



AI-Driven Geospatial Analysis of Indoor Radon Levels: A Case Study in Chungcheongbuk-do, South Korea

Liadira Kusuma Widya^{1,2} · Fatemeh Rezaie^{3,4} · Jungsub Lee⁵ · Jongchun Lee⁵ · Bo Ram Park⁵ · Juhee Yoo⁵ · Woojin Lee⁶ · Saro Lee^{3,4}

Received: 25 July 2024 / Revised: 12 December 2024 / Accepted: 29 December 2024

© King Abdulaziz University and Springer Nature Switzerland AG 2025

Abstract

Radon is a naturally occurring radioactive gas found in many terrestrial materials, including rocks and soils. Due to the potential health risks linked to persistent exposure to high radon concentrations, it is essential to investigate indoor radon accumulation. This study generated indoor radon index maps for Chungcheongbuk-do, South Korea, selected factors such as lithology, soil depth texture, drainage, material composition, surface texture, soil thickness, calcium oxide and strontium levels, slope, topographic wetness index, wind exposure, valley depth, and the LS factor. These factors were analyzed using frequency ratios (FRs) to assess the influence on indoor radon distribution. The resulting maps were validated with several techniques, including FR, convolutional neural network, long short-term memory, and group method of data handling. The establishment of a geospatial database provided a basis for the integration and analysis of indoor radon levels, along with relevant geological, soil, topographical, and geochemical data. The study calculated the correlations between indoor radon and diverse factors statistically. The indoor radon potential was mapped for Chungcheongbuk-do by applying these techniques, to assess the potential radon distribution. The robustness of the validated model was assessed using the area under the receiver operating curve (AUROC) for both training and testing datasets.

Keywords Artificial Intelligence (AI) · Convolutional Neural Networks (CNN) · Geospatial Analysis · Group Method of data Handling (GMDH) · Indoor Radon Level · Long short-term Memory (LSTM)

1 Introduction

Radon, a naturally occurring radioactive gas produced by the decay of uranium and thorium in rocks and soils, poses a significant health risk when accumulated indoors due to its association with lung cancer (Mezquita et al. 2019; Lorenzo-González et al. 2019; Riudavets et al. 2022). Studies have shown that indoor radon is the second leading cause of lung cancer, with its impact being particularly pronounced in certain demographics such as males and individuals aged 40–45 (Qiang et al. 2023). Other health issues linked to radon exposure include respiratory problems such as potential aggravation of asthma (Mukharesh et al. 2022; Banzon et al. 2023). To address these risks, it is crucial to monitor and control indoor radon concentrations, as recommended by the European Union Basic Safety Standards, which emphasize the importance of proper ventilation in reducing radon levels in dwellings and workplaces (Qiang et al. 2023). Computational fluid dynamics simulations have demonstrated that indoor radon concentrations can

✉ Saro Lee
leesaro@kigam.re.kr

¹ Department of Science Education, Kangwon National University, Chuncheon-si 24341, Gangwon-do, Republic of Korea

² Department of Civil Engineering, Sunan Bonang University, Tuban 62313, Indonesia

³ Geoscience Data Center, Korea Institute of Geoscience and Mineral Resources (KIGAM), 124, Gwahak-ro, Yuseong-gu, Daejeon 34132, Republic of Korea

⁴ Department of Geophysical Exploration, Korea University of Science and Technology, Daejeon, Korea

⁵ Indoor Environment and Noise Research Division, Environmental Infrastructure Research Department, National Institute of Environmental Research, Seo-gu, Incheon 22689, Republic of Korea

⁶ College of AI Convergence, Dongguk University-Seoul, Jung-gu, Seoul 04620, Republic of Korea

be significantly influenced by ventilation rates, with higher concentrations near the floor and lower values near air inlets (Qiang et al. 2023). The South Korean Indoor Air Quality Control in Public Use Facilities Act and U.S. Environmental Protection Agency both recommend an upper limit of 148 Bq/m³. These findings emphasize the importance of regular monitoring and specific decisions to mitigate the risks associated with indoor radon exposure.

The likelihood of radon being present and its distribution on radon potential maps is intricately linked to many geological factors (Friedmann and Gröller 2010; Ciotoli et al. 2017; Haneberg et al. 2020), geochemical concentrations (Cho et al. 2015; Cho and Choo 2019), soil conditions (Panahi et al. 2022), topography (Szabó et al. 2014; Hwang et al. 2017), and climate (Petermann et al. 2021). Geological settings, such as the presence of uranium-bearing rocks and soils, can significantly affect radon levels due to the radioactive decay of uranium isotopes. Geochemical variation in the soil can also affect radon emanation rates, with higher radium and uranium concentrations leading to increased radon emissions. Soil conditions play a crucial role in radon migration and accumulation, as soil permeability and moisture levels can influence the transport of radon gas into indoor spaces. Topography, including elevation and slope, can also affect radon entry pathways into buildings, and indoor radon levels. Climate factors, such as temperature and humidity, can influence radon levels by affecting soil gas movement and the ventilation of indoor spaces. Changes in atmospheric pressure and temperature gradients can also influence radon transport. These interconnected factors underscore the complexity of radon distribution and the importance of considering multiple variables in radon risk assessment and mapping efforts.

Several recent studies using geospatial technologies and artificial intelligence (AI) approaches have shown the importance of statistical approaches for developing radon potential maps (Friedmann and Gröller 2010). Misclassification has the potential to misrepresent the severity of risks and undermine the reliability of indoor radon maps (Selamat et al. 2021). These studies have used geospatial approaches (Ciotoli et al. 2017; Coletti et al. 2022), spatial data on the frequency ratio (FR) method, which calculates the ratio between the frequency of a specific factor (e.g., geological or topographical characteristic) with radon occurrence in groundwater (Hwang et al. 2017), support vector machines, multivariate adaptive regression splines, random forest models (Petermann et al. 2021), recurrent neural networks, long short-term memory (LSTM) (Panahi et al. 2022), extreme learning machines, and random vector functional link algorithms (Rezaie et al. 2021). Machine learning is a powerful tool for solving complicated challenges involving several variables, and is effective for handling missing data,

noise, and big data processing (L'Heureux et al. 2017; Zhou et al. 2017).

Indoor radon levels in South Korea are influenced by various factors. Studies have highlighted the prevalence of elevated indoor radon levels in some regions of South Korea, pointing to the significance of geological factors like lithology, faults, and lineament in radon emissions (Park et al. 2019b; Rezaie et al. 2023). Other studies of indoor radon in South Korea have emphasized the roles of the type of house, building materials and walls, and the resident's ventilation habits in exacerbating indoor radon accumulation, particularly in urban areas with limited air circulation (Park et al. 2018, 2019a). Collectively, these studies underscore the complex interplay of geological, structural, and environmental factors that contribute to indoor radon levels in South Korea, emphasizing the importance of targeted mitigation strategies and regulatory measures to address radon exposure risks in residential and commercial buildings.

Radon exposure is a public health concern in South Korea, prompting the need for robust radon management strategies (Yoon et al. 2016). This study focused on predicting the indoor radon level distribution in Chungcheongbuk-do by examining the interplay between indoor radon and environmental factors such as geology, geochemistry, soil conditions, and topographic characteristics. The radon potential was visualized through the application of FR analysis with advanced AI approaches such as convolutional neural networks (CNN), long short-term memory (LSTM), and the group method of data handling (GMDH). Given the susceptibility of South Korea to radon exposure, the development of scientifically informed radon index maps is important for guiding national environmental decision-making and assessing regional radon exposure. By establishing a spatial radon database and using advanced spatial analysis techniques underpinned by AI approaches, this study seeks to provide a rigorous scientific basis for radon management, facilitating public safety and environmental protection.

2 Materials and methods

2.1 Materials

2.1.1 Study Area

Chungcheongbuk-do is located in the central part of South Korea, 127°29'57"E, 36°38'30"N (Fig. 1); it is bordered by Gyeonggi-do to the north, Gangwon-do to the east, Gyeongsangbuk-do to the southeast, and Jeollabuk-do to the southwest. It encompasses three cities and eight counties with a total population of 1,597,709 living in 703,916 households across an area of 7,407.7 km² (Chungcheongbuk-do

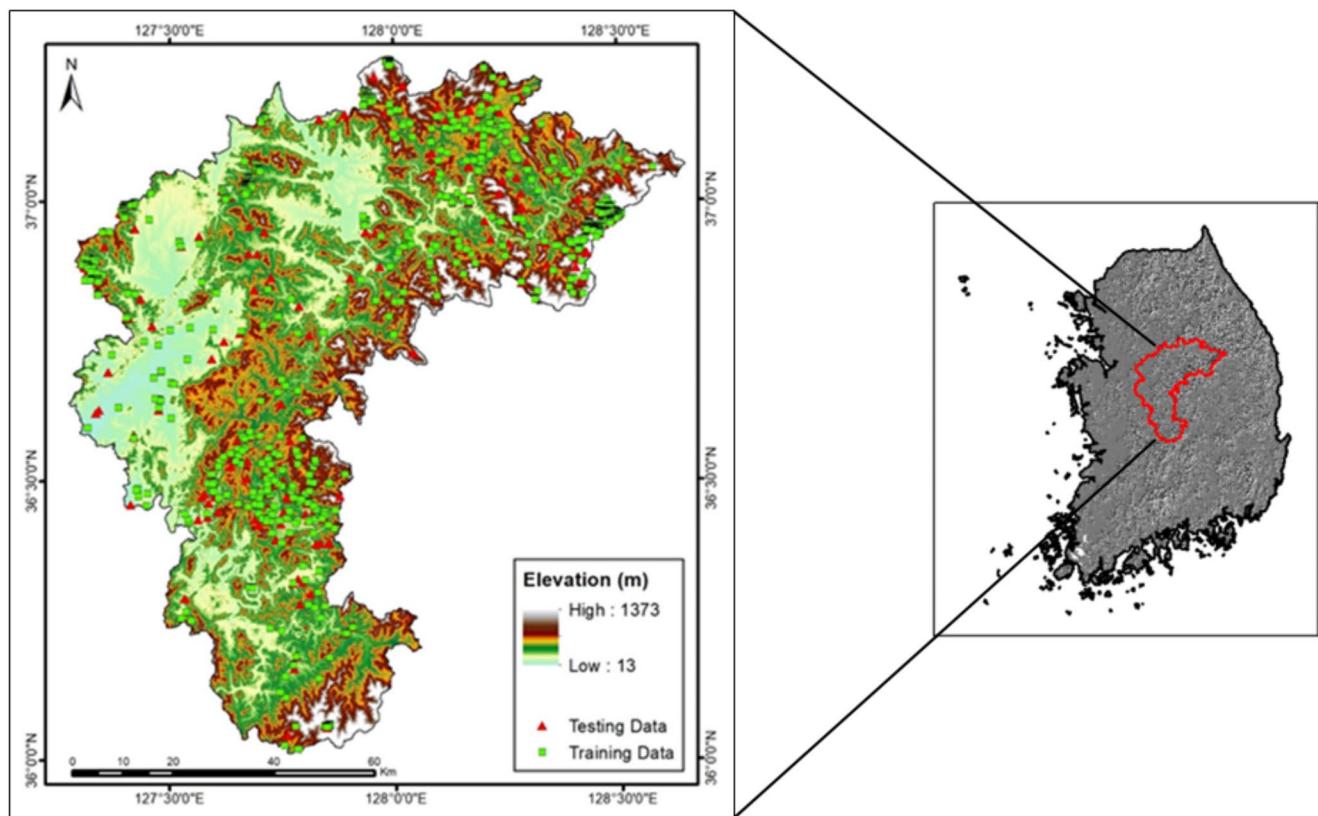


Fig. 1 Study location and measurements distributions (training and testing data) of Chungcheongbuk-do, South Korea, highlighting its exact geographical coordinates and spatial context

Government 2024). The geography of the region features diverse topography, including the Sobaek and Worak Mountains, plains and river valleys. The soil composition varies throughout the province, with fertile soils in the plains supporting agricultural activities. Geologically, Chungcheongbuk-do has a mix of formations influenced by its mountainous landscapes and proximity to fault lines, contributing to its geological diversity and potential geological impact on the area. Chungcheongbuk-do has a temperate climate with distinct seasons characterized by hot humid summers and cold dry winters. This area is known for its high agricultural yield and wide variety of plant types.

