OPEN ACCESS
Journal of Multidisciplinary Applied Natural Science Vol. xx No. xx (xxxx) https://doi.org/10.47352/jmans.2774-3047.356
Check for updates
Research Article

Topography and Soil Indices Predict Environmental *Burkholderia pseudomallei* in Paddy Fields using Interpretable Machine Learning

Wacharapong Saengnill, Jutharat Jittimanee, Suwaporn Dandee, Jaruwan Wongbutdee*, and Pongthep Thongsang

Received : August 14, 2025	Revised : November 15, 2025	Accepted : December 7, 2025	Online : February 15, 2026

Abstract
To understand the environmental distribution of *Burkholderia pseudomallei*, it is essential to study the causative agent of melioidosis for effective public health risk assessment. This study integrates geostatistical analysis and machine learning to predict the spatial distribution of *Burkholderia pseudomallei* in paddy soils of northeastern Thailand. A total of 92 soil samples were collected and analysed using culture-based methods. Environmental covariates were derived from remote sensing and topographic data, including land surface temperature, normalised difference salinity index, bare soil index, digital elevation model, distance to water, slope, aspect, and soil drainage. Indicator kriging was used to generate a spatial probability map of *Burkholderia pseudomallei* presence. An extreme gradient boosting machine learning model was applied to predict bacterial presence. Of the 92 soil samples analysed, 40.22% tested positive for *Burkholderia pseudomallei*. Indicator kriging demonstrated clustered distributions primarily in low-lying, poorly drained areas. The extreme gradient boosting model achieved an F1-score of 0.70 on the testing dataset. Shapley additive explanations analysis highlighted the digital elevation model, bare soil index, and slope as the most influential predictors. The resulting risk maps provide valuable tools for identifying high-risk areas, supporting targeted surveillance and public health interventions in melioidosis-endemic regions.

Keywords: environmental health, extreme gradient boosting (XGBoost), health geography, melioidosis, risk mapping, spatial epidemiology

1. INTRODUCTION

Understanding the distribution of pathogens in soil environments is critical for safeguarding public health and improving environmental management practices [1][2]. Soil-borne pathogens pose significant challenges owing to their complex interactions with environmental factors, the difficulty of direct detection, and their potential to cause widespread outbreaks. Accurate identification of high-risk areas can support targeted interventions, reduce human exposure, and enhance resource allocation for disease prevention and control. For instance, high-risk clusters have been identified for the incidence of leptospirosis [3], melioidosis [4], and soil-transmitted helminth infections [5]-[7]. In addition, Hatta et al. [8] identified a hotspot for leptospirosis and enteric fever in Malaysia, suggesting the need for multi-disease intervention strategies. *Burkholderia pseudomallei*, the causative agent of melioidosis, exemplifies these challenges. Responsible for approximately 165,000 infections and 89,000 deaths annually, primarily in tropical regions [9], this bacterium thrives in soil and water, particularly in agricultural environments. Paddy fields, with their wet and nutrient-rich conditions, provide an ideal habitat for *B. pseudomallei*, which poses a significant threat to both human health and agricultural productivity. Despite its importance to public health, the spatial distribution and ecological drivers of *B. pseudomallei* remain poorly understood, which limits the effectiveness of mitigation strategies [10].

Environmental factors affect the survival and are associated with the presence of *B. pseudomallei* bacteria. The bacterium can survive across a range of temperatures, although refrigeration can induce a proportion of the population into a viable but non-culturable state [11]. In particular, *B. pseudomallei* exhibits the highest growth rate at 37 °C, with modest reductions observed at 30, 40, and 42 °C [12]. Similarly, soil salinity affects microbial communities and has been linked to the increased persistence of *B. pseudomallei* in specific soil types.

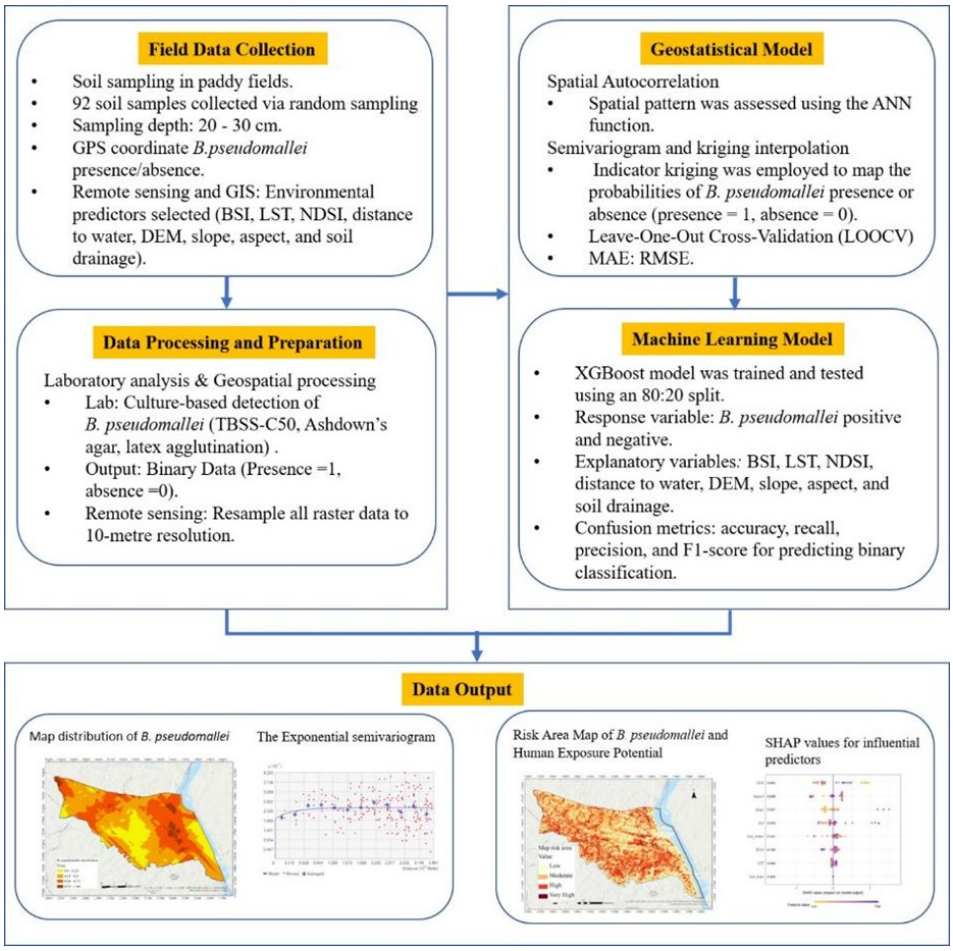

Figure 1. Framework illustrating the integrated workflow for predicting the spatial distribution of *B. pseudomallei*. The process begins with (1) data acquisition and preparation, followed by the (2) parallel modelling phase, and ends with (3) data output synthesising results from both models to generate spatial risk maps. GIS, geographical information system; BSI, bare soil index; LST, land surface temperature; NDSI, normalised difference salinity index; DEM, digital elevation model; ANN, average nearest neighbour; LOOCV, leave-one-out-cross-validation; MAE, mean absolute error; RMSE, root mean square error; XGBoost, extreme gradient boosting; SHAP, SHapley Additive exPlanations.

The ability of *B. pseudomallei* to thrive in saline conditions poses a significant public health risk, particularly in endemic regions. These adaptations may lead to an increased incidence of melioidosis, emphasising the need to monitor and control environmental salinity levels to mitigate outbreaks [13]. Conversely, while salinity enhances the persistence and virulence of *B. pseudomallei*, other factors, such as iron availability, also play a crucial role in its growth dynamics, indicating a complex interplay of environmental conditions that influence bacterial behaviour [14]. Soil with a water content of less than 10% led to the death of *B. pseudomallei* within 70 days, whereas soil with a water content of more than 40% maintained bacterial life for 726 days [15]. In addition, in soil media with a water content below 15%, *B. pseudomallei* did not grow until after 60 days of incubation, suggesting that soil water content is an important factor in determining its growth rate [16].

Remote sensing (RS) technologies further enhance these capabilities by providing high-resolution environmental datasets that are critical for modelling pathogen habitats. Key RS-derived variables, such as land surface temperature (LST), normalized difference vegetation index, soil moisture, and soil salinity, play pivotal roles in shaping the ecological niche of soil bacteria. There is a significant correlation between remotely sensed environmental factors and outbreaks of leptospirosis, highlighting the potential of these technologies for disease surveillance and early

Table 1. Summary of remote sensing and topographic data sources and parameters.

