EVALUATING MACHINE LEARNING METHODS VOTING SYSTEM FOR PREDICTING THE OCCURRENCE OF LUMPY SKIN CONDTION

Amandeep Kaur, Student, **Karandeep Singh**, Assistant Professor Department of Computer Science and Engineering, Punjabi University, Patiala

Abstarct

Lumpy skin disease is a viral disease that primarily affects cattle. It is caused by the lumpy skin disease, which belongs to the Capripoxvirus genus. LSD is characterized by the formation of nodules or lumps on the skin of infected cattle. This study aimed to evaluate how well some machine learning algorithms could predict the likelihood of LSDV infection based on geographical and meteorological characteristics. Geospatial and climatic factors are crucial to the disease's epidemiology because of their close connection to the persistence of arthropod vectors. The results of this study suggest that combining geographical and meteorological factors, voting system of different machine learning output utilized to forecast the occurrence of LSDV infection with high precision. In locations where there is a high risk of LSDV infection, applying the predicting capability of these methodologies could be very helpful in implementing screening and awareness programs as well as preventative efforts like vaccination.

Keywords: Lumpy diseases, Machine learning models, Features extraction, Feature selection.

1. Introduction

Infection with the LSDV, which causes acute or subacute sickness in the population of cattle and water buffalo, poses a serious threat to the cattle industry. All types of cattle can contract the disease, although calves and cows towards the height of their milk production are most vulnerable [1].The lumpy skin disease in cattle is contagious, eruptive, and occasionally lethal. It is identified by lumps on the skin and other body parts. Secondary bacterial infection typically makes things worse in this case.LSD was first noted in Zambia in 1929 [2]. Over the next 85 years, it progressively spread over Africa and the Middle East. Only the South and East of Africa were affected by lumpy skin disease, but in the 1970s it moved from the northwest of the continent into sub-Saharan West Africa. It has spread to various Middle Eastern countries since 2000. It also made it to Turkey and several Balkan nations in 2013 [3]. Recently, cases of lumpy skin disease have been found in Bangladesh, the Republic of China, Georgia, and Russia. The increasing global spread of lumpy skin disease has caused concern internationally. The Western Hemisphere, Australia, and New Zealand have not recorded any cases of the disease. Lumpy skin disorders may manifest regularly or unexpectedly.

1.1 Characteristics of Lumpy Disease

Some key characteristics of lumpy skin disease in cattle:

a) Clinical significance: Fever is the normal first symptom of the illness, which is then followed by the development of skin lumps or nodules. These nodules may be solid or soft to the touch, and their sizes might vary.
b) General symptoms: Other symptoms of infected cattle include decreased appetite, weight loss, decreased milk output, weakness, and resistance to movement. In severe cases, pneumonia, internal organ inflammation, and even death, as well as more severe symptoms and complications, are possible outcomes [4].
c) Transmission: The main ways that LSD is spread are by insects, especially biting flies like mosquitoes and ticks. These insects can pick up the virus when they feed on diseased animals, then spread it to healthy cattle. The disease can also spread through direct contact with diseased animals or by contact with contaminated objects, such as clothing or equipment [4].
d) Geographic distribution: Primarily in sub-Saharan areas of Africa, lumpy skin condition was prevalent. The Middle East, Europe, and Asia have all seen an increase in the disease's distribution

in recent years. Concerns about its geographic spread have prompted heightened surveillance and control efforts in the affected areas [5]

e) Economic impact: LSD may have a substantial negative impact on cattle ranchers' bottom lines. Animals with this condition produce less milk and grow less weight, which costs money. Furthermore, the profitability of the cattle business may be impacted by the presence of skin lesions due to lower hide value [6].

1.2 Literature survey

Performing a complete literature survey on lumpy skin disease detection in cattle can be an extensive task, as research in this field continues to evolve.

To forecast the incidence of disease in unobserved (test) data, EhsanallahAfshariSafavi [7] evaluated the capacity of a number of machine learning algorithms to predict the occurrence of LSDV infection based on climatic and geological factors, meteorological, animal population density, dominant land cover, and topography parameters. Several machine learning methods demonstrated up to 97% accuracy in predicting the existence of LSDV in test data. The findings of this study imply that voting scheme for predicting LSDV foreseen with great accuracy.

