Vital Annex: International Journal of Novel Research in Advanced Sciences
ISSN: 2751-756X
Volume 04 Number 07 (2025)
https://innosci.org/IJNRAS

Article

Lumpy Skin Disease Classification on Imbalanced Data: A Comprehensive Approach Using Machine and Ensemble Learning

Shahla U. Umar

1. Department of Software, College of Computer Science and Information Technology, University of Kirkuk, Kirkuk, Iraq

* Correspondence: Shahla_umar@uokirkuk.edu.iq

Abstract: Lumpy Skin Disease (LSD) is a major problem of veterinary care that requires proper, prompt diagnosis and classification to enable effective treatment. However, medical datasets are often affected by such challenges as class imbalance, missing values, outliers, and high dimensionality, which makes it challenging to create effective diagnostic tools. This study presents a hybrid machine learning (ML) model which aims to address these typical data complexities. We integrate the Synthetic Minority Over-Sampling Technique (SMOTE) as an effective method of dealing with data imbalance with ensemble learning, Bagging in particular, to improve the precision of the classification. In our research, three well-known ML algorithms Decision Trees (DT), Logistic Regression (LR) and Naive Bayes (NB) are rigorously tested together with our Bagging model. The evaluation of the performance was performed on a wide range of measures: accuracy, precision, recall, F-score, and Matthews Correlation Coefficient (MCC). The findings are conclusive in the sense that the Bagging ensemble has been shown to be better than the rest of the models in the sense that it had the highest accuracy of 89.5% and the highest precision and recall rates as well. Moreover, SMOTE was very useful in reducing the bias in the dataset thus making model training more accurate. These results highlight the significant performance of SMOTE and ensemble learning combination towards producing very reliable and accurate diagnostic tools. Such an original method can bring a major breakthrough in LSD treatment and has an enormous potential to enhance the diagnosis of other multifactorial disorders in veterinary medicine.

Keywords: Machine learning, Imbalanced data, Classification, SMOTE, Lumpy Skin disease

Citation: Umar, S. U. Lumpy Skin Disease Classification on Imbalanced Data: A Comprehensive Approach Using Machine and Ensemble Learning. Vital Annex: International Journal of Novel Research in Advanced Sciences 2025, 4(7), 263-270

Received: 10th May 2025
Revised: 16th Jun 2025
Accepted: 24th Jul 2025
Published: 31th Jul 2025

1. Introduction

Lumpy Skin Disease (LSD) classification in the healthcare sector is one of the most crucial issues. Therefore, Machine learning (ML) algorithms incorporating data mining techniques that help to strengthen the healthcare systems are widely used today. Choosing relevant features and using ensembles that involve several ML classifiers, which is a method that captures complex data structures, proves to make up the distribution of data and the modeling accuracy high. It ensures bias-free decisions in the landslide classification. This approach takes precision diagnostics a notch higher but also plants confidence that expert doctors use to act upon.

It becomes more challenging to diagnose LSD because of class imbalance in biomedical datasets, the presence of missing values, outliers, and high dimensionality, which affect artificial classification systems. The consequences ascribed to the problem were resolved by deploying the Synthetic Minority Over-sampling Technique (SMOTE) used on the medical datasets [1]. Nevertheless, some outdated metrics, such as classification accuracy, fail to measure the effectiveness while the dataset is class imbalanced. These illustrations demonstrate that it takes place without new methods of diseases, such as LSD.

A robust analysis of LSD is feasible but may be biased through imbalance of class, the presence of missing values, outliers, and high dimensionality medical datasets, which may affect the performance of various classifiers. However, medical datasets and Common Data Curation issues were solved using the SMOTE algorithm to sample the minority-over-sampled data sets [2]. Still, in face datasets that show unequal distribution of classes, accuracy metrics alone cannot assess the model's performance. Therefore, the techniques to adjust the diagnostic for LSD and other illnesses are endorsed.

SMOTE is necessary for bid data imbalance to create synthetic minority class instances. It is observed that when the research is applied in one class, performing better; the delay is in the classification performance, as in. Apart from its purpose of reducing bias in datasets and increasing accuracy, SMOTE also plays a role in catering to the minority class by diminishing their observations, hence increasing the performance of ML algorithms. SMOTE is both a classifier-independent approach and is easy to apply. Therefore, it can be used in data partitioning.

