

Disease Prediction Model to Assess the Impact of Changes in Precipitation Level on the Risk of Anthrax Infectiousness among the Livestock Hosts in Karnataka, India

**Sushma Bylaiah^{1*}, Seema Shedole¹, Kuralayanapalya Puttahonnappa Suresh²,
Leena Gowda³, Bindya Shivananda⁴, Chandan Shivamallu⁵, Sharanagouda S Patil⁶**

1. Department of Computer Science & Engineering, M S Ramaiah Institute of Technology, Matthikere, Bengaluru, Karnataka, INDIA.
2. Spatial Epidemiology Laboratory, Indian Council of Agricultural Research (ICAR) - National Institute of Veterinary Epidemiology and Disease Informatics (NIVEDI), Yelahanka, Bengaluru, Karnataka, INDIA.
3. Department of Veterinary Public Health and Epidemiology, Veterinary College, Hebbal, Bengaluru, Karnataka, INDIA.
4. Department of Chemistry, Sri Jayachamarajendra College of Engineering, JSS Science and Technology University, Mysuru, Karnataka, INDIA.
5. Department of Biotechnology and Bioinformatics, Faculty of Life Sciences, JSS Academy of Higher Education & Research, Mysuru- 570 015, Karnataka, INDIA.
6. Virology Laboratory, Indian Council of Agricultural Research (ICAR) - National Institute of Veterinary Epidemiology and Disease Informatics (NIVEDI), Yelahanka, Bengaluru, Karnataka, INDIA.

*Corresponding author: Sushma Bylaiah, email: sushma.b3@gmail.com

Abstract

Anthrax is a one of the zoonotic diseases existing in India. Early detection of anthrax outbreaks is crucial for minimizing anthrax morbidity and death, as well as the risk of anthrax transmission in the population. Objective of the present research is to develop a disease prediction model by employing Machine-Learning techniques to assess the risk of anthrax analogous to the impact of changes in precipitation level that can benefit as an early warning system for detecting future anthrax outbreaks among livestock across Karnataka. By considering the disease incidence data during 2000 to 2019, livestock population data and the ecological parameters, the machine learning model was successful in identifying the next outbreak susceptible areas and the parameters that contribute significantly to the disease outbreak. Machine learning model was developed by R statistical software version 3.1.3 using different data mining regression and classification models viz., GLM, GAM, MARS, FDA, CT, SVM, NB, ADA, RF, GBM and ANN. Disease incidence data was collected from Department of animal husbandry, Bengaluru, Karnataka. Disease incidence data was divided in two groups based on average annual precipitation above and below normal (1151mm) for the risk assessment and study the impact of changes in precipitation level. Data with average annual-precipitation above normal was predicted with high risk in the north, northern east and the state's southern region. Whereas data with average annual-precipitation below normal was predicted with high risk in south, northern east and the state's central region. Cohen's Kappa, ROC curve, True Skill Statistics (TSS), and ACCURACY was used to assess the models'

performance. Further, this model can be intensified and validated using the anthrax outbreak data available at national level which will be useful for policymakers to formulate control strategies.

Keywords: -Anthrax; Disease prediction; Karnataka; Livestock; Machine Learning; Outbreak.

1. Introduction

Machine learning is increasingly becoming more predominant in the area of healthcare. Machine learning is a subdivision of artificial intelligence that employs algorithms, such as statistical models, to complete specific task without the use of explicit instructions, relying instead on patterns and inference. Novel applications for machine learning approaches are also being developed, such as the organization of disease case details into epidemiological info that can be used in outbreak investigations and risk modelling [1]. Epidemiologists are adopting machine learning techniques to study about various animal and zoonotic diseases by including Geographical Information Systems (GIS). The ability to predict outbreaks early is critical in order to reduce morbidity and mortality by implementing disease prevention as well as control measures on time [2]. Anthrax is one among the infectious diseases that mainly affects grazing animals, but it can also infect humans when contact with sick animals or animal products, such as skin, meat and fur may cause serious illness. Inhalational, cutaneous, gastrointestinal, and intravenous/injectional infections are the four main ways in which humans can become infected. The disease's cutaneous form is caused by spores entering into the skin through a scrape or cut. The gastrointestinal form is commonly caused by eating spore-infested meat that hasn't been properly cooked. Inhaling aerosol particles with diameters ranging from 1 to 5 micro meters causes inhalational anthrax. Intravenous drug-related anthrax with disproportionate tissue edema causes atypical, serious localized soft tissue infections [3]. It is one of humanity's oldest threats, and it's endemic in animals all over the world. Anthrax has become less common in developed countries, but it still poses a serious health threat in developing nations. Anthrax is caused by *B. anthracis*, a Gram-positive, endospore-forming, and rod-shaped bacterium[4]. The bacilli are huge, non-motile, and can be found in single or in short chains. *B. anthracis* can be found in the environment as dormant spores that can live for decades in dry soil and are immune to temperature and disinfectants[5]. *B. anthracis* has two large plasmids (pxo1 and pxo2) crucial for maximum pathogenicity, the absence of either plasmid significantly reduces the pathogenicity of *B. anthracis*. The pathogenic activity of *B. anthracis* is primarily expressed via the capsule and the development of a toxic compound involving three proteins known as edema factor (EF), lethal factor (LF), and protective antigen (PA) [6]. In India, livestock disease outbreaks have a negative impact on the economy of animal husbandry farmers [7]. Predicting the incidence of anthrax disease outbreaks can be extremely beneficial to India's long-term growth. Previous disease prediction study has largely relied on conventional statistical models, which have varying degrees of prediction accuracy [8]. Climate and environmental factors appear to be limiting the disease's geographic spread. Though it is found all over India, its higher endemicity in south India is because of the warm humid climate and alkaline soil that favors survival and germination of spores. Many areas of Eastern Europe, Southern Europe, Africa,

South America, Central Asia, and South East Asia, including India, are prevalent to anthrax. According to reports it is one among the top ten livestock diseases in India [9] [10].