Between 2014 and 2017, Golden et al. [8] collected soil and waste specimens from 11 grazing poultry enterprises in the USA. They developed random forest and gradient boosting machine prediction models to forecast the incidence of Listeria spp. in samples based on meteorological factors such as temperature, wind speed, storm speed, relative humidity, and moisture at the agricultural region. The AUC performance ratings for the random forest and gradient boosting machine models of faecal samples were 0.905 and 0.855, respectively.

Rony et al. [9] employed various CNN architectures, including traditional deep CNN, Inception-V3, and VGG-16 in deep learning, to proactively identify the most common external diseases. The complete process of using the algorithms for detecting illnesses, from data collection through process and outcome, is fully described in the study. The suggested method has been shown to be effective, producing results with a 95% accuracy rate, which may reduce errors in diagnosis and be helpful to doctors and livestock farmers in identifying diseases.

Machine learning techniques were utilised by Liang et al. [10] to predict African swine fever outbreaks globally utilizing bio-climatic data. When only relevant climatic features presented in the subset dataset, the support vector machine method showed the highest accuracy (76.02%), whereas the random forest algorithm outperformed other methods with 80.4% accuracy in the dataset encompassing all predictive factors.

The LSD dataset from Mendeley Data is used by Suparyati et al. [11] in their study, which aims to predict the presence of LSD in a certain area. Because lumpy infections are uncommon, data is skewed, which makes it challenging to train machine-learning systems. In sampling techniques (SMOTE), the imbalanced data are handled using the random under-sampling approach and the produced minority over-sampling strategy. The Random Forest classification algorithm was trained using sample data in order to predict when lumpy infections will emerge. When applied to data that has been both over- and under-sampled, the Random Forest classifier works well. According to the analysis of performance indicators, SMOTE outperforms Random Under-sampling by 1% to 2%. How to balance disparate data classes is the main subject of the author's research.

To forecast the weekly total of infectious human diarrhoea cases in Shanghai, China.A feed-forward back-propagation neural network model was created by Wang et al. [12] using meteorological information as prediction characteristics. When compared to nonlinear models like neural networks, support vector regression, and random forests regression, multiple linear regression performed poorly. When all performance evaluation factors were considered at once, neural networks produced the most satisfactory results.

The lumpy illness detection architectural disease is a suggestion made by Gaurav et al. [13]. The features required to build this framework were extracted using VGG-16, VGG-19andInception-v3.The research is then assessed utilizing our dataset and a variety of cutting-edge approaches, such as k-nearest neighbor, support vector machine, Nave Bayes, and artificial neural networks, which results in excellent feature extraction performance. The Inception-v3 model, which employs an artificial neural network and has a greater precision of 92.5% on the testing dataset, is the best model, according to the authors.

In order to estimate the incidence of LSDV infection in countries with a prior history of disease outbreak reported between 2011 and 2021, research was conducted to develop predictive models employing some strong machine learning algorithms based on meteorological and geospatial features. This was done because insects are important in LSDV transmission, and they depend on climatic and geographical features.

1.3 Contribution of research

The contribution of proposed research is voting system used for prediction of LSDV. The benefits are:

1. Greater accuracy: A voting mechanism frequently increases the total prediction's accuracy by integrating the forecasts of several different models. This is because each model is probably better at forecasting some events than others, and the voting system can benefit from the strengths of each model by combining the forecasts.
2. Reduced biased: Voting procedures can also aid in reducing bias in machine learning models. This is because each model is trained on a different dataset, and the voting mechanism can assist to average out any biases that may exist in each individual model by pooling the predictions.
3. Increased robustness: Machine learning models can become more resistantto changes in the data by using voting systems. This is because each model will probably be impacted by data changes in a unique way, and the voting method can help to lessen the impact of any changes by pooling the forecasts.

2. Methodology

Figure 1. shows a summary of the procedures conducted in the materials and techniques, while the following sections go into more depth about each action.

