nueva clasificación se empleó la escala propuesta por (Wentworth, 1922), de esta manera se clasificó granulométricamente las partículas sedimentarias que componen las diferentes unidades geológicas.

4.6.2.3. Elaboración del mapa de Radón promedio

Para evaluar la relación entre la litología y los niveles de radón en el suelo, resulta útil establecer un nivel promedio de concentración de radón en el suelo en función a los tipos de litología predominantes y las características de cada contexto. Este enfoque permite identificar patrones en las concentraciones de radón y anticipar áreas con potenciales niveles altos o bajos de radón en función del tipo de roca y su distribución en el área de estudio (Ndjana Nkoulou et al., 2024).

Los datos necesarios para la generación del mapa de radón base son obtenidos a partir de las mediciones in situ realizadas por el INGEMMET. Para esto se realiza un

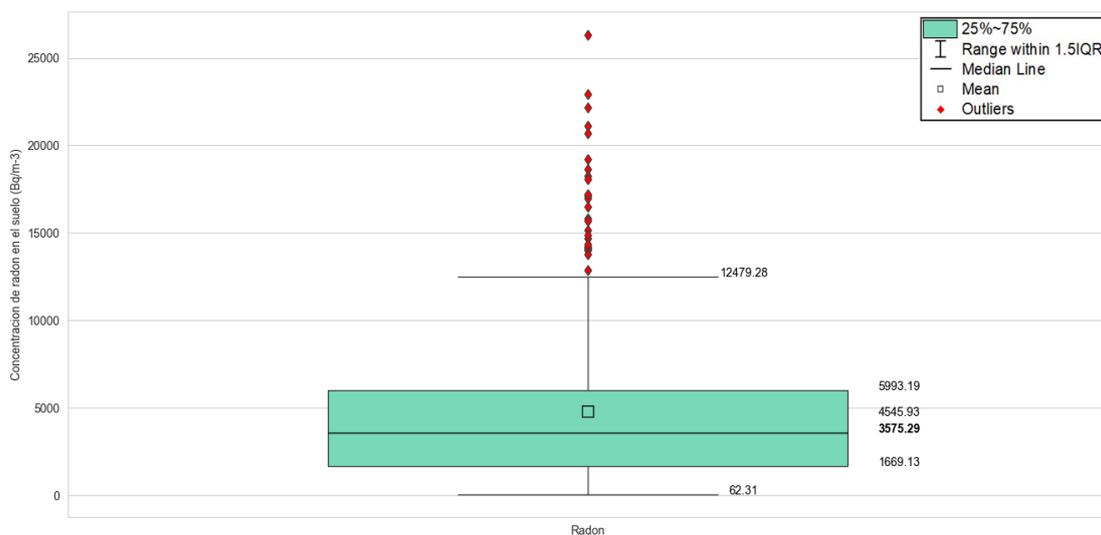


Figura 4.7: Diagrama de cajas, concentración de radón en suelo

análisis mediante un gráfico de cajas (Fig. 4.7). Este método estadístico visual nos permitirá identificar los outliers. Estos outliers son valores que se alejan significativamente del resto de los datos en una distribución, estos se encuentran relacionados a eventos excepcionales o características propias. Por otro lado se realiza el calculo del rango intercuartil (IQR) tabla 4.1 que es una medida de dispersión que describe la extensión de la mitad central del conjunto de datos.

	IQR	IQR Superior	IQR Inferior
Concentracion radon (Bq/m^3)	4032.5	11368.75	-4761.25

Tabla 4.1: Estadísticas descriptivas, rango intercuartil (IQR), rango intercuartil superior y rango intercuartil inferior

Identificados y descartados los outliers que en este caso son valores que se entiende pueden estar relacionados con la presencia de radón técnicamente incrementado, se procede a generar el mapa por medio de una interpolación espacial de los valores puntuales de concentración de radón en el gas del suelo. El método de interpolación empleado fue el de distancia inversa ponderada (IDW) empleado para estimar valores en puntos no medidos a partir del uso de valores conocidos en puntos de muestra (Khanjer and Abdullha, 2017),

4.6.2.4. Elaboración del mapa de Densidad de Bulk

Usualmente lo ideal para el modelamiento de la concentración de radón en suelo es el uso de la permeabilidad como un predictor, ya que esta muestra la capacidad de los suelos para permitir el flujo de un fluido sin alterar su composición. En este caso, fue

imposible obtener data de permabilidad del suelo en la zona de estudio. En cambio se opto por utilizar un parámetro asociado a la permeabilidad de suelo que es la densidad de Bulk o densidad aparente. La densidad de Bulk es la relación entre la masa de un material y su volumen total, incluyendo los espacios vacíos (poros) (Blake, 1965). Se optó por tomar este parámetro por la relación que tiene con otras propiedades físicas como la porosidad y permeabilidad de suelos. La densidad de Bulk posee una relación inversamente proporcional con la porosidad, este fenómeno se debe a que los espacios vacíos ocupan un volumen mayor, lo que conlleva a una reducción en la cantidad de masa contenida en cada unidad de volumen (Šperl and Trčková, 2008). Por otro lado, la relación entre la porosidad y permeabilidad se sustenta en que los materiales con una mayor porosidad exhiben una permeabilidad superior, dado que ofrecen una mayor capacidad para el flujo de fluidos a través de su estructura debido a la presencia de un mayor espacio disponible.

Los valores de Densidad de Bulk fueron obtenidos recopilando información de la base de datos de “Soilgrids” que es un sistema de mapeo digital de suelos basado en una recopilación global de datos de perfiles de suelos (WoSIS) y capas ambientales (Poggio et al., 2021). Se obtuvo el archivo raster de estos datos para la zona de estudio, esta fue procesada utilizando el software Qgis (Version 3.36.3).

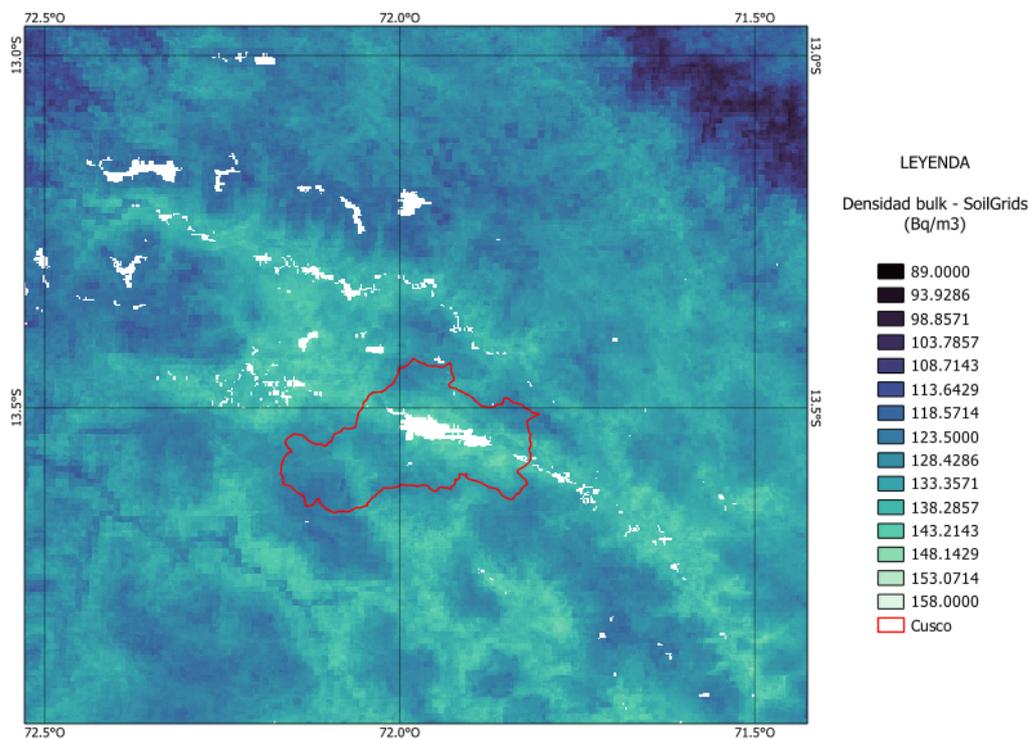


Figura 4.8: Raster densidad de bulk, (Poggio et al., 2021)

Al procesar este se observaron posibles espacios vacíos (celdas sin datos) en el mapa. Se realizó un procesamiento de la imagen ráster empleando el algoritmo “Fill Nodata” derivado de GDAL (GDAL/OGR contributors, 2024). Este algoritmo rellena regiones de un ráster que contienen valores “NoData” por medio de una interpolación desde los bordes (Fig. 4.9). Los nuevos valores en estas regiones “NoData” se calculan

empleando los valores de los píxeles circundantes por medio de una ponderación

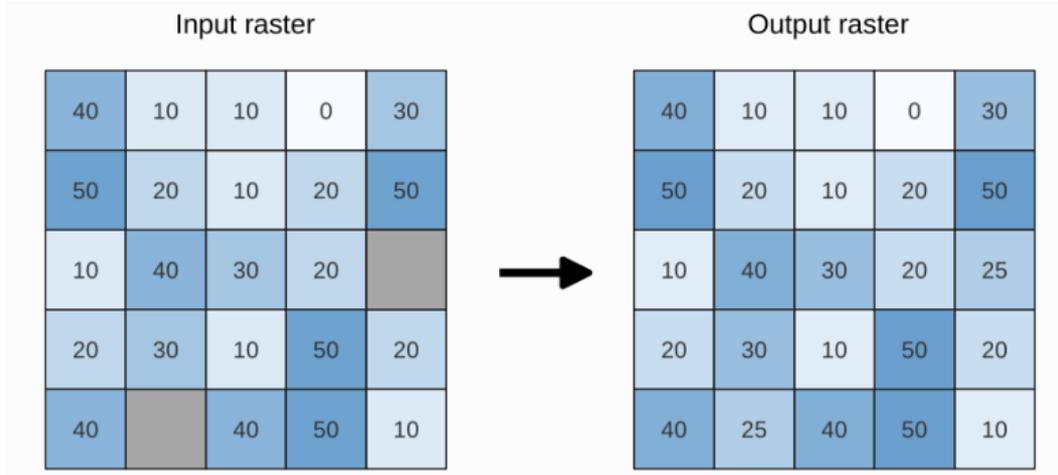


Figura 4.9: Relleno de valores sin datos (en gris) de un ráster

4.6.2.5. Elaboración del mapa de Proximidad a fallas

Para la elaboración de un mapa de proximidad a fallas de la zona, es necesario conocer la ubicación de fallas geológicas identificadas. Esta información sobre la presencia de fallas geológicas se obtuvo por medio de los estudios de neotectónica realizados por el INGEMMET (Benavente et al., 2013). Estas fallas identificadas se encuentran disponibles como vectores en los archivos Shapefile para cada cuadrángulo (Fig. 4.10).

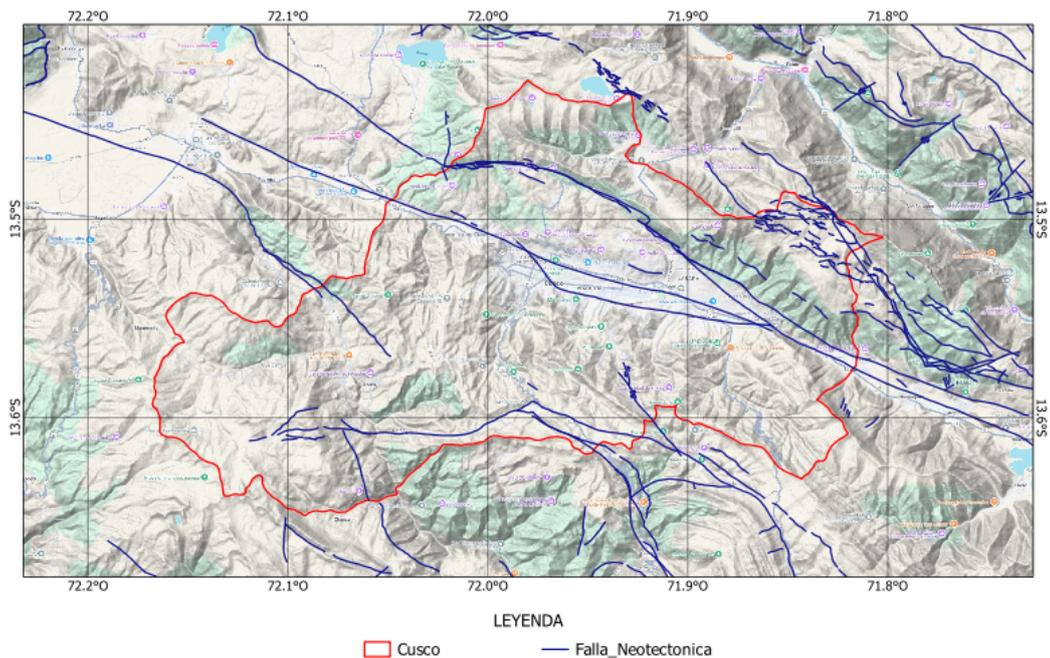


Figura 4.10: Fallas geológicas activas identificadas Benavente et al. (2013)

Para generar el mapa de proximidad a fallas es necesario realizar un procesado de estos vectores que representan las fallas, como estos son una capa vectorial primero debemos convertirla a un ráster, después procedemos al análisis de proximidad por medio del algoritmo "gdal proximity"(GDAL/OGR contributors, 2024). Este proceso genera un ráster que muestra la distancia entre cada píxel y el píxel objetivo más cercano. Los píxeles objetivo se definen como aquellos en el ráster original con valores específicos dentro de un conjunto establecido, en ese caso las líneas azules (Fig. 4.10)

4.6.3. Preprocesamiento de los datos

El preprocesamiento de datos es una serie de técnicas y prácticas que se implementan a los datos crudos (RAW) antes de poder utilizarlos, los algoritmos de machine learning requieren datos de entrada en un formato específico. El problema surge en la mayoría de conjuntos de datos, ya que estos suelen provenir de diversas fuentes como encuestas, sensores, registros históricos o procesos experimentales, debido a esto, esta data en forma cruda por lo general no representan conjuntos adecuados para su uso directo. Para esto el preprocesamiento de datos se encarga de acondicionar los datos para su utilización en algoritmos.

En machine learning es importante garantizar que los datos sean de alta calidad y consistentes, ya que la posible existencia de data que contenga errores u omisiones pueden llevar a conclusiones erróneas y modelos pobres o deficientes. Para esto se emplea un conjunto de tareas destinadas a limpiar, transformar y organizar los datos para que los algoritmos de Machine Learning logren extraer información precisa, significativa y confiable. Para esto primeramente empleamos el uso de estadísticos descriptivos para las variables no categóricas. La tabla 4.2 proporciona una visión clara de la distribución, tendencia central y dispersión de los datos.

	Granulometria (mm)	Distancia a fallas (m)	Radon base (Bq/m^3)	Densidad bulk (cg/m^3)
count	465	465	465	465
mean	61.6430	548.9859	2968.5615	140.0300
std	211.4758	927.6304	1678.7741	4.6123
min	0.0010	1.0000	123.2249	123.9565
25 %	0.0050	105.5130	1659.8130	137.9641
50 %	0.0200	229.2008	2961.4597	142.0358
75 %	0.1250	605.2834	4143.2835	143.0180
max	1024.0000	6430.2388	6930.8353	145.0000

Tabla 4.2: Estadísticas descriptivas de variables de entrada

4.6.3.1. Limpieza de datos

La limpieza de datos es un paso fundamental en el preprocesamiento, implica identificar y corregir errores, incoherencias e imprecisiones como valores faltantes, atípicos entre otros. En este caso el primer procedimiento fue la eliminación de valores faltantes, para ello se empleo el lenguaje de programación Python (Version 3.12.4) (Van Rossum and Drake, 2009) en el entorno de desarrollo de Visual Studio Code (Version 1.91), en específico se utilizó la librería Pandas (version 2.2.2) (Wes McKinney, 2010), como los valores faltantes son pocos o nulos se optó por el método de eliminación de estos en caso existieran (Fig. 4.11), en el caso de que estos hubieran representado una cantidad significativa es posible realizar una imputación de estos por medio de técnicas apropiadas.

```
import pandas as pd

data_files = pd.read_csv("data.csv", delimiter=';')
print(data_files)
data_x = data_files[["CODI", "Mapa_proximidad_fallas", "Granulomet", "Densidad_bulk", "Radon_directas"]]
data_x = data_x.dropna()
data_y = data_files[["Radon"]]
data_y = data_y.dropna()
print(data_x.describe())
print(data_y.describe())
```

Figura 4.11: Código para la eliminación de valores nulos

A continuación se hizo la detección de valores atípicos, esto implica la identificación de valores que resultan inusuales y presentan una desviación significativa en la distribución del conjunto de datos, esto fue evaluado y los datos anómalos fueron excluidos del conjunto de datos. Así mismo se verificó la consistencia de los datos garantizando que estos sean coherentes, para esto se revisó el formato y las unidades para cada conjunto, ya que como se indicó previamente estos al provenir de múltiples fuentes son susceptibles a diversos tipos de errores. Finalmente se procedió con la identificación y eliminación de posibles registros duplicados (Fig. 4.12) ya que la presencia de estos puede incurrir en sesgos o relaciones erróneas posteriormente.

```
import pandas as pd

data_files = pd.read_csv("data.csv", delimiter=';')
print(data_files)
data_x = data_files[["CODI", "Mapa_proximidad_fallas", "Granulomet", "Densidad_bulk", "Radon_directas"]]
data_x = data_x.dropna()
data_y = data_files[["Radon"]]
data_y = data_y.dropna()
print(data_x.describe())
print(data_y.describe())
data_x.drop_duplicates()
data_y.drop_duplicates()
```

Figura 4.12: Código para la eliminación de valores duplicados

La tabla 4.3 muestra el conjunto de datos (data set) después del proceso de

eliminación de datos duplicados, valores nulos o faltantes.

