JOIV : Int. J. Inform. Visualization, 8(1) - March 2024 456-464

INTERNATIONAL JOURNAL
ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

A Comparative Analysis of Combination of CNN-Based Models with Ensemble Learning on Imbalanced Data

Xiaoling Gao [a,b], Nursuriati Jamil [a], Muhammad Izzad Ramli [a,*], Syed Mohd Zahid Syed Zainal Ariffin [a]

[a] *College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Shah Alam 40450, Selangor, Malaysia*
[b] *Ningxia University, Yinchuan, Ningxia, 750021, China*
Corresponding author: [*]izzadramli@uitm.edu.my

Abstract*— This study investigates the usefulness of the Synthetic Minority Oversampling Technique (SMOTE) in conjunction with convolutional neural network (CNN) models, which include both single and ensemble classifiers. The objective of this research is to handle the difficulty of multi-class imbalanced image classification. The application of SMOTE in imbalanced picture datasets is still underexplored, even though CNNs have been shown to be successful in image classification and that ensemble learning approaches have improved their performance. To investigate whether or not SMOTE can increase classification accuracy and other performance measures when combined with CNN-based classifiers, our research makes use of a CIFAR-10 dataset that has been artificially step-imbalanced and has varying imbalanced ratios. We conducted experiments using five distinct models, namely AdaBoost, XGBoost, standalone CNN, CNN-AdaBoost, and CNN-XGBoost, on datasets that were either imbalanced or SMOTE-balanced. Metrics such as accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC) were included in the evaluation process. The findings indicate that SMOTE dramatically improves the accuracy of minority classes, and that the combination of ensemble classifiers with CNNs and oversampling techniques significantly improves overall classification performance, particularly in situations when there is a high-class imbalance. When it comes to enhancing imbalanced classification tasks, this study demonstrates the potential of merging oversampling techniques with CNN-based ensemble classifiers to minimize the impacts of class imbalance in picture datasets. This suggests a promising direction for future research in this area.

***Keywords*— Deep learning; ensemble learning; SMOTE; imbalanced data; image classification.**

Manuscript received 4 Oct. 2023; revised 31 Jan. 2024; accepted 20 Feb. 2024. Date of publication 31 Mar. 2024.

I. INTRODUCTION

Image classification is a promising field in computer vision, but real-world datasets often have imbalanced distributions, with the majority classes having most samples and the minority classes being scarce[1]. The classification of imbalanced data is a subject of interest in data mining and machine learning, with applications in fields as diverse as finance, computer vision, biology, and medicine. It is also an urgent issue in the domain of image classification. When classifying imbalanced data, most classifiers will generally experience a certain degree of performance loss. In other words, although the classifier has obtained a high overall classification accuracy, it ignores the identification of minority samples.

Imbalanced data generally include two types: two-class and multi-class [2]. Multi-class imbalanced data presents more significant complex challenges than two-class data due to uneven distribution, overlapping multiple categories, and sample noise. Class imbalance presents a notable obstacle for image classification algorithms, as they are prone to exhibit bias towards the majority class. Consequently, this bias leads to subpar performance when accurately classifying minority classes. The challenge of class imbalance garnered the attention of researchers starting in the 1990s [3]. Subsequently, many learning methodologies have been devised and implemented to address this problem [4]. These methodologies encompass sampling, cost-sensitive approaches, ensemble methods, active learning, and one-class classification. In the present study, we extensively emphasize two prominent processes: sampling techniques and ensemble methods.

Data-level methods employ resampling approaches to obtain a balanced data distribution across various classes. The above approaches can be classified into three primary techniques: under-sampling, over-sampling, and hybrid sampling. Over-sampling is the process of duplicating present samples or adding additional samples to minority classes. The simplest method is random over-sampling, which involves raising the number of minority class instances via random replication. However, this approach is prone to overfitting. To address these issues, Chawla introduced SMOTE[5], in which new minority class samples are generated by using the k-nearest neighbors. Next, the algorithm calculates the dissimilarity between the specific feature vector and its nearest neighboring vector. The variation is multiplied by a stochastic variable ranging from 0 to 1. Ultimately, the multiplication output is combined with the specific feature vector to increase the occurrences of the minority class. Most over-sampling approaches apply improved SMOTE to increase samples from the minority class within the initial data set [6], including borderline-SMOTE [7], safe-level-SMOTE [8], SMOTE-IPF [9] and SMOTE-RSB [10]. However, these methods are mainly aimed at two-class imbalanced problems, and most data sets are tabular.

Ensemble learning uses multiple model predictions to improve predictive performance. Ensemble learning strategically combines classifiers or expert models for regression and classification [11]. Ensemble learning includes bagging, boosting, and stacking [12]. Since boosting aims to decrease bias, we focus on boosting ensemble learning. Boosting adds ensemble members sequentially to correct the predictions of the previous model and produce a weighted average [13]. AdaBoost (Adaptive Boosting)[14], Gradient Boosting [15], and Extreme Gradient Boosting (XGBoost) [16] are a few standard boosting techniques. AdaBoost is a machine learning technique that falls under the ensemble methods. AdaBoost aims to enhance empirical performance by combining multiple "weaker learners". XGBoost is a decision tree ensemble algorithm that employs gradient boosting and is renowned for its exceptional scalability.

For class imbalance issues, the ensemble algorithm is popular. Chawla et al[17] introduced the SMOTEBoost algorithm, which integrates SMOTE and the AdaBoost algorithm to improve generalization ability. The assessment index level primarily aims to explore and optimize the index of classification algorithms. Lv [18] used the oversampling SMOTE method, and the AdaBoost algorithm was used to process the unbalanced consumption data of credit cards. Empirical results show that the SMOTE-AdaBoost method outperforms the traditional AdaBoost method. Ileberi [19] proposed a machine learning-based system for credit card fraud detection, and the dataset is resampled using SMOTE. AdaBoost approaches can be combined with SVMs, Logistic Regression, Random Forests, Extreme Gradient Boosting, Decision Trees, and Extra Trees. The experimental results suggest that the AdaBoost-boosted model produces the best outcomes. Sainin et al. [20] found that the proposed ensemble classifier using sampling, feature selection, and the AdaboostM1 method with random forest can improve the performance. Su [21] proposed a model based on SMOTE-AdaBoost. The results have a better recognition effect on the intention of the combat target under the unbalanced data. The ensemble method proposed by Gao et al. [22] integrates various data balance techniques, such as adaptive synthetic sampling (ADSYN), borderline SMOTE, SMOTE-ENN, and SMOTE-Tomek combined with the XGBoost algorithm for gene identification. In summary, compared with the method using a single classifier, the boosting ensemble method can effectively overcome the class imbalance machine learning problem [23], showing superior performance[24].

