Original Research Article

Exploring the Impact of Climate Variables on Livestock Anthrax Outbreaks: A Machine Learning Approach

.

ABSTRACT

Anthrax, a globally significant disease, poses substantial threats to both livestock and human populations. Timely identification of anthrax outbreaks is paramount to mitigate its impact on animal health, human health, and public safety. This study aims to construct a predictive model for livestock anthrax disease occurrence. By leveraging the potential of advanced Machine-Learning techniques, we projected the likelihood of anthrax outbreaks across India, through incorporating a diverse set of meteorological, and remote sensing parameters. The ultimate goal is to establish a spatial risk map that can serve as an early warning system, aiding in the anticipation and management of future anthrax outbreaks in India's livestock population. Our analysis revealed elevated risk zones for anthrax outbreaks in the southern and north-eastern regions of India, contrasting with medium to low-risk areas in the central parts. Notably, Enhanced Vegetation Index (EVI), Normalized Difference Vegetation Index (NDVI), rainfall, soil moisture, and wind speed emerged as pivotal variables driving the model's predictive accuracy. Among the employed models, the random forest, adaptive boosting, and classification tree analysis approaches showcased superior performance in livestock anthrax risk assessment. The risk map was generated using significant variables by exploiting best fitted models. These findings hold profound implications for policymakers, guiding the targeted deployment of control strategies against anthrax outbreaks. The dynamic risk maps generated through this study enhance public awareness, equipping decision-makers with vital insights for informed action. By spotlighting risk management endeavours, these maps further enhance governance and risk mitigation efforts.

Keywords:

Anthrax, Livestock, Machine Learning, Meteorological Variables, Remote Sensing Factors, Risk Assessment, Risk Management.

1. INTRODUCTION

Anthrax, a zoonotic disease caused by the spore-forming bacterium *Bacillus anthracis*, has been a longstanding concern due to its impact on livestock, wildlife, and human populations (Turnbull, 1998; Fasanella et al., 2010). Its capacity to persist in the environment for extended periods presents a distinctive challenge to global public health, veterinary medicine, and livestock industries. Despite extensive research, there still exists a complex interaction between ecological conditions, climatic factors, and the epidemiology of anthrax, especially in regions where it is endemic.

Anthrax transmission dynamics involve intricate interactions between the bacterium, hosts, and the environment. Human exposure to anthrax occurs through contact with contaminated animal products, while livestock and wildlife can be affected through various routes such as ingestion, inhalation, and percutaneous exposure (Blackburn et al., 2010). This zoonotic disease has marked its presence across geographies, manifesting a particular burden in regions like Central Asia and West Africa, where livestock management practices often contribute to its spread. Despite its potential as a bioterrorism agent, anthrax remains categorized as a neglected disease by global health authorities, leading to underreporting and insufficient attention to its true impact (FAO, 2006).

Many endemic diseases worldwide, including anthrax, are particularly sensitive to long-term climate shifts. Incorporating ecological data into disease studies not only demonstrates the relationship between disease and the environment but also aids in identifying potential risk factors for disease outbreaks (Myers et al., 2000; Michael et al., 2018). The spread of anthrax across geographical regions is influenced by a combination of diverse climatic and environmental elements (Blackburn et al., 2007; Hugh-Jones and Blackburn, 2009). The causative agent of anthrax B. anthracis, has the capability to endure for extended periods in the environment, potentially lasting for years under congenial ecological conditions (Turnbull, 1998; Michael et al., 2018). There is ongoing debate regarding whether anthrax exclusively occurs during specific periods of the year or not. The presence of anthrax in a specific area can be attributed to the favorable conditions of temperature, precipitation, soil quality, vegetation, as well as the occurrence of drought. Nevertheless, the literature indicates that the factors influencing epidemics differ significantly from one region to another (Turnbull, 1998; Hugh-Jones and Blackburn, 2009). Despite the effectiveness of routine anthrax vaccination and proper outbreak response in controlling the disease, underreporting frequently distorts its true burden and geographical distribution, making the implementation of adequate vaccination campaigns challenging (Turnbull, 1998).

Climate change with its profound influence on ecosystems and the environment is recognized as a key player in shaping the epidemiology of various diseases, including anthrax. Changes in temperature, precipitation patterns, and habitat alterations can impact the occurrence and distribution of anthrax outbreaks. However, the exact relationship between climatic factors, ecological conditions, and the dynamics of anthrax outbreaks remains intricate and region-specific. Despite its historical prominence, the spatial ecology of B. anthracis, the causative agent of anthrax, remains relatively enigmatic. Specific geographical and environmental conditions that facilitate the prolonged survival of anthrax are still not well understood (Smith et al., 2000; Hugh-Jones and De Vos, 2002). The prevailing literature emphasizes B. anthracis ubiquitous presence as a soil-borne bacterium, its resilience bolstered by higher calcium levels, elevated temperatures, increased humidity, a slightly alkaline pH, and higher levels of decomposed organic matter. Additionally, the organism's capability to thrive in harsh environments contributes to the extended persistence of its spores in soil (Smith et al., 2000; Coker, 2002; Assefa et al., 2020). Consequently, certain regions experience more frequent anthrax outbreaks than others. Notably, in anthrax-endemic areas with warm climates, such as Turkey, Ethiopia, and South Africa, significant outbreaks tend to occur during dry and warm periods following heavy precipitation (Carlson et al., 2019). Remarkably, anthrax can also endure in colder regions. The extensive geographical spread of anthrax, along with its potential recurrence after years or even decades, can be attributed to the robust resistance of spores to unfavourable conditions and their capacity for efficient multiplication (Driks, 2009).

In this study, we address the gap in understanding the influence of climatic factors on anthrax outbreaks and geographical distribution in India. By harnessing the power of machine learning, we aim to develop a predictive model that integrates meteorological data,

remote sensing information, livestock population dynamics, and historical anthrax outbreak statistics. The results of this investigation hold promise for informing animal health authorities and policymakers in devising effective control strategies to mitigate the impact of anthrax on livestock and human populations in India.