	Granulometría (<i>mm</i>)	Proximidad fallas (<i>m</i>)	Radon base (<i>Bq/m³</i>)	Densidad bulk (<i>cg/m³</i>)	CODI	Radon (<i>Bq/m³</i>)
0	0.5	266.3907	1939.2475	136.9549	391	1932.20
1	0.005	936.0534	583.1302	136.0561	3	540.97
2	256	2616.8643	227.3141	132.2932	52	387.77
3	0.005	1290.0330	154.2461	135.7194	3	153.95
4	0.005	549.5980	195.7638	135.3068	3	226.58
..
460	0.005	1580.3496	129.2406	136.1375	3	101.89
461	0.005	223.8057	361.0758	135.5252	3	349.36
462	0.015	1288.6473	1615.5952	133.3389	2601	1604.97
463	0.001	5835.7568	1653.2126	137.9358	288	1640.07
464	0.125	177.5049	5123.3732	130.8098	454	12466.36

Tabla 4.3: Tabla del conjunto de datos posterior a la limpieza de data

4.6.3.2. Transformación de datos

La transformación de datos es necesaria para preparar los conjuntos adecuadamente, de manera que estos tengan la forma ideal para los algoritmos. Para este proceso empezamos con la normalización de los datos, esta es crucial para garantizar que todas las variables contribuyan de manera equitativa en el entrenamiento del modelo evitando la presencia de múltiples escalas que puedan influir en el dominio de alguna variable. Existen diversos métodos de normalización básicamente la idea es poder escalar los datos para que estos se encuentren en un rango específico para este caso se optó por una normalización min-max escalado en un rango de (0, 1) para esto empleamos la librería de Python scikit-learn (versión 1.5.1) (Pedregosa et al., 2011) Fig.4.13, en específico la función "MinMaxScaler". esta normalización está basada en la ecuación 4.1 donde x_{min} es el valor mínimo y x_{max} es el valor máximo del conjunto de datos.

$$x_{std} = \frac{(x - x_{min})}{(x_{max} - x_{min})} \quad (4.1)$$

Muchas funciones de activación utilizadas en redes neuronales (como sigmoid y tanh) tienden a ser sensibles a los valores de entrada (inputs) cercanos a cero. La normalización ayuda a asegurar que las entradas estén dentro de este rango óptimo. Finalmente en la tabla 4.4 observamos los inputs (distancia a fallas, granulometría, densidad de bulk, radón base y CODI) y el target (concentración de radón) normalizados

Por otro, al tratarse de variables estas no siempre son representaciones de tipo cuantitativas. Podemos encontrar variables categóricas que son un tipo común de datos los cuales representan atributos cualitativos, estos a diferencias de las variables cuantitativas tienen un conjunto finito de valores discretos. Por ende para que estas variables puedan ser procesadas es necesario realizar una codificación de variables categóricas.

```

from sklearn.model_selection import train_test_split
import pandas as pd
from sklearn.preprocessing import MinMaxScaler

data_files = pd.read_csv("data.csv", delimiter=';')
print(data_files)
data_x = data_files[["CODI", "Mapa_proximidad_fallas", "Granulomet", "Densidad_bulk", "Radon_directas"]]
data_y = data_files[["Radon"]]

scaler = MinMaxScaler(feature_range=(0, 1))
scaler_1 = MinMaxScaler(feature_range=(0, 1))

data_x1 = scaler.fit_transform(data_x)
dx1 = pd.DataFrame(data_x1, columns=["CODI", "Mapa_proximidad_fallas", "Granulomet", "Densidad_bulk", "Radon_directas"])
data_y1 = scaler_1.fit_transform(data_y)
dy1 = pd.DataFrame(data_y1, columns=["Radon"])
dff= pd.concat([dx1, dy1], axis=1)
print(dff)

```

Figura 4.13: Código para la normalización de los datos, rango de 0 a 1

	CODI	Mapa_proximidad_fallas (<i>m</i>)	Granulometria (<i>mm</i>)	Densidad bulk (<i>cg/m³</i>)	Radon base (<i>Bq/m³</i>)	Radon (<i>Bq/m³</i>)
0	0.0894	0.0413	0.0005	0.6177	0.2668	0.0712
1	0.0000	0.1454	0.0000	0.5750	0.0676	0.0182
2	0.0113	0.4069	0.2500	0.3962	0.0153	0.0124
3	0.0000	0.2005	0.0000	0.5590	0.0046	0.0035
4	0.0000	0.0853	0.0000	0.5394	0.0107	0.0063
..
460	0.0000	0.2457	0.0000	0.5788	0.0009	0.0015
461	0.0000	0.0347	0.0000	0.5498	0.0349	0.0109
462	0.5985	0.2003	0.0000	0.4459	0.2192	0.0588
463	0.0657	0.9075	0.0000	0.6643	0.2247	0.0601
464	0.1039	0.0275	0.0001	0.3257	0.7345	0.4725

Tabla 4.4: Conjunto de datos normalizado

En nuestro caso la variable categórica que empleamos esta relacionada con la geología de la zona, esta variable describe el tipo de formación geológica presente.

Para la codificación de esta variable se opto por la técnica de codificación de etiquetas donde se asigna un valor numérico único arbitrario a cada categoría (Petermann et al., 2021). Se utilizo para esta codificación los valores de ID asignados por el INGEMMET para cada categoría según corresponde como se observa en la Tabla 4.5 en la columna CODI.

CODI	NAME	UNIDAD
3	Qh-al	Depósito aluvial
4	Qh-cl	Depósito coluvial
52	PET-m	Grupo Mitu
121	Q-cl	Depósito coluvial
161	Q-gl	Depósito glaciar
182	P-di,tn	Diorita, tonalita
193		Pueblo
194		Laguna
227	Ks-puq	Formación Puquín
288	PEc-c	Grupo Copacabana
391	P-m	Formación Muñani
419	D-c	Grupo Cabanillas
453	Qp-ss	Formación San Sebastián
454	NpQ-r	Formación Rumicolca
455	Np-chin	Formación Chincheros
459	Kis-ayb	Formación Ayabacas
461	Ki-pau	Formación Paucarbamba
462	Ki-hua	Formación Huancané
550	Q-fl	Depósito fluvial
551	Q-glfl	Depósito glaciar, fluvial
1399	PoNm-pu/2	Formación Punacancha - Miembro 2
1458	PoNm-pu/1	Formación Punacancha - Miembro 1
1459	Peo-so/s	Formación Soncco - Miembro superior
1460	Peo-so/i	Formación Soncco - Miembro inferior
1462	Ti-p	Formación Pachatusán
1844	Kis-ma	Formación Maras
1895	PN-t	Grupo Tacaza
2572	Q-al	Depósito aluvial
2580	Peo-so	Formación Soncco
2581	Kis-a/s	Formación Arcurquina - Miembro superior
2582	Kis-a/m	Formación Arcurquina - Miembro medio
2583	Kis-a/i	Formación Arcurquina - Miembro inferior
2601	KsP-a	Formación Auzangate
3277	Ks-vi	Formación Vilquechico
4052	PN-se-cmdi	Unidad Senja - cuarzomonzodiorita
4344	Peo-k	Formación Kayra

Tabla 4.5: Tabla de codificación por etiquetas de la litología del valle del cusco, (Carlotto et al., 2011)

4.6.3.3. Reducción de dimensionalidad

Una de las técnicas para la reducción de la dimensionalidad es la selección de características, existen diferentes métodos para la selección de estas, en este caso al tratarse de un sistema que involucra distintos factores espaciales como lo son la litología, permeabilidad del suelo, porosidad del suelo, tamaño de grano, elevación, pendiente, distancia a fallas, distancia a lineamientos, distribución de radionuclidos como uranio o radio, propiedades geoquímicas del suelo entre otros (Rezaie et al., 2023; Petermann et al., 2021) y factores temporales como temperatura, humedad relativa, velocidad de viento, precipitación, presión barométrica entre otros (Petermann et al., 2021; Elío et al., 2023).

Para el modelado de su comportamiento, todos estos en conjunto son factores que influyen en la concentración de radón en el gas del suelo, pero, no todos estos atributos son igual de importantes, por medio de la selección de características se identifica los mas relevantes para el problema en cuestión. Basado en los objetivos del estudio se opto por solo trabajar con factores espaciales, por medio de una correlación de pearson se determino las relaciones presentes entre las variables elegidas para nuestro estudio, como (Rezaie et al., 2023) demostró, para la mayoría de las factores existen correlaciones lineales débiles, esto indica que la mayoría de las variables son independientes.

4.6.3.4. División del conjunto de datos

La división del conjunto de datos para el entrenamiento de una red neuronal es un proceso fundamental. Un problema común que puede surgir en el entrenamiento de esta es el sobreajuste, ocurre cuando el modelo aprende demasiado bien los detalles del conjunto de datos, como consecuencia este no generaliza y resulta en un rendimiento deficiente en datos nuevos. La división del conjunto de datos ayuda a prevenir que el modelo se adapte demasiado a los detalles y el ruido de este, así el modelo no memoriza los datos de entrenamiento y en su lugar pueda aprender patrones. Generalmente, se divide en tres subconjuntos: entrenamiento, validación y prueba (testeo) como se observa en la Fig.4.14.

- Conjunto de entrenamiento: Utilizado para ajustar los pesos y biases de la red neuronal, permitiendo aprender patrones y características de de los datos.
- Conjunto de validación: Utilizado para evaluar el rendimiento de la red durante el entrenamiento y ajustar los hiperparámetros. Ayuda a prevenir el sobreajuste y determina cuándo detener el entrenamiento.
- Conjunto de prueba (testeo): Utilizado para evaluar el rendimiento final del modelo entrenado, proporciona una estimación imparcial del rendimiento del modelo en datos no vistos.

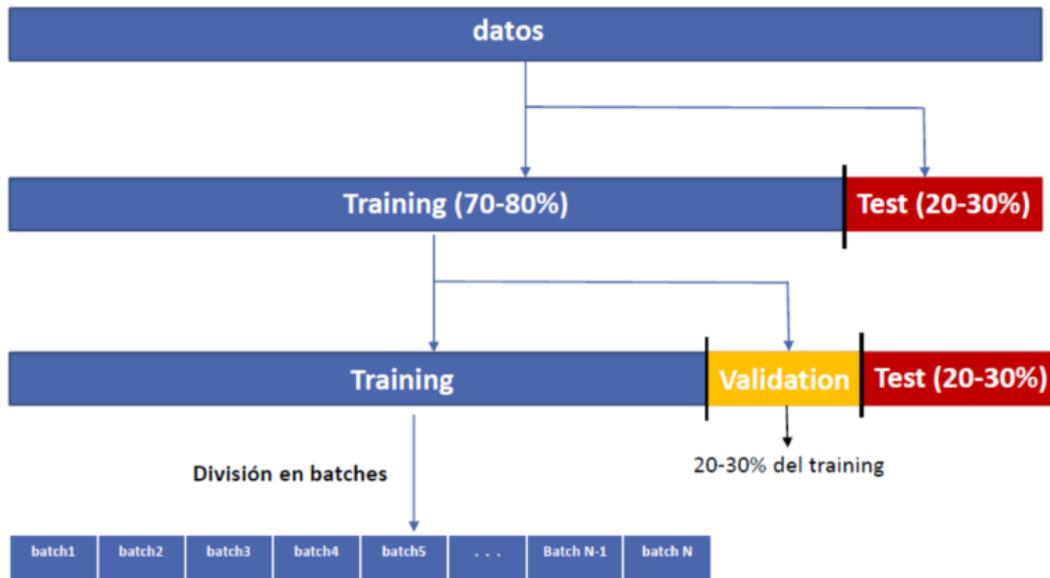


Figura 4.14: Relleno de valores sin datos (en gris) de un ráster

Las proporciones de división del conjunto de datos pueden variar según el tamaño del conjunto de datos y el problema específico que se está tratando. Para este caso las proporciones de división fueron 80 % para el conjunto de entrenamiento, 10 % para el conjunto de validación y 10 % para el conjunto de prueba. La tabla 4.6 muestra el conjunto de datos final después del proceso de limpieza, transformación y reducción. Sobre este conjunto se realizara la división.

	Longitud	Latitud	Granulometría	Proximidad a fallas	Radón base	Densidad de bulk	CODI	Radón
0	-71.935944	-13.488457	0.5	266.3907	1939.2475	136.9549	391	1932.20
1	-71.911536	-13.470378	0.005	936.0534	583.1302	136.0561	3	540.97
2	-71.877954	-13.452074	256	2616.8643	227.3141	132.2932	52	387.77
3	-71.927629	-13.482708	0.005	1290.0330	154.2461	135.7194	3	153.95
4	-71.908239	-13.469131	0.005	549.5980	195.7638	135.3068	3	226.58
...
460	-71.926213	-13.480532	0.005	1580.3496	129.2406	136.1375	3	101.89
461	-71.905403	-13.468146	0.005	223.8057	361.0758	135.5252	3	349.36
462	-71.886597	-13.460647	0.015	1288.6473	1615.5952	133.3389	2601	1604.97
463	-71.862813	-13.426488	0.001	5835.7568	1653.2126	137.9358	288	1640.07
464	-71.954397	-13.425956	0.125	177.5049	5123.3732	130.8098	454	12466.36

Tabla 4.6: Conjunto de datos obtenido después del pre procesamiento

La división se puede realizar empleando diversas librerías de Python como Scikit-learn, Pandas o TensorFlow, en este caso se empleo Scikit-learn usando la función "train test split"(Fig.4.15). Esta se encarga de dividir matrices en subconjuntos de datos de prueba y de entreno basado en la recepcion de dos parametros (X) e (Y), siendo en la Fig.4.15 "X"los predictores e "Y"los valores de concentracion de radón obtenidos por mediciones insitu.

```

from sklearn.model_selection import train_test_split
import pandas as pd

data_files = pd.read_csv("data.csv", delimiter=';')
print(data_files)
data_x = data_files[["CODI", "Mapa_proximidad_fallas", "Granulometria", "Densidad_bulk", "Radon_base"]].values
data_y = data_files[["Radon"]].values
Xtrain, X_test, y_train, y_test = train_test_split(data_x, data_y, test_size=0.2, random_state=0)

```

Figura 4.15: Código para la división del conjunto de datos

4.6.4. Desarrollo de la red neuronal

4.6.4.1. Diseño del modelo LSTM

El diseño de un modelo Long Short-Term Memory (LSTM) involucra un proceso multietapa, abarca desde la definición de la arquitectura, la configuración de los hiperparámetros, el entrenamiento y la evaluación de esta. La arquitectura seleccionada para este caso de estudio fue el uso de redes LSTM y se seleccionó Python como lenguaje de programación principal para implementar todo el proceso de desarrollo, entrenamiento y predicción. Python dada su versatilidad y la amplia gama de bibliotecas disponibles para el procesamiento de datos, visualización y, en particular, el aprendizaje profundo es una excelente opción. En este contexto, se eligió Keras, una API de alto nivel especializada en aprendizaje profundo construida sobre TensorFlow, fue la herramienta escogida para la implementación, desarrollo y entreno del modelo LSTM debido a su facilidad de uso y eficiencia. De esta manera esta combinación de herramientas (Python, Keras y TensorFlow) (Fig.4.16) proporcionó un entorno de desarrollo robusto y flexible para abordar el problema planteado.

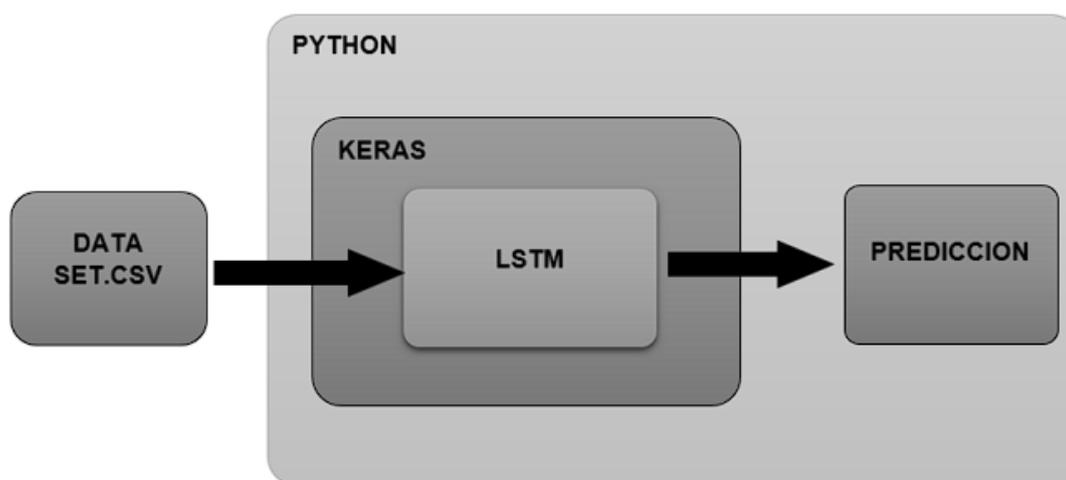


Figura 4.16: Arquitectura modelo LSTM

Por otro lado en el diseño de la red LSTM por capas se optó por una arquitectura LSTM bidireccional apilada Brownlee (2017). Esta se basa en los conceptos de arquitecturas LSTM apiladas Fig.4.17 y arquitecturas LSTM bidireccionales Fig.4.18. En la

primera el modelo se conforma por múltiples capas LSTM ocultas y cada una con múltiples celdas de memoria (Brownlee, 2017), estas capas ocultas tiene por consecuencia que el modelo sea mas profundo.

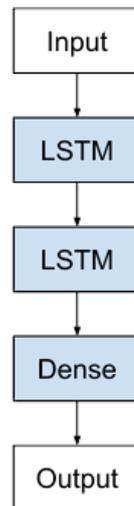


Figura 4.17: Arquitectura LSTM apilada; (Brownlee, 2017)

En la segunda el modelo se basa en la suposición de que toda la secuencia está disponible (Brownlee, 2017), esto recorriendo los pasos de tiempo de la secuencia hacia adelante (Forward LSTM) y también hacia atrás (Backward LSTM).

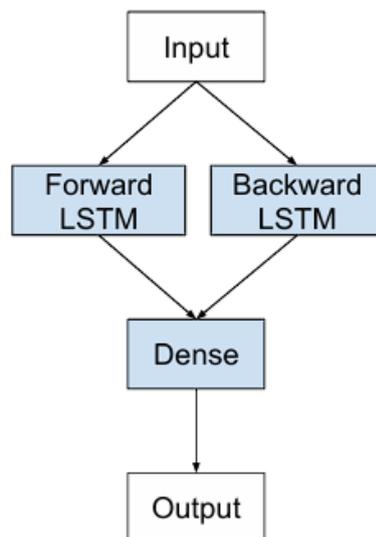


Figura 4.18: Arquitectura LSTM bidireccional; (Brownlee, 2017)

Como se menciona el diseño de la arquitectura de la red LSTM es por medio de capas, primero se genera la capa de entrada (Fig.4.19) (input layer) en la cual es

necesario definir la dimensionalidad para las entradas (inputs) de la red, que para este caso corresponden a la cantidad de variables (predictores) seleccionados, luego se generan las capas ocultas (Fig.4.19) (hidden layers), es necesario establecer el número de capas LSTM y el número de nodos ocultos por cada capa, finalmente la capa de salida (Fig.4.19) (output layer) que se compone de una única capa y un solo nodo. Para cada una de estas tres capas se define los parámetros necesarios.