The National Institute of Environmental Science, South Korea conducted comprehensive indoor radon measurements for this study, while studying the associated influencing factors, to support the development of radon management plans by local governments. A total of 212 indoor radon measurements (Fig. 1) were selected across the Chungcheongbuk-do region during the winter of 2020–2021, employing RADUET-type detectors (Radosys Ltd., Budapest, Hungary). The winter season was chosen for data collection to capture higher radon levels typically observed during colder months. The winter season was chosen for data collection to capture higher radon levels typically

observed during colder months when buildings are more sealed, reducing ventilation and allowing radon to accumulate indoors. The detectors were installed in various residential buildings. To minimize potential thoron interference, detectors were positioned at least 1 m from walls in frequently occupied rooms. During data processing, discrepancies in coordinate values and missing data were corrected. For analysis, indoor radon values exceeding 148 Bq/L were classified as indicating elevated radon risk. This systematic approach integrated indoor radon measurements that it took itself, emphasizing specific factors to analyze the indoor radon level distribution map. For multi-media radon data, discrepancies in coordinate values and missing radon data were identified and rectified, ensuring the accuracy and reliability of the segmented radon data.

Table 1 gives detailed descriptions of geological, soil, geochemical, and topographic factors that can affect radon levels. Data related to radon were collected and analyzed for Chungcheongbuk-do using different methods. The geological data were from 1:250,000 scale maps of the Korea Institute of Geoscience and Mineral Resources (KIGAM). These data provide lithological information that contributes to determine the bedrock, which affects radon concentrations. Soil properties were from the Korean National

Table 1 Data sources

Data sources	Type of factors	Factors
Korea Institute of Geoscience and Mineral Resources (KIGAM), Korea	Geological	Lithology
National Institute of Agricultural Sciences, Korea	Soil Characteristics	Soil drainage Soil material Soil depth Deep soil texture Surface soil texture
Korea Institute of Geoscience and Mineral Resources	Geochemical	CaO Sr
National Geographic Information Institute, South Korea	Topographic	Slope TWI Wind exposition Valley depth LS factor

Institute of Agricultural Sciences included soil drainage, material, soil depth, deep soil texture, and surface soil texture. These factors influence the soil's permeability and its potential to transport radon concentrations. Geochemical mapping used data from the KIGAM included concentrations of calcium oxide (CaO) and strontium (Sr). These components may be used as indicators of the presence of materials that include uranium, which produce radon. Topographic mapping was conducted using the SAGA GIS program with digital elevation model inputs from the National Geographic Information Institute, topographic factors such as slope, Topographic Wetness Index (TWI), wind exposition, valley depth, and LS factor were analyzed. These factors influence the accumulation and movement of radon gas. After data collection and analysis, the FR method was used to identify factors associated with elevated radon levels by calculating the ratio of a factor's presence in high-radon areas. This method helps in understanding the spatial distribution factors of radon concentrations, which are listed in Table 1. The FR method's application in radon studies is supported by various research findings that highlight the influence of geological and environmental factors on radon levels (Hwang et al. 2017).

Assessing indoor radon levels entails the consideration of various factors influencing its accumulation within enclosed environments. Geological attributes, particularly lithology characteristics, influence radon concentration levels due to the presence of uranium in rocks and soils. Radon, a decay product of uranium, is released from geological formations and can accumulate in indoor environments. The geological composition, including rock types and soil characteristics, plays a role in determining radon levels (Cho et al. 2015; Ciotoli et al. 2017). Radon levels in indoor environments are primarily affected by the local geology. Figure 2 provides a detailed overview of the lithological characteristics

in Chungcheongbuk-do, which is characterized by widespread Daebu granite, a Jurassic granite (Jgr) recognized for its impact on radon levels, which contributes to radon levels due to its higher uranium content (Cho and Choo 2019). The decay of these radioactive elements, such as uranium, contributes to radon production. The alteration of these minerals can further enhance radon release, as structural changes increase permeability and facilitate radon escape (Hwang et al. 2024). This phenomenon is influenced by various factors, including mineral composition, geological processes, and environmental conditions.

Soil factors include deep texture, drainage, material, surface texture, and thickness, (Fig. 3). These play roles in the movement of radon from the soil into indoor spaces and variation in radon concentrations. Radon risk can be determined by amalgamating the radon potential associated with these soil properties; with different soil types exhibiting distinct transport mechanisms (Cosma et al. 2013a; Dicu et al. 2023). For instance, clay soils typically show diffusive radon transport, while sandy soils exhibit convective transport, with the latter having higher radon flux densities (Gavriliev et al. 2022).

The geochemical data, which can influence the movement of radon indoors, were acquired in a survey that considered the presence of calcium oxide (CaO) and strontium (Sr) (Fig. 4). Geochemical maps were generated using spatial analysis tools with interpolation. CaO and Sr are pertinent indicators in the geochemical composition within the soil through their effects on soil properties (Liu et al. 2013; Wen et al. 2020) and influence radon migration (Cho et al. 2015). The geochemical composition of soil and rock can affect the mineralogical environment in which uranium occurs, thereby impacting indoor radon levels (Nunes et al. 2023). For example, uranium incorporated into secondary minerals, which form coatings on soil grains, may have higher radon emanation rates compared to uranium trapped within the crystal lattice of primary minerals (Chitra et al. 2021). The presence of CaO and Sr can indicate the geochemical conditions favorable for radon release, as they are often associated with processes that lead to the redistribution and concentration of uranium in the environment (Plechacek et al. 2022).

Topographic factors (Fig. 5) are another facet of indoor radon level analysis. Slope, the topographic wetness index (TWI), wind exposition, valley depth, and slope length and the steepness (LS) factor collectively contribute to how the surrounding landscape influences radon entry into buildings, influencing radon migration pathways and concentration variation (Cho et al. 2015; Rezaie et al. 2023). The slope influences radon migration by affecting the flow of groundwater and surface runoff, which can carry radon to lower-lying areas. Steeper slopes may facilitate faster radon

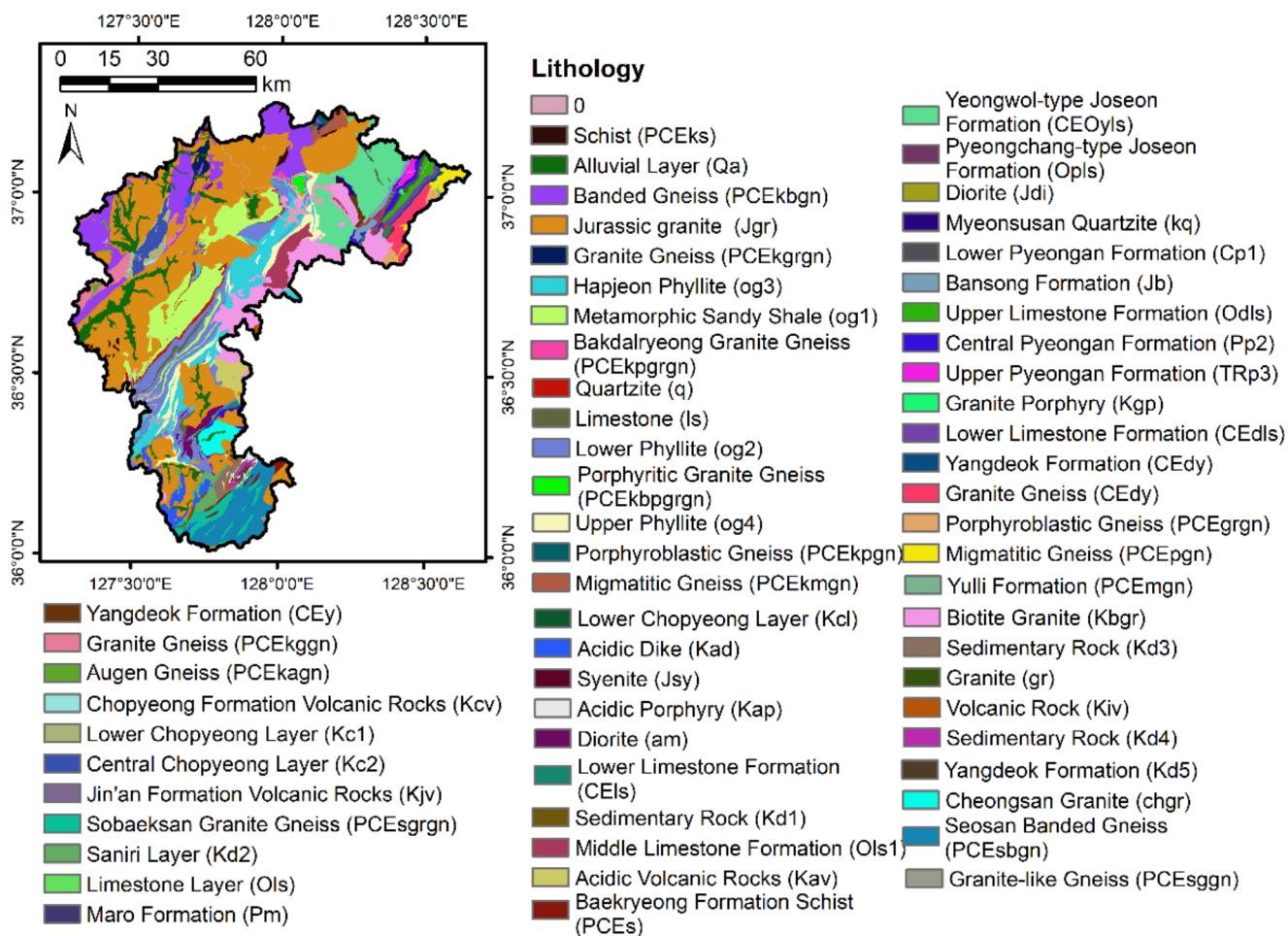


Fig. 2 Lithology factor, as an influencing factor related to indoor radon concentration, detailing the various types of lithology

dispersion, while flatter areas can lead to radon accumulation, especially if they coincide with low-permeability soils.

The TWI indicates the potential for water accumulation in the landscape, which can trap radon near the surface when soil becomes saturated. Higher TWI values are typically associated with poorly drained areas (Allende-Prieto et al. 2024). Conversely, areas with low TWI values are more likely to allow radon to escape into the atmosphere. Wind exposition influences radon dispersion, with areas shielded from wind potentially experiencing higher radon levels due to reduced air movement, while locations with greater wind exposure can facilitate radon dilution (Doering 2019).