Precipitation is a highly dynamic phenomena that changes in intensity, occurrence, and interval as a result of geographical characteristics such as altitude, flora-fauna, land-use cover, and other variables. Karnataka's coastal region is one of the major foci of all pervasive precipitation action on the west shoreline. The coastal Karnataka region, which is located on the upwind side of the Western Ghats mountain range and dense forest, has a significant longitudinal and time-based erraticism of precipitation. Severe water shortages and flooding are predicted as a result of changes in precipitation patterns [11]. The flooding probably brings anthrax spores to the surface. The dry weather that follows will have caused evaporation of water and resulting in spore concentration [12]. The interrelation between precipitation and temperature is that precipitation affects soil moisture, which intern affects surface temperature [13]. Machine learning (ML) has had a major effect on infectious disease prediction in the biomedical field. Machine learning promises to improve disease detection and prediction in the biomedical field, as well as the impartiality of the decision-making [14]. Thus, we used different regression and classification ML techniques for efficient prediction of anthrax outbreak among livestock by using large dataset of Karnataka state. The performances of all these techniques were estimated and compared on various perspectives.

This study was led to assess the risk of anthrax in Karnataka, India analogous to the changes in precipitation level by employing machine learning models and examine the potential for establishing climate based early warning and surveillance system which could be an input to the veterinarians to take necessary steps to control the infections and reduce the cost of diagnosis and can also help policymakers and other public health organizations to better target resources to areas of greatest need.

2. Materials and methods

Hot and dry environments play a vital role in anthrax outbreaks [12]. Long-term pathogen survival in the environment occurs in soil. Precipitation events influence soil conditions and fluctuation, insufficient precipitation might result in dry soil [15]. Graze becomes scarce during the dry season, having poor nutritional value and being increasingly abrasive. Livestock are forced to consume heavily used short grass or plants, exposing them to anthrax by ingestion or inhalation of spore-infested soil [16]. Changes in precipitation will be one of the greatest critical factors defining the overall impact of climate variation and disease risk pattern [17]. Monitoring precipitation patterns and changes in precipitation amounts is an important aspect of determining disease risk. Above all, a thorough grasp of the precipitation pattern in a changing environment will aid in better decision-making [18]. Andhence motivation to study the impact of changes in precipitation level on the risk of anthrax among livestock in Karnataka.

2.1 Description of the study area

This disease modelling is developed for Karnataka. Karnataka is a state in India with a total area of 191,791 km², 181 blocks distributed across 30 districts, and 29340 villages with a total of 27.7 million livestock animals.

2.2 Livestock and anthrax incidence data

The data on livestock population in five major animal species, namely cattle, buffalo, sheep, goat, and pigs, was collected at the district level from India's 20th livestock census. The current study includes the incidence data of anthrax among livestock throughout Karnataka. The village level data on anthrax outbreaks in livestock over a 20-year period (2000–2019) was collected from the Dept. of Animal Husbandry in Bengaluru, Karnataka, India. Disease incidence data was divided into two different sets based on average annual precipitation above and below normal (1151 mm) between 2000-2019 and subjected to climate-disease modelling to understand the disease pattern over the years [19].

2.3 Risk factors data

Meteorological variables such as air temperature (k), potential evaporation rate (w/m^2), precipitation ($kg/m^2/s$), soil moisture (kg/m^2), specific humidity (kg/kg), surface pressure (pa), and wind speed (m/s) were extracted into a network common data format (netCDF) from the Global Land Data Assimilation System (New and reprocessed GLDAS version 2: <https://ldas.gsfc.nasa.gov/gldas>), to read and extract data, the R package `ncdf4` was used [20]. The remote sensing variables like Normalized Difference vegetative index (NDVI, 16 day interval, 500 meter), Enhanced Vegetation Index (EVI, 16 day interval), Leaf Area Index (LAI in m^2/m^2 , 16 day interval, 500 meter), Land Surface temperature (LST in degree Celsius, 8 day interval, 1KM) and Potential Evapotranspiration (PET in mm, 16 day interval, 500 meter) were extracted with atmospheric corrections using Moderate Resolution Imaging Spectro-radiometer (MODIS) satellite image products such as MOD13A1, MOD11A2, MOD16A2 and MOD15A2H [21] [22]. The R package 'modis' and 'gdalutils' were used to excerpt the data from HDF files by translating into GeoTIFF files. Using R package 'raster,' all of the variables are grouped into raster (grid) files. The soil parameters Ec (Electrical conductivity), pH, Zinc, Potassium, Sulphur, Boron, Carbon, and phosphorus were retrieved from the Karnataka soil health data database, ICRISAT Development center, Government of Karnataka, 2016 [15].

2.4 Spatial autocorrelation

Spatial autocorrelation is the analysis of the spatial arrangement of the values associated with a set of spatial units [23]. The G_i^* statistic is well-known as a valuable method for determining the degree of local spatial autocorrelation and detecting disease clusters. G_i^* values are presented as a regular normal variant with a z-score distribution probability [24]. In the present study, spatial autocorrelation was used to resolve issues with spatial units that have observable attributes. The Getis Ord's Index was calculated using R statistical software. A clustered pattern is appeared when the G_i^* value is larger than zero and a distributed pattern appeared when the G_i^* value is smaller than zero.

2.5 Space-Time Cluster Analysis

The Poisson based clustering models using space-time scan statistics were executed in SaTScan software v9.6 to detect the temporal, spatial, and space-time clusters of anthrax in Karnataka over the course of 20 years. The model was run on each year's outbreak dataset, adjusting for the underlying population of each village by using the total number of epidemic

outbreaks in that year per village and for all SaTScan clusters, statistical significance was set at a P -value ≤ 0.05 .

2.6 Linear Discriminant Analysis

LDA was first found by Fisher (1936) as a tool for determining the linear combinations of variables best divided observations into groups, or classifications [25][26]. LDA was used to find variations in risk factors between places that were known as a persistent space-time cluster in SaTScan and those that were non persistent cluster. The dependent variable was a binary response (0 or 1) with clustered locations equal to 1 and non-clustered locations equal to 0. Using a binary response (0 or 1), a total of 12 environmental, soil, and remote sensing variables were considered for LDA. The statistical significance of parameters was set at P -value ≤ 0.05 in this study, which was carried out in R software.