Deep learning-based methods have been widely used in the image classification literature with notable achievements . Many researchers have proposed deep learning-based methods to identify multi-class imbalanced data due to neural networks' excellent durability and fault tolerance [25]. In recent years, many studies have combined CNN and ensembles in the field of imbalanced data. Taherkhani et al. [26] introduced AdaBoost-CNN to solve the dataset imbalance problem in multi-category scenarios. Jiang et al. [27] generated a new adaptive imbalance classification framework, combining HAFL-Boosting and ConvNeXt for high-dimensional image classification. The proposed framework extends boosting methods and deep neural networks to multi-class imbalanced classification situations and improves the model's imbalance handling. Babayomi et al. [28] suggest a model C-XGBoost for detecting brain cancers early on by combining CNNs and XGBoost. It has a lower model complexity than CNNs, making it quicker to train and less prone to overfitting in handling an imbalanced medical picture. Lv et al. [29] proposed an ERFS-based Ensemble CNN (EECNN) technique to solve the HSI classification multi-class imbalance problem. EECNN generates training samples using ROS with a dynamic sampling rate, enriches RFS data to provide a balanced training set, and uses CNN as a sub-classifier to develop an integrated classification learning model. According to experiments, this method solves the HSI multi-class imbalance problem better than CNN, RF, and deep learning ensemble methods. Choudhary et al.[30] proposed a multi-label ensemble classifier for heterogeneous and imbalanced brain CT investigations of ICH using CNN, SVM, and XG-Boost. The suggested method trains a CNN and feeds its extracted feature output to CNN, XG-Boost, and SVM for classification. Experimental results reveal that the model outperforms CNN, CNN with SVM, and CNN with XG-Boost in ICH detection. Zhao et al. [31] proposed a novel pulsar candidate classification framework, AdaBoost-MICNN, which combines the powerful ability of CNN to learn the mapping between input and output and the advantages of the AdaBoost algorithm to deal with unbalanced samples and can quickly and accurately find valid pulsar information under the dataset. Chinta et al. [32] proposed a classification framework pipeline that extracts features using a pre-trained ResNet 50 model and augments data using hybrid sampling with under-sampling and over-sampling. XGB beats the state-of-the-art in classification accuracy compared to Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN).

Due to the remarkable ability of CNNs to analyze and discover patterns in large amounts of data, combining CNNs and Ensembles can produce effective classification results. However, few papers have further compared simultaneously the differential performance of CNN combined with AdaBoost and XGBoost in imbalanced image classification in the case of oversampling methods. In this study, we aim to investigate the role and impact of the SMOTE method combined with single classifiers and ensemble classifiers in imbalanced image classification under different levels of imbalance. Therefore, we apply over-sampling, CNN, and ensemble methods to explore an efficient combination technique that deals with class imbalance issues in real-world scenarios and ensures fair representation of all classes for image classification.

II. Materials and Method

The execution of each experiment is thoroughly discussed in this section. First, the overall method is introduced. Next, the data set preprocessing and experimental setup are introduced. Finally, the observed metrics are analyzed.

A. The overall Method

This section describes the experiment using the SMOTE with CNN-based ensemble learning to investigate the problem of handling imbalanced image data. Our suggested methodology is summarized in Fig. 1. First, we utilize the benchmark image dataset CIFAR-10. Although the original dataset is balanced, according to the paper [33] we get an imbalanced CIFAR-10. To compare the performance estimation score achieved by each classifier, we use the test set with the same distribution as the training set to test the performance of the algorithms. Then, the SMOTE over-sampling method is used in an imbalanced train set. After that, we present the different classification methods in the balanced train set. The five methods include AdaBoost, XGBoost, CNN, CNN-AdaBoost, and CNN-XGBoost. Finally, the result of these five methods is compared with evaluation metrics.

