scientific reports

Check for updates

OPEN

# The global epidemiology, risk factors, and mortality prediction of nocardiosis: an easily missed opportunistic infection

Bingqian Du[1], Ziyu Song[1,2], Zhiqiang Ren[3], Deming Tang[4], Jirao Shen[1], Jiang Yao[1,5], Xiaotong Qiu[1], Shuai Xu[1], Min Yuan[1], Zhiguo Liu[1] & Zhenjun Li[1]✉

**This study was to comprehensively investigate the epidemiology of nocardiosis worldwide and develop an interpretable machine learning (ML) model to predict mortality in patients with nocardiosis. The PubMed and Web of Science databases were searched for the literature review using the keywords: "*Nocardia*" or "nocardiosis" through 31 August 2024, 9,750 cases of nocardiosis were reported. Nine ML algorithms were employed to predict the mortality in patients with nocardiosis. A total of 9,750 reported cases were identified and included. Most cases were from North America and Asia. The mean age of patients was 50.4 ± 19.3, with a male predominance (64.7%). The overall all-cause mortality rate was 19.8%, although disseminated infections were associated with a higher mortality rate of 31.7%. Since 2000, the number of reported nocardiosis cases has increased markedly, while the all-cause mortality rate has decreased significantly and stabilized. The distribution of *Nocardia* species exhibited regional variation. Advanced age, male, underlying diseases, disseminated infections, infection type, clinical features, and use of corticosteroids or immunosuppressants had a higher risk of all-cause mortality [Odds Ratio (*ORs*) = 1.35–2.63, $P < 0.05$]. The stochastic gradient boosting (SGBT) model outperformed eight other machine learning models, accurately predicting mortality in patients with nocardiosis across both training and test datasets. This study provides a comprehensive overview of the global epidemiology and species distribution of nocardiosis, highlighting distinct regional patterns. An interpretable ML model was developed and validated that helps clinicians identify high-risk patients early and provides a basis for developing personalized treatment plans.**

**Keyword** *Nocardia*, Nocardiosis, The global epidemiololgy

*Nocardia* is a gram-positive, facultative aerobic actinomycete that is a widespread pathogen in the environment and can cause serious infections of the lungs, skin, brain, and central nervous system[1,2]. Nocardiosis is primarily an opportunistic infection in immunocompromised individuals, but immunocompetent people are also at risk[3]. Clinical diagnosis faces multiple challenges due to its lack of specificity in clinical features, vague imaging features, low bacterial load, and difficulty in culturing pathogens[1,4]. In terms of treatment, nocardiosis usually requires months to years of antimicrobial treatment and has a high recurrence rate, especially in patients with underlying diseases[5,6].

In recent years, with the global population growth and the increase in the number of immunosuppressed people (such as organ transplant and cancer chemotherapy patients), the number of *Nocardia* infection case reports and detections has increased significantly[1,4]. This change is not only attributed to the expansion of the susceptible population, but also related to the increase in detection rate brought about by advances in diagnostic technology[6–8]. However, there are still many data gaps in the global incidence and epidemic characteristics

[1]National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102200, China. [2]Wenzhou Key Laboratory of Sanitary Microbiology, Key Laboratory of Laboratory Medicine, Ministry of Education, China, School of Laboratory Medicine and Life Sciences, Wenzhou Medical University, Wenzhou 325035, China. [3]School of Public Health and Emergency Management, Southern University of Science and Technology, Shenzhen 518055, China. [4]Dongcheng Center for Disease Control and Prevention, Beijing 100009, China. [5]Department of Clinical Laboratory, The Quzhou Affiliated Hospital of Wenzhou Medical University (Quzhou People's Hospital), Quzhou 324002, China. ✉email: lizhenjun@icdc.cn

of nocardiosis at the national and regional levels, highlighting the urgency of strengthening monitoring and research.

*Nocardia* infection carries a high risk of mortality, with a reported one-year mortality rate of approximately 25%[9]. Risk factors for mortality include immunosuppression, use of glucocorticoids, disseminated infection, cavitary or pleural involvement, and infection with specific species[9,10]. Case reports and case series describe the experiences of individual or small groups of patients in detail and are increasingly used to build predictive models in medical research, especially for rare diseases where large-scale data collection is difficult[11]. It is worth noting that machine learning (ML) technology has shown significant advantages in early detection, diagnosis, and prognosis prediction of infectious diseases[12]. If a mortality prediction model for nocardiosis can be developed, it will provide key support for clinical diagnosis and personalized treatment strategies.

To fill the knowledge gap in the epidemiology, risk factors, and mortality prediction of nocardiosis, this study systematically reviewed the literature (including case reports and case series) published between January 1, 1950 and August 31, 2024, focusing on the epidemiological characteristics, clinical and laboratory manifestations, risk factors, mortality prediction indicators, strain distribution, and antimicrobial susceptibility of nocardiosis. The results of this study will provide a scientific basis for optimizing the surveillance system of nocardiosis and formulating precise prevention and control measures.

## Methods

### Protocol and registration

This study was conducted in strict accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for study selection and data reporting (Supplementary File 1)[13]. The protocol is registered in the PROSPERO database (Identifier: CRD42025633469).

### Search strategies

This systematic review focused on case reports and case series published between January 1, 1950, and August 31, 2024. We searched PubMed and Web of Science databases to identify all reported cases of nocardiosis, using the search terms "*Nocardia*" and "nocardiosis". All studies were identified through a manual review of the reference lists of included articles (Supplementary File 4). Two researchers independently screened and systematically assessed all retrieved references to exclude (a) Duplicate or irrelevant articles. (b) Reviews, editorials, conference abstracts, or studies without full-text availability. (c) Incomplete data on age, sex, country, or site of infection. Discrepancies were resolved through discussion with co-authors until a consensus was reached.

### Case inclusion and exclusion criteria

Cases were included in the final analysis if they met the following criteria: (1) Isolation of *Nocardia* species from clinical samples [sputum, bronchoalveolar lavage fluid (BALF), cerebrospinal fluid (CSF), abscess, serum, or tissue samples], or identified by genomic sequencing or molecular methods. (2) The diagnosis of *Nocardia* infection supported by clinical and radiological evidence.

### Data extraction

For the case review, researchers employed a standardized method to extract, code, and analyze relevant data from case reports using Microsoft Excel. Specifically, two researchers (B.D. and Z.S.), working independently and blinded to each other, coded detailed information for each case, including host factors (age, gender, underlying conditions, geographical distribution), microbial factors {types of *Nocardia* species, multidrug resistance (MDR) [In this study, we adopted the definition proposed by Robert Schlaberg et al., in which MDR *Nocardia* isolates are defined as those exhibiting resistance or intermediate resistance to two or more of the commonly used empirical antibiotics: amikacin (AMK), ceftriaxone (CRO), imipenem (IMP), and trimethoprim-sulfamethoxazole (TMP-SMX)[14].], TMP-SMX resistance}, infection related factors (type of infection, infection site), clinical and treatment-related factors (clinical features, previous medication types, treatment drugs, antibiotic or oprative treatment, all-cause mortality), sample sources, vital signs, and blood parameters. However, due to the long study period and variability in laboratory protocols, drug susceptibility testing (DST) methods were not standardized across reports. In particular, some earlier studies employed diffusion-based methods that are no longer recommended by current Clinical and Laboratory Standards Institute (CLSI) guidelines, and DST methodologies were not always specified. These inconsistencies may influence the comparability and accuracy of reported resistance data. Additionally, climate type and annual average precipitation were matched to individual cases using city-level data from WorldClim (https://www.worldclim.org/data/index.html).

### Data classification and grouping

Cases of nocardiosis were classified into two groups based on the extent of infection: (1) single site infection, where the infection was confined to one anatomical site, and (2) disseminated infection, where the infection involved two or more noncontiguous sites[9]. Furthermore, patients were stratified into two outcome-based groups: survival group and deceased group, based on standardized criteria for vital status at the end of follow-up. Vital status was determined from the original case reports, typically corresponding to documented cure and hospital discharge or survival at approximately one year after the initiation of antibiotic therapy.

### ML model

Risk factors and data preprocessing: we reviewed the literature on *Nocardia* infection to identify variables unnecessary for mortality prediction, aiming to simplify the model and mitigate overfitting. The severity and mortality of *Nocardia* infection are mainly affected by the following factors: host immune status, severity and extent of infection, microbial factors, and timeliness and effectiveness of diagnosis and treatment. Data for

categorical variables were processed using the one-hot encoding technique. The entire data split ratio of the training set and the test set is 80%: 20%.

Selection of variables: in this study, recursive feature elimination (RFE) was employed for variable selection to enhance predictive performance and improve model stability. To ensure the robustness of the feature selection process and the generalizability of the resulting models, we implemented 10 repetitions of tenfold cross-validation throughout the RFE procedure.

Model development and validation: nine ML algorithms—logistic regression (LR), decision tree (DT), random forest (RF), K-nearest neighbor (KNN), support vector machine (SVM), naive Bayes (NB), extreme gradient boosting (XGB), stochastic gradient boosting (SGBT), and neural network (NNET)—were employed to predict the mortality in patients with nocardiosis. To optimize model performance, we conducted hyperparameter tuning using the default grid search method implemented in the 'caret' package, applied to the optimal feature subset. This process was performed within 10 repetitions of tenfold cross-validation to ensure robust internal validation.

Model comparison and explanation: the reliability of the models was evaluated using several widely accepted indicators, including the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, brier score, accuracy, precision, negative predictive value (NPV), and F1 score. Calibration curves were also employed to assess the agreement between predicted probabilities and observed outcomes. The optimal prediction model was identified based on its performance across these evaluation indicators in both the training and test data. The SHapley Additive exPlanations (SHAP) method is a game theory-inspired approach designed to quantify feature importance and enhance interpretability of predictive models.

### Statistical analysis

Descriptive statistics were reported as counts and percentages for categorical variables and as means ± standard deviations (SDs) for continuous variables. The chi-square test was used to compare categorical variables, while the *t*-test was applied to compare continuous variables. Risk factors associated with all-cause mortality in patients with nocardiosis were identified through univariable and multivariable logistic regression analyses of the entire dataset. Covariates for the models, including age, gender, geographical distribution, temperature zone, and climate type, were selected based on a literature review. A two-tailed $P$-value < 0.05 was considered statistically significant. Statistical analyses were conducted using SAS version 9.4 software (SAS Institute Inc., Cary, NC, USA). Furthermore, we selected patients between 2005 and 2024 as subgroups for the above statistical analysis.

Nine machine learning models were developed using R version 4.4.3 and the "caret" (version 7.0–1) package, which offers a unified interface for various algorithms. The models were trained via the train function with specified method parameters: logistic regression (method = "glm"), decision tree (method = "rpart"), random forest (method = "ranger"), support vector machine with radial basis kernel (method = "svmRadial"), k-nearest neighbors (method = "knn"), naive Bayes (method = "naive bayes"), extreme gradient boosting (method = "xgbTree"), gradient boosting (method = "gbm"), and neural network (method = "nnet").