When we look at combination learning, which is an example of the Bagging Ensemble approach, the said has been perceived to improve classification for LSD cases with imbalanced data. Bagging, which combines the prediction algorithms of separate models by attempting to over-correcters, creates distorted datasets. Moreover, other ensemble techniques were employed to add a new classification layer, substantially increasing the accuracy rate. Therefore, These strategies are used to surmount the data imbalance issue, leading to sensitizing detection and prediction. Implementing the Bagging algorithm in an ML model has significantly increased the accuracy and sensitivity of LSD classification, thus becoming valuable for boosting the model performance.

This study focuses on the Lumpy Skin Disease classification accuracy via DT, LR, and NB performance. The first goal is to test if the algorithms narrow specific lumpy skin diseases. The SMOTE approach applies, and its influence on classification is established. Besides, ensemble learning is employed to help ML models develop higher accuracy and robustness with an array of the latest models by applying the Bagging method. This broad-based analysis will explore the use of traditional ML algorithms in building appropriate and accurate LSD analysis devices for dermatology.

The study direction will be as follows: section 2 will go over reviewed for related research work. In section 3, the procedure and possible measures for lesson delivery will be presented. Section 4 will be twofold; results and discussions will be presented here. The next section focuses on discussing an outcomes section and suggestions for future works.

1.1 Related Works

Compared to the traditional LSD technique, classification algorithms manipulate one's ability to classify the features of skin conditions. ML algorithms can detect patterns in skin images or, perhaps, inpatient data to ensure the disease was indeed identified [3]. ML techniques and ensemble methods have been the most efficient and reliable in medicine for disease diagnosis . Neural networks, and ensemble learning have been employed in previous research works for diabetes diagnosis [4]. One of the best ways to point out the most crucial problems that affect the effectiveness of ML algorithms for LSD classification is feature selection, data balancing, and hyperparameter optimization [5], [6]. On the other hand, healthcare providers can be empowered by using ML algorithms and other innovative diagnostic tools to act swiftly and dependably in finding quick solutions.

However, Zheng et al. [7] proposed an unequal probability sampling method to classify imbalanced data. The algorithms were performed on 14 datasets (classification test) with standard classification methods. The best results in the AUC were given by various proposed sampling algorithms, which proved to be dominant relative to random sampling and were also superior to the AUC metric. Apart from that, Zhang et al. constructed a system that used the SMOTE-RFE-XGBoost method to diagnose spinal disorders. For instance, SMOTE, RFE, XGBoost, and tree-based feature selection methods were used. Communications were around this point. Fitriyani [8] suggested a model for chronic diseases based on the DBSCAN algorithm, with SMOTE-ENN and Random Forest being added, and these methods were compared with traditional approaches such as the LR, DT, MLP, and SVM, which led to better results for the heart disease and diabetes type 2 prediction. LR was frequently adopted as a linear classification predictive model, but classifiers were found for DT to predict targets. The results suggest the central role of DT, LR, and NB in being effective in one-class and multi-class classification tasks.

The procedures for the imbalance in data as the foundation behind disease identification, including LSDS, have been found in Rahmi et al. [9]. Methods like SMOTE, Random Oversampling, undersampling, and ensemble learning have eliminated group bias by facilitating increased classification performance in medical data. This imbalance in LSD datasets among the classes provides such an issue that the minority class can be overseen. In this case, classification algorithms tend to prefer the majority class, which may have an opposite result for accuracy [10]. Liu et al. highlighted a significant issue with class imbalance as they provided a case where the model biased and made it difficult to make valid inferences is cited. Ensemble methods focusing on feature selection, undersampling, and cost-sensitive learning techniques have demonstrated promising outcomes in various healthcare studies to overcome class imbalance and provide a better detection process [11]. To inducement a unique understanding, the research relies on the customized solution of unbalanced data for the multi-class identification of LSD.

An enhancement of the collection of papers has proven that SMOTE and its widely used variants are the most frequent algorithms of data balancing in the classification of LSD. SMOTE is an oversampling technique to artificially increase the size of the dataset by generating synthetic data. This resolves class imbalance. For the next stage, SMOTE algorithm was implemented in conjunction with the ensemble method SVM of Yang and Guan to help classify imbalanced datasets of cerebrovascular disorders. As a result, special accuracy and classification performance were achieved. Not only does the SMOTENN technique that was applied to handle imbalanced data in heart disease prediction seem to improve model accuracy and performance, but they are also known to [12]. The impact of SMOTE, as well as its modifications for the specific task of data balancing for ML-based LSD diagnosis and other medical concerns, has been proven to result in high prediction accuracy.

