The Indonesian Journal of Computer Science
www.ijcs.net
Volume 14, Issue 5, October 2025
https://doi.org/10.33022/ijcs.v14i5.5020

A Comparative Study of Multi-class Classification Based on Imbalanced Data: A Review

Rojan Zaki Abdulkreem[1], Adnan Mohsin Abdulazeez [2]
rojan.zaki@auas.edu.krd [1], adnan.mohsin@dpu.edu.krd [2]
[1]Akre University for Applied Science / Technical College of Informatics – Akre / Department of Information Technology – Kurdistan Region – Iraq
[2]Duhok Polytechnic University – Kurdistan Region

Article Information

Received : 7 Oct 2025
Revised : 18 Oct 2025
Accepted : 23 Oct 2025

Keywords

Imbalanced data,
Multiclass classification,
DL model,
Coast-sensitive learning,
Ensemble methods.

Abstract

Classification of unbalanced multiclass datasets is still a major challenge in machine learning in many fields of applications, including medical diagnostics, fraud detection, and picture classification, where minority classes are the most crucial, but at the same time under-represented. Classical classification algorithms designed for balanced data tend to overfit the majority classes deeming a large number of minority classes misclassified and, as a result, compromising the model's performance. This review covers the main state-of-the-art techniques for class imbalance problems including under-sampling and over-sampling techniques, ensemble approaches, cost-sensitive learning, and producing synthetic data via SMOTE (synthetic minority oversampling technique). Recently, GANs (Generative Adversarial Networks) have also been employed to generate synthetic data, specifically valuable for complex datasets where realistic data augmentation is needed. Each of these techniques is analyzed in terms of their capability of dealing with imbalanced data through conventional metrics such as accuracy and specific metrics for imbalanced datasets such as F1-score, G-mean, and others. Recent advancements, such as hybrid approaches and learning from deep learning models are also discussed as viable solutions given the complexities associated with big data (high dimensional and large) and their corresponding models. Such comparative analysis should facilitate the construction of more robust models that handle complex data in modern applications.

A. Introduction

The multi-class classification problem – the problem of classifying an observation into two or more classes – from an imbalanced data perspective, has received considerable attention in the machine-learning field since this problem is widespread in real-world applications [1] [2]. When the classes in a dataset are not equally represented – in other words, when an imbalance between classes exists – it can be a major cause of the poor performance of a classification algorithm in general, and those that were designed for the case of balanced classes in particular [3]. The problem is particularly critical in medical diagnosis and fraud detection cases [4]. Or image classification, where the less-represented class (the minority class) is of higher importance, but its representation in the data is lower [5].
Studies also show that ordinary classification methods tend to be biased towards favoring the majority class, causing minority classes to have high rates of misclassification [6]. This can cause the developed models to fail to generalize well because this minority class is often the class of interest [7]. For instance, some studies show that across a whole suite of problems, simply having class imbalance damages the overall performance metrics, typically based on accuracy or area under the curve (AUC). This has led to the use of special techniques to overcome these biases [8] [9].

Several ideas have been proposed for mitigating the effects of class imbalance in multi-class classification. [10]. including resampling approaches, involving either over-sampling of the minority class and under-sampling of the majority class, that is, sampling a balanced version of the data; and ensemble approaches, where multiple classifiers are combined into a single classifier to improve the collective results of separate classifiers by leveraging strengths of each algorithm to compensate the other [11]. More sophisticated techniques, such as deep learning and reinforcement learning have been studied as well, and show promise to achieve the same purpose of balancing the accuracy of minority classes as a traditional cost-sensitive learning solution does [12]

Recent research on how to tackle the problem of multi-class classification on imbalanced datasets mostly makes use of cost-sensitive learning, an approach that assigns high misclassification costs to minority classes, so that they get more attention in the training phase [13]. Similarly, recent developments indicate that cost-sensitive learning can improve the classification of minority classes, especially in critical areas like medical diagnosis, fraud detection, and tobacco studies, where errors have potentially severe consequences [14]. Further combining this approach with resampling techniques, such as either under-sampling the majority class or oversampling the minority class, has been proven to provide better model performance [15].

Another good way to deal with imbalanced data is to make fake data using methods like SMOTE [16]. SMOTE creates new samples for the smaller class, which helps balance the dataset and makes the model work better on new data [17]. New research has started to use GANs to make realistic fake data. This has been helpful in tricky areas like image sorting and medical information [18]. These methods are key to making classification models perform better when real-world data is very uneven [19] [20].

What's more, picking the right evaluation metrics plays a key role when working with imbalanced datasets [21]. Regular metrics like accuracy can be misleading [22]. So, it's better to use metrics that focus on how well the model handles minority classes such as G-mean and F1-score, to get a more accurate picture of how the model is doing [23]. This change highlights the need to create evaluation strategies that show how classifiers perform with imbalanced data [24].

Tackling the problems of multi-class classification in uneven datasets needs a many-sided approach. This includes using special algorithms resampling methods, and the right evaluation measures. We need to keep studying this area to build stronger models. These models should handle the complexities of uneven data well. In the end, this will lead to better results in key applications.

This study looks at and compares different ways to boost multi-class classification on uneven datasets. By checking out methods like resampling, cost-sensitive learning, and cutting-edge models, it tries to show how well they work to make under-represented classes perform better in key areas like spotting fraud and diagnosing medical conditions.

The hint of this paper is organized as follows. Section 2 reviews different methods for handling imbalanced multiclass data, counting resampling approach, cost-sensitive learning, and deep learning approaches. Section 3 instant the results of the comparative analysis and highlights the strengths and circumspection of each method. A discussion of these findings is conferred in Section 4. Finally, the conclusion in Section 5 summarizes the key acumen and offers suggestions for future research.

B. Theoritical Background

B.1 Multi-class Classification

Multi-class Classification Predicting one of three or more potential class labels for a particular instance is known as multi-class classification [25]. Unlike binary classification, which has only two possible outcomes, multi-class classification must manage more intricate decision boundaries across several classes. It is extensively used in domains such as medical diagnostics, natural language processing, and picture identification. Increased processing complexity and possible class overlap are major obstacles that can make precise categorization challenging. [26].

B.1.1 Techniques

- **One-vs-Rest (OVR):** This method allows binary classifiers to be reused by dividing the multiclass problem into multiple binary classification tasks, one for each class, as illustrated in Figure A.
- **The One-vs-One (OVO):** approach creates classifiers for each pair of classes, producing a system that selects the class with the most "wins.", as illustrated in Figure B [27].
- **Native Multi-class Algorithms:** Algorithms such as decision trees, k-nearest neighbors (KNN), and neural networks are naturally suitable for

multiclass classification without the need for disintegration into binary tasks [28].

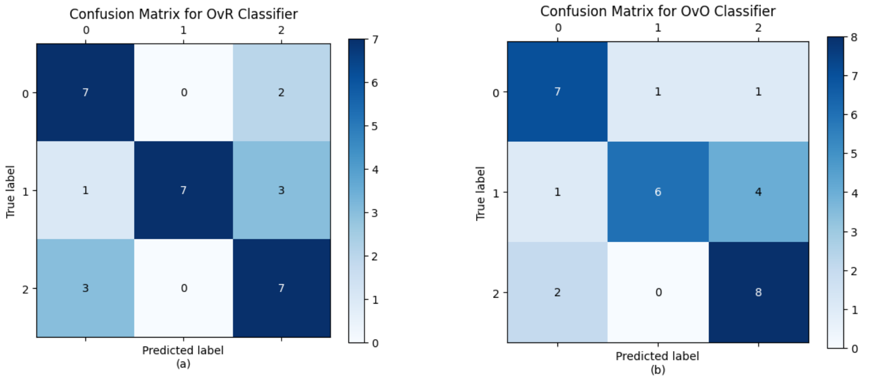

Figure 1. The confusion matrices for (a) One-vs-One (OvO) classifier and (b) One-vs-Rest (OvR) classifier.