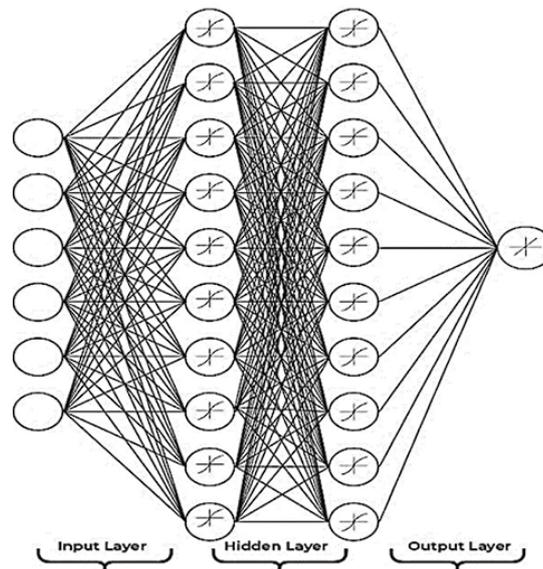


Figura 4.19: Estructura de red neuronal profunda

De esta manera se desarrollo un modelo de aprendizaje profundo que emplea una estructura de red neuronal recurrente. Concretamente redes neuronales recurrentes bidireccionales. El componente central de esta red como se menciona es el uso de capas LSTM bidireccionales. La Fig.4.20 muestra la implementación de las distintas capas mencionadas por medio de Tensorflow y Keras.

```
import tensorflow as tf
from keras.layers import Layer
from keras.layers import Conv2D, MaxPooling2D, Flatten, Dense, Dropout
from keras.layers import LSTM, Dense, Bidirectional
from keras.initializers import HeNormal

model = tf.keras.models.Sequential(
    [
        tf.keras.layers.InputLayer((1,5)),
        tf.keras.layers.Bidirectional(LSTM(10, return_sequences=True, activation="relu")),
        tf.keras.layers.Bidirectional(LSTM(50, return_sequences=True, activation="relu",kernel_initializer=HeNormal())),
        tf.keras.layers.Bidirectional(LSTM(50, return_sequences=True, activation="relu",kernel_initializer=HeNormal())),
        tf.keras.layers.Dense(1, activation="sigmoid", kernel_initializer=HeNormal())
    ],
)

model.compile(
    loss = tf.keras.losses.MeanSquaredError(),
    metrics = tf.keras.metrics.RootMeanSquaredError(),
    optimizer = tf.keras.optimizers.legacy.Nadam(learning_rate=0.001, decay=1e-6)
)
```

Figura 4.20: Implementacion del codigo para generar la red

4.6.4.2. Entrenamiento del modelo LSTM

El entrenamiento de la red LSTM bidireccional, se realiza utilizando el conjunto de datos previamente dividido.

```
import tensorflow as tf
from keras.layers import Layer
from keras.layers import Conv2D, MaxPooling2D, Flatten, Dense, Dropout
from keras.layers import LSTM, Dense, Bidirectional
from keras.initializers import HeNormal

model = tf.keras.models.Sequential(
    [
        tf.keras.layers.InputLayer((1,5)),
        tf.keras.layers.Bidirectional(LSTM(10, return_sequences=True, activation="relu")),
        tf.keras.layers.Bidirectional(LSTM(50, return_sequences=True, activation="relu", kernel_initializer=HeNormal())),
        tf.keras.layers.Bidirectional(LSTM(50, return_sequences=True, activation="relu", kernel_initializer=HeNormal())),
        tf.keras.layers.Dense(1, activation="sigmoid", kernel_initializer=HeNormal())
    ],
)
model.compile(
    loss = tf.keras.losses.MeanSquaredError(),
    metrics = tf.keras.metrics.RootMeanSquaredError(),
    optimizer = tf.keras.optimizers.legacy.Nadam(learning_rate=0.001, decay=1e-6)
)
model.fit(
    x = X,
    y = Y,
    epochs = 350,
    batch_size = 2,
    #shuffle = False
)
```

Figura 4.21: Diseño red LSTM

Se alimentara la red LSTM con los datos de entrenamiento donde la matriz de datos (X) son los datos de entrada compuestos por las variables predictoras (densidad de bulk, granulometría, proximidad a fallas, CODI y radón base). Estos fueron obtenidos a partir de los mapas insumo generados. Por otro lado la matriz de datos (Y) que contiene los targets asignados (concentracion de radon en suelo), estos fueron obtenidos por medio de las mediciones insitu. Se necesita que los datos se encuentren en forma de secuencias. Esto implica dividir los datos en pequeñas ventanas (secuencias) que la red pueda procesar. Es necesario definir algunos parámetros iniciales para el entrenamiento como el tamaño de lote (batch size), el numero de épocas de entrenamiento, el optimizador, la tasa de aprendizaje y la función de perdida. La Fig.4.21 muestra la implementacion de estos parametros para este caso se opto por un tamaño de lote de 2, un numero de 350 epocas de entrenamiento, la función Nadam como optimizador con una tasa de aprendizaje de 0.001 y MeanSquaredError como funcion de perdida.

4.6.5. Evaluación de la red neuronal

4.6.5.1. Mean absolute error - MAE

El Error Absoluto Medio (MAE, por sus siglas en inglés) es una métrica de evaluación que se utiliza en modelos predictivos. Es una de las métricas de error más sencillas y es ampliamente usada debido a su interpretación intuitiva. Esta nos indica

qué tan lejos, en promedio, están las predicciones de nuestro modelo en comparación con los valores reales.

$$MAE = \frac{1}{n} \sum_{i=1}^n |u_i - \hat{u}_i| \quad (4.2)$$

El MAE se calcula como el promedio de las diferencias absolutas entre los valores reales y los valores predichos (ecuación 4.2). Es decir, se toman las diferencias entre cada valor real y su correspondiente valor predicho, se convierten en valores absolutos y finalmente se calcula el promedio de todas estas diferencias.

4.6.5.2. Root mean square error - RMSE

El Root Mean Square Error (RMSE, por sus siglas en inglés). El RMSE se utiliza para medir el grado de desviación estándar entre la predicción y los valores de observación (Feng et al., 2022). Mediante el cálculo de la raíz cuadrada del promedio de los errores al cuadrado entre los valores predichos y los valores observados (ecuación 4.3), se utilizan para evaluar la compatibilidad predictiva entre modelos (Rezaie et al., 2023). La ventaja del RMSE es que amplifica los errores grandes, lo que puede ser útil en aplicaciones donde los errores grandes son particularmente problemáticos.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (u_i - \hat{u}_i)^2} \quad (4.3)$$

4.6.6. Procedimiento de generación de un mapa de concentración de gas radón en suelo por medio de Machine Learning

La generación de un mapa de concentración de gas radón en el suelo por medio de Machine Learning es un proceso que permite estimar y visualizar los niveles de radón en áreas donde no se cuenta o se complica la realización de mediciones directas empleando un modelo previamente entrenado. Este proceso involucra la recopilación de datos de concentración de radón en ubicaciones específicas y su asociación con diversas características o variables predictivas del contexto como factores ambientales y otros parámetros geológicos y físicos. Una vez entrenado el modelo, este se convierte en una herramienta predictiva capaz de estimar la concentración de radón en nuevos puntos basándose en estas variables. El modelo de Machine Learning ajusta los patrones identificados en la fase de entrenamiento a puntos nuevos. Permitiendo así generar predicciones de concentración en ubicaciones donde no existen datos de monitoreo directo. Para generar las predicciones en puntos nuevos, el modelo de Machine Learning toma las coordenadas de cada punto, junto con las características geológicas y ambientales, y produce una estimación de la concentración de radón en el suelo (Fig.4.22).

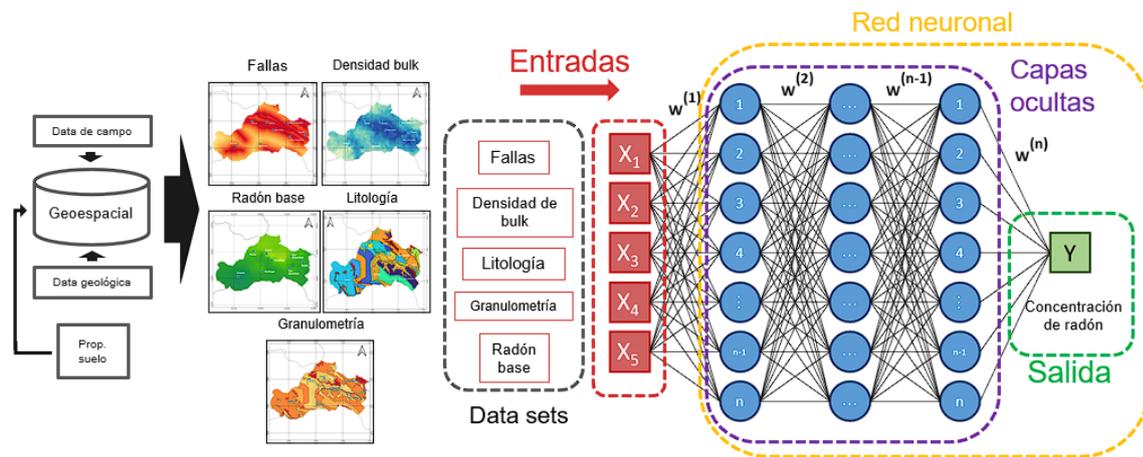


Figura 4.22: Esquema para la obtención de las predicciones

Posterior a esto, con las predicciones de las concentraciones generadas en los nuevos puntos se emplea técnicas de interpolación en GIS para obtener superficies continuas a partir de puntos discretos (Fig.4.23).



Figura 4.23: Esquema para la elaboración del mapa de radón empleando las predicciones

Para la generación de este mapa primeramente se realizó un grillado (Fig.4.24) sobre la provincia del Cusco. Para este grillado se optó por tomar cuadrantes de 1KM x 1KM distribuido uniformemente a lo largo del área de estudio que servirá como base para la estratificación del muestreo y la posterior interpolación de los resultados.

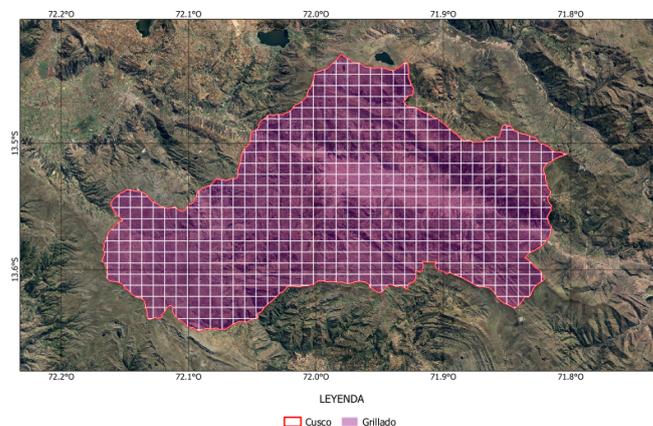


Figura 4.24: Provincia del Cusco con grilla de 1Km x 1Km

Basados en el grillado generado se generaran puntos aleatorios por cada cuadrante. Para cada punto, se extraerán datos de las diversas capas temáticas (capa de litoestratigrafía, capa de granulometría, capa de densidad de Bulk, capa de radón promedio, capa de proximidad a fallas) emepleando el plugin de Qgis "Point sampling tool" que extrae atributos de polígonos y valores ráster de múltiples capas en puntos de muestreo definidos. Este requiere una capa de puntos que contenga las ubicaciones de los puntos de muestreo, en nuestro caso seria la capa de puntos aleatorios generados para cada cuadrante. Finalmente el complemento genera una nueva capa de puntos, conservando las ubicaciones originales y añadiendo los atributos derivados de los polígonos y/o capas ráster.

Capítulo 5

Resultados

5.1. Radón en el gas del suelo en la ciudad de Cusco

5.1.1. Contexto geológico

La composición geológica de la región del Cusco es sumamente compleja y diversa, producto de millones de años de procesos geológicos que han dado lugar a una amplia variedad de litologías (Carlotto et al., 2011).

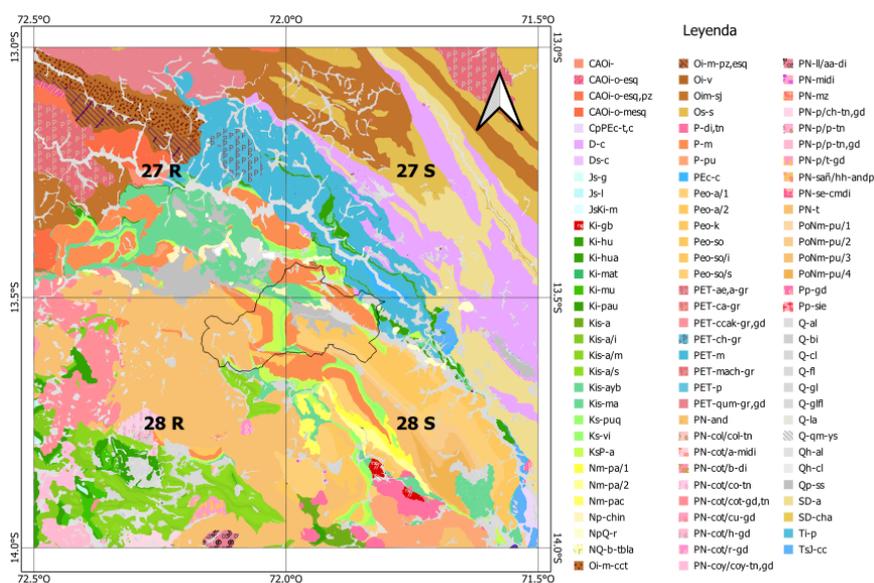


Figura 5.1: Mapa geológico de los cuadrangulos 27r, 27s, 28r, 28s correspondiente a la región Cusco, (Carlotto et al., 2011).

En la región andina, específicamente en la Cordillera Oriental y el Altiplano, se encuentran varias formaciones geológicas que abarcan desde el Siluro-Devónico hasta el Cuaternario (Carlotto et al., 2011). La composición de la zona varía según la unidad, en resumen en la zona de estudio se ha registrado en gran medida rocas ígneas y sedimentarias así como depósitos inconsolidados compuestos de arcillas, arenas y gravas.

5.1.2. Contexto geológico estructural

Por otro lado, la región de Cusco está marcada por una serie de estructuras geológicas que incluyen fallas y pliegues. Estas reflejan los procesos de geodinámica interna relacionados con una importante actividad sísmica cortical debido a los sistemas de fallas activas. Las principales estructuras activas de la zona son las fallas Tambomachay, Pachatusan, Cusco, Alto Qosqo y Qoricocha (Benavente et al., 2013). Estas fallas han controlado la evolución tectónica y sedimentaria pliocuaternaria de la región y particularmente de las cuencas Cusco y Ccatca (Carlotto et al., 2011).

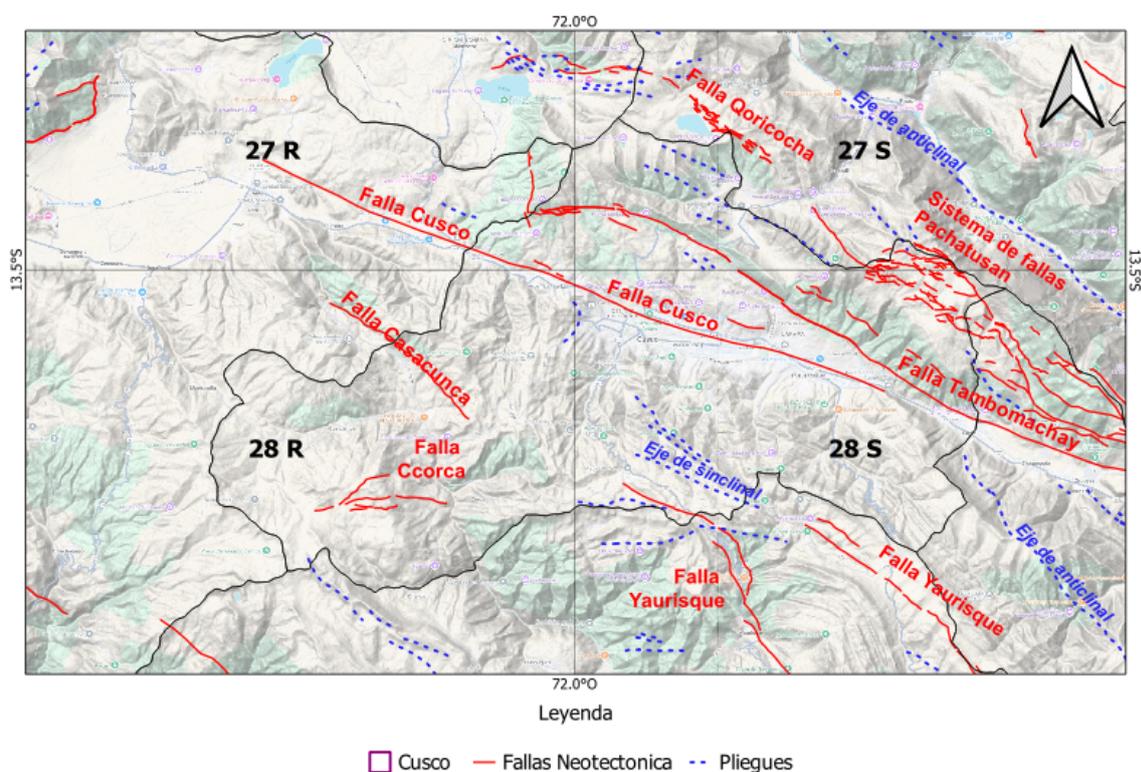


Figura 5.2: Mapa Neotectónico de la provincia del Cusco

5.1.3. Contexto litológico

La litología de Cusco es una compleja mezcla de distintos tipos de rocas (Tabla 5.1). El mapa litológico de la región Cusco (Fig. 5.3) a escala 1:500,000 de INGEMMET proporciona una representación detallada de las diferentes unidades y tipos de rocas presentes. Este mapa muestra la distribución espacial de las litologías brindando información valiosa para la comprensión de la litología de la zona y su concentración de radón en el gas del suelo.