Valley depth significantly affects radon accumulation, as deeper valleys tend to have stronger upvalley winds and more pronounced vertical circulation cells, which can lead to the formation of elevated inversion layers that trap radon near the valley floor (Wagner et al. 2015). The LS factor reflects the combined effects of slope length and steepness on erosion potential, which can also influence soil permeability and radon migration (Das et al. 2022). Regions with high LS values may exhibit increased soil erosion,

potentially exposing uranium-rich layers and enhancing radon release.

2.2 Methodology

The methodological framework used to generate indoor radon level maps for Chungcheongbuk-do, South Korea, by integrating various factors that influence radon distribution with advanced AI approaches. The key factors affecting indoor radon levels, which are grouped into four categories: geology, soil, geochemical, and topography. The methodology involved splitting the radon data into two parts for model development and validation: 70% of the data was used for training the models, while the 30% was set aside for testing. To evaluate the significance of different factors contributing to indoor radon levels, the FR method was applied. This approach plotted an indoor radon level map for Chungcheongbuk-do using the FR and AI approaches, including GMDH, CNN, and LSTM models. This integrated approach allowed for the capture of spatial and environmental variables that influence indoor radon distribution.

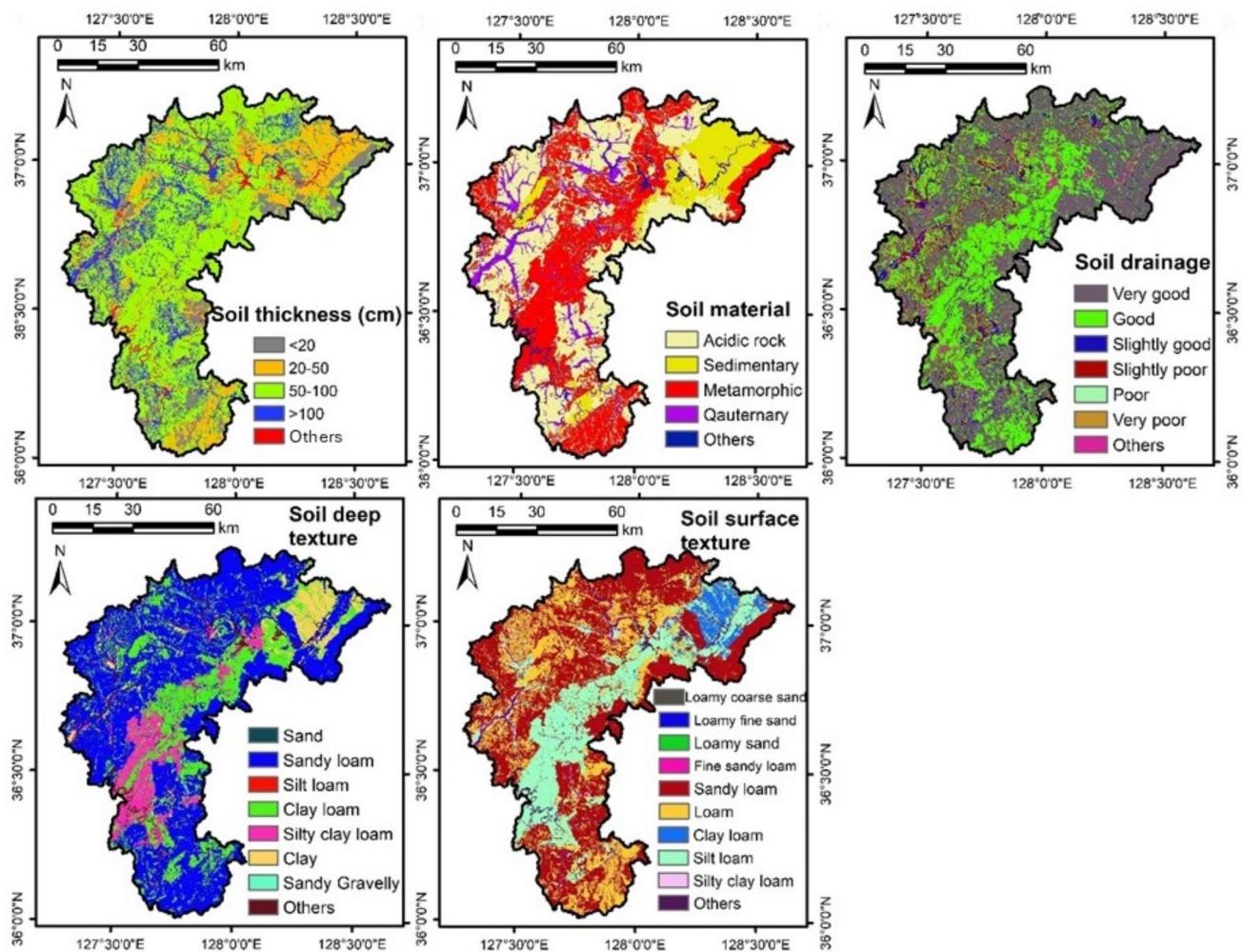


Fig. 3 Soil characteristics that are associated with indoor radon concentration

FR analysis applied to identify key factors influencing radon levels, while GMDH, CNN, and LSTM models provided data handling for generating indoor radon probability maps. The model prediction accuracy was evaluated using the area under the receiver operating characteristic curve (AUROC). Figure 6 shows the methodological steps employed to achieve accurate radon level mapping for the region.

2.2.1 Identifying Influencing Factors of Indoor Radon Level Variation Using the FR

The FR is used to identify potential statistical relationships between a given phenomenon and its associated variables (Lee and Talib 2005). This study determined the FR of variables related to indoor radon and categorized them based on each class. In the correlation analysis, FR is a statistical method used to quantify the relationship between indoor radon occurrence and specific factors, such as soil, geochemical, geological, and topography characteristics, by

comparing the proportion of areas with elevated radon levels to those factors. The presence or absence of radon must first be defined and are set to 0 and 1 based on a threshold value of the indoor radon index of 148 Bq/m³ to indicate areas with high and low radon concentrations. For indoor radon, FR was calculated using 1 for the indoor radon index in Chungcheongbuk-do. The FR approach was used to evaluate the correlation between the indoor radon index and various factors and was calculated by dividing the area related to each indoor radon variable in a subclass and the entire study area within that subclass, as shown in Eq. 1 (Huang et al. 2020).

$$FR = \frac{N_{(I_i)} / N_{(F_i)}}{N_{(I)} / N_{(A)}} \quad (1)$$

where $N_{(I_i)}$ is the pixels of the indoor radon in subclass i of the factor; $N_{(F_i)}$ is the total number of pixels in subclass i ;

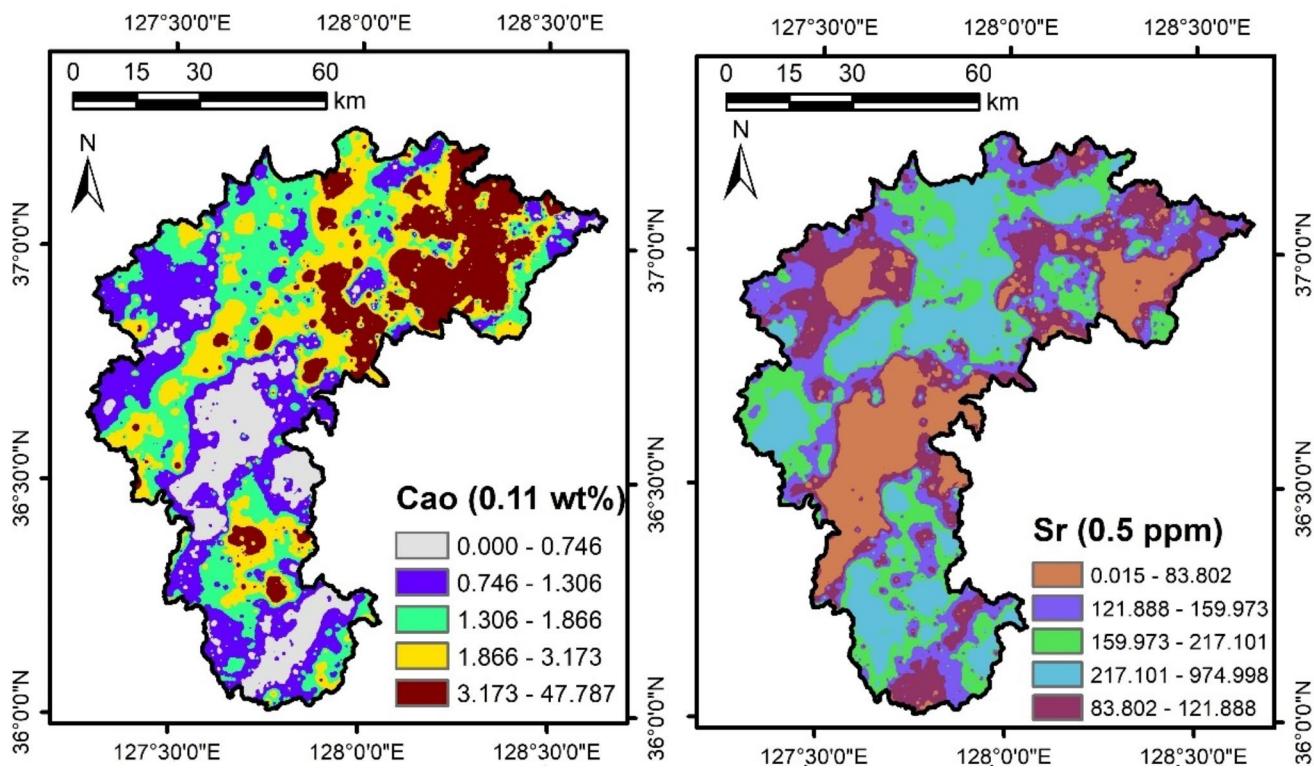


Fig. 4 Geochemical aspects, specifically focusing on variables such as CaO and Sr, which have been correlated with indoor radon concentration

$N_{(I)}$ is the total indoor radon distribution of the factor; and $N_{(A)}$ is the total area.

2.2.2 Model Description

2.2.2.1 Frequency Ratio The FR model is a statistical model that is used to analyze the frequency of events or patterns in a dataset. The indoor radon level map was developed using the FR approach, which is weighted based on the significance of each factor in contributing to indoor radon levels, allowing for a weighted sum of the factors to generate an indoor radon level probability map (Megahed et al. 2023). The weighted sum tool allows strategic weighting and the amalgamation of various factors to produce an indoor radon level map.