B.1.2 Challenges

- **Complexity:** This problem requires more processing because the model must consider additional class barriers.
- **Class Overlap**: There is a greater incidental of class overlaps when there are more classes, which may result in incorrect classification [29].

B.2 Imbalanced data in multiclass classification

Imbalanced datasets in multiclass classification appear when certain classes have a disproportionately low number of instances. This imbalance may distort model performance, favoring majority classes while misclassifying or overlooking minority classes. Opposition classes frequently perform crucial events in applications, such as disease diagnosis or fraud detection; thus, it is essential to number this imbalance to agree on accurate and decent forecasts. [30].

B.2.1 Consequences of Class Imbalance

- **Majority Class Bias:** Models often predict the majority class more often, overlooking the minority class. This happens even when the minority class represents key outcomes in fields like fraud detection or medical diagnosis.
- **Skewed Decision Boundaries:** When data isn't balanced, the lines between classes can get pushed closer to minority class examples. This makes it harder for the model to learn general rules and leads to wrong classifications [31].

B.2.2 Multi-class Imbalance Challenges

In contrast to the binary classification, imbalances can occur in varying amounts across several classes. Resolving imbalances is more difficult because every minority class may require different treatments. [32].

B.3 Existing Approaches for Handling Imbalanced Data

B.3.1 Under-sampling / Oversampling

These techniques adjust the class distribution of the dataset to balance it out by summing less of the majority class (under-sampling) or more of the minority class (oversampling). These techniques are shown in Figure 2: Relationship between Under-sampling and Oversampling Techniques, which shows the influence of each one on the minority and majority classes in a 2D dataset. Subsampling will lose data but oversampling might risk overfitting with duplicates [33].

Figure 2. compares under-sampling and over-sampling methods.

B.3.2 SMOTE

smote addresses class bias by calculating artificial minorities through interpolation rather than duplicated data points [34]. This approach, as shown in Figure 3: SMOTE Process, generates new, heterogeneous instances of the minority class, maximizing the model's exposure to the minority class and eliminating the possibility of overfitting [35].

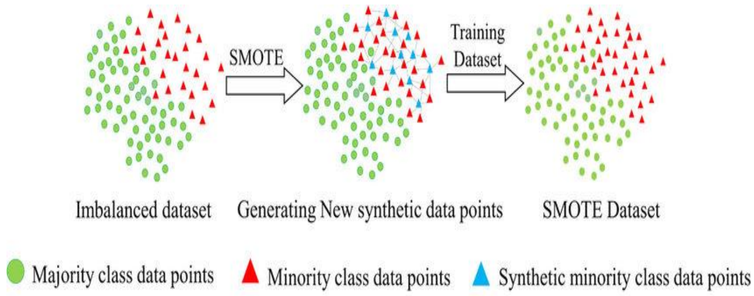

Figure 3. Illustrates the SMOTE Process.

B.3.3 GAN

GANs are increasingly used to generate artificial data samples and are useful for class imbalance in applications like medical imaging and object recognition [36]. GANs are two neural networks (the generator and the discriminator) that compete against each other to generate realistic data that simulate the minority class, thereby broadening its image. This has proven useful in producing high-quality data, particularly for larger datasets where traditional approaches cannot generate persuasive minority samples [37].

B.3.4 Ensemble

- **Boosting techniques:** Misclassified examples of minority-class instances can be assigned extra weight by techniques such as AdaBoost and Gradient Boosting, and performance can be boosted as a result [38].
- **Bagging:** bagging combines a bunch of training subsets with decision-tree classification to improve both the accuracy of classification and deal with imbalances [39].

B.3.5 Algorithmic Adjustments

- **Ensemble Methods:** Techniques such as boosting (e.g., AdaBoost, Gradient Boosting) and bagging (e.g., Random Forests) improve performance on imbalanced datasets by combining multiple models. Boosting emphasizes learning from misclassified instances, which often include minority class samples [40].
- **Cost-conscious Learning:** This method modifies the algorithm to make the model more sensitive to minority classes by giving them larger misclassification penalties [41].

B.4 Evaluation Metrics for Imbalanced Data

Therefore, when the data are imbalanced, the traditional accuracy metrics become extremely misleading see Figure 4 below. As a result, a model can achieve deceptively high accuracy levels by simply predicting the majority class, almost all of the time. Instead, precision, recall, F1-score (the harmonic means of precision and recall), and area under the receiver operating characteristic curve or AUC-ROC (a number between 0 and 1, which measures the performance of a binary classifier) are far more reliable performance metrics when working with imbalanced data [42]. These metrics report on the performance of the model on the minority class and are sensitive to the class distribution. Performance of a model on an imbalanced dataset, with confusion matrix, key metrics, and ROC curve depicting precision, recall, and overall classification accuracy [43][44].

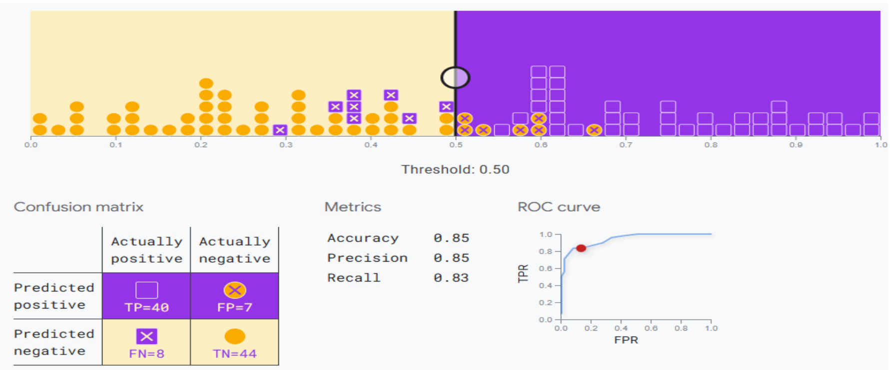

Figure 4. Iillustrates the model's performance metrics, confusion matrix, and ROC curve at a threshold of 0.50.

B.5 Recent advances in multiclass classification for imbalanced data

More recently, research has shifted attention to sophisticated approaches to classify unbalanced datasets by introducing a more sensitive model to minority group samples. Specifically, cost-sensitive learning imposes additional penalties for misclassifying different classes (a generally higher penalty for the minority ones) or adapting deep learning models by tweaking the loss function or through hybrid approaches [45].

B.5.1 Cost-sensitive Learning and Algorithmic Innovations

Cost-efficient learning modifies model training to ensure the correct classification of minority classes by penalizing misclassification. This method is extremely important for unbalanced multiclass environments, where standard algorithms generally are not robust to underrepresented classes [46]. Realistic, cost-effective models can dynamically weight minority classes more heavily, causing the learning process to focus more on these examples. For example, focal loss, a loss function tailored to a specific class, reduces the effect of easily classifiable examples while increasing the impact of hard-to-classify samples, such as minority class samples. Weighted cross-entropy functions work in much the same way, but they tweak error propagation to privilege minority classes. This value-based paradigm is especially effective in fields such as fraud detection, where the minority group constitutes unusual, costly instances. Furthermore, innovations enable cost-efficient tuning of deep neural networks, making the approach flexible and extensible in different data domains [47].

