International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Novel Ensemble Learning Technique for Lumpy Skin Disease Classification

Venkata Pratyusha Kumari Sunkara[1], Ramesh Chappa*[2], JagadeeswaraRao G.[3]

Submitted: 28/01/2024 **Revised**: 06/03/2024 **Accepted**: 14/03/2024

Abstract: Lumpy skin disease is a highly contagious viral disease that affects cattle and has significant economic implications for the livestock industry. It has a direct relationship with climate, as the latter plays a major role in studying the infection and the pattern of transmission followed by it. This study illustrates how different climate parameters contribute to the diagnosis of LSD in cattle in a certain nation or location. In recent years, there has been a growing interest in using machine learning algorithms and geographical data to predict and mitigate the spread of LSD. Several studies have been conducted to explore the potential of machine learning (ML) models for predicting LSD outbreaks based on geographical data. In this paper, we built an effective ensemble ML model called gradient boosting (GB) to classify LSD. Our model outperforms other standalone ML models like random forest (RF), support vector machine (SVM), k-nearest neighbors (KNN), decision tree (DT), logistic regression (LR), and extreme GB (XGB). Further, we have compared our results against the recent past study results, and our model outperformed with accuracy (98.97) and ROC-AUC (0.99). By using these techniques, one can improve one's capacity to predict different diseases and natural occurrences.

Keywords: *ensemble method, gradient boosting, lumpy skin, machine learning*

1. Introduction

The lumpy skin diseases (LSD) virus, which is related to the poxviridae and capripoxvirus genera, respectively, is the extremely infectious agent that causes lumpy skin disorders [1-2]. Sucking insects carry the parasite that causes lumpy skin illnesses, which are mostly found in Africa but have lately expanded to regions of Asia and Europe [3]. Due to animal deaths, decreased milk production, damage to hides, and a high incidence of infertility, lumpy skin disorders have resulted in significant financial losses. The World Organisation for Animal Health lists lumpy skin disorders as the infectious illnesses that are most evident in the most impacted or epidemic nations [4]. The virus that causes lumpy skin disorders has impacted 29966, 8837, and 2471 cattle in Africa, Asia, and Europe, respectively, between 2005 and 2021. The transmission of lumpy skin disorders in cattles and buffaloes may occur via a variety of routes. These include scabs and crusts, skin nodules, and other materials that can be removed for up to 35 days (Fig. 1). Measures such as the diagnosis of index cases of lumpy skin illnesses and the broad implementation of vaccination campaigns should help reduce the number of occurrences of this disease, according to records of recent epidemic cases in Western Asia.

[1] *Department of Computer Science and Engineering, Aditya Institute of Technology and Management, Tekkali – 532201, India*
ORCID ID: 0009-0007-4391-3108
[2] *Department of Computer Science and Engineering, Aditya Institute of Technology and Management, Tekkali – 532201, India*
ORCID ID: 0000-0001-9829-0146
[3] *Department of Information Technology, Aditya Institute of Technology and Management, Tekkali – 532201, India*
ORCID ID: 0000-0003-0171-6337
** Corresponding Author Email: chappa.ramesh01@gmail.com*

These machine learning methods are used to create more accurate prediction models, regardless of whether they are learned under supervision or not [5]. A branch of machine learning algorithms known as reinforcement learning is the science of decision-making. It uses artificial intelligence to teach itself via trial and error and feedback from its actions [6]. For improved prediction performance, recent research has encouraged the use of ensemble approaches, which combine two or more algorithms and are used across several fields. Large amounts of clinical data need to be stored and organised, analysed, and developed into data mining techniques to uncover hidden patterns and unearth new information from clinical data. These tasks are becoming more important for health care applications [7]. ML plays a crucial role in genetics by aiding in various aspects such as predicting genetic disorders, extracting gene-disease associations, and modeling gene-gene interactions. In the context of genetic disorder prediction, a machine learning model was developed to predict the presence of genetic disorders and specify the disorder subclass, showcasing the importance of such models in healthcare.

Afshariet al. [8] used meteorological and geographic data to create a unique artificial neural network model that effectively predicts lumpy skin disorders. The constructed model has training and testing accuracies of 94% and 97%, respectively, which allowed it to diagnose lumpy skin disorders well.In addition, writers in [9] used k-nearest neighbour, RF, SVM, and multilayer perception (MLP) algorithms to categorise the incidence of LSD infection. With a 97.7% accuracy rate, the RF algorithm outperforms the others. The goal of this study is to use machine learning

classifiers in conjunction with feature wiz feature selection on a dataset of several detected characteristics related to lumpy skin conditions. Various ML methods have been used in various publications to predict lumpy skin disorders (LSD), applied a variety of machine learning techniques to the LSD dataset containing meteorological and geographic features in order to predict lumpy skin diseases [10]. These techniques included LR, SVM, DT, RF, and artificial neural network. The pre-processed features in the dataset were separated using a percentage split method and one hot encoding technique. According to an evaluation of the model, the artificial neural network outperformed

Fig. 1. Excerpts of LSD virus [20]

all other algorithms, showing 97%, 88%, 94%, and 97% accuracy, precision, and f1 score, respectively.