2.2.2.2 Group Method of Data Handling The GMDH is a robust approach to mathematical modeling and data analysis developed by Alexey G. Ivakhnenko in the 1970s that has been applied in various fields (Ivakhnenko 1970). The

GMDH algorithm uses a self-organization principle to identify the optimal model complexity by systematically evaluating numerous models that meet the specified criteria (Ivakhnenko 1978). The GMDH algorithm consists of multiple functions that effectively handle several issues and enhance the precision of problem-solving outcomes. The functions include linear, polynomial, and ratio-polynomial variations (Ivakhnenko and Ivakhnenko 2000). The relationship between input and output variables can be described by a complex discrete form of the Volterra functional series, commonly referred to as the Kolmogorov-Gabor polynomial (Farlow 1984). Equation 2 gives the correlation between the input and output variables of the model. GMDH starts with a set of input data that includes multiple variables. Each variable represents a feature or attribute of the dataset. The variables are grouped into different sets or layers. Within each group or layer, mathematical models are developed to represent the relationships among variables. The process continues until a predefined stop criterion is met or until the best possible model is achieved. The resulting model can be used for making predictions on new data.

$$y = a_0 + \sum_{i=1}^n a_i x_i + \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_{ijk} x_i x_j x_k + \dots \quad (2)$$

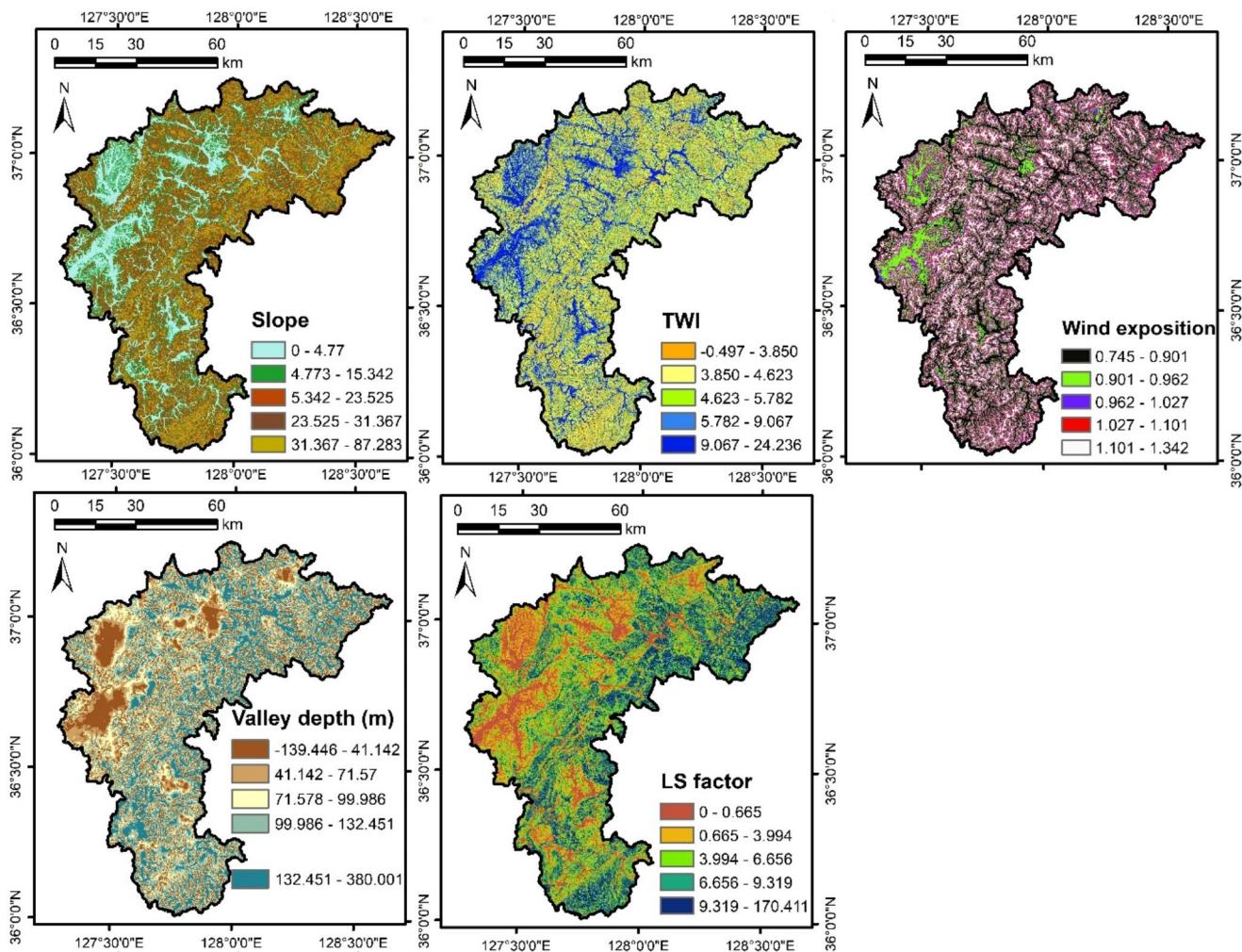


Fig. 5 Topography features, including Slope, the topographic wetness index (TWI), wind exposure, valley depth, and LS factor, which have been identified as influencing associations with indoor radon concentration

Where y is the prediction result; x is the input variables vector; a is the coefficient calculated using the least squares error approach; and n denotes the number of input variables.

2.2.2.3 Convolutional Neural Network A CNN is a deep-learning algorithm that belongs to the broader category of AI, deep learning approaches. Deep learning is a specialized area of machine learning that emphasizes the use of artificial neural networks that include a layer committed to the convolution operation. The fundamental architecture of the CNN model consists of convolution, pooling, and fully connected layers (Lecun et al. 1998; Yamashita et al. 2018).

The input layer serves as the initial interface for raw data, converting it into a numerical form, often expressed as a tensor, which is a multi-dimensional data array. The core part of CNN is a convolutional layer functions, where convolution operations are performed on the inputted numerical data through a set of adaptable filters, also known as kernels

(Thi Ngo et al. 2021). Upon completion of the convolution operations, an activation function is employed on a per-element basis to the resultant data from the convolutional layer. This function injects non-linearity into the computational model, which is essential for the network's ability to capture and represent complex data relationships (Berkani et al. 2023).

Pooling operations contribute to the reduction of computational demand and bolster the model's invariance to shifts in position. By downsampling the feature maps from the convolutional layers, the pooling layer condenses the dimensionality of the data (Barata et al. 2019). General techniques include max pooling, which isolates the highest value within a designated subsection, and average pooling, which determines the sectional average (Chawshin et al. 2022).

The fully connected layer plays a crucial role in rendering predictions by integrating the high-level features derived from antecedent layers (Ma et al. 2021). Meanwhile, the

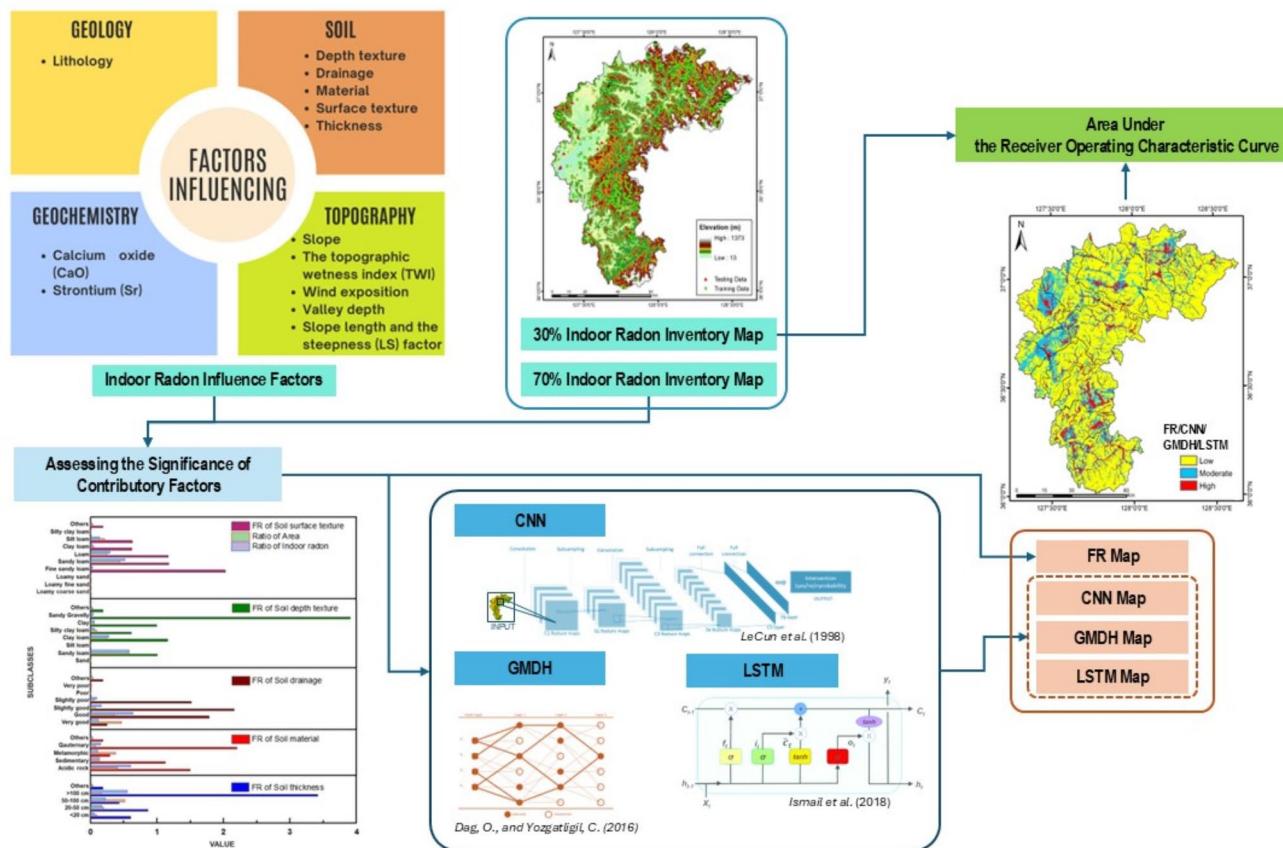


Fig. 6 The study applied a methodological framework for developing an indoor radon map by applying FR and AI approaches

output layer in a CNN, responsible for producing the ultimate output, generally comprises neurons that align with each of the defined categories.

2.2.2.4 Long Short-Term Memory LSTM is another type of deep neural network algorithm in which the network output is fed back into the network as subsequent input (Kong et al. 2019). The structural framework of LSTM networks excels in identifying intricate spatial configurations and temporal trends across diverse settings. LSTM networks are structured around unique memory units that incorporate gating mechanisms, which enable the model to preserve and modify data over extended sequences, ensuring crucial information is not discarded.

The key components of an LSTM cell include input gates, forget gates, cells, output gates, and cell outputs state (Graves 2012). The input gate is responsible for managing how new data is assimilated into the cell state, effectively serving as a filter for incoming information. The forget gate critically assesses which parts of the stored information are no longer pertinent and thus should be removed, taking into account both the new input and the preceding hidden state.

Concurrently, the candidate cell state, poised with potential information for addition to the cell state, hinges on the decisions made by both the input and forget gates. Following this, the cell's output state, a key component of the LSTM's memory architecture, is refreshed with this vetted information, thereby encapsulating the crux of the input sequence at that specific moment. The output gate's role is pivotal in dictating the quantum of updated information that is conveyed from the cell state, thereby modulating the balance between memory retention and attrition across time steps. This mechanism of selective information management within the LSTM's memory cells facilitates the maintenance of a continuous and relevant data stream across prolonged sequences, effectively mitigating the challenge of information degradation over extended periods.

2.2.3 Model Performance Assessment Using AUROC

The AUROC is a powerful metric for evaluating the accuracy of predictions made by AI models. The AUROC method was used to validate the indoor radon index potential maps. For validation, training (70% of the total radon data) and testing (30% of the total radon data) were selected randomly

to generate radon index maps using the training data. In the training and testing phase, the AUROC can be used as a performance indicator to evaluate the efficacy of the FR and AI algorithms (Bradley 1997). The ROC expresses the predictive index value obtained from the prediction map as the ratio of the location of radon data to the total area. The x -axis of the graph shows the upper percentage of areas with a high indoor radon index and the y -axis shows the lower percentage of areas with training or testing radon data (Pencina et al. 2008). The model's predictive performance is categorized into five ranges based on the AUROC: fail (0.5–0.6), poor (0.6–0.7), fair (0.7–0.8), good (0.8–0.9), and excellent (0.9–1.0) (Carter et al. 2016). The AUROC is used to assess model performance in many fields, such as flood susceptibility maps (Dodangeh et al. 2020), groundwater potential (Panahi et al. 2020), and radon distribution maps (Rezaie et al. 2022).