B.5.2 Hybrid and Deep Learning Approaches

Hybrid approaches use resampling along with deep learning algorithms to control the data imbalance. Resampling techniques like SMOTE initially create artificial minority samples by interpolating over previously existing samples, thus achieving greater diversity without redundancy. The weighted data representations are passed into deep learning structures (CNN for images, RNN for sequential data) which provide an excellent way to model complex data patterns. Second, (GANs) have redefined synthetic data generation by producing authentic, high-quality minority class samples through adversarial training where two networks (a generator and a discriminator) continuously improve on one another [48].Such GAN-generated data reduces overfitting risks posed by traditional oversampling, which is particularly useful in domains such as medical imaging where we cannot get much data from minority classes (e.g., rare diseases). Furthermore, transfer learning extends these hybrid approaches by leveraging pre-trained models as a starting point, tailoring them to the characteristics of imbalanced datasets with few data to increase the generalization power of the model to minority classes. Hybrid methods combine all of these methods to create a more balanced class representation, thus facilitating the model to better handle real-world, high-dimensional, complex datasets [49].

C. Related Work

Table 1: Summary of multiclass classification studies and comparative analyses.

Ref	Paper	Type of classification	Dataset	Technique	Metrics	Results	Conclusion	Strengths	Limitations	Future directions
[1]	Bearing Fault Classification Using Multi-Class Machine Learning (ML) Techniques (2023).	Binary: defective vs. non-defective; 3-class: non-defective, inner ring defect, roller defect; 7-class: no defect, three inner ring defects, three roller defects.	Data from the Politecnico Di Torino bearing test rig includes 6-channel acceleration impulses from two accelerometers, recorded at 51.2 kHz for 10 seconds under varying load and speed.	Classical ML algorithms: LR, K-NN, RF, SVC, KSVM. Features extracted using tsfresh.	Accuracy, Precision, Recall, F1-score, ROC, AUC.	KSVM and K-NN achieved F1 scores of 0.9997 and 0.9986 in binary classification, 0.9993 and 0.9926 in 3-class, and 0.9996 and 0.9955 in 7-class classification.	Machine learning techniques enabled precise bearing defect categorization, with K-NN and KSVM outperforming all others.	Near-perfect classification with ML techniques. Successfully handled multiple fault conditions and operational conditions.	Natural evolving bearing problems have not yet been included in the current work, which focuses on artificially created defects.	Expand research to include natural defects with unsupervised learning, and examine ML's impact on time-evolving bearing flaws.
[2]	Uncertainty estimation-based adversarial attack in multi-class classification (2023).	They specifically use a 'one-versus-all' approach to convert the problem into binary classifications for each class.	The experiments used MNIST, featuring handwritten numbers, and Cora, a graph-structured dataset of categorized scholarly publications.	Monte-Carlo-based adversarial attack (MC-AA) for uncertainty estimation, compared against other methods like MC-dropout and Deep Ensemble.	AUC scores were 0.889 for Cora (LEConv) and 0.98 for MNIST (CNN), indicating effective class distinction.	MC-AA excelled at estimating model uncertainty near decision boundaries but struggled in multi-class settings, resolved by converting to binary classifications.	MC-AA provides accurate predictions and manages uncertainties in overlapping class distributions, making it effective for multi-class tasks.	MC-AA outperforms traditional deep learning techniques, manages overlapping class distributions well, and achieves high accuracy in uncertainty estimation.	MC-AA faces challenges with multi-class problems without binary conversion and can be computationally intensive. Performance may falter in highly overlapping classes.	To increase accuracy and resilience, future studies could construct hybrid models, investigate more datasets, and optimize MC-AA for multi-class tasks.
[3]	Efficient Multiclass Classific	They specifically focus on datasets with	Utilizes high-dimensional datasets across	A hybrid feature-selection approach combining	Accuracy, computational	Achieved over 75% feature reduction and a	By maximizing accuracy and	Superior accuracy, a notable decrease in feature	Limited effectiveness in addressing storage	Additional investigation into the optimization of

	ation Using Feature Selection in High-Dimensional Datasets (2023).	numerous features.	various fields such as bioinformatics and engineering.	filter and wrapper methods, specifically (MISFS).	efficiency, and feature reduction ratio.	maximum accuracy of 97%, outperforming existing methods.	minimizing features, the MISFS algorithm efficiently tackles problems with high-dimensional datasets.	dimensionality, and a shorter computation time than current techniques.	space issues and other complexities in high-dimensional data	computing space and the use of the method for additional categorization issues.
[4]	CPS Attack Detection under Limited Local Information in Cyber Security: An Ensemble Multi-Node Multi-Class Classification Approach (2024).	to different cyber-physical system (CPS) attacks (e.g., Fuzzers, DoS, Reconnaissance, etc.).	Uses datasets such as UNSW-NB15 and Edge-IIoT for cyber-attack detection, using data from several dispersed nodes with local data filtering, where each node includes only a few classes.	Ensemble learning with local classifiers (e.g., Logistic Regression, SVM, Random Forest) and (GMM) for density estimation.	Precision, Recall, Accuracy, KL divergence for distribution drift analysis.	High attack detection precision and recall (around 100%) were attained, and the ensemble method occasionally outperformed the full-data approach because it was better able to handle data imbalances.	The suggested ensemble multi-node method eliminates the necessity for raw data sharing between nodes and effectively classifies CPS attacks with little local knowledge.	Avoiding data sharing helps to address concerns about data privacy and communication costs; this works even when data is censored.	The method may struggle with imbalanced datasets and requires a sufficient number of samples for effective local classification.	Investigate increasingly intricate data structures and evaluate the strategy using bigger datasets and actual CPS implementations.
[5]	Parallel CNN-ELM: A multiclass classification of chest X-ray images to identify seventeen lung diseases including	17 lung conditions, including atelectasis, cardiomegaly, pneumonia, COVID-19, and tuberculosis, were multiclassified from chest X-	Includes 40,490 CXR images of 17 lung diseases from sources like CXR14, Kaggle, and BIMCV, with 32,463 for training and 8,028 for testing.	Hybrid model combining Parallel CNN for feature extraction and ELM for classification. Grad-CAM is used for the visualization of important regions.	Precision, Recall, F1-Score, Accuracy, Area Under the Curve (AUC).	Achieved 90.92% accuracy and 96.93% AUC for classifying 17 diseases. High accuracy for specific diseases: 99.37% for COVID-19 and 99.98% for tuberculosi	The model outperformed SOTA models on balanced and imbalanced datasets, with a mobile app prototype demons	Fast processing (0.9 ms per image), low complexity, high accuracy for lung illnesses, and effective Grad-CAM use for interpretability.	A model generalization for rare diseases may be impacted by dataset imbalance, particularly for conditions like hernia (only 88 pictures).	Enhance the model's performance with more data for uncommon illnesses and prepare it for use in mobile medical apps.