The global frequency distribution of nocardiosis and the corresponding distribution of *Nocardia* species were visualized using ArcGIS version 10.7 software (Environmental Systems Research Institute, Redlands, CA, USA).

## Results

### Study selection and characteristics

As illustrated in Supplementary File 1, a total of 10,832 relevant records were retrieved from two databases. After removing duplicates, 10,108 records underwent further eligibility screening. Following the exclusion of ineligible studies, this analysis included 2,077 studies, encompassing 9,750 cases of nocardiosis across 76 countries and regions worldwide. The included studies identified 79 distinct clinical *Nocardia* species isolates.

### Global epidemiology and distribution of nocardiosis

The global epidemiology and distribution of nocardiosis are illustrated in Fig. 1 (Fig. 1A 1950 to 2024 and Fig. 1B 2005 to 2024). Fourteen countries, including the United States, China, Canada, Japan, India, Spain, France, Israel, Australia, Iran, South Korea, Pakistan, Italy, and the United Kingdom, reported over 100 cases each. The distribution of clinical *Nocardia* species exhibited regional variation. Specifically, *N. asteroides* and *N. farcinica* were widely distributed, with *N. asteroides* predominating in the United States, Canada, Spain, the United Kingdom, and Saudi Arabia, and *N. farcinica* being more prevalent in China, Japan, France, and Germany. Significantly, some isolates that were previously identified as *N. asteroides* have now been reclassified and given new or different species. Therefore, the wide distribution of *N. asteroides* may be overstated. Moreover, *N. brasiliensis* was primarily reported in Latin America and parts of Asia. Over time, the epidemiology of nocardiosis has changed significantly compared with before 2000: the number of reported cases has increased substantially, while the all-cause mortality rate has decreased notably and stabilized (Fig. 2). Furthermore, since 2000, the identification of novel clinical *Nocardia* species has steadily risen (Fig. S1).

### Epidemiological and clinical characteristics of patients with nocardiosis

Only 3,035 of the 9,750 cases had detailed microbial, infection, clinical, and treatment-related data, so subsequent analyses were based on this population. Table 1 presents descriptive data for 3,035 patients with nocardiosis, comparing those with single-site infections (N = 2,032) to those with disseminated infections (N = 1,003), as well as outcomes of survival (N = 2,433) versus deceased (N = 602). For host factors, the mean age of patients was 50.4 ± 19.3, with cases reported across a wide age range, from a few days old to 98 years. Older age, particularly aged 70 years or above, and male gender were significantly associated with both disseminated infection (both $P < 0.001$) and death ($P < 0.001$ and $P = 0.015$, respectively). Underlying conditions conferring immunosuppression were strongly prevalent and linked to deceased outcomes. Cancer and organ/stem cell

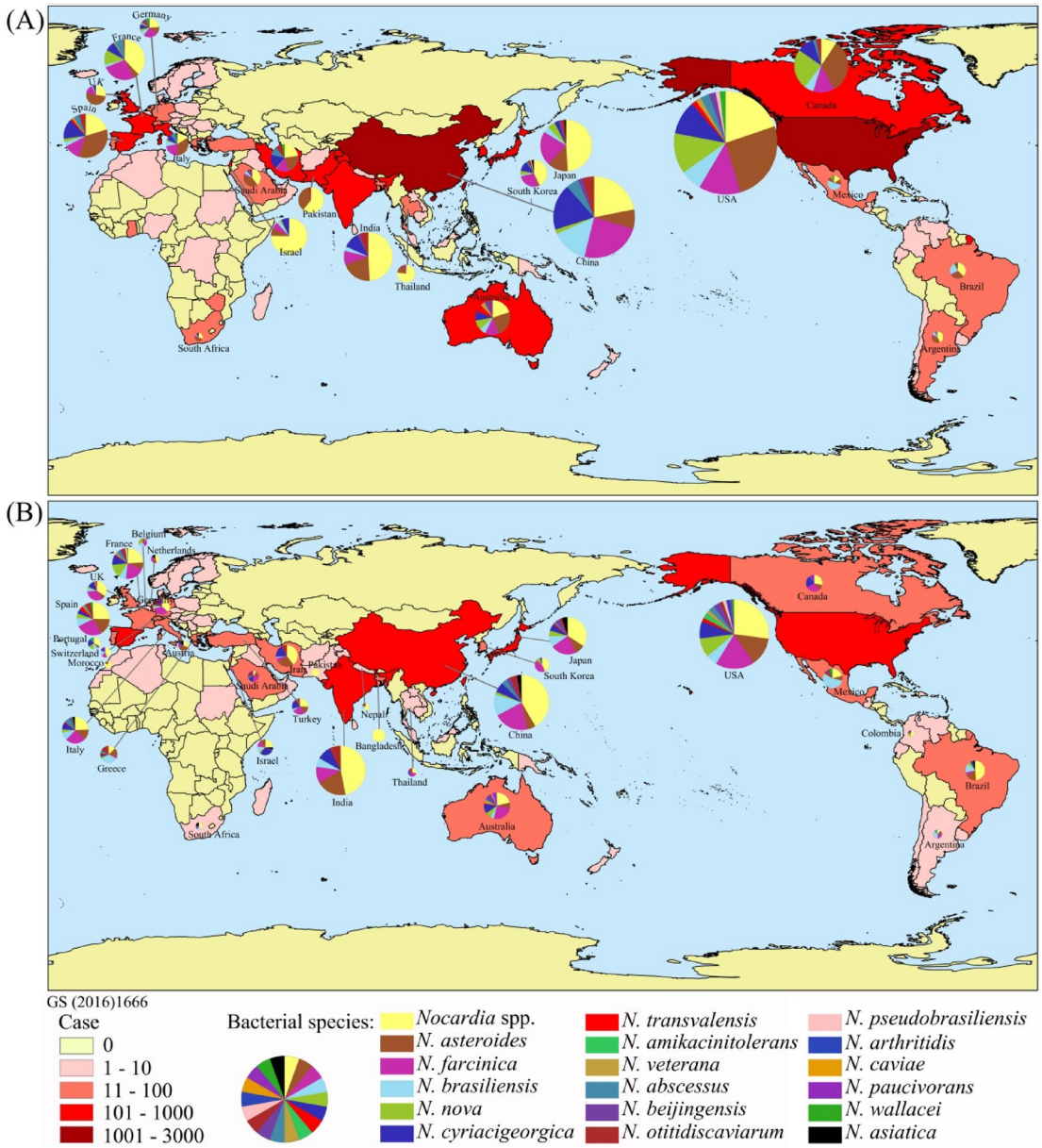


**Fig. 1**. The map of the global nocardiosis frequency distribution and corresponding *Nocardia* species from (**A**) 1950 to 2024 and (**B**) 2005 to 2024. [Note: Frequency from 1950 to 2024 and 2005 to 2024 were used to depict the map of global human nocardiosis, the light yellow indicates areas with no available case report. The map from Standard Map Services Website (http://bzdt.ch.mnr.gov.cn/), and approval number of map: GS (2016)1666.

transplants were significantly more common in patients with disseminated infection (both $P<0.001$) and deceased group ($P<0.001$ and $P=0.007$, respectively). Other conditions like cardiovascular diseases, renal disease, hepatopathy, chronic respiratory diseases, and autoimmune diseases also showed significant associations with dissemination (all $P<0.05$), while cardiovascular diseases, hepatopathy, and chronic respiratory diseases were additionally linked to increased mortality ($P<0.05$). Geographically, most patients were reported from North America (34.3%) and Asia (34.2%), followed by Europe (22.9%), Oceania (4.2%), and South America (2.9%), with only 47 patients (1.5%) reported from African countries. Patients were more likely to come from temperate (1,798, 59.2%) and monsoon (1,241, 40.9%) climate zones. For microbial factors, while *N. asteroides* was the most frequent isolate (28.5%), likely reflecting historical taxonomic limitations. Furthermore, patients infected with *N. asteroides* (26.6%) and *N. farcinica* (23.4%) exhibited higher mortality compared to those infected with other *Nocardia* species. For infection related factors, the mortality of disseminated infection was significantly higher than that of single-site infection, occurring in 31.6% of disseminated infection group compared to 14.0% of single-site infection group. Pulmonary infection was the most common infection site (55.5%), along with brain infection (28.2%). Brain infection was markedly higher in disseminated patients (58.3% vs 23.3%, $P<0.001$) and

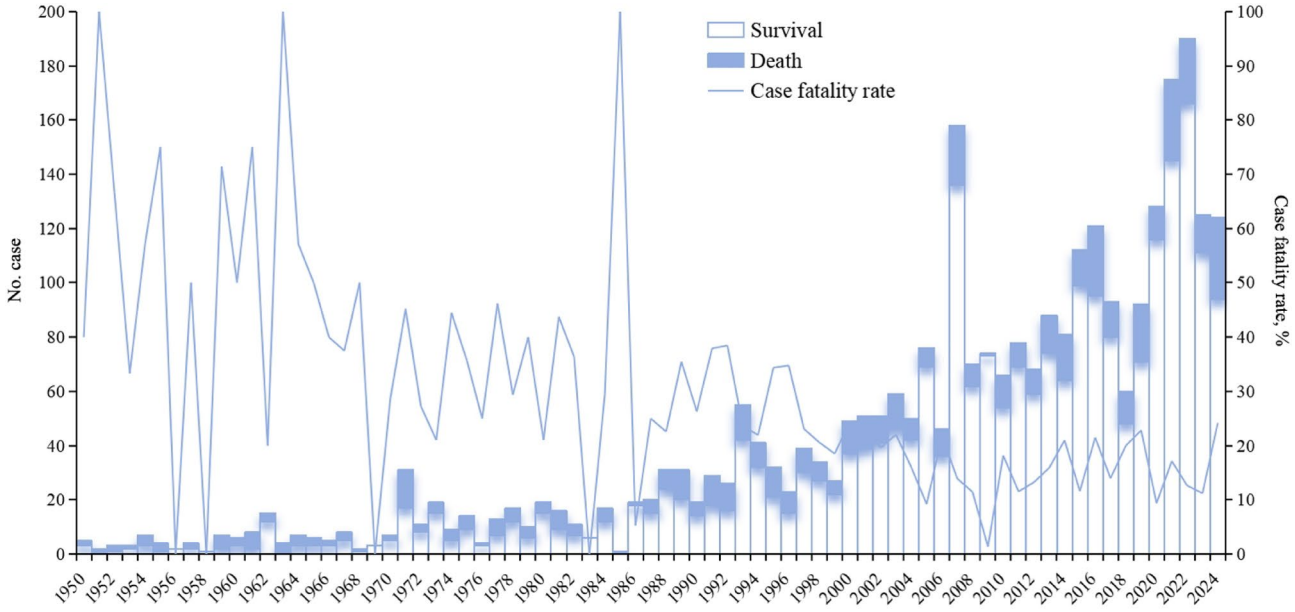


**Fig. 2**. Frequency of global human nocardiosis and case-fatality rates, 1950–2024.

fatal outcomes (43.3% vs 24.4%, $P<0.001$). However, due to limitations in the original reference, the proportion of brain infection in disseminated infections may have been underestimated. Moreover, among survival cases, the mean treatment duration was 185.9 days (SD 168.5) for single-site infections and 265.1 days (SD 169.8) for disseminated infections ($t=-7.350$, $P<0.001$). Fever was the most common clinical presentation in patients with nocardiosis, observed in 34.7% of cases. Other frequent manifestations included cough and/or expectoration (25.7%), skin symptoms (24.3%), dyspnoea (16.1%), brain abscess (14.5%), node (12.8%), eye symptoms (12.0%) and headache (10.4%) (Table S1). The prevalence of several symptoms was significantly higher in cases of disseminated infection compared to single-site infection, including fever, cough and/or expectoration, skin symptoms, brain abscess, dyspnea, node, and delirium ($P<0.05$). Similarly, the prevalence of fever, cough and/or expectoration, dyspnea, brain abscess, and level of capability/altered sensorium in the death group was also higher than that in the survival group ($P<0.05$). The use of corticosteroids or immunosuppressants was more common in patients with disseminated infection and deceased group (both $P<0.001$). Furthermore, patients with disseminated infections were more at risk of all-cause mortality relative to single-site infections (31.7% vs. 14.0%; $P<0.001$). Out of 3,035 cases, vital signs and laboratory characteristics were measured on 1,210 cases. The disseminated infection and death groups had significantly higher average body temperatures, neutrophil percentages, and C-reactive protein (CRP) levels compared to the single-site infection and survival groups, respectively (both $P<0.05$). Furthermore, white blood count (WBC) in the disseminated infection group was higher than in the single-site infection, respiratory rate in the death group was higher than the survival group (both $P<0.05$) (Table S2).