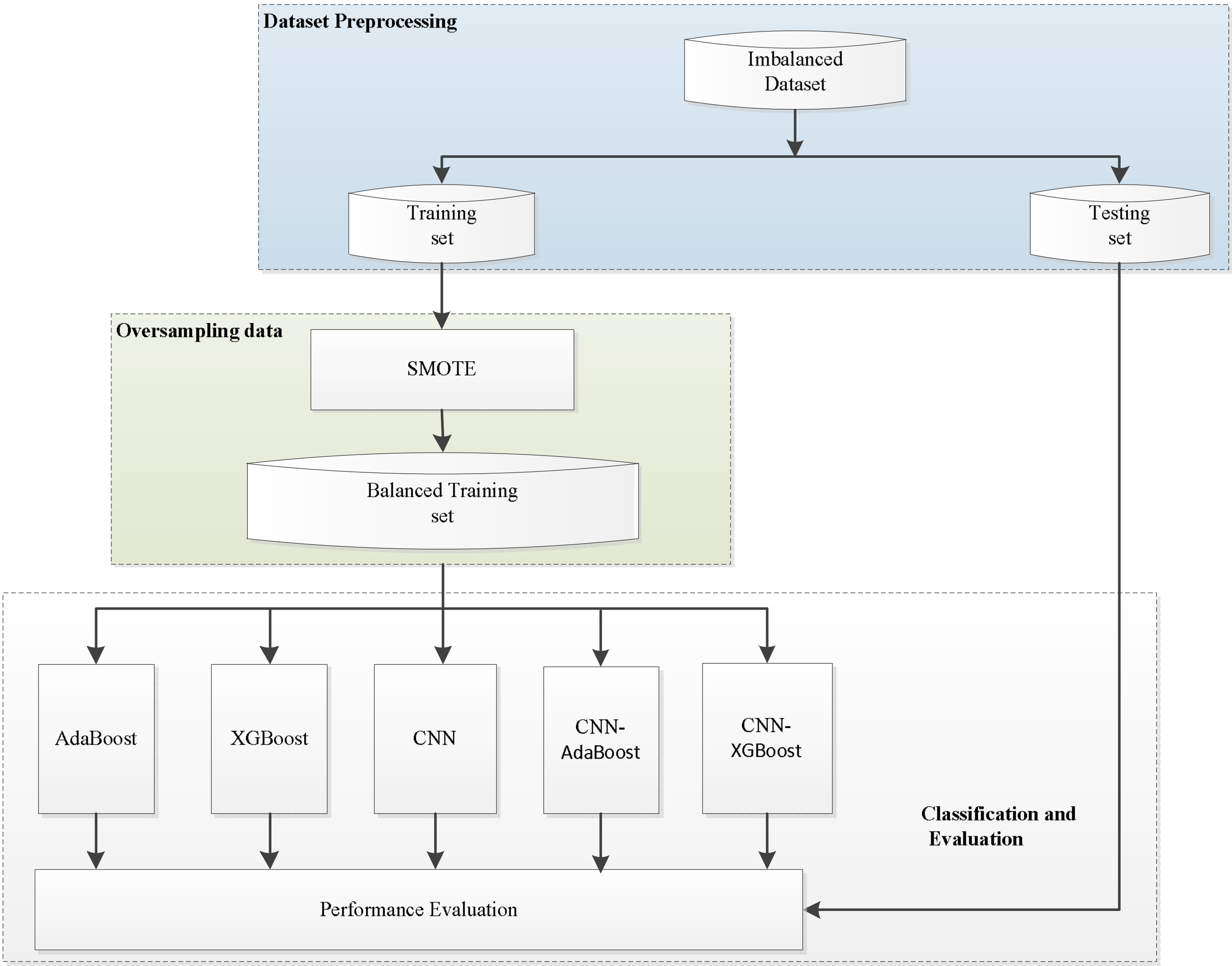


Fig. 1 Proposed classification model

B. Datasets Preprocessing

For evaluation, the CIFAR-10 benchmark data set was adopted. The CIFAR-10 dataset comprises 60,000 images, partitioned into a training set of 50,000 images and a testing set of 10,000 images. The images in the CIFAR-10 data set contain ten categories of natural objects. It is worth noting that these images do not demonstrate any inherent imbalance in their natural distribution. The Imbalanced CIFAR-10 is built on the original CIFAR-10 dataset by reducing the number of training samples per class, and the test set has the same ratio. An imbalance ratio is the ratio between sample sizes of the majority class and the minority class. That is the imbalance ratio IR = Nmax/Nmin.

Two types of imbalances, step imbalance, and long-tailed imbalance, are generated by implementing the methodology outlined in the references [34]. However, we only focus on the step imbalance in this study. For the step imbalance setting, all five minority classes have the same sample size as all five majority classes. The proportion of the minority class accounts for half of the total class. The IR we used in our experiments are 5, 10, and 100. We experimented on three imbalance ratios, and the goal is to investigate the performance variance of various classifiers under various IR of class imbalance. A test set's class distribution follows that of a training set to test the performance of the algorithms. The overview of the step imbalance dataset used in our experiments is shown in Table I.

TABLE I
OVERVIEW OF THE CIFAR-10 STEP IMBALANCE DATASET

Dataset	IR	Majority	Minority	Dimension
Train set	5, 10, 100	5000	1000, 500, 50	32X32
Test set	5, 10, 100	1000	200, 100, 10	32X32

C. Class Balance

Before classification algorithms were used, the dataset was first preprocessed with the SMOTE algorithm. We utilize the Normalize function to convert the data into a standard normal distribution, making the model easier to converge, where the three values of mean and std represent the three channels of the image. Normalization processing can speed up the convergence of the neural networks.

Once the data are separated into testing and training sets, the SMOTE is exclusively employed on the samples within the training set. To ensure that all classes in the training set can achieve a balanced distribution, SMOTE is used to create fresh artificial samples, which are subsequently added to the minority classes of the training set. After the SMOTE algorithm, the balanced training set is shown in Table II.

TABLE II
VARIATIONS ON THE TRAIN SET WITH SMOTE

IR	Imbalanced Train	Balanced Train	Test
5	30000	50000	6000
10	27500	50000	5500
100	25250	50000	5050

In this study, we used SMOTE to balance the data distribution, and subsequent work will be done on the balanced data set. However, SMOTE cannot be used directly for high-dimensional data. We modify SMOTE to handle image data. One solution to this problem is to apply the entire image as a feature vector. Thus, if the image's width is w pixels, its height is h pixels, and there are three channels, the feature vector will be $c \times w \times h$.

D. Classification Model

1) CNN-AdaBoost: The main idea of AdaBoost is to compute classification results. And assign weights to each base classifier. Among them, the number of iterations and the learning rate are the critical parameters in AdaBoost to improve the classification results. Each iteration produces a base classifier that has been trained with different sample learned coefficients. According to the training results, the sample learning factor will be modified at each iteration. When samples are misclassified with more weight, the classifier will prioritize these samples in the subsequent iterations. This study uses the CNN-AdaBoost model as shown in Fig. 2. Among them, three CNNs are set as the base classifiers for extracting image features, AdaBoost obtains three basic classifiers after three iterations, and then AdaBoost gets the final output by voting on these basic classifiers.