The outcomes of machine learning models for the prediction of lumpy skin disorders were compared by the authors in reference [11]. The lumpy skin diseases dataset was subjected to ten different classifiers in order to identify the methods that would provide the greatest prediction performance. With a 98% F1 score, the findings demonstrated that the RF classifier combined with the light gradient boosted machine classifier performed better than the other models. The prediction of lumpy skin illnesses was compared between RF and KNN algorithms by pal et al. in [12]. Because standard medical diagnostic methods are more time- and money-consuming, a unique model was created to classify the major incidence of lumpy skin illnesses. KNN beat RF in the constructed model assessment, with accuracy and F1 score of 95.23% and 95.98%, respectively.By using RF and hyperparameter adjustment, a unique model was created [13]. After the model was assessed, it was discovered that using RF in conjunction with SMOTE and GA improved the area under the curve (AUC) from 94% to 98% and the recall from 90% to 99%.

In order to effectively forecast lumpy skin illnesses based on picture datasets, the authors of [14] used a variety of machine learning methods, including DT, SVM, boosted trees, k-means clustering, convolutional neural networks, and genetic algorithms. The convolution neural networks performed better than other machine learning techniques. Another innovative model for the accurate prediction of lumpy skin disorders was created by [15] using a convolutional neural network. A classification accuracy of 95% was achieved for the input system.For the purpose of predicting lumpy skin disorders, models including ARIMA, FTS, and NNAR were introduced by [16]. Out of seven datasets used, fuzzy time series outperformed the other models in five of them.By integrating Long Short-Term Memory (LSTM) with MobileNetV2, skin illnesses were categorised by [17]. The created model has 86.57%, 93.34%, 92.68%, and 86.34% for accuracy, precision, f-measure, and recall, respectively. Kumar et la. [18] focused on utilising several machine learning models to predict lumpy skin disease in cattle and water buffalo. The assessment was based on metrics like RMSE, MAE, MSE, and accuracy. The capability of the RF and Poly-F models to predict LSD was shown; RF had the best accuracy, at 97.10%, while Poly-F came in second, at 96.77%. The prediction of lumpy skin disorders using the RF and KNN algorithms was compared by the authors of [19]. In the developed model, KNN outperformed RF, obtaining an accuracy of 95.23% and an F1 score of 95.98%.

2. Materials and methods

This section discusses the dataset, methods used and methodology followed in our work.

2.1. About Dataset

LSD occurrence based on a meteorological and geospatial feature dataset obtained from the database contributed by Ehsanallah Afshari Safavi [21]. The dataset consists of 24803 records and 20 features (19 independent and one dependent). As a first step in the data pre-processing, we removed a few irrelevant features such as X, Y (latitude and longitude coordinates of a data point), region, country, reportingDate, X5_Ct_2010_Da, and X5_Bf_2010_Da from the dataset, resulting in 13 features (including the target column) remaining in our dataset. The description of the LSD dataset with 13 features is shown in Table 1. The features F1 to F12 are the independent variables with numerical data, and C01 is a target class with categorical data with two classes (class 0 indicates ‘not lumpy’, class 1 indicates ‘lumpy’). The data was collected from different geographical regions, such as Asia, Africa, and Europe.

Where the features represent several climatic and environmental conditions of the data points.

Table 1. LSD dataset description

Feature No. (Feature)	*Description*	*Data Type*	*Min*	*Max*	*Mean±(Std)*
F1 (cld)	cloud cover	Numerical	0	98.7	59.453±(19.423)
F2 (dtr)	daytime temperature range		2.000	20.60	9.107 ±(2.988)
F3 (frs)	frost occurrence		0	31.00	23.978 ±(11.518)
F4 (pet)	Potential Evapotranspiration		0	7.5	0.804±(1.173)
F5 (pre)	precipitation		0	341.9	26.272±(33.631)
F6 (tmn)	minimum temperature		-52.10	23.9	-15.79±(17.588)
F7 (tmp)	temperature		-48.1	28.5	-11.22±(28.5)
F8 (tmx)	maximum temperature		-44.2	36.4	-6.682±(18.541)
F9 (vap)	vapor pressure		0	28.6	3.729±(4.953)
F10 (wet)	wetness		0	30.92	8.543±(6.205)
F11 (elevation)	elevation		66	249	164.769±(19.679)
F12 (dominant_land_cover)	land_cover		0	11	4.416±(2.40)
C01(lumpy) (Target Class)	Lumpy (yes/no)	Categorical	0	1	-

2.2. Data Preprocessing

As the preprocessing of the data is a critical foundation for successful ML models. It improves the model's performance, training time, and prevents overfitting. With the well-prepared data one can ensure building an accurate and generalized model. In our work, we have done the following preprocessing steps in order to improve the dataset quality.