Through the utilization of the Bagging ensemble learning techniques, the prior LSD domains' classification precision is improved. A study revealed that bagging, boosting, and stacking ensembles are the most constant algorithms for imbalanced data identification problems in classifying disease tasks [13]. An ensemble model of SVM-SMOTE with integrated learning is proposed as a class imbalance-removing method for the cerebrovascular disease dataset, thus allowing disease recognition and prediction of future trends to be achieved with higher accuracy [14]. These ensemble models, which employ various techniques, e.g., Bagging, data level, and algorithm level, can surpass a single ordinary classifier, thus leading to better detection performance . These works look personalized and mostly focus on students' learning styles. Their main goal is to use Bagging technique to improve LSD accuracy.

One of the major issues seen in the light from the in-depth comparison of ML algorithms used in LSD classification relates to the issue of class imbalance and feature selection. The researchers in the study have used classifiers RF, SVM, DT, LR, NB for accurate LSD classification . The list of performance metrics involves accuracy, sensitivity, specificity, precision, and F1-score as the measure of the model's capability in handling an imbalanced dataset [15]. The ensemble method of combining various classifiers was specifically proper, improved classification performance, and addressed the overfitting problem . The researches pay attention to feature selection, hyperparameter tuning, and data balancing techniques for ML models to be more accurate and reliable for LSD categorization.

While the diversity of algorithms used for the classification of aluminum-ingot disease is still present in ensemble studies despite this, the point of interest is that many of the studies miss the inclusion of particularly DT, LR, and NB models in addition to the application of SMOTE and bagging ensemble methods in LSD. The essence of the study is in choosing and adapting algorithms and techniques; however, there is no representation of the effect that combining the algorithms can have in improving the classification precision in the analysis of LSD [16], [17].

2. Materials and Methods

This section will discuss our methodology and proposed work in detail. The proposed work in this study is illustrated in Figure 1.

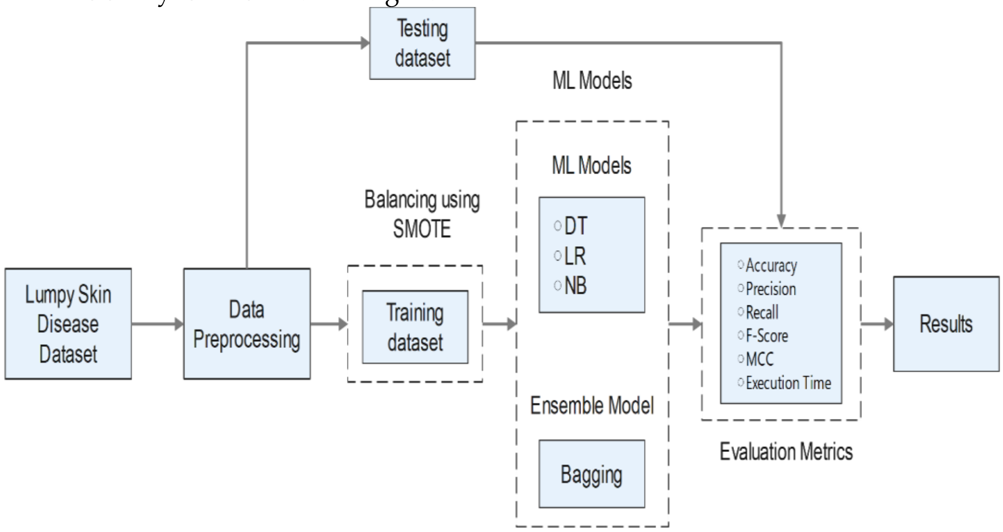


Figure 1. The proposed methodology

2.1. Dataset and Data Preprocessing

This study used a LSD publicly available at Kaggle, presumably containing relevant features connected with the disease [18]. The dataset contains 24803 samples and 20 columns. The target is binarily represented by 1 and 0 where 1 represents that person has lumpy LSD and 0 means he has not been affected. Next, we performed preprocessing on the dataset, including finding missing values and normalizing features, to enhance computational efficiency and model performance.