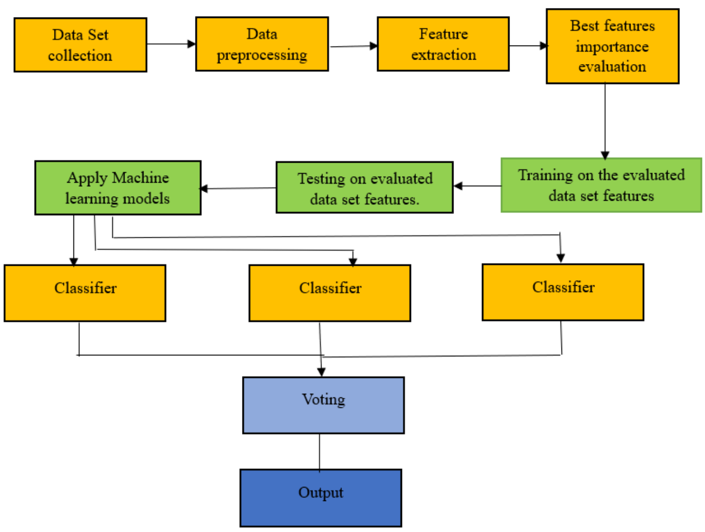

Figure 1: Block diagram of the proposed methodology

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, No. 3, June : 2023

2.1 Data set availability

Data on the outbreak of lumpy skin disease The Global Animal Disease Information System of the FAO was used to obtain the geographic locations of outbreaks of Lumpy Skin Disease the exact moment when the outbreak started, as well as the latitude and longitude of the epidemic spot, for the period between January 2011 and March 2021 obtained [14]

2.2. Feature extraction

Three types of data are extracted from the dataset named as meteorological data, animal density data, land cover data, elevation data as shown in figure 2.

a) Meteorological data: This data used to extract features such as temperature, precipitation, humidity, wind speed, and wind direction. These features can be used to predict the behaviour of animals, the growth of vegetation, and the spread of diseases. This information was taken from the Climatic Research Unit (CRU TS4.04) of the University of East Anglia. [15]

b) Animal density data: This data used to extract features such as the number of animals per unit area, the type of animals, and the age and sex of the animals. These features can be used to predict the impact of animals on the environment, such as the spread of diseases and the destruction of vegetation. These features are gathered from Gridded Livestock of the World (GLW 3) database [16]

c) Land cover data: This data can be used to extract features such as the type of land cover, the amount of vegetation, and the slope of the land. These features can be used to predict the potential for flooding, erosion, and landslides.

d) Elevation data: This data can be used to extract features such as the elevation of the land, the slope of the land, and the presence of water bodies. These features can be used to predict the potential for flooding, erosion, and landslides.

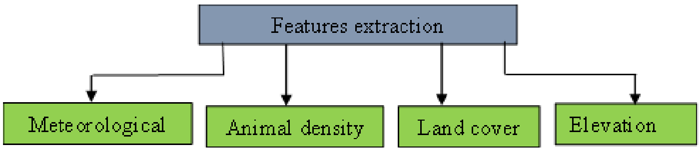

Figure 2: Feature extraction

2.3 Feature selection

A statistical measure called correlation shows how closely two variables are related to one another. Perfect positive correlation is indicated by a correlation coefficient of 1, perfect negative correlation by a correlation coefficient of 1, and no correlation by a correlation coefficient of 0. Machine learning features can be chosen using correlation [17]. The goal is to choose features that have a strong correlation with the target variable while being uncorrelated with one another. Table 1 shows the correlation among features.

Table 1: Feature importance

Features	Correlation
x	-0.408407
y	-0.146576
region	-0.920090
country	-0.826448
reportingDate	-0.833443
cld	0.237754
dtr	-0.216253
frs	-0.172834
pet	0.061748
pre	0.419686
tmn	0.308560
tmp	0.283436
tmx	0.258099
vap	0.169412
wet	0.099620
elevation	-0.112408
dominant_land_cover	-0.163045
X5_Ct_2010_Da	0.068884
X5_Bf_2010_Da	-0.038905

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, No. 3, June : 2023

2.4. Algorithms for machine learning are utilised in the testing and training phases.

a) Support Vector Machine:

Support vector machines (SVMs) are learning algorithms that analyse data for classification and regression analysis. They are supervised learning models. An SVM training algorithm creates a model that categorises fresh data points into one of the classes given a set of labelled training data. SVMs are among the most widely used machine learning algorithms for regression and classification. They are utilised to address a wide range of issues and are renowned for their precision and robustness. Finding a hyperplane that best separates the two classes of data is the fundamental tenet of SVMs. A plane or line known as a hyperplane divides the data space into two parts. Finding a hyperplane with the biggest margin the separation between the hyperplane and the nearest data points on either sideis the objective [18].