La litología de la zona de estudio (Fig. 5.3) descrita por el INGEMMET se divide en depósitos inconsolidados (I), rocas intrusivas (II), rocas volcánicas (III), rocas volcano-sedimentarias (IV), rocas sedimentarias (V) y rocas metamórficas (VI) Vélchez

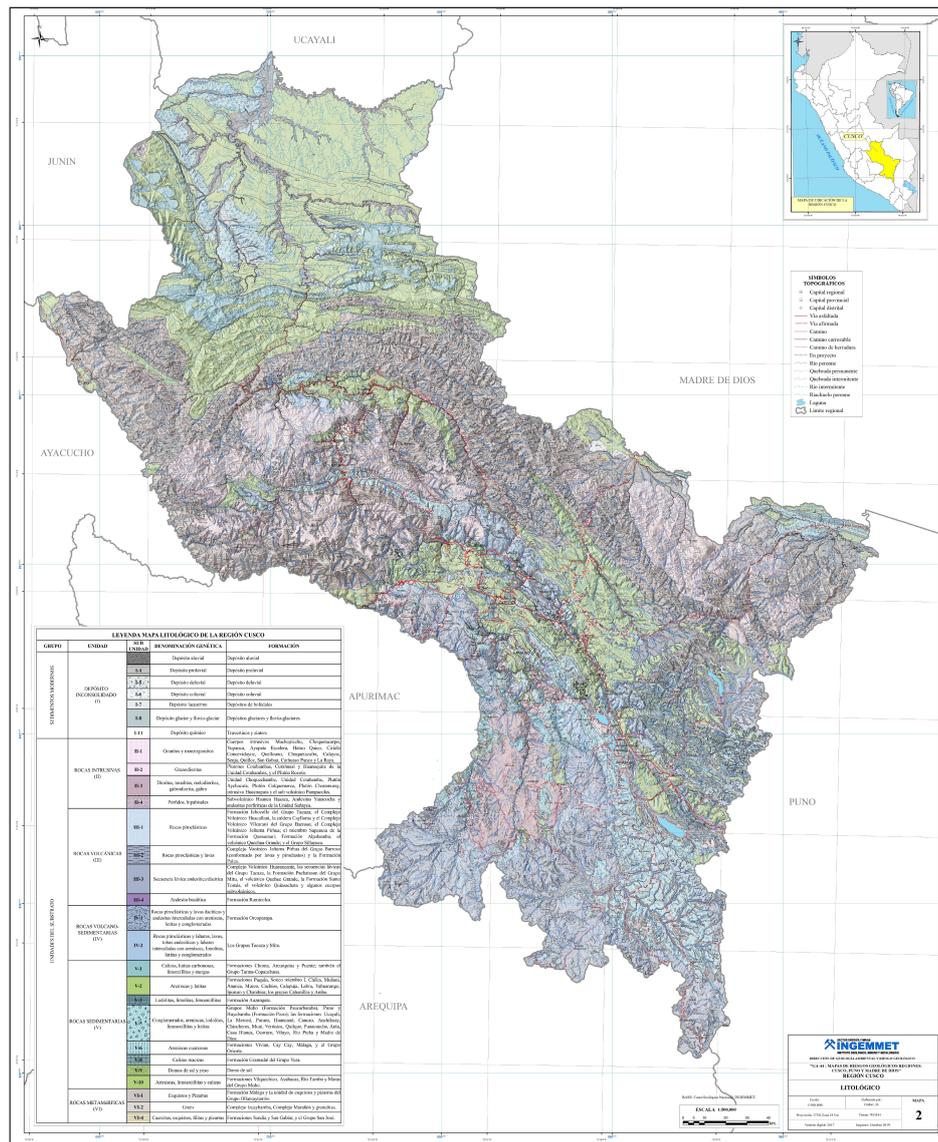


Figura 5.3: Mapa litológico del Departamento del Cusco (Vílchez et al., 2020)

et al. (2020). De esta manera se presenta una visión general de la litología predominante en el valle del Cusco. Debido a la diversidad de esta (Tabla 5.1) se observa la necesidad de evaluar el impacto que puede originar en la distribución de concentración de radón en el gas del suelo.

Litología	Descripción	Edad	Principales manifestaciones
Rocas sedimentarias	Rocas que se forman a partir de la acumulación de sedimentos.	Paleozoico a Cenozoico	Subcuenca de Vilcabamba, Valle del Cusco
Rocas volcánicas	Rocas formadas por la erupción de volcanes.	Cenozoico	Cordillera de los Andes
Rocas metamórficas	Rocas formadas por la transformación de otras rocas mediante calor y presión.	Paleozoico a Cenozoico	Valle del Cusco, subcuenca de Vilcabamba

Tabla 5.1: Litología departamento del Cusco

5.1.4. Niveles de concentración en el valle del Cusco

Se tomaron 465 valores de concentración de radón en suelo. En la (Fig. 5.4) se muestran los niveles y la distribución de los valores de radón obtenidos. El promedio de la concentración de radón en la zona de estudio fue 4545 (Bq/m^3) aproximadamente. El 50% de los datos de concentración de radón en el gas del suelo del valle del Cusco estuvieron por debajo de 3740 (Bq/m^3). Por otro lado, el siguiente 25% de los datos estuvieron entre 3575 y 5993 (Bq/m^3) (relacionado con el tercer cuartil de los datos). Las concentraciones más altas entre el 75% y 99% están entre 5993 y 12479 (Bq/m^3). Por último, se registran valores atípicos (valores en color rojo) con valores superiores a 12479 (Bq/m^3) que representan el 1% de todos los valores.

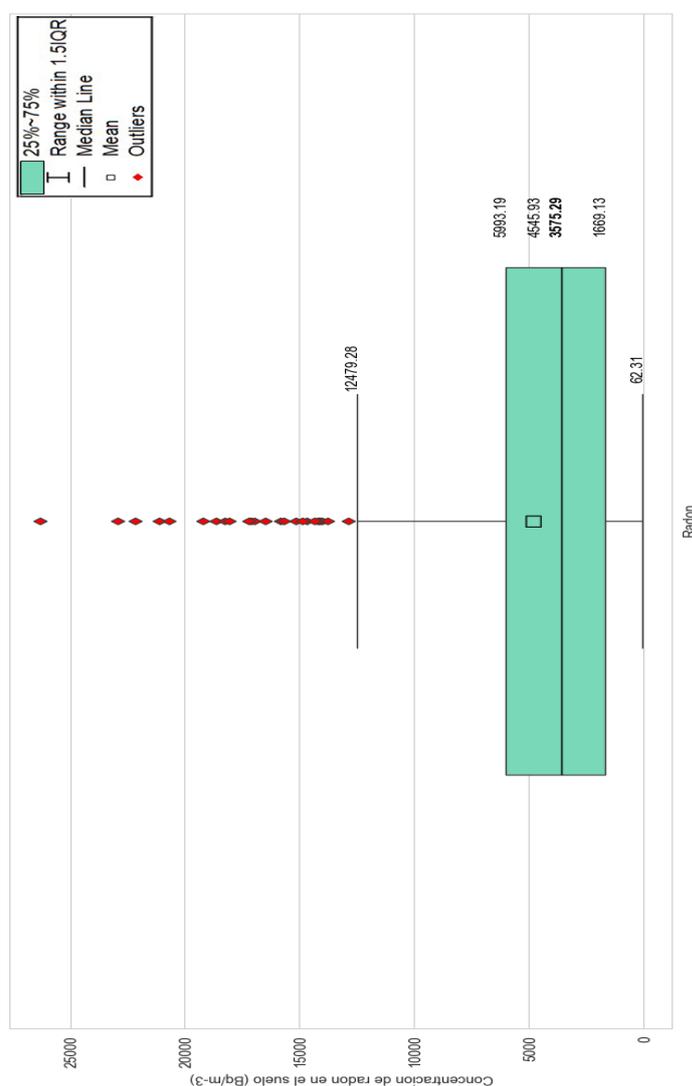


Figura 5.4: Diagrama de cajas de la concentración de radón en el gas del suelo en el Valle del Cusco

Los niveles de fondo usualmente suelen ser determinados por medio del cálculo de los valores umbrales utilizando los valores medios más dos desviaciones estándar después de eliminar los valores anómalos. En este caso se optó por realizar un análisis por medio del gráfico cuantil-cuantil (Q-Q) (Fig. 5.5). Este método proporciona otra manera para distinguir diferentes poblaciones entre valores de fondo y anómalos (Fu et al., 2017). El valor umbral resultante fue de aproximadamente 10 kBq/m^3 . La población de fondo se encuentra en el rango aproximado de 0 a 5 kBq/m^3 . Una zona de transición se identifica entre 5 y 10 kBq/m^3 , marcando el cambio hacia valores anómalos. Por encima de 10,000, se observa claramente la población anómala posiblemente asociada a características geológicas específicas como zonas de fractura o cambios litológicos significativos.

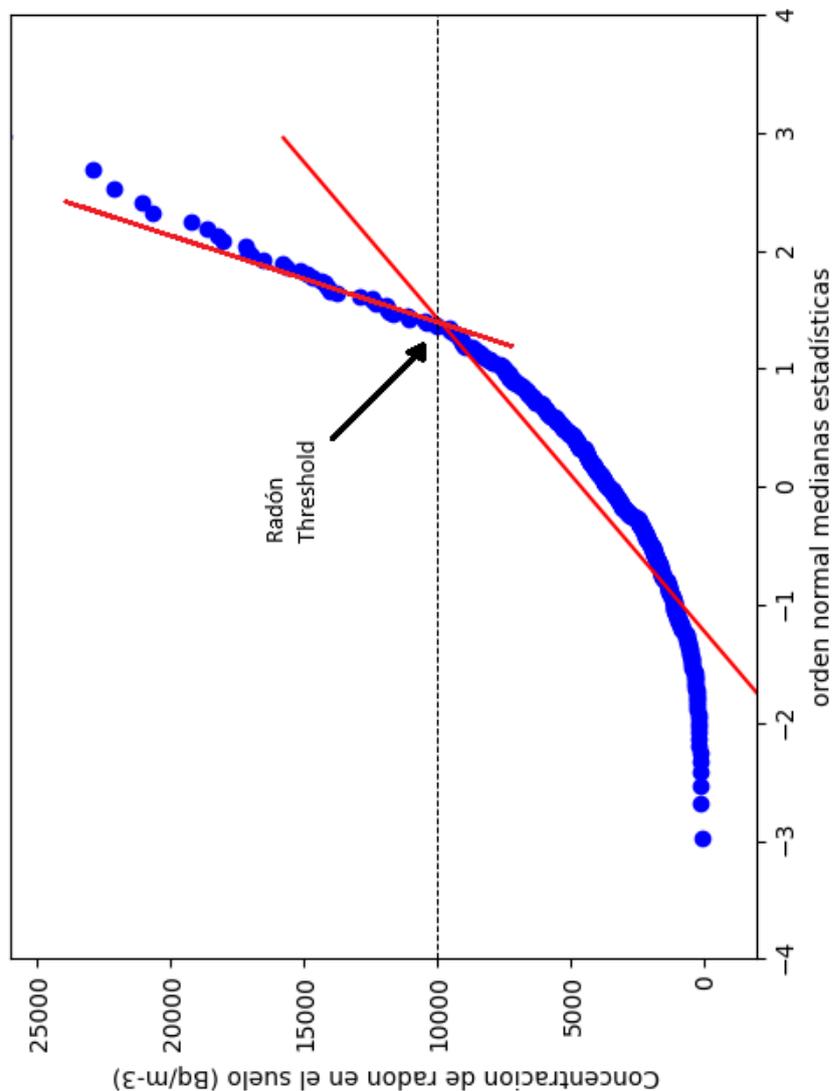


Figura 5.5: Gráfico de cuantil-cuantil de las concentraciones de radón en suelo del valle del Cusco

5.2. Mapas insumo generados

5.2.1. Mapa de Unidades Litoestratigráficas

La Figura 5.6 es el mapa de unidades litoestratigráficas generado a partir de información sobre el cuadrángulo Cusco a escala 1:50000 (Carlotto et al., 2011). El mapa proporciona una representación visual de la distribución y relación espacial de las diversas capas geológicas presentes en la región. Se identificaron 36 unidades que se encuentran clasificadas según el código (CODI) que aparece en la tabla 5.2

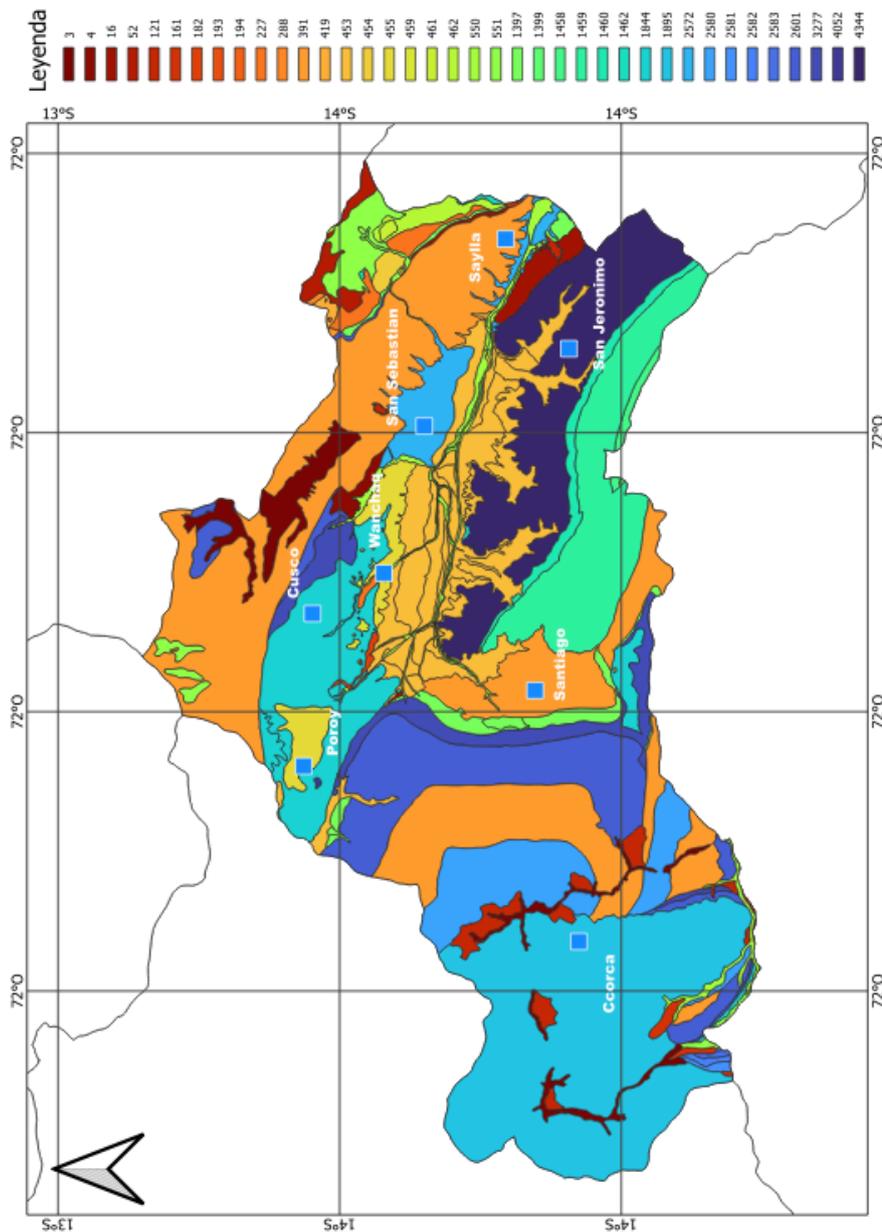


Figura 5.6: Mapa Litoestratigráfico de la provincia de Cusco

CODI	NAME	UNIDAD	DESCRIP
3	Qh-al	Depósito aluvial	Acumulación de grava, arena, limo y arcilla con clastos subangulosos a angulosos de diferente composición.
4	Qh-cl	Depósito coluvial	Constituida por bloques rocosos heterométricos y homogéneos angulosos a sub angulosos.
52	PET-m	Grupo Mitu	Areniscas conglomerados, brechas, lavas y tobas de cenizas
121	Q-cl	Depósito coluvial	Constituida por bloques rocosos heterométricos y homogéneos angulosos a sub angulosos.
161	Q-gl	Depósito glaciar	Depósitos morrénicos, bloques angulosos rellenos con arcillas, limos y arenas.
182	P-di,tn	Diorita, tonalita	Diorita, tonalita
193		Pueblo	Ciudad / Pueblo
194		Laguna	Laguna
227	Ks-puq	Formación Puquín	Lutitas rojas y moradas con carofitas, yesos. Lutitas Negras marinas muy poco profundas y areniscas
288	PEc-c	Grupo Copacabana	Calizas bioclásticas, areniscas y lutitas negras
391	P-m	Formación Muñani	Arenisca beige friable y limoarcillitas rojizas
419	D-c	Grupo Cabanillas	Esquistos cuarzosos feldespáticos, filitas y pizarras, limoarcillitas cremas, filitas, areniscas finas.
453	Qp-ss	Formación San Sebastián	Arcillas, diatomitas y turba.
454	NpQ-r	Formación Rumicolca	Shoshonitas, andesitas
455	Np-chin	Formación Chincheros	Arenas, gravas
459	Kis-ayb	Formación Ayabacas	Calizas masivas grises y marrones, limoarcillitas rojas.
461	Ki-pau	Formación Paucarbamba	Calizas gris oscuras, margas y lutitas paradas en capas delgadas.
462	Ki-hua	Formación Huancané	Areniscas cuarzosas blanquecinas en capas gruesas a medianas tabulares de grano medio a fino, con estructuras sedimentarias de estratificación sesgada y laminación interna paralela.
550	Q-fl	Depósito fluvial	Acumulaciones de gravas, arenas, limos y arcillas transportadas en el cauce de río y quebradas.
551	Q-glfl	Depósito glaciar, fluvial	Gravas, arenas en matriz limoarenosas. Arenas y materiales residuales no consolidados.
1399	PoNm-pu/2	Formación Punacancha - Miembro 2	Areniscas, conglomerados con clastos volcánicos.
1458	PoNm-pu/1	Formación Punacancha - Miembro 1	Limolitas rojas, areniscas y un nivel volcánico.
1459	Peo-so/s	Formación Soncco - Miembro superior	Areniscas y conglomerados polimícticos
1460	Peo-so/i	Formación Soncco - Miembro inferior	Conglomerados polimícticos
1462	Ti-p	Formación Pachatusán	Riolitas de color rojo violáceo, coladas, aglomerados y brechas volcánicas de basalto.
1844	Kis-ma	Formación Maras	Arcillas, lutitas, y algunos estratos de calizas y yeso
1895	PN-t	Grupo Tacaza	Flujos andesíticos, areniscas y limolitas gris violáceas con conglomerados.
2572	Q-al	Depósito aluvial	Acumulación de grava, arena, limo y arcillas con clastos subangulosos a angulosos de diferente composición.
2580	Peo-so	Formación Soncco	Pelitas y limolitas rojas y pardas, areniscas de grano medio con clastos blandos de pelitas rojas, conglomerados
2581	Kis-a/s	Formación Arcurquina - Miembro superior	Calizas de color gris en estratos gruesos
2582	Kis-a/m	Formación Arcurquina - Miembro medio	Calizas de color gris en estratos delgados, bien estratificadas.
2583	Kis-a/i	Formación Arcurquina - Miembro inferior	Calizas gris oscuras, calizas arcillosas.
2601	KsP-a	Formación Auzangate	Arenisca cuarzoza con lodilitas rojizas laminadas en capas delgadas
3277	Ks-vi	Formación Vilquechico	Areniscas cuarzosas blancas en estratos gruesos, areniscas rojizas con areniscas cuarzosas.
4052	PN-se-cmdi	Unidad Senja - cuarzomonzodiorita	Cuarzomonzodiorita
4344	Peo-k	Formación Kayra	Areniscas feldespáticas, lutitas rojas, conglomerados y bancos arenosos

Tabla 5.2: Tabla de unidades geológicas de la provincia del Cusco; (Carlotto et al., 2011)

5.2.2. Mapa de Granulometría

La Figura 5.7 representa el mapa de granulometria para la provincia del Cusco. Es una representación gráfica que muestra la distribución de los tamaños de partículas. La tabla 5.3 muestra el tamaño de grano para cada unidad. Se observa una heterogeneidad significativa en la distribución de tamaños de partículas en el área de estudio. Los valores se encuentran en el rango de 0 a 1024 mm, siendo las capas rojas del Paleogeno las que poseen el menor tamaño de grano y las rocas volcanicas del Neogeno el mayor tamaño de grano. La granulometrias predominante de la zona es de 0.5mm.