2. MATERIAL AND METHODS

2.1 Study area

The anthrax is a disease of utmost zoonotic importance and sporadic cases of anthrax continue to be reported from many parts of the world. From India, both sporadic cases and outbreaks are being reported regularly. Most of the anthrax outbreak reports are from southern, central and north-eastern states of India like, Andhra Pradesh, Assam, Chhattisgarh, Jharkhand, Karnataka, Kerala, Madhya Pradesh, Odisha, Puducherry, Tamil Nadu, Telangana, and West Bengal. The initiation of this specific study was prompted by non-availability of data in other states of India and data availability of anthrax outbreak situation in the above-mentioned states during 2000-2022 emphasized to commence this particular study. In an effort to mitigate the likelihood of false positive predictions, we deliberately excluded regions lacking sufficient data points from our study.

2.2 Disease data

In India, anthrax is considered a notifiable disease, necessitating the obligatory reporting of all observed and verified instances of anthrax outbreaks in both animals and humans. The confirmation process relies on clinical indicators and the microscopic analysis of blood smears, and bacterial culture growth analysis. The outbreak and attacks data were retrieved from the respective State Department of Animal Husbandry and Veterinary Services. This information covers the time span from 2000 to 2022 and was supplemented by data from existing literature. The outbreak details were organized in terms of spatial and temporal references, with careful verification of coordinates (X, Y), species, district codes, and occurrence month.

2.3 Livestock data

India possesses a significant livestock population, consisting of a collective count of approximately 535.78 million animals. Within this group, cattle make up 192.49 million, while exotic and crossbred animals contribute 50.42 million. “Of particular note, the nation accommodates 109.85 million buffaloes, 9.06 million pigs, 148.88 million goats, 74.26 million sheep, and 0.44 million Mithun and yaks, (Department of Animal Husbandry and Dairying-DAHD-20th livestock census of India). The livestock population data throughout study region in five major animal species, i.e., cattle, buffalo, sheep, and goats, were collected from the 20th livestock census of India at the village level”.(Muluken et al., 2015; Marlize et al., 2009)

2.4 Risk factors data

2.4.1 Meteorological data:“The Meteorological parameters such as soil moisture (kg/m2), potential evaporation rate (w/m2), specific humidity (kg/kg), rainfall (kg/m2/s), air temperature (k), wind speed (m/s), and surface pressure (pa) were extracted from the Global Land Data Assimilation System (GLDAS-2)” (Rodell et al., 2004). “GLDAS-2 deploys advanced land modelling and data integration methods to capture satellite and ground-based observed data with a spatial resolution of 0.25◦ × 0.25◦ and a temporal resolution retrieved in network common data format (netCDF). This includes metadata as well as data

that have a multidimensional array and data dimensions. The data were restored using the 'ncdf4' package in the R tool".(Muluken et al., 2015; Marlize et al., 2009)

2.4.2 Remote sensing data: Moderate Resolution Imaging Spectroradiometer (MODIS) satellite was employed as a source of remote sensing variables (Justice et al., 2002). Widely used remote sensing parameters like the enhanced vegetation index (EVI, 16-day interval), potential evapotranspiration (PET, 16-day interval, 500 m), land surface temperature (LST, 8-day interval, 1 km), normalized difference vegetation index (NDVI, 16-day interval, 500 m), potential leaf area index (LAI, 16-day interval, 500 m), and were extracted with image products such as MOD16A2, MOD11A2, MOD13A1, and MOD15A2H. These products are available in HDF format at various spatial and temporal resolutions. The R packages "gdalutils" and "modis" were used to extract data in HDF files and refine them into GeoTIFF files. Through the R package "raster" all the variables were organized in raster (grid) type files and each predictor must be a raster layer reflecting a variable of the concern.

2.5 Data pre-processing and feature engineering

Data collection from different sources could be internal and /or external to satisfy the objectives of forewarning requirements, data can be in any format, CSV, XML, JSON, etc. In this process of data and feature engineering, we focus mainly on understanding the given data set and cleaning up the dataset, better understanding of features and their relationships, extracting essential variables, handling missing values and human error, identifying outliers, transforming features if there are outliers, so that either truncate a data above threshold or transform the data using log or any other transformation, scaling the features extracted. This process would be maximizing the insights into a dataset.

2.6 Spatio-temporal Endemicity

The annual occurrence of anthrax attacks was examined to gain insights into the spatial and temporal patterns of disease prevalence. The analysis aimed to identify possible shifts in the distribution of reported disease cases over both geographical and time dimensions. The cumulative instances of anthrax cases reported in India for each year spanning from 2000-2022 were visualized by creating a map that illustrates the incidence rate at the district level.

2.7 Space-Time Cluster Analysis

" SaTScan software version 9.6 was used to develop Poisson-based clustering models based on space-time scan statistics in order to identify the temporal, geographical, and space-time clusters of anthrax in the study area. In case of SaTScan, to detect spatial clusters across a study area, a series of moving windows with varying diameters were used, likewise, temporal clusters are detected and it places ellipses or circles of constantly varying sizes over a three-dimensional study area" (Muluken et al., 2015; Marlize et al., 2009). The circles with observed values that were higher than expected values reported as clusters. For the SaTScan analyses, village wise latitude and longitude coordinates were considered to perform cluster analysis on dataset where each parameter has a disease status (case vs control), as well as spatial and temporal attributes. Using the total number of cases in a given year per epi unit (village), the model was ran on each year's case dataset while accounting for the underlying population of each epi unit. For all SaTScan clusters, the p-value cutoff for statistical significance was established at 0.05.

2.8 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a machine learning algorithm rooted in Fisher's linear discriminant theory, utilized to distinguish between multiple classes. Through discriminant analysis, risk parameters have been thoroughly examined, establishing a linear relationship among them. This correlation forms a robust foundation for accurately understanding the attribute's impact on computation and assessment. SaTScan was employed to detect regions with both significant and non-significant space-time clusters, allowing for the identification of risk occurrences. LDA was then utilized to assess variations in risk factors within these identified regions. The binary response (0 or 1) was assigned based on clustering status, with 1 denoting clustered regions and 0 representing non-clustered ones. The LDA was performed using R, with a statistical significance level set at ($p \leq 0.05$) for the 12 variables under consideration in this study.