	COVID-19 (2023).	rays.				s.	trating real-world use.			
[6]	Kurdish News Dataset Headlines (KNDH) through multiclass classification (2024).	Multiclass classification of news headlines into five categories: Social, Sport, Health, Economic, and Technology.	34 news outlets provided 50,000 Kurdish news headlines, evenly split across five classifications.	Used (Kurdish LPT) for tokenization, spell-checking, stemming, and preprocessing.	Precision, recall, F1-score, and accuracy.	The model achieved high accuracy in classifying news headlines into the correct categories.	The study effectively showed that effective text categorization models could be constructed using the dataset.	Large and balanced dataset for Kurdish news, preprocessed for NLP tasks, includes proper linguistic features.	Restricted to news headlines and the Kurdish language (Sorani dialect); might not transfer to other text formats.	Expand the dataset to include more categories and different dialects of the Kurdish language, and enhance preprocessing tools.
[7]	Multi-Class Skin Cancer Classification Using Vision Transformer Networks and Convolutional Neural Network-Based Pre-Trained Models (2023).	Skin cancer lesions.	Used data augmentation strategies to correct the class imbalance in the HAM10000 dataset that was obtained via Kaggle.	Vision Transformer (ViT) and various CNN-based transfer learning models (like ResNet, DenseNet, and VGG).	Accuracy, recall, precision, f1-score, AUC.	Achieved an accuracy of 92.14% with the ViT model, outperforming CNN-based transfer learning methods across multiple metrics.	The ViT model is a competitive substitute for conventional CNNs since it successfully categorizes different forms of skin cancer and corrects class imbalance through data augmentation.	High classification accuracy, efficient management of class imbalance, and a thorough comparison with the most advanced models.	Dependence on high-quality images; potential overfitting due to class imbalance despite augmentation; requires significant computational resources for training.	Examine more sophisticated augmentation techniques, incorporate more datasets for better generalization, and look into real-time implementation in medical environments.
[8]	A heuristic method for discovering multi-class classification rules from	Using multi-source data for complex tabular datasets across fields like healthcare, life sciences, and	Uses public datasets like UCI's Australia, Germany, Breast Cancer, Mushroom, Wine, and Lending Club, featuring diverse	HEA, GA, and PSO to generate classification rules without requiring extensi	Accuracy, ACC, AUC, number of rules, ANAs per rule.	The proposed method achieved better or comparable performance in classification tasks across various datasets,	The heuristic approach greatly improves performance and explainability	Handles mixed, imbalanced, and incomplete multi-source data effectively, generating fewer, simpler	Performance may differ in circumstances of missing values and very unbalanced datasets, necessitating	Investigate sophisticated approaches to managing mixed data, enhance rule simplification for better interpretab

	multi-source data in cloud–edge system (2023).	finance.	classes, feature types, and imbalance rates.	ve preprocessing.		using fewer and simpler rules than existing methods.	over traditional and SOTA classifiers, providing precise predictions with clear principles.	rules with high accuracy and generalization.	additional optimization in practical settings.	ility, and expand the method's applicability to more datasets and domains.
[9]	Comparison of multiclass classification techniques using dry bean dataset (2024).	The seven distinct types of dry beans: are Seker, Barbunya, Bombay, Cali, Horoz, Sira, and Dermason.	Dry bean dataset from UCI Machine Learning Repository, containing 13,611 samples, with 16 features (e.g., area, perimeter, axis lengths) and imbalanced classes.	KNN, DT, LR, NB, RF, XGB, SVM, MLP, and ADASYN algorithms were used for balancing the dataset.	ACC, SE, SP, F1-Score, AUC, Cohen's Kappa (Kappa), MSE, FPR.	XGB achieved the highest accuracy: 93% without ADASYN and 95.4% with ADASYN. KNN and RF also performed well, with ACC of 95% and 94% respectively, when ADASYN was applied.	KNN and RF were similar on balanced datasets, but XGB excelled on both balanced and skewed data. ADASYN boosted accuracy for all models, especially with unbalanced data.	Effectively addresses imbalanced datasets with ADASYN, enhancing accuracy and highlighting key features: ShapeFactor2, Minor Axis Length, and ShapeFactor1.	The study excluded the beans' texture and color and solely looked at their geometric characteristics. Additional testing is required for different seed types or more complicated datasets, as the sample was limited to dry beans.	For a more thorough study, use extra elements like texture or color. To increase the model's generalization and robustness, test the approach on more agricultural datasets or expand the dataset.
[10]	Multiclass skin lesion classification in dermoscopic images using the Swin transformer model (2024).	Skin lesion.	ISIC 2019 dataset with 25,331 images across eight diagnostic categories: AK, BCC, BKL, DF, NV, etc.	CNN and transformer architectures are integrated into the Swin transformer model.	Sensitivity: 82.3%, Specificity: 97.9%, Accuracy: 97.2%, Balanced Accuracy: 82.3%.	Obtained the best-balanced accuracy when compared to the most advanced techniques.	The proposed method effectively improves classification performance in multiclass skin lesion diagnosis.	Weighted cross-entropy loss is used to address class imbalance; it has good sensitivity and specificity.	Challenges remain with intraclass variation and low contrast in dermoscopic images; and reliance on large datasets for training.	Examine more diagnostic categories, real-time applications, and preprocessing/postprocessing improvements.

[11]	Multi-Class Classification of Plant Leaf Diseases Using Feature Fusion of Deep Convolutional Neural Network and Local Binary Pattern (2023).	To plant leaf diseases, including apple, tomato, and grape leaf diseases with multiple disease classes for each leaf type.	Three PlantVillage datasets were used: Apple Leaf (4,645 images, 4 classes), Grape Leaf (4,639 images, 4 classes), and Tomato Leaf (18,160 images, 10 classes), augmented to address imbalances.	The method combined hand-crafted LBP features with deep CNN features utilizing max-pooling layers and separable convolutions, then fused the features and performed SoftMax classification.	Accuracy, Precision, Recall, F1-Score, AUC-ROC Curve, Confusion Matrix.	Achieved 99%, 96.6%, and 98.5% validation accuracies, and 98.8%, 96.5%, and 98.3% test accuracies for Apple, Tomato, and Grape datasets, outperforming traditional CNNs and transfer learning.	The proposed deep and LBP feature combination enhances plant leaf disease categorization accuracy, outperforming existing methods while reducing computational costs.	High accuracy, low complexity, better feature extraction with deep learning and LBP fusion, and effective unbalanced dataset handling via data augmentation.	The method is only tested on three datasets, and there is potential for overfitting in real-time applications. Future testing on other plant leaf datasets is required.	Investigate the use of LBP variants for more robust feature extraction, expand the method to classify diseases in other crops, and test the model's performance in real-time edge computing systems.
[12]	Skin Lesion Segmentation and Multiclass Classification Using Deep Learning Features and Improved Moth Flame Optimization (2021).	Classification into seven skin lesion classes (e.g., melanoma, benign keratosis, basal cell carcinoma, etc.).	ISBI 2016, ISBI 2017, ISBI 2018, PH2 for segmentation; HAM10000 for classification.	The project involves image enhancement with LCcHIV, segmentation using a 10-layer CNN, feature extraction with ResNet 101 and DenseNet201, and optimization with IMFO and KELM.	Accuracy, F1-score, AUC	Segmentation accuracy: ISBI 2016 (95.38%), ISBI 2017 (95.79%), ISIC 2018 (92.69%), PH2 (98.70%). Classification accuracy on HAM10000: 90.67%.	The suggested approach outperformed numerous cutting-edge methods in terms of multiclass classification and skin lesion segmentation, achieving excellent accuracy in	High segmentation accuracy with DSS and feature optimization. Effective combination of CNN feature extraction and optimization techniques.	Data augmentation was necessary due to the imbalance in the classification dataset (HAM10000).	Explore unsupervised methods and apply the method to real-time medical applications, especially with evolving datasets or natural skin lesions.