2.2 Model Training

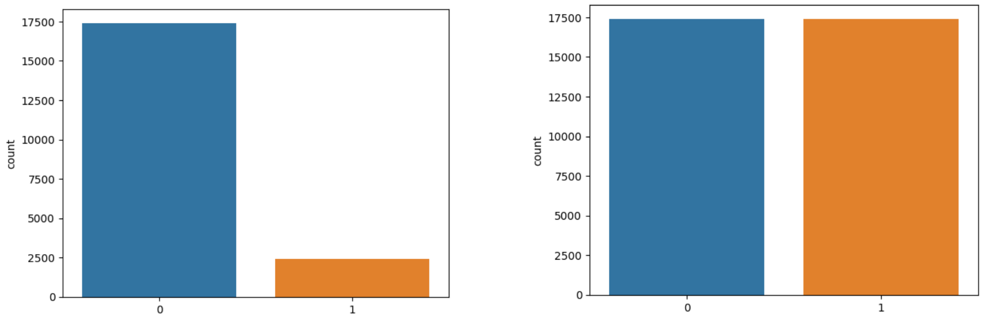
After the preprocessing step, the dataset was divided into two different sets: a training set and a testing set. A 80:20 split ratio was used, and this is a common approach to ML where a huge majority of the data is used to train the model and only a smaller, unseen part is used to evaluate. This created 19,842 training samples and 4,961 testing samples.

After looking at the training subset, a large class imbalance was encountered. The "zero" class (not affected by LSD) had 17,410 samples and the "one" class (affected by LSD) had only 2,432 samples. Such an imbalance may lead to the introduction of bias in the model and the model would tend to favor the majority class and perform in a poor way in the minority class which can be the class of interest in medical diagnosis.

In order to solve this important problem, the Synthetic Minority Over-sampling Technique (SMOTE) was used on the training data. SMOTE is a more advanced upsampling algorithm, which, rather than simply copying the existing instances of the minority classes, creates new, synthetic ones along the line segments between the existing minority class neighbors. This method allows balancing the distribution of classes without information loss. Following SMOTE, a balanced training set was achieved, with 17,410 samples per class, after having the minority class (one) over-sampled to match the zero class. This balancing procedure was then proven to have worked visually in figure 2, which confirms the state of the target variable before and after. The testing dataset was kept as it was, i.e. imbalanced deliberately, so that the model performance can be analyzed on a data that can be considered close to the real world data. "The number of training samples before and after applying SMOTE is summarized in Table 1."

Table 1. Number of samples in the training part regarding the target after splitting the dataset before and using SMOTE

Classes	Before using SMOTE	After using SMOTE
	Training	
0	17,410	17410
1	2,432	17410



Before using SMOTE After using SMOTE

Figure 2: The target class before and after applying SMOTE to balance the dataset

2.3 ML Models

In this study, a variety of individual and ensemble ML models were selected to carry out the classification of Lumpy Skin Disease. In this way, a baseline performance can be measured with well known individual classifiers, and then performance improvement can be demonstrated with a more sophisticated ensemble method. The three baselines algorithms tested include Decision Tree (DT), Logistic Regression (LR), and Naive Bayes (NB). DT is a non-parametric supervised learning algorithm that produces a model to forecast a target variable learning simple, hierarchical decision rules that are inferred in the data features. LR is an established statistical algorithm in binary classification which models a probability of a discrete result of an input variable . NB The third model that is founded on the application of Bayes theorem with a naive assumption of conditional independence between any two features. These particular algorithms were selected due to their demonstrated durability and common usage when solving real-life classification tasks, which gives a solid and stable performance baseline.

In the current research, a set of individual and ensemble ML models was selected to conduct the classification of Lumpy Skin Disease. In this way, it is possible to evaluate the baseline performance on the well-known individual classifiers and then demonstrate the improvement of performance with the more sophisticated ensemble method.

In order to add more on this baseline and have a higher predictive power, an ensemble learning model was also used. Precisely, the Bagging model was employed in maximizing the accuracy of the final prediction. Bagging is a meta-algorithm which is used to enhance the stability and accuracy of ML algorithms [19]. It works by making several random subsamples of the training set and training a distinct base classifier on each. The results are then averaged or a majority vote of the predictions by each individual classifier. The major objective of employing Bagging model is to minimize variance and prevent overfitting, and thus maximize the correctness of the final prediction. The success of this method can be accentuated by the fact that it is extensively used in various disciplines, including time series analysis, sentiment analysis, and fraud detection [20], [21].