## Risk factors for mortality in patients with nocardiosis

After selecting clinically relevant variables based on domain knowledge, we retained the most direct risk factors or predictors. Table 2 presents risk factors associated with all-cause mortality in patients with nocardiosis. Using univariate logistic regression analysis, 21 potential risk factors associated with mortality were identified. After adjusting for potential confounders, 17 factors remained significantly associated with increased odds of mortality. Increasing age and male were an independent risk factors. Among underlying conditions, cancer, organ/stem cell transplant, cadiovascular diseases, hepatopathy, and chronic respiratory diseases were significant independent risk factors. Disseminated infection and five specific infection sites all showed a significantly higher risk of mortality (all $P<0.01$). Clinical features independently associated with increased risk included dyspnoea, delirium, brain abscess, and hemiplegia (all $P<0.01$). Furthermore, prior use of immunosuppressive medications, including corticosteroids, immunosuppressants, and chemotherapy/radiation, were strongly associated with all-cause mortality (all $P<0.01$).

## Prediction models for mortality in patients with nocardiosis

According to the 17 risk factors identified by the multivariate logistic regression analysis, we further used the RFE strategy for feature selection. The RFE variable selection process for ML model is visualized in Supplementary Fig. S2. The optimal feature subset for ML model was determined to be 10 variables using the RFE method. We then carried out ten repeats of tenfold internal cross-validation to construct nine ML models. The results showed that the XGB model performed the best in terms of AUC (Median = 0.718, IQR: 0.686–0.745) and specificity (AUC = 0.985, IQR: 0.974–0.990), followed by SGBT (AUC = 0.714, IQR: 0.689–0.735; specificity = 0.985, IQR: 0.974–0.990) and LR (AUC = 707, IQR: 0.680–0.728; specificity = 0.980, IQR: 0.974–0.990) (Tables S3 and S4).

| Characteristic | Total (N = 3035) | Single-site infection (N = 2032) | Disseminated infection (N = 1003) | $\chi^2$ | *P*-value | Survival (N = 2433) | Deceased(N = 602) | $\chi^2$ | *P*-value |
|---|---|---|---|---|---|---|---|---|---|
| *Host factors* | | | | | | | | | |
| Age (years) | | | | 36.737 | **< 0.001** | | | 30.497 | **< 0.001** |
| ≤ 18 | 201 (6.6) | 166 (8.2) | 35 (3.5) | | | 175 (7.2) | 26 (4.3) | | |
| 19–29 | 267 (8.8) | 192 (9.4) | 75 (7.5) | | | 220 (9.0) | 47 (7.8) | | |
| 30–39 | 376 (12.4) | 258 (12.7) | 118 (11.8) | | | 298 (12.2) | 78 (13.0) | | |
| 40–49 | 498 (16.4) | 310 (15.3) | 188 (18.7) | | | 406 (16.7) | 92 (15.3) | | |
| 50–59 | 591 (19.5) | 374 (18.4) | 217 (21.6) | | | 481 (19.8) | 110 (18.3) | | |
| 60–69 | 588 (19.4) | 399 (19.6) | 189 (18.8) | | | 477 (19.6) | 111 (18.4) | | |
| 70–79 | 386 (12.7) | 245 (12.1) | 141 (14.1) | | | 292 (12.0) | 94 (15.6) | | |
| ≥ 80 | 128 (4.2) | 88 (4.3) | 40 (4.0) | | | 84 (3.5) | 44 (7.3) | | |
| Gender | | | | 10.403 | **0.001** | | | 5.871 | **0.015** |
| Female | 1071 (35.3) | 757 (37.3) | 314 (31.3) | | | 884 (36.3) | 187 (31.1) | | |
| Male | 1964 (64.7) | 1275 (62.7) | 689 (68.7) | | | 1549 (63.7) | 415 (68.9) | | |
| Underlying conditions | | | | | | | | | |
| Cancer | 389 (12.8) | 219 (10.8) | 170 (16.9) | 22.89 | **< 0.001** | 246 (10.1) | 143 (23.7) | 80.388 | **< 0.001** |
| Organ/stem cell transplant | 504 (16.6) | 287 (14.1) | 217 (21.6) | 27.357 | **< 0.001** | 382 (15.7) | 122 (20.2) | 7.262 | **0.007** |
| Diabetes mellitus | 291 (9.6) | 178 (8.8) | 113 (11.3) | 4.866 | **0.027** | 221 (9.0) | 70 (11.6) | 3.604 | 0.058 |
| Cardiovascular diseases | 299 (9.9) | 174 (8.6) | 125 (12.5) | 11.499 | **< 0.001** | 225 (9.3) | 74 (12.3) | 5.037 | **0.025** |
| HIV | 177 (5.8) | 109 (5.4) | 68 (6.8) | 2.450 | 0.118 | 133 (5.4) | 44 (7.3) | 2.983 | 0.084 |
| Renal disease | 378 (12.5) | 184 (9.1) | 194 (19.3) | 65.173 | **< 0.001** | 296 (12.1) | 82 (13.6) | 0.937 | 0.333 |
| Hepatopathy | 98 (3.2) | 53 (2.6) | 45 (4.5) | 7.582 | **0.006** | 63 (2.6) | 35 (5.8) | 16.059 | **< 0.001** |
| Chronic respiratory diseases | 329 (10.8) | 240 (11.8) | 89 (8.9) | 5.996 | **0.014** | 245 (10.1) | 84 (14.0) | 7.531 | **0.006** |
| Autoimmune diseases [a] | 452 (14.9) | 251 (12.4) | 201 (20.0) | 31.311 | **< 0.001** | 365 (15.0) | 87 (14.5) | 0.115 | 0.734 |
| Geographical distribution | | | | 45.403 | **< 0.001** | | | 79.871 | **< 0.001** |
| Asia | 1039 (34.2) | 763 (37.5) | 276 (27.5) | | | 924 (38.0) | 115 (19.1) | | |
| Europe | 695 (22.9) | 434 (21.4) | 261 (26.0) | | | 527 (21.7) | 168 (27.9) | | |
| North america | 1040 (34.3) | 673 (33.1) | 367 (36.6) | | | 794 (32.6) | 246 (40.9) | | |
| South america | 88 (2.9) | 63 (3.1) | 25 (2.5) | | | 61 (2.5) | 27 (4.5) | | |
| Africa | 47 (1.5) | 35 (1.7) | 12 (1.2) | | | 33 (1.4) | 14 (2.3) | | |
| Oceania | 126 (4.2) | 64 (3.1) | 62 (6.2) | | | 94 (3.9) | 32 (5.3) | | |
| Temperature zone | | | | 74.76 | **< 0.001** | | | 15.338 | **< 0.001** |
| Tropical climate | 437 (14.4) | 371 (18.3) | 66 (6.6) | | | 380 (15.6) | 57 (9.5) | | |
| Subtropical climate | 800 (26.4) | 519 (25.5) | 281 (28.0) | | | 639 (26.3) | 161 (26.7) | | |
| Temperate climate | 1798 (59.2) | 1142 (56.2) | 656 (65.4) | | | 1414 (58.1) | 384 (63.8) | | |
| Climate type | | | | 24.99 | **< 0.001** | | | 11.168 | **0.025** |
| Continental climate | 858 (28.3) | 560 (27.6) | 298 (29.7) | | | 681 (28.0) | 177 (29.4) | | |
| Mediterranean climate | 422 (13.9) | 276 (13.6) | 146 (14.6) | | | 328 (13.5) | 94 (15.6) | | |
| Maritime climate | 393 (12.9) | 230 (11.3) | 163 (16.3) | | | 299 (12.3) | 94 (15.6) | | |
| Monsoon climate | 1241 (40.9) | 875 (43.1) | 366 (36.5) | | | 1026 (42.2) | 215 (35.7) | | |
| Rainforest/savanna climate | 121 (4.0) | 91 (4.5) | 30 (3.0) | | | 99 (4.1) | 22 (3.7) | | |
| Annual average precipitation (mm) | 967.63 ± 447.46 | 976.76 ± 462.23 | 949.12 ± 415.53 | 2.57 | 0.109 | 967.63 ± 447.46 | 960.76 ± 447.53 | 2.400 | 0.122 |
| *Microbial factors* | | | | | | | | | |
| Type of *nocardia* species [b] | | | | 152.277 | **0.002** | | | 79.938 | **< 0.001** |
| *N. Asteroides* | 866 (28.5) | 583 (28.7) | 283 (28.2) | | | 636 (26.1) | 230 (38.2) | | |
| *N. Farcinica* | 423 (13.9) | 228 (11.2) | 195 (19.4) | | | 324 (13.3) | 99 (16.4) | | |
| *N. Brasiliensis* | 275 (9.1) | 214 (10.5) | 61 (6.1) | | | 252 (10.4) | 23 (3.8) | | |
| *N. Cyriacigeorgica* | 169 (5.6) | 111 (5.5) | 58 (5.8) | | | 136 (5.6) | 33 (5.5) | | |
| *N. Nova* | 120 (4.0) | 77 (3.8) | 43 (4.3) | | | 98 (4.0) | 22 (3.7) | | |
| *N. Otitidiscaviarum* | 88 (2.9) | 64 (3.1) | 24 (2.4) | | | 70 (2.9) | 18 (3.0) | | |
| *N. Abscessus* | 45 (1.5) | 31 (1.5) | 14 (1.4) | | | 40 (1.6) | 5 (0.8) | | |
| *N. Beijingensis* | 45 (1.5) | 32 (1.6) | 13 (1.3) | | | 41 (1.7) | 4 (0.7) | | |
| *N. Transvalensis* | 33 (1.1) | 18 (0.9) | 15 (1.5) | | | 23 (0.9) | 10 (1.7) | | |
| *N. Asiatica* | 33 (1.1) | 19 (0.9) | 14 (1.4) | | | 32 (1.3) | 1 (0.2) | | |
| *N. Veterana* | 27 (0.9) | 21 (1.0) | 6 (0.6) | | | 21 (0.9) | 6 (1.0) | | |