							both areas.			
[13]	Improving Multiclass Classification of Fake News Using BERT-Based Models and ChatGPT-Augmented Data (2022).	Multiclass classification	1876 messages in the CheckThat! Lab datasets from CLEF-2022 were classified as true, somewhat false, false, and other.	BERT-based models (SBERT, RoBERTa, mBERT) with ChatGPT-augmented data for class balance.	Macro-averaged F1-score	exceeded current methods in terms of performance.	The study demonstrates the potential of transformer models and AI-generated data in enhancing fake news detection accuracy across multiple classes.	Improves class balance by using AI-generated data and cutting-edge transformer models.	The practical application of FND needs further improvements; challenges in achieving high accuracy in multiclass scenarios remain.	To increase classification accuracy, more research should be done on dataset diversity and model improvements.
[14]	Multiclass blood cancer classification using deep CNN with optimized features (2023).	the Acute Lymphoblastic Leukemia (ALL) into four classes: Benign, Early Pre-B, Pre-B, and Pro-B.	3262 blood smear photos from 89 patients—25 healthy and 64 suspected of having ALL—are included in the collection. A Zeiss microscope was used to take pictures at a magnification of 100x.	combo of pre-trained CNN models (VGG19, ResNet 50, InceptionV3) for feature extraction, followed by feature selection approach(PCA, LDA, SVC) and ML classifiers like SVM and LR.	Accuracy, Precision, Recall, F1-Score, Confusion Matrix.	Reached a maximum accuracy of 99.84% using the Logistic Regression classifier, ResNet50, and SVC feature selection.	The suggested model successfully combined CNN architectures with feature optimization strategies to improve the accuracy of leukemia classification.	High accuracy, the capacity to handle multiclass classification of blood cancer, and the use of nature-inspired algorithms (PSO, CSO) for feature selection.	Takes a lot of time and computing power to train; it may have trouble with smaller datasets or other kinds of medical imaging.	Examine more sophisticated optimization methods, make use of bigger and more varied datasets, and look at the usage of cancer classification schemes other than leukemia.
[15]	High-precision multiclass classification of lung	Multiclass classification of 15 different lung diseases using	The ChestX-ray14 dataset consists of 112,120 frontal-view X-rays	MobileLungNetV2, a variant of MobileNetV3, was	Accuracy, Precision, Recall, Specificity, FPR, FNR, and F1-score.	MobileLungNetV2 achieved 96.97% accuracy, 96.71% precision, 96.83%	MobileLungNetV2 excelled in lung disease classification	The fine-tuned MobileLungNetV2 achieved high accuracy (96.97%),	The dataset is imbalanced, leading to over-representation of	To increase the robustness of the boost model: 1) increase the size of the dataset,

	disease through customized MobileNetV2 from chest X-ray images (2023).	deep learning methods (specifically CNN-based).	of the lungs from 30,805 patients and is annotated with 15 classes of diseases, including cardiomegaly, pneumonia, and emphysema.	fine-tuned and compared with the transfer learning models: InceptionV3, AlexNet, and VGG19.		recall, and 99.78% specificity, with Emphysema at 99.92% specificity and Pneumothorax at 98.06% recall.	from X-rays, outperforming other transfer learning models in accuracy and precision.	precision, recall, and specificity across classes, outperforming other CNN models.	some classes and under-representation of others, which could skew performance.	and 2) incorporate more advanced optimizations such as the monarch butterfly and HHO algorithms.
[16]	A machine learning model for multi-class classification of quenched and partitioned steel microstructure type by the k-nearest neighbor algorithm (2023).	Single-label multiclass classification problem involving six types of microstructures: {M, B, RA}, {M, F, B, RA}, {M, F, RA}, {M, F, RA, C}, {M, RA}, and {M, RA, C}.	A dataset of 348 samples was compiled from 107 articles, including compositional and heat treatment parameters. Features include C, Mn, Si, TMAE (total microalloying elements), Ac1, Ac3, Ms, QT, and PT.	K-NN	Achieved an overall performance of 97.7% in the training dataset and 77.7% in the testing dataset measured as f1-Score.	The martensite-retained austenite {M, RA} type was found to be the most perplexing class, despite the model's excellent accuracy in classifying microstructure types.	The study showed machine learning effectively classifies steel microstructures and revealed how composition and heat treatment affect microstructure evolution.	High classification accuracy; comprehensive dataset; ability to explore compositional effects on microstructure; potential for industrial applications in steel design.	Between training and testing datasets, the model's performance differed considerably; certain classes were harder to correctly categorize.	Further research could involve expanding the dataset with more diverse samples and exploring other machine-learning techniques to improve classification accuracy and generalization capabilities.
[17]	An effective multiclass skin cancer classification approach based on deep convolutional neural network (2023).	Skin cancer lesions.	Utilizes the HAM10000 (10,015 images across 7 classes) and ISIC-2019.	DCNN compared with VGG16, VGG19, DenseNet121, DenseNet201, and MobileNetV2.	Accuracy, recall, precision, f1-score, specificity, AUC.	Achieved 98.5% accuracy on HAM10000 and 97.1% on ISIC-2019, outperforming other models.	Class imbalance is successfully addressed by the suggested DCNN model, which also improves skin cancer detection classification	High classification accuracy, robust against class imbalance, applicable to diverse datasets.	Preprocessing methods are not thoroughly explored, and any overfitting problems are not addressed.	Further assessment with additional datasets, exploration of real-time application in clinical settings, and integration with other diagnostic tools.

							accuracy.			
[18]	A GAN-Based Data Augmentation Method for Imbalanced Multi-Class Skin Lesion Classification (2024).	The skin lesions into seven categories, including melanoma, keratosis, nevus, etc.	The dataset with 10,015 dermoscopic images across seven skin disease classes, is heavily imbalanced (e.g., the largest class has 6705 images, and the smallest class has 115 images).	The two-stage GAN approach (STGAN) uses a ResNet 50 for classification after class-specific generation and an unconditional GAN for universal knowledge.	Accuracy, Sensitivity, Precision, F1-score, Specificity, Fréchet Inception Distance (FID), Inception Score.	98.23% accuracy, 88.85% sensitivity, 90.23% precision, 89.48% F1-score, and 98.34% specificity were attained. Produced pictures with a 15.63% increase in Inception Score and a 15.71% increase in FID over baseline.	By successfully addressing the mode collapse issue in GANs for minority classes, STGAN enhances classification performance and image generation quality.	Improves classification for imbalanced datasets, generates high-quality and diverse images, boosts CNN classifier performance, and maintains high precision and recall for minority classes.	Classification performance may be impacted if generated images contain artifacts from the original dataset, such as hair and veins.	Explore image inpainting techniques to remove natural artifacts from generated images and further enhance the classification performance.
[19]	Pitfalls of assessing extracted hierarchies for multi-class classification (2023).	HMC	OpenML and UCI public datasets (such as Collins, Abalone, MNIST, and Covertype)	HAC, RBD, CBD	Accuracy, Error Propagation	RBD and CBD hierarchies improved accuracy on complex datasets, with random hierarchies sometimes outperforming single classifiers.	Informed hierarchies capture class relationships well, yet random hierarchies can outperform non-hierarchical methods.	Analyzing hierarchy extraction clarifies its impact on HMC performance, separating hierarchy quality from HMC benefits.	The power of classifiers may lessen the influence of hierarchy quality, making it more difficult to assess the hierarchy's actual worth.	Investigate further into optimizing hierarchy extraction methods and balancing classifier complexity with hierarchy quality.
[20]	Multi-Class Unlearning for Image Classification via Weight Filtering (2023).	To image datasets where multiple classes can be unlearned selectively (e.g., image classes like airplanes, cats, and dogs)	Evaluated on MNIST (60k/10k, 10 classes), CIFAR-10 (60k, 10 classes), and ImageNet subset (5k, 1k classes).	WF-Net uses a Weight-Filtering layer to unlearn specific classes in one round, working with both CNNs	Accuracy on retained and forgotten sets, Zero Retrain Forgetting (ZRF) score, Relearn Time (RT), Insertion score, and Deleti	WF-Net achieved near-zero accuracy on forgotten classes, high accuracy on retained ones, strong ZRF scores, and faster relearn times; ViT models	WF-Net ensures explainability while enabling effective multi-class unlearning in one round and is generali	Provides explainability, reduces costs by unlearning multiple classes at once, and works with ViTs and CNNs while retaining relevant	ViT models showed slightly lower retention accuracy, with larger datasets like ImageNet exhibiting greater performance gaps	Extend unlearning to large datasets, optimize for Transformers, and explore privacy-focused applications.