2.4 Model Evaluation

Each ensemble model's performance is strictly evaluated using a comprehensive set of metrics, including accuracy, precision, recall, F-score, MCC (Matthews Correlation Coefficient), and execution time. These metrics comprehensively consider the models' abilities, since various characteristics like correct classification rate, ability to avoid false positives, balance between precision and recall, and computational efficiency.

3. Results and Discussion

In this section, the results of the application of the ML algorithms on the Lumpy Skin Disease dataset after balancing the training data using the SMOTE algorithm are presented and analyzed. Everything was done on the basic plan of Google Colab. The DT, LR and NB models are compared with the Bagging ensemble model. The main findings were summarized in Table 2 and plotted in ROC curve in Figure 3. The results are evident that Bagging ensemble model performed better in almost all the evaluation measures.

Table 2 shows the performance indicators and the execution times of the four Ensemble and ML models. Bagging ensemble was the most suitable model on all assessment measures among the four models. It obtained an accuracy of 0.895, implying that it correctly identified the highest proportion of instances. The Bagging ensemble also exhibited the highest precision of 0.959, which implies a false positive rate that is very low and is essential in diagnostics to prevent unneeded interventions. Furthermore, its recall of 0.847, and F-score of 0.900 show that it has a great capacity for recognizing positive cases and finding a balance between precision and recall. Besides, the Bagging ensemble obtained the highest MCC value of 0.797, representing a durable and robust positive correlation between the predicted and actual outcomes, which is a reliable measure for imbalanced data.

Table 2. Results of applied ML and Ensemble models after balancing the dataset

Model	Accuracy	Precision	Recall	F-Score	MCC	Execution Time
DT	0.85	0.935	0.784	0.853	0.714	0.0083
LR	0.84	0.876	0.829	0.852	0.679	0.0061
NB	0.815	0.878	0.775	0.823	0.636	0.0018
Bagging	0.895	0.959	0.847	0.900	0.797	0.0790

The Bagging model's better performance can be explained by its ensemble character. Bagging or Bootstrap Aggregating trains a number of classifiers on different random subsets of the training data and averages the predictions. The process can be applied in reducing individual model variability like DT, which can over fit the training samples. Combining the outputs in this way gives a more robust and generalized classifier that is less sensitive to the specific noise and variance of the training set, and hence more accurate and robust on unseen test data.

Comparatively, the DT and LR models were also resilient with accuracy of 0.85 and 0.84 respectively. The DT scored high with precision of 0.935 and MCC of 0.714 and the LR had a fair recall of 0.829 implying that both models are capable of accurate classification of cases. There was however a noticeable trade-off in computational speed. NB classifier was the least accurate and the fastest in execution time of 0.0018 seconds followed by LR with 0.0061 seconds and DT with 0.0083 seconds. Conversely, the Bagging ensemble took the most time to complete at 0.0790 seconds, which can be attributed to its ensemble character, which involves numerous training and prediction iterations.(Figure 3)