2.2.1. Handling Outliers

Outliers are data points that deviate greatly from the rest of the data's regular trend. They may result from inaccurate data collection or measurement, or they may just reflect uncommon but real occurrences. The outliers in the dataset need to be handled before it supplied to a ML model to avoid distorted results, hidden insights, mislead decisions, and overfitting [22]. There are several techniques to deal with the outliers such as detect and remove or impute them with a computed value. In our work we have imputed the outliers using a inter quartile range (IQR) method [23] shown in Figure 2.

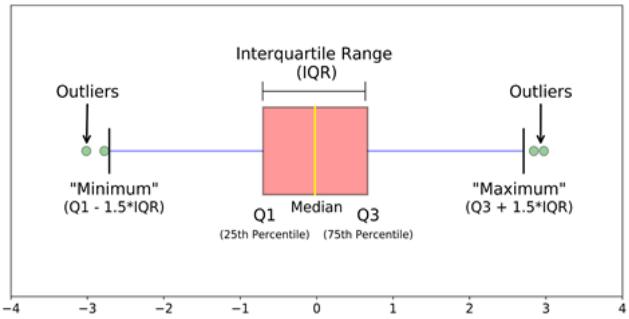

Fig. 2. Detecting outliers using IQR method [24]

The process of IQR method is as follows:

i. sort the dataset in ascending order

ii. compute Q1 (first quartile) and Q3 (third quartile).

iii. compute IQR = Q3-Q1

iv. then, find lower bound = $(Q1 - 1.5 * IQR)$,

upper bound = $(Q3 + 1.5 * IQR)$

v. For all data points, do if it falls under the lower bound and above the upper bound, then mark it as an outlier.

2.2.2. Scaling and Balancing dataset

Scaling the dataset ensures all features have a fair say in the learning process, helps algorithms learn more efficiently, and caters to the specific needs of certain machine learning

models. It's a fundamental step in preparing your data for building robust and accurate machine learning models [25]. Except the target variable (C01), rest of the features were scaled using mean-based scaling method. The features were scaled using the formula shown in (1).

$$X_{new} = \frac{X_i - X_m}{sd} \quad (1)$$

Here, X_i andX_m represents i[th] feature, and its mean value.sd is the standard deviation of X_i. X_{new} is a new transformed feature.

Our dataset is completely imbalanced as there is a huge difference in the number of instances in each class. Balancing the data is a good procedure especially when dealing with classification tasks [26]. Imbalanced dataset leads to bias towards majority class and poor performance with minority classes. In our dataset, class 0 (not lumpy) is a majority class with 21764 instances and class 1 (lumpy) is a minority class with 3039 instances. Here, we used a technique called synthetic minority oversampling technique (SMOTE) in order to balance our dataset. It creates similar samples like minority class instances to balance with majority class instances. The working of SMOTE is as follows:

i. identifies the minority samples

ii. For each data point in minority samples, find the nearest neighbour (NN) from the same class.

iii. Randomly selects a NN and create a synthetic data point that aligns along the line segment connecting them.

iv. step ii and iii will be repeated for number of times upto the minority class samples equal to majority class.

SMOTE will helps in improving performance of ML models by making them learn more effectively from all the classes equally. Also, it reduces the bias towards majority class, leads to fairer predictions.

2.2.3. Feature selection

Feature selection (FS) is a crucial process in data analysis, which involves selecting subset of the most relevant features from the data. It helps in improving accuracy, reducing overfitting, and faster training by models. There are several approaches for feature selection like supervised and unsupervised. Supervised FS selects the features based on the target variable. Where as in unsupervised FS completely relies on the inherent characteristics of the data. We have used an unsupervised FS technique called principal component analysis (PCA) for selecting features. PCA is a dimensionality reduction technique helps us to condense the data to a lower dimensional space. It does this by identifying new features, called principal components (PCs), that represent the directions of greatest variance in the data. While using with ML models PCA reduces the complexity and improves the model performance.

2.3. ML models

In our work we have experimented on various supervised ML models such as SVM, logistic regression, decision tree, KNN, RF and several boosting ensemble models like GB, and XG Boost.

2.3.1. Boosting ensemble models

Boosting is an ensemble technique that enhances the performance of a weak learner by iteratively training and combining multiple weak learners into a single strong learner.By combining multiple learners, boosting can significantly improve the overall accuracy of the model compared to a single weak learner. It can be effective for imbalanced datasets by giving more weight to the minority class during training. Adaboost and gradient boosting (GB) are two commonly used boosting algorithms. GB uses the gradients of a loss function to guide the training process. The loss function measures how well the current model is performing. The gradient indicates the direction of the steepest descent for minimizing the loss. It follows a sequential learning process in which models are trained sequentially with each one attempting to correct the errors made by the previous model(s) based on the loss function's gradient. Algorithm of GB is given in Table 2.

Table 2. Algorithm for GB

Step 1: Train a weak learner, like a small decision tree, on the entire dataset.
Step 2: Compute the gradients of the loss function for each data point based on the initial model's predictions. These gradients represent the errors the model made on each data point.
Step 3: Train a new weak learner to minimize these errors, focusing on the data points with higher gradients (larger errors).
Step 4: Combine the predictions from all the weak learners using weighted summation, where the weights are determined by the model's performance on the training data.
Step 5: Repeat steps 2-4 until a stopping criterion (e.g., maximum number of trees, minimal loss reduction) is met.