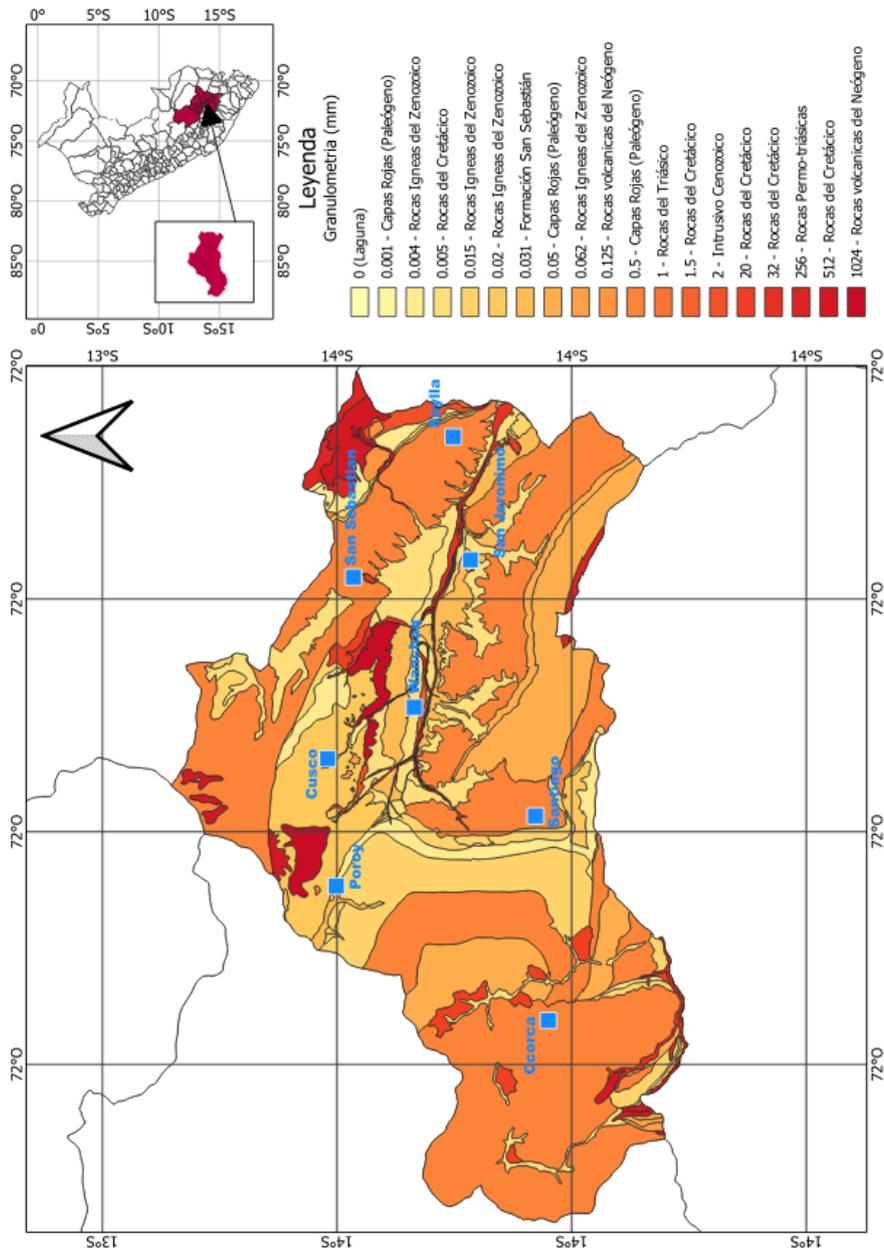


Figura 5.7: Mapa Granulometrico Cusco

UNIDAD	NAME	Granulometria
Laguna		0
Ciudad		0
Grupo Copacabana	PEc-c	0.001
Grupo Cabanillas	D-c	0.004
Formación Puquin	Ks-puq	0.004
Formación Vilquechico	Ks-vi	0.004
Depósito aluvial	Q-al	0.005
Depósito aluvial	Qh-al	0.005
Formación Auzangate	KsP-a	0.015
Formación Quilque	Pp-qu	0.015
Formación Maras	Kis-ma	0.02
Formación Chilca	Pp-chil	0.031
Formación San Sebastian	Qp-ss	0.031
Formación Soncco	Peo-so	0.05
Formación Soncco - Miembro inferior	Peo-so/i	0.05
Formación Soncco - Miembro superior	Peo-so/s	0.05
Formación Arcurquina - Miembro inferior	Kis-a/i	0.062
Formación Arcurquina - Miembro medio	Kis-a/m	0.062
Formación Arcurquina - Miembro superior	Kis-a/s	0.062
Formación Ayabacas	Kis-ayb	0.062
Formación Paucarbamba	Ki-pau	0.125
Formación Rumicolca	NpQ-r	0.125
Formación Kayra	Peo-k	0.5
Formación Muñani	P-m	0.5
Unidad Senja - cuarzomonzodiorita	PN-se-cmdi	0.5
Grupo Tacaza	PN-t	0.5
Formación Pachatusan	Ti-p	1
Depósito biogenico	Q-bi	1.5
Formación Huancane	Ki-hua	2
Diorita, tonalita	P-di,tn	2
Depósito coluvial	Q-cl	20
Depósito fluvial	Q-fl	32
Grupo Mitu	PET-m	256
Formación Punacancha - Miembro 1	PoNm-pu/1	256
Formación Punacancha - Miembro 2	PoNm-pu/2	256
Depósito glaciario	Q-gl	512
Depósito glaciario, fluvial	Q-glfl	512
Formación Chincheros	Np-chin	1024

Tabla 5.3: Tabla de clasificación granulométrica de la provincia de Cusco

5.2.3. Mapa de Radón promedio

El mapa de radón promedio esperado (Fig. 5.8), se obtuvo a partir del análisis estadístico de 465 muestras in situ basado en el IQR. Se ha determinado que el 50% de las muestras presentan concentraciones de radón por debajo de 4032.5 (Bq/m^3). La interpolación de estos datos ha permitido generar un mapa continuo que muestra una distribución espacial de las concentraciones. Se observa una sutil asociación entre las concentraciones de radón y la litología de la zona. Las áreas con predominancia de granitos y rocas metamórficas presentan concentraciones de radón significativamente más altas en comparación con las zonas con rocas sedimentarias.

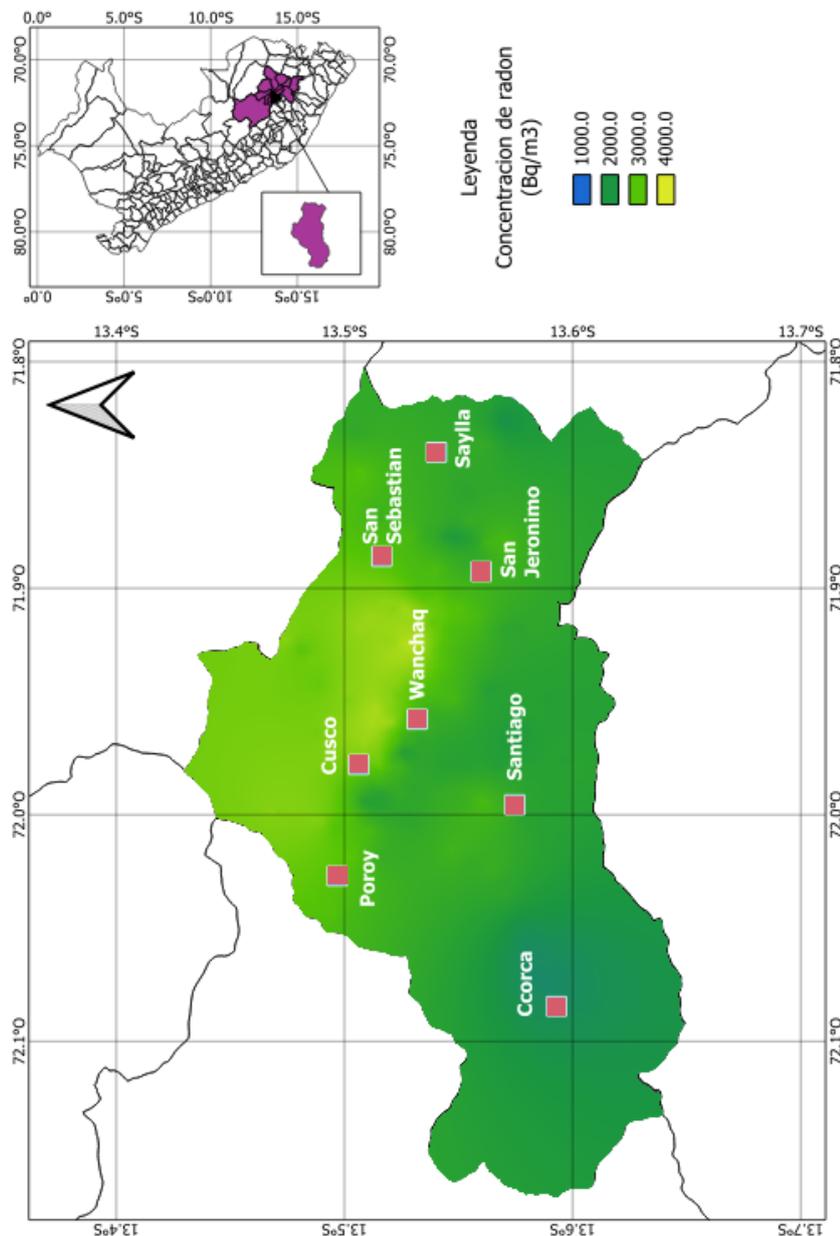


Figura 5.8: Mapa radón base Cusco

5.2.4. Mapa de Densidad de Bulk

El mapa de densidad de Bulk (Fig. 5.9), se obtuvo mediante la recopilación y procesamiento de datos del sistema de cartografía digital del suelo SoilGrids. Los valores de densidad de Bulk obtenidos para la zona de estudio varían entre 119 y 145 cg/cm^3 . La zona de cuenca presenta los valores mas elevados del área, esto muestra una clara relación entre la densidad de Bulk y la topografía del terreno. Los valores más altos se concentran en las zonas de menor pendiente que pueden estar asociados a procesos de compactación donde predominan los suelos más finos. Por otro lado, en las zonas de mayor pendiente predominan suelos más gruesos y porosos presentando valores más bajos.

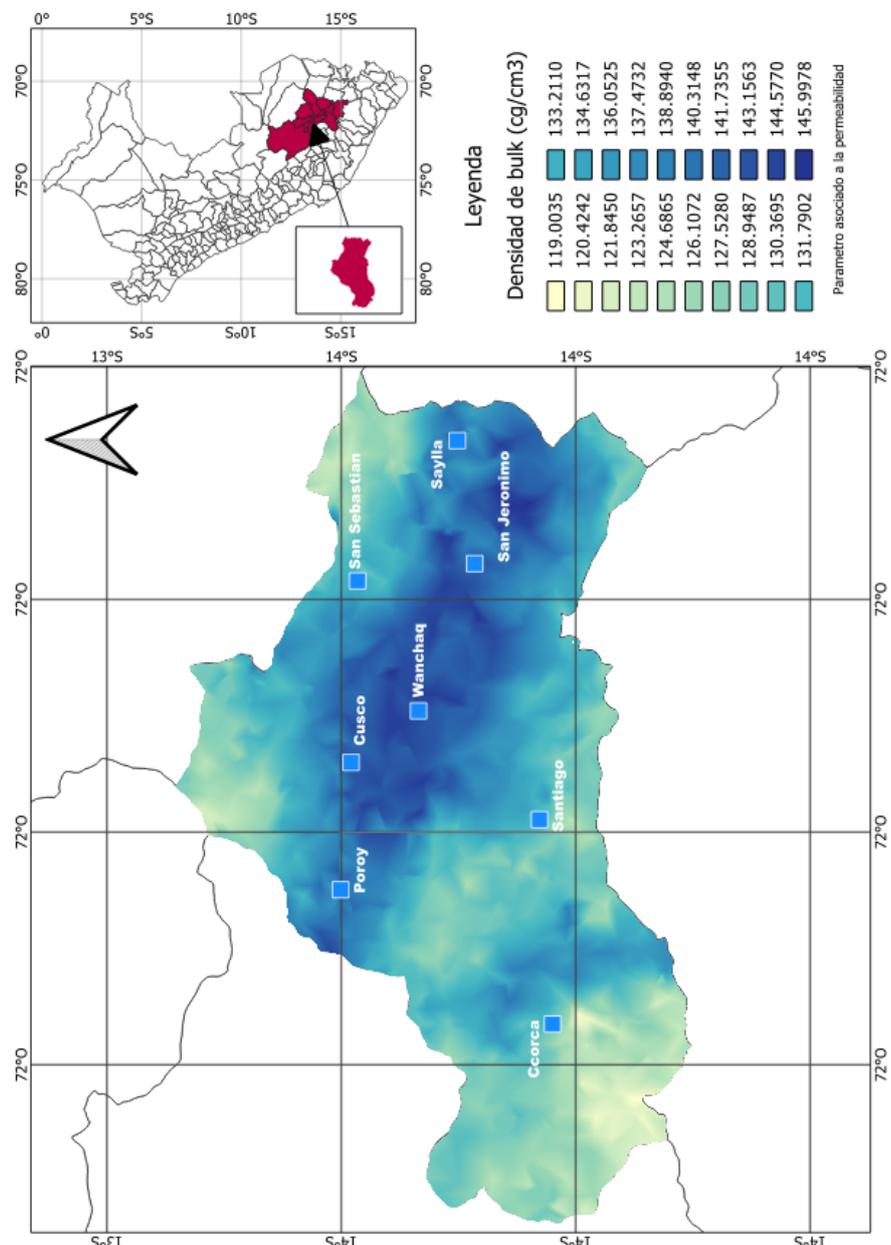


Figura 5.9: Mapa Densidad de bulk de suelo Cusco

5.2.5. Mapa de Proximidad de fallas

La Figura 5.10 muestra el mapa de proximidad de fallas, elaborado a partir de los estudios neotectónicos del INGEMMET. Representa la distancia a la que se estima la ubicación de una falla geológica donde cada píxel resultante en el mapa de proximidad tiene un valor asociado a la distancia al píxel objetivo más cercano. Los píxeles objetivo son los vectores de falla rasterizados teniendo un valor de 0. Esto ha permitido identificar una compleja red de fallas geológicas activas de tipo normal y de rumbo en la zona de estudio. Los resultados obtenidos muestran una alta densidad de fallas y una notable cercanía entre los diferentes sistemas. Destaca la presencia de fallas que se localizan a distancias muy cortas del área urbana como la falla Cusco y la falla Tambomachay.

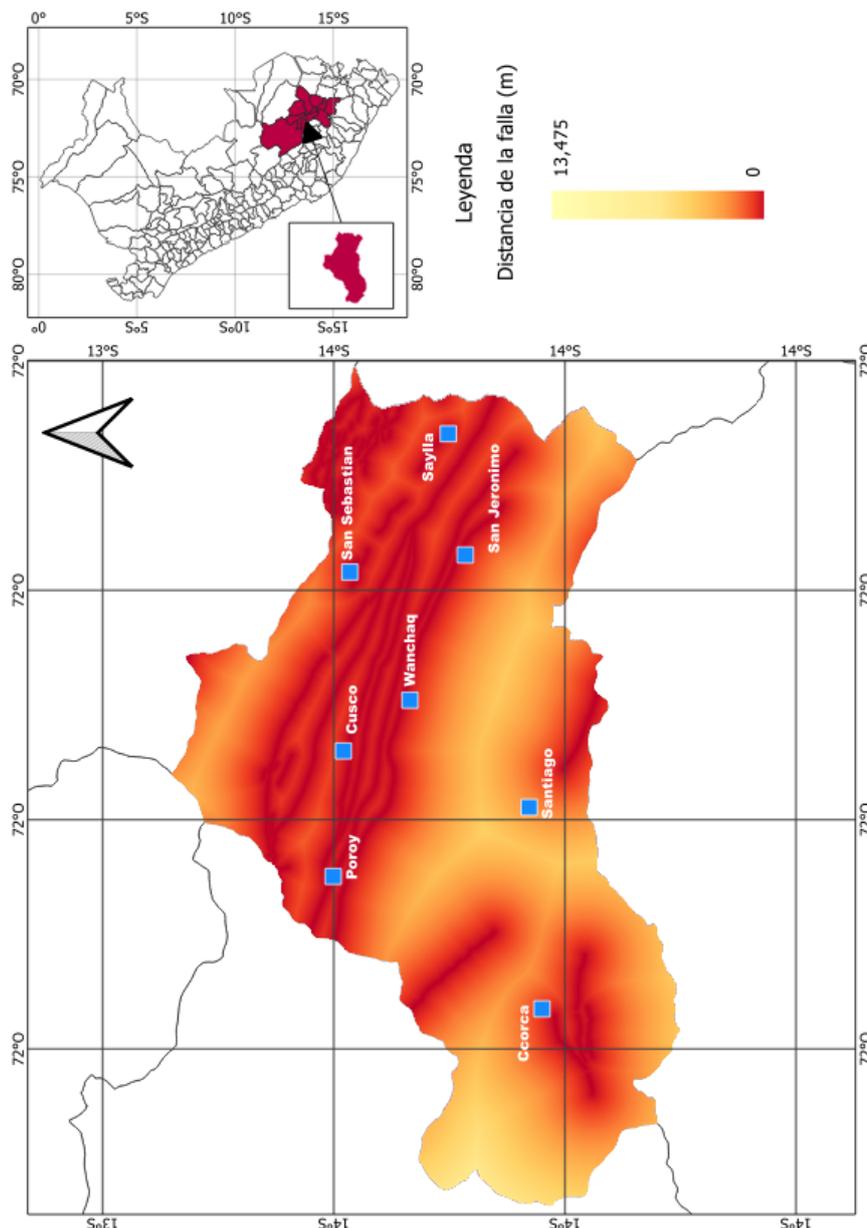


Figura 5.10: Mapa proximidad fallas Cusco

5.3. Modelo de red neuronal artificial

5.3.1. Estructura de la red

Se desarrolló un modelo de red neuronal recurrente bidireccional (Bi-LSTM) para predecir la concentración de radón en el suelo. La arquitectura de la red neuronal, presentada en la Tabla 5.4, consta de tres capas ocultas bidireccionales y una capa de salida densa. El input layer toma una secuencia de datos de entrada y los procesa a través de múltiples capas BiRNN para extraer características y patrones relevantes. Finalmente la capa densa asigna estas características extraídas a un único valor de salida. Cabe resaltar que para este caso se dejó que la ANN autogenerara los pesos y sesgos para cada una de las neuronas en cada capa.