2.9 Risk modelling and mapping by accomplishing machine learning

Risk modelling and mapping were conducted using data spanning from 2000 to 2022, aggregated at the grid level. The severity of anthrax was forecasted by creating a risk map for the study area through climate-disease relationship modelling, which predicted the spatial occurrence of anthrax outbreaks. The dataset encompassing risk factors was collected, pre-processed, and annotated with disease conditions, as well as latitude and longitude information. Risk estimate was carried out using machine learning algorithms to identify the most accurate prediction model with improved performance. Disease modelling was executed with nine machine learning approaches, such as random forest (RF), generalized linear models (GLM), generalized additive models (GAM), flexible discriminant analysis (FDA), support vector machine (SVM), multiple adaptive regression splines (MARS), naive Bayes (NB), classification tree analysis (CT), and adaptive boosting (ADA).

2.9.1 Hyper parameterization: The ability of a model to provide accurate outputs for unseen input data, known as generalization, is a key objective in Machine Learning. A well-generalized model strikes a balance between under fitting and over fitting. Training and testing data play pivotal roles in regulating model performance. The training data enables the algorithm to discern patterns, cross-validation ensures accuracy, and the test data assesses predictive capability with new information. Over fitting occurs when a model excessively learns noise in the training data, impairing its performance on new data. Non-parametric and non-linear models, while more flexible, are more susceptible to over fitting. Conversely, an under fit model cannot effectively model the training data or generalize to new data. Striking the right balance between memorization and generalization is a common challenge in machine learning algorithms. Regularization techniques are employed to mitigate over fitting. In this study, all models were assessed for over fitting or under fitting, and to optimize coefficient estimation, p-values, and R-Square values, the data was randomly split into a 70% training set and a 30% testing set. This approach ensures a robust evaluation of model performance in the present study.

2.9.2 Model evaluation and Ensemble techniques: In this study, predictions based on various combinations of risk factors were generated using different model artefacts. Response graphs were developed to facilitate the interpretation and evaluation of these predicted results. A comprehensive set of evaluation metrics including the Receiving Operating Characteristic (ROC) curve, True Skill Statistics (TSS), Cohen's Kappa (Heidke Skill Score), Area Under the ROC Curve (AUC), F1 score, error rate, accuracy, and logistic loss (LOGLOSS) were employed to assess the discriminative capacity of the fitted models. These metrics were utilized to evaluate the accuracy of prediction models based on the presence (1) or absence (0) of data. In this study, the outcomes of separate forecasts from multiple model methods were aggregated using a Raster Stack approach (Liu et al., 2009). Rather than relying on a single best model, it is recommended to combine predictions from

different models, which provide scores ranging from 0 to 1. Averaging these scores yielded the most accurate prediction (Huppert and Katriel, 2013; Omri Allouche, 2006). The average model score was derived by considering models that met the criteria of kappa > 0.60, ROC > 0.90, and TSS > 0.80 for further assessment of disease risk (Abdrakhmanov et al., 2017). This approach ensures a robust evaluation and aggregation of predictions for a more accurate risk assessment.

2.10 Basic reproduction number (R_0)

The estimated number of additional infectious disease cases that resulted from the initial incident in a community that is susceptible is known as the basic reproduction number (R_0). If $R_0 > 1$, the number of individuals affected will increase, additionally, if $R_0 < 1$, the number will decline. R_0 represents the disease transmission rate. In the current study, the R0 was estimated using a maximum likelihood estimation (ML) method (Mahmud and Patwari, 2020). A visible and comprehensive view of the possibility and impact of a disease in a certain location was obtained by superimposing the R_0 on the risk map projected using the density of livestock, meteorological, and remote sensing parameters.

2.11 Statistical software

The statistical analyses, risk maps, and disease forecasts were carried out using R statistical software version 3.1.3 (version 3.4.3, Vienna, Austria: R Foundation for Statistical Computing). Using R as a comprehensive suite, data mining, computing, and graphical display were accomplished. With the assistance of R packages such as plyr, dplyr, rgdal, raster, data.table, openxlsx, tmap, sp, spdep, sf, BAMM tools, foreign, geosphere, MASS, biomod2, dsimo, mgcv, randomforest, mda, gbm, earth data extraction data alignment, annotation, analysis, model fitting, and validation were achieved. Risk mapping and hotspot analysis were performed using Getis ord's Index and to acquire the geographical and temporal clusters in the relevant study area, SaTScan v9.6 was used.

3. RESULTS AND DISCUSSION

3.1 Spatial Endemicity of Anthrax

The dataset encompassing anthrax incidence within the study period (2000-2022) underwent comprehensive analysis to ascertain the endemicity status of the disease across the study regions. Fig. 1 provides a visual representation of the geographical distribution and endemicity levels of anthrax at the district level. Notably, Koraput district in Odisha and Murshidabad district in West Bengal reported a notably high incidence of anthrax, surpassing 401 cases. In contrast, Chamarajanagar and Kolar districts in Karnataka, Chittoor and Visakhapatnam districts in Andhra Pradesh, Bankura and Barddhaman districts in West Bengal, Simdega district in Jharkhand, and Sundargarh district in Odisha experienced a medium risk of anthrax, with reported cases ranging from 51 to 400. Similarly, numerous districts in various states such as Karnataka (including Bangalore Rural, Bellary, Chikkaballapura, Davanagere, Hasan, Kodagu, and Mysore), Andhra Pradesh (Anantapur and Vizianagaram), Odisha (Debagarh and Rayagada), Chhattisgarh (Durg), Tamil Nadu (Erode, Ramanathapuram, Tirunelveli, Vellore, and Viluppuram), Assam (Golaghat), Jharkhand (Gumla), West Bengal (Hugli, Jalpaiguri, Koch Bihar, and Nadia), Telangana (Nalgonda), Puducherry (Puducherry), and Kerala (Thiruvananthapuram) reported a low incidence of anthrax, with cases ranging from 1 to 50.