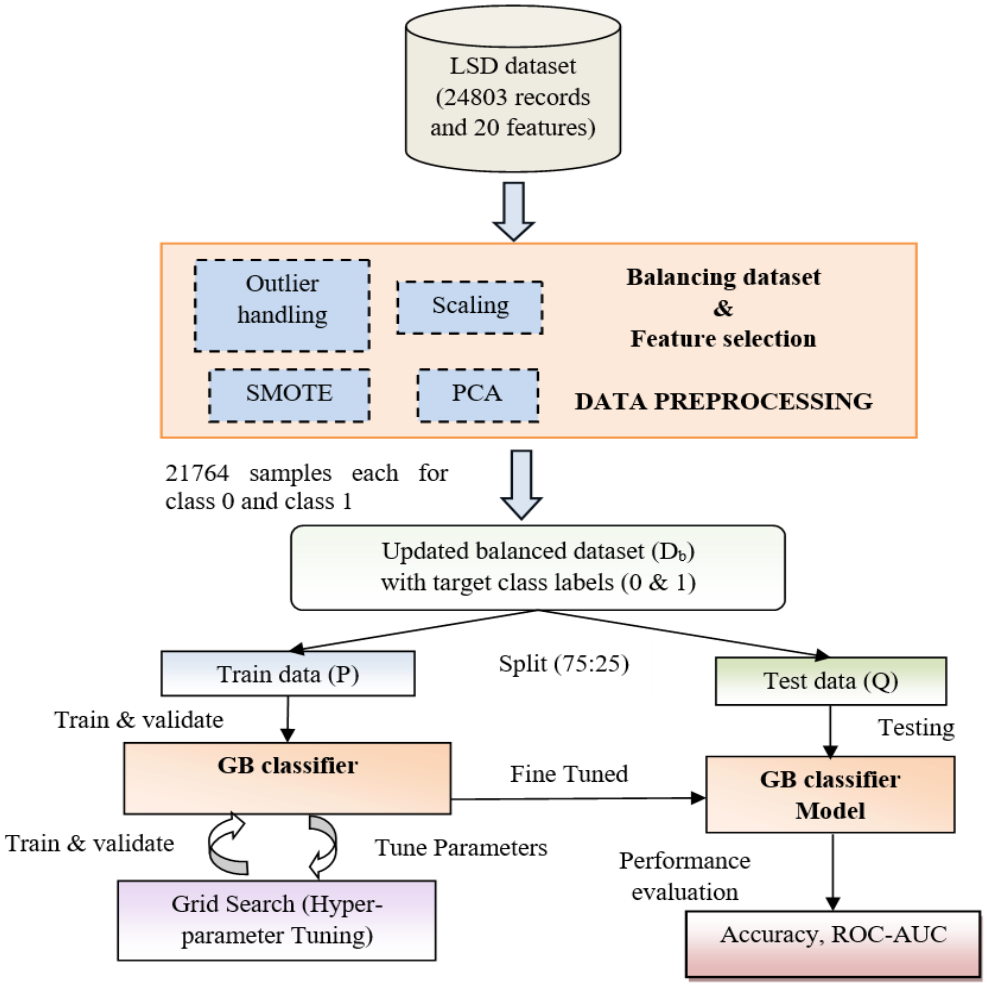

Fig. 3. Methodology of our work

2.4. Model Performance Evaluation

The metrics used to evaluate the ML model's classification performance is confusion matrix, accuracy, precision, recall, F1-score, receiver operating characeristics – area under curve (ROC-AUC) and mathew's correlation coefficient (MCC). Figure 3 shows the complete methodology of our work. The class-wise distribution of the classification model's output is called a confusion matrix (CM). True positive (TP), false positive (FP), true negative (TN), and false negative (FN) are examples of common classes in CM. The ratio of properly identified samples to all samples in the dataset indicates a model's accuracy [26]. Equations (2), (3), (4), and (5) show the accuracy, precision, f1-score, and recall assessment metrics used to evaluate the new model (M). The following are the mathematical formulas for certain measures. Equation (2) expresses accuracy as the ratio of the number of assessed cases to the number of anticipate distances. The ratio of accurately predicted positive cases to all positive instances in a positive class is known as precision. The estimate of a positive event that is accurately identified as such is known as recall. Equation (5) is used to compute the F1 score, an assessment measure that assesses a dataset's correctness.

$$M_{accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (2)$$

$$M_{precision} = \frac{TP}{(TP+FP)} \quad (3)$$

$$M_{recall} = \frac{TP}{(TP+FN)} \quad (4)$$

$$M_{F1-score} = \frac{2}{\left(\frac{1}{M_{precision}} + \frac{1}{M_{recall}}\right)} \quad (5)$$

The false positive rate (FPR) of the model is the ratio between FP and the total number of samples that are actually negative. It is calculated using (6), shown below.