3 Results

3.1 Parameter Selection and Experimental Analysis

CNN was employed with tuning of hyperparameters to optimize performance of the model. The network architecture consisted of a convolutional layer with a 14×2 kernel and 30 filters, followed by batch normalization, ReLU activation, dropout, and average pooling. The fully connected layers included 10 neurons in the first layer and 1 in the output layer. Key hyperparameters such as learning rate 0.02, mini-batch size 8, and training epochs 100 were selected after experimentation, with early stopping implemented to prevent overfitting. The 70–30% split for training and testing validation was used to ensure the model's robustness.

3.2 Analysis of Influential Factors

In this study, correlations were calculated using the indoor radon with various geological, soil, geochemical, and topographic factors using databases established to produce radon index maps. The FR approach was used to select and evaluate the correlations between the indoor radon index and the different factors. Finally, 13 factors were selected to be used for the indoor radon index map. Figure 7 illustrates the FR of the geological factor lithology class. In the lithology class, the subclasses Jgr and Yeongwol-type Joseon Formation (CEOyls) have high FR values of indoor radon 1.489 and 1,717, respectively.

Figure 8 shows the FR values of soil factors including deep texture, drainage, material, surface texture, and thickness. The clay loam subclass within the soil deep texture class had a high FR of 1.156. The slightly good subclass

within the soil drainage class had the highest FR of 2.152. The acidic rock subclass of soil material dominated the indoor radon index with an FR of 1.156. The indoor radon index was mostly influenced by the sandy loam subclass of soil surface texture, with an FR of 1.156. The soil thickness subclass of the > 100 cm had a high FR of 3.415.

For indoor radon index distribution, the Sr geochemistry was related to the patterns of indoor radon occurrence. As depicted in Fig. 9, CaO and Sr influenced the indoor radon index map. Notably, the CaO class with the subclass range of 3.173–47.787 had a high FR of 1.323, which affected the indoor radon index distribution. The 217.101–974.998 subclass within the Sr class had a high FR of 1.607.

The results of the topographic analysis, that are illustrated in Fig. 10, summarize of the variables that contributed to the indoor radon index distribution. Analyzing each subclass of each topographic parameter showed notable patterns. The slope factor within the range of 0–4.773 had an FR value of 2.676, and influenced indoor radon index levels. Similarly, locations characterized by the higher TWI subclass (9.067–24.236) had an FR of 1.936, indicating an impact on the indoor radon distribution. Furthermore, areas with lower wind exposition (0.745–0.901) had a high FR of 3.072. The Valley depth in the range of 132.451–380.001 had an FR value of 2.391. Areas with lower LS Factor values (0.665–3.994) had an FR of 2.238.

3.3 Potential Indoor Radon Level Map

Incorporating a comprehensive set of factors, our study used four distinct modeling approaches to analyze indoor radon levels in Chungcheongbuk-do, including probability models, FRs, and sophisticated LSTM, CNN, and GMDH AI techniques to generate a suite of radon index maps. Each map assigned an index value to every 10×10 m grid segment, categorizing the radon concentration into high, moderate, and low. To enhance their interpretability, the maps represented high, moderate, and low levels in red, blue, and yellow, respectively, red areas constituting 10% of the total, blue 20%, and yellow 70%. The distribution of the indoor radon level across Chungcheongbuk-do is illustrated in Fig. 11.

3.4 Model Validation

The model prediction was validated by a thorough review of AUROC metrics. In the training phase, the AUROC values for the FR, CNN, LSTM, and GMDH, models were 90%, 89%, 90%, and 82%, respectively (Table 2). The calculations represent the ability of the models to recognize and understand patterns in the training dataset. In the next validation stage using the testing dataset, the FR, CNN, LSTM,

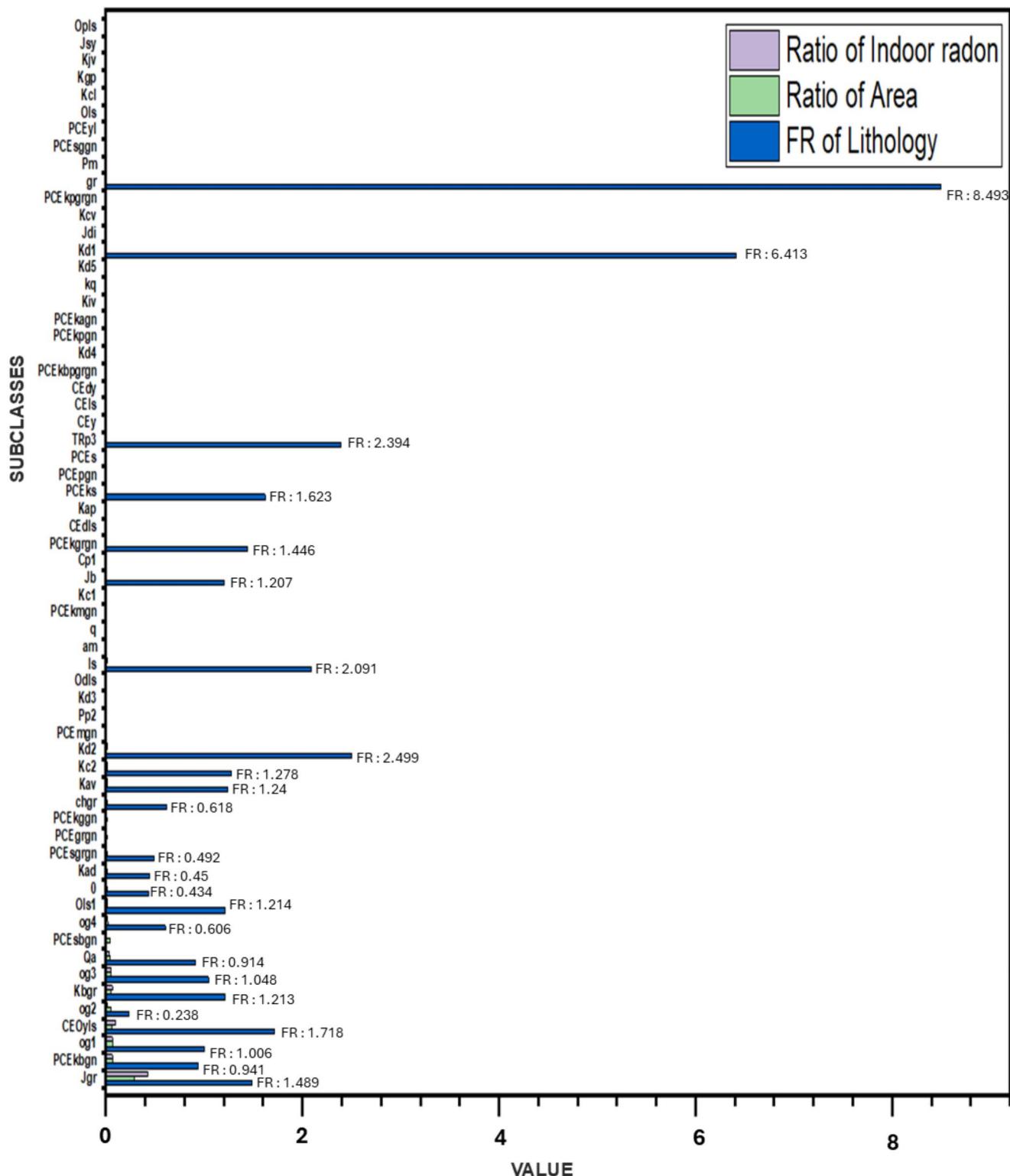


Fig. 7 FR of lithology within each subclass and its influence on indoor radon level

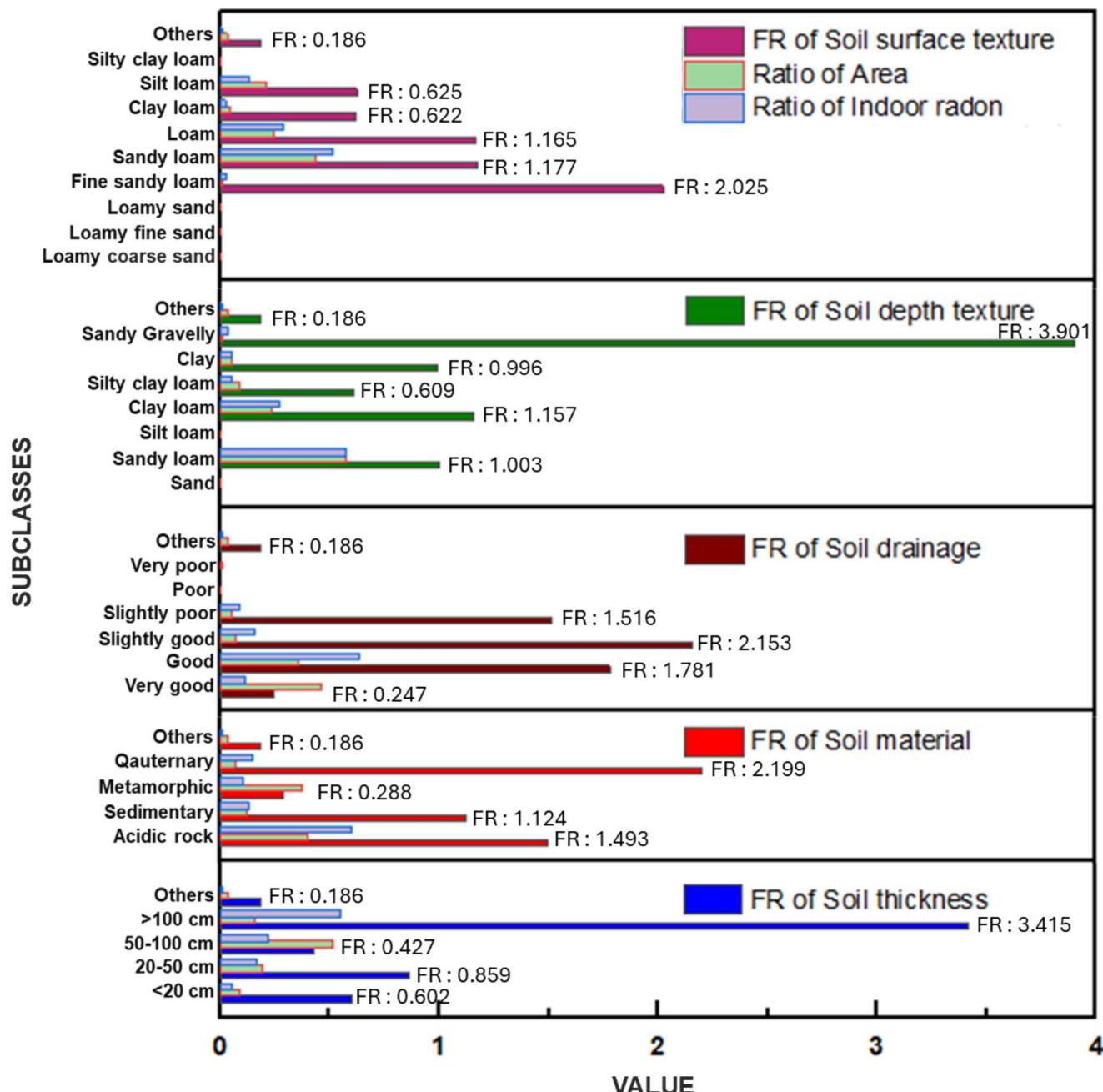


Fig. 8 FR of soil categories, including thickness, material, drainage, depth texture, surface texture for each subclass, and their influence on indoor radon level

and GMDH models had respective AUROCs of 88%, 86%, 88%, and 81%. The AUROC values obtained during testing are essential criteria that demonstrate the power of the models to predict indoor radon level maps. The AUROC metrices calculated in this study are shown in Table 2 and have accuracies generally exceeding 80%, indicating that the index maps are valid.