Parameter	Data Source	Acquisition Date	Spatial Resolution (m)
Land surface temperature (LST)	Landsat 8 OLI/TIRS	April 13 2024	30
Normalised difference salinity index (NDSI)	Sentinel-2A Level-2	April 27 2024	10 / 20
Bare soil index (BSI)	Sentinel-2A Level-2	April 27 2024	10 / 20
Distance to water bodies	Sentinel-2A Level-2 (MNDWI)	April 27 2024	10
Digital elevation model (DEM), Slope, aspect	ESA Copernicus/ OpenTopography	April 27 2024	30

Note: MNDWI, modified normalised difference water index; OLI, Operational Land Imager; TIRS, Thermal Infrared Sensor.

warning systems [17][18]. In addition, LST significantly influences the distribution of soil-transmitted helminths in Brazil and Southeast Asia [19][20]. The relationship between *B. pseudomallei* presence in the soil and environmental factors derived from satellite imagery remains unclear, resulting in limited knowledge and data on spatial factors affecting the persistence of the organism in soil. Therefore, it is necessary to identify the physical environmental factors that affect the persistence or distribution of *B. pseudomallei* in soil.

Recent advancements in artificial intelligence (AI) and machine learning (ML) have expanded the potential for predictive disease mapping in both agricultural and environmental contexts. AI-driven approaches have been successfully applied to disease detection in crops such as corn, apples, and rice using multispectral and hyperspectral data [21]-[23]. These studies illustrate how AI-based classification can detect subtle spectral variations associated with biological or environmental stress. However, their application to soil-borne human pathogens remains limited, particularly for *B. pseudomallei*, whose ecological complexity requires interpretable and data-efficient models. Machine learning algorithms such as Random Forest, Light Gradient Boosting Machine (LightGBM), and Extreme Gradient Boosting (XGBoost) have shown excellent capability in modelling non-linear relationships and identifying variable importance in environmental data [24]-[26]. Recent studies have also demonstrated the value of hybrid and interpretable ML frameworks for environmental and biomedical prediction problems, highlighting their adaptability to complex datasets [27]. XGBoost, in particular, combines high predictive power with interpretability through SHapley Additive exPlanations (SHAP) values, which quantify each variable's contribution to model predictions. This feature makes XGBoost a robust tool for elucidating the influence of topographic and soil indices on *B. pseudomallei* distribution.

This study hypothesizes that the spatial distribution of *B. pseudomallei* in paddy soils is significantly influenced by topographic and soil surface indices derived from remote sensing data. The objectives of this study are: (i) to generate

spatial distribution maps of *B. pseudomallei* presence in paddy soils; (ii) to develop and evaluate interpretable ML models that incorporate RS-derived environmental indices to produce spatial risk maps of *B. pseudomallei* distribution; and (iii) to identify and quantify the key environmental and topographic predictors associated with *B. pseudomallei* presence in paddy soils. The findings of this study provide new insights into the ecological determinants of *B. pseudomallei*, demonstrate the utility of interpretable ML for environmental pathogen mapping, and contribute to data-driven strategies for melioidosis surveillance and prevention in vulnerable agricultural regions.

2. MATERIALS AND METHODS

2.1. Study Area

The study was conducted in Natan Subdistrict, Natan District, Ubon Ratchathani Province, Thailand. Situated in the northeastern region of the country, this area forms part of the larger Kong River Basin, a critical agricultural hub dominated by rice cultivation. The study area was centred at geographic coordinates 15°54'18"N latitude and 105°16'36"E longitude, covering approximately 56.15 km^2 of predominantly rural landscapes. Most of the land is dedicated to rice paddy cultivation, while the remainder comprises forested areas, small-scale vegetable farms, and residential zones. Natan Subdistrict experiences a tropical monsoon climate with three distinct seasons: a rainy season from May to October, characterised by heavy rainfall and high humidity; a cool season from November to February, with milder temperatures; and a hot season from March to April, marked by high temperatures and drier conditions. The melioidosis cases reported between 2018 and 2022 were 10, 10, 9, 9, and 13, respectively, which were obtained from the Ubon Ratchathani Provincial Public Health Office, Ministry of Public Health, Thailand [28].

2.2. Ethics Approval

Biosafety protocol was approved by The Ubon Ratchathani University- Institutional Biosafety Committee-IBC (ID# IBC 01/2567). Human protocol was approved by the Ubon Ratchathani University-Human Ethics Committee (ID# UBU-REC 171/2565).

2.3. Conceptual Framework

This study outlined a comprehensive framework for predicting the spatial distribution of *B. pseudomallei* by integrating field-collected soil data with RS and topographic information (Figure 1). The methodology began with data acquisition, in which 92 soil samples were analysed for bacterial presence and linked to key environmental predictors. The core of the study involved a dual analytical approach, simultaneously employing a geostatistical model (indicator kriging) to assess spatial dependency and an ML model (XGBoost) to identify complex environmental relationships. In the final phase, the outputs from both models were synthesised to generate actionable risk maps that identified high-risk zones and the most influential predictors, thereby providing a robust tool for public health surveillance and intervention strategies against melioidosis.

2.4. Soil Observations

In total, 92 soil samples were collected based on a statistically determined sample size calculated using an appropriate probability formula (Equation (1)). Sampling was conducted on April 26–27 2024, employing a random sampling strategy within the paddy fields. Soil samples were systematically collected from the topsoil layer at depths ranging from 20 to 30 cm. At each sampling point, approximately 100 g of soil was collected using a sterilised shovel and was thoroughly disinfected with 70% alcohol before sampling to prevent cross-contamination. The collected soil samples were stored in durable three-layered plastic bags designed to minimise external contamination. Immediately after collection, the samples were placed in a cool insulated box to maintain their integrity during transportation. All samples were transported to the laboratory on the same day for subsequent analysis. Geographic coordinates for each sampling location were recorded using a Garmin eTrex 30x GPS device with the World Geodetic System (WGS) 1984 datum and Universal Transverse Mercator (UTM) Zone 48N projection, ensuring the precise spatial referencing necessary for accurate geospatial analyses;