$$M_{FPR} = \frac{FP}{(FP+TN)} \quad (6)$$

The ROC-AUC metric quantifies how well the model can be distinguished between classes. Better model predictions are indicated by a higher AUC value [27]. Plotting of ROC takes place between TPR and FPR on the Y and X axes. Since our problem involves a multi-class classification,

binarizing the expected output is necessary to achieve both FPR and TPR. The One versus. One (OvO) approach or the One vs. Rest (OvR) method can be used to accomplish this. Every class is compared to every other class in the first procedure. The second method contrasts each distinct pairwise pairing of classes. We used the OvR approach for binarization in our study.For our model evaluation, we employed MCC. The MCC will evaluate the categories' quality. It is applicable to multiclass

and binary classifications [28]. It is the most effective way to condense the confusion matrix. Equation (7) shown below, is used to determine a model's MCC value.

$$M_{MCC} = \frac{(TN \times TP) - (FN \times FP)}{\sqrt{(FP+TP)(FN+TP)(TN+FN)(TN+FP)}} \qquad (7)$$

Out of the twelve features, eight had outliers, while the remaining features had none at all. Table 3 lists the number of outliers found and imputed for each feature using the quartile technique. The degree and direction of a linear link between two variables are quantified by correlation analysis. It's an essential tool for exploratory data analysis. The correlation among the features in our independent data (x) is shown as a heatmap in Figure 3. From the figure, it is worth noting that the feature F3 (frs) is having a strong negative correlation with features F9 (vap), F4 (pet), F7 (tmp), F8 (tmx) and F6 (tmn) with correlation value -0.91, -0.89, -0.86, -0.86, and -0.85 respectively. The feature F6 (tmn) having a strong positive correlation with F7 (tmp), F8 (tmx), F9 (vap) and F4 (pet).

3. Results and Discussion

Table 3. Dataset Outlier Information

Feature Name	F2 (dtr)	F3 (frs)	F4 (pet)	F5 (pre)	F9 (vap)	F10 (wet)	F11 (elevation)	F12 (dominant_land_cover)
No. of outliers	170	4788	2503	1981	2166	3	844	5078