The SVM method uses the optimisation technique to locate the hyperplane that maximises the margin. A mathematical method called optimisation can be used to identify the ideal answer to a problem. Finding the hyperplane that maximises the margin is the objective in the case of SVMs.

b) Linear regression

A continuous value can be predicted using the supervised machine learning technique of linear regression using a collection of independent factors. However, by employing a threshold value to change the predicted continuous value into a discrete value, linear regression can also be applied to classification issues [19]

c) AdaBoost

The AdaBoost algorithm works by iteratively training a series of weak learners, each of which is trained on a weighted version of the training data. The weights are assigned such that instances that are misclassified by previous learners are given more weight. The output of each weak learner is then combined into a weighted sum, with the weights of each learner being adjusted based on its performance [20].

The final output of the AdaBoost classifier is the sign of the weighted sum.AdaBoost has been shown to be effective in a variety of classification problems, including spam filtering, image classification, and natural language processing. It is a relatively simple algorithm to implement, and it is available in many machine learning libraries. Table 2. showing the tuning parameters of above-mentioned classifiers.

Table 2: showing the tuning parameters of above-mentioned classifiers.

2.5 Evaluation parameters

Model	Method	Tuning Parameter(s)
AdaBoost	AdaBoost	n_estimators=700, algorithm='SAMME.R', learning_rate=0.1
SVM	Ksvm	Kernel Radial Basis
Linear Regression	regression	No. of observation=150 , DiscrimType: ‘linear’

There are many evaluation parameters that can be used to evaluate the performance of a classifier [21]. Some of the most common parameters include:

- Accuracy: This is the percentage of instances that are correctly classified.
- Precision: This represents the proportion of positive situations that actually are positive.
- Recall: This is the percentage of instances that are actually positive that are classified as positive.
- F1-score: This is a weighted average of precision and recall.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, No. 3, June : 2023

- Area under the ROC curve (AUC): This is a measure of the classifier's ability to distinguish between positive and negative instances.

2.6 Result analysis

The results of this study showed that the incidence of LSDV infection projected in test samples with high accuracy by utilising machine learning algorithms and employing different climatic and geospatial factors as predictive variables. Figure 3-5 represents the graphical analysis of classifiers performance.