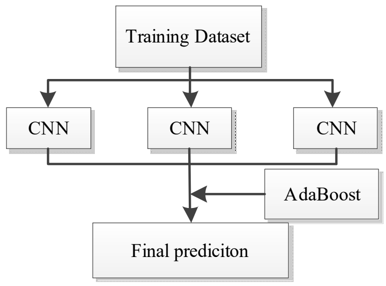


Fig. 2 CNN-Adaboost process flow chart

2) CNN-XGBoost: Several researchers have already proposed the CNN-XGBoost model, which combines the benefits of CNN and XGBoost. The CNN model automatically extracts features of varying levels, and the resulting feature vectors are input to the XGBoost model for event detection. The structure of the CNN-XGBoost network is depicted in Fig. 3. The CNN network consists of two convolution layers and two sampling layers, with the XGBoost classifier serving as the final output layer. The entire network is separated into two sections. The CNN model is initially used to extract the features, and the output of CNN's fully connected layer is fed into the XGBoost classifier. XGBoost will generate image classification results.

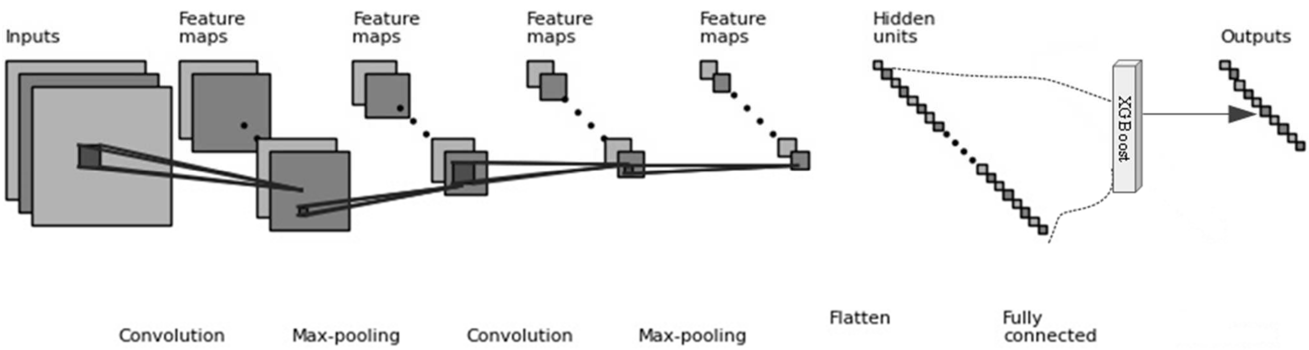


Fig. 3 Representation of the combined CNN-XGBoost technique

E. Experimental Setup

The application is developed on the Kaggle platform with the Python programming language, which is ideal for machine learning. This platform is accessible via the Internet. It maintains all of the Python libraries available. It provides considerable programming convenience in this approach.

For SMOTE, we used the imbalanced-learn package of the scikit-learn library [35]. Our empirical studies, or those reported in the relevant published literature were used to select parameter values. Cross-validation is crucial for model evaluation and comparison with other models, as it ensures that all samples appear at least once during the training and testing phases. The StratifiedKFold function uses a stratified random sampling method, and the proportion of different categories in the verification set is consistent with the proportion of the original sample. Therefore, in this experiment, we use 5-Fold to split the imbalanced training dataset, dividing the entire training set into five disjoint subsets while maintaining the sample category ratio. The percentages for each class are the same in the training and test sets. SMOTE is performed on every fold.

We experimented with 32 × 32 RGB images, as the original CNN implementation is readily available for use at the required resolution. The architecture of CNN used in the experiments is shown in Table III. The CNN-AdaBoost and CNN-XGBoost methods use the same CNN architecture as the base classifier. Using these parameters, the accuracy of the three models was compared based on the same underlying resources. To address the overfitting problem, the optimizer was served by stochastic gradient descent (SGD) with the weight decay set to 0.001. Cross-entropy loss (CEL) is used to evaluate the loss between the actual and predicted classes. We tested different learning rates, and 0.001 is a local optimal value. We take 64 pictures as a batch and whole pictures as iterations to calculate the classification accuracy for 50 iterations. The principal parameters of the experiment are described in Table IV as follows:

TABLE III
THE ARCHITECTURE OF CNN USED IN EXPERIMENTS

Layer	Kernel size	Stride	Padding	Channels	Activation
Input	N/A	N/A	N/A	3	N/A
Convolution 2D	3 x 3	1	SAME	32	ReLu+Batch Norm2d(32)
Convolution 2D	3 x 3	1	SAME	32	ReLu+Batch Norm2d(32)
Maxpooling 2D	2 x 2	2	VALID	32	Dropout(0.2)
Convolution 2D	3 x 3	1	SAME	64	ReLu+Batch Norm2d(64)
Convolution 2D	3 x 3	1	SAME	64	ReLu+Batch Norm2d(64)
Maxpooling 2D	2 x 2	2	VALID	64	Dropout(0.3)
Convolution 2D	3 x 3	1	SAME	128	ReLu+Batch Norm2d(128)
Convolution 2D	3 x 3	1	SAME	128	ReLu+Batch Norm2d(128)
Maxpooling 2D	2 x 2	2	VALID	128	Dropout(0.4)
Fully connected	1 x 1	N/A	N/A	128	ReLu+ Dropout(0.5)
Fully connected	1 x 1	N/A	N/A	10	N/A

TABLE IV
HYPER-PARAMETER TUNNING OF CLASSIFICATION MODELS

Model	Hyper-Parameter	Range
AdaBoost	Learning rate	0.1,0.01,0.001
	n_estimators	(5,50)
XGBoost	Learning rate	(0.05,0.1)
	Maximum tree depth	(7,10)
	Subsample ratio	(0.1,1)
	Column subsample ratio	(0.1,1)
	Minimum child weight	(1,10)
	Gamma	(0,0.01)
CNN	batch size	16,32,64,128
	Learning rate	0.1,0.01,0.001

F. Evaluation Methods

In this research, we are attempting to resolve a multiclassification issue. As a result, we evaluate our models using a multi-class confusion matrix [36]. To assess the efficacy of our proposed techniques, five indicators are used to study model performance, including accuracy, precision, recall, F1 score, and area under a receiver operating characteristic curve (AUC). Precision is the total number of correct predictions made by the model. Accuracy, Sensitivity, Specificity, and F1 score are mathematically described in terms of the confusion matrix, as shown in Table V:

TABLE V
THE EVALUATION METRICS USED IN THE EXPERIMENT

Metrics	Calculation
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
F1-score	$\frac{2TP}{2TP + FP + FN}$

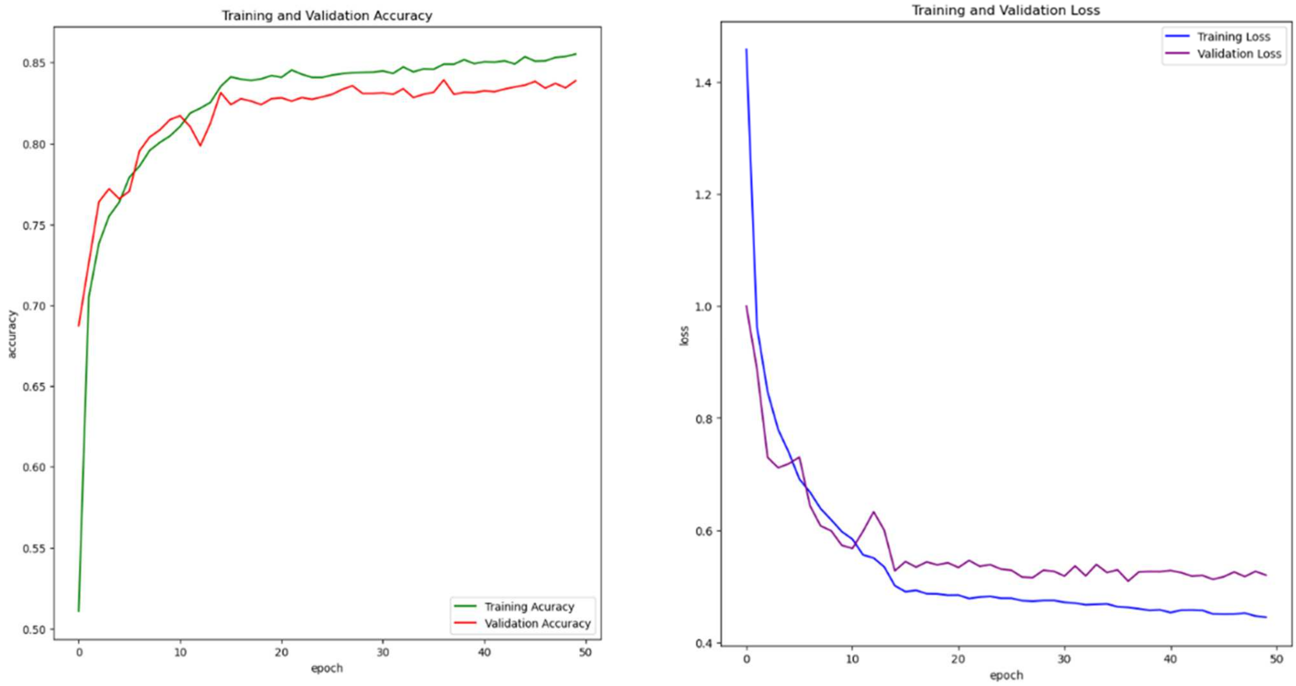
where the terms TP, TN, FP, and FN represent True Positive, True Negative, False Positive, and False Negative, respectively.

III. RESULTS AND DISCUSSION

In this section, the results of all experiments are described in detail. Finally, the experimental performance metrics are compared and analyzed.

A. Results of Classification

This research proposed five different classification algorithms and their ensembles to deal with sizeable imbalanced image datasets efficiently. To investigate which method performs consistently better than others, a comparison is made with CNN on the imbalanced data set as a baseline. This indicates that the training data set for the baseline does not contain replicated or created data. It should be noted that to analyze the impact of SMOTE on the imbalanced data set, the data set in the whole experiment was not additional data augmentation. We run the same tests under the same setting for datasets with and without SMOTE and then compare various performance evaluation metrics. The learning curve is one of the most essential evaluation metrics for the deep learning algorithm. As shown in Fig. 4, the baseline CNN model reduces training and validation losses as the number of epochs increases.



(a) (b)

Fig. 4 Loss and accuracy of the CNN model (IR=10)

As shown in Table VI, the CNN-AdaBoost classification approach produces an overall test accuracy that exceeds 75%, higher than that of CNN- XGBoost and CNN. The single CNN model outperforms the single XGBoost and AdaBoost models. Whether SMOTE is used, the CNN-AdaBoost model has higher classification accuracy than the AdaBoost model. The same happens with CNN-XGBoost. The CNN-XGBoost model has higher classification accuracy than the XGBoost model. The results show that CNN has superior feature extraction capabilities and can improve the classification performance of AdaBoost and XGBoost.

TABLE VI
CLASSIFICATION ACCURACY ON DIFFERENT IR

Methods	IR=5		IR=10		IR=100	
	Imb.	Bal.	Imb.	Bal.	Imb.	Bal.
AdaBoost	.4408	.3328	.4640	.3024	.3152	.2586
XGBoost	.6395	.6262	.6853	.6629	.7442	.7382
CNN	.7585	.7386	.7941	.7652	.8588	.8338
CNN-AdaBoost	**.7682**	**.7520**	**.8133**	**.7909**	**.8717**	**.8636**
CNN-XGBoost	.7405	.7052	.7821	.7427	.8533	.8147

In the CNN-AdaBoost and CNN-XGBoost investigation, the Boosting part outperforms the CNN part regarding feature extraction quality. This demonstrates that the Boosting model could train its weak learners more effectively and that their assembly could fill in some gaps in the CNN features initially retrieved and enhance the classification outcome. In addition, after using the SMOTE algorithm for five models, they all got lower accuracy than before. As the SMOTE approach gives more weight to the minority class, the model becomes biased towards it. The model will forecast the minority class more accurately, but overall accuracy will decline, as shown in Fig. 5.