4 Discussion

Identifying influencing factors is the initial step for mapping the indoor radon distribution. This study used the information gain ratio (IGR) to identify the quantitative predictive power of the factors influencing prediction (Abedini et al. 2019). The IGR approach enhances our comprehension of

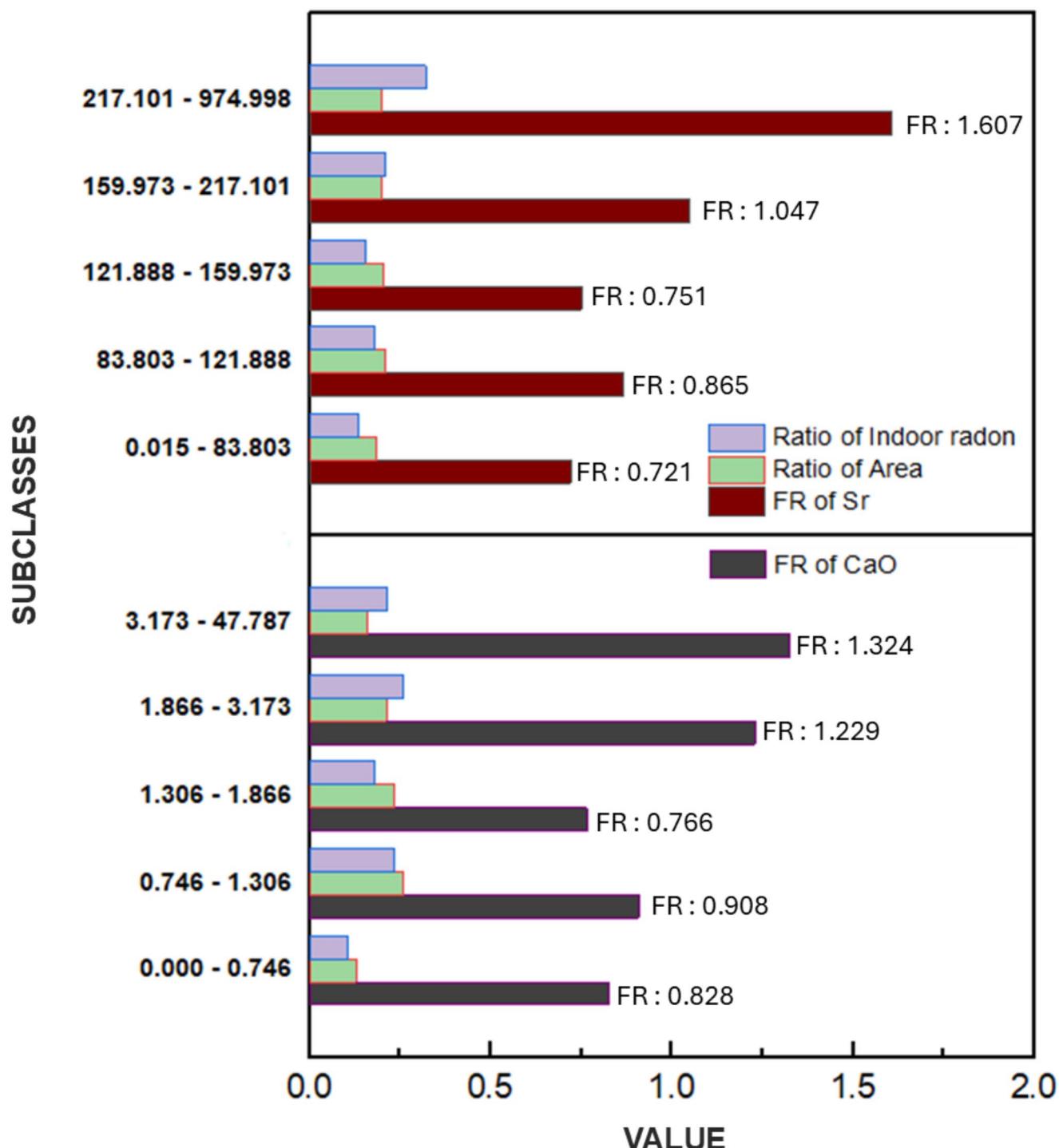


Fig. 9 FR of geochemical categories, including CaO and Sr, within each subclasses, along with their impact on indoor radon level

the effects of diverse environmental and architectural attributes on indoor radon levels. As shown in Table 3, this study used the IGR technique to identify 13 factors that affect indoor radon levels in Chungcheongbuk-do: lithology, soil deep texture, soil drainage, soil material, soil

surface texture, and soil thickness, CaO, Sr, topographic slope, TWI, wind exposition, valley depth, and LS factor. Lithologically, Daebu Granite, which is common in the center of the Korean Peninsula, dominates Chungcheongbuk-do (Cho and Choo 2019). As shown in Fig. 2 and the FR

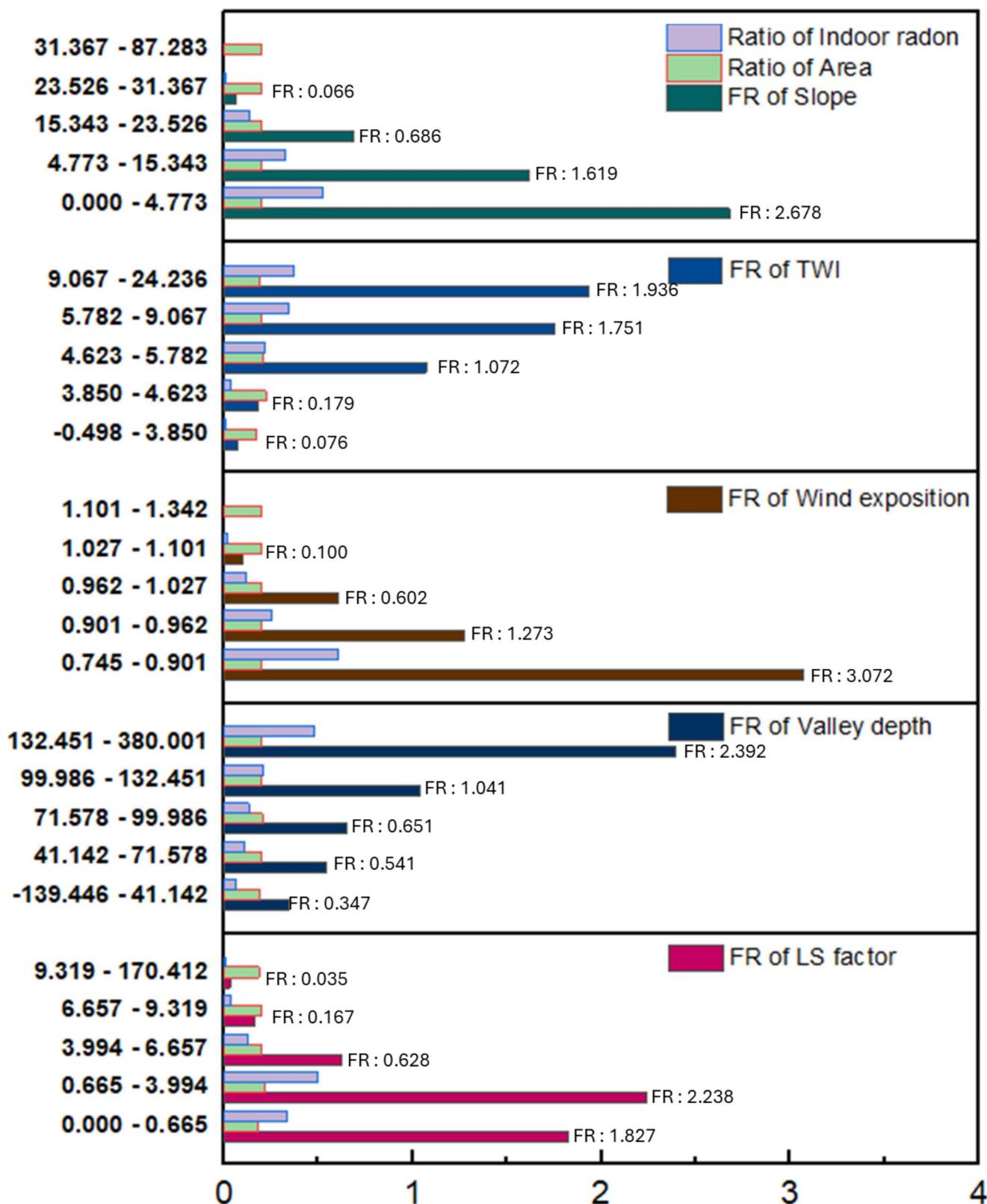


Fig. 10 FR of topography categories, encompassing LS factor, valley depth, wind exposition, TWI, and Slope, within each subclass, and their respective impacts on indoor radon level

