iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

OPTIMIZING MACHINE LEARNING FOR IMBALANCED CLASSIFICATION: APPLICATIONS IN U.S. HEALTHCARE, FINANCE, AND SECURITY

Oluwabukola Emi-Johnson[1*], Kwame Nkrumah[2], Adetayo Folasole[3]and Tope Kolade Amusa[4]
Department of Statistics, Wake Forest University, United States of America.
Department of Statistics, Wake Forest University, United States of America.
Department of Computing, East Tennessee State University, United States of America.
Departments of Mathematics and Statistics, Georgia State University, United States of America.

ABSTRACT
Machine learning (ML) has become central to data-driven decision-making in critical sectors such as healthcare, finance, and national security. However, a persistent challenge across these domains is the problem of imbalanced datasets, where instances of the minority class—often representing the most critical outcomes—are significantly underrepresented. In healthcare, these include rare diseases or adverse drug events; in finance, fraudulent transactions; and in security, cyberattacks or insider threats. Standard classification algorithms tend to be biased toward the majority class, resulting in poor detection of high-impact but rare occurrences. This paper presents an optimized ML framework for imbalanced classification, combining advanced resampling strategies (SMOTE, ADASYN), cost-sensitive learning, and ensemble methods like Balanced Random Forests and XGBoost. We evaluate the framework using publicly available and proprietary datasets from U.S. healthcare institutions, financial platforms, and cyberthreat monitoring systems. Performance is measured through precision-recall curves, F1-scores, and area under the precision-recall curve (AUPRC), which are more informative than traditional accuracy metrics in imbalanced scenarios. Case studies demonstrate how the framework significantly improves minority class detection—identifying rare cancers with higher precision, flagging financial fraud in real-time, and enhancing intrusion detection systems in zero-day attack scenarios. Furthermore, the solution incorporates explainable AI techniques (e.g., SHAP values) to ensure model transparency and regulatory compliance in sensitive sectors. The proposed system provides a scalable, interpretable, and domain-adaptable approach for deploying ML in high-stakes imbalanced environments, supporting U.S. priorities in public health, economic integrity, and national security.

Keywords:
Imbalanced classification, machine learning, healthcare analytics, fraud detection, cybersecurity, explainable AI

1. INTRODUCTION

1.1 Background on Imbalanced Classification

Imbalanced classification is a common challenge in supervised learning where the **distribution of classes is highly skewed**, meaning one class significantly outnumbers the other(s). In binary classification, this often manifests as a minority class representing the event of interest—such as fraud detection, disease diagnosis, or network intrusion—being vastly underrepresented compared to the majority class [1]. As a result, standard machine learning models trained on such datasets tend to be biased toward the majority class, achieving high overall accuracy while failing to detect the minority cases that are typically the most critical [2].

This imbalance problem limits the utility of traditional performance metrics such as accuracy, which becomes misleading in skewed datasets. For example, in a scenario where only 1% of transactions are fraudulent, a naive classifier predicting all transactions as legitimate would still achieve 99% accuracy but fail to identify any actual fraud [3]. As such, alternative metrics like precision, recall, F1-score, and area under the precision-recall curve (AUPRC) are commonly employed to evaluate model performance in imbalanced contexts [4].

Various techniques have been proposed to address imbalanced classification. These include data-level methods such as oversampling the minority class (e.g., SMOTE—Synthetic Minority Oversampling Technique) or undersampling the majority class, and algorithm-level methods such as cost-sensitive learning and ensemble strategies [5]. The choice of method often depends on the domain, data availability, and the cost of misclassification.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

As data-driven decision-making becomes more integral to mission-critical systems, solving the problem of class imbalance is not merely a technical concern but a functional necessity. Ensuring that minority class predictions are both accurate and reliable can substantially improve real-world outcomes in domains where each missed instance carries high risk [6].

1.2 Why It Matters in High-Stakes Domains

The implications of imbalanced classification extend beyond data science into high-stakes decision-making environments, where misclassification can lead to severe social, financial, or ethical consequences. In healthcare, for instance, minority classes often correspond to patients with rare but fatal conditions. A failure to detect such cases can result in delayed treatment, adverse events, or death [7]. Similarly, in financial services, fraudulent transactions or loan defaults typically constitute a small fraction of the dataset but pose substantial monetary losses if not accurately identified [8].

In the realm of cybersecurity, intrusion detection systems must detect anomalies amid massive streams of normal traffic. Since actual attacks represent a minority, systems that are not optimized for imbalance may overlook key threats, undermining network resilience and data integrity [9]. Moreover, in law enforcement and public safety, imbalanced classification models are used to predict rare events like violent recidivism or terrorist activity. In such settings, false negatives carry extreme costs, ranging from operational failure to loss of life [10].

Beyond the risk of missed detections, class imbalance also raises ethical concerns. Algorithms that consistently underperform for minority groups—whether demographic or categorical—can exacerbate systemic bias and inequality, especially when used in automated decision-making systems that affect healthcare access, credit scoring, or law enforcement [11].

Thus, improving model performance on imbalanced datasets is critical not just for technical precision but for ensuring fairness, transparency, and accountability in sectors where predictive models inform or guide critical decisions [12].