		.		and Vision Transformers (ViTs).	on score.	performed slightly worse.	zable to other datasets and systems.	classes.	than MNIST or CIFAR-10.	
[21]	Deep Learning-Based Skin Lesion Multi-Class Classification with Global Average Pooling Improvement (2023).	The skin lesions into seven classes, including melanoma, basal cell carcinoma, squamous cell carcinoma, benign keratosis, and others.	The HAM10000 dataset includes 10,015 images of seven skin lesion types, with 6,705 in the largest class (nv) and 115 in the smallest (pdf), making it imbalanced.	The DCNN model employed global average pooling, black hat filtering, histogram equalization, and data augmentation, compared to VGG16 and others.	Accuracy, precision, recall, F1-score, ROC-AUC.	The model achieved 97.2% accuracy, 0.9969 ROC-AUC, and 0.97 for precision, recall, and F1-score, outperforming VGG16, MobileNetV2, ResNet50, and DenseNet121.	With thorough pre-processing and global pooling, the DCNN model was able to attain a higher classification accuracy for skin lesions.	High accuracy, low cost, improved handling of imbalanced datasets via augmentation, and reduced overfitting with global average pooling.	Only the HAM10000 dataset was used to train the model; more training on a variety of datasets is necessary. Black hair and other natural artifacts may still lead to certain incorrect diagnoses.	Training on additional datasets to improve generalization, optimizing hyperparameters, and applying attention mechanisms to enhance the classification of more complex datasets.
[22]	Deep Learning Approach for Automated Data Augmentation and Multi-class Classification of Pap Smear Images (2023).	The cervical cell types from Pap smear images into four categories: NILM, HSIL, LSIL, and SCC.	The dataset contains 963 liquid-based cytology (LBC) images divided into four classes: NILM (613 images), LSIL (163), HSIL (113), and SCC (74).	Utilized GANs for data augmentation and EfficientNet architecture for classification.	Precision, Recall, F1-Score, and Accuracy.	Achieved 99.1% testing accuracy using GAN for data augmentation and EfficientNet for classification, with precision of 99.2% and recall of 99.4%.	Using GANs for synthetic data generation enhanced classification accuracy over traditional data augmentation methods.	GAN-assisted automated data augmentation improved classification by increasing dataset variety.	A concentration on just four cervical cell classifications, a small dataset size, and the possibility of overfitting even with GAN augmentation.	Expand the dataset, include more diverse medical imaging types, and apply the approach to different medical image classification tasks.
[23]	Ensemble Learnin	The masses are into	The dataset contains 101,242 B-	Ensemble learnin	Precision, Sensitivity (Recall),	The WAV (top 7 classifiers)	Ensemble learnin	Enhanced classification	The emphasis was on	Explore ensemble learning for

	g of Multiple Models Using Deep Learning for Multiclass Classification of Ultrasound Images of Hepatic Masses (2023).	four classes: BLT, LCY, MLC, and PLC.	mode ultrasound images of (FLLs) from 30,873 cases, collected between 2018–2021 across 11 medical centers in Japan, with 26,440 images used for training and testing in four categories.	g using 16 different CNNs (including EfficientNet, ResNeXt, Xception, etc.). Techniques include SV, WAV, WHV, and stacking (ST) as ensemble methods.	Specificity, F1 score, and Accuracy. ROC-AUC values were calculated for some ensemble techniques like SV, WAV, and ST.	achieved 78.3% accuracy, surpassing ResNeXt101's 71.9%, and had the highest sensitivity, specificity, and F1 score.	g greatly improves CNN performance for hepatic mass classification with ultrasound images, with WAV as the top performer among seven classifiers.	accuracy using ensemble learning methods. Outperformed individual CNN models. Extensive multicenter dataset ensuring generalizability.	grayscale B-mode images, which might not include all the information needed for contrast-enhanced images or other imaging modalities.	other types of liver masses and investigate applications of ensemble learning with real-time ultrasound diagnostics.
[24]	Multiclass classification by Min–Max ECOC with Hamming distance optimization (2023).	Min–Max ECOC method.	Fashion-MNIST is one of four genuine datasets used for testing.	DNN, Min-Max ECOC matrix.	Accuracy, confusion matrix metrics, and two new indicators were defined in the study.	The suggested approach fared better than the current approaches in terms of error reduction and classification accuracy.	A promising strategy for enhancing ensemble learning in multiclass classification tasks is the Min–Max ECOC method.	The method theoretically optimizes the design of ECOC matrices and demonstrates superior performance on complex datasets.	Computational complexity increases with larger ECOC matrices; optimal values are challenging to calculate for large datasets.	Future studies might look into modifying the Min–Max ECOC approach for use in real-time applications and other machine learning frameworks.

D. Discussion and Comparison

The results of this study revealed notable differences in how various techniques handle multiclass classification with imbalanced data. Resampling techniques, particularly SMOTE, were effective in improving minority class performance but suffered from limitations such as overfitting and loss of valuable majority class data in under-sampling methods.

For jobs where the minority class is crucial, such as medical diagnosis, cost-sensitive learning has proven to be very advantageous. Although it occasionally decreased the overall accuracy of the model, it significantly increased the recall for the minority classes.

Overall, ensemble approaches such as AdaBoost and Random Forest performed well, with AdaBoost being particularly good at increasing minority class recognition, and Random Forest providing the most balanced performance. The capacity of these techniques to concentrate on examples that are challenging to categorize allowed them to routinely outperform conventional classifiers.

The best overall performance was obtained by the deep learning models, specifically CNNs with weighted cross-entropy loss and GAN-based synthetic data creation. These models were perfect for large, complicated datasets because they balanced high accuracy and F1 scores for both majority and minority classes. In contrast to other methods, they require more careful tweaking and have high computational demands.

Comparatively, although simpler techniques such as SMOTE and cost-sensitive learning are effective for smaller datasets or where computational resources are limited, ensemble and deep learning methods are better suited for large-scale, high-dimensional problems, providing superior generalization and handling of imbalanced data.

E. Conclusion

This review discusses the issue of handling classification in imbalanced datasets, in crucial fields like medical diagnosis and fraud detection where minority classes are significant factors to consider. To enhance classification accuracy for imbalanced data sets various approaches including resampling techniques, ensemble models, and cost-sensitive learning have been investigated. While methods like SMOTE are successful in tackling class imbalances, some challenges, like overfitting, may arise as a result. Ensemble techniques, like Random Forest and AdaBoost, consistently enhance classification accuracy by combining classifiers; however, they may need adjustments for performance. Cost-sensitive learning offers a solution by giving priority to minority classes through varying misclassification costs.