Continued

| Characteristic | Total (N = 3035) | Single-site infection (N = 2032) | Disseminated infection (N = 1003) | $\chi^2$ | *P*-value | Survival (N = 2433) | Deceased(N = 602) | $\chi^2$ | *P*-value |
|---|---|---|---|---|---|---|---|---|---|
| *N. Paucivorans* | 21 (0.7) | 8 (0.4) | 13 (1.3) | | | 20 (0.8) | 1 (0.2) | | |
| *N. Caviae* | 17 (0.6) | 12 (0.6) | 5 (0.5) | | | 12 (0.5) | 5 (0.8) | | |
| *N. Pseudobrasiliensis* | 15 (0.5) | 12 (0.6) | 3 (0.3) | | | 13 (0.5) | 2 (0.3) | | |
| *N. Wallacei* | 15 (0.5) | 7 (0.3) | 8 (0.8) | | | 10 (0.4) | 5 (0.8) | | |
| *N. Amikacinitolerans* | 13 (0.4) | 13 (0.6) | 0 (0.0) | | | 12 (0.5) | 1 (0.2) | | |
| *N. Arthritidis* | 12 (0.4) | 8 (0.4) | 4 (0.4) | | | 12 (0.5) | 0 (0.0) | | |
| *N. Elegans* | 11 (0.4) | 6 (0.3) | 5 (0.5) | | | 9 (0.4) | 2 (0.3) | | |
| Unidentified and other species | 807 (26.6) | 568 (28.0) | 239 (23.8) | | | 672 (27.6) | 135 (22.4) | | |
| Multidrug resistance | | | | 7.345 | **0.025** | | | 3.934 | 0.140 |
| Yes | 281 (9.3) | 172 (8.5) | 109 (10.9) | | | 213 (8.8) | 68 (11.3) | | |
| No | 491 (16.2) | 316 (15.5) | 175 (17.5) | | | 392 (16.1) | 99 (16.5) | | |
| Unknown | 2263 (74.5) | 1544 (76.0) | 719 (71.6) | | | 1828 (75.1) | 435 (72.2) | | |
| Tmp-smx resistance | | | | 9.27 | **0.010** | | | 2.034 | 0.362 |
| Yes | 119 (3.9) | 73 (3.6) | 46 (4.6) | | | 90 (3.7) | 29 (4.8) | | |
| No | 522 (17.2) | 324 (15.9) | 198 (19.7) | | | 414 (17.0) | 108 (17.9) | | |
| Unknown | 2394(78.9) | 1635 (80.5) | 759 (75.7) | | | 1929 (79.3) | 465 (77.2) | | |
| *Infection related factors* | | | | | | | | | |
| Type of infection | | | | | | | | 130.516 | **< 0.001** |
| Single site infection | 2032 (66.9) | | | | | 1747 (71.8) | 285 (47.3) | | |
| Disseminated infection | 1003 (33.1) | | | | | 686 (28.2) | 317 (52.7) | | |
| Infection site | | | | | | | | | |
| Pulmonary infection | 1686 (55.5) | 865 (42.6) | 821 (81.1) | 419.739 | **< 0.001** | 1258 (51.7) | 428 (71.0) | 73.487 | **< 0.001** |
| Skin infection | 859 (28.3) | 272 (13.4) | 385 (38.3) | 75.036 | **< 0.001** | 726 (29.8) | 133 (22.0) | 14.272 | **< 0.001** |
| Brain infection | 857 (28.2) | 474 (23.3) | 585 (58.3) | 669.261 | **< 0.001** | 596 (24.4) | 261 (43.3) | 84.702 | **< 0.001** |
| Eye infection | 326 (10.7) | 230 (11.3) | 96 (9.5) | 2.139 | 0.1436 | 298 (12.2) | 28 (4.6) | 29.051 | **< 0.001** |
| Bacteremia | 174 (5.7) | 12 (0.5) | 162 (16.1) | 300.878 | **< 0.001** | 109 (4.4) | 65 (10.7) | 35.636 | **< 0.001** |
| *Clinical and treatment-related factors* | | | | | | | | | |
| Clinical features | | | | | | | | | |
| Fever | 1056 (34.7) | 602 (29.2) | 454 (45.2) | 72.385 | **< 0.001** | 792 (32.5) | 264 (43.8) | 27.168 | **< 0.001** |
| Headache | 318 (10.4) | 158 (7.7) | 160 (15.9) | 47.864 | **< 0.001** | 240 (9.8) | 78 (12.9) | 4.92 | 0.027 |
| Pectoralgia | 263 (8.6) | 165 (8.1) | 98 (9.7) | 2.312 | 0.128 | 190 (7.8) | 73 (12.1) | 11.363 | **< 0.001** |
| Cough and/or expectoration | 781 (25.7) | 462 (22.7) | 319 (31.8) | 28.896 | **< 0.001** | 591 (24.2) | 190 (31.5) | 13.348 | **< 0.001** |
| Hemoptysis | 110 (3.6) | 72 (3.5) | 38 (3.7) | 0.116 | 0.734 | 84 (3.4) | 26 (4.3) | 1.0371 | 0.309 |
| Dyspnoea | 491 (16.1) | 288 (14.1) | 203 (20.2) | 18.222 | **< 0.001** | 347 (14.2) | 144 (23.9) | 33.195 | **< 0.001** |
| Pleural effusion | 173 (5.7) | 75 (3.6) | 98 (9.7) | 46.178 | **< 0.001** | 113 (4.6) | 60 (9.9) | 25.432 | **< 0.001** |
| Nausea and vomiting | 148 (4.8) | 77 (3.7) | 71 (7.0) | 15.664 | **< 0.001** | 109 (4.4) | 39 (6.4) | 4.155 | 0.042 |
| Diarrhoea or abdominal discomfort | 115 (3.7) | 53 (2.6) | 62 (6.1) | 23.519 | **< 0.001** | 75 (3.0) | 40 (6.6) | 16.795 | **< 0.001** |
| Anorexia | 117 (3.8) | 62 (3.0) | 55 (5.4) | 10.719 | **< 0.001** | 86 (3.5) | 31 (5.1) | 3.395 | 0.065 |
| Delirium | 287 (9.4) | 116 (5.7) | 171 (17.0) | 100.861 | **< 0.001** | 184 (7.5) | 103 (17.1) | 51.372 | **< 0.001** |
| Brain abscess | 443 (14.5) | 162 (7.9) | 281 (28.0) | 216.417 | **< 0.001** | 305 (12.5) | 138 (22.9) | 41.773 | **< 0.001** |
| Weakness | 297 (9.7) | 148 (7.2) | 149 (14.8) | 43.613 | **< 0.001** | 201 (8.2) | 96 (15.9) | 32.288 | **< 0.001** |
| Eye symptoms | 367 (12.0) | 266 (13.0) | 101 (10.0) | 5.765 | 0.016 | 324 (13.3) | 43 (7.1) | 17.306 | **< 0.001** |
| Skin symptoms | 740 (24.3) | 450 (22.1) | 290 (28.9) | 16.682 | **< 0.001** | 635 (26.0) | 105 (17.4) | 19.619 | **< 0.001** |
| Node | 389 (12.8) | 214 (10.5) | 175 (17.4) | 28.746 | **< 0.001** | 320 (13.1) | 69 (11.4) | 1.235 | 0.267 |
| Epilepsy | 110 (3.6) | 40 (1.9) | 70 (6.9) | 48.266 | **< 0.001** | 84 (3.4) | 26 (4.3) | 1.037 | 0.309 |
| Hemiplegia | 105 (3.4) | 55 (2.7) | 50 (4.9) | 10.437 | 0.001 | 71 (2.9) | 34 (5.6) | 10.766 | 0.001 |
| Loss of weight | 164 (5.4) | 91 (4.4) | 73 (7.2) | 10.298 | 0.001 | 126 (5.2) | 38 (6.3) | 1.213 | 0.271 |
| Previous medication types | | | | | | | | | |
| Corticosteroids use | 961 (31.7) | 544 (26.8) | 417 (41.6) | 68.012 | **< 0.001** | 719 (29.6) | 242 (40.2) | 25.284 | **< 0.001** |
| Immunosuppressants use [c] | 1584 (52.2) | 902 (44.4) | 682 (68.0) | 149.974 | **< 0.001** | 1177 (48.4) | 407 (67.6) | 71.532 | **< 0.001** |
| Chemotherapy drugs/radiation | 233 (7.9) | 122 (6.0) | 111 (11.1) | 24.286 | **< 0.001** | 134 (5.51) | 99 (16.45) | 81.455 | **< 0.001** |
| Calcineurin inhibitors | 153 (5.0) | 79 (3.9) | 74 (7.4) | 17.087 | **< 0.001** | 119 (4.89) | 34 (5.65) | 0.577 | 0.447 |
| Treatment drugs | | | | | | | | | |
| Trimethoprim sulfamethoxazole | 1818 (59.90) | 1162 (57.19) | 656 (65.40) | 18.885 | **< 0.001** | 1498 (61.57) | 320 (53.16) | 14.224 | **< 0.001** |
| Continued | | | | | | | | | |