$$n = \frac{P(1-P)Z^2_{\alpha/2}}{e^2} \quad (1)$$

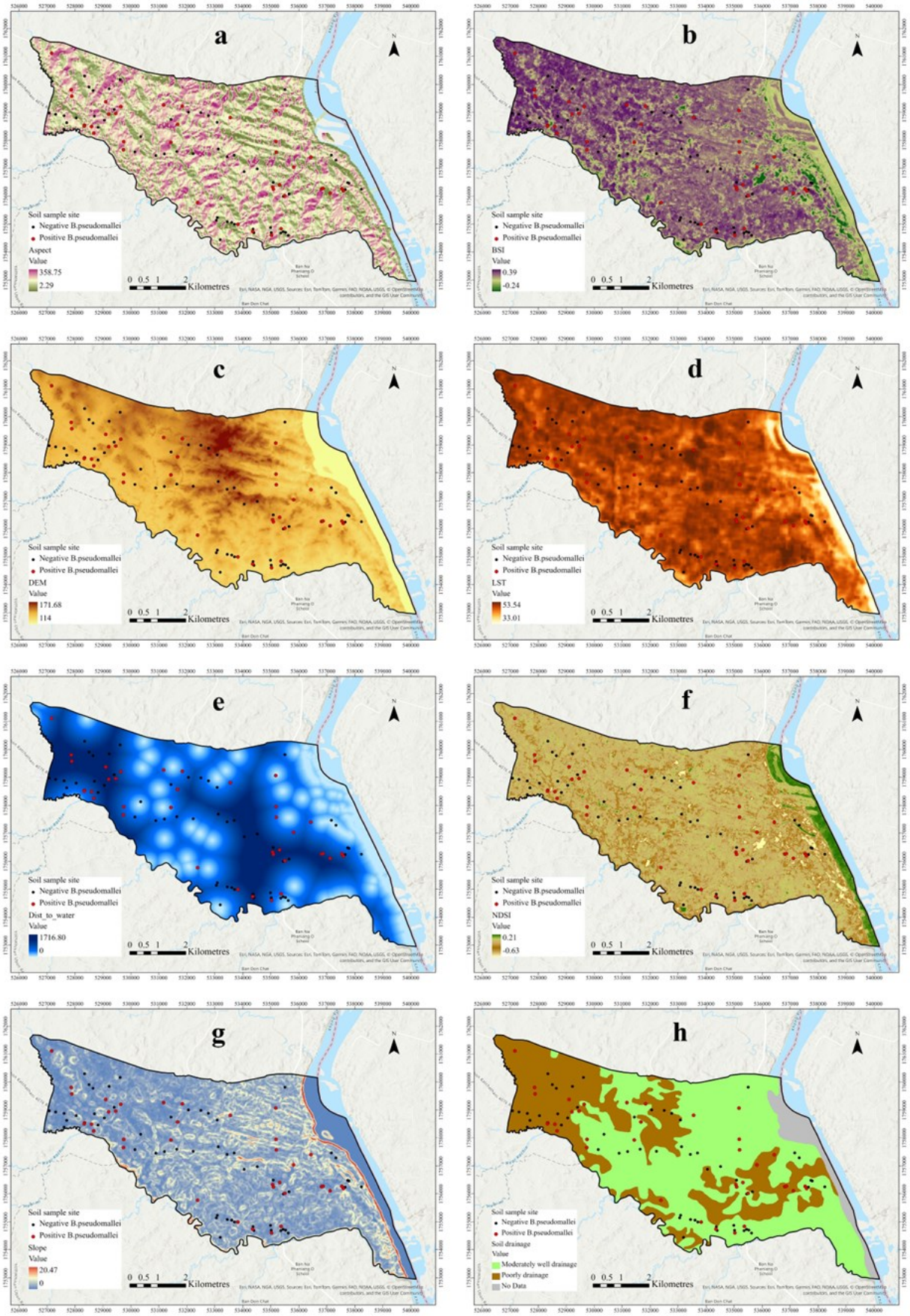

Figure 2. Maps of environmental factors: (a) BSI, (b) LST, (c) NDSI, (d) distance to water, (e) DEM, (f) slope, (g) aspect, and (h) soil drainage.

where p is the proportion of *B. pseudomallei* in paddy soils (p = 0.39) [29], e is the margin of error for estimating p (acceptable error = 0.1), and z is the confidence level at 95% (1.96).

2.5. Environmental Parameters

This study applied advanced RS techniques to identify key environmental variables that significantly influence the spatial distribution of *B. pseudomallei* (Table 1). Environmental parameters were selected based on their known ecological relevance to soil-borne pathogens and processed using rigorous geospatial methods, following established procedures from previous studies [15][16][30]-[32]. The LST data were obtained from Landsat 8 Operational Land Imager/Thermal Infrared Sensor imagery captured on April 13 2024, with a spatial resolution of 30 m. The processing involved a JavaScript-based workflow within the Google Earth Engine platform, including image filtering for cloud-free conditions, radiometric corrections, and conversion of digital numbers to temperature values. LST is an essential indicator that provides insights into soil surface heat, which significantly affects *B. pseudomallei*'s survival and proliferation.

Sentinel-2A Level-2 imagery from the European Space Agency's (ESA's) Copernicus Open Access Hub was acquired on 27 April 2024. The normalised difference salinity index (NDSI), used to assess soil and water salinity, was calculated using shortwave infrared (SWIR) reflectance bands; higher NDSI values denoted elevated salinity. The bare soil index (BSI) was derived from spectral bands, including SWIR, red, near-infrared, and blue, effectively distinguishing bare soils from other land cover types, with a higher BSI indicating greater bare soil coverage. The distance to water bodies was calculated by generating a water mask from Sentinel-2 imagery using the modified normalised difference water index (MNDWI), computed as MNDWI = (Green – SWIR1)/(Green + SWIR1), where Green corresponds to Band 3 (10 m resolution) and SWIR1 corresponds to Band 11 (20 m resolution). SWIR1 was resampled to match the resolution of Band 3 via bilinear interpolation. A binary water mask was produced using a visually determined threshold (MNDWI > 0 or NDWI > 0.2). Subsequently, small pixel clusters were filtered using the majority filter tool in ArcGIS Pro version 3.5.2. The Euclidean distance was calculated from each non-water pixel to the nearest water pixel, providing distance measurements (in metres) between the sampling sites and water bodies.

Topographic variables, including a digital elevation model (DEM), slope, and aspect, were obtained from the European Space Agency (ESA) via the Copernicus Open Access Hub and the OpenTopography portal. The dataset provided high-resolution elevation data (30 m), ensuring precise terrain characterisation. The slope was calculated in degrees to assess terrain steepness and its potential impact on water flow and soil stability, which are factors influencing *B. pseudomallei* habitats. The aspect was computed to identify the directional orientation of slopes, which is relevant for analysing microclimatic factors, such as solar exposure and soil moisture retention, both critical to pathogen survival. Soil drainage, reflecting the rate and extent of water movement through the soil, was derived from digital soil mapping at a scale of 1:50,000, referenced to UTM Zone 48 and the WGS-1984 datum. Drainage conditions were

Table 2. Configuration parameters for extreme gradient boosting.

Parameter	Description	XGBoost
Nrounds	Number of boosting iterations	150
Max depth	Maximum tree depth	4
Eta	Learning rate	0.01
Gamma	Minimum loss reduction	0
Colsample by tree	Subsample ratio of columns	1
Min child weight	Minimum sum of instance weight (Hessian)	1
Subsample	Subsample ratio of the training set	1

Table 3. Number and percentage of soil samples positive for *B. pseudomallei* detection.

***B. pseudomallei* detection**	**Number (N)**	**Percentage (%)**
B. pseudomallei positive	37	40.22
B. pseudomallei negative	55	59.78
Total	**92**	**100.00**

categorised into poorly drained and moderately well -drained classes.

2.6. Methodology

2.6.1 B. pseudomallei Detection

Ten grams of soil from each sample were aseptically transferred into sterile 50 mL universal tubes, and 10 mL of threonine-basal salt solution containing 50 mg/L of colistin (TBSS-C50) was added. Each mixture was thoroughly homogenised using a vortex mixer and incubated aerobically at 40 °C for 48 h. Post-incubation, a 10 µL aliquot from the upper enrichment layer was streaked onto Ashdown's selective agar to isolate colonies. The plates were incubated at 40 °C and visually examined daily for up to 7 days. Suspected colonies of *B. pseudomallei* were identified by their distinctive purple, flat, dry, and wrinkled appearance, alongside additional morphotypes described by Chantratita et al. [33] These colonies were subjected to a preliminary oxidase test, and positive colonies were confirmed using a latex agglutination test, as described by Wuthiekanun et al. [34] and Anuntagool et al. [35].

2.6.2. Data Preparation for Spatial Statistics

The locations of the soil samples were georeferenced and linked to the detection data for *B. pseudomallei*. The detection data were classified into two binary variables: positive (1) and negative (0). Spatial autocorrelation analysis was conducted to evaluate spatial patterns (clustered, random, or dispersed) and lag distances using the average nearest neighbour (ANN) method. Subsequently, indicator kriging was applied to generate a probabilistic risk map of *B. pseudomallei* distribution, relying on spatial dependency characterised by a semivariogram model. The resulting probability map was categorised into 4 risk classes. Predictive modelling was then performed using the XGBoost algorithm, incorporating environmental parameters as predictors of the presence or absence of *B. pseudomallei* in paddy soils. All raster data from the satellite images were resampled to a consistent spatial resolution of 10 m, and the raster values at the 92 sample locations were extracted. Collinearity among environmental variables was assessed via Pearson's correlation analysis, and variables exhibiting correlation coefficients exceeding 0.7 were excluded to mitigate multicollinearity. Ultimately, eight environmental parameters—BSI, LST, NDSI, distance to water, DEM, slope, aspect, and soil drainage—were used as model inputs (Figure 2).

2.6.3. Geostatistical Methods

The spatial distribution of *B. pseudomallei* was analysed using geostatistical methods to evaluate spatial autocorrelation and identify distribution patterns. The analysis determined whether the presence of *B. pseudomallei* was clustered, dispersed, or random, and quantified the spatial relationships between sampling locations. This process involved calculating distances between sampling points and analysing their spatial correlation using semivariogram equations and geospatial tools. The spatial autocorrelation and geostatistics model were analyzed using ArcGIS Pro version 3.5.2.

2.6.3.1. Spatial Autocorrelation

The selection of lag size significantly affects the empirical semivariogram. For example, if the lag size is excessively large, short-range autocorrelation may be masked. Conversely, if the lag size is too small, many bins may be empty, and the sample sizes within bins will be insufficient to obtain representative averages. To determine spatial autocorrelation, the relationship between distance and spatial patterns was assessed using the ANN function. This distance method will include at least a few pairs of points within a unit area, based on the

Euclidean distance metric. Therefore, the results showed a significant clustered spatial relationship of *B. pseudomallei* (p-value < 0.001), with an ANN distance of 288.41 m.