Cutting-edge techniques, like combining methods and advanced deep learning models, hold the potential for handling complex and extensive datasets effectively. Nevertheless, these strategies demand fine-tuning. Pose computational challenges. Moreover, using evaluation metrics such as the F1 score and Recall, alongside accuracy measures ensures a more accurate reflection of model performance when dealing with imbalanced data.

Ultimately, no one method can offer a solution for all cases. The strategy is reliant on the specific dataset and context of the application. Future research should focus on enhancing these techniques to develop more versatile, efficient, and widely applicable models for handling imbalanced multiclass classification tasks.

F. Acknowledgment

The authors thank Akre University for Applied Sciences and Duhok Polytechnic University for their support, and Prof. Dr. Adnan Mohsin Abdulazeez for his valuable guidance.

G. References

[1] C. Sujatha and A. Mohan, "Bearing Fault Classification Using Multi-Class Machine Learning (ML) Techniques," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 11, no. 1, 2024, doi: 10.4108/eetsis.3895.

[2] I. Alarab and S. Prakoonwit, "Uncertainty estimation based adversarial attack in multi-class classification," *Multimed Tools Appl*, vol. 82, no. 1, pp. 1519–1536, Jan. 2023, doi: 10.1007/s11042-022-13269-1.

[3] A. Kumar, A. Kaur, P. Singh, M. Driss, and W. Boulila, "Efficient Multiclass Classification Using Feature Selection in High-Dimensional Datasets," *Electronics (Switzerland)*, vol. 12, no. 10, May 2023, doi: 10.3390/electronics12102290.

[4] J. Liu, Y. Tang, H. Zhao, X. Wang, F. Li, and J. Zhang, "CPS Attack Detection under Limited Local Information in Cyber Security: An Ensemble Multi-Node Multi-Class Classification Approach," *ACM Trans Sens Netw*, vol. 20, no. 2, Jan. 2024, doi: 10.1145/3585520.

[5] F. Riaz, S. Abdulla, H. Suzuki, S. Ganguly, R. C. Deo, and S. Hopkins, "Accurate Image Multi-Class Classification Neural Network Model with Quantum Entanglement Approach," *Sensors*, vol. 23, no. 5, Mar. 2023, doi: 10.3390/s23052753.

[6] "Specifications Table", doi: 10.17632/kb7vvkg2th.2.

[7] M. A. Arshed, S. Mumtaz, M. Ibrahim, S. Ahmed, M. Tahir, and M. Shafi, "Multi-Class Skin Cancer Classification Using Vision Transformer Networks and Convolutional Neural Network-Based Pre-Trained Models," *Information (Switzerland)*, vol. 14, no. 7, Jul. 2023, doi: 10.3390/info14070415.

[8] J. Shang, Z. Xiao, T. Tao, J. Wang, and Z. Wu, "A heuristic method for discovering multi-class classification rules from multi-source data in cloud–edge system," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 2, Feb. 2024, doi: 10.1016/j.jksuci.2024.101962.

[9] M. Salauddin Khan *et al.*, "Comparison of multiclass classification techniques using dry bean dataset," *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 6–20, Jun. 2023, doi: 10.1016/j.ijcce.2023.01.002.

[10] S. Ayas, "Multiclass skin lesion classification in dermoscopic images using swin transformer model," *Neural Comput Appl*, vol. 35, no. 9, pp. 6713–6722, Mar. 2023, doi: 10.1007/s00521-022-08053-z.

[11] K. M. Hosny, W. M. El-Hady, F. M. Samy, E. Vrochidou, and G. A. Papakostas, "Multi-Class Classification of Plant Leaf Diseases Using Feature Fusion of Deep Convolutional Neural Network and Local Binary Pattern," *IEEE Access*, vol. 11, pp. 62307–62317, 2023, doi: 10.1109/ACCESS.2023.3286730.

[12] M. A. Khan, M. Sharif, T. Akram, R. Damaševičius, and R. Maskeliūnas, "Skin lesion segmentation and multiclass classification using deep learning

features and improved moth flame optimization," *Diagnostics*, vol. 11, no. 5, 2021, doi: 10.3390/diagnostics11050811.
[13] E. Shushkevich, M. Alexandrov, and J. Cardiff, "Improving Multiclass Classification of Fake News Using BERT-Based Models and ChatGPT-Augmented Data," *Inventions*, vol. 8, no. 5, Oct. 2023, doi: 10.3390/inventions8050112.
[14] W. Rahman, M. G. G. Faruque, K. Roksana, A. H. M. S. Sadi, M. M. Rahman, and M. M. Azad, "Multiclass blood cancer classification using deep CNN with optimized features," *Array*, vol. 18, Jul. 2023, doi: 10.1016/j.array.2023.100292.
[15] F. J. M. Shamrat, S. Azam, A. Karim, K. Ahmed, F. M. Bui, and F. De Boer, "High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images," *Comput Biol Med*, vol. 155, Mar. 2023, doi: 10.1016/j.compbiomed.2023.106646.
[16] A. K. Gupta, S. Chakroborty, S. K. Ghosh, and S. Ganguly, "A machine learning model for multi-class classification of quenched and partitioned steel microstructure type by the k-nearest neighbor algorithm," *Comput Mater Sci*, vol. 228, Sep. 2023, doi: 10.1016/j.commatsci.2023.112321.
[17] E. H. Houssein, D. A. Abdelkareem, G. Hu, M. A. Hameed, I. A. Ibrahim, and M. Younan, "An effective multiclass skin cancer classification approach based on deep convolutional neural network," *Cluster Comput*, Dec. 2024, doi: 10.1007/s10586-024-04540-1.
[18] Q. Su, H. N. A. Hamed, M. A. Isa, X. Hao, and X. Dai, "A GAN-Based Data Augmentation Method for Imbalanced Multi-Class Skin Lesion Classification," *IEEE Access*, vol. 12, pp. 16498–16513, 2024, doi: 10.1109/ACCESS.2024.3360215.
[19] P. del Moral, S. Nowaczyk, A. Sant'Anna, and S. Pashami, "Pitfalls of assessing extracted hierarchies for multi-class classification," *Pattern Recognit*, vol. 136, Apr. 2023, doi: 10.1016/j.patcog.2022.109225.
[20] S. Poppi, S. Sarto, M. Cornia, L. Baraldi, and R. Cucchiara, "Multi-Class Unlearning for Image Classification via Weight Filtering," 2024.
[21] P. V. S. P. Raghavendra, C. Charitha, K. G. Begum, and V. B. S. Prasath, "Deep Learning–Based Skin Lesion Multi-class Classification with Global Average Pooling Improvement," *J Digit Imaging*, vol. 36, no. 5, pp. 2227–2248, Oct. 2023, doi: 10.1007/s10278-023-00862-5.
[22] S. Nayar, D. Priya, and J. Vinitha Panicker, "Deep Learning Approach for Automated Data Augmentation and Multi-class Classification of Pap Smear Images," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 2205–2214. doi: 10.1016/j.procs.2024.04.209.
[23] N. Nakata and T. Siina, "Ensemble Learning of Multiple Models Using Deep Learning for Multiclass Classification of Ultrasound Images of Hepatic Masses," *Bioengineering*, vol. 10, no. 1, Jan. 2023, doi: 10.3390/bioengineering10010069.
[24] G. Szűcs, "Multiclass classification by Min–Max ECOC with Hamming distance optimization," *Visual Computer*, vol. 39, no. 9, pp. 3949–3961, Sep. 2023, doi: 10.1007/s00371-022-02540-z.