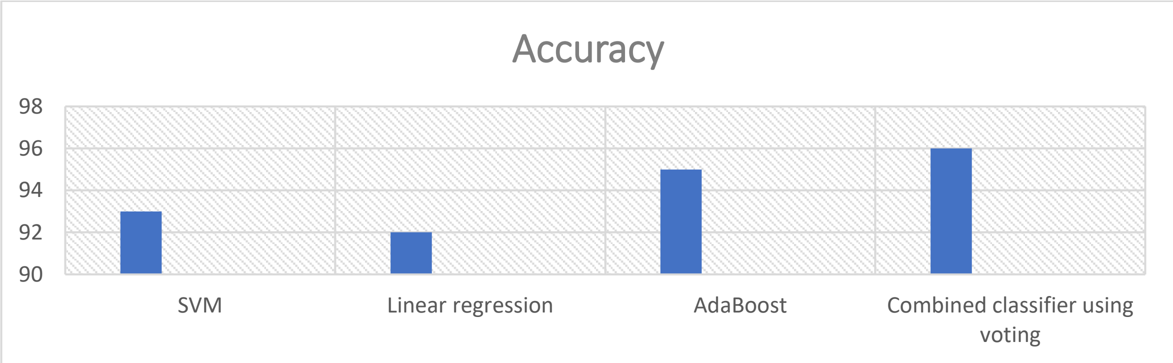

Figure 3: Accuracy predictive analysis

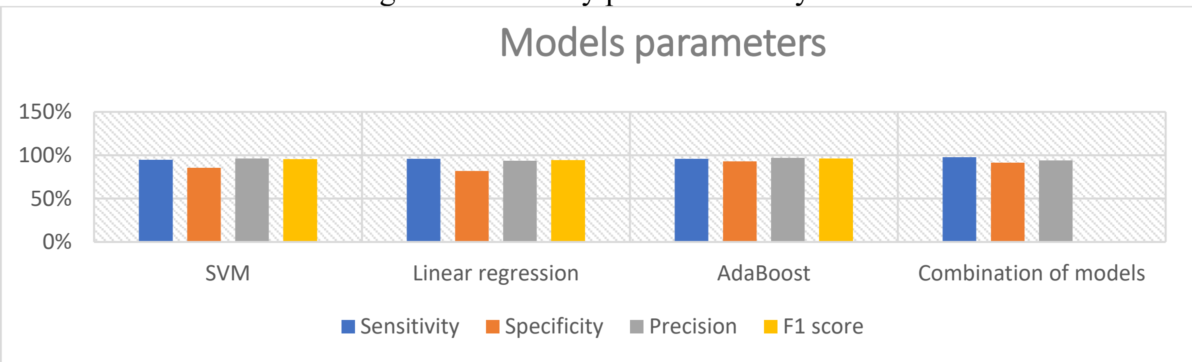

Figure 4: Model evaluation parameters

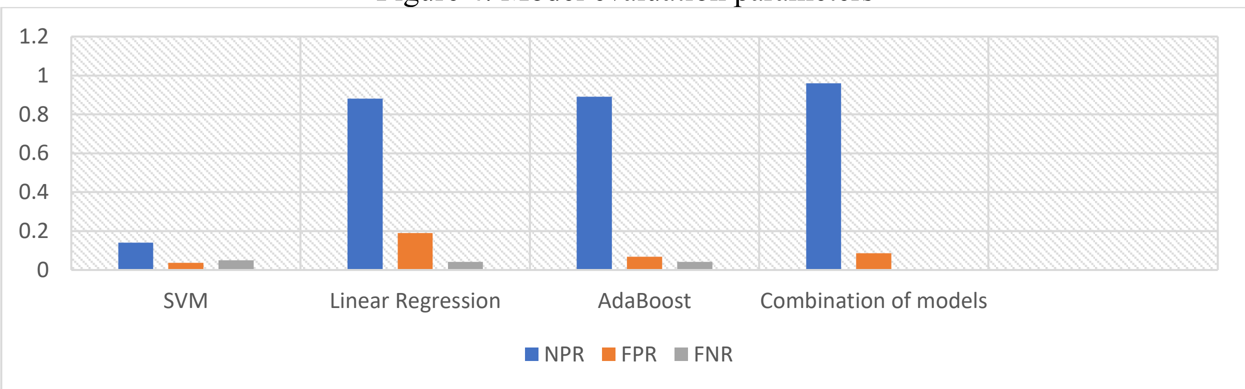

Figure 5: Positive and negative rate prediction

3. Discussion and conclusion

However, it is important to note that the majority of the passive accounts from veterinary facilities across several nations made up the LSDV outbreak data used in the current study. When examining the results, several limitations of employing passive monitoring data should be taken into consideration. Reporting is hampered in some nations by the compensation plans' presence or quality, the efficiency and accessibility of veterinary services, the isolation of some locations, and farmer visibility .However, the absence of LSDV reports in several regions of the investigated nations may be explained by the absence of favourable environmental factors for the spread of the illness there. The

current study has some limitations, including the little amount of data collected, the few predictor variables included, and the possibility that the disease has spread to other regions of the studied countries with differing meteorological and geographic conditions since completing this research.
In conclusion, although the voting system produces better results, various machine learning algorithms, such as ANN, Decision tree, Random Forest, etc., can accurately forecast the occurrence of LSDV infection based on a few geographical and climatic variables. Using this technique could be highly beneficial to establish monitoring and awareness programmes as well as preventive measures like vaccine campaigns in areas where LSDV infection is a high risk.