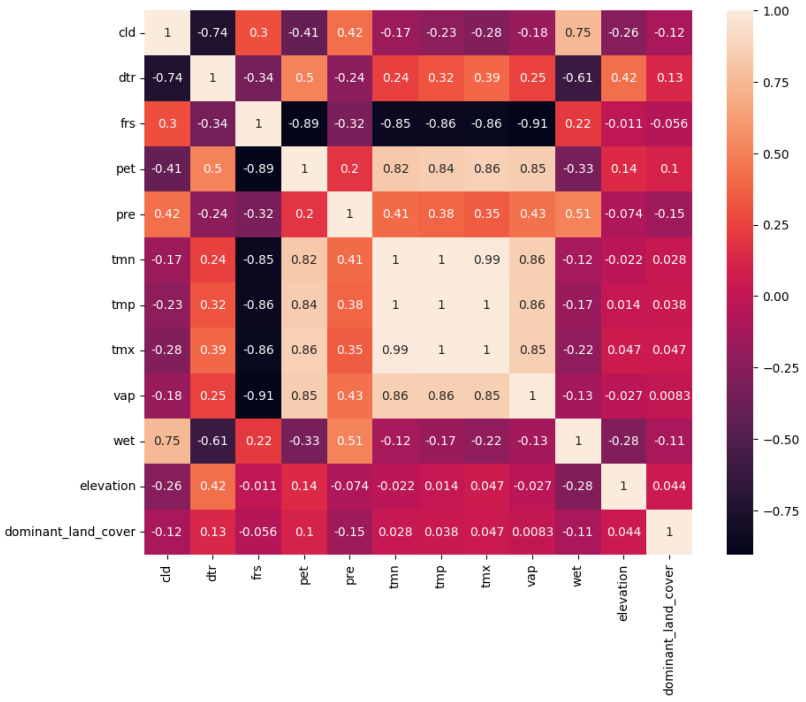

Fig. 3. Correlation heatmap of features

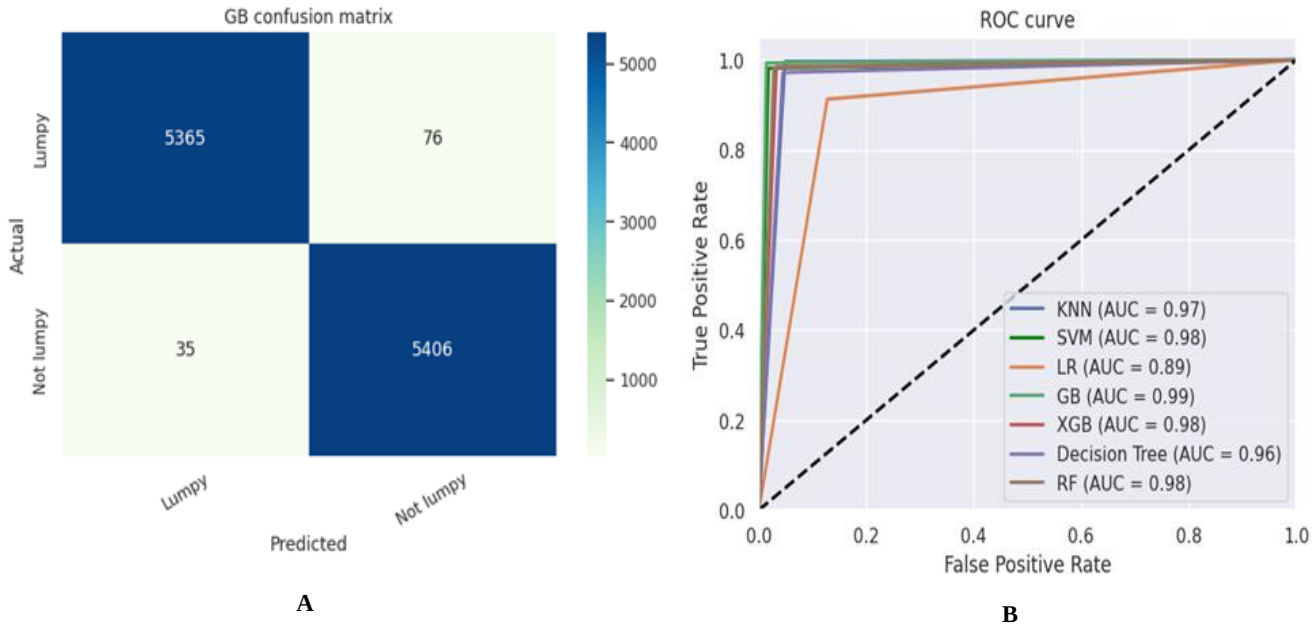

Fig. 4. Models Performance evaluation. (A) Confusion Matrix of GB model (B) ROC-AUC of various models experimented

Table 4. Performance metric values of ML models

Train-Test split Ratio	*Model*	*Train Accuracy*	*Test Accuracy*	*Precision*	*Recall*	*F1-score*	*ROC-AUC*	*MCC*
	SVM	99.326	98.189	0.98	0.98	0.98	0.98	0.97
	KNN	97.972	97.39	0.97	0.97	0.97	0.97	0.95
	RF	99.99	97.75	0.97	0.97	0.98	0.98	0.96
75-25	XGB	98.95	97.61	0.97	0.97	0.98	0.98	0.95
	GB	**98.93**	**98.97**	**0.99**	**0.99**	**0.99**	**0.99**	**0.98**
	DT	100	96.38	0.96	0.96	0.96	0.96	0.93
	LR	89.93	89.27	0.9	0.9	0.89	0.89	0.79
	SVM	99.254	98.22	0.98	0.98	0.98	0.98	0.96
	KNN	97.92	97.23	0.98	0.98	0.98	0.97	0.95
	RF	100	97.779	0.98	0.98	0.98	0.98	0.96
70-30	XGB	99.035	97.514	0.98	0.98	0.98	0.98	0.95
	GB	**98.96**	**98.88**	**0.98**	**0.98**	**0.99**	**0.99**	**0.98**
	DT	100	96.5	0.97	0.97	0.97	0.96	0.93
	LR	89.79	89.48	0.89	0.89	0.89	0.89	0.79

We have tested with two train-test split ratios of 75-25 and 70-30, utilizing different machine learning models. Table 4 above displays the values for the ML models' performance measures. According to the findings, the GB model outperformed the others, with test accuracy values of 98.97 and 98.88 in train-test split ratios of 75–25 and 70–30, respectively. It is worth noting that the GB model is also supported by the remaining performance measures, such as recall, accuracy, and ROC-AUC values. In comparison to other models, the MCC value for the GB model in both splits is likewise high (0.98). Figure 4(A) displays the confusion matrix of the GB model for the train-test split (75–25). The comparison of ROC-AUC curves of our ML models were shown in Figure 4(B). Table 4 shows the tuned

hyperparameters using grid search method for all our experimented ML models.

Table 4. Hyperparameters tuned in models

Model name	*Hyperparameters tuned*
GB	learning_rate = 0.1, n_estimators =100, max_depth = 9, subsample = 0.5
KNN	n_neighbors = 5, metric = 'euclidean', weights = 'uniform'
RF	random_state = RANDOM_SEED, n_estimators = 200, max_features = 'log2'
XGB	learning_rate = 0.01, n_estimators =100, max_depth =5, subsample = 0.5
LR	solver = 'lbfgs', max_iter =100
SVM	kernel = 'rbf', gamma = 8
DT	max_depth = 2, random_state = 42, min_samples_leaf = 50

3.1 Comparative Analysis

Here, we are comparing our results with the recent past literature results of LSD classification shown in Table 5. From the table it is evident that our model is outperforming to the other LSD classification models.