2.6.3.2. Semivariogram and Kriging Interpolation

Geostatistics provides a robust framework for analysing spatial dependencies among sampled observations, leveraging spatial autocorrelation to estimate values at unmeasured locations. By interpolating continuous surfaces, this method enables the prediction of the spatial probability of *B. pseudomallei* presence across the study area. In this study, the distribution of *B. pseudomallei* in soil was modelled using semivariogram analysis and kriging interpolation. The semivariogram quantifies spatial variability as a function of the separation distance (lag distance) between sample points. Mathematically, it measures half the mean squared difference between paired samples separated by distance h (Equation (2)). The semivariogram function is characterised by three key parameters: 1) Sill: Represents the total variance at which spatial autocorrelation stabilises. 2) Nugget: Captures small-scale variability or measurement errors occurring at minimal separation distances. 3) Range: Defines the maximum distance over which spatial dependence exists; beyond this, samples are considered uncorrelated. Mathematically, the semivariogram is expressed as Equation (2);

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (2)$$

Where $N(h)$ is the number of sample pairs separated by distance h, and $z(x_i)$ and $z(x_i + h)$ are the observed values at locations x_i and $(x_i + h)$, respectively.

Empirical semivariograms were generated to assess the spatial relationships between sampling locations. Various theoretical models, including spherical, exponential, and Gaussian functions, were fitted to the empirical semivariograms. Model selection was based on prediction accuracy metrics, including the mean absolute error (MAE) and root mean square error (RMSE). The exponential model was determined to provide the best fit, optimizing the prediction of B. pseudomallei distribution.

Spatial autocorrelation, as defined by a semivariogram, was used to weigh the relationships between the sampling points. Indicator kriging was employed to map the probabilities of B. pseudomallei presence or absence, treating the detection data as binary (presence = 1; absence = 0). The indicator-kriging model estimates the probabilities of unmeasured locations, thereby producing a continuous probability map (Equation (3)). Values approaching 1 indicated a high

Table 4. Semivariogram model and cross-validation of *B. pseudomallei*. ME, mean error.

Semivariogram	Nugget	Sill	Range (m)	ME	RMSE	RMSSE
Exponential	0.194	0.053	1,112	0.005	0.503	0.988

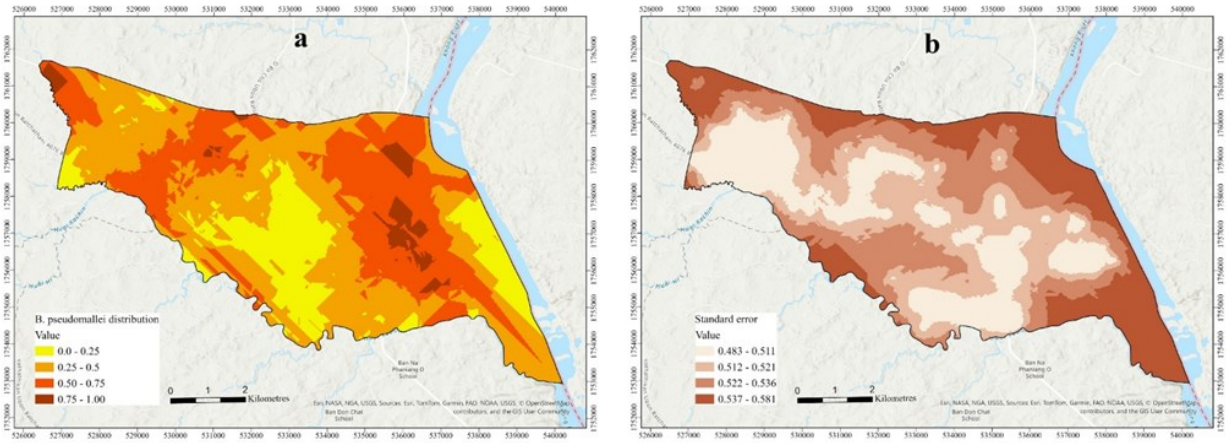

Figure 3. The probability map for public health risk mapping: (a) distribution map of *B. pseudomallei*, (b) standard error map of *B. pseudomallei*.

Table 5. Descriptive analysis of environmental factors. IQR, interquartile range.

Environmental factors	Positive				Negative			
	Median	IQR	Q1	Q3	Median	IQR	Q1	Q3
BSI	0.16	0.05	0.13	0.18	0.16	0.04	0.14	0.18
LST	48.57	1.80	47.65	49.45	48.73	1.86	47.65	49.51
NDSI	-0.15	0.05	-0.18	-0.13	-0.15	0.04	-0.18	-0.14
Distance to water	679.32	534.17	480.68	1014.85	660.5	560.17	406.4	966.57
DEM	135	8.36	131.68	140.04	136.37	12.83	128.86	141.69
Slope	1.43	1.00	1.12	2.11	1.71	1.33	1.19	2.51
Aspect	173.89	144.09	106.17	250.26	168.99	95.34	122.73	218.07
Soil drainage	Number (%)				Number (%)			
1. Moderately well drained	20 (21.74)				32 (34.78)			
2. Poorly drained	17 (18.48)				23 (25.00)			

probability of B. pseudomallei presence, whereas values closer to 0 indicated its absence. The resulting probability map delineated high-risk areas for B. pseudomallei outbreaks, facilitating the identification of vulnerable zones. The indicator kriging Equation (3).

$$I(x) = \begin{cases} 1, & if\ z(x) \geq z_k \\ 0, & if\ z(x) < z_k \end{cases} \quad (3)$$

2.6.3.3. Validation

To assess the accuracy of the model and the association between predicted and observed values, the *B. pseudomallei* prediction map was validated using a cross-validation approach. Leave-one-out-cross-validation was employed to evaluate the performance of the model. This technique involved using all but one data point to train the prediction model, while reserving the excluded point for testing. This process was repeated for each data point in the dataset to ensure that every observation was used for both training and testing. Model accuracy was evaluated using two error metrics: (1) MAE represents the average of the absolute differences between observed and predicted values, providing a direct measure of error magnitude regardless of direction and (2) RMSE reflects the square root of the mean of the squared differences between observed and predicted values, emphasizing larger errors and providing a more sensitive measure of prediction quality.

2.6.3.4. Extreme Gradient Boosting

The XGBoost algorithm was employed as the primary supervised machine learning model to predict the presence or absence of B. pseudomallei. Before training, data imbalance between positive and negative samples was addressed using the Synthetic Minority Oversampling Technique (SMOTE). This preprocessing step generated synthetic samples for the minority (positive) class, ensuring balanced representation and improving model generalizability. The balanced dataset was then split into training (80%) and testing (20%) subsets. Hyperparameters, including the number of estimators (Nrounds), maximum tree depth (Max depth), learning rate (Eta), and subsampling ratio (Table 2), were optimized using a grid search combined with 10-fold cross-validation. The software used for analyzing and visualizing the

Table 6. Performance of the XGBoost model.

XGBoost model	Accuracy	Recall	Precision	F1-score
Training dataset	0.648	0.563	0.600	0.581
Testing dataset	0.666	0.538	0.100	0.700

XGBoost model was R v. 4.4.3 with the xgboost, SHAPforxgboost, and DMwR packages.

The model evaluation stratifies each fold according to a specific parameter set, applying a suite of metrics, including accuracy, recall, precision, and F1-score, to predict the binary classification [36][37]. These metrics were computed individually for each fold and then aggregated to provide a comprehensive assessment of model performance. Accuracy measures the overall correctness of the model and is defined as Equation (4).

$$\text{Accuracy} = \text{TP+TN} / (\text{TP+TN+FP+FN}) \quad (4)$$

Recall, also known as sensitivity, is crucial for detecting as many true-positive cases as possible, and is defined as Equation (5);

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (5)$$

where TP represents true positives, and FN represents false negatives. Precision helps minimise false positives and ensures that detected cases are correct. It is calculated as Equation (6);

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (6)$$

where FP represents false positives. The F1-score, which balances recall and precision, is useful for evaluating performance on imbalanced datasets and is calculated as Equation (7);

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (7)$$

where true negatives (*TN*) are the number of correct negative predictions, true positives (*TP*) are the number of correct positive predictions, false positives (*FP*) are the number of incorrect positive predictions, and false negatives (*FN*) are the number of positive cases that were predicted to be negative.

3. RESULTS AND DISCUSSIONS

3.1. B. pseudomallei Detection

The 92 soil samples analyzed, 37 (40.22%) tested positive for *B. pseudomallei*, while 55 (59.78%) were negative (Table 3). This high detection rate indicates a substantial environmental reservoir of the pathogen, consistent with previous reports from the region [29][38]. Additionally, several studies have found *B. pseudomallei* in paddy soils in Southeast Asia, including Laos [39], Malaysia [40][41], and Myanmar [42]. This distribution underscores the substantial abundance of *B. pseudomallei* in the examined paddy soils. Furthermore, soil sampling was conducted in April during the dry season, a period shown to yield higher *B. pseudomallei* isolation rates than the wet season [43].