[25] S. Szeghalmy and A. Fazekas, “A comparative study on noise filtering of imbalanced data sets,” *Knowl Based Syst*, vol. 301, Oct. 2024, doi: 10.1016/j.knosys.2024.112236.
[26] A. Galán-Cuenca, A. J. Gallego, M. Saval-Calvo, and A. Pertusa, “Few-shot learning for COVID-19 chest X-ray classification with imbalanced data: an inter vs. intra domain study,” *Pattern Analysis and Applications*, vol. 27, no. 3, Sep. 2024, doi: 10.1007/s10044-024-01285-w.
[27] E. F. Swana, W. Doorsamy, and P. Bokoro, “Tomek Link and SMOTE Approaches for Machine Fault Classification with an Imbalanced Dataset,” *Sensors*, vol. 22, no. 9, May 2022, doi: 10.3390/s22093246.
[28] W. Chandra, B. Suprihatin, and Y. Resti, “Median-KNN Regressor-SMOTE-Tomek Links for Handling Missing and Imbalanced Data in Air Quality Prediction,” *Symmetry (Basel)*, vol. 15, no. 4, Apr. 2023, doi: 10.3390/sym15040887.
[29] S. Gholampour, “Impact of Nature of Medical Data on Machine and Deep Learning for Imbalanced Datasets: Clinical Validity of SMOTE Is Questionable,” *Mach Learn Knowl Extr*, vol. 6, no. 2, pp. 827–841, Jun. 2024, doi: 10.3390/make6020039.
[30] T. M. Alam *et al.*, “An Efficient Deep Learning-Based Skin Cancer Classifier for an Imbalanced Dataset,” *Diagnostics*, vol. 12, no. 9, Sep. 2022, doi: 10.3390/diagnostics12092115.
[31] M. Ahmad, M. Abdullah, H. Moon, and D. Han, “Plant Disease Detection in Imbalanced Datasets Using Efficient Convolutional Neural Networks with Stepwise Transfer Learning,” *IEEE Access*, vol. 9, pp. 140565–140580, 2021, doi: 10.1109/ACCESS.2021.3119655.
[32] G. Sambasivam and G. D. Opiyo, “A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks,” *Egyptian Informatics Journal*, vol. 22, no. 1, pp. 27–34, Mar. 2021, doi: 10.1016/j.eij.2020.02.007.
[33] A. M. Sowjanya and O. Mrudula, “Effective treatment of imbalanced datasets in health care using modified SMOTE coupled with stacked deep learning algorithms,” *Applied Nanoscience (Switzerland)*, vol. 13, no. 3, pp. 1829–1840, Mar. 2023, doi: 10.1007/s13204-021-02063-4.
[34] S. Gholampour, “Impact of Nature of Medical Data on Machine and Deep Learning for Imbalanced Datasets: Clinical Validity of SMOTE Is Questionable,” *Mach Learn Knowl Extr*, vol. 6, no. 2, pp. 827–841, Jun. 2024, doi: 10.3390/make6020039.
[35] Q. Su, H. N. A. Hamed, M. A. Isa, X. Hao, and X. Dai, “A GAN-Based Data Augmentation Method for Imbalanced Multi-Class Skin Lesion Classification,” *IEEE Access*, vol. 12, pp. 16498–16513, 2024, doi: 10.1109/ACCESS.2024.3360215.
[36] W. Shafqat and Y. C. Byun, “A Hybrid GAN-Based Approach to Solve Imbalanced Data Problem in Recommendation Systems,” *IEEE Access*, vol. 10, pp. 11036–11047, 2022, doi: 10.1109/ACCESS.2022.3141776.
[37] J. Tanha, Y. Abdi, N. Samadi, N. Razzaghi, and M. Asadpour, “Boosting methods for multi-class imbalanced data classification: an experimental review,” *J Big Data*, vol. 7, no. 1, Dec. 2020, doi: 10.1186/s40537-020-00349-y.

[38] J. Tanha, Y. Abdi, N. Samadi, N. Razzaghi, and M. Asadpour, "Boosting methods for multi-class imbalanced data classification: an experimental review," *J Big Data*, vol. 7, no. 1, Dec. 2020, doi: 10.1186/s40537-020-00349-y.
[39] M. Khushi *et al.*, "A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data," *IEEE Access*, vol. 9, pp. 109960–109975, 2021, doi: 10.1109/ACCESS.2021.3102399.
[40] S. V. Amanoul, A. M. Abdulazeez, D. Q. Zeebare, and F. Y. H. Ahmed, "Intrusion Detection Systems Based on Machine Learning Algorithms," in *2021 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2021 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., Jun. 2021, pp. 282–287. doi: 10.1109/I2CACIS52118.2021.9495897.
[41] H. T. Sadeeq, S. Y. Ameen, and A. M. Abdulazeez, "Cancer Diagnosis based on Artificial Intelligence, Machine Learning, and Deep Learning," in *2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, IEEE, Nov. 2022, pp. 656–661. doi: 10.1109/3ICT56508.2022.9990784.
[42] A. Arafa, N. El-Fishawy, M. Badawy, and M. Radad, "RN-SMOTE: Reduced Noise SMOTE based on DBSCAN for enhancing imbalanced data classification," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 8, pp. 5059–5074, Sep. 2022, doi: 10.1016/j.jksuci.2022.06.005.
[43] A. F. Jahwar and A. Mohsin Abdulazeez, "Segmentation and Classification for Breast Cancer Ultrasound Images Using Deep Learning Techniques: A Review," in *2022 IEEE 18th International Colloquium on Signal Processing & Applications (CSPA)*, IEEE, May 2022, pp. 225–230. doi: 10.1109/CSPA55076.2022.9781824.
[44] S. Bagui and K. Li, "Resampling imbalanced data for network intrusion detection datasets," *J Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-020-00390-x.
[45] D. Dablain, K. N. Jacobson, C. Bellinger, M. Roberts, and N. V. Chawla, "Understanding CNN fragility when learning with imbalanced data," *Mach Learn*, vol. 113, no. 7, pp. 4785–4810, Jul. 2024, doi: 10.1007/s10994-023-06326-9.
[46] A. S. Qureshi and T. Roos, "Transfer Learning with Ensembles of Deep Neural Networks for Skin Cancer Detection in Imbalanced Data Sets," *Neural Process Lett*, vol. 55, no. 4, pp. 4461–4479, Aug. 2023, doi: 10.1007/s11063-022-11049-4.
[47] D. Dablain, B. Krawczyk, and N. V. Chawla, "DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data," *IEEE Trans Neural Netw Learn Syst*, vol. 34, no. 9, pp. 6390–6404, Sep. 2023, doi: 10.1109/TNNLS.2021.3136503.
[48] T. Wongvorachan, S. He, and O. Bulut, "A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining," *Information (Switzerland)*, vol. 14, no. 1, Jan. 2023, doi: 10.3390/info14010054.

[49] M. Nahiduzzaman *et al.*, “Parallel CNN-ELM: A multiclass classification of chest X-ray images to identify seventeen lung diseases including COVID-19,” *Expert Syst Appl*, vol. 229, Nov. 2023, doi: 10.1016/j.eswa.2023.120528.