Table 5. Comparative analysis with earlier studies

Ref.	*ML model(s) used*	*Accuracy*	*ROC*
Olaniyan et al. [29], 2023	Stacked ensemble	97.69%	NA
Afshari et al. [30], 2022	Artificial neural networks (ANN)	97%	0.97
Dofadar et al. [11], 2022	RF and Light Gradient Boosted Machine	98%	NA
Our Study	**Gradient Boosting**	**98.97%**	**0.99**

The existing model [11] with RF and light GB machine classifiers gives accuracy of 98%. The other two studies [29][30] used stacked ensemble and ANN models, gives accuracy of 97.6% and 97% respectively.

4. Conclusion and Future Scope

In our work we have built an effective classification model for LSD classification based on meteorological and geospatial conditions. We have carefully pre-processed the data for classification by handling outliers, balancing the data, and selecting the relevant features. We have experimented several ML models by tuning hyperparameters using grid search method. Our ensemble model GB is performing well in classifying LSD when compared to other existing models in the literature. These methods can be used to enhance the ability to anticipate various illnesses and natural events. Consequently, future research can incorporate the deep learning method and ensemble approaches' predictive capabilities.

Author contributions

Venkata Pratyusha B: Conceptualization, Methodology, Software, Literature, Writing-Reviewing and Editing **Ramesh Chappa:** Data curation, Software, Validation., Field study **JagadeeswaraRao G:** Original draft preparation, Validation, Visualization, Investigation.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Afshari, S. & Ehsanallah. (2022). Assessing machine learning techniques in forecasting lumpy skin disease occurrence based on meteorological and geospatial features. Trop. Anim. Health Prod., 1(54), 12- 37, doi: 10.1007/s11250-022-03073-2.

[2] Anwar, A., Na-Lampang, K., Preyavichyapugdee, N. & Punyapornwithaya, V. (2022). Lumpy Skin Disease Outbreaks in Africa, Europe, and Asia (2005–2022): Multiple Change Point Analysis and Time Series Forecast. Viruses, 10(14), 12-32, doi: 10.3390/v14102203.

[3] Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Comput.Sci., 3(2), 1–21, 2021, doi: 10.1007/s42979-021-00592-x.

[4] Molla, W., de Jong, M .C., Gari, G. Frankena, K. (2017). Economic impact of lumpy skin disease and cost effectiveness of vaccination for the control of outbreaks in Ethiopia. Prev. Vet. Med., 7(14), 100–107, doi: 10.1016/j.prevetmed.2017.09.003.

[5] Reddy, P., Viswanath, P. & Reddy, E. B. (2018). Semi-supervised learning: a brief review. Int. J. Eng. Technol., 18(7), 1- 18, doi: 10.14419/ijet.v7i1.8.9977.

[6] Singh, V., Chen, S. S., Singhania, B. Nanavati, A. kumarkar, & Gupta, A. (2022). How are reinforcement learning and deep learning algorithms used for big data based decision making in financial industries–A

review and research agenda. Int. J. Inf. Manag. Data Insights, 2(2), 100 - 109, doi: 10.1016/j.jjimei.2022.100094.

[7] Saeed, S., Shaikh, A., M. Memon, M. A.& Naqvi, S. M. Impact of Data Mining Techniques to Analyze Health Care Data. J. Med. Imaging Heal. Informatics, 4(8), 682–690, doi: 10.1166/jmihi.2018.2385.

[8] Afshari, S. & Ehsanallah. (2022). Assessing machine learning techniques in forecasting lumpy skin disease occurrence based on meteorological and geospatial features. Trop. Anim. Health Prod., 1(54), 12- 37, doi: 10.1007/s11250-022-03073-2.

[9] Anwar, A., Na-Lampang, K., Preyavichyapugdee, N. & Punyapornwithaya, V. (2022). Lumpy Skin Disease Outbreaks in Africa, Europe, and Asia (2005–2022): Multiple Change Point Analysis and Time Series Forecast. Viruses, 10(14), 12-32, doi: 10.3390/v14102203.

[10] Ujjwal, N.,Singh, A., Jain, A. K. & Tiwari, R. G. (2022). Exploiting Machine Learning for Lumpy Skin Disease Occurrence Detection. 2022 10th Int. Conf. Reliab. Infocom Technol. Optim. (Trends Futur. Dir. (ICRITO), Noida, India, 1–6, doi: doi: 10.1109/ICRITO56286.2022.9964656.

[11] Dofadar, D. F., Abdullah, H.,. Khan, R. H., Rahman, R. & Ahmed, M. S. (2022). A Comparative Analysis of Lumpy Skin Disease Prediction Through Machine Learning Approaches. 2022 IEEE Int. Conf. Artif. Intell. Eng. Technol., doi: 10.1109/IICAIET55139.2022.9936742.

[12] Pal, O. K. (2021). Skin Disease Classification: A Comparative Analysis of K-Nearest Neighbors (KNN) and Random Forest Algorithm. Int. Conf. Electron. Commun. Inf. Technol. (ICECIT), Khulna, Bangladesh, 1-5, doi 10.1109/ICECIT54077.2021.9641120

[13] Suparyati, E. U. & Muhammad, A. H. (2022). Lumpy Skin Disease Prediction Based on Meteorological and Geospatial Features using Random Forest Algorithm with Hyperparameter Tuning.. 5th Int. Conf. Inf. Commun. Technol., p. Yogyakarta, Indonesia, 99–104, doi: 10.1109/ICOIACT55506.2022.9971807.