3.2. Geostatistical Model

The spatial dependency of the *B. pseudomallei* distribution was assessed using an exponential semivariogram model. Table 4 summarises the semivariogram parameters and cross-validation metrics, indicating a range of 1,112 m. Cross-validation yielded a root mean square standardized error (RMSSE) of 0.988. This distance provides a critical spatial scale for public health, suggesting that if the bacterium is found at one location, the risk is elevated over a radius of more than one kilometre in the surrounding area. This is consistent with previous findings demonstrating the localised clustering of environmental *B. pseudomallei* populations. The indicator kriging map identified high-probability zones predominantly in flat and poorly drained areas, corroborating the results of previous studies on the environmental drivers of melioidosis [44][45]. Additionally, this finding was reinforced by Seng et al. [46], who reported genetic similarities among environmental isolates and clinical cases from patients living nearby. The implications for public health risk mapping of *B. pseudomallei* distribution and the variance map are

presented in Figures 3(a) and 3(b), respectively. The probability map of *B. pseudomallei* infection showed probability values ranging between 0 and 1. Areas where *B. pseudomallei* infection was most likely to occur are shown in brown, while areas farther away where infection was less likely to occur are shown in yellow. Geostatistical modelling enables the identification of high-risk zones and facilitates focused surveillance, land-use planning, and health education. In resource-limited settings, where field-based pathogen surveillance is constrained, geostatistical tools—combined with RS—offer cost-effective strategies to inform targeted interventions and early warning systems [38][47]-[49].

3.3. Machine Learning Model

Descriptive analysis of environmental variables between *B. pseudomallei*-positive and *B. pseudomallei*-negative samples demonstrated minor variations in the median and interquartile ranges (IQRs) (Table 5). Variables such as the BSI, LST, NDSI, distance to water, DEM, slope, and aspect exhibited overlapping distributions between the two classes. In terms of soil drainage, 21.74% of the positive samples were located in moderately well-drained soils and 18.48% in poorly drained soils, compared with 34.78% and 25.00% of the negative samples, respectively.

Table 6 summarises the performance of the XGBoost model on both the training and testing datasets. The SMOTE preprocessing improved the model's sensitivity to minority (positive) samples, reflected in the enhanced F1-score of 0.700. These results emphasize that *B. pseudomallei* occurrence is strongly governed by terrain structure and soil surface conditions. Figure 4(a) presents the relative importance of environmental variables identified by the XGBoost–SMOTE model for predicting the presence of *B. pseudomallei* in paddy soils. The model, despite its moderate overall performance, effectively revealed key predictors that shape the bacterium's spatial distribution. The digital elevation model (DEM) was identified as the most influential variable, followed by slope, BSI, NDSI, and aspect. In contrast, distance to water, LST, and soil drainage contributed less to the prediction. The dominance of DEM, slope, and BSI underscores the importance of terrain configuration and surface exposure in defining suitable habitats for *B. pseudomallei*. DEM likely captures elevation-driven variations in drainage and moisture retention, which directly influence soil saturation—a critical factor for bacterial persistence. Aspect may also reflect microclimatic differences in solar radiation and temperature gradients that shape

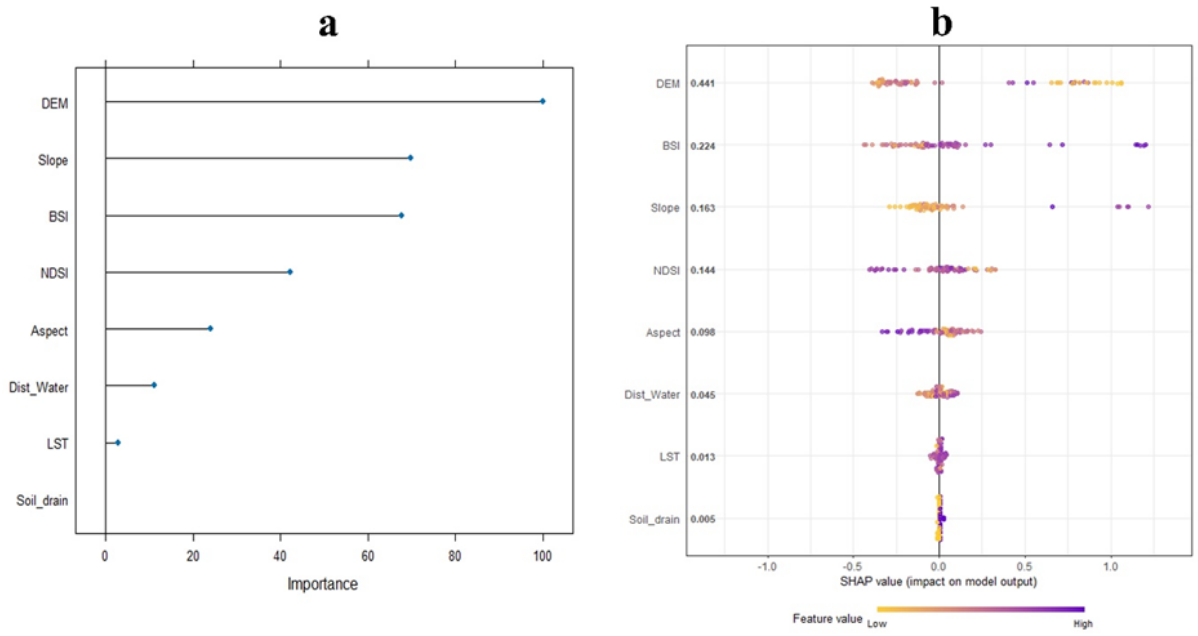

Figure 4. Relative importance and interpretability of environmental predictors in the XGBoost-SMOTE model: (a) relative importance of environmental factors in predicting *B. pseudomallei* presence; (b) SHAP values illustrating the contribution of influential predictors.

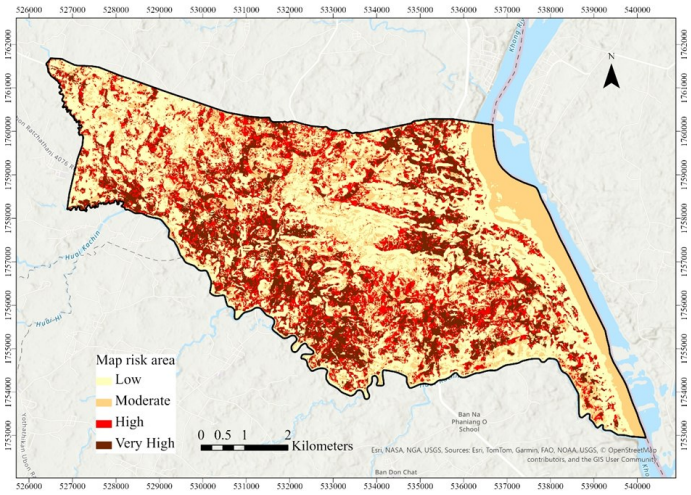

Figure 5. Risk area map of *Burkholderia pseudomallei* distribution and human exposure potential. The probability values were classified into four categories: low (0.00-0.25), moderate (0.25-0.50), high (0.50-0.75), and very high (0.75-1.00).

microbial ecology in tropical environments. These findings are consistent with Pongmala et al. [50], who reported a negative correlation between *B. pseudomallei* occurrence and elevation, indicating a preference for low-lying, water-accumulating areas conducive to bacterial survival.

A major finding of this study was the identification of topographic features, specifically the DEM, BSI, and slope, as the most influential predictors of *B. pseudomallei* presence, a conclusion strongly supported by the SHAP analysis of our XGBoost-SMOTE model (Figure 4 (b)). The primacy of DEM aligns with the established ecological understanding of *B. pseudomallei*. This bacterium thrives in moist environments, and lower-elevation areas typically function as drainage basins, accumulating water and organic matter from surrounding landscapes [32] [42][50][51]. Our indicator kriging map supports this, visually correlating higher-probability zones with low-lying terrain. This finding is consistent with research on other soil-transmitted pathogens, such as leptospirosis, in which elevation and topography are critical determinants of habitat suitability [17]. The additional influence of slope and aspect refines this interpretation, suggesting that micro-topographic variation governs localized soil moisture retention and solar exposure. Such variations create favorable microhabitats that promote bacterial persistence, particularly during the hot–dry season when samples were collected. These findings collectively underscore the critical role of landscape structure in shaping the environmental niche of *B. pseudomallei* and highlight the value of topographic variables as proxies for soil hydrological conditions in spatial risk modeling.