Capa (tipo)	Forma de salida	# parametros
Input_layer (Bidirectional)	(None, 1, 5)	220
Hidden_layer (Bidirectional)	(None, 1, 10)	440
Hidden_layer (Bidirectional)	(None, 1, 50)	7200
Hidden_layer (Bidirectional)	(None, 1, 50)	15200
Output_layer (Dense)	(None, 1, 1)	51

Tabla 5.4: Estructura de la red neuronal

5.3.2. Métricas de evaluación

Para evaluar el desempeño del modelo de red neuronal propuesto, se calcularon métricas de error estándar. Los resultados obtenidos se presentan en la Tabla 5.5, se obtuvo un valor de 0.0933 para el Error Absoluto Medio (MAE) y un valor de 0.1810 para la Raíz del Error Cuadrático Medio (RMSE). Los valores más bajos de MAE indican una buena precisión de la predicción (Duong et al., 2021). Por otro lado, cuanto menor sea el valor de RMSE, más cerca estarán los valores de predicción de los valores de observación (Feng et al., 2022). Así mismo, los mejores modelos tienen valores de RMSE más bajos, mientras que los valores más altos indican incoherencia entre las predicciones de los modelos (Rezaie et al., 2023).

Estos resultados indican que el modelo presenta un buen desempeño en la predicción de la concentración de radón, dado que los valores de MAE y RMSE son relativamente bajos.

Métrica	Valor
Error Absoluto Medio (MAE)	0.0933
Raíz del Error Cuadrático Medio (RMSE)	0.1810

Tabla 5.5: Resultados de las métricas de evaluación para la RNA

5.3.3. Predicciones y valores observados experimentalmente

Para evaluar la precisión del modelo en la predicción de la concentración de radón en suelo, se realizó una comparación entre los valores predichos por el modelo, los cuales posteriormente fueron interpolados y los valores medidos experimentalmente en campo por el INGEMMET. Los resultados de esta comparación se representan en un mapa, superponiendo el mapa generado por machine learning basado en los valores predichos por el modelo y los puntos de medición in situ (Fig. 5.11).

Las Figuras 5.11a, 5.11b, 5.11c, 5.11d y 5.11e constituyen una forma de validación visual de la concordancia entre los valores de concentración de radón predichos por el modelo de aprendizaje automático y los datos experimentales obtenidos en campo. Al examinar detalladamente cada subfigura, se observa consistentemente que los puntos de medición activa (representados por círculos) se superponen de manera muy estrecha con los colores de fondo del mapa de predicción (representados por el gradiente de color). En la Figura 5.11a, que abarca una sección más amplia del emplazamiento urbana del área de estudio, se aprecia cómo los puntos experimentales de baja concentración (azules y verdes) coinciden con las áreas predichas de menor concentración, mientras que los puntos de alta concentración (naranjas y rojos) se sitúan precisamente sobre las zonas del mapa modelado con valores elevados de radón. De igual manera el resto de figuras magnifican esta correspondencia, revelando que incluso en escalas más finas, la transición de colores en los puntos medidos refleja con gran precisión la distribución espacial de la concentración predicha por el modelo.

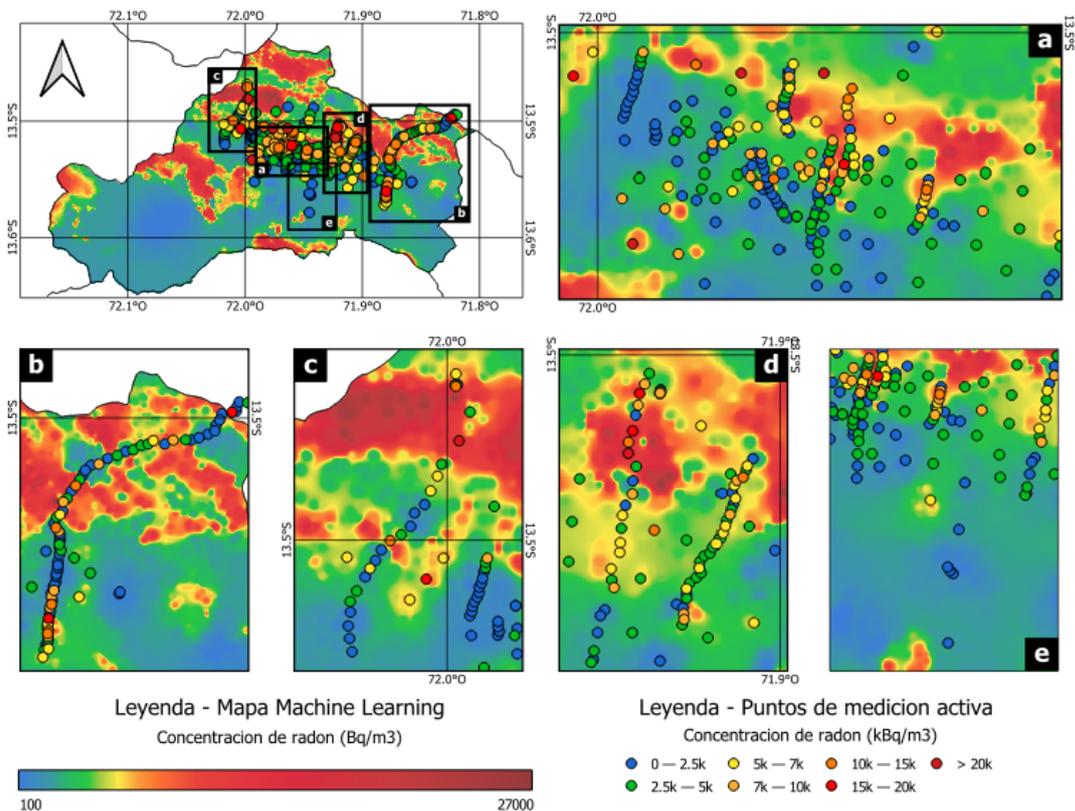


Figura 5.11: Mapa generado por machine learning y valores experimentales

5.4. Mapa de radón generado con Machine Learning

Basado en el grillado se generaron puntos en cada cuadrante (Fig. 5.12). En total se generaron 6365 puntos sobre toda la superficie. Con estos puntos se procederá a realizar el cálculo de la concentración de radón en suelo empleando el modelo previamente entrenado. Posteriormente se realizó una interpolación por el método de IDW y se obtuvo el mapa de concentración de radón en suelo de la provincia de Cusco (Fig. 5.17).

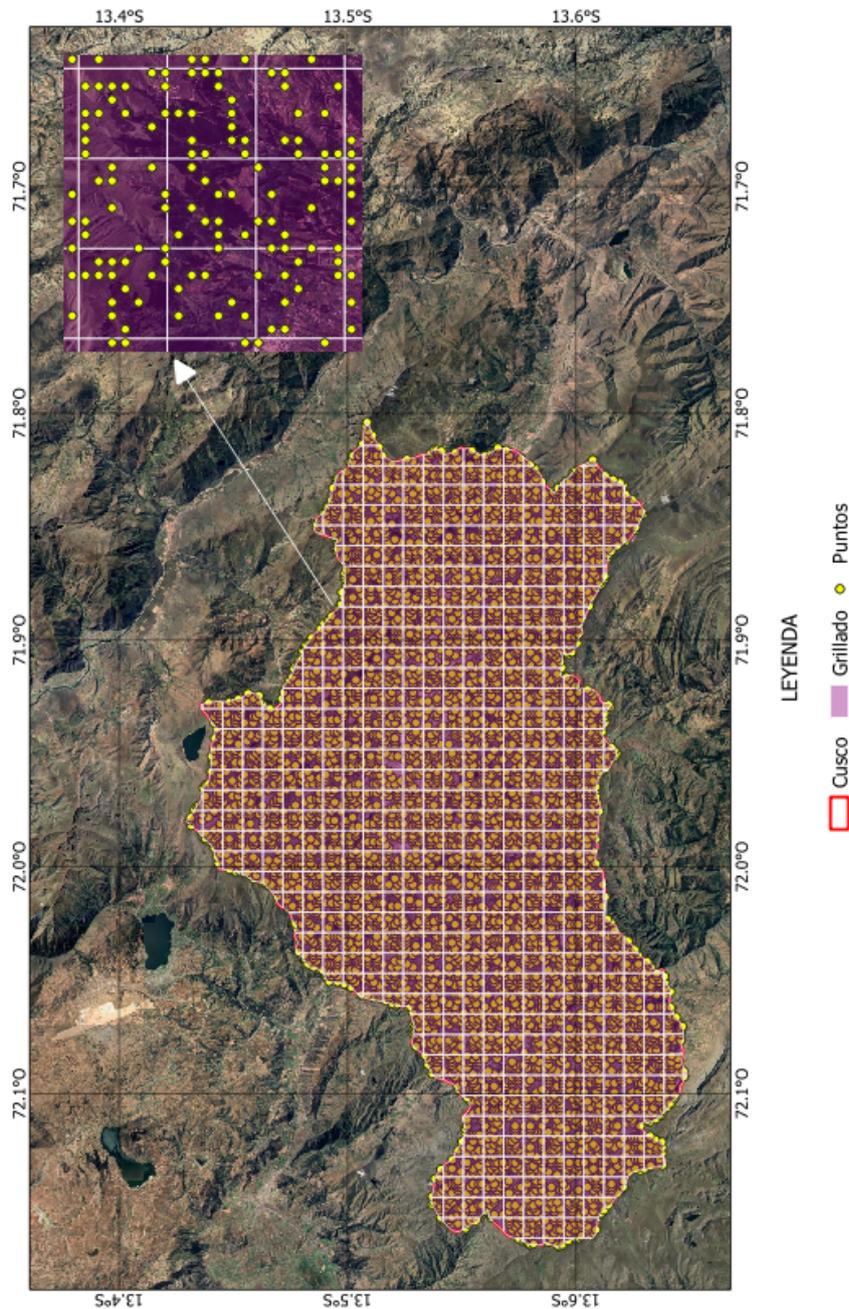


Figura 5.12: Esquema de grillado de la provincia del Cusco

Los niveles obtenidos en la zona de estudio son variables. De esta manera podemos observar la Fig. 5.13a correspondiente a la zona de Ccorca donde predomina concentraciones relativamente bajas en especial en la zona urbana de este distrito, aunque se llegan a observar ciertas áreas con concentraciones mas elevadas. Por otro lado, la Fig. 5.13b corresponde a la zona de Poroy que, a diferencia de Ccorca, se evidencia mayor cantidad de áreas con concentraciones elevadas. Cabe resaltar que estas se ubican en la parte superior del distrito alejadas del área urbana del mismo.

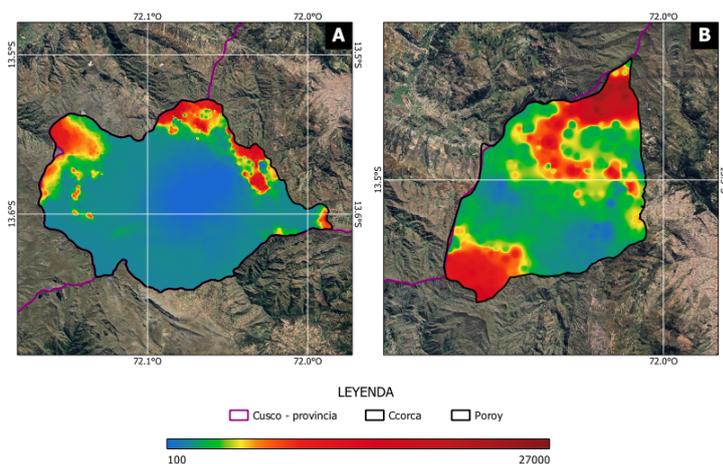


Figura 5.13: a) Mapa de concentración de radón distrito de Ccorca, b) Mapa de concentración de radón distrito de Poroy

La Fig. 5.14a muestra el distrito del Cusco, es la zona que describe la mayor cantidad de áreas con niveles elevados de concentración de radón. Teniendo los valores mas elevados en la zona de Tambomachay y algunos segmentos en Saqsaywaman. El área netamente urbana tiene niveles mas reducidos a excepción de alto qosqo. La Fig. 5.14b correspondiente al distrito de Wanchaq es la que presenta la menor cantidad de valores elevados en la concentración de radón.

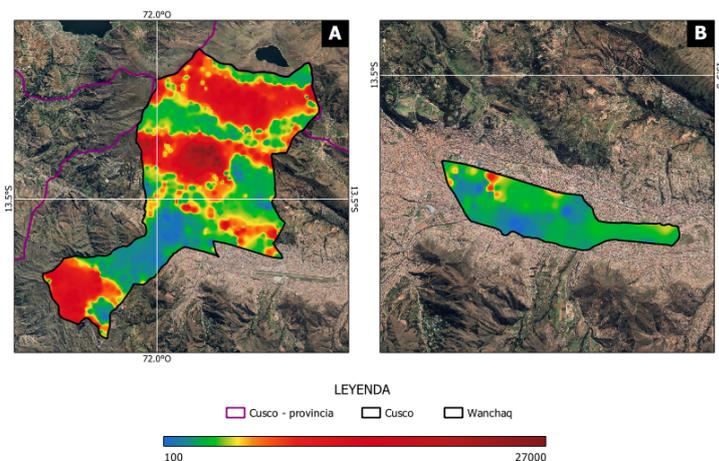


Figura 5.14: a) Mapa de concentración de radón distrito de Cusco, b) Mapa de concentración de radón distrito de Wanchaq

La Fig. 5.15a, distrito de Santiago, tiene una zona urbana con niveles predominantemente bajos a moderados aunque si posee niveles elevados en otras áreas. El distrito de San Sebastián (Fig. 5.15b) muestra dos zonas características. La parte baja del distrito se caracteriza por niveles bajos de concentración. Por otro lado la parte superior del distrito posee zonas con niveles de concentración mas elevados en áreas cercanas a la falla Tambomachay. Una mayor porción de la zona urbana de este distrito es la que se encuentra con niveles medios a altos.

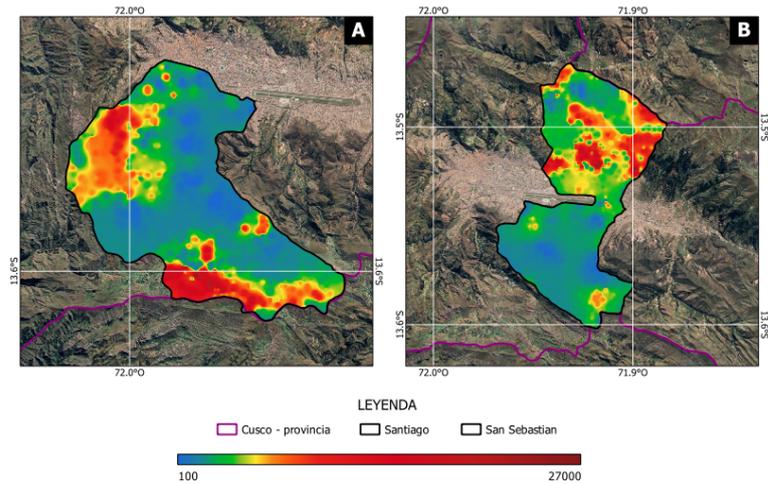


Figura 5.15: a) Mapa de concentración de radón distrito de Santiago, b) Mapa de concentración de radón distrito de San Sebastián

El distrito de San Jerónimo (Fig. 5.16a) de manera análoga al distrito de San Sebastián tiene el mismo comportamiento una zona con mayor concentración que otra. En este caso, el área urbana de San Sebastián tiene niveles bajos a moderados y presenta los niveles mas elevados por la zona del sistema de fallas Pachatusan. Finalmente el distrito de Saylla (Fig. 5.16b) con un área urbana con niveles de concentración bajos.