References

1. 1.Namazi, F., &KhodakaramTafti, A. (2021). Lumpy skin disease, an emerging transboundary viral disease: A review. Veterinary Medicine and Science, 7(3), 888-896.
2. Roche, X., Rozstalnyy, A., TagoPacheco, D., Pittiglio, C., Kamata, A., Beltran Alcrudo, D., ... & Sumption, K. (2021). Introduction and spread of lumpy skin disease in South, East and Southeast Asia: Qualitative risk assessment and management. Food & Agriculture Org.
3. Chihota, C., Rennie, L., Kitching, R., & Mellor, P. (2003). Attempted mechanical transmission of lumpy skin disease virus by biting insects. Medical and Veterinary Entomology, 17(3), 294-300.
4. KA.Al-Salihi,IQ.Hassan,“LumpyskindiseaseinIraq:studyofthediseaseemergence” Transboundaryand EmergingDiseases62,457–462,2015.
5. A. Sprygin, Ya Pestova, D. B. Wallace, E. Tuppurainen, A. V. Kononov. "Transmission oflumpyskin diseasevirus:A shortreview."Virus research269 (2019):197637.
6. Cornelius Henry Annandale, Dietmar Erik Holm, Karen Ebersohn, and Estelle HildegardVenter. "Seminal transmission of lumpy skin disease virus in heifers." Transboundary andemerging diseases 61, no. 5 (2014):443-448.
7. EhsanallahAfshariSafavi,."Assessingmachine learningtechniquesinforecastinglumpyskindiseaseoccurrencebasedonmeteorologicalandgeospatial features." TropicalAnimalHealthand Production54, no. 1 (2022):1-11.
8. Golden, C. E., Rothrock Jr, M. J., & Mishra, A. (2019). Comparison between random forest and gradient boosting machine methods for predicting Listeria spp. prevalence in the environment of pastured poultry farms. Food research international, 122, 47-55
9. RonyMd,DolaBarai,ZahidHasan."CattleExternalDiseaseClassificationUsingDeepLearningTechniques."In202112thInternationalConferenceonComputingCommunicationandNetworking Technologies(ICCCNT), pp. 1-7. IEEE,2021. 23
10. Liang, Y., Li, J., Gu, J., Zhou, Z., & Wang, Z. (2015). Artifcial neural networks for infectious diarrhea prediction using meteorological factors in Shanghai (China). Applied Soft Computing, 35, 280-290.
11. Suparyati,EmmaUtami,AlvaHendiMuhammad,”ApplyingDifferentResampling Strategies In Random Forest Algorithm To Predict Lumpy Skin Disease” In Random ForestAlgorithm To Predict Lumpy Skin Disease. Jurnal RESTI (RekayasaSistem dan TeknologiInformasi), 6(4)(2022), 555-562.
12. Gaurav Rai, Naveen, Aquib Hussain, Amit Kumar and Rahul Nijhawan, “A Deep LearningApproachtoDetectLumpySkinDiseaseinCows”,InComputerNetworks,BigDataandIoT (pp. 369-377)(2021).Springer, Singapore.
13. Wang, Y., Li, J., Gu, J., Zhou, Z., & Wang, Z. (2015). Artifcial neural networks for infectious diarrhea prediction using meteorological factors in Shanghai (China). Applied Soft Computing, 35, 280-290.
14. Gilbert, M., Nicolas, G., Cinardi, G., Van Boeckel, T. P., Vanwambeke, S. O., Wint, G. W., & Robinson, T. P. (2018). Global distribution data for cattle, bufaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. Scientifc data, 5(1), 1-11

15. Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. Scientifc data, 7(1), 1-18.
16. Latham, J., Cumani, R., Rosati, I., &Bloise, M. (2014). Global land cover share (GLC-SHARE) database beta-release version 1.0– 2014. FAO: Rome, Italy.
17. Senan, E. M., Abunadi, I., Jadhav, M. E., &Fati, S. M. (2021). Score and Correlation Coefficient-Based Feature Selection for Predicting Heart Failure Diagnosis by Using Machine Learning Algorithms. Computational and Mathematical Methods in Medicine, 2021.
18. Scholkopf, B. (1998). Support vector machines: a practical consequence of learning theory. IEEE Intelligent systems, 13
19. Cox, D. R. (1958). The regression analysis of binary sequences. Journal of the Royal Statistical Society: Series B (Methodological), 20(2), 215-232.
20. Freund, Y., &Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences, 55(1), 119-139
21. Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems: O'Reilly Media.