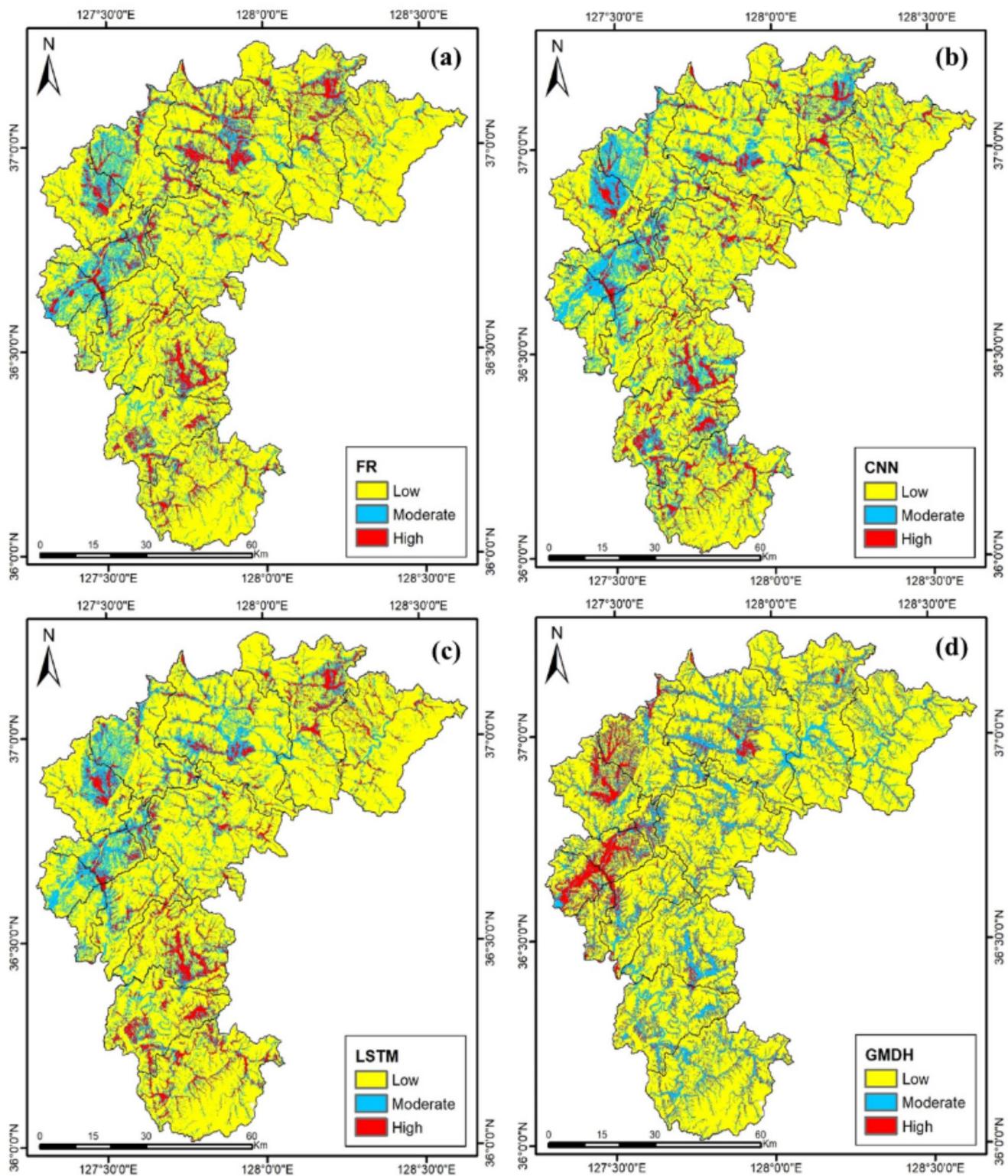


Fig. 11 Indoor radon level Map for (a) FR, (b) CNN, (c) LSTM, and (d) GMDH maps

Table 2 The AUROC values of FR, CNN, LSTM and GMDH

Train/Test	Model	Indoor Radon
Train Data	FR	90%
	CNN	89%
	LSTM	90%
	GMDH	82%
Test Data	FR	88%
	CNN	86%
	LSTM	88%
	GMDH	81%

Table 3 The IGR value for each factors

Classes	IGR
Lithology	0.72
Soil thickness	0.697
Soil surface texture	0.694
Soil depth texture	0.779
Soil drainage	0.788
Soil material	0.808
Sr	0.256
Cao	0.135
LS factor	0.806
Valley depth	0.419
Wind exposition	0.878
TWI	0.673
Slope	0.89

analysis, Daebo Granite groundwater has high radon levels (Cho et al. 2019). Daebo Granite and similar formations often contain moderate to high levels of uranium, which is the primary source of radon gas. The decay of uranium in these rocks leads to the production of radon, which can then migrate into the surrounding environment (Lee et al. 2021; Hashemi et al. 2024). A previous study developed models for geogenic radon that integrated soil hydraulic properties, soil physical characteristics, uranium concentrations, and chemical properties, the SAGA wetness index, and climate data (Petermann et al. 2021).

We applied several AI algorithms and meticulously assessed their performance to determine their efficacy in predictive modeling. The LSTM has demonstrated high performance compared to CNN and GMDH models when assessed using the AUROC. This aligns with previous research highlighting the robustness of LSTM at capturing the attribute information of conditioning factors and powerful sequential modelling capabilities to handle the spatial relationship. The capabilities of the FR, CNN, and GMDH models were also estimable on the AUROC values. This disparity underscores the significance of algorithm selection based on the specific nature of the dataset and the desired model outcome. However, the GMDH result was low compared to others. Noted that GMDH has limitations and may be sensitive to noise in the data (Elbagoury et al. 2023). The success of GMDH depends on careful parameter tuning and

appropriate validation techniques to ensure the reliability of the developed models.

The utilisation of GMDH, CNN, and LSTM AI algorithms gives strengths and addresses challenges in the development of indoor radon level maps. AI approaches substantially improve the predictive accuracy of the distribution of indoor radon levels. By proactively identifying areas with high radon concentrations, residents can implement appropriate preventative steps. Radon exposure can be minimized by installing radon mitigation or improving the construction of houses or buildings. The development of indoor radon maps has advantages for public health, environmental monitoring, and urban planning. However, these algorithms also have limitations. The datasets used to generate radon potential maps might be limited at capturing the impact of various factors on radon levels. These factors include climate change (Petermann et al. 2021) and specific building characteristics such as the number of floors, window types, presence of elevators or basements, building materials, heating systems, renovations, room types, construction period, and ventilation systems, all of which can affect radon concentrations (Cosma et al. 2013b; Ivanova et al. 2017). Based on this study, further in-depth research is needed to optimize the factors needed for future standardized radon index mapping.

5 Conclusion

This study showed that it is the imperative to investigate indoor radon level distributions by mapping radon concentrations. Using sophisticated AI techniques such as CNN, LSTM, and GMDH, the study successfully generated and validated indoor radon level maps for Chungcheongbuk-do, South Korea. We analyzed the factors affecting indoor radon levels by combining various geological, soil, topographical, and geochemical datasets. Comprehensive indoor radon level distribution maps were generated using statistical approaches and AI algorithms. Our findings indicate that LSTM, as a AI, deep learning approach, indeed outperformed other models in terms of predictive accuracy. Validation using AUROC analysis shows the robustness of the developed models. The LSTM achieved AUROC values of 90% for training and 88% for testing datasets. This study enhances our comprehension of indoor radon distributions and allows for recommendations for future geospatial research and the development of effective mitigation plans, especially in radon-prone areas such as Chungcheongbuk-do, South Korea.

Author Contributions Liadira Kusuma Widya: Writing – original draft, writing-review and editing, Methodology, Validation, Visualization; Fatemeh Rezaie: Formal analysis, Methodology, Visualiza-

tion; Jungsub Lee: Data curation, writing-review and editing; Bo Ram Park: Formal analysis, and writing-review; Jongchun Lee: Data curation, writing-review and editing; Juhee Yoo: Data curation, writing-review and editing; Woojin Lee: Methodology, writing-review and editing; Saro Lee: Conceptualization, Supervision, Funding acquisition, Project administration.

Funding This research was supported by the Basic Research Project of the Korea Institute of Geoscience and Mineral Resources (KIGAM) and the National Research Foundation of Korea (NRF) grant funded by Korea government (MSIT) (No. 2023R1A2C1003095). Also, this research was commissioned by the National Institute of Environmental Research (NIER) fund by the Ministry of Environment (MOE) of the Republic of Korea (NIER-2020-04-01-001).

Declarations

Conflict of interest The authors declare no competing interests.

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

References

- Abedini M, Ghasemian B, Shirzadi A, Bui DT (2019) A comparative study of support vector machine and logistic model tree classifiers for shallow landslide susceptibility modeling. Environ Earth Sci 78:560. <https://doi.org/10.1007/s12665-019-8562-z>
- Allende-Prieto C, Roces-García J, Sañudo-Fontaneda LÁ (2024) The high-resolution calibration of the Topographic Wetness Index using PAZ Satellite Radar Data to determine the optimal positions for the Placement of Smart Sustainable Drainage systems (SuDS) in Urban environments. Sustainability 16:598. <https://doi.org/10.3390/su16020598>
- Banzon TM, Greco KF, Li L et al (2023) Effect of radon exposure on asthma morbidity in the School Inner-City Asthma study. Pediatr Pulmonol 58:2042–2049. <https://doi.org/10.1002/ppul.26429>
- Barata F, Kipfer K, Weber M et al (2019) Towards device-agnostic mobile cough detection with convolutional neural networks. In: 2019 IEEE International Conference on Healthcare Informatics (ICHI). IEEE, pp 1–11
- Berkani S, Guermah B, Zakroum M, Ghogho M (2023) Spatio-temporal forecasting: a survey of data-driven models using exogenous data. IEEE Access 11:75191–75214. <https://doi.org/10.1109/ACCESS.2023.3282545>
- Bradley AP (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit 30:1145–1159. [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2)
- Carter JV, Pan J, Rai SN, Galanduk S (2016) ROC-ing along: evaluation and interpretation of receiver operating characteristic curves. Surgery 159:1638–1645. <https://doi.org/10.1016/j.surg.2015.12.029>
- Chawshin K, Berg CF, Varagnolo D, Lopez O (2022) Automated porosity estimation using CT-scans of extracted core data. Comput Geosci 26:595–612. <https://doi.org/10.1007/s10596-022-10143-9>
- Chitra N, Sundar SB, Valan II et al (2021) Modeling and experiments to Estimate Radon emanation factor in soil—grain size and moisture effect. Radiat Prot Dosimetry 194:104–112. <https://doi.org/10.1093/rpd/ncab087>
- Cho BW, Choo CO (2019) Geochemical behavior of uranium and radon in groundwater of jurassic Granite Area, Icheon, Middle Korea. Water (Basel) 11:1278. <https://doi.org/10.3390/w11061278>
- Cho B-W, Choo CO, Kim MS et al (2015) Spatial relationships between radon and topographical, geological, and geochemical factors and their relevance in all of South Korea. Environ Earth Sci 74:5155–5168. <https://doi.org/10.1007/s12665-015-4526-0>
- Cho BW, Kim HK, Kim MS et al (2019) Radon concentrations in the community groundwater system of South Korea. Environ Monit Assess 191:189. <https://doi.org/10.1007/s10661-019-7301-y>
- Chungcheongbuk-do Government (2024) Administrative districts. In: <https://www.chungbuk.go.kr/wwweng/index.do>
- Ciotoli G, Voltaggio M, Tuccimei P et al (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. J Environ Radioact 166:355–375. <https://doi.org/10.1016/j.jenvrad.2016.05.010>
- Coletti C, Ciotoli G, Benà E et al (2022) The assessment of local geological factors for the construction of a Geogenic Radon Potential map using regression kriging. A case study from the Euganean Hills volcanic district (Italy). Sci Total Environ 808:152064. <https://doi.org/10.1016/j.scitotenv.2021.152064>
- Cosma C, Cucoş-Dinu A, Papp B et al (2013a) Soil and building material as main sources of indoor radon in Băița-Ștei radon prone area (Romania). J Environ Radioact 116:174–179. <https://doi.org/10.1016/j.jenvrad.2012.09.006>
- Cosma C, Cucoş-Dinu A, Papp B et al (2013b) Soil and building material as main sources of indoor radon in Băița-Ștei radon prone area (Romania). J Environ Radioact 116:174–179. <https://doi.org/10.1016/j.jenvrad.2012.09.006>
- Das S, Bora PK, Das R (2022) Estimation of slope length gradient (LS) factor for the sub-watershed areas of Juri River in Tripura. Model Earth Syst Environ 8:1171–1177. <https://doi.org/10.1007/s40808-021-01153-0>
- Dicu T, Cucoş A, Botoş M et al (2023) Exploring statistical and machine learning techniques to identify factors influencing indoor radon concentration. Sci Total Environ 905:167024. <https://doi.org/10.1016/j.scitotenv.2023.167024>
- Dodangeh E, Panahi M, Rezaie F et al (2020) Novel hybrid intelligence models for flood-susceptibility prediction: Meta optimization of the GMDH and SVR models with the genetic algorithm and harmony search. J Hydrol (Amst) 590:125423. <https://doi.org/10.1016/j.jhydrol.2020.125423>
- Doering C (2019) A new background subtraction method for assessing public Radiation exposure due to Radon Transport from a Uranium Mine. Radiat Prot Dosimetry 186:530–535. <https://doi.org/10.1093/rpd/ncz141>
- Elbagoury BM, Vladareanu L, Vlăduceanu V et al (2023) A hybrid stacked CNN and residual feedback GMDH-LSTM Deep Learning Model for Stroke Prediction Applied on Mobile AI Smart Hospital platform. Sensors 23:3500. <https://doi.org/10.3390/s23073500>
- Farlow SJ (1984) Self-organizing method in modeling: GMDH type algorithm. Marcel Dekker Inc., New York
- Friedmann H, Gröller J (2010) An approach to improve the Austrian radon potential map by bayesian statistics. J Environ Radioact 101:804–808. <https://doi.org/10.1016/j.jenvrad.2009.11.008>
- Gravriliev S, Petrova T, Miklyev P (2022) Factors influencing radon transport in the soils of Moscow. Environ Sci Pollut Res 29:88606–88617. <https://doi.org/10.1007/s11356-022-21919-y>
- Graves A (2012) Long short-term memory. In: Supervised sequence labelling with recurrent neural networks. Studies in computational