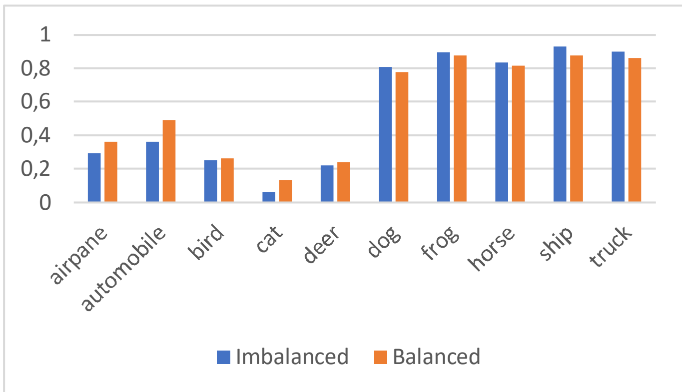
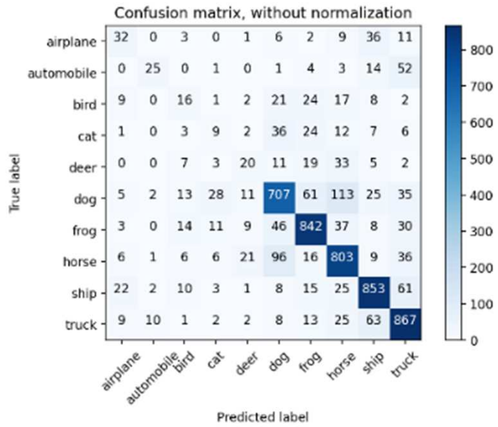


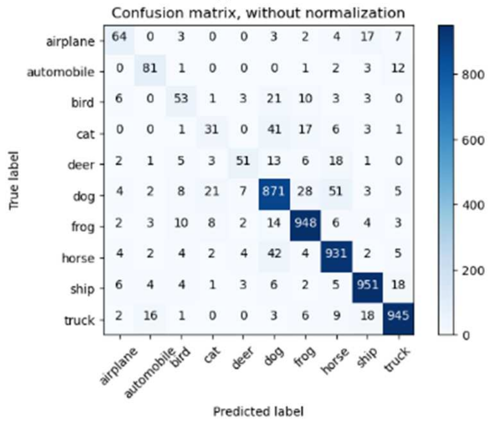
Fig. 5 Performance of each class in CNN-AdaBoost (IR=10)

The confusion matrix above in Fig.6 shows that the color of the main diagonal gradually deepens, and the number of tuples correctly predicted for each class grows to some amount. On the other hand, the number of prediction error classes of tuples outside the main diagonal rapidly diminishes. The testing accuracy for CNN-AdaBoost is higher than the best validation accuracy produced by CNN's basic classification model. This demonstrates that the primary classification method of AdaBoost can improve the accuracy made by the basic classification approach. However, the test accuracy of CNN-

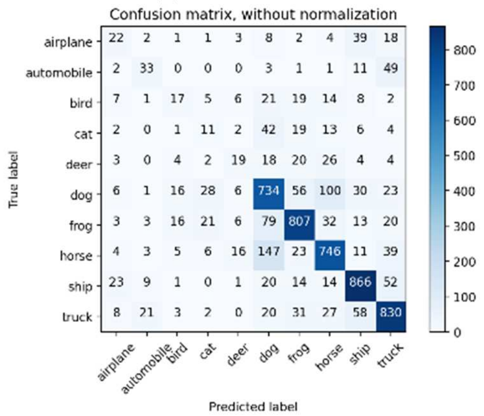
XGBoost is lower than CNN. When the base classifier is not weak, XGBoost will not improve the results. This demonstrates the distinction between AdaBoost and XGBoost when combined with the CNN model.



(a) CNN



(b) CNN-AdaBoost



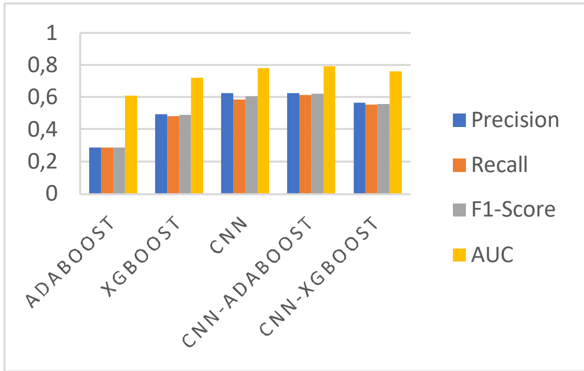
(c) CNN-XGBoost

Fig. 6 Confusion matrix (IR=10)

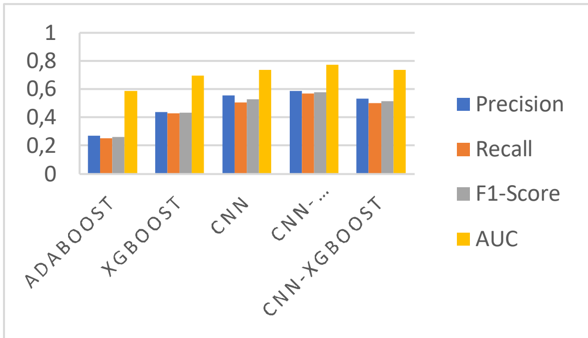
B. Performance comparison

In this study, the SMOTE method was combined with five single and CNN-based ensemble classifiers and 5-fold cross-validation was used to compare the performance of the classification models based on the accuracy, sensitivity, specificity, and F1 score obtained from experiments. The experimental results are shown in Fig.7.