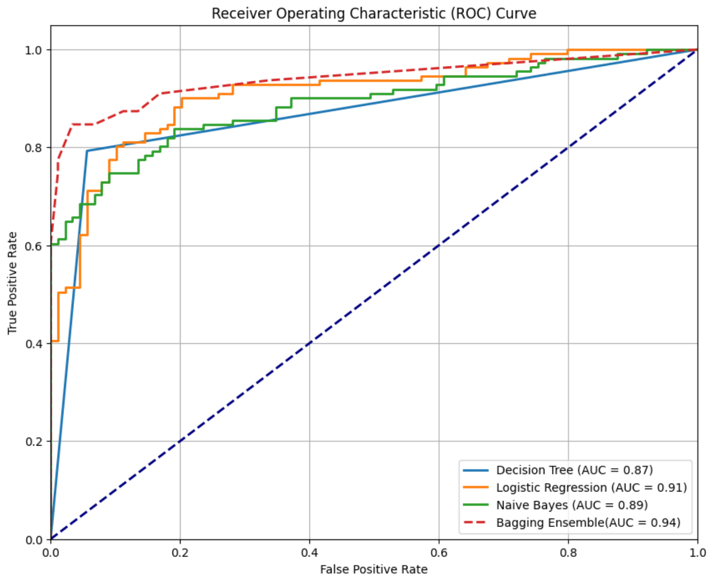


Figure 3. ROC Curve for ML and Ensemble models

Figure 3 represents the Receiver Operating Characteristic (ROC) of the ML and ensemble models discussed in this paper. The ROC curve draws a graph of the true positive rate against the false positive rate to give a picture of the diagnostic capability of a model, with the Area Under the Curve (AUC) being used as a summative performance indicator. The findings indicate clearly that the Bagging ensemble model had the best AUC of 0.94, which means it has the best ability to classify between the classes than the individual models. LR was the best of the standalone classifiers with an AUC of 0.91 followed by NB with an AUC of 0.89 and DT with an AUC of 0.87. The increased AUC value of the ensemble confirms the usefulness of the combination of models in order to improve the accuracy of the prediction and the overall robustness to this classification problem.

4. Conclusion

This study discuss how ML models classify Lumpy Skin Disease and how to handle the class imbalance which is a common problem in medical data. To overcome this issue, we applied SMOTE technique to balance the data and also ensemble learning techniques especially Bagging in order to improve the performance of the model. Bagging performed better than single models (DT, LT, and NB) and obtained the highest scores in terms of accuracy (89.5 %), precision, recall, F1 score, and MCC. The findings indicate that the integration of data-level balancing with model-level strategies has potentials of improving the diagnostic accuracy in veterinary health applications. This strategy contributed to the elimination of the bias that was caused by the lack of balance in the data and made the predictions more stable and reliable. Although this study was conducted on LSD, this approach may be helpful in other medical issues in which the issue of data imbalance exists. In future, we will increase the size of the dataset and will test on other ensemble techniques like Boosting and Stacking. Testing feature selection methods and deep learning models could also be promising in terms of new information and more accurate results. Ultimately, this would be translated into a useful diagnostic tool that can be used to identify LSD outbreak in the real world scenario in a fast and efficient manner.

REFERENCES

[1] Y. C. Wang and C. H. Cheng, “A multiple combined method for rebalancing medical data with class imbalances,” *Comput. Biol. Med.*, vol. 134, no. May, p. 104527, 2021, doi: 10.1016/j.compbiomed.2021.104527.

[2] C. Azad, B. Bhushan, R. Sharma, A. Shankar, K. K. Singh, and A. Khamparia, “Prediction model using SMOTE, genetic algorithm and decision tree (PMSGD) for classification of diabetes mellitus,” *Multimed. Syst.*, vol. 28, no.

4, pp. 1289–1307, 2022, doi: 10.1007/s00530-021-00817-2.

[3] N. Sultan, M. Hasan, M. F. Wahid, H. Saha, and A. Habib, "Cesarean Section Classification Using Machine Learning With Feature Selection, Data Balancing, and Explainability," *IEEE Access*, vol. 11, no. July, pp. 84487–84499, 2023, doi: 10.1109/ACCESS.2023.3303342.

[4] M. Zheng *et al.*, "An automatic sampling ratio detection method based on genetic algorithm for imbalanced data classification," *Knowledge-Based Syst.*, vol. 216, p. 106800, 2021, doi: 10.1016/j.knosys.2021.106800.

[5] S. U. Umar, M. R. Baker, and K. H. Jihad, "Machine Learning for Enhanced Diabetes Prediction: A SMOTE-Based Comparative Study," in *Lecture Notes in Networks and Systems*, 2025, vol. 1243 LNNS, pp. 123–133. doi: 10.1007/978-3-031-81080-0_12.

[6] A. S. H. Alwazy, G. Buyrukoglu, S. Buyrukoglu, and M. R. Baker, "Evaluating machine learning and statistical learning techniques for cancer classification and diagnosis," *Iran J. Comput. Sci.*, 2025, doi: 10.1007/s42044-025-00233-z.

[7] N. S. Rahmi, N. W. S. Wardhani, M. B. Mitakda, R. S. Fauztina, and I. Salsabila, "SMOTE Classification and Random Oversampling Naive Bayes in Imbalanced Data : (Case Study of Early Detection of Cervical Cancer in Indonesia)," *Proc. 2022 IEEE 7th Int. Conf. Inf. Technol. Digit. Appl. ICITDA 2022*, pp. 1–6, 2022, doi: 10.1109/ICITDA55840.2022.9971421.

[8] A. Ishaq *et al.*, "Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques," *IEEE Access*, vol. 9, pp. 39707–39716, 2021, doi: 10.1109/ACCESS.2021.3064084.