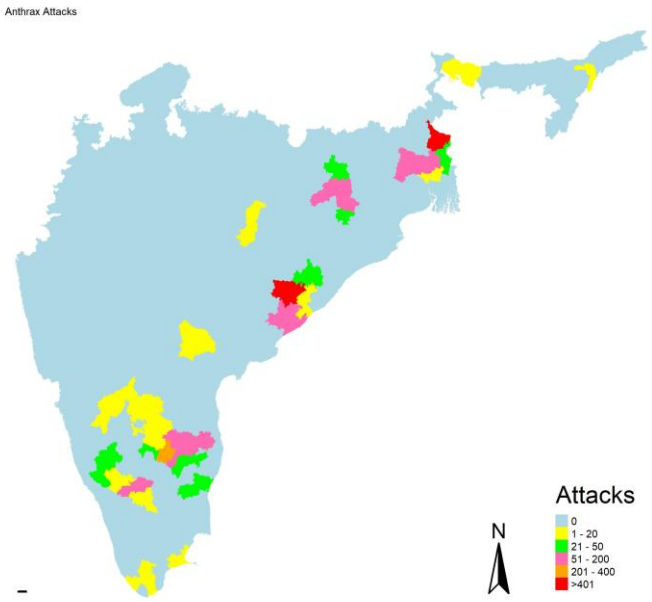

Fig. 1. A map of India delineates the study counties, highlighting the district-wise cumulative occurrences of anthrax spanning from 2000 to 2022.

3.2 Temporal distribution of anthrax

The highest number of anthrax outbreaks was observed between 2012 and 2016. In the remaining years under study, the incidence of anthrax outbreaks remained relatively stable. However, a notable decrease in anthrax outbreaks was observed from 2020 onwards (Fig. 2). When examining the monthly distribution of anthrax outbreaks (Fig. 3), it becomes evident that the pre-monsoon and monsoon period, spanning from May to August, recorded the highest occurrences, with the peak observed in the month of May.

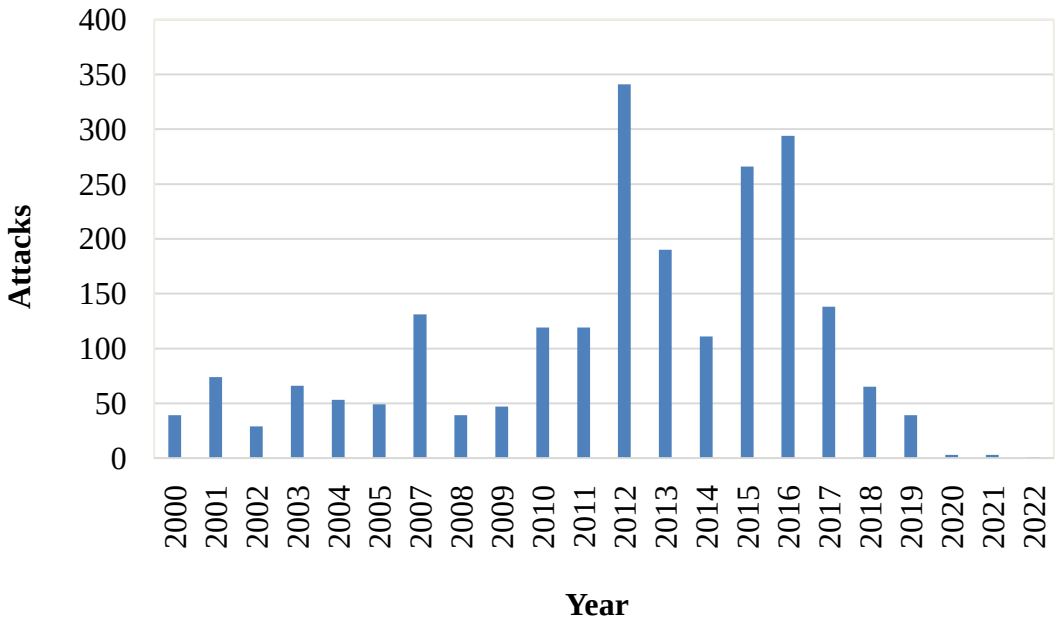

Fig. 2. Yearly incidence of livestock anthrax outbreaks within the designated study regions of India

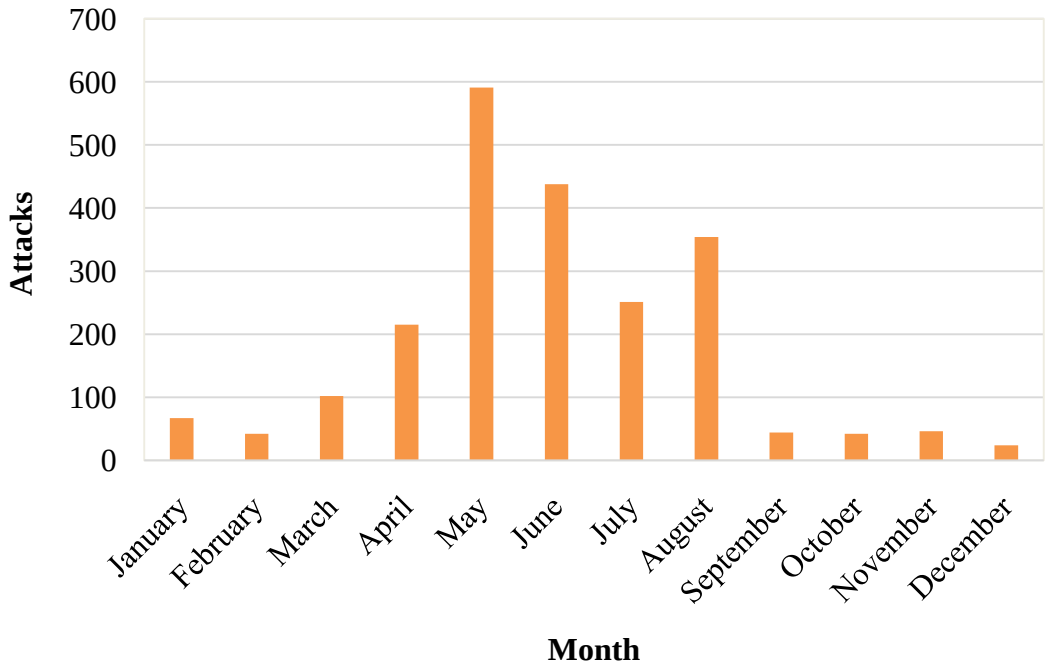

Fig. 3. Monthly frequency of livestock anthrax attacks across the selected study areas in India.

3.3 Space-Time cluster analysis

Utilizing space-time cluster analysis, disease clusters were identified in both the southern and north-eastern regions of India. In terms of spatial variation, six noteworthy clusters exhibiting high risk and one extensive cluster indicating low risk were discerned (Fig. 4). The village-level disease clustering was established over the period from 2000 to 2022. Disease incidence is represented by red dots within significant red circles, denoting villages with a high risk of disease incidence. Conversely, blue dots within significant blue circles form clusters indicating villages with a low risk of disease incidence. These findings align with the spatial endemicity observations. Locations that experienced a higher frequency of outbreaks between 2000 and 2022 were situated within high-risk clusters in southern India, while areas with fewer outbreaks formed a single cluster in the north-eastern region.