2.7 Risk modelling and mapping

The proposed model to assess the risk of anthrax in Karnataka, analogous to the impact of changes in precipitation level using Machine-Learning techniques is depicted in the Figure 1.

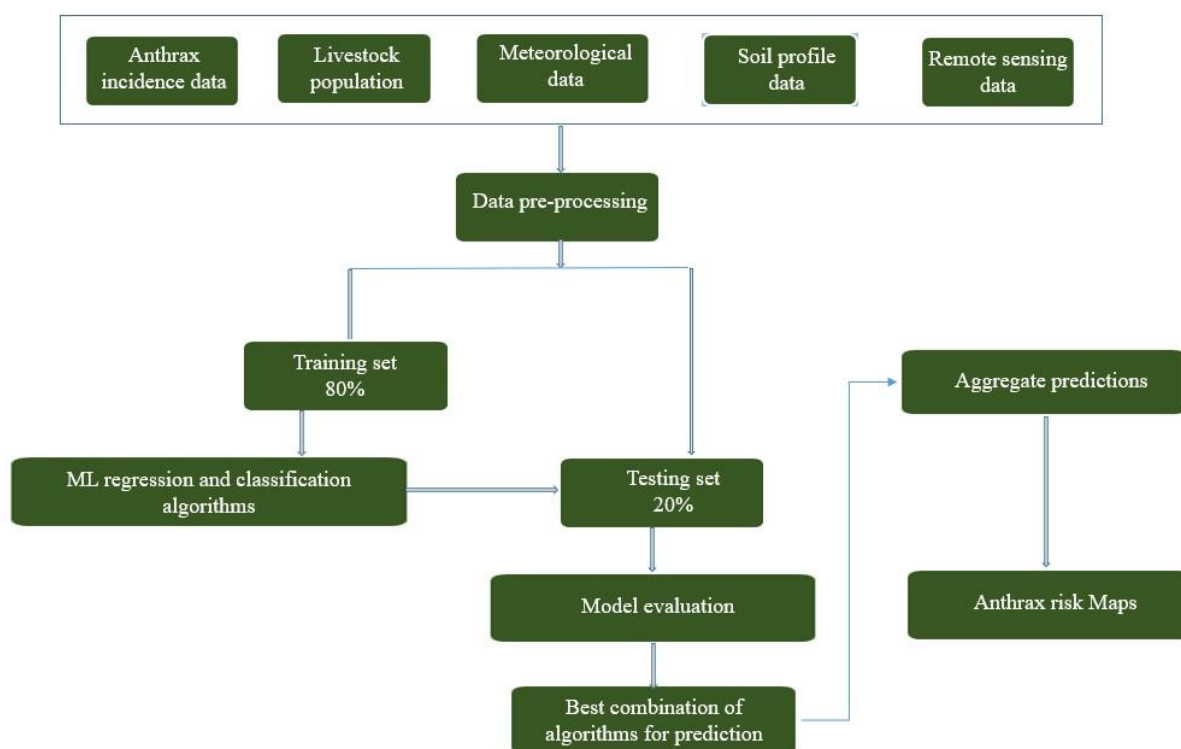


Figure 1. A flowchart depicting potential data included and the model-building process

The risk factor data generated for a period of 20 years (2000-2019) were aggregated to grid level. Anthrax severity was predicted by generating the risk map (2000-2019) for Karnataka state by performing climate-disease relationship modelling which predicts the spatial occurrence of the ailment in accordance with changes in precipitation levels. The data pertaining to risk variables were obtained, pre-processed and were annotated with disease

condition along with the respective latitude and longitude. The risk estimation is done by using machine learning models to produce an optimal prediction with improved performance. In the present study the predictive modelling was developed using several data mining models *viz.* Generalized Linear Models (GLM), Generalized Additive Models (GAM), Classification Tree Analysis (CT), Support Vector Machine (SVM), Artificial Neural Network (ANN), Multiple Adaptive Regression Splines (MARS), Flexible Discriminant Analysis (FDA), Naïve Bayes (NB), Random Forest (RF), Gradient Boosting Machine (GBM) and Adaptive Boosting (ADA) [27]. All the models were evaluated for over-fitting because it can cause misleading of estimated co-efficient, *P*-values and R-Squares values [27]. In this study, the cross-validation procedure was adopted to assess the over-fitting of models by keeping 80% data on the training dataset and 20% data on the test dataset [28]. The discriminating power of the fitted models was evaluated using the Receiver Operating Characteristic (ROC) curve, Cohen's Kappa (Heildke Skill Score), Skill Statistics (TSS), and ACCURACY. The outcomes of separate predictions by multiple model methods were combined using Raster Stack. Rather than depend on a single best model, it is recommended to use the combined prediction results of different models, which are in the scale of 0 to 1. The best prediction was obtained by averaging the scores and average score was obtained by considering the model satisfying the criteria with kappa >0.60, ROC >0.90, TSS >0.80, and ACCURACY >0.90 [29] [30]. Response plots were created to explore and understand model predictions.

2.8 Statistical software

R statistical software version 3.1.3 (<http://www.R-project.org/>) was used to conduct statistical analyses, predictions and risk maps. For data mining, calculation, and graphical display, R was used as an integrated suite. R packages such as *asada*, *biomod2*, *BAMM* tools, *caret*, *dplyr*, *data.table*, *dismo*, *earth*, *e1071*, *foreign*, *geosphere*, *gbm*, *inputemissings*, *kernlab*, *MASS*, *mgcv*, *mda*, *nnet*, *openxlx*, *openxlsx*, *plyr*, *psych*, *pROC*, *qdap*, *raster*, *rgdal*, *randomforest*, *rpart*, *spdep*, *sf*, *sp*, *SDM* tools, and *tmap* were used for data processing, combining, annotating and model fitting. For data extraction, data alignment, annotation, analysis, modelling, and risk mapping, *Gaia's ord's Index* was used. The spatial and temporal clusters across a study area were detected using *SaTScan v9.6*.