| Characteristic | Total (N = 3035) | Single-site infection (N = 2032) | Disseminated infection (N = 1003) | $\chi^2$ | *P*-value | Survival (N = 2433) | Deceased(N = 602) | $\chi^2$ | *P*-value |
|---|---|---|---|---|---|---|---|---|---|
| Imipenem | 453 (14.93) | 237 (11.66) | 216 (21.54) | 51.540 | < **0.001** | 335 (13.77) | 118 (19.60) | 12.928 | < **0.001** |
| Meropenem | 261 (8.60) | 128 (6.30) | 133 (13.26) | 41.398 | < **0.001** | 209 (8.59) | 52 (8.64) | 0.001 | 0.970 |
| Amikacin | 508 (16.74) | 296 (14.57) | 212 (21.14) | 20.797 | < **0.001** | 410 (16.85) | 98 (16.28) | 0.114 | 0.736 |
| Cefazolin | 359 (11.83) | 196 (9.65) | 163 (16.25) | 28.095 | < **0.001** | 275 (11.30) | 84 (13.95) | 3.251 | 0.071 |
| Clarithromycin | 48 (1.58) | 29 (1.43) | 19 (1.89) | 0.942 | 0.332 | 39 (1.60) | 9 (1.50) | 0.036 | 0.849 |
| Linezolid | 273 (9.00) | 135 (6.64) | 138 (13.76) | 41.529 | < **0.001** | 220 (9.04) | 53 (8.80) | 0.034 | 0.855 |
| Ciprofloxacin | 132 (4.35) | 83 (4.08) | 49 (4.89) | 1.035 | 0.309 | 106 (4.36) | 26 (4.32) | 0.002 | 0.968 |
| Moxifloxacin | 83 (2.73) | 48 (2.36) | 35 (3.49) | 3.208 | 0.073 | 75 (3.08) | 8 (1.33) | 5.580 | **0.018** |
| Minocycline | 240 (7.91) | 160 (7.87) | 80 (7.98) | 0.010 | 0.922 | 200 (8.22) | 40 (6.64) | 1.646 | 0.200 |
| Amoxicillin clavulanic acid | 143 (4.71) | 93 (4.58) | 50 (4.99) | 0.249 | 0.618 | 131 (5.38) | 12 (1.99) | 12.360 | < **0.001** |
| Antibiotic treatment | | | | 105.055 | < **0.001** | | | 99.595 | < **0.001** |
| Single | 985 ( 32.5) | 778 (38.3) | 207 (20.6) | | | 834 (34.3) | 151 (25.1) | | |
| Combination | 1681 (55.4) | 1005 (49.5) | 676 (67.4) | | | 1337 (55.0) | 344 (57.1) | | |
| None | 61 (2.0) | 37 (1.8) | 24 (2.4) | | | 20 (0.8) | 41 (6.8) | | |
| Unknown | 308 (10.1) | 212 (10.4) | 96 (9.6) | | | 242 (9.9) | 66 (11.0) | | |
| Operative treatment | | | | 2.381 | 0.123 | | | 3.338 | 0.068 |
| Yes | 800 (26.4) | 518 (25.5) | 282 (28.1) | | | 659 (27.1) | 141 (23.4) | | |
| No | 2235 (73.6) | 1514 (74.5) | 721 (71.9) | | | 1774 (72.9) | 461 (76.6) | | |
| Mean duration of treatment (days)[d] | 218.5 ± 170.8 | 185.9 ± 168.5 | 265.1 ± 169.8 | − 7.350 | < **0.001** | | | | |
| Outcome | | | | 131.210 | < **0.001** | | | | |
| Deceased | 602 (19.8) | 285 (14.0) | 317 (31.6) | | | | | | |
| Survival | 2433 (80.2) | 1747 (86.0) | 686 (68.4) | | | | | | |

**Table 1**. Characteristics of 3035 patients with nocardiosis. The Chi-square test was used for categorical variables, the *t*-test was used for continuous variables. Results are expressed as number (percent) or mean ± standard deviation (Mean ± SD). [a]Autoimmune diseases, including systemic lupus erythematosus, Sjogren's syndrome, rheumatoid arthritis, scleroderma, polyarteritis nodosa and Wegener granulomatosis and so on. [b]Species names are reported as originally described in the source literature, without grouping into species complexes. However, some listed species belong to recognized complexes based on current taxonomic understanding. For example: The *N. nova* complex includes *N. nova*, *N. veterana*, and *N. elegans*. The *N. transvalensis* complex includes *N. transvalensis* and *N. wallacei*. The *N. abscessus* complex includes *N. abscessus*, and *N. asiatica*. [c]Immunosuppressants, including cancer chemotherapy, calcineurin inhibitors and other immunosuppressants. [d]Only 1035 patients have information on the duration of treatment.

In the training set, the RF model performed the best in terms of discrimination and calibration, with an AUC of 0.822 and a calibration score of 0.136, followed by the SGBT model (Fig. 3A,C). The RF model had the highest accuracy (0.833), sensitivity (0.996), specificity (0.176), NPV(0.914), and F1 score (0.905) (Table S4). But in the test set, the SGBT model had the best performance, with an AUC of 0.756 and a calibration degree of 0.136, followed by XGB (Fig. 3B,D). The SGBT model had the highest accuracy (0.819), F1 score (0.897), as well as good sensitivity (0.988), specificity (0.133), precision (0.822), NPV(0.727) (Table S4). This indicates that the SGBT model has good generalization ability, maintaining stable performance on unseen data. In summary, the SGBT model performed the best in both the training set and the test set and is thus recommended as the preferred model for the prediction of the risk of mortality in patient with nocardiosis (Fig. S3), followed by the XGB model.

## Model explanation

To increase clinical interpretability, we used the SHAP method to explain the output of the final model by calculating the contribution of each variable to the prediction. This explainable method provides two types of explanations: a global explanation of the model at the feature level and a local explanation at the individual level. The global explanation describes the overall functionality of the model. As shown in the SHAP summary bar plot (Fig. 4A), the contribution of features to the model was evaluated using the mean SHAP values, which were displayed in descending order: disseminated infection, age, skin infection, use immunosuppressants and delirium were the five most important features in the prediction model. In addition, the SHAP summary dot plot (Fig. 4B) visually shows the direction and strength of the influence of each feature on the model prediction: features such as disseminated infection, advanced age, use immunosuppressants, delirium, use chemotherapy drugs/radiation, have cancer, chronic respiratory diseases and hepatopathy significantly increased the risk of mortality. In addition, local explanations help us make personalized treatment choices by calculating and displaying the contribution of each feature to the prediction results of a single patient. The SHAP waterfall plot (Fig. 4C) illustrates the contribution of each feature to the model's prediction for a patient with nocardiosis. The specific value of each

| Variable | Univariate analysis OR (95% CI, *P*-value) | Multivariate analysis OR (95% CI, *P*-value) |
|---|---|---|
| Age (years) | 1.01 (1.01–1.01, $P<0.01$)** | 1.01 (1.01–1.02, $P<0.01$)** |
| Gender | | |
| Female | Refer | Refer |
| Male | 1.27 (1.05–1.53, $P=0.02$)* | 1.20 (0.99–1.46, $P=0.07$) |
| Underlying diseases | | |
| Cancer | 2.77 (2.20–3.48, $P<0.01$)** | 2.45 (1.93–3.11, $P<0.01$)** |
| Organ/stem cell transplant | 1.37 (1.09–1.71, $P=0.01$)* | 1.16 (0.92–1.47, $P=0.22$) |
| Diabetes mellitus | 1.32 (0.99–1.75, $P=0.06$) | 1.34 (0.99–1.80, $P=0.06$) |
| Cardiovascular diseases | 1.38 (1.04–1.82, $P=0.03$)* | 1.18 (0.89–1.58, $P=0.26$) |
| HIV [a] | 1.36 (0.96–1.94, $P=0.09$) | 1.34 (0.93–1.95, $P=0.12$) |
| Renal disease | 1.14 (0.88–1.48, $P=0.33$) | 1.18 (0.90–1.55, $P=0.23$) |
| Hepatopathy | 2.32 (1.52–3.55, $P<0.01$)** | 2.27 (1.47–3.51, $P<0.01$)** |
| Chronic respiratory diseases | 1.45 (1.11–1.89, $P=0.01$)* | 1.34 (1.02–1.77, $P=0.04$)* |
| Autoimmune diseases [b] | 0.96 (0.74–1.23, $P=0.74$) | 1.02 (0.79–1.33, $P=0.88$) |
| Multidrug resistant | | |
| Yes | 1.33 (0.99–1.77, $P=0.05$) | 1.24 (0.92–1.67, $P=0.16$) |
| No | 1.03 (0.81–1.30, $P=0.84$) | 1.05 (0.82–1.34, $P=0.71$) |
| Unknown | 0.86 (0.71–1.05, $P=0.15$) | 0.88 (0.71–1.08, $P=0.21$) |
| TMP-SMX resistance | | |
| Yes | 1.32 (0.86–2.02, $P=0.21$) | 1.38 (0.89–2.14, $P=0.15$) |
| No | 1.07 (0.84–1.35, $P=0.59$) | 1.02 (0.80–1.29, $P=0.89$) |
| Unknown | 0.89 (0.72–1.10, $P=0.27$) | 0.91 (0.73–1.14, $P=0.42$) |
| Type of infection | | |
| Single-site infection | refer | refer |
| Disseminated Infection | 2.83 (2.36–3.40, $P<0.01$)** | 2.68 (2.22–3.24, $P<0.01$)** |
| Infection site | | |
| Pulmonary infection | 2.30 (1.89–2.79, $P<0.01$)** | 2.20 (1.81–2.66, $P<0.01$)** |
| Skin infection | 0.67 (0.54–0.82, $P<0.01$)** | 0.66 (0.53–0.82, $P<0.01$)** |
| Brain infection | 2.36 (1.96–2.84, $P<0.01$)** | 2.31 (1.90–2.83, $P<0.01$)** |
| Eye infection | 0.35 (0.24–0.52, $P<0.01$)** | 0.40 (0.26–0.60, $P<0.01$)** |
| Bacteremia | 2.58 (1.87–3.56, $P<0.01$)** | 2.40 (1.73–3.34, $P<0.01$)** |
| Clinical features | | |
| Hemoptysis | 1.26 (0.81–1.98, $P=0.31$) | 1.25 (0.78–1.99, $P=0.36$) |
| Dyspnoea | 1.89 (1.52–2.35, $P<0.01$)** | 1.89 (1.51–2.37, $P<0.01$)** |
| Delirium | 2.52 (1.95–3.27, $P<0.01$)** | 2.26 (1.73–2.96, $P<0.01$)** |
| Brain abscess | 2.08 (1.66–2.60, $P<0.01$)** | 1.90 (1.50–2.41, $P<0.01$)** |
| Epilepsy | 1.26 (0.81–1.98, $P=0.31$) | 1.16 (0.74–1.84, $P=0.52$) |
| Hemiplegia | 1.99 (1.31–3.03, $P<0.01$)** | 2.05 (1.33–3.15, $P<0.01$)** |
| Loss of weight | 1.23 (0.85–1.79, $P=0.27$) | 1.09 (0.75–1.60, $P=0.65$) |
| Previous medication types | | |
| Corticosteroids use | 1.60 (1.33–1.93, $P<0.01$)** | 1.54 (1.27–1.86, $P<0.01$)** |
| Immunosuppressants use[c] | 2.23 (1.85–2.69, $P<0.01$)** | 2.11 (1.74–2.56, $P<0.01$)** |
| Chemotherapy drugs/radiation | 3.38 (2.56–4.46, $P<0.01$)** | 2.93 (2.20–3.89, $P<0.01$)** |
| Calcineurin inhibitors | 1.16 (0.79–1.72, $P=0.45$) | 1.16 (0.78–1.74, $P=0.47$) |

**Table 2**. Risk factors associated with all-cause mortality in 3035 patients with nocardiosis. Adjusted for age, gender, and geographical distribution. *OR*, Odds ratio; *CI*, Confidence interval; *P*-value < 0.05 considered significant. *$P<0.05$, **$P<0.01$. [a]HIV, human immunodeficiency virus. [b]Autoimmune diseases, including systemic lupus erythematosus, Sjogren's syndrome, rheumatoid arthritis, scleroderma, polyarteritis nodosa and Wegener granulomatosis and so on. [c]Immunosuppressants, including cancer chemotherapy, calcineurin inhibitors and other immunosuppressants.