Fig. 4. Space-Time cluster analysis of Anthrax for area under study on a map of India. Red spots represent high risk of disease incidence and blue spots represent incidence with negligible risk.

3.4 Linear discriminant analysis

Linear discriminant analysis (LDA) was employed to identify the key risk factors, including meteorological and remote sensing variables, contributing to the occurrence of anthrax. The results of the linear discriminant analysis are presented in Table 1. The findings indicate that EVI (0.016), LST (0.0004), NDVI (0.009), rainfall precipitation rate (0.002), soil moisture (0.001), and wind speed (0.00006) (p value < 0.05) emerged as significant risk factors associated with anthrax outbreaks. Furthermore, this study revealed that regions characterized by specific environmental conditions, such as average EVI (0.269), LST (32.963), NDVI (0.416), rainfall precipitation rate (5.38 mm), soil moisture (24.531 kg/m^2), and wind speed (2.224 m/s), were conducive for disease outbreaks. These primary significant risk metrics, positively influencing disease incidence, were further integrated into

disease modelling and risk mapping efforts. These results are consistent with various literature reports, which highlight a higher incidence of anthrax during dry and warm periods following intensive precipitation (Carlson et al., 2019). Moreover, our findings align with previous studies that have identified LST, NDVI, and rainfall as significant factors associated with anthrax, with higher outbreaks occurring during the monsoon months of August, September, and October (Suma et al., 2017). In another study, air temperature, wind speed, and potential evaporation rate were identified as potential risk indicators during El Nino years, whereas during La Nina years, air temperature, EVI, NDVI, specific humidity, and wind speed were found to be significant contributors to anthrax in the Karnataka region (Suresh et al., 2022). These observations further underscore the complexity of the interplay between environmental variables and anthrax dynamics. B. anthracis, an extracellular pathogen, exhibits rapid replication within the bloodstream, leading to the onset of disease. The survival of its spores is believed to be influenced by soil pH, organic calcium, potassium, and zinc concentrations (Jayachandran, 2002). Additionally, the dissemination of spores is facilitated by precipitation and wind speed (Jocelyn et al., 2011). Animals typically come into contact with these spores through behaviours such as grazing on low or scarce grass close to the surface, or by being herded into restricted areas during periods of water scarcity (Turnbull, 1998). These interactions contribute significantly to the transmission dynamics of anthrax.

Table 1: Results of Linear Discriminant Analysis (LDA)

Parameter	Mean (Presence)	SD	F-Value	p-Value	95 % CI
Air Temperature (k)	24.496	3.305	0.494	0.483	24.06 to 24.94
Enhanced Vegetation Index (EVI)	0.269	0.130	5.839	**0.016***	0.25 to 0.29
Leaf area index (LAI)	0.163	0.469	0.099	0.754	0.10 to 0.23
Land Surface Temperature (LST)	32.963	7.679	12.919	**0.0004***	31.94 to 33.99
Normalized Difference Vegetative Index (NDVI)	0.416	0.178	6.925	**0.009***	0.39 to 0.44
Potential Evapotranspiration (PET)	1351.214	1289.485	2.144	0.144	1179.25 to 1523.18
Potential evaporation rate (w/m2)	218.266	76.211	0.115	0.735	208.10 to 228.43
Rainfall Precipitation rate (mm)	5.38	3.90	10.093	**0.002***	4.86 to 5.90
Soil moisture (kg/m^{-2})	24.531	6.204	12.206	**0.001***	23.70 to 25.36
Specific Humidity (kg/kg)	0.015	0.002	2.829	0.094	0.014 to 0.015
Surface Pressure (pa)	93744.991	2676.907	0.033	0.856	93388.00 to 94101.99
Wind speed (m/s)	2.224	1.263	16.696	**0.00006**	2.06 to 2.39

Where, SD= Standard deviation, CI= Confidence interval, and * 5 percent level of significance

3.5 Anthrax risk assessment and estimation

The significant ecological and environmental risk factors identified through LDA underwent climate-disease modelling. Maps were generated based on areas affected (cases) and

unaffected (controls) by anthrax (Fig. 5). In the map (Fig. 5A-C), case data is denoted by red circles, signifying locations with reported disease incidences at various thresholds, while control data is represented by blue dots, indicating places without anthrax incidence.

The random forest machine learning model demonstrated superior performance over other models in predicting anthrax outbreaks. It exhibited satisfactory evaluation metrics, including Kappa (0.751), ROC (0.995), TSS (0.936), AUC (0.995), accuracy (0.973), F1 score (0.949), and lower error rate (0.027), as well as log loss (0.242). This was followed by adaptive boosting with Kappa (0.512), ROC (0.735), TSS (0.469), AUC (0.735), accuracy (0.795), F1 score (0.646), lower error rate (0.205), and log loss (7.093), and classification tree analysis with diminishing evaluation metrics: Kappa (0.476), ROC (0.820), TSS (0.467), AUC (0.820), accuracy (0.714), F1 score (0.609), lower error rate (0.286), and log loss (0.465) (Table 2).

The most reliable predictions were achieved through the ensemble method, which involved averaging the scores from the random forest, adaptive boosting, and classification tree analysis models. These three models demonstrated superior performance compared to the others employed. Based on this ensemble approach, risk predictions for the study area were estimated at the district level. Previous disease prediction research predominantly relied on traditional statistical models, which exhibited varying degrees of prediction accuracy (Suma et al., 2017; Sinkie et al., 2016; Osman et al., 2018). By incorporating Geographical Information Systems (GIS), epidemiologists are deploying machine learning techniques to examine drivers of animal and zoonotic diseases. In the present study ensembling of random forest, adaptive boosting, and classification tree analysis models was found to be highly efficient in prediction of anthrax with high accuracy. B. anthracis considered to have a high dispersing capacity. High dispersion capacity enhances the likelihood of the species being reported in a location where adequate conditions do not exist (Pearson, 2007). This provides a difficulty to modelling because the models require presence records to locate areas with adequate conditions, which may lead to model mistakes. Furthermore, different genotypes of B. anthracis have been demonstrated to have varied requirements for soil factors such as pH and calcium content and environmental factors (Smith et al., 2000; Hugh-Jones and De Vos, 2002) therefore modelling based on the specific genotype(s) may improve model performance. In future there is need to focus these limiting factors for the development of effective predictive modelling for anthrax. Nevertheless, the use of machine learning algorithms in modelling in current study has resulted in delineating more detailed *B. anthracis* ecological niche and high-risk regions in the study sites of India.