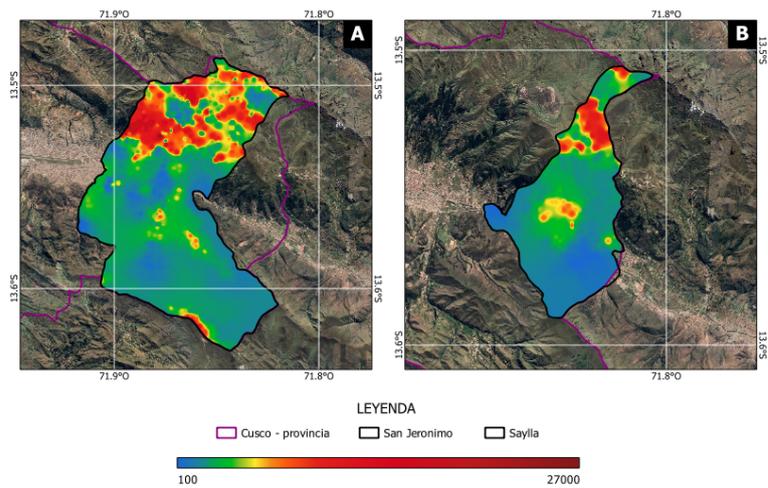


Figura 5.16: a) Mapa de concentración de radón distrito de San Jerónimo, b) Mapa de concentración de radón distrito de Saylla

5.4. Mapa de radón generado con Machine Learning

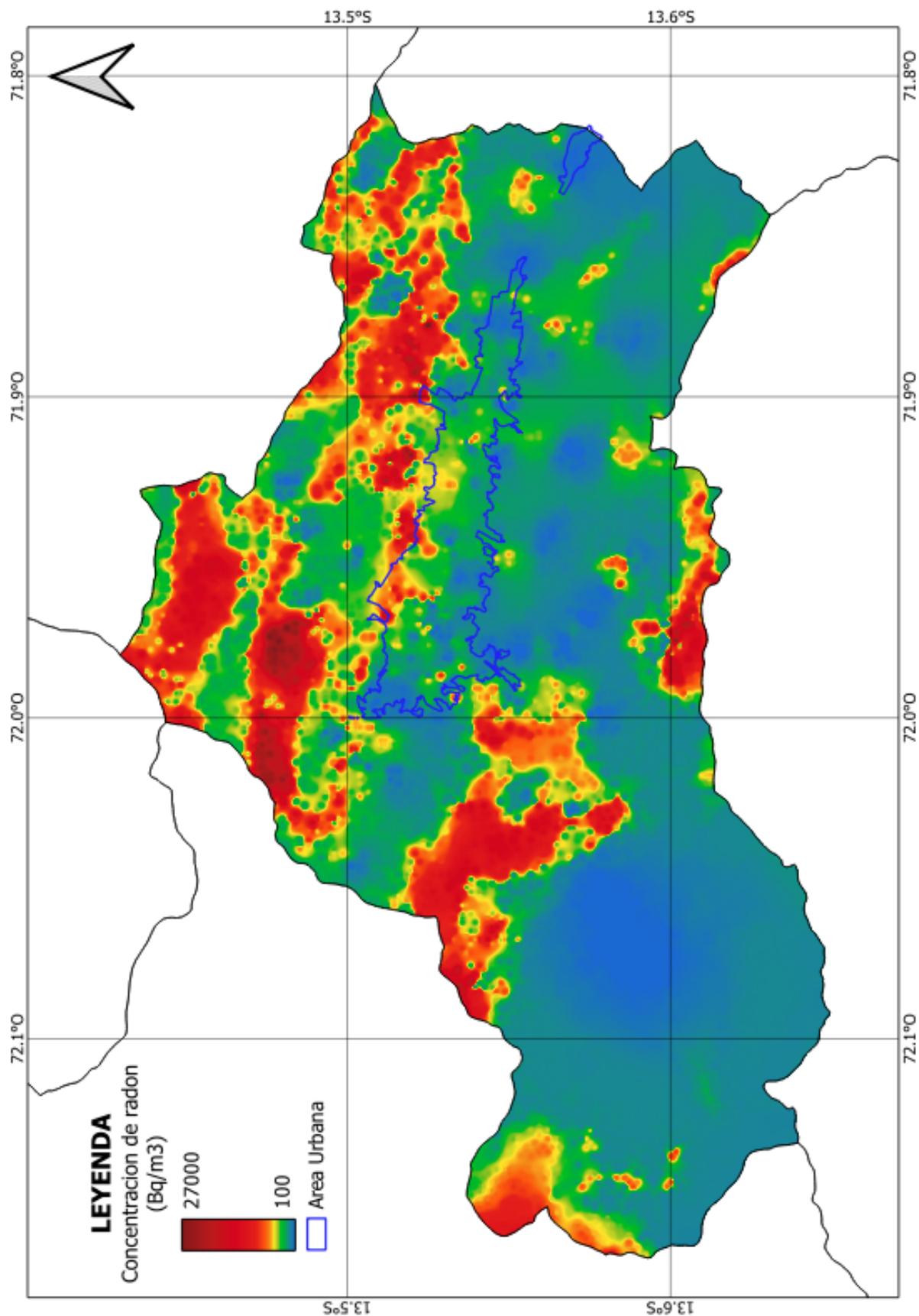


Figura 5.17: Mapa de concentración de radón en suelo de la provincia del Cusco por medio de Machine Learning

Capítulo 6

Discusiones y Conclusiones

6.1. Discusiones

6.1.1. Gas radón como herramienta para la detección de fallas geológicas

Los resultados obtenidos en este estudio revelan una interesante relación entre la concentración de radón en el suelo y la presencia de fallas geológicas. Como se indicó previamente las fallas geológicas son vías de migración que favorecen la migración de geogases del suelo (Chen et al., 2018),

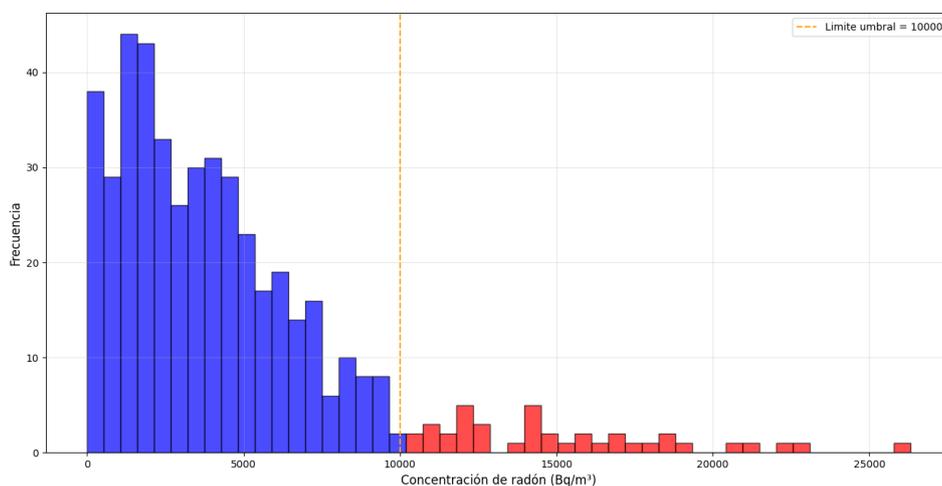


Figura 6.1: Histograma de la distribución de concentración de radón

Basado en los resultados obtenidos por mediciones activas y ML en los distritos de la provincia del Cusco. Se observó una variabilidad significativa en diferentes zonas, los valores de concentración de radón en suelo se encuentran en un rango de 0.1 – 27 (kBq/m^3), se pudo identificar anomalías a partir de 10 (kBq/m^3) (Fig. 6.1)

Al comparar los valores de concentración de radón obtenidos mediante mediciones

directas en campo y los valores predichos por el modelo de aprendizaje automático, se observa una clara correlación espacial entre las zonas de alta concentración de radón y las áreas identificadas o inferidas como zonas de falla por el INGEMMET. Esta correlación puede explicarse por diversos mecanismos. En primer lugar, las fallas geológicas representan zonas de debilidad en la corteza terrestre debido a su mayor grado de permeabilidad y porosidad en relación con las rocas circundantes. En segundo lugar, las zonas de fracturamiento pueden actuar como conductos para la circulación de fluidos subterráneos, lo que puede favorecer la liberación de radón hacia la atmósfera del suelo (Cothern and Smith, 1987).

Las mediciones activas forman parte de un primer muestreo exploratorio en la zona. En la Figura 6.2 se presentan los puntos de muestreo de gas radón en el suelo identificados como anomalías y las fallas geológicas o segmentos de fallas identificados por el INGEMMET. Se observa como en proximidades a la falla Cusco existen múltiples puntos con valores anómalos (puntos rojos Figura 6.2) así como otros asociados a anomalías (puntos naranjas Figura 6.2). De igual manera, también se observan estas anomalías en segmentos de la falla Tambomachay y algunos próximos al sistema de fallas Pachatusan.

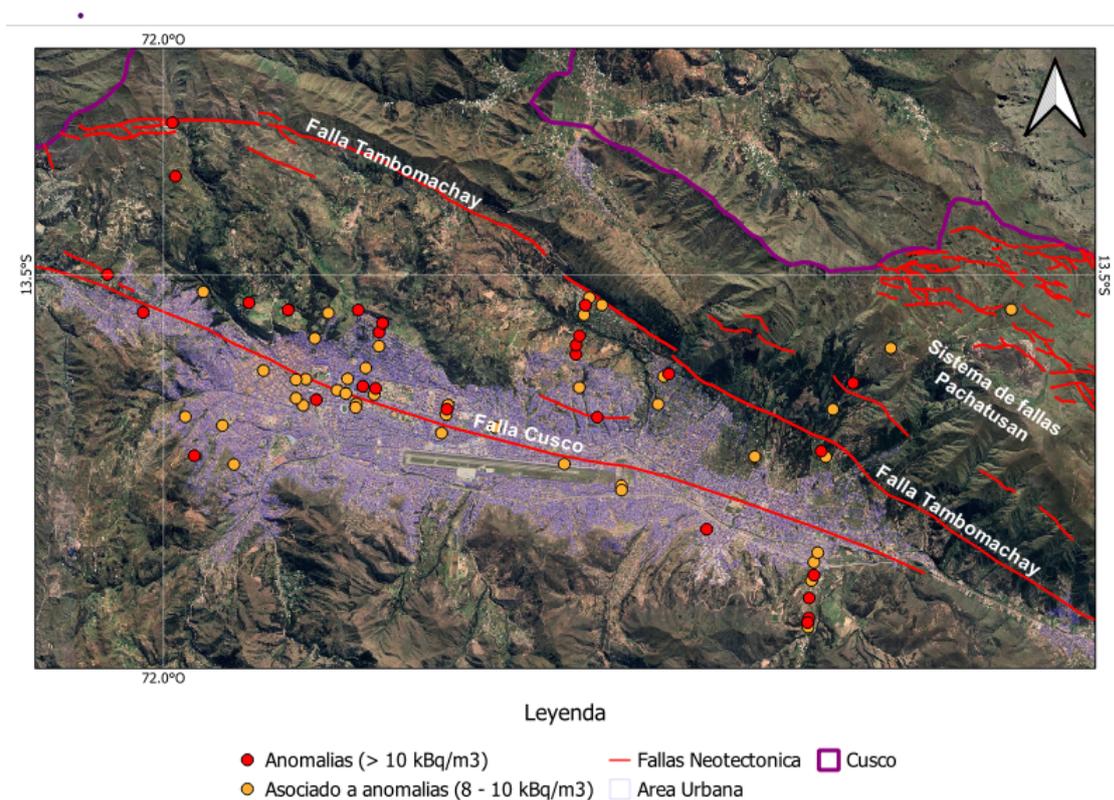


Figura 6.2: Mapa de fallas y puntos de muestreo mediciones activas

Si bien los resultados obtenidos muestran una relación entre las fallas Cusco, Tambomachay y anomalías de concentración de radón. Es importante observar que existe presencia de otras zonas con elevadas concentraciones de radón que no coinciden directamente con el trazado existente de estas fallas. Esto sugiere la posible existencia

de otros segmentos de falla o incluso de fallas completamente nuevas que no habían sido identificadas previamente.

Por otro lado, la Figura 6.3 presenta las áreas con niveles de concentración de radón en suelo anómalos obtenido por medio de Machine Learning y las fallas geológicas o segmentos de fallas identificados por el INGEMMET. En este caso se puede apreciar una mayor cantidad de áreas con posibles niveles elevados de radón en proximidades de fallas identificadas. En comparación con las mediciones in situ en este caso se observa valores anómalos en proximidades de la falla Casacunca y la falla Yaurisque. Al igual que en las mediciones in situ se observa áreas anómalas en proximidades de la falla Cusco, la falla Tambomachay y el sistema de falla Pachatusan.

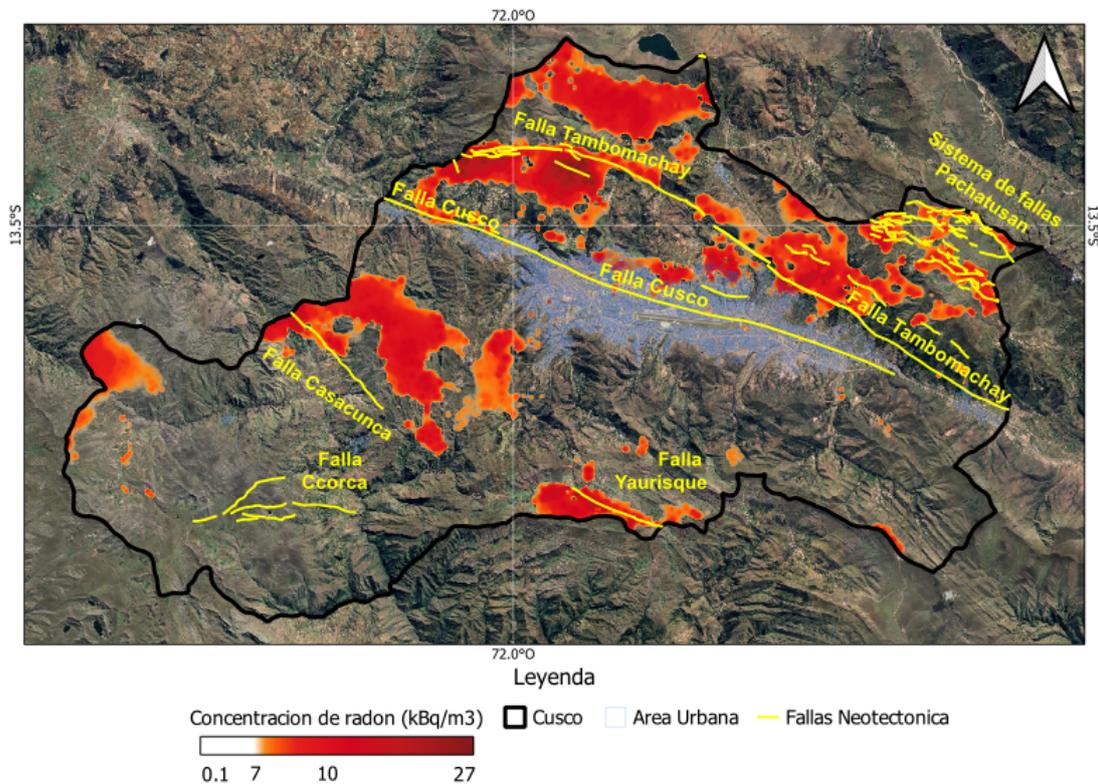


Figura 6.3: Mapa de fallas y áreas con niveles anómalos generados por ML

Basado en los resultados de este estudio, estos sugieren que los niveles de concentración de radón en suelo pueden funcionar como una herramienta complementaria para la detección e identificación de fallas geológicas, especialmente en áreas donde la información geológica disponible es limitada. Al identificar zonas con anomalías en la concentración de radón, es posible inferir la presencia de estructuras geológicas ocultas (Zheng et al., 2012; Davidson et al., 2016; López et al., 2016). Dando la posibilidad de realizar estudios geofísicos más detallados para confirmar su existencia. Por otro lado, en la Gestión de Riesgos de Desastres (GRD) la identificación de fallas geológicas es fundamental para la evaluación de los riesgos geológicos, como sismos y deslizamientos de tierra. Los resultados de este estudio sugieren que los mapas de concentración de radón pueden ser utilizados como una herramienta de fácil implementación para identificar zonas con mayor potencial de radón que a su vez podrían estar relacionadas con

zonas de fracturamiento.

6.1.2. Concentración de gas radón en suelo y en interiores

Ya que la principal fuente de radón en interiores es el suelo, proveniente de la desintegración radiactiva del uranio presente en las rocas y el suelo subyacente. Se realizó una comparación entre los resultados obtenidos en este estudio, donde se empleó algoritmos de machine learning para estimar las concentraciones de radón en el suelo, y los datos de concentración de radón en interiores reportados por Arteaga (2023), donde se empleó el uso de detectores de trazas nucleares expuestos en una cámara de difusión. En este contexto, Arteaga (2023) reportó valores de concentración medios para la zona urbana de Cusco empleando dos tipos de detectores, para CR-39 la concentración media encontrada fue de $59,94 \pm 10,53$ (Bq/m^3) y para LR-115 fue de $55,8 \pm 12,4$ (Bq/m^3). Además de indicar una relación entre la concentración de radón en interiores con ciertos factores estacionales y de materiales de construcción. Este estudio, al emplear un enfoque basado en redes neuronales artificiales para estimar la concentración de radón en el suelo nos permite tener una perspectiva complementaria.

Al comparar nuestros mapas de distribución de radón en el suelo con los resultados de Arteaga (2023), observamos una concordancia general en términos de los niveles de concentración presentes en las zonas. Es necesario resaltar que el estudio de interiores se realizó únicamente en la zona urbana de la ciudad. A pesar de la significativa concordancia que se observa es necesario destacar que el mapa generado por machine learning identifica ciertas zonas con niveles elevados de concentración en la zona urbana, cosa que no se ve reflejado en el estudio de interiores. No obstante, es importante destacar que el mapa representa una estimación de la concentración de radón en el suelo y no reflejan directamente los niveles de radón en interiores. Aunque nuestros resultados corroboran en gran medida los hallazgos de estudios previos, se requiere de una mayor cantidad de datos de medición de radón en interiores para establecer una relación causal más sólida y generar posibles mapas de riesgo más detallados.

6.2. Conclusiones

La implementación de un protocolo estandarizado para las mediciones in situ de las concentraciones de radón en suelo permitió obtener datos confiables y representativos de la zona de estudio. Estas mediciones constituyen la primera base de datos sistemática de concentraciones de radón en suelo para el Valle del Cusco, estableciendo un precedente importante para futuros estudios en la región.

La elaboración de mapas de parámetros geológicos y físicos que influyen en la distribución del radón en el suelo mediante sistemas de información geográfica (GIS) representa una herramienta cartográfica integrada para la comprensión de la dinámica del radón en el contexto geológico local.

El mapa de concentración de radón generado a partir del modelo de redes neuronales artificiales proporciona una visualización espacial de la distribución de radón

en el suelo en el Valle del Cusco. Constituyendo una herramienta para la identificación de zonas de alto riesgo radiológico, planificación urbana e implementación de medidas preventivas en áreas críticas.

La aplicación de un modelo de predicción basado en redes neuronales artificiales ha demostrado ser un método adecuado para estimar los niveles de radón en el gas del suelo proporcionando resultados de una primera idea sobre los niveles estimados de concentración que se pueden encontrar en la zona de estudio. En algunos casos existe la posibilidad de que el modelo no tenga un ajuste correcto. Esto puede estar relacionado con factores antropogénicos no contemplados en el desarrollo del modelo o con un posible mal ajuste por parte del modelo, ya que aún se encuentra en fase de desarrollo.