3. Results

3.1 Spatial autocorrelation

The data was checked for the existence of any clusters in the entire dataset before conducting hotspot analysis using Getis-Ord G_i^* technique, which is a tool for evaluating global spatial autocorrelation. The presence of clusters within the dataset is required for hotspot analysis, and the output value is stated as a z-score, with a high z-score indicating the presence of hotspots or clusters. A year with a high z-score indicated the existence of hotspots, while a year with a low z-score indicated the presence of cold spots as in Figure 2(a) & (b). For further research and modelling, the Getis-Ord G_i^* analysis is being used to classify villages/districts with a high risk of disease incidence.

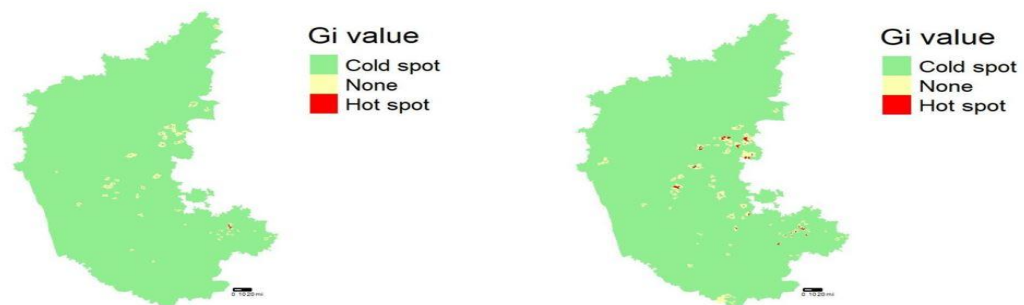


Figure 2. Hotspot analysis (2000-2019), a) Average annual precipitation above normal & (b) Average annual precipitation below normal, respectively.

3.2 Space-Time Cluster Analysis

Space time cluster analysis has shown the existence of disease cluster in the central east region and the southern region during the average annual precipitation above normal and the average annual precipitation below normal revealed the existence of disease cluster in the central east region and the southern east region. Two and three major clusters of high risk were identified in the spatial variation, leading to an increasing trend in anthrax as shown in Figures 3(a) and (b). Between 2000 and 2019, village-level disease clustering was established, with disease incidence characterized by red colored dots inside significant red circles indicating villages with a high risk of disease incidence, and pink circle dots indicating villages with disease incidence but not part of significant clusters.

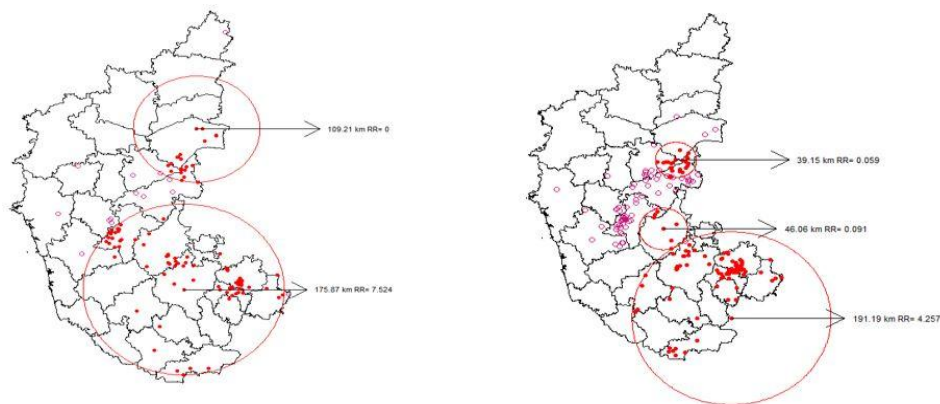


Figure 3. Space-Time Analysis (2000-2019), a) Average annual precipitation above normal & (b) Average annual precipitation below normal, respectively.

3.3 Linear Discriminant Analysis

The discovery of significant disease clusters using a space-time cluster model encouraged us to inspect the ecological, environmental, and other risk factors that contributed to the development of the significant clusters. As a result, we attempted to use Linear Discriminant Analysis to identify significant risk factors (climatic, remote sensing, soil profiles and host) responsible for the development of disease clusters at the village level in the following step. The identified risk parameters were then utilized to model and predict spatial risk. LDA results are tabulated for the average annual precipitation above and below normal as presented Table 1(a) & (b). For disease risk modelling, environmental variables

that were observed to be significantly linked to disease incidence at a p-value ≤ 0.05 were considered.

Table 1(a). Discriminant analysis results of average annual precipitation above normal (2000-2019).

Parameter	Sum sq	Mean sq	F value	P value
Air-temperature	1.230513	1.230513	11.68928	0.000845
EVI	0.112667	0.112667	1.070282	0.302848
LAI	0.031328	0.031328	0.297605	0.586344
LST	0.430313	0.430313	4.087763	0.045296
NDVI	0.092897	0.092897	0.882474	0.349308
PET	0.045199	0.045199	0.42937	0.513484
Potential-evaporation-rate	0.192502	0.192502	1.828679	0.178687
Rain-precipitation-rate	0.001559	0.001559	0.014811	0.903329
Soil-moisture	0.005875	0.005875	0.055814	0.81362
Specific-humidity	0.409899	0.409899	3.893845	0.050635
Surface-pressure	0.011399	0.011399	0.108289	0.742643
Wind-speed	3.288176	3.288176	31.23609	1.33E-07

Table 1(b). Discriminant analysis results of average annual precipitation below normal (2000-2019).

Parameter	Sum sq	Mean sq	F value	P value
Air-temperature	3.915114	3.915114	23.77104	1.82E-06
EVI	1.085331	1.085331	6.589705	0.010779
LAI	0.153783	0.153783	0.933711	0.334737
LST	0.890814	0.890814	5.408673	0.020754
NDVI	0.563229	0.563229	3.419704	0.06548
PET	0.34991	0.34991	2.124515	0.146083
Potential-evaporation-rate	0.05324	0.05324	0.32325	0.570119
Rain-precipitation-rate	1.080929	1.080929	6.562978	0.010938
Soil-moisture	0.103008	0.103008	0.625425	0.42971
Specific-humidity	4.205072	4.205072	25.53155	7.88E-07
Surface-pressure	0.159209	0.159209	0.966652	0.326368
Wind-speed	9.009321	9.009321	54.70106	1.65E-12

The findings of the analysis showed possible risk factors such as air-temperature, LST, specific-humidity and wind-speed when the average annual precipitation was above normal. And when the average annual precipitation was below normal, air-temperature, EVI, LST, rain-precipitation-rate, specific-humidity and wind-speed was turned to be a significant risk factors as shown in Table. 1(a) & (b). The significant clusters were superimposed over the major significant risk parameters, which were found to have a positive impact on disease incidence.