feature and its corresponding SHAP value in the plot indicate the positive and negative impact of the feature on the prediction result. The plot analyzes the case of a 75-year-old, immunosuppressed patient with disseminated infection. Disseminated infection, using immunosuppressants and age contributed + 0.166, + 0.041, and + 0.031 to the positive prediction results, respectively; eye infection contributed − 0.095 to the negative prediction

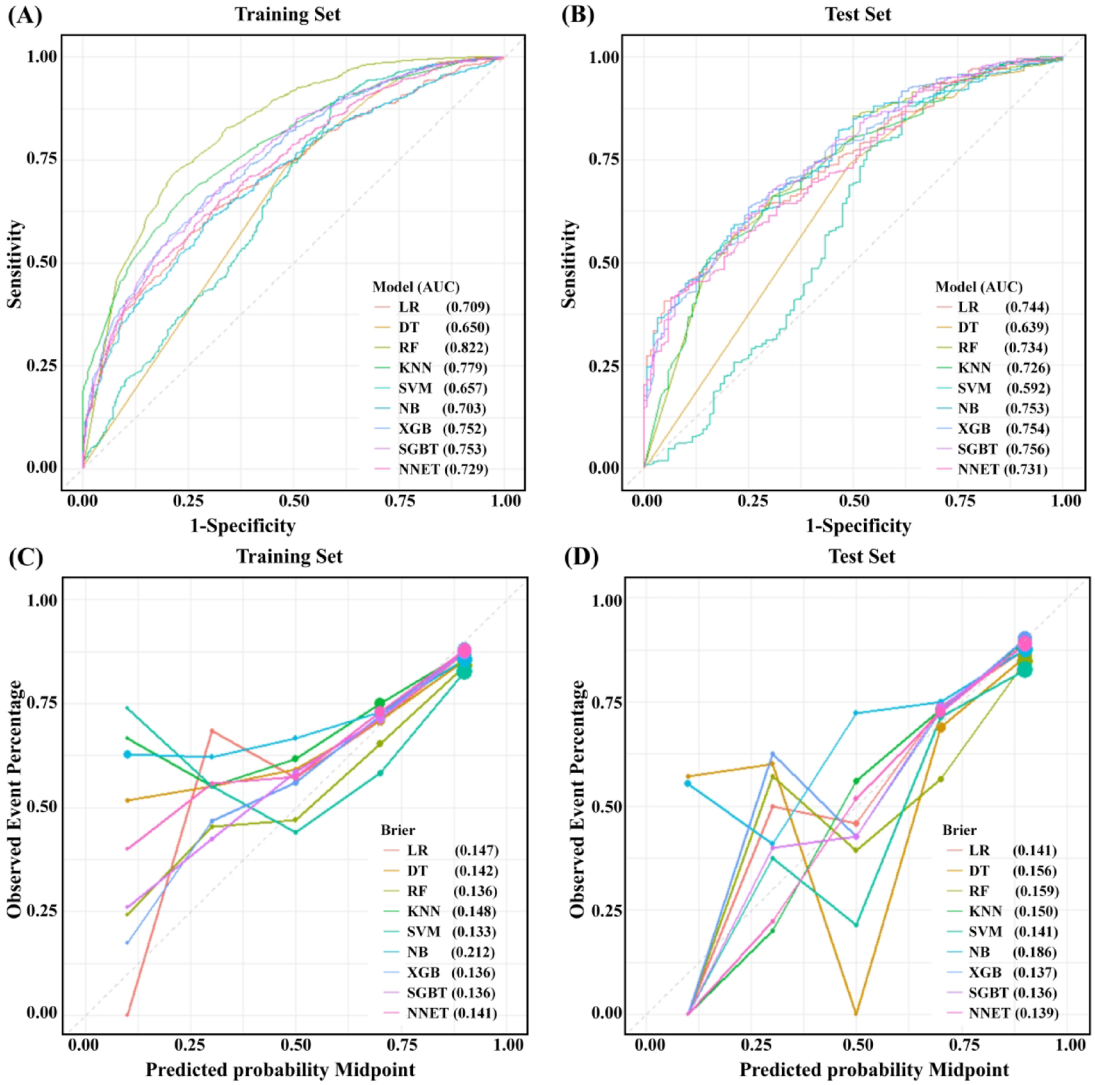


**Fig. 3**. Performance of machine learning (ML) models in predicting mortality in patients with nocardiosis in the training and test sets. Receiver operating characteristic (ROC) curve analysis is shown in (**A**) and (**B**), and calibration curve analysis in (**C**) and (**D**). Abbreviations: LR: logistic regression; DT: decision tree; RF: random forest; SVM: support vector machine; KNN: K-nearest neighbors; NB: naive bayes; XGB: extreme gradient boosting; SGBT: stochastic gradient boosting; NNET: neural network.

results. Other features, such as cancer, chemotherapy drugs/radiation, hepatopathy, etc., also had different degrees of influence on the prediction results. This allows clinicians to understand why an elderly patient is classified as high-risk and may prompt earlier intensive monitoring, adjustment of immunosuppressive therapy, earlier initiation of combination antibiotic therapy, consideration of longer treatment durations or adjunctive interventions. By accumulating SHAP values, the waterfall plot visually demonstrates the formation process of the prediction result for a specific patient, helping clinicians understand the importance of different variables and make personalized treatment decisions based on the patient's condition. In addition, the SHAP dependence plot helps to understand how individual features affect the output of the prediction model. Figure 4D compares the actual values and the SHAP values of these 10 features, where features with SHAP values greater than zero correspond to positive predictions in the model.

## Diagnosis and species distribution of nocardiosis

In the majority of cases, nocardiosis was diagnosed based on distinctive histopathological features and culturing results. Imaging diagnosis was conducted on 1,437 cases. The molecular diagnosis was established in 712 cases by 16S RNA sequencing methods, Polymerase Chain Reaction (PCR), Metagenomics Next Generation Sequencing (mNGS), and Matrix-Assisted Laser Desorption/Ionization-Time of Flight (MALDI-TOF). Sample sources included skin and soft tissue pus (20.9%), sputum (19.6%), brain abscess or cerebrospinal fluid (15.3%), bronchoalveolar lavage fluid (BALF) (13.5%), and other (Table S1). Out of 3,035 cases, species identification was made in 2,379 cases. *N. asteroides* (28.5%) was the most common species, followed by *N. farcinica* (13.9%), *N.*

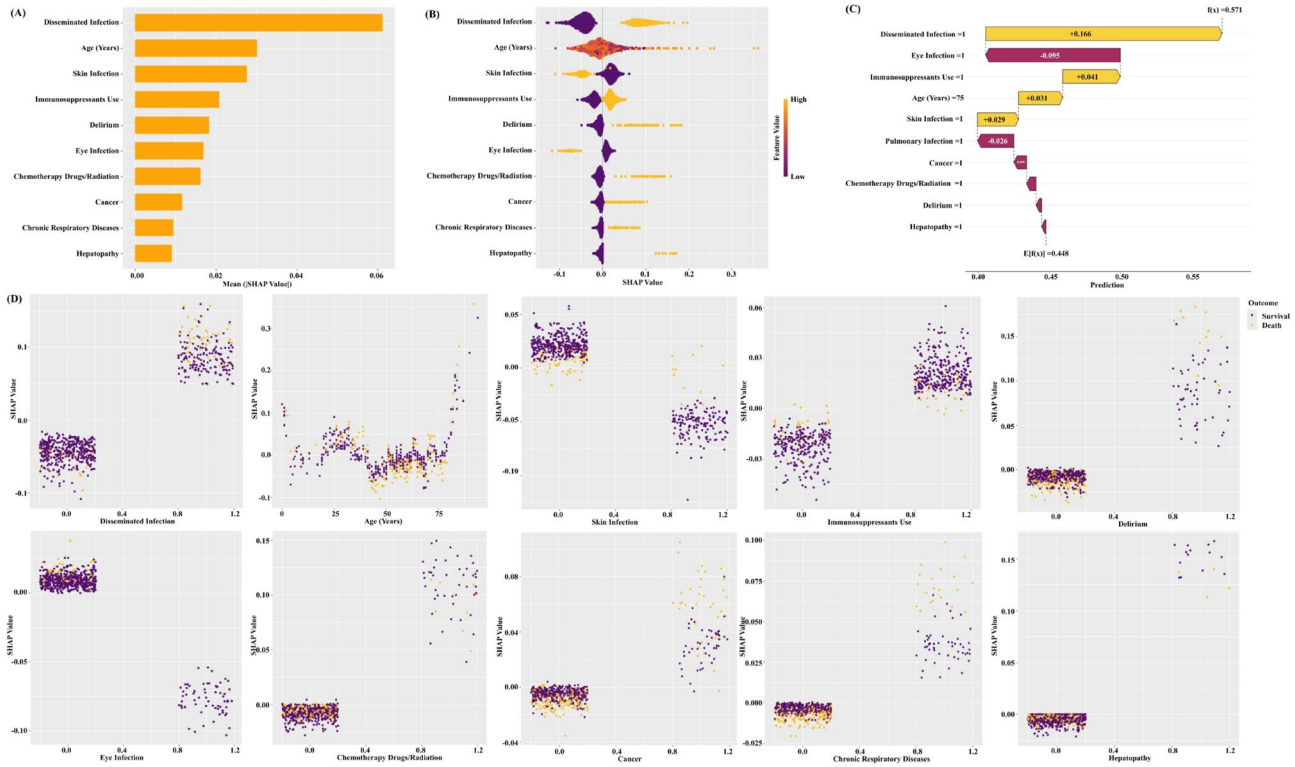


**Fig. 4**. Global and local model explanation by the SHapley Additive exPlanation (SHAP) method. (**A**) SHAP summary bar plot. This plot evaluates the contribution of each feature to the model using mean SHAP values, displayed in descending order. (**B**) SHAP summary dot plot. The probability of mortality increases with the SHAP values of the features. Each dot represents a patient's SHAP value for a given feature, with orange indicating higher feature values and purple indicating lower values. Dots are stacked vertically to show density. (**C**) SHAP waterfall plot. This plot shows the contribution of each feature to the prediction result of a particular patient using the Stochastic Gradient Boosting Tree (SGBT) model. Orange bars indicate features that contribute positively to the prediction, while purple bars indicate negative contributions. Feature values are shown alongside their SHAP values, highlighting key features such as disseminated infection (+0.166), eye infection (− 0.095), immunosuppressants use (+0.041), age (years) = 75 (+0.031), skin infection (+0.029) and pulmonary infection (− 0.026). The overall contribution is 0.571, with a baseline contribution of 0.448. (**D**) SHAP dependence plot. Each dependence plot shows how a single feature affects the model's output, with each point representing a patient. SHAP values are on the y-axis, and actual feature values are on the x-axis. Features with SHAP values above zero push the decision towards the "mortality" class.

*brasiliensis* (9.1%), and *N. cyriacigeorgica* (5.6%) (Table 1). Particularly, the reported prevalence of *N. asteroides* in earlier studies is likely inflated and should be interpreted with caution. Additionally, the distribution of species varies among different infection types. In general, *N. asteroides* and *N. farcinica* were the most common isolates causing pulmonary infection, disseminated infection, brain/CNS infection, and eye infection, but *N. brasiliensis* (37.3%) was the most frequent isolates causing skin infection (Fig. S4).