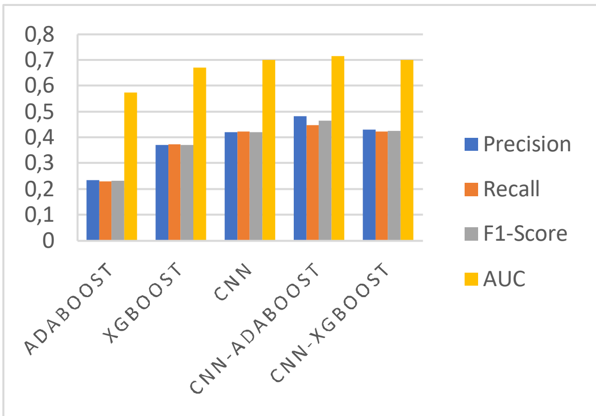
As previously concluded in the literature, CNNs outperform conventional reinforcement learning statistically for image classification under imbalanced conditions. However, after using SMOTE, CNN-AdaBoost obtains the best F1, AUC. Moreover, as the IR value increases, CNN-AdaBoost achieves the highest score, regardless of the imbalanced or balanced training set. The results show that CNN-AdaBoost outperforms CNN and CNN-XGBoost for the classification of multi-class imbalanced image datasets, demonstrating the advantages of the CNN-AdaBoost algorithm in dealing with imbalanced image classification.



(a) IR=5



(b) IR=10



(c) IR=100

Fig. 7 Classification performance on different IR

In addition, according to the findings, an increase in IR decreases AUC, recall, and F-measure. Using SMOTE oversampling increases the TP rate while reducing the FN rate, resulting in significantly higher AUC and recall for all classifiers. Furthermore, CNN-AdaBoost leads by 0.06%, considering the accuracy of the five classification algorithms. This shows that CNN-AdaBoost outperforms CNN-XGBoost and CNN in optimistic class prediction. Besides, CNN-AdaBoost performs better than 0.02% in terms of recall. This indicates that CNN-AdaBoost is also good at distinguishing positive classes. In addition, using the AUC value, the CNN-AdaBoost model was evaluated better on the balanced training set with SMOTE than the imbalanced training set, as it yielded a higher AUC value of 79.18%. Only the CNN-AdaBoost model was combined with the SMOTE algorithm. It performs better on balanced datasets than on imbalanced datasets.

IV. CONCLUSION

Imbalanced data is expected in real-world applications. The performance of traditional classification methods is frequently poor when dealing with imbalanced image data. In this paper, we propose an imbalanced image data classification model based on CNN, including SMOTE technique and Boosting ensemble learning, which can effectively cope with the problem of multi-class imbalance. The effectiveness of the recently proposed combination approaches CNN-AdaBoost and CNN-XGBoost in handling multi-class imbalanced image classification issues is investigated, and a comprehensive comparison is performed.

The results demonstrate that the application of the SMOTE approach can increase minority class accuracy. Based on the findings in Fig. 5, 6, and 7, both CNN-AdaBoost and CNN-XGBoost combined SMOTE approaches have produced positive outcomes for handling multi-class imbalanced images. This combination typically outperforms individual classifiers in resolving multi-class imbalance, even at high IR. The CNN-AdaBoost algorithm showed higher accuracy and performance among the five algorithms. Although the suggested CNN combines Boosting and SMOTE classification methods perform outstandingly, it could be observed from the experiments that the method still has limitations. This model must be validated on another skewed image dataset. Furthermore, the CNN-AdaBoost method cannot obtain a significant advantage over other algorithms on all evaluation metrics.

In terms of our future work, we will propose the improved SMOTE approach for imbalanced image data as well as investigate the use of contrastive learning to classify long-tailed imbalanced images, such as the Cifar-10-TL and iNaturalist data sets, to improve the effectiveness of ensemble learning in other computer vision classification datasets.

ACKNOWLEDGMENT

This study is funded by Ministry of Higher Education Malaysia through Fundamental Research Grant Scheme (FRGS/1/2022/ICT02/UITM/02/8). Due acknowledgements are also accorded to the College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Shah Alam, for supporting this study.

REFERENCES

[1] Yakshit, G. Kaur, V. Kaur, Y. Sharma, and V. Bansal, "Analyzing various Machine Learning Algorithms with SMOTE and ADASYN for Image Classification having Imbalanced Data," *Proceedings of 2022 IEEE International Conference on Current Development in Engineering and Technology, CCET 2022*, no. December 2022, 2022, doi: 10.1109/CCET56606.2022.10080783.

[2] Z. Yang, R. O. Sinnott, J. Bailey, and Q. Ke, *A survey of automated data augmentation algorithms for deep learning-based image classification tasks*, vol. 65, no. 7. Springer London, 2023. doi:10.1007/s10115-023-01853-2.

[3] A. Ali, S. M. Shamsuddin, and A. L. Ralescu, "Classification with class imbalance problem: A review," *International Journal of Advances in Soft Computing and its Applications*, vol. 7, no. 3, pp. 176–204, 2015.

[4] M. Han, A. Li, Z. Gao, D. Mu, and S. Liu, "A survey of multi-class imbalanced data classification methods," *Journal of Intelligent and Fuzzy Systems*, vol. 44, no. 2, pp. 2471–2501, 2023, doi: 10.3233/JIFS-221902.

[5] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, no. Sept. 28, pp. 321–357, 2002, [Online]. Available: https://arxiv.org/pdf/1106.1813.pdf%0Ahttp://www.snopes.com/horrors/insects/telamonia.asp

[6] G. Kovács, "Smote-variants: A python implementation of 85 minority oversampling techniques," *Neurocomputing*, vol. 366, pp. 352–354, 2019, doi: 10.1016/j.neucom.2019.06.100.

[7] H. Han, W. Y. Wang, and B. H. Mao, "Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning," *Lecture Notes in Computer Science*, vol. 3644, no. PART I, pp. 878–887, 2005, doi: 10.1007/11538059_91.

[8] C. Bunkhumpornpat, K. Sinapiromsaran, and C. Lursinsap, "Safe-level-SMOTE: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5476 LNAI, pp. 475–482, 2009, doi: 10.1007/978-3-642-01307-2_43.

[9] J. Stefanowski and F. Herrera, "SMOTE-IPF : Addressing the noisy and borderline examples problem in imbalanced classification by a re-sampling method with filtering," no. September, 2014, doi:10.1016/j.ins.2014.08.051.