3.4 Risk assessment and estimation

The major ecological and environmental risk factors identified by LDA were subjected to climate-disease modelling. Anthrax-affected (case) and unaffected (control) areas were used to produce maps as shown in Figure 4. The case data is represented by red circles, while the control data is represented by blue dots on the map as shown in Figure 4(a), (b), & (c), red circles represent the places having the disease incidences at different thresholds and blue dots represents the places without incidence of anthrax.

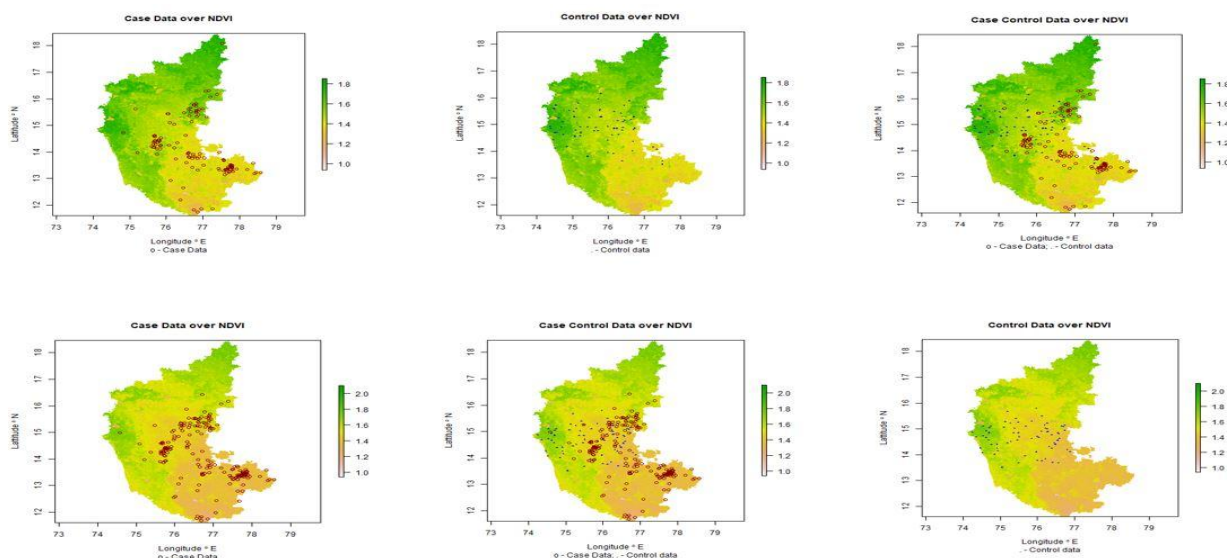


Figure 4(a), (b) & (c).Karnataka map showing the occurrence of outbreak for average annual precipitation above and below normal, respectively. (a) Case Data- Red color circles indicate the places having anthrax incidence, (b) Control Data- blue dots indicating the places having no anthrax incidence, and (c) Case-Control Data- shows both presence-absence of anthrax incidence.

The case-control data were subjected to various models, the best fitted models for the average annual precipitation above normal were RF, GBM & ADA and for the average annual precipitation below normal were RF, GBM, CT& ADA. The best fitted models were selected based on the statistically specified evaluation criteria namely Cohen's Kappa, ROC, TSS, and ACCURACY. Models shown relatively low error (2.2) for training set, whereas for test set shown high error (2.8) which was considered as over fitting. After performing cross validation, training and test set error was 2.4 and was chosen as good fit. Average score was calculated and recorded for the average annual precipitation above and below normal as shown in Table 2(a) & (b).

Table 2(a). Model evaluation table (Average annual precipitation above normal)

Sl. No	Model	Model Specification	KAPPA	ROC	TSS	ACCURACY
1	GLM	$E(Y X) = \mu = g^{-1}(X\beta)$ Y- Expected Value, X-Conditional, $X\beta$ - Linear Predictor, g-Link Function	0.249	0.684	0.25	0.616949

2	GAM	$g(E(Y))$ $= \beta_0 + f_1(x_1)$ $+ f_2(x_2) + \dots + f_i(x_i)$ Y-Response Variable, g-Link Function, f_i -Specified Parametric Form, x_i -Predictor Variable	0.249	0.684	0.25	0.616949
3	RF	$Y = \sum_{i=1}^n f(t_n)$ Y- Average of aggregated predictions of the multiple decision trees, t_n – multiple decision trees trained on different subset of the same training data	0.855	0.999	0.979	0.986441
4	GBM	$f(x)$ $= \arg \min_{\theta} \sum_{i=1}^n L(y_i, \theta)$ $+ \sum_{m=1}^M \eta \rho_m \phi_m(x)$ m- Iteration, η -Learning Rate, ρ_m - Step length	0.691	0.986	0.863	0.932203
5	NNET	$Y = f\left(\sum_{i=1}^n x_i w_i\right) + b$ Y-Output, x_i -Inputs, w_i - Weights, b - Bias	-0.001	0.495	0.001	0.338983
6	MARS	$\hat{f}(x) = \sum_{i=1}^k c_i B_i(x)$ c_i - Constant Coefficient, $B_i(x)$ - Basis Function	0.638	0.956	0.782	0.888136
7	FDA	$\eta_l(x) = X^T \beta_l$	0.324	0.642	0.283	0.735593
8	CT	$f(x) = \sum_{j=1}^T w_j I(x \in R_j)$	0.723	0.937	0.766	0.871186
9	SVM	$\{x: f(x) = x^T \beta + \beta_0 = 0\}$	0.694	0.934	0.788	0.898305
10	NB	$P(c x) = \frac{P(x c)P(c)}{P(x)}$ $P(c x)$ -Posterior Probability, $P(x c)$ -Likelihood, $P(c)$ -Class Prior Probability, $P(c)$ -Predictor Prior Probability	-0.442	0.816	- 0.211	0.325424