Table 2: Machine learning model's evaluation metrices

Models	KAPPA	ROC	TSS	AUC	Accuracy	ERROR RATE	F1 SCORE	LOGLOSS
GLM	0.222	0.656	0.267	0.656	0.607	0.393	0.192	0.615
GAM	0.222	0.656	0.267	0.656	0.607	0.393	0.192	0.615
RF	0.751	0.995	0.936	0.995	0.973	0.027	0.949	0.242
MARS	0.347	0.746	0.389	0.746	0.679	0.321	0.523	0.558
FDA	0.063	0.525	0.05	0.525	0.661	0.339	0.095	11.719
CT	0.476	0.82	0.467	0.820	0.714	0.286	0.609	0.465
SVM	0.412	0.785	0.458	0.785	0.723	0.277	0.182	0.846
NB	-0.035	0.48	0.064	0.480	0.527	0.473	0.496	0.922
ADA	0.512	0.735	0.469	0.735	0.795	0.205	0.646	7.093

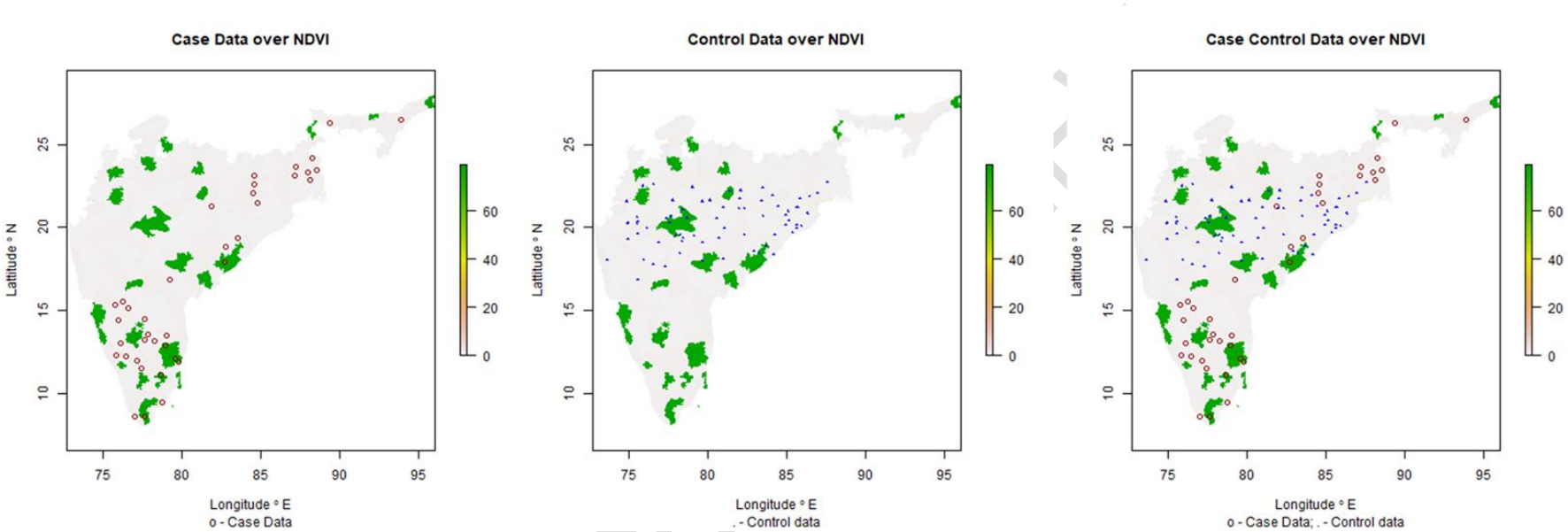

Fig. 5. Anthrax outbreaks case-control data are depicted on a map of area under study. (A) Case data: red-coloured circles denote locations where anthrax has been reported, (B) control data: blue-coloured dots denote locations where anthrax has not been reported, and (C) case-control data: displays both the existence and absence of anthrax incidence

3.6 Anthrax risk prediction and mapping

Risk maps provide an advanced digital platform for a comprehensive assessment of the likelihood and potential impact of diseases, enabling the development of synergistic strategies within a specific study area. In this study, an ensemble model consisting of the Random Forest and Classification Tree models, along with significant environmental risk factors identified through LDA analysis, was utilized for the generation of a risk map. The spatial prediction of the anthrax ecological niche in the study sites across India is depicted in Fig. 6. The predicted high-risk areas were predominantly concentrated in the southern regions of India, specifically in districts of Karnataka, Andhra Pradesh, Kerala, Tamil Nadu, and Maharashtra, exhibiting the most favorable conditions for the persistence of B. anthracis. Additionally, some areas with conducive environmental conditions for anthrax outbreaks were also identified in the northeastern regions of India. Conversely, in the central and northern parts of India, the conditions conducive for anthrax establishment and the risk of anthrax outbreaks were predicted to be very low or negligible.