BSI also emerged as a significant predictor. High BSI values correspond to areas with less vegetation. In a paddy-dominated landscape, this could represent post-harvest fields or areas with sparser crop growth. These conditions may lead to increased direct solar radiation and higher soil temperatures, which, contrary to the expected desiccation, might select for hardier, persistent strains of *B. pseudomallei* known to tolerate a range of temperatures [12]. Alternatively, the absence of dense root systems could alter soil structure and water percolation in ways that favour the bacterium.

Interestingly, variables traditionally associated with *B. pseudomallei*, such as LST and soil drainage, exhibited lower predictive importance in our model. This does not necessarily diminish their biological relevance but may reflect the specific conditions of our study. Notably, the LST data, captured at a single point in time, may not fully represent the microclimatic temperature variations experienced by the bacteria within the soil profile. Similarly, while our soil drainage data (poorly vs. moderately well-drained) were informative, their coarse resolution may have been superseded by the fine-scale hydrological effects captured by the high-

resolution DEM and slope data. In addition, in undisturbed environments, *B. pseudomallei* is frequently found in areas with abundant grasses, such as spear grass, as well as in moist soils and near streams [38][52]. The presence of roots and red-brown or red-grey soil also indicates a higher likelihood of the bacterium's presence [52]. It is important to remember that, while the MaxEnt model offers valuable insights, all ecological models have limitations. For instance, they may not account for every variable, and the environmental conditions that influence *B. pseudomallei* distribution can change over time [47]. For a complex, spatially diverse organism such as *B. pseudomallei*, a model that successfully identifies its ecological drivers is invaluable for strategic public health planning. The SHAP analysis was particularly useful, providing a clear, measurable ranking of the importance of each predictor, which aligned well with established ecological principles.

3.4. Public Health Implications and Applications

The risk maps generated in this study represent a significant step towards evidence-based public health in melioidosis-endemic regions (Figure 5). These are not simply academic exercises; they are practical tools that can be used immediately by the Ubon Ratchathani Provincial Public Health Office. These maps can guide targeted environmental surveillance, allowing officials to prioritise sampling efforts in the high-risk zones identified in the southern and central parts of the subdistrict. This increases the efficiency of resource allocation and the likelihood of pathogen detection. Furthermore, these maps can support the development of geographically targeted public awareness campaigns, advising farmers and residents in high-probability areas to adopt protective measures, such as wearing boots during cultivation, and to recognise the early symptoms of melioidosis.

3.5. Limitations and Future Directions

Although robust, this study has several limitations that suggest avenues for future research. First, the soil samples were collected at a single time point during the dry season. The distribution and viability of *B. pseudomallei* are known to fluctuate with seasonal rainfall and soil moisture; therefore, longitudinal studies with sampling across both wet and dry seasons are needed to capture these temporal dynamics. Second, our analysis was based on RS and topographical data. Although powerful, these datasets do not include crucial soil physicochemical properties, such as pH, organic carbon content, or iron availability, which are known to influence *B. pseudomallei* survival [14]. Future studies should integrate field-based soil chemistry analyses with RS data to develop more comprehensive predictive models. Finally, our models predict environmental presence, not necessarily human risk. The crucial next step is to validate these environmental risk maps against epidemiological data on human melioidosis cases to establish a direct link between environmental hotspots and disease incidence.

4. CONCLUSIONS

This study combined remote sensing (RS), geostatistics, and interpretable machine learning (XGBoost with SMOTE preprocessing) to model the spatial distribution of *B. pseudomallei* in paddy soils of northeastern Thailand. The model successfully identified key environmental predictors—digital elevation model (DEM), slope, and bare soil index (BSI)—as the most influential variables governing bacterial presence. By addressing data imbalance with SMOTE, model sensitivity improved, achieving an F1-score of 0.700 and enhancing detection of minority (positive) samples. The findings confirm the hypothesis that topographic and surface soil indices derived from RS data significantly influence *B. pseudomallei* distribution. Spatial risk maps developed from the model highlight environmentally favorable areas for bacterial persistence, providing actionable information for melioidosis surveillance and targeted interventions. Beyond these specific outcomes, the study demonstrates the value of combining interpretable ML with geostatistical approaches for environmental pathogen modeling. This integrative framework can be extended to other soil- or water-borne diseases to support early warning systems, public health planning, and sustainable agricultural management. Future work should include larger and temporally distributed datasets and explore

ensemble ML methods to further refine predictive performance and generalizability.

AUTHOR INFORMATION

Corresponding Author

Jaruwan Wongbutdee — College of Medicine and Public Health, Ubon Ratchathani University, Ubonratchathani-34190 (Thailand);
orcid.org/0009-0009-1884-9006
Email: jaruwan.w@ubu.ac.th

Authors

Wacharapong Saengnill — College of Medicine and Public Health, Ubon Ratchathani University, Ubonratchathani-34190 (Thailand);
orcid.org/0000-0001-7293-0635

Jutharat Jittimanee — College of Medicine and Public Health, Ubon Ratchathani University, Ubonratchathani-34190 (Thailand);
orcid.org/0000-0001-9998-1387

Suwaporn Dandee — College of Medicine and Public Health, Ubon Ratchathani University, Ubonratchathani-34190 (Thailand);
orcid.org/0000-0002-9359-5298

Pongthep Thongsang — Department of Geology, Chulalongkorn University, Bangkok - 10330 (Thailand);
orcid.org/0009-0008-9955-2628

Author Contributions

W. S. Project administration, Supervision, Conceptualization, Visualization, Methodology, Formal analysis, Data curation, and Writing – review & editing; J. J. and S. D. Data curation and Methodology (Labolatory); J. W. Conceptualization, Data collection, Laboratory, Writing – original draft, and Writing – review & editing; P. T. Methodology.

Conflicts of Interest

The authors declare no conflict of interest.

ACKNOWLEDGEMENT

This research supported by The National Science, Research and Innovation Fund (Fundamental Fund, FF67), grant to Wacharapong Saengnill.

DECLARATION OF GENERATIVE AI

Not applicable.

REFERENCES

[1] I. Grishkan. (2024). "Soil as a Source of Fungi Pathogenic for Public Health". *Encyclopedia.* **4** (3): 1163-1172. 10.3390/encyclopedia4030075.

[2] Z.-L. Li, H. Wu, S.-B. Duan, W. Zhao, H. Ren, X. Liu, P. Leng, R. Tang, X. Ye, J. Zhu, Y. Sun, M. Si, M. Liu, J. Li, X. Zhang, G. Shang, B.-H. Tang, G. Yan, and C. Zhou. (2023). "Satellite Remote Sensing of Global Land Surface Temperature: Definition, Methods, Products, and Applications". *Reviews of Geophysics.* **61** (1): e2022RG000777. 10.1029/2022RG000777.

[3] L. A. K. Melchior, K. R. C. da Silva, A. E. P. Silva, and F. Chiaravalloti-Neto. (2024). "Spatial, Temporal, and Space-Time Analysis of Leptospirosis Cases in Acre, 2001-2022". *Revista Brasileira de Epidemiologia.* **27** : e240063. 10.1590/1980-549720240063.

[4] J. Wongbutdee, J. Jittimanee, and W. Saengnill. (2023). "Spatiotemporal Distribution and Geostatistically Interpolated Mapping of the Melioidosis Risk in an Endemic Zone in Thailand". *Geospatial Health.* **18** (2): 1189. 10.4081/gh.2023.1189.

[5] D. J. F. Gerber, S. Dhakal, M. N. Islam, A. Al Kawsar, M. A. Khair, M. M. Rahman, M. J. Karim, M. S. Rahman, M. M. Aktaruzzaman, C. Tupps, M. Stephens, P. M. Emerson, J. Utzinger, and P. Vounatsou. (2023). "Distribution and Treatment Needs of Soil-Transmitted Helminthiasis in Bangladesh: A Bayesian Geostatistical Analysis of 2017-2020 National Survey Data". *PLoS Neglected Tropical Diseases.* **17** (11): e0011656. 10.1371/journal.pntd.0011656.

[6] R. J. Soares Magalhães, M. S. Salamat, L. Leonardo, D. J. Gray, H. Carabin, K. Halton, D. P. McManus, G. M. Williams, P. Rivera, O. Saniel, L. Hernandez, L. Yakob, S. T. McGarvey, and A. C. A. Clements. (2015). "Mapping the Risk of Soil-Transmitted

Helminthic Infections in the Philippines". *PLoS Neglected Tropical Diseases.* **9** (9): e0003915. 10.1371/journal.pntd.0003915.