[9] R. Nithya, T. Kokilavani, and T. L. A. Beena, "Balancing cerebrovascular disease data with integrated ensemble learning and SVM-SMOTE," *Netw. Model. Anal. Heal. Informatics Bioinforma.*, vol. 13, no. 1, 2024, doi: 10.1007/s13721-024-00447-4.

[10] G. Mulugeta, T. Zewotir, A. S. Tegegne, L. H. Juhar, and M. B. Muleta, "Classification of imbalanced data using machine learning algorithms to predict the risk of renal graft failures in Ethiopia," *BMC Med. Inform. Decis. Mak.*, vol. 23, no. 1, pp. 1–17, 2023, doi: 10.1186/s12911-023-02185-5.

[11] L. Liu, X. Wu, S. Li, Y. Li, S. Tan, and Y. Bai, "Solving the class imbalance problem using ensemble algorithm: application of screening for aortic dissection," *BMC Med. Inform. Decis. Mak.*, vol. 22, no. 1, pp. 1–16, 2022, doi: 10.1186/s12911-022-01821-w.

[12] M. R. Baker *et al.*, "Comparison of Machine Learning Approaches for Detecting COVID-19-Lockdown-Related Discussions During Recovery and Lockdown Periods," *J. Oper. Intell.*, vol. 1, no. 1, pp. 11–29, Oct. 2023, doi: 10.31181/jopi1120233.

[13] A. Abdellatif, H. Abdellatef, J. Kanesan, C. O. Chow, J. H. Chuah, and H. M. Gheni, "An Effective Heart Disease Detection and Severity Level Classification Model Using Machine Learning and Hyperparameter Optimization Methods," *IEEE Access*, vol. 10, no. August, pp. 79974–79985, 2022, doi: 10.1109/ACCESS.2022.3191669.

[14] B. Zhang, X. Dong, Y. Hu, X. Jiang, and G. Li, "Classification and prediction of spinal disease based on the SMOTE-RFEXGBoost model," *PeerJ Comput. Sci.*, vol. 9, pp. 1–20, 2023, doi: 10.7717/PEERJ-CS.1280.

[15] N. L. Fitriyani, M. Syafrudin, G. Alfian, C. K. Yang, J. Rhee, and S. M. Ulyah, "Chronic Disease Prediction Model Using Integration of DBSCAN, SMOTE-ENN, and Random Forest," *2022 ASU Int. Conf. Emerg. Technol. Sustain. Intell. Syst. ICETSIS 2022*, pp. 289–294, 2022, doi: 10.1109/ICETSIS55481.2022.9888806.

[16] J. Yang and J. Guan, "A Heart Disease Prediction Model Based on Feature Optimization and Smote-Xgboost Algorithm," *Inf.*, vol. 13, no. 10, 2022, doi: 10.3390/info13100475.

[17] K. R. Mahmudah, F. Indriani, Y. Takemori-sakai, Y. Iwata, T. Wada, and K. Satou, "Classification of imbalanced data represented as binary features," *Appl. Sci.*, vol. 11, no. 17, 2021, doi: 10.3390/app11177825.

[18] S. Shahane, "Lumpy Skin Disease Dataset," *Kaggle Dataset*, 2023. https://www.kaggle.com/datasets/saurabhshahane/lumpy-skin-disease-dataset

[19] O. Nooruldeen, M. R. Baker, A. M. Aleesa, A. H. Ghareeb, and E. H. Shaker, "Strategies for predictive power: Machine learning models in city-scale load forecasting," *e-Prime - Adv. Electr. Eng. Electron. Energy*, vol. 6, 2023, doi: 10.1016/j.prime.2023.100392.

[20] E. F. Aziz and M. R. Baker, "Enhancing Multi-Class Password Strength Prediction Through Machine Learning and Ensemble Techniques," *Int. J. Saf. Secur. Eng.*, vol. 14, no. 5, pp. 1635–1645, 2024, doi: 10.18280/ijsse.140530.

[21] K. H. Jihad, M. R. Baker, M. Farhat, and M. Frikha, "Machine Learning-Based Social Media Text Analysis: Impact of the Rising Fuel Prices on Electric Vehicles," *Lect. Notes Networks Syst.*, vol. 647 LNNS, pp. 625–635, 2023, doi: 10.1007/978-3-031-27409-1_57/TABLES/5.