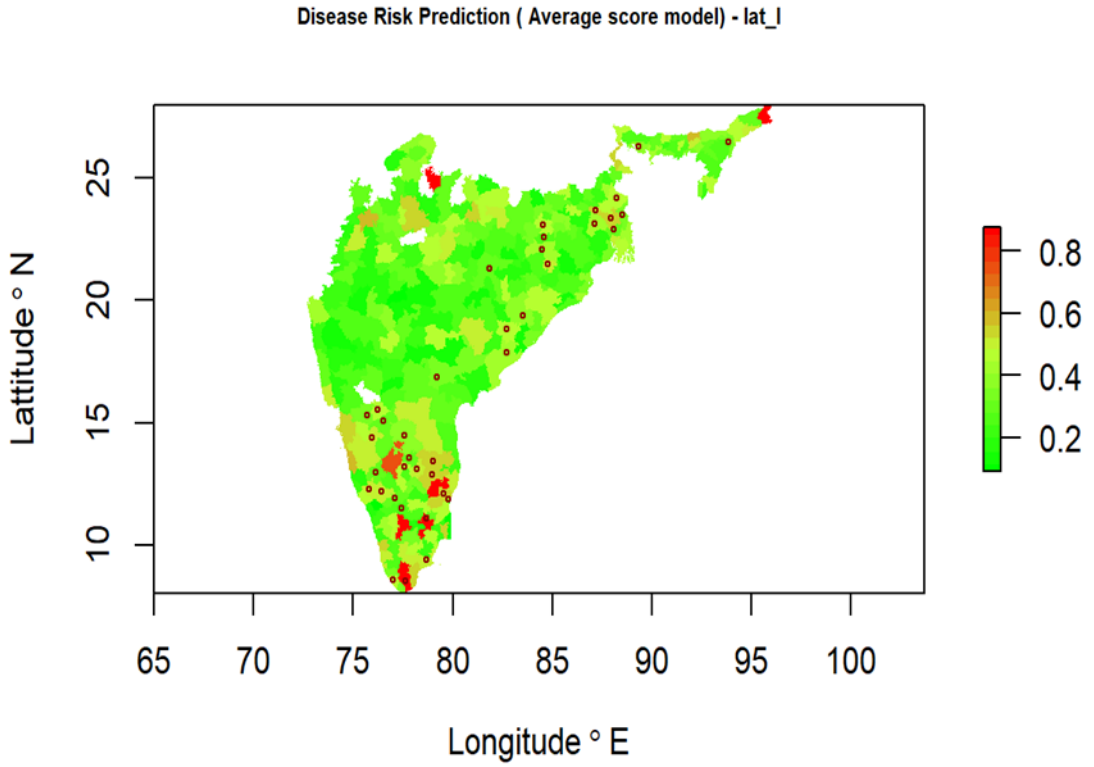

Fig. 6. Anthrax risk prediction map generated for area under study. Red indicates areas with high risk of anthrax, while yellow and green indicate areas of medium and low suitability respectively.

3.7 Estimation of basic reproduction number (R_0) of Anthrax

In this current study, the basic reproduction number (R_0) was computed for the districts falling within the significantly clustered zones identified by the SaTScan analysis. The R_0 values were calculated at the conclusion of the risk assessment and subsequently overlaid onto the risk map (Fig. 7). R_0 values exceeding 1.00 indicate areas or districts where the prevalence of the disease is increasing, while R_0 values below 1.00 indicate areas or

districts where the disease prevalence is decreasing. The R_0 values, as depicted in Fig. 7, ranged from 0.75 (Thiruvananthapuram district of Kerala) to 5.37 (Koraput district of Odisha). This indicates that districts in the southern regions and north-eastern states are more likely to experience anthrax outbreaks, rendering them more vulnerable to anthrax. Moreover, regions with initially low R_0 values may potentially transition to higher R_0 values in the near future, potentially due to the migration of infected animals from one location to another.

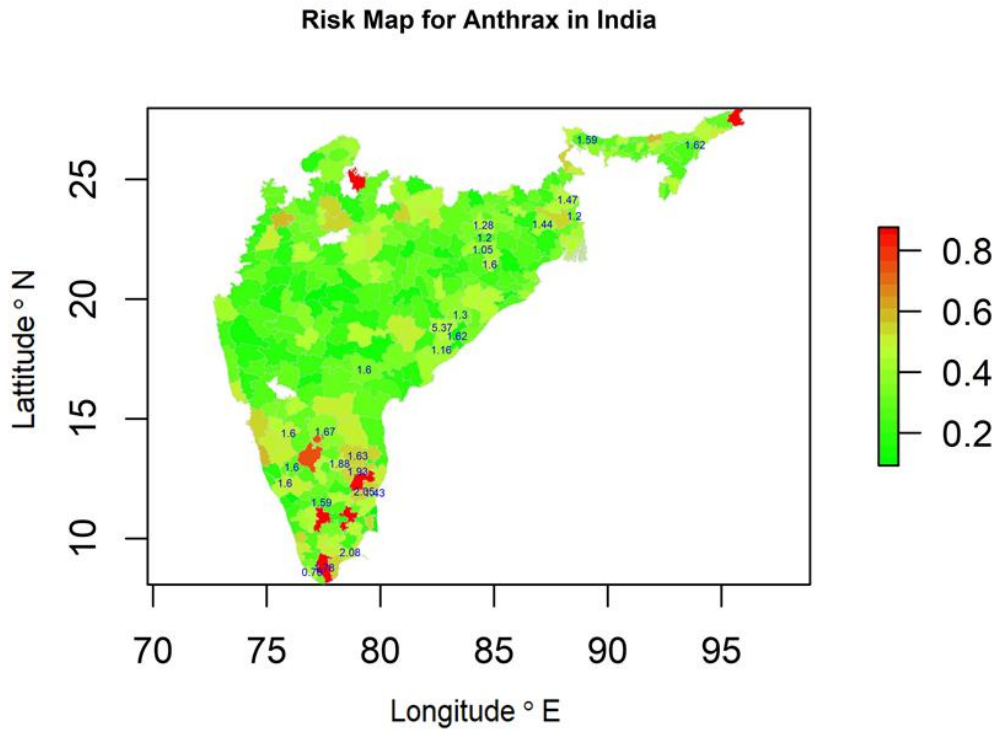

Fig. 7. Anthrax R0 values on risk prediction map

4 Conclusion

In conclusion, our study advances the understanding of anthrax epidemiology in India by demonstrating the utility of environmental factors in assessing anthrax risk. By integrating advanced analytical techniques and environmental data, we provide valuable insights that can inform targeted control measures and enhance preparedness in regions vulnerable to anthrax outbreaks. These findings have practical implications for public health authorities and policymakers in the formulation of effective strategies for anthrax prevention and control. Increased surveillance measures are also recommended in identified probable regions with acceptable conditions where outbreaks have not been observed. These efforts, when combined with technological advancements and continued research, have the potential to significantly contribute to anthrax management and prevention strategies in India.

REFERENCES

Abdrakhmanov, S. K. Mukhanbetkaliyev, Y. Y. Korennoy, F. I. Sultanov, A. A. Kadyrov, A. S. Kushubaev, D. B. Bakishev, T. G. 2017. Maximum entropy modelling risk of anthrax in the Republic of Kazakhstan. Prev. Vet. Med., 144:149-157.