## Antimicrobial susceptibility of *Nocardia* isolates

Antimicrobial susceptibility data were available for 772 *Nocardia* isolates, determined using either the paper disk diffusion or broth microdilution method (Table S5). MDR was observed in 36.4% of isolates (281/772). Susceptibility rates for key antimicrobials were as follows: linezolid (LNZ), 99.4%; AMK, 93.9%; TMP-SMX, 81.4%; IMP, 80.9%; minocycline (MINO), 73.1%; moxifloxacin (MXF), 70.8%; CRO, 67.7%; doxycycline (DOXY), 65.8%; tigecycline (TYGACIL), 65.4%; cefoxitin (CFX), 61.0%; ciprofloxacin (CPFX), 59.7%; cefepime (FEP), 58.4%; amoxicillin-clavulanic acid (AMC), 56.9%; tobramycin (TOB), 54.4%; and clarithromycin (CLA), 45.2%. Susceptibility profiles varied across *Nocardia* species. Most isolates exhibited high susceptibility to LNZ and AMK but lower susceptibility to FEP, CPFX, and CLA. Notably, *N. cyriacigeorgica* (35.0%), *N. nova* (18.8%), and *N. otitidiscaviarum* (16.7%) demonstrated relatively low susceptibility to AMC, while *N. farcinica* (18.6%) and *N. nova* (31.3%) showed reduced susceptibility to TOB. In contrast, *N. cyriacigeorgica* (100%) and *N. brasiliensis* (75.9%) were highly susceptible to TOB. However, interpretation of susceptibility trends should be made with caution due to heterogeneity in testing methods.

## Subgroup analysis

The results showed that the selection of a particular study (cases from 2005 to 2024) does not significantly alter the overall results, indicating that an evaluation of the results had considerable stability (Table S6, S7).

## Discussion

This systematic review synthesized data from 9,750 cases of nocardiosis across 2,077 studies, spanning 76 countries and regions and encompassing 79 clinically relevant *Nocardia* species. This study provides a comprehensive overview for the global epidemiology of nocardiosis, including the distribution and characteristics of clinical isolates. Potential risk factors were extensively studied, and mortality was predicted using multiple ML methods. These findings lay the foundation for understanding the disease and provide a reference for early diagnosis and personalized treatment of the disease.

The epidemiology of nocardiosis and the distribution of *Nocardia* species exhibit distinct geographical patterns. Cases were predominantly reported from North America (34.3%) and Asia (34.2%), followed by Europe (22.9%), Oceania (4.2%), and South America (2.9%), with notably fewer cases from African regions (1.5%). This apparent scarcity of data from Africa does not necessarily indicate a low prevalence of *Nocardia* but may reflect limited socioeconomic development and weaker healthcare infrastructure, potentially leading to underdiagnosis or underreporting[15]. Fourteen countries, including the United States, China, and Canada, each reported over 100 cumulative cases of nocardiosis. Species such as *N. asteroides*, *N. farcinica*, and *N. cyriacigeorgica* appear widely distributed across multiple countries and regions[7,16,17], though their relative proportions vary significantly. *N. asteroides* predominates in North American and Southern European countries (e.g., the United States, Canada, Spain, and the United Kingdom), whereas *N. farcinica* is more prevalent in Asian and Western European countries (e.g., China, Japan, South Korea, France, and Germany). Notably, the high proportion of *N. asteroides* may reflect the limitations of historical taxonomic classification of *Nocardia* and should be interpreted with caution. In contrast, *N. brasiliensis* is primarily reported in Latin America and parts of Asia. These differences may be influenced by environmental factors such as climate and soil pH; for instance, *N. brasiliensis* has been associated with tropical environments[18], while *N. asteroides* and *N. brasiliensis* are often isolated from soils with pH ranges of 7.0–8.0 and 8.0–9.0, respectively[19]. However, these ecological associations remain unconfirmed in human infections. Additionally, our analysis revealed a significant increase in reported nocardiosis cases over time, alongside a notable decrease and stabilization in the all-cause mortality rate since 2000. These are likely due to a combination of improved diagnostics, changes in population susceptibility, environmental factors, and better clinical management. Concurrently, the identification of new clinical *Nocardia* species has steadily risen, consistent with updates in the List of Prokaryotic Names with Standing in Nomenclature (LPSN) (http://www.bacterio.net/genus/nocardia). This trend may benefit from advancements in molecular and genomic techniques, improved taxonomic classification, increased clinical awareness and testing and data sharing.

In this study, approximately 65% of nocardiosis cases were male, consistent with findings from previous research[20,21]. This gender disparity may be associated with differences in exposure to risk factors, such as occupational or environmental hazards, though the precise mechanisms remain unclear. Nocardiosis cases spanned all age groups, a pattern that may also reflect varying risk profiles across the population. However, the absence of data from healthy controls in this study limits our ability to quantify the relative risk of nocardiosis in the general population. Additionally, our analysis revealed that cases from temperate and maritime climate zones exhibited significantly higher rates of disseminated infection and mortality compared to other regions ($P < 0.05$). A similar trend was observed among cases involving corticosteroids or immunosuppressants, corroborating prior studies[22]. In contrast, some earlier research suggests that nocardiosis is less prevalent in temperate climates[23], while other studies link warm, humid conditions—such as those in maritime climates—to increased rates of nocardial keratitis and actinomycetoma[24,25]. These maritime environments, characterized by stable temperatures and high soil moisture, may favor the survival and proliferation of *Nocardia* species[26], though this hypothesis remains debated and requires further investigation.

In this study, pulmonary nocardiosis emerged as the most common clinical manifestation, with cancer, organ or stem cell transplantation, hepatopathy, and chronic respiratory diseases identified as the predominant underlying conditions. Fever was the most frequent clinical presentation (34.7%), followed by cough and/or expectoration (25.7%), skin symptoms (24.3%), dyspnea (16.1%), brain abscess (14.5%), node (12.8%), eye symptoms (12.0%), and headache (10.4%), consistent with prior studies[9,27,28]. Notably, the reported frequency of brain abscesses may be underestimated, potentially influenced by diagnostic limitations, heterogeneity in evaluation approaches, and inherent constraints of retrospective data collection from the literature. Disseminated *Nocardia* infection was independently associated with an increased risk of mortality, with a all-cause mortality rate of 31.7%, a finding that underscores its severity. Untreated or inadequately managed single-site infections, such as pulmonary or cutaneous nocardiosis, may progress to disseminated disease, significantly elevating mortality risk[4]. Similarly, brain or CNS infections and bacteremia exhibited high mortality rates of 30.4% and 37.3%, respectively, highlighting their critical nature. Particularly, disseminated *Nocardia* infection and CNS infection strongly linked to immunosuppression, healthy individuals are rarely affected. Although skin and eye infections were associated with lower mortality, they can infect healthy individuals.especially those engaged in gardening and agricultural production. Primary skin and eye nocardiosis is a localized skin/eye infection caused by *Nocardia* through direct inoculation (e.g., trauma), presenting as pustules, nodules, abscesses, or keratitis. It responds well to 3–6 months of antibiotics (e.g., TMP-SMX) with a favorable prognosis. However, if the host's immune response is compromised or treatment is delayed, it may also have serious consequences[1,4]. Patients infected with *N. asteroides* (26.6%) and *N. farcinica* (23.4%) exhibited higher mortality compared to those infected with other *Nocardia* species. While *N. farcinica* has been previously associated with disseminated infections and increased one-year mortality[9,29], potentially due to inherent virulence factors, whereas similar evidence for *N. asteroides* is lacking. Specifically, the observed higher mortality in *N. asteroides* cases should be

interpreted with caution, given the historical taxonomic ambiguities and potential misidentification associated with this species. Monitoring inflammatory markers and vital signs, such as neutrophil percentage, CRP levels, WBC count, respiratory rate, and body temperature—is critical in managing nocardiosis. Antimicrobial therapy was identified as a protective factor, with combination regimens demonstrating greater efficacy than monotherapy.

Our analysis revealed MDR rate of 36.4% among *Nocardia* isolates, posing challenges for treatment selection. Susceptibility profiles varied significantly across species[25]. Most isolates exhibited high susceptibility to LNZ (99.4%), AMK (93.9%), and TMP-SMX (81.4%), but lower susceptibility to FEP (58.4%), CPFX (59.7%), and CLA (45.2%). Specifically, *N. cyriacigeorgica*, *N. nova*, and *N. otitidiscaviarum* showed reduced susceptibility to AMC, corroborating earlier findings[25]. These results emphasize the importance of species-level identification and susceptibility testing to guide antimicrobial selection. In the absence of timely species identification or susceptibility data, empirical treatment with LNZ, AMK, and TMP-SMX may be warranted. It should be noted that variability in susceptibility testing approaches across studies may affect the reliability of aggregated resistance data. Differences in methodology, particularly between older and more recent reports, could have introduced inconsistencies that limit the precision of resistance rate estimates. Therefore, these findings should be interpreted with caution. Furthermore, among survival cases, the mean treatment duration was 185.9 days (SD 168.5) for single-site infections and 265.1 days (SD 169.8) for disseminated infections, which may be related to bacterial load and recurrence risk, indicating that prolonged therapy for disseminated infection is recommended to ensure complete clearance of pathogens[30].

To our knowledge, this is the first ML model based on large-sample case or series report data from systematic review predicting the risk of mortality in patients with nocardiosis. Multivariate logistic regression analysis revealed that age, gender, underlying diseases (cancer, organ/stem cell transplant, cardiovascular diseases, hepatopathy, chronic respiratory diseases), disseminated infection, infection site (pulmonary, skin, brain, eye, bacteremia), clinical features (dyspnoea, delirium, brain abscess, hemiplegia), previous medication types (corticosteroids, immunosuppressants, chemotherapy drugs/radiation) were factors independently associated with the risk for mortality in patients with nocardiosis. To date, studies on predicting the risk of mortality in nocardiosis have never been reported. This is mainly because nocardiosis is a rare disease, data are often fragmented, and cases are usually in the form of case series or case reports rather than large cohort. With the emerging of ML in the field of infectious diseases, the application of ML or artificial intelligence (AI) to predict the outcomes in rare infections such as nocardiosis is a cutting-edge area[31]. This study timely fills a critical gap in the field of *Nocardia* infectious disease research and clinical decision support.

Among the nine machine learning models, the SGBT model had the highest AUC, accuracy, and F1 score in test set. The SGBT algorithm can improve the prediction accuracy and stability of the model by building an ensemble of decision trees sequentially, where each tree corrects errors of the previous ones, optimizing a loss function using gradient descent. Several studies have shown that the SGBT method is very valuable for predictive models in the medical field[32,33]. In this study, we used the SGBT algorithm to build a final model with 10 predictors. These predictors underscore the importance of host immunocompromise, disseminated infection, and underlying medical conditions like cancer and chronic respiratory diseases as key determinants of morbidity and mortality in *Nocardia* infections. These features are obvious characteristics of *Nocardia* patients and can be easily obtained clinically, making them an effective tool for effectively predicting the risk of death in *Nocardia* patients. Another major advantage of the present study is that we used the SHAP method to reasonably elucidate the "black box" of the ML model. The SHAP method clarifies the functionality of the model by providing global and local explanations, detailing how the personalized input data are used to make specific predictions for individual patients.