[7] T. Tsheten, K. A. Alene, A. C. Restrepo, M. Kelly, C. Lau, A. C. A. Clements, D. J. Gray, C. Daga, V. J. Mapalo, F. E. Espino, and K. Wangdi. (2024). "Risk Mapping and Socio-Ecological Drivers of Soil-Transmitted Helminth Infections in the Philippines: A Spatial Modelling Study". *The Lancet Regional Health - Western Pacific.* **43** : 100974. 10.1016/j.lanwpc.2023.100974.

[8] H. Mohd Hatta, K. I. Musa, N. M. H. Mohd Fuzi, and P. Moraga. (2024). "Spatial Interaction Between Leptospirosis and Enteric Fever in Kelantan, Malaysia: A 2016-2022 Notification Registry Analysis". *Asia-Pacific Journal of Public Health.* **36** (8): 738-745. 10.1177/10105395241286118.

[9] D. Limmathurotsakul, N. Golding, D. A. B. Dance, J. P. Messina, D. M. Pigott, C. L. Moyes, D. B. Rolim, E. Bertherat, N. P. J. Day, S. J. Peacock, and S. I. Hay. (2016). "Predicted Global Distribution of Burkholderia pseudomallei and Burden of Melioidosis". *Nature Microbiology.* **1** : 15008. 10.1038/nmicrobiol.2015.8.

[10] S. Smith, P. Horne, S. Rubenach, R. Gair, J. Stewart, L. Fairhead, and J. Hanson. (2021). "Increased Incidence of Melioidosis in Far North Queensland, Queensland, Australia, 1998-2019". *Emerging Infectious Diseases.* **27** (12): 3119-3123. 10.3201/eid2712.211302.

[11] T. J. J. Inglis and J. L. Sagripanti. (2006). "Environmental Factors That Affect the Survival and Persistence of Burkholderia pseudomallei". *Applied and Environmental Microbiology.* **72** (11): 6865-6875. 10.1128/AEM.01036-06.

[12] S. Paksanont, K. Sintiprungrat, T. Yimthin, P. Pumirat, S. J. Peacock, and N. Chantratita. (2018). "Effect of Temperature on Burkholderia pseudomallei Growth, Proteomic Changes, Motility, and Resistance to Stress Environments". *Scientific Reports.* **8** (1): 9167. 10.1038/s41598-018-27356-7.

[13] P. Pumirat, M. Vanaporn, U. Boonyuen, N. Indrawattana, A. Rungruengkitkun, and N. Chantratita. (2017). "Effects of Sodium Chloride on Heat Resistance, Oxidative Susceptibility, Motility, Biofilm, and Plaque Formation of Burkholderia pseudomallei". *MicrobiologyOpen.* **6** (4): e493. 10.1002/mbo3.493.

[14] S. Wang-Ngarm, S. Chareonsudjai, and P. Chareonsudjai. (2014). "Physicochemical Factors Affecting the Growth of Burkholderia pseudomallei in Soil Microcosm". *The American Journal of Tropical Medicine and Hygiene.* **90** (3): 480-485. 10.4269/ajtmh.13-0446.

[15] S. Tong, S. Yang, Z. Lu, and W. He. (1996). "Laboratory Investigation of Ecological Factors Influencing the Environmental Presence of Burkholderia pseudomallei". *Microbiology and Immunology.* **40** (6): 451-453. 10.1111/j.1348-0421.1996.tb01092.x.

[16] Y. S. Chen, S. C. Chen, C. M. Kao, and Y. L. Chen. (2003). "Effects of Soil pH, Temperature and Water Content on the Growth of Burkholderia pseudomallei". *Folia Microbiologica.* **48** (2): 253-256. 10.1007/BF02930965.

[17] P. W. Dhewantara, W. Hu, W. Zhang, W.-W. Yin, F. Ding, A. Al Mamun, and R. J. Soares Magalhães. (2019). "Climate Variability, Satellite-Derived Physical Environmental Data and Human Leptospirosis: A Retrospective Ecological Study in China". *Environmental Research.* **176** : 108523. 10.1016/j.envres.2019.06.004.

[18] A. Luenam and N. Puttanapong. (2020). "Modelling and Analyzing Spatial Clusters of Leptospirosis Based on Satellite-Generated Measurements of Environmental Factors in Thailand During 2013-2015". *Geospatial Health.* **15** (2): 856. 10.4081/gh.2020.856.

[19] S. Brooker, P. Singhasivanon, J. Waikagul, S. Supavej, S. Kojima, T. Takeuchi, T. V. Luong, and S. Looareesuwan. (2003). "Mapping Soil-Transmitted Helminths in Southeast Asia and Implications for Parasite Control". *The Southeast Asian Journal of Tropical Medicine and Public Health.* **34** (1): 24-36.

[20] R. G. C. Scholte, N. Schur, M. E. Bavia, E. M. Carvalho, F. Chammartin, J. Utzinger,

and P. Vounatsou. (2013). "Spatial Analysis and Risk Mapping of Soil-Transmitted Helminth Infections in Brazil, Using Bayesian Geostatistical Models". *Geospatial Health.* **8** (1): 97-110. 10.4081/gh.2013.58.

[21] F. Alpsalaz, Y. Özüpak, E. Aslan, and H. Uzel. (2025). "Classification of Maize Leaf Diseases with Deep Learning: Performance Evaluation of the Proposed Model and Use of Explainable Artificial Intelligence". *Chemometrics and Intelligent Laboratory Systems.* **262** : 105412. 10.1016/j.chemolab.2025.105412.

[22] E. Aslan and Y. Özüpak. (2024). "Diagnosis and Accurate Classification of Apple Leaf Diseases Using Vision Transformers". *Computer and Decision Making: An International Journal.* **1** : 1-12. 10.59543/comdem.v1i.10039.

[23] Y. Özüpak, F. Alpsalaz, E. Aslan, and H. Uzel. (2025). "Hybrid Deep Learning Model for Maize Leaf Disease Classification with Explainable Artificial Intelligence". *New Zealand Journal of Crop and Horticultural Science.* **53** (5): 2942-2964. 10.1080/01140671.2025.2519570.

[24] W. Alsabhan and A. Alfadhly. (2025). "Effectiveness of Machine Learning Models in Diagnosis of Heart Disease: A Comparative Study". *Scientific Reports.* **15** (1): 24568. 10.1038/s41598-025-09423-y.

[25] H. Belhad, A. Bourbia, and S. Boughanja. (2025). "Chronic Diseases Prediction Using Machine Learning and Deep Learning Methods". *arXiv.* 10.48550/arXiv.2505.00189.

[26] T. Chen and C. Guestrin. (2016). "XGBoost: A Scalable Tree Boosting System". *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 785-794. 10.1145/2939672.2939785.

[27] E. Aslan and Y. Özüpak. (2024). "Detection of Road Extraction from Satellite Images with Deep Learning Method". *Cluster Computing.* **28** (1): 72. 10.1007/s10586-024-04880-y.

[28] C. Ruangroj. (2023). "Melioidosis Report". Ubon Ratchathani Provincial Public Health, Office of Public Health, Thailand.

[29] W. Saengnill, K. Charoenjit, K. Hrimpeng, and J. Jittimanee. (2020). "Mapping the Probability of Detecting Burkholderia pseudomallei in Rural Rice Paddy Soil Based on Indicator Kriging and Spatial Soil Factor Analysis". *Transactions of the Royal Society of Tropical Medicine and Hygiene.* **114** (7): 521-530. 10.1093/trstmh/traa029.

[30] P. L. Bulterys, M. A. Bulterys, K. Phommasone, M. Luangraj, M. Mayxay, S. Kloprogge, T. Miliya, M. Vongsouvath, P. N. Newton, R. Phetsouvanh, C. T. French, J. F. Miller, P. Turner, and D. A. B. Dance. (2018). "Climatic Drivers of Melioidosis in Laos and Cambodia: A 16-Year Case Series Analysis". *The Lancet Planetary Health.* **2** (8): e334-e343. 10.1016/S2542-5196(18)30172-4.

[31] X. Liu, L. Pang, S. H. Sim, K. T. Goh, S. Ravikumar, M. S. Win, G. Tan, A. R. Cook, D. Fisher, and L. Y. A. Chai. (2015). "Association of Melioidosis Incidence with Rainfall and Humidity, Singapore, 2003-2012". *Emerging Infectious Diseases.* **21** (1): 159-162. 10.3201/eid2101.140042.