11	ADA	$F_T(x) = \sum_{t=1}^T f_t(x)$ f_t - Weak Learner, x - Input, T - T^{th} Positive or Negative Classifier	0.878	0.937	0.874	0.945763
----	-----	--	-------	-------	-------	----------

Table 2(b). Model evaluation table (Average annual precipitation below normal)

Sl. No	Model	Model Specification	KAPPA	ROC	TSS	ACCURACY
1	GLM	$E(Y X) = \mu = g^{-1}(X\beta)$ Y- Expected Value, X-Conditional, $X\beta$ - Linear Predictor, g-Link Function	0.367	0.855	0.612	0.793258
2	GAM	$g(E(Y))$ $= \beta_0 + f_1(x_1)$ $+ f_2(x_2) + \dots + f_i(x_i)$ Y-Response Variable, g-Link Function, f_i -Specified Parametric Form, x_i -Predictor Variable	0.367	0.855	0.612	0.793258
3	RF	$Y = \sum_{i=1}^n f(t_n)$ Y- Average of aggregated predictions of the multiple decision trees, t_n – multiple decision trees trained on different subset of the same training data	0.87	0.999	0.988	0.988764
4	GBM	$f(x)$ $= \arg \min_{\theta} \sum_{i=1}^n L(y_i, \theta)$ $+ \sum_{m=1}^M \eta \rho_m \phi_m(x)$ m- Iteration, η -Learning Rate, ρ_m - Step length	0.587	0.983	0.879	0.916854
5	NNET	$Y = f(\sum_{i=1}^n x_i w_i) + b$ Y-Output, x_i -Inputs, w_i - Weights, b- Bias	0.001	0.495	0.001	0.224719
6	MARS	$\hat{f}(x) = \sum_{i=1}^k c_i B_i(x)$ c_i - Constant Coefficient, $B_i(x)$ - Basis Function	0.559	0.97	0.812	0.905618

7	FDA	$\eta_l(x) = X^T \beta_l$	0.51	0.708	0.416	0.817978
8	CT	$f(x) = \sum_{j=1}^T w_j I(x \in R_j)$	0.793	0.971	0.819	0.934831
9	SVM	$\{x: f(x) = x^T \beta + \beta_0 = 0\}$	0.659	0.939	0.798	0.889888
10	NB	$P(c x) = \frac{P(x c)P(c)}{P(x)}$ $P(c x)$ -Posterior Probability, $P(x c)$ -Likelihood, $P(c)$ -Class Prior Probability, $P(c)$ -Predictor Prior Probability	-0.215	0.71	- 0.124	0.229213
11	ADA	$F_T(x) = \sum_{t=1}^T f_t(x)$ f_t - Weak Learner, x - Input, T - T^{th} Positive or Negative Classifier	0.936	0.975	0.95	0.955056

3.5 Risk prediction and mapping

Data with average annual precipitation above normal was predicted with high disease risk in the north, northern east and the state's southern region as shown in Figure 5(a). And the data with average annual precipitation below normal was predicted with high disease risks in south, northern east and the state's central region indicating severe disease can be expected in these areas as shown in Figure 5(b). While the western region of the state was recognized as having a low risk in both groups. Risk map is a digital tool that provides a more holistic view of the likelihood and effect of disease, as well as the opportunity to build synergies in a specified study area. It raises public awareness and encourages policymakers and planners to take proactive action to minimize risk to life and strengthen risk management and governance by highlighting risk management initiatives. A new statistical method for risk mapping was introduced in this study to improve the accuracy of short-term risk prediction. The risk map was generated using the biomod2 package in R software. Machine learning models were used to model the disease data with significant predictor variables defined by the LDA function, such as ecological data, remote sensing data, and soil parameters.

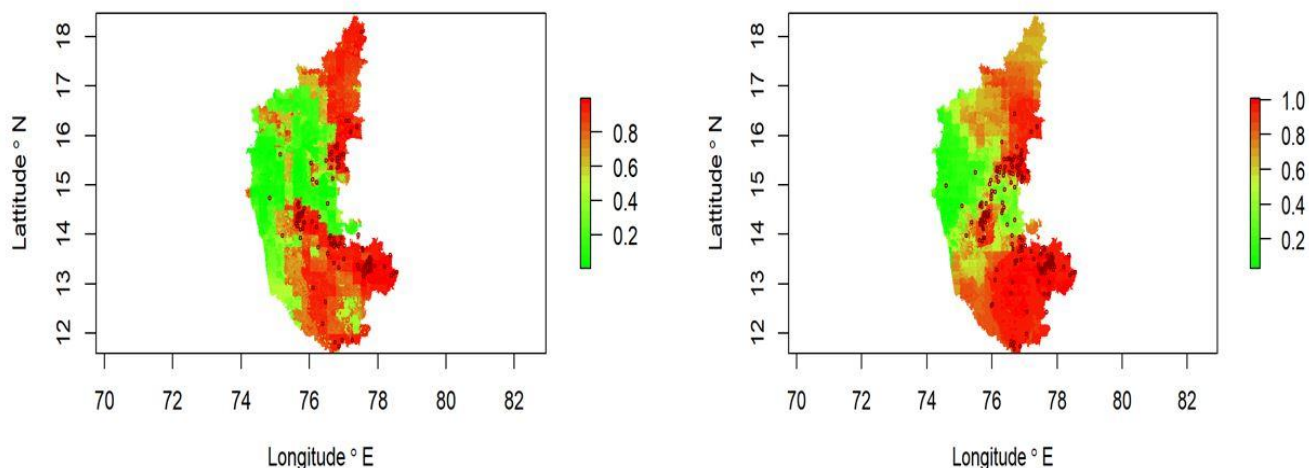


Figure 5. Anthrax risk prediction map (2000-2019) a) Average annual precipitation above normal & (b) Average annual precipitation below normal, respectively.