Assefa, A. Bihon, A. Tibebu, A. 2020. Anthrax in the Amhara regional state of Ethiopia; spatiotemporal analysis and environmental suitability modelling with an ensemble approach. Prev. Vet. Med., 184: 105155.

Blackburn, J. McNyset, K. Curtis, A. Hugh-Jones, M. 2007. Modelling the geographic distribution of Bacillus anthracis, the causative agent of anthrax disease, for the contiguous United States using predictive ecologic niche modelling. Am, J. Trop. Med., Hyg 77: 1103.

Blackburn, J. K. Curtis, A. Hadfield, T. L. O'Shea, B. Mitchell, M. A. Hugh-Jones, M. E. 2010. Confirmation of Bacillus anthracis from flesh-eating flies collected during a West Texas anthrax season. J. Wildl. Dis., 46(3): 918e922.

Carlson, C. J. Kracalik, I.T. Ross, N. Alexander, K.A. Hugh-Jones, M. E. Fegan, M. et al. 2019. The global distribution of Bacillus anthracis and associated anthrax risk to humans, livestock and wildlife. Nat. Microbiol., 4: 1337-1343.

Coker, P. R. 2002. Bacillus anthracis Spore Concentrations at Various Carcass Sites. Doctoral dissertation, Louisiana State University. 77pp.

Driks, A. 2009. The Bacillus anthracis spore. Mol Aspects Med., 30: 368-373.

Fasanella, A. Galante, D. Garofolo, G. Jones, M. H. 2010. Anthrax undervalued zoonosis. Vet. Microbiol., 140: 318-331.

Food and Agriculture Organization of the United Nations, Great Britain, Department for International Development, Animal Health Programme, International Office of Epizootics, World Health Organization. The Control of neglected zoonotic diseases: a route to poverty alleviation: report of a joint WHO/ DFID-AHP meeting, 20 and 21 September 2005, WHO Headquarters, Geneva, with the participation of FAO and OIE. Geneva, Switzerland: World Health Organization. 2006.

Hugh-Jones, M. Blackburn, J. 2009. The ecology of Bacillus anthracis. Mol. Aspects Med., 30: 356-367.

Hugh-Jones, M. E. De Vos, V. 2002. Anthrax and wildlife. OIE scientific and technical review, International Office of Epizootics. 21(2): 359-383.

Huppert A, Katriel, G. (2013) Mathematical modelling and prediction in infectious disease epidemiology. Clin. Microbiol. Infect., 19: 999-1005.

Jayachandran, R. 2002. Anthrax: Biology of Bacillus anthracis. Curr. Sci., 82: 1220-1226.

Jocelyn, M. Larissa, L. Alim, A. Yerlan, P. Mathew, V. E. Jason, K. B. 2011. Ecological Niche Modelling of the Bacillus anthracis A1. a sub-lineage in Kazakhstan. BMC Ecol., 11: 32.

Justice, C. O. Townshend, J. R. G. Vermote, E. F. Masuoka, E. Wolf, R. E. Saleous, N. et al. 2002. An overview of MODIS land data processing and product status. Remote Sens. Environ., 83: 3-15.

Liu, C. White, M. Newell, G. 2009. Measuring the accuracy of species distribution models: a review. 18 World IMACS, MODSIM Congress, Caims, Australia, 13-17 July.

Marlize, C. Michael, C. Aaron, M. M. Gerdalize, K. Maureen, C. David, N. D. 2009. Using the SaTScan method to detect local malaria clusters for guiding malaria control programmes. Malar. J., 8: 68.

Michael, G. W. de Smalen, A. W. Siobhan, M. M. 2018. Climatic influence on anthrax suitability in warming northern latitudes. Sci. Rep., 8: 9269.

Muluken, A. Abera, K. Alemayehu, W. Bagtzoglou, A. C. 2015. Childhood diarrhea exhibits spatiotemporal variation in Northwest Ethiopia: A SaTScan spatial statistical analysis. PLoS ONE, 10: e0144690.

Myers, M. F. Rogers, D. J. Cox, J. Flahault, A. Hay, S. I. 2000. Forecasting disease risk for increased epidemic preparedness in public health. Adv. Parasitol., 47: 309-330.

Omri, A. Asaf, T. and Ronen, K. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and true skill statistic (TSS). J. Appl. Ecol., 43: 1223-1232.

Osman, Shaibu, Oluwole Daniel Makinde, David Mwangi Theuri. 2018. Mathematical modelling of transmission dynamics of anthrax in human and animal population. Math. Theory Model, 8(6): 47-67.

Pearson, R. G. 2007. Species distribution modelling for conservation educators and practitioners, synthesis. Bull. Am. Mus. Nat. Hist., 3: 54-89.

Rodell, M. Houser, P. R. Jambor, U. E. Gottschalck, J. Mitchell, K. Meng, C. J. Arsenault, K. Cosgrove, B. Radakovich, J. Bosilovich, M. Entin, J. K. 2004. The global land data assimilation system. Bull. Am. Meteorol. Soc., 85(3): 381-394.

Sinkie, Zerihun Mamo, Narasimha Murthy. 2016. Modelling and simulation study of anthrax attack on environment. Differ. Equ., 1: 2.

Smith, K. L. De Vos, V. Price, L. B. Hugh-Jones, M. E. Keim, P. 2000. Bacillus anthracis diversity in Kruger national park. J. Clin. Microbiol., 38(10): 3780-3784.

Suma, A. P. Suresh, K. P. Gajendragad, M. R. Kavya, B. A. 2017. Outbreak prediction of anthrax in Karnataka using Poisson, negative-binomial and zero truncated models. Inter. J. Sci. Res., 6(3): 32-36.

Suresh, K. P. Bylaiah, S. Patil, S. Kumar, M. Indrabalan, U. B. Panduranga, B. A. et al. 2022. A new methodology to comprehend the effect of El Nino and La Nina oscillation in early warning of anthrax epidemic among livestock. Zoonotic Dis., 2(4): 267-290.

Turnbull, P. C. B. 1998. Guidelines for the surveillance and control of anthrax in humans and animals; WHO/EMC/ZDI/98.6, Wiltshire SP40JG; World Health Organization: Geneva, Switzerland.