[32] S. Palasatien, R. Lertsirivorakul, P. Royros, S. Wongratanacheewin, and R. W. Sermswan. (2008). "Soil Physicochemical Properties Related to the Presence of Burkholderia pseudomallei". *Transactions of the Royal Society of Tropical Medicine and Hygiene.* **102** : S5-S9. 10.1016/S0035-9203(08)70003-8.

[33] N. Chantratita, V. Wuthiekanun, K. Boonbumrung, R. Tiyawisutsri, M. Vesaratchavest, D. Limmathurotsakul, W. Chierakul, S. Wongratanacheewin, S. Pukritiyakamee, N. J. White, N. P. J. Day, and S. J. Peacock. (2007). "Biological Relevance of Colony Morphology and Phenotypic Switching by Burkholderia pseudomallei". *Journal of Bacteriology.* **189** (3): 807-817. 10.1128/JB.01258-06.

[34] V. Wuthiekanun, N. Anuntagool, N. J. White, and S. Sirisinha. (2002). "Short Report: A Rapid Method for the Differentiation of

Burkholderia pseudomallei and Burkholderia thailandensis". *The American Journal of Tropical Medicine and Hygiene.* **66** (6): 759-761. 10.4269/ajtmh.2002.66.759.

[35] N. Anuntagool, P. Naigowit, V. Petkanchanapong, P. Aramsri, T. Panichakul, and S. Sirisinha. (2000). "Monoclonal Antibody-Based Rapid Identification of Burkholderia pseudomallei in Blood Culture Fluid from Patients with Community-Acquired Septicaemia". *Journal of Medical Microbiology.* **49** (12): 1075-1078. 10.1099/0022-1317-49-12-1075.

[36] E. Aslan and Y. Özüpak. (2025). "Comparison of Machine Learning Algorithms for Automatic Prediction of Alzheimer Disease". *Journal of the Chinese Medical Association.* **88** (2). 10.1097/JCMA.0000000000001188.

[37] E. Aslan, Y. Özüpak, F. Alpsalaz, and Z. M. S. Elbarbary. (2025). "A Hybrid Machine Learning Approach for Predicting Power Transformer Failures Using Internet of Things-Based Monitoring and Explainable Artificial Intelligence". *IEEE Access.* **13** : 113618-113633. 10.1109/ACCESS.2025.3583773.

[38] D. Limmathurotsakul, V. Wuthiekanun, N. Chantratita, G. Wongsuvan, P. Amornchai, N. P. J. Day, and S. J. Peacock. (2010). "Burkholderia pseudomallei Is Spatially Distributed in Soil in Northeast Thailand". *PLoS Neglected Tropical Diseases.* **4** (6): e694. 10.1371/journal.pntd.0000694.

[39] D. A. B. Dance, M. Knappik, S. Dittrich, V. Davong, J. Silisouk, M. Vongsouvath, S. Rattanavong, A. Pierret, P. N. Newton, P. Amornchai, V. Wuthiekanun, S. Langla, and D. Limmathurotsakul. (2018). "Evaluation of Consensus Method for the Culture of Burkholderia pseudomallei in Soil Samples from Laos". *Wellcome Open Research.* **3** : 132. 10.12688/wellcomeopenres.14851.2.

[40] H. I. Musa, L. Hassan, Z. H. Shamsuddin, C. Panchadcharam, Z. Zakaria, and S. Abdul Aziz. (2016). "Physicochemical Properties Influencing Presence of Burkholderia pseudomallei in Soil from Small Ruminant Farms in Peninsular Malaysia". *PLoS One.* **11** (9): e0162348. 10.1371/journal.pone.0162348.

[41] R. Shaharudin, N. Ahmad, M. A. Kamaluddin, and Y. Veloo. (2016). "Detection of Burkholderia pseudomallei from Post-Flood Soil Samples in Kelantan, Malaysia". *The Southeast Asian Journal of Tropical Medicine and Public Health.* **47** (5): 951-956.

[42] M. M. M. Swe, M. M. Win, J. Cohen, A. P. Phyo, H. N. Lin, K. Soe, P. Amornchai, T. T. Wah, K. K. N. Win, C. Ling, D. M. Parker, D. A. B. Dance, E. A. Ashley, and F. Smithuis. (2021). "Geographical Distribution of Burkholderia pseudomallei in Soil in Myanmar". *PLoS Neglected Tropical Diseases.* **15** (5): e0009372. 10.1371/journal.pntd.0009372.

[43] L. Manivanh, A. Pierret, S. Rattanavong, O. Kounnavongsa, Y. Buisson, I. Elliott, J. Maeght, K. Xayyathip, J. Silisouk, M. Vongsouvath, R. Phetsouvanh, P. N. Newton, G. Lacombe, O. Ribolzi, E. Rochelle-Newall, and D. A. B. Dance. (2017). "Burkholderia pseudomallei in a Lowland Rice Paddy: Seasonal Changes and Influence of Soil Depth and Physico-Chemical Properties". *Scientific Reports.* **7** (1): 3031. 10.1038/s41598-017-02946-z.

[44] A. L. Baker and J. M. Warner. (2016). "Burkholderia pseudomallei Is Frequently Detected in Groundwater That Discharges to Major Watercourses in Northern Australia". *Folia Microbiologica.* **61** (4): 301-305. 10.1007/s12223-015-0438-3.

[45] M. L. Corkeron, R. Norton, and P. N. Nelson. (2010). "Spatial Analysis of Melioidosis Distribution in a Suburban Area". *Epidemiology and Infection.* **138** (9): 1346-1352. 10.1017/S0950268809991634.

[46] R. Seng, N. Saiprom, R. Phunpang, C. J. Baltazar, S. Boontawee, T. Thodthasri, W. Silakun, and N. Chantratita. (2019). "Prevalence and Genetic Diversity of Burkholderia pseudomallei Isolates in the Environment Near a Patient's Residence in Northeast Thailand". *PLoS Neglected Tropical Diseases.* **13** (4): e0007348. 10.1371/journal.pntd.0007348.

[47] J. F. V. Abrantes, Z. A. P. Cariño, H. L. S. Mercado, F. N. Vicencio, G. R. S. Sosa, M. A. M. Habaña, and N. H. A. Dagamac. (2025). "Identification of Environmental Determinants Involved in the Distribution of Burkholderia pseudomallei in Southeast Asia Using MaxEnt Software". *PLoS Neglected Tropical Diseases.* **19** (1): e0012684. 10.1371/journal.pntd.0012684.

[48] T. Sarin, V. Vinoj, G. P. Gujjula, B. Behera, J. Jena, and S. Mohanty. (2025). "The Spatial Mapping of Melioidosis Exposure in the Eastern Indian State of Odisha". *Current Research in Microbial Sciences.* **8** : 100346. 10.1016/j.crmicr.2025.100346.

[49] J. Wongbutdee, J. Jittimanee, S. Daendee, P. Thongsang, and W. Saengnill. (2024). "Exploring the Relationship Between Melioidosis Morbidity Rate and Local Environmental Indicators Using Remotely Sensed Data". *International Journal of Environmental Research and Public Health.* **21** (5): 614. 10.3390/ijerph21050614.

[50] K. Pongmala, A. Pierret, P. Oliva, A. Pando, V. Davong, S. Rattanavong, N. Silvera, M. Luangraj, L. Boithias, K. Xayyathip, L. Menjot, M. Macouin, E. Rochelle-Newall, H. Robain, A. Vongvixay, A. J. H. Simpson, D. A. B. Dance, and O. Ribolzi. (2022). "Distribution of Burkholderia pseudomallei Within a 300-cm Deep Soil Profile: Implications for Environmental Sampling". *Scientific Reports.* **12** (1): 8674. 10.1038/s41598-022-12795-0.

[51] T. Shaw, K. Assig, C. Tellapragada, G. E. Wagner, M. Choudhary, A. Göhler, V. K. Eshwara, I. Steinmetz, and C. Mukhopadhyay. (2022). "Environmental Factors Associated With Soil Prevalence of the Melioidosis Pathogen Burkholderia pseudomallei: A Longitudinal Seasonal Study From South West India". *Frontiers in Microbiology.* **13** : 902996. 10.3389/fmicb.2022.902996.

[52] M. Kaestli, M. Mayo, G. Harrington, L. Ward, F. Watt, J. V. Hill, A. C. Cheng, and B. J. Currie. (2009). "Landscape Changes Influence the Occurrence of the Melioidosis Bacterium Burkholderia pseudomallei in Soil in Northern Australia". *PLoS Neglected Tropical Diseases.* **3** (1): e364. 10.1371/journal.pntd.0000364.