4. Discussion

In support of the literature findings, it is consistently observed globally that anthrax is a hot season disease [10]. Anthrax seasons are said to be marked by hot, dry weather, which distress animals and lowers their natural resistance to infection [16]. Hot weather can affect host resistance indirectly by causing changes in nutrition and behaviour, as well as altering nonspecific local resistance of the skin and mucous membranes, allowing pathogen to enter more easily. UV light, temperature extremes, drying, and high pH are all highly resistant to *B. anthracis* spores [5]. Anthrax has a broad geographical range, with some areas having a high risk of persistence infection and others having a low risk [9]. Anthrax can be restricted to specific areas within landscapes that have similar weather patterns. Anthrax outbreaks affect a wide variety of animals, and those hot and dry environments play an important role in outbreaks, whether by changes in animal behaviour or changes in pathogen ecology. Since soil is the site of long-term pathogen survival in the ecosystem, regional outbreak restriction and permanently infected anthrax zones may be attributed to soil conditions and their fluctuation caused by precipitation events [22]. In support of the present findings, it is observed that a relatively a smaller number of districts were classified as high-risk regions in the data with average annual precipitation above normal compared to the data with average annual precipitation below normal. Environmental optimum conditions for sustaining the presence of *Bacillus anthracis* in soil have been attributed to the existence of high-risk areas for anthrax, which include a neutral or slightly alkaline pH of the soil, adequate nutrients, relative humidity, adequate temperature, wind speed, land surface temperature, Enhanced vegetation index, and precipitation rate. As the precipitation is below average the weather will consistently become hot and dry leads to the high risk of disease incidence. The results also predicted, there is a chance of disease incidence in low-risk regions as well. This may happen due to the infected animal movement between the regions. If there is no pathogen existence in the region but soil and environment is conducive and pathogen is introduced via animal movement, then the chance of infection occurrence will be high. If both soil and environment is non-conducive and pathogen is introduced then the chance disease incidence will be minimal. If either soil or environment is conducive and pathogen is introduced there

will be a risk of disease incidence. And if both environment and soil is conducive but no pathogen is induced then the chance of disease occurrence may be negotiable.

As per the predicted results, approximately 43% and 57% of the area is under high risk during the average annual precipitation above and below normal respectively. The livestock population (Cattle, Buffalo, Sheep and Goat) in these areas was found to be 2,39,668 and 6,44,435. The literature findings show 25% of anthrax prevalence in Asia [10]. Henceforth 59,917 and 1,61,108 livestock may be susceptible to anthrax. There are 1917 and 2372, total veterinary institutions functioning in the predicted areas [31] as in Table 3. Thus, with better medical resources and with an effective vaccination program in the high-risk regions it would be possible to control the infections and to reduce the cost of the diagnosis.

Table 3. Impact analysis

Sl. No	Description	Average annual precipitation above normal	Average annual precipitation below normal
1	% Of area covered under risk area >0.60	43%	57%
2	Population (C, B S, G) in these area	239668	644435
3	Prevalence of Anthrax in Asia	25%	25%
4	Total cases predicted as per prevalence	59917	161108
5	Total veterinary Institutions functioning in the Predicted Districts	1917	2372
6	Response/actions to be taken	Effective vaccination program	Effective vaccination program

However, risk assessment modelling based on ecological, climate, and host data has the capacity to predict disease outbreaks in advance and provide opportunities to reduce disease transmission, this approach is likely to have a much lower predictive value due to the uncertainties associated with most climate–disease relationships and the confounding influences of other factors. Although precise disease outbreak predictions cannot be formed merely based on environmental observations and host distribution, this information could be utilized to issue a warning that environmental conditions are favorable to disease outbreak. To achieve an effective risk reduction function, a forewarning system should be viewed as an information system that allows relevant national and regional institutions to make informed decisions, also vulnerable dairy farmer groups to take proactive measures to minimize the impacts of impending livestock disease outbreaks. Machine learning models have shown to be effective at mining data from numerous sources to detect geographic hotspots and epidemic risks. Despite the power of climate, ecological and disease risk models, considerable uncertainties remain, identifying these uncertainties, highlighting importance of improved data may improve the model accuracy, realism, confidence, together with translating uncertainties in model inputs into uncertainties in model outputs, are important benefits of modeling.

5. Conclusion

The current study used anthrax incidence data and the risk factor data to generate risk maps with different levels of risk. The results exhibited that the changes in precipitation level have positive impact on the disease risk. Southern, northern and eastern Karnataka was predicted with higher risk of anthrax outbreaks than the west of the state. Soil type, soil pH, air temperature, EVI, LST, specific humidity, wind speed and precipitation rate were observed to be the most important factors in anthrax persistence. The characteristic of geographic distribution and seasonal variations of anthrax are the prima facia evidence of linkages of weather and climate and potentially affect the timing and intensity of disease outbreaks. The study findings help to develop early warning system includes vulnerability and risk analysis, feasible response plans and strategies for effective veterinary and public health communication. Study also helps to develop the linkages between climate and infectious disease, can be greatly aided by the focussed efforts to employ the recent technological advances such as remote sensing of ecological changes and high-speed computational modelling to track the geographic distribution of anthrax. However, the inadequacy of high-quality epidemiological data is a major obstacle to enhance our understanding of the climate-disease relationship. A concerted exertion should be made to gather long term spatially resolved disease surveillance data, along with suitable set of meteorological and ecological observations for developing and validating robust disease models.

Acknowledgment

We would like to thank the Spatial Epidemiology lab, Indian Council of Agricultural Research (ICAR) - National Institute of Veterinary Epidemiology and Disease Informatics (NIVEDI), Department of Veterinary Public Health and Epidemiology, Veterinary College and Outreach project on Zoonotic diseases, ICAR for providing necessary support to carry out this research work.

Ethical statement

Authors declare that ethical statement is not applicable as we have not collected any animal samples for the study.

Author's contributions

SB: Collected the data, conducted the research work and wrote the manuscript, SS: Performed review and supervision. KPS: Guided the research steps, LG: Performed writing review and editing, SSP: Performed writing review and editing. All authors read and approved the final manuscript.

Conflict of interest

The authors declare no conflict of interest.

Funding statement

No fund was received for this particular study.