- intelligence, vol 385. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-24797-2_4
- Haneberg WC, Wiggins A, Curl DC et al (2020) A geologically based indoor-radon potential map of Kentucky. *GeoHealth*. <https://doi.org/10.1029/2020GH000263>
- Hashemi S, Shin I, Kim S-O, Lee W-C, Lee S-W, Jeong DH, Kim M, Kim H, Yang J (2024) Health risk assessment of uranium intake from private residential drinking groundwater facilities based on geological characteristics across the Republic of Korea. *Sci Total Environ* 913:169252. <https://doi.org/10.1016/j.scitotenv.2023.169252>
- Huang F, Cao Z, Guo J et al (2020) Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. *Catena (Amst)* 191:104580. <https://doi.org/10.1016/j.catena.2020.104580>
- Hwang J, Kim T, Kim H et al (2017) Predictive radon potential mapping in groundwater: a case study in Yongin, Korea. *Environ Earth Sci* 76:515. <https://doi.org/10.1007/s12665-017-6838-8>
- Hwang J, Lee J-Y, Viaroli S (2024) Occurrence and geochemistry of altered radioactive accessory minerals as sources of radionuclide in mesozoic granite aquifers (Korea). *Chemosphere* 359:142326. <https://doi.org/10.1016/j.chemosphere.2024.142326>
- Ivakhnenco AG (1970) Heuristic self-organization in problems of engineering cybernetics. *Automatica* 6:207–219. [https://doi.org/10.1016/0005-1098\(70\)90092-0](https://doi.org/10.1016/0005-1098(70)90092-0)
- Ivakhnenco AG (1978) The group method of data handling in long-range forecasting. *Technol Forecast Soc Change* 12:213–227. [https://doi.org/10.1016/0040-1625\(78\)90057-4](https://doi.org/10.1016/0040-1625(78)90057-4)
- Ivakhnenco AG, Ivakhnenko GA (2000) Problems of further development of the group method of data handling algorithms. Part I, mathematical theory of pattern recognition. *Pattern Recognition Image Anal* 10(2):187–194
- Ivanova K, Stojanovska Z, Tsanova M, Kunovska B (2017) Building-specific factors affecting indoor radon concentration variations in different regions in Bulgaria. *Air Qual Atmos Health* 10:1151–1161. <https://doi.org/10.1007/s11869-017-0501-0>
- Kong W, Dong ZY, Jia Y et al (2019) Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Trans Smart Grid* 10:841–851. <https://doi.org/10.1109/TSG.2017.2753802>
- L'Heureux A, Grolinger K, Elyamany HF, Capretz MAM (2017) Machine learning with big data: challenges and approaches. *IEEE Access* 5:7776–7797. <https://doi.org/10.1109/ACCESS.2017.2696365>
- Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86:2278–2324. <https://doi.org/10.1109/5.726791>
- Lee S, Talib JA (2005) Probabilistic landslide susceptibility and factor effect analysis. *Environ Geol* 47:982–990. <https://doi.org/10.1007/s00254-005-1228-z>
- Lee H, Lee J, Yoon S, Lee C (2021) 222Rn exhalation rates from some Granite and Marble used in Korea: preliminary study. *Atmosphere* 12(8):1057. <https://doi.org/10.3390/atmos12081057>
- Liu W-J, Liu C-Q, Zhao Z-Q et al (2013) Elemental and strontium isotopic geochemistry of the soil profiles developed on limestone and sandstone in karstic terrain on Yunnan-Guizhou Plateau, China: implications for chemical weathering and parent materials. *J Asian Earth Sci* 67–68:138–152. <https://doi.org/10.1016/j.jseas.2013.02.017>
- Lorenzo-González M, Torres-Durán M, Barbosa-Lorenzo R et al (2019) Radon exposure: a major cause of lung cancer. *Expert Rev Respir Med* 13:839–850. <https://doi.org/10.1080/17476348.2019.1645599>
- Ma Z, Mei G, Piccialli F (2021) Machine learning for landslides prevention: a survey. *Neural Comput Appl* 33:10881–10907. <https://doi.org/10.1007/s00521-020-05529-8>
- Megahed HA, Abdo AM, AbdelRahman MAE et al (2023) Frequency ratio Model as Tools for Flood susceptibility mapping in urbanized areas: a Case Study from Egypt. *Appl Sci* 13:9445. <https://doi.org/10.3390/app13169445>
- Mezquita L, Benito A, Ruano-Ravina A et al (2019) Indoor radon in EGFR- and BRAF-mutated and ALK-rearranged non-small-cell lung cancer patients. *Clin Lung Cancer* 20:305–312e3. <https://doi.org/10.1016/j.cllc.2019.04.009>
- Mukharesh L, Greco KF, Banzon T et al (2022) Environmental radon and childhood asthma. *Pediatr Pulmonol* 57:3165–3168. <https://doi.org/10.1002/ppul.26143>
- Nunes LJR, Curado A, Lopes SI (2023) The relationship between Radon and Geology: sources, transport and indoor Accumulation. *Appl Sci* 13:7460. <https://doi.org/10.3390/app13137460>
- Panahi M, Sadhasivam N, Pourghasemi HR et al (2020) Spatial prediction of groundwater potential mapping based on convolutional neural network (CNN) and support vector regression (SVR). *J Hydrol (Amst)* 588:125033. <https://doi.org/10.1016/j.jhydrol.2020.125033>
- Panahi M, Yariyan P, Rezaie F et al (2022) Spatial modeling of radon potential mapping using deep learning algorithms. *Geocarto Int* 37:9560–9582. <https://doi.org/10.1080/10106049.2021.2022011>
- Park TH, Kang DR, Park SH et al (2018) Indoor radon concentration in Korea residential environments. *Environ Sci Pollut Res* 25:12678–12685. <https://doi.org/10.1007/s11356-018-1531-3>
- Park JH, Lee CM, Kang DR (2019a) A deterministic model for estimating indoor radon concentrations in South Korea. *Int J Environ Res Public Health* 16:3424. <https://doi.org/10.3390/ijerph16183424>
- Park N-W, Kim Y, Chang B-U, Kwak G-H (2019b) County-level indoor radon concentration mapping and uncertainty assessment in South Korea using geostatistical simulation and environmental factors. *J Environ Radioact* 208–209:106044. <https://doi.org/10.1016/j.jenvrad.2019.106044>
- Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS (2008) Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 27:157–172. <https://doi.org/10.1002/sim.2929>
- Petermann E, Meyer H, Nussbaum M, Bossew P (2021) Mapping the geogenic radon potential for Germany by machine learning. *Sci Total Environ* 754:142291. <https://doi.org/10.1016/j.scitotenv.2020.142291>
- Plechacek A, Scott SR, Gotkowitz MB, Ginder-Vogel M (2022) Strontium and radium occurrence at the boundary of a confined aquifer system. *Appl Geochem* 142:105332. <https://doi.org/10.1016/j.apgeochem.2022.105332>
- Qiang Z, Yao Y, Li Z et al (2023) Risk Assessment of Lung Cancer caused by indoor Radon exposure in China during 2006–2016: a multicity, longitudinal analysis. *Indoor Air* 2023:1–11. <https://doi.org/10.1155/2023/6943333>
- Rezaie F, Kim SW, Alizadeh M et al (2021) Application of machine learning algorithms for Geogenic Radon potential mapping in Danyang-Gun, South Korea. *Front Environ Sci* 9. <https://doi.org/10.3389/fenvs.2021.753028>
- Rezaie F, Panahi M, Lee J et al (2022) Radon potential mapping in Jangsu-gu, South Korea using probabilistic and deep learning algorithms. *Environ Pollut* 292:118385. <https://doi.org/10.1016/j.envpol.2021.118385>
- Rezaie F, Panahi M, Bateni SM et al (2023) Spatial modeling of geogenic indoor radon distribution in Chungcheongnam-do, South Korea using enhanced machine learning algorithms. *Environ Int* 171:107724. <https://doi.org/10.1016/j.envint.2022.107724>
- Riudavets M, Garcia de Herreros M, Besse B, Mezquita L (2022) Radon and lung cancer: current trends and future perspectives. *Cancers (Basel)* 14:3142. <https://doi.org/10.3390/cancers14133142>

- Selamat FE, Tagusari J, Matsui T (2021) Mapping of transportation noise-induced health risks as an alternative tool for risk communication with local residents. *Appl Acoust* 178:107987. <https://doi.org/10.1016/j.apacoust.2021.107987>
- Szabó KZ, Jordan G, Horváth Á, Szabó C (2014) Mapping the geo-genic radon potential: methodology and spatial analysis for central Hungary. *J Environ Radioact* 129:107–120. <https://doi.org/10.1016/j.jenvrad.2013.12.009>
- Thi Ngo PT, Panahi M, Khosravi K et al (2021) Evaluation of deep learning algorithms for national scale landslide susceptibility mapping of Iran. *Geosci Front* 12:505–519. <https://doi.org/10.1016/j.gsf.2020.06.013>
- Wagner JS, Gohm A, Rotach MW (2015) The impact of valley geometry on daytime thermally driven flows and vertical transport processes. *Q J R Meteorol Soc* 141:1780–1794. <https://doi.org/10.1002/qj.2481>
- Wen Y, Li W, Yang Z et al (2020) Enrichment and source identification of cd and other heavy metals in soils with high geochemical background in the karst region, Southwestern China. *Chemosphere* 245:125620. <https://doi.org/10.1016/j.chemosphere.2019.125620>
- Yamashita R, Nishio M, Do RKG, Togashi K (2018) Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 9:611–629. <https://doi.org/10.1007/s13244-018-0639-9>
- Yoon JY, Lee J-D, Joo SW, Kang DR (2016) Indoor radon exposure and lung cancer: a review of ecological studies. *Ann Occup Environ Med* 28:15. <https://doi.org/10.1186/s40557-016-0098-z>
- Zhou L, Pan S, Wang J, Vasilakos AV (2017) Machine learning on big data: opportunities and challenges. *Neurocomputing* 237:350–361. <https://doi.org/10.1016/j.neucom.2017.01.026>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.