[14] Mohammed, S. S. & Al-Tuwaijari, J. M. (2021). Skin Disease Classification System Based on Machine Learning Technique: A Survey. IOP Conf. Ser. Mater. Sci. Eng., doi: 10.1088/1757-899x/1076/1/012045.

[15] Lake, B., Getahun, F. & Teshome, F. (2022). Application of Artificial Intelligence Algorithm in Image Processing for Cattle Disease Diagnosis. J. Intell. Learn. Syst. Appl., 71–88, doi: 10.4236/jilsa.2022.144006.

[16] Punyapornwithaya V. (2023). Forecasting of daily new lumpy skin disease cases in Thailand at different stages of the epidemic using fuzzy logic time series, NNAR, and ARIMA methods. Preventive Veterinary Medicine, https://doi.org/10.1016/j.prevetmed.2023.105964.

[17] Kshirsagar, P. R., Manoharan, H., Shitharth, S., Alshareef, A. M., Albishry, N. & Balachandran, P. K. (2022). Deep Learning Approaches for Prognosis of Automated Skin Disease. Life 2022, 12, 426, https://doi.org/10.3390/life12030426.

[18] A. Kumar, B. Kumar and H. S. Negi, "Predicting Lumpy Skin Disease using Various Machine Learning Models," *2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, Greater Noida, India, 2023, pp. 412-416, doi: 10.1109/CISES58720.2023.10183604.

[19] A. K. Jain, R. G. Tiwari, N. Ujjwal and A. Singh, "An Explainable Machine Learning Model For Lumpy Skin Disease Occurrence Detection," *2022 International Conference on Data Analytics for Business and Industry (ICDABI)*, Sakhir, Bahrain, 2022, pp. 503-508, doi: 10.1109/ICDABI56818.2022.10041665.

[20] *lumpy skin disease - Google Search*. (n.d.). accessed on 28 Mar, 2024 https://www.google.com/search?sca_esv=65d2b102cd6037ef&sxsrf=ACQVn09vsuDxqGyO37WWdieHxztnX4iFyg:1709540060046&q=lumpy+skin+disease&tbm=isch&source=lnms&sa=X&ved=2ahUKEwjmkIiAldqEAxV7y6ACHcXBBCAQ0pQJegQIEhAB&biw=1366&bih=625&dpr=1

[21] Afshari Safavi, Ehsanallah (2021), “Lumpy Skin disease dataset”, Mendeley Data, V1, doi: 10.17632/7pyhbzb2n9.1

[22] Yasi, Dani., Agus, Gunawan., Sapto, Wahyu, Indratno. (2022). Detecting Online Outlier for Data Streams using Recursive Residual. doi: 10.1109/ICIC56845.2022.10006916

[23] Magar, V., Ruikar, D., Bhoite, S., & Mente, R. (2024). Innovative Inter Quartile Range-based Outlier Detection and Removal Technique for Teaching Staff Performance Feedback Analysis. *Journal of Engineering Education Transformations*, *37*(3).

[24] Kavitha, K., Chittineni, S., Andhavarpu, N. K., & Chappa, V. (2022, January). Textual Sentimental Classification using Convolution Neural Network Algorithm. In *2022 International Conference on Computing, Communication and Power Technology (IC3P)* (pp. 246-249). IEEE.

[25] Vital, T. P. (2023). Intellectual Gestational Diabetes Diagnosis System Using MLP-Whale Optimization Algorithm including Statistical Analysis. *International Journal of Computing and Digital Systems*, *14*(1), 1-1.

[26] JagadeeswaraRao, G., Sivaprasad, A. An integrated ensemble learning technique for gene expression classification and biomarker identification from RNA-seq data for pancreatic cancer prognosis. *Int. j. inf. tecnol*. **16**, 1505–1516 (2024). https://doi.org/10.1007/s41870-023-01688-8

[27] Rao GJ, Prasad AS. Classification of gene expression from RNA-seq data for pancreatic cancer prognosis using ensemble learning. J App Biol Biotech. 2024;12(3):45-53. DOI: 10.7324/JABB.2024.171755

[28] Matthews, B. W. (1975). "Comparison of the predicted and observed secondary structure of T4 phage lysozyme". Biochimica et Biophysica Acta (BBA) - Protein Structure. 405 (2): 442–451. doi:10.1016/0005-2795(75)90109-9. PMID 1180967.

[29] Olaniyan, O. M., Adetunji, O. J., & Fasanya, A. M. (2023). Development of a model for the prediction of lumpy skin diseases using machine learning techniques. *ABUAD Journal of Engineering Research and Development*, *6*(2), 100-112.

[30] Afshari, S. & Ehsanallah. (2022). Assessing machine learning techniques in forecasting lumpy skin disease occurrence based on meteorological and geospatial features. Trop. Anim. Health Prod., 1(54), 12- 37, doi: 10.1007/s11250-022-03073-2.