This study represents the first systematic review of reported cases to comprehensively evaluate the global epidemiology of nocardiosis and used ML models in mortality prediction, offering a viable approach for rare disease research and personalized medicine. However, several limitations must be acknowledged. Firstly, data were collected over an extended period, resulting in heterogeneity among cases and risk factors due to variations in diagnostic practices, reporting standards, and patient populations. To enhance the reliability of our findings, we applied stringent inclusion and exclusion criteria and employed robust statistical methods, including logistic regression and subgroup analyses. Nevertheless, the risk factor analysis excluded numerous cases due to incomplete information on treatment and follow-up, limiting the generalizability of these results. Consequently, extrapolations from this study should be interpreted with caution. In addition, the exclusive use of published case reports and series introduces a risk of publication bias, as severe or atypical cases are more likely to be reported. This may lead to an overestimation of mortality rates and the apparent strength of certain risk factors. Consequently, extrapolations from this study should be interpreted with caution. Secondly, case reports often suffer from publication bias and heterogeneity, which can affect model performance. In the future, governments and health departments should be considered in monitoring nocardiosis, improving disease reporting systems and facilitating more accurate assessments of its burden. Thirdly, another limitation of our study lies in the assignment of geographic climate zones based solely on the reported locations of clinical isolates and cases, without considering potential temporal changes in climate classifications over the extended study period (1950–2024). Climate zones may have shifted due to global climate change and other environmental factors, which we did not account for in this analysis. Consequently, this may introduce some bias when interpreting geographic trends. Future studies could benefit from incorporating historical climate data to more accurately assess the impact of climate variation on the distribution of isolates. Fourthly, all model evaluations were performed using internal validation methods, and no external test cohort was available. As such, the generalizability of the predictive model remains to be verified. Future studies should focus on prospective data collection and external validation in independent cohorts to confirm the model's performance across different populations and clinical settings.

## Conclusion

In conclusion, this study provided a comprehensive overview of the global epidemiology of nocardiosis and clinical isolates distribution. The data determined some risk factors and successfully developed an explainable ML model to predict the risk of mortality in patients with nocardiosis based on case reports. In addition, given their consistently high susceptibility rates, LNZ (99.4%), AMK (93.9%), and TMP-SMX (81.4%) are recommended for empirical treatment of suspected *Nocardia* infections pending species identification and susceptibility testing.

## Data availability

The data that support the findings of this study are available from the corresponding author, Zhenjun Li, upon reasonable request.

Received: 16 May 2025; Accepted: 27 October 2025
Published online: 26 November 2025

## References

1. Traxler, R. M. et al. Updated review on *Nocardia* Species: 2006–2021. *Clin. Microbiol. Rev.* **35**, e0002721. https://doi.org/10.1128/cmr.00027-21 (2022).
2. McNeil, M. M. & Brown, J. M. The medically important aerobic actinomycetes: Epidemiology and microbiology. *Clin. Microbiol. Rev.* **7**, 357–417. https://doi.org/10.1128/CMR.7.3.357 (1994).
3. Thipmontree, W. & Suputtamonkol, Y. Comparison of clinical outcomes of pulmonary nocardiosis between AIDS and non-AIDS patients. *BMC Infect. Dis.* **24**, 649. https://doi.org/10.1186/s12879-024-09519-2 (2024).
4. Fatahi-Bafghi, M. Nocardiosis from 1888 to 2017. *Microb. Pathog.* **114**, 369–384. https://doi.org/10.1016/j.micpath.2017.11.012 (2018).
5. Wilson, J. W. Nocardiosis: Updates and clinical overview. *Mayo Clin. Proc.* **87**, 403–407. https://doi.org/10.1016/j.mayocp.2011.11.016 (2012).
6. Coussement, J. et al. *Nocardia* infection in solid organ transplant recipients: A multicenter European case-control study. *Clin. Infect. Dis.* **63**, 338–345. https://doi.org/10.1093/cid/ciw241 (2016).
7. Minero, M. V. et al. Nocardiosis at the turn of the century. *Medicine (Baltimore)* **88**, 250–261. https://doi.org/10.1097/MD.0b013e3181afa1c8 (2009).
8. Kageyama, A. et al. Nocardial infections in Japan from 1992 to 2001. *Eur. J. Epidemiol.* **19**, 383–389. https://doi.org/10.1023/b:ejep.0000024706.02325.c0 (2004).
9. Yetmar, Z. A. et al. Mortality after nocardiosis: Risk factors and evaluation of disseminated infection. *Open Forum Infect. Dis.* **10**, ofad409. https://doi.org/10.1093/ofid/ofae122 (2023).
10. Yamagata, M. et al. Clinical characteristics of *Nocardia* infection in patients with rheumatic diseases. *Clin. Dev. Immunol.* **2013**, 818654. https://doi.org/10.1155/2013/818654 (2013).
11. Bradley, L. A., Haddow, H. R. M. & Palomaki, G. E. Treatment of mucopolysaccharidosis type II (Hunter syndrome): A systematic evidence review. *Genet. Med.* **19**, 1187–1201. https://doi.org/10.1038/gim.2017.30 (2017).
12. Alqaissi, E. Y., Alotaibi, F. S. & Ramzan, M. S. Machine-learning predictive models for diagnosing infectious diseases. *Comput. Math. Methods Med.* **2022**, 6902321. https://doi.org/10.1155/2022/6902321 (2022).
13. Page, M. J. et al. PRISMA 2020 explanation and elaboration. *BMJ* **372**, n160. https://doi.org/10.1136/bmj.n160 (2021).
14. Schlaberg, R., Fisher, M. A. & Hanson, K. E. Susceptibility profiles of *Nocardia* isolates based on current taxonomy. *Antimicrob. Agents Chemother.* **58**, 795–800. https://doi.org/10.1128/AAC.01531-13 (2014).
15. Liu, Z. et al. A worldwide epidemiological evolution of human brucellosis. *Emerg. Microbes Infect.* **13**, 2290839. https://doi.org/10.1080/22221751.2023.2290839 (2024).
16. Hashemi-Shahraki, A. et al. Genetic diversity and antimicrobial susceptibility of *Nocardia* species. *Sci. Rep.* **5**, 17862. https://doi.org/10.1038/srep17862 (2015).
17. Nenoff, P. et al. Eumycetoma and actinomycetoma: an update. *J. Eur. Acad. Dermatol. Venereol.* **29**, 1873–1883. https://doi.org/10.1111/jdv.13008 (2015).
18. Saubolle, M. A. & Sussland, D. Nocardiosis: Review of clinical and laboratory experience. *J. Clin. Microbiol.* **41**, 4497–4501. https://doi.org/10.1128/JCM.41.10.4497-4501 (2003).
19. Kachuei, R. et al. Diversity of Nocardia spp. in soil of Isfahan Iran. *Asian Pac. J. Trop. Biomed.* **2**, 474–478. https://doi.org/10.1016/S2221-1691(12)60079-3 (2012).
20. Kurahara, Y. et al. Pulmonary nocardiosis: Analysis of 59 cases. *Respir. Investig.* **52**, 160–166. https://doi.org/10.1016/j.resinv.2013.09.004 (2014).
21. Smego, R. A. & Gallis, H. A. Clinical spectrum of *Nocardia* brasiliensis. *Rev. Infect. Dis.* **6**, 164–180. https://doi.org/10.1093/clinids/6.2.164 (1984).
22. Lebeaux, D. et al. Outcome and treatment of nocardiosis after solid organ transplantation. *Clin. Infect. Dis.* **64**, 1396–1405. https://doi.org/10.1093/cid/cix124 (2017).
23. Brown-Elliott, B. A. et al. Clinical and lab features of *Nocardia* spp. *Clin. Microbiol. Rev.* **19**, 259–282. https://doi.org/10.1128/CMR.19.2.259-282.2006 (2006).
24. Stapleton, F. et al. Climate, disease severity, and microbial keratitis. *Am. J. Ophthalmol.* **144**, 690–698. https://doi.org/10.1016/j.ajo.2007.06.037 (2007).
25. Bonifaz, A. et al. Mycetoma: Experience of 482 cases. *PLoS Negl. Trop. Dis.* **8**, e3102. https://doi.org/10.1371/journal.pntd.0003102 (2014).
26. Venegas, R., Acevedo, J. & Treml, E. Ocean warming impacts on marine ecosystems. *Deep Sea Res. II* **212**, 105318. https://doi.org/10.1016/j.dsr2.2023.105318 (2023).
27. Meena, D. S. et al. CNS Nocardiosis: A systematic review. *Med. Princ. Pract.* **31**, 333–341. https://doi.org/10.1159/000525509 (2022).
28. Yang, J. et al. Clinical characteristics, susceptibility, and treatment of nocardiosis. *Int. J. Infect. Dis.* **130**, 136–143. https://doi.org/10.1016/j.ijid.2023.02.023 (2023).
29. Soueges, S. et al. Prognosis of disseminated nocardiosis: multicenter study. *J. Infect.* **85**, 130–136. https://doi.org/10.1016/j.jinf.2022.05.029 (2022).
30. Margalit, I. et al. How do I manage nocardiosis?. *Clin. Microbiol. Infect.* **27**, 550–558. https://doi.org/10.1016/j.cmi.2020.12.019 (2021).
31. Peiffer-Smadja, N. et al. Machine learning for infectious diseases. *Clin. Microbiol. Infect.* **26**, 584–595. https://doi.org/10.1016/j.cmi.2019.09.009 (2020).
32. Zhang, Z. et al. Predictive analytics with gradient boosting. *Ann. Transl. Med.* **7**, 152. https://doi.org/10.21037/atm.2019.03.29 (2019).

33. Thadani, S. et al. ICU mortality prediction in children using ML. *Crit. Care Explor.* **6**, e1188. https://doi.org/10.1097/CCE.0000000000001188 (2024).

## Acknowledgements
We are grateful to all authors who have previously published data on *Nocardia* worldwide.

## Author contributions
B.D. and Z.S. performed the data collection and processed and drafted the manuscript. B.D., Z.R., Zhiguo Liu, Z.S., D.T., J.S., and J.Y. participated in the design of the study. Zhenjun Li, M.Y., S.X. and X.Q. critically reviewed the manuscript. Zhenjun Li managed the project. All the authors have read and approved the final manuscript.

## Funding
This research was supported by the National Key Research and Development Program of China (2024YFC2309300) and the National Natural Science Foundation of China (82073624).

## Declarations

## Competing interests
The authors declare no competing interests.

## Additional information
**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-26244-1.

**Correspondence** and requests for materials should be addressed to Z.L.

**Reprints and permissions information** is available at www.nature.com/reprints.

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