References

- [1] EE Rees et al., “Risk assessment strategies for early detection and prediction of infectious disease outbreaks associated with climate change,” *CCDR*, vol. 45, no. 5, pp. 119-126, 2019.
- [2] D. P. Kshirsagar et al., “Disease alerts and forecasting of zoonotic diseases: an overview,” *Veterinary World*, vol. 6, pp. 889-896, 2013.
- [3] B. I. a. J. R. Vladimir Savransky, “Current Status and Trends in Prophylaxis and Management of Anthrax Disease,” *MDPI Pathogens*, 2020.
- [4] Joakim A gren et al., “Microevolution during an Anthrax Outbreak Leading to Clonal Heterogeneity and Penicillin Resistance,” *PLOS ONE*, p. Volume 9 Issue 2, 2014.
- [5] Serkan Oncu et al., “Anthrax – an overview,” *Med Sci Monit*, vol. 9, no. 11, pp. 276-283, 2003.
- [6] Sushma Bylaiah et al., “Relative Analysis of Codon Usage and Nucleotide Bias between Anthrax Toxin Genes Subsist Inpxo1 Plasmid of Bacillus Anthracis,” *Annals of R.S.C.B*, pp. 5758 - 5774, 2021.
- [7] P Krishnamoorthy et al., “Trends and future prediction of livestock diseases outbreaks by periodic regression analysis,” *Indian Journal of Animal Sciences*, p. 369–376, 2019.
- [8] K P Suresh et al., “Application of Artificial Intelligence for livestock disease prediction,” *Indian Farming*, pp. 60-62, 2019.
- [9] Suma A P et al., “Outbreak Prediction of Anthrax in Karnataka using Poisson, Negative-Binomial and Zero Truncated Models,” *INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH*, pp. 505-509, 2017.
- [10] Bylaiah Sushma et al., “An Estimate of Global Anthrax Prevalence in Livestock: A Meta-analysis,” *Veterinary World*, pp. vol 14, EISSN: 2231-0916, 2021.
- [11] H. M. Vinaya Kumar et al., “Rainfall Trend Analysis and Adaptation Strategies to Manage Climate-Induced Crisis in Coastal Zone of Karnataka, India,” *Journal of Scientific Research & Reports*, pp. 13(5): 1-11, 2017.
- [12] Z. Maksimovic et al., “The apparent role of climate change in a recent anthrax outbreak in cattle,” *Rev. Sci. Tech. Off. Int. Epiz*, pp. 36 (3), 959-963, 2017.
- [13] Rong-Gang Cong et al., “The Interdependence between Rainfall and Temperature: Copula Analyses,” *The Scientific World Journa*, pp. 405675, 11, 2012.
- [14] A.K.M Sazzadur Rahman et al., “A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning Algorithms,” *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, pp. 419-422, 2019.
- [15] Michael G. Walsh et al., “Climatic influence on anthrax suitability in warming northern latitudes,” *SCIeNTIFIC Reports*, p. 8:9269, 2018.
- [16] S.M. Chikerema et al., “Spatial modelling of Bacillus anthracis ecological niche in Zimbabwe,” *Preventive Veterinary Medicine* , pp. 25-30, 2013.
- [17] Bushra Praveen et al., “Analyzing trend and forecasting of rainfall changes in India

- using non parametrical and machine learning approaches,” *Scientific Reports*, p. 10:10342, 2020.
- [18] Rong-Gang Cong et al., “The Interdependence between Rainfall and Temperature: Copula Analyses,” *The Scientific World Journal*, p. 405675, 2012.
- [19] V. M. S. & S. R. M.N.Chakari, “ANNUAL SEASONAL RAINFALL & AREA COVERAGE DURING 2011 IN KARNATAKA,” Government of Karnataka, Bangalore, 2012.
- [20] Ian T. Kracalik et al., “Modeling the environmental suitability of anthrax in Ghana and estimating populations at risk: Implications for vaccination and control,” *PLOS Neglected Tropical Diseases*, 2017.
- [21] C.O. Justice et al., “An overview of MODIS Land data processing and product status,” *Elsevier Science Inc*, p. 3 – 15, 2002.
- [22] Michael H. Norris et al., “Linking Geospatial and Laboratory Sciences to Define Mechanisms behind Landscape Level Drivers of Anthrax Outbreaks,” *Int. J. Environ. Res. Public Health*, pp. 16, 3747, 2019.
- [23] Praprut Songchitruksa et al., “Getis–Ord Spatial Statistics to Identify Hot Spots by Using Incident Management Data,” *Journal of the Transportation Research Board*, p. 42–51, 2010.
- [24] Ian T. Kracalik et al., “Analysing the spatial patterns of livestock anthrax in Kazakhstan in relation to environmental factors: a comparison of local (Gi*) and morphology cluster statistics,” *Geospatial Health*, pp. 111-126, 2012.
- [25] Peter Boedeker et al., “Linear Discriminant Analysis for Prediction of Group Membership: A User-Friendly Primer,” *Advances in Methods and Practices in Psychological Science*, p. Vol. 2(3) 250–263, 2019.
- [26] A. Izenman, “Linear Discriminant Analysis,” *Springer Texts in Statistics*, 2013.
- [27] Pankaj Mehta et al., “A high-bias, low-variance introduction to Machine Learning for physicists,” *Elsevier*, pp. 1-124, 2019.
- [28] Johnson Kolluri et al., “Reducing Overfitting Problem in Machine Learning Using Novel L1/4 Regularization Method,” in *Proceedings of the Fourth International Conference on Trends in Electronics and Informatics* (, 2020).
- [29] S.K. Abdrakhmanov et al., “Maximum entropy modeling risk of anthrax in the Republic of Kazakhstan,” *Elsevier-Preventive Veterinary Medicine*, pp. 144, 149–157, 2017.
- [30] Omri Allouche et al., “Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS),” *Journal of Applied* , p. 1223–1232, 2006.
- [31] G. o. Karnataka, “VETERINARY INSTITUTIONS FUNCTIONING IN THE STATE DISTRICTWISE, TALUKAWISE, PLACEWISE,” DEPARTMENT OF ANIMAL HUSBANDRY AND VETERINARY SERVICES, 2017.