La metodología desarrollada en este estudio demuestra el potencial de las técnicas de machine learning, específicamente las redes neuronales artificiales, para la predicción y mapeo de variables ambientales complejas como la concentración de radón en suelo. Este enfoque puede ser adaptado y replicado en otras regiones con características geológicas similares. Los resultados obtenidos contribuyen significativamente al conocimiento sobre la distribución del radón en el Valle del Cusco y establecen una base metodológica para futuras investigaciones en el campo de la radioactividad natural y su monitoreo mediante técnicas de inteligencia artificial.

6.2.1. Limitaciones

- La precisión del modelo de redes neuronales artificiales está condicionada por la cantidad de puntos de medición disponibles en el área de estudio.
- Las mediciones in situ de radón pueden verse afectadas por variaciones temporales y estacionales que no fueron consideradas en el modelo actual.
- La resolución espacial de algunos parámetros geológicos utilizados en el SIG podría limitar la precisión del mapeo en ciertas áreas específicas.

6.2.2. Recomendaciones

Considerando las importantes contribuciones de este estudio y las limitaciones identificadas, se sugieren las siguientes recomendaciones para futuras investigaciones y aplicaciones:

1. Ampliación y Densificación de la Base de Datos de Radón:

- Incrementar el número de puntos de medición para mejorar significativamente la precisión y la robustez del modelo de redes neuronales, es crucial expandir la red de muestreo in situ especialmente en áreas poco representadas o con variabilidad geológica compleja.
- Diversificar la distribución espacial de los puntos asegurando una cobertura más homogénea y representativa.

2. Incorporación de Variables Temporales y Estacionales:

- Implementar estudios de monitoreo a largo plazo dada la influencia de las variaciones temporales y estacionales en la concentración de radón (temperatura, humedad del suelo, presión atmosférica, etc.), se recomienda establecer un programa de monitoreo continuo en puntos estratégicos a lo largo de diferentes estaciones del año. Esto permitirá desarrollar modelos que capturen la dinámica temporal del radón y no solo su distribución espacial.

3. Refinamiento y Actualización de Parámetros Geológicos y Físicos en el SIG:

- Explorar la inclusión de otros parámetros que puedan influir en la concentración de radón, como la densidad de fallas geológicas, la presencia de fracturas, la composición mineralógica específica de rocas y suelos, e incluso datos de la capa freática.

4. Investigación y Cuantificación de Factores Antropogénicos:

- Realizar estudios para identificar y cuantificar el impacto de factores humanos (como el uso del suelo, la presencia de edificaciones, la profundidad de cimientos, alteraciones del suelo por obras civiles, etc.) que no fueron contemplados y que podrían explicar algunas de las desviaciones del modelo.
- Integrar estas variables en futuras iteraciones del modelo para mejorar su capacidad predictiva en entornos urbanos y semiurbanos.

5. Optimización y Validación del Modelo de Redes Neuronales:

- Probar diferentes arquitecturas de redes neuronales (p. ej., redes neuronales convolucionales para datos geospaciales, diferentes funciones de activación, optimizadores) y parámetros de entrenamiento para buscar un ajuste aún mejor.

6. Fomento de la Colaboración Interinstitucional:

- Establecer alianzas con agencias de salud, planificación urbana y geológicas locales para integrar los resultados de este estudio en la toma de decisiones políticas y la gestión del riesgo radiológico en el Valle del Cusco.
- Promover la educación y la concienciación pública sobre los riesgos del radón y las medidas preventivas, utilizando el mapa de concentración como una herramienta didáctica.

Bibliografía

- Al-Khateeb, H., Aljarrah, K., Alzoubi, F., Alqadi, M., and Ahmad, A. (2017). The correlation between indoor and in soil radon concentrations in a desert climate. *Radiation Physics and Chemistry*, 130:142–147.
- Arteaga, R. (2023). Evaluación de la concentración de radón en aire en viviendas de cusco metropolitano.
- Baldwin, J. and Grantham, V. (2015). Radiation hormesis: historical and current perspectives. *Journal of Nuclear Medicine Technology*, 43(4):242–246.
- Baskaran, M. (2016a). *Radon: A tracer for geological, geophysical and geochemical studies*, volume 367. Springer.
- Baskaran, M. (2016b). *Radon: A tracer for geological, geophysical and geochemical studies*, volume 367. Springer.
- Benà, E., Ciotoli, G., Ruggiero, L., Coletti, C., Bossew, P., Massironi, M., Mazzoli, C., Mair, V., Morelli, C., Galgaro, A., et al. (2022). Evaluation of tectonically enhanced radon in fault zones by quantification of the radon activity index. *Scientific Reports*, 12(1):21586.
- Benavente, C. L., Delgado Madera, G. F., Taípe Maquerhua, E. L., Audin, L., and Pari Pinto, W. (2013). Neotectónica y peligro sísmico en la región cusco [boletín c 55].
- Bishop, C. M. and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*, volume 4. Springer.
- Blake, G. (1965). Bulk density. *Methods of soil analysis: Part 1 physical and mineralogical properties, including statistics of measurement and sampling*, 9:374–390.
- Brosed, A. and de Física Médica, S. E. (2011). *Fundamentos de física médica: Medida de la radiación*. Sociedad Española de Física Médica.
- Brownlee, J. (2017). *Long short-term memory networks with python: develop sequence prediction models with deep learning*. Machine Learning Mastery.
- Carlotto, V. S., Cárdenas Roque, J. D., and Carlier, G. (2011). Geología del cuadrángulo de cusco, hoja 28-s, escala 1:50,000 - [boletín a 138].
- Cember, H. (2009). *Introduction to health physics*.

- Chen, Z., Li, Y., Liu, Z., Wang, J., Zhou, X., and Du, J. (2018). Radon emission from soil gases in the active fault zones in the capital of china and its environmental effects. *Scientific reports*, 8(1):16772.
- Cinelli, G., Tollefsen, T., Bossew, P., Gruber, V., Bogucarskis, K., De Felice, L., and De Cort, M. (2019). Digital version of the european atlas of natural radiation. *Journal of environmental radioactivity*, 196:240–252.
- Ciotoli, G., Voltaggio, M., Tuccimei, P., Soligo, M., Pasculli, A., Beaubien, S., and Bigi, S. (2017). Geographically weighted regression and geostatistical techniques to construct the geogenic radon potential map of the lazio region: A methodological proposal for the european atlas of natural radiation. *Journal of environmental radioactivity*, 166:355–375.
- Cothorn, C. R. and Smith, J. E. (1987). *Environmental radon*, volume 35. Springer Science & Business Media.
- Davidson, J. R. J., Fairley, J., Nicol, A., Gravley, D., and Ring, U. (2016). The origin of radon anomalies along normal faults in an active rift and geothermal area. *Geosphere*, 12(5):1656–1669.
- De Miguel, E., Barrio-Parra, F., Elío, J., Izquierdo-Díaz, M., García-González, J. E., Mazadiego, L. F., and Medina, R. (2018). Applicability of radon emanometry in lithologically discontinuous sites contaminated by organic chemicals. *Environmental Science and Pollution Research*, 25:20255–20263.
- DOE, United states (1993). *DOE Fundamentals Handbook: Nuclear Physics and Reactor Theory*. US Department of Energy.
- Doss, M. (2018). Are we approaching the end of the linear no-threshold era? *Journal of Nuclear Medicine*, 59(12):1786–1793.
- Duong, V.-H., Ly, H.-B., Trinh, D. H., Nguyen, T. S., and Pham, B. T. (2021). Development of artificial neural network for prediction of radon dispersion released from sinquyen mine, vietnam. *Environmental Pollution*, 282:116973.
- Elío, J., Crowley, Q., Scanlon, R., Hodgson, J., and Long, S. (2017). Logistic regression model for detecting radon prone areas in ireland. *Science of The Total Environment*, 599:1317–1329.
- Elío, J., Petermann, E., Bossew, P., and Janik, M. (2023). Machine learning in environmental radon science. *Applied Radiation and Isotopes*, 194:110684.
- ESR, E. S. o. R. (2015). Summary of the european directive 2013/59/euratom: essentials for health professionals in radiology. *Insights into imaging*, 6:411–417.
- Farid, S. (1997). Indoor and soil radon measurements in swaziland by track detectors. *Journal of environmental radioactivity*, 34(1):29–36.
- Feng, X., Zhong, J., Yan, R., Zhou, Z., Tian, L., Zhao, J., and Yuan, Z. (2022). Groundwater radon precursor anomalies identification by emd-lstm model. *Water*, 14(1):69.

- Fick, A. (1855). Poggendorff's flannel. *Physik*, 94(59):297.
- Fu, C.-C., Yang, T. F., Chen, C.-H., Lee, L.-C., Wu, Y.-M., Liu, T.-K., Walia, V., Kumar, A., and Lai, T.-H. (2017). Spatial and temporal anomalies of soil gas in northern taiwan and its tectonic and seismic implications. *Journal of Asian Earth Sciences*, 149:64–77.
- Garavaglia, M., Dal Moro, G., and Zadro, M. (2000). Radon and tilt measurements in a seismic area: temperature effects. *Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy*, 25(3):233–237.
- García-Talavera San Miguel, M., Martín Matarranz, J., Gil de Mingo, R., García Cardieno, J., and Suárez Mahou, E. (2013). El mapa predictivo de exposición al radón en españa. *Madrid: Consejo de Seguridad Nuclear*.
- GDAL/OGR contributors (2024). *GDAL/OGR Geospatial Data Abstraction software Library*. Open Source Geospatial Foundation.
- Gilmore, G. (2008). *Practical gamma-ray spectroscopy*. John Wiley & Sons.
- Giustini, F., Ciotoli, G., Rinaldini, A., Ruggiero, L., and Voltaggio, M. (2019). Mapping the geogenic radon potential and radon risk by using empirical bayesian kriging regression: A case study from a volcanic area of central italy. *Science of the Total Environment*, 661:449–464.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.
- Gregorič, A., Drinovec, L., Ježek, I., Vaupotič, J., Lenarčič, M., Grauf, D., Wang, L., Mole, M., Stanič, S., and Močnik, G. (2020). The determination of highly time-resolved and source-separated black carbon emission rates using radon as a tracer of atmospheric dynamics. *Atmospheric Chemistry and Physics*, 20(22):14139–14162.
- Guevara, C. (2019). Medición de radón 222 en lima metropolitana utilizando tres tipos de monitores con detectores de trazas nucleares. *Magíster en Física. Pontificia Universidad Católica del Perú*.
- Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.
- Hosoda, M., Tokonami, S., Suzuki, T., and Janik, M. (2020). Machine learning as a tool for analysing the impact of environmental parameters on the radon exhalation rate from soil. *Radiation Measurements*, 138:106402.
- Hughes, M. B., Elío, J., and Crowley, Q. (2022). A user's guide to radon priority areas, examples from ireland. *Journal of the European Radon Association*, 3:7586.
- IAEA, W. H. O. (2014). *Radiation Protection and Safety of Radiation Sources: International Basic Safety Standards*. Number GSR Part 3 in General Safety Requirements. INTERNATIONAL ATOMIC ENERGY AGENCY, Vienna.

- IAEA, W. H. O. (2015). *Protection of the Public Against Exposure Indoors due to Radon and Other Natural Sources of Radiation*. Number SSG-32 in Specific Safety Guides. INTERNATIONAL ATOMIC ENERGY AGENCY, Vienna.
- IARC, W. G. o. t. E. o. C. R. t. H. et al. (1988). Man-made mineral fibres. In *Man-made Mineral Fibres and Radon*. International Agency for Research on Cancer.
- INEI (2017). Censos nacionales 2017: Xii de población, vii de vivienda y iii de comunidades indígenas. Instituto Nacional de Estadística e Informática del Perú.
- IPEN, S. R. R. (1997). Oficina técnica de la autoridad nacional. *Instituto Peruano de Energía Nuclear*.
- Janik, M., Bossew, P., and Kurihara, O. (2018). Machine learning methods as a tool to analyse incomplete or irregularly sampled radon time series data. *Science of The Total Environment*, 630:1155–1167.
- Khanjer, E. F. and Abdullha, W. (2017). Using inverse distance weighting interpolation method to prepare maps of radon concentrations distribution in wasit province.
- Kumar, A. and Chauhan, R. (2014). Measurement of indoor radon–thoron concentration and radon soil gas in some north indian dwellings. *Journal of Geochemical Exploration*, 143:155–162.
- Levin, I., Karstens, U., Hammer, S., DellaColetta, J., Maier, F., and Gachkivskyi, M. (2021). Limitations of the radon tracer method (rtm) to estimate regional greenhouse gas (ghg) emissions—a case study for methane in heidelberg. *Atmospheric Chemistry and Physics*, 21(23):17907–17926.
- Liza Neciosup, R. A. (2017). Mapeo de los niveles de radón 222 en el distrito de san martín de porres (lima-perú) en el periodo 2015-2016.
- López, J., Ornelas, O. D., Sajo-Bohus, L., Rodriguez, G., and Chavarria, I. (2016). Correlation between underground radon gas and dormant geological faults. *Journal of Nuclear Physics, Material Sciences, Radiation and Applications*, 4(1):265–275.
- Mose, D. G. and Mushrush, G. W. (1997). Variable spacial and seasonal hazards of airborne radon. *Atmospheric Environment*, 31(21):3523–3530.
- Ndjana Nkoulou, J., Saïdou, Hamadou Yerima, A., Oumar Bobbo, M., Nugraha, E., Ipan, A., Najam, L., Suárez-Navarro, J., Yoho, B., and Kwato Njock, M. (2024). Lithological impact on radon levels: A study of indoor and soil gas radon in the centre region of cameroon. *Journal of Geophysical Research: Atmospheres*, 129(13):e2024JD040843.
- Nielson, K. and Rogers, V. C. (1982). Mathematical model for radon diffusion in earthen materials. Technical report, Rogers and Associates Engineering Corp., Salt Lake City, UT (USA).

- Oni, O. M., Aremu, A. A., Oladapo, O. O., Agboluaje, B. A., and Fajemiroye, J. A. (2022). Artificial neural network modeling of meteorological and geological influences on indoor radon concentration in selected tertiary institutions in southwestern nigeria. *Journal of Environmental Radioactivity*, 251:106933.
- Otton, J. K. (1992). *The geology of radon*. Washington: Government Printing Office.
- Partington, J. (1957). Discovery of radon. *Nature*, 179(4566):912–912.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Pereyra, P., López, M., and Vilcapoma, L. (2015). Concentration measurements of radon 222 indoors in lima—peru. *International Journal of Physics*, 3(4):165–169.
- Petermann, E., Meyer, H., Nussbaum, M., and Bossew, P. (2021). Mapping the geogenic radon potential for germany by machine learning. *Science of the Total Environment*, 754:142291.
- Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G., Kempen, B., Ribeiro, E., and Rossiter, D. (2021). Soilgrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *Soil*, 7(1):217–240.
- Rezaie, F., Panahi, M., Bateni, S. M., Kim, S., Lee, J., Lee, J., Yoo, J., Kim, H., Kim, S. W., and Lee, S. (2023). Spatial modeling of geogenic indoor radon distribution in chungcheongnam-do, south korea using enhanced machine learning algorithms. *Environment International*, 171:107724.
- Saeed, S. and Hassan, S. (2015). Determination of radon, uranium and other radioactive isotopes concentration in different types of natural water in nenava governorate. 8:227–244.
- Sakoda, A., Ishimori, Y., and Yamaoka, K. (2011). A comprehensive review of radon emanation measurements for mineral, rock, soil, mill tailing and fly ash. *Applied Radiation and Isotopes*, 69(10):1422–1435.
- Soria, A. F. (2020). *Física nuclear y de partículas*, volume 62. Universitat de València.
- Šperl, J. and Trčková, J. (2008). Permeability and porosity of rocks and their relationship based on laboratory testing. *Acta Geodyn Geomater*, 5(149):41–47.
- Steiner, A. (2016). Radiación: efectos y fuentes programa de las naciones unidas para el medio ambiente. *Pnuma*, page 68.
- Suarez, E., Fernández, J., Baeza, A., Moro, M., García, D., Pozo, J., and Lanaja, J. (2000). Proyecto marna. *Mapa de Radiación Gamma Natural*.

- Tareen, A. D. K., Asim, K. M., Kearfott, K. J., Rafique, M., Nadeem, M. S. A., Iqbal, T., and Rahman, S. U. (2019). Automated anomalous behaviour detection in soil radon gas prior to earthquakes using computational intelligence techniques. *Journal of environmental radioactivity*, 203:48–54.
- Thumvijit, T., Chanyotha, S., Sriburee, S., Hongsriti, P., Tapanya, M., Kranrod, C., and Tokonami, S. (2020). Identifying indoor radon sources in pa miang, chiang mai, thailand. *Scientific Reports*, 10(1):17723.
- UNSCEAR (2022). Sources, effects and risks of ionizing radiation unscear 2020/2021 report.
- US Cancer Statistics, W. G. et al. (2016). United states cancer statistics: 1999–2013 incidence and mortality web-based report. *Atlanta: US Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute*, page 201.
- USEPA, I. E. D. (2002). *A Citizen's Guide to Radon: The Guide to Protecting Yourself and Your Family from Radon*. US Environmental Protection Agency, Indoor Environments Division.
- Valcarce, D., Alvarellos, A., Rabuñal, J. R., Dorado, J., and Gestal, M. (2022). Machine learning-based radon monitoring system. *Chemosensors*, 10(7):239.
- Van Rossum, G. and Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.
- Vaupotič, J., Andjelov, M., and Kobal, I. (2002). Relationship between radon concentrations in indoor air and in soil gas. *Environmental Geology*, 42:583–587.
- Vílchez, M. S., Sosa Senticala, N. L., Pari Pinto, W., and Peña Laureano, F. (2020). Peligro geológico en la región cusco - [boletín c 74].
- Wentworth, C. K. (1922). A scale of grade and class terms for clastic sediments. *The journal of geology*, 30(5):377–392.
- Wes McKinney (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61.
- WHO (2016). Ionizing radiation, health effects and protective measures.
- WHO, W. H. O. (2009). *WHO handbook on indoor radon: a public health perspective*. World. Health. Organization.
- Wiegand, J. (2001). A guideline for the evaluation of the soil radon potential based on geogenic and anthropogenic parameters. *Environmental Geology*, 40:949–963.
- Zafirir, H., Barbosa, S. M., and Malik, U. (2013). Differentiation between the effect of temperature and pressure on radon within the subsurface geological media. *Radiation measurements*, 49:39–56.

-
- Zhang, A., Lipton, Z. C., Li, M., and Smola, A. J. (2021). Dive into deep learning. *arXiv preprint arXiv:2106.11342*.
- Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, 14(1):35–62.
- Zheng, Q., Yu, B., Wang, S., and Luo, L. (2012). A diffusivity model for gas diffusion through fractal porous media. *Chemical Engineering Science*, 68(1):650–655.