[10] E. Ramentol, Y. Caballero, and R. Bello, "SMOTE-RSB ∗ : a hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory," 2011, doi: 10.1007/s10115-011-0465-6.

[11] X. Dong and Z. Yu, "A survey on ensemble learning," vol. 14, no. 2, pp. 241–258, 2020.

[12] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 8, no. 4, pp. 1–18, 2018, doi: 10.1002/widm.1249.

[13] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," *Eng Appl Artif Intell*, vol. 115, no. June, p. 105151, 2022, doi: 10.1016/j.engappai.2022.105151.

[14] M. Rahman, R. Prodhan, Y. Shishir, and S. Ripon, "Analyzing and Evaluating Boosting-Based CNN Algorithms for Image Classification," *2021 International Conference on Intelligent Technologies, CONIT 2021*, pp. 1–6, 2021, doi: 10.1109/CONIT51480.2021.9498328.

[15] S. Badirli, X. Liu, Z. Xing, A. Bhowmik, K. Doan, and S. S. Keerthi, "Gradient Boosting Neural Networks: GrowNet," 2020, [Online]. Available: http://arxiv.org/abs/2002.07971

[16] T. Chen and C. Guestrin, "XGBoost : A Scalable Tree Boosting System," pp. 785–794, 2016, doi: 10.1145/2939672.2939785.

[17] N. V Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "LNAI 2838 - SMOTEBoost: Improving Prediction of the Minority Class in Boosting," 2003.

[18] M. Lv, Y. Ren, and Y. Chen, "Research on imbalanced data : based on SMOTE-adaboost algorithm," *2019 IEEE 3rd International Conference on Electronic Information Technology and Computer Engineering, EITCE 2019*, pp. 1165–1170, 2019, doi:10.1109/EITCE47263.2019.9094859.

[19] E. Ileberi, Y. Sun, and Z. Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost," *IEEE Access*, vol. 9, pp. 165286–165294, 2021, doi:10.1109/ACCESS.2021.3134330.

[20] M. S. Sainin, R. Alfred, and F. Ahmad, "Ensemble Meta Classifier with Sampling and Feature Selection for Data with Multiclass Imbalance Problem Faculty of Computing and Informatics School of Computing , Universiti Utara Malaysia , Malaysia," vol. 2, no. 2, pp. 103–133, 2021.

[21] Q. Su *et al.*, "Target Intention Recognition Model Based on SMOTE-AdaBoost under Unbalanced Samples," vol. 6027, pp. 227–231, 2022.

[22] Q. Gao *et al.*, "Identification of Orphan Genes in Unbalanced Datasets Based on Ensemble Learning," *Front Genet*, vol. 11, no. October, 2020, doi: 10.3389/fgene.2020.00820.

[23] J. Edward, M. M. Rosli, Y. Chua, N. Alicezah, M. Kasim, and H. Nawawi, "Classification Prediction of Familial Hypercholesterolemia using Ensemble-based Classifier with Feature Selection and Rebalancing Technique," pp. 278–283, 2022.

[24] R. Panthong, “Combining SMOTE and OVA with Deep Learning and Ensemble Classifiers for Multiclass Imbalanced,” 2022, doi:10.3844/jcssp.2022.732.742.

[25] X. Jiang, J. Wang, and J. Wang, “An adaptive multi-class imbalanced classification framework based on ensemble methods and deep network,” *Neural Comput Appl*, vol. 35, no. 15, pp. 11141–11159, 2023, doi: 10.1007/s00521-023-08290-w.

[26] A. A. Khan, O. Chaudhari, and R. Chandra, “A review of ensemble learning and data augmentation models for class imbalanced problems: combination, implementation and evaluation,” 2023, [Online]. Available: http://arxiv.org/abs/2304.02858

[27] A. Taherkhani, G. Cosma, and T. M. McGinnity, “AdaBoost-CNN: An adaptive boosting algorithm for convolutional neural networks to classify multi-class imbalanced datasets using transfer learning,” *Neurocomputing*, vol. 404, pp. 351–366, 2020, doi:10.1016/j.neucom.2020.03.064.

[28] Babayomi M, Olagbaju OA, Kadiri AA, “Convolutional XGBoost (C-XGBOOST) Model for Brain Tumor Detection,” Jan.2023, arXiv preprint arXiv:2301.02317.

[29] Q. Lv *et al.*, “Enhanced-Random-Feature-Subspace-Based Ensemble CNN for the Imbalanced Hyperspectral Image Classification,” *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 14, pp. 3988–3999, 2021, doi: 10.1109/JSTARS.2021.3069013.

[30] B. Choudhary, A. Babu, and U. T. B, *An Ensemble MultiLabel Classifier for Intra-Cranial Haemorrhage Detection from Large, Heterogeneous and Imbalanced Database*, vol. 2. Springer Nature Switzerland. doi: 10.1007/978-3-031-27609-5.

[31] H. Zhao, J. Jin, Y. Liu, Y. Shen, and Y. Jiang, “AdaBoost-MICNN: A new network framework for pulsar candidate selection,” *Mon Not R Astron Soc*, vol. 513, no. 2, pp. 2869–2883, 2022, doi:10.1093/mnras/stac619.

[32] U. Chinta and A. Atyabi, “A Framework Pipeline to Address Imbalanced Class Distribution Problem in Real-world Datasets,” *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 746–753, 2023, doi:10.1109/CCWC57344.2023.10099163.

[33] H. P. Chou, S. C. Chang, J. Y. Pan, W. Wei, and D. C. Juan, “Remix: Rebalanced Mixup,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12540 LNCS, pp. 95–110, 2020, doi:10.1007/978-3-030-65414-6_9.

[34] Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, *106*, 249–259. doi: 10.1016/j.neunet.2018.07.011

[35] J. Hao and T. K. Ho, “Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language,” *Journal of Educational and Behavioral Statistics*, vol. 44, no. 3, pp. 348–361, 2019, doi: 10.3102/1076998619832248.

[36] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Inf Process Manag*, vol. 45, no. 4, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.