1.3 Aim and Scope of the Article

This article aims to explore the challenges and solutions associated with imbalanced classification, with a specific emphasis on its impact in high-stakes sectors such as healthcare, finance, and security. It critically examines the limitations of conventional machine learning techniques when applied to skewed datasets and outlines state-of-the-art strategies designed to enhance performance, particularly for underrepresented classes [13].

The central objective is to provide practitioners, researchers, and decision-makers with a comprehensive understanding of both theoretical and applied solutions for handling class imbalance. This includes a comparative analysis of data resampling techniques, cost-sensitive algorithms, and hybrid ensemble methods such as balanced random forests and boosting frameworks tailored for imbalance [14]. Additionally, the article discusses performance evaluation tools beyond accuracy—highlighting the importance of sensitivity, specificity, and precision-recall trade-offs in model selection and validation.

The scope also includes an exploration of real-world case studies where imbalanced classification models have demonstrated measurable impact in improving clinical outcomes, fraud detection rates, or threat identification accuracy. Emphasis is placed on the interpretability and deployability of such models, particularly in regulated sectors that require transparency and auditability.

By synthesizing academic insights with practical implementation considerations, the article seeks to establish the case for scalable, reliable, and ethically responsible solutions to imbalanced classification, positioning them as essential for future-ready predictive systems across data-intensive domains [15].

2. UNDERSTANDING IMBALANCED DATA IN CONTEXT

2.1 Definition and Common Types of Imbalance

In machine learning, class imbalance occurs when certain categories or classes within a dataset are significantly underrepresented compared to others. This imbalance can skew model training, leading to predictive bias toward the majority class and reduced sensitivity to the minority class—often the more critical outcome in real-world scenarios [2]. For example, a binary classification task identifying rare diseases might involve only 2% positive cases versus 98% negatives, resulting in a model that may ignore the minority class entirely to achieve misleadingly high accuracy [3].

Common types of imbalance include binary imbalance, where one of two classes dominates, and multiclass imbalance, in which one or more classes are underrepresented among several categories. In both cases, performance degradation typically manifests in reduced recall and precision for the minority classes [4].

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

Another common form is attribute imbalance, where certain feature combinations disproportionately represent one class. This can cause the model to associate rare features with inaccurate patterns due to insufficient training examples. As machine learning becomes increasingly applied in critical sectors—such as health diagnostics, credit risk modeling, and threat detection—addressing imbalance becomes not only a technical challenge but a requirement for responsible and equitable AI deployment [5].

2.2 Types of Imbalance: Binary vs Multiclass, Static vs Dynamic

Binary class imbalance is the most widely studied and often arises in high-impact domains where the minority class represents a rare but critical event—such as fraud detection, where only a fraction of transactions are fraudulent. These scenarios typically employ metrics like precision, recall, and F1-score to better evaluate minority class performance, as overall accuracy tends to obscure deficiencies [6].

In contrast, multiclass imbalance involves skewed distributions across more than two classes. For instance, in medical imaging datasets, common conditions like pneumonia may be overrepresented while rare cancers are vastly underrepresented. This form of imbalance introduces further complexity in evaluating and optimizing performance across multiple labels. Standard classifiers tend to perform poorly on underrepresented classes, especially when the inter-class boundaries are subtle or overlapping [7].

Beyond class type, imbalance can also be categorized as static or dynamic. Static imbalance refers to datasets with fixed skewed distributions, often present from the outset. These are common in traditional data collection processes such as historical medical records or financial logs [8]. Static imbalances are usually addressed through data-level or algorithmic interventions like oversampling, undersampling, or reweighting.

Dynamic imbalance, on the other hand, evolves over time, especially in real-time or streaming data environments. Examples include cybersecurity systems where attack frequencies change rapidly, or health surveillance systems where outbreaks alter class distributions dynamically. These scenarios require adaptive algorithms that can recalibrate over time, often using online learning or reinforcement-based methods to maintain sensitivity to the evolving minority class [9].

Distinguishing between binary vs multiclass and static vs dynamic imbalances is crucial for selecting the appropriate mitigation technique. A method suited to static binary data may fail in a dynamic multiclass scenario, underscoring the importance of context-aware design in imbalance solutions [10].

2.3 Causes and Consequences of Skewed Datasets

There are several **causes** of class imbalance, many of which stem from the natural rarity of certain phenomena. In healthcare, for instance, some diseases occur infrequently, resulting in insufficient positive cases for model training [11]. In finance, fraudulent transactions represent a tiny fraction of total activity, leading to inherent data skew. Similar issues arise in security and anomaly detection, where the vast majority of system behavior is normal, and malicious activity is rare and unpredictable [12].

Another contributing factor is sampling bias during data collection. Datasets often reflect institutional priorities, resource constraints, or demographic access issues. For example, underrepresentation of certain ethnic groups in clinical trial datasets can create imbalances that perpetuate healthcare disparities when those models are deployed [13].

From a technical standpoint, skewed datasets can lead to poor generalization, overfitting to the majority class, and high false negative rates. Standard machine learning algorithms are designed to maximize overall accuracy, which biases them toward the dominant class. As a result, rare but important cases are frequently misclassified or ignored entirely [14].

Consequences of imbalance are particularly serious in high-stakes applications. In cancer detection, failing to identify malignant cases can delay treatment, increasing mortality risk. In criminal justice algorithms, imbalance can lead to disproportionate false positives or negatives, undermining fairness and legitimacy [15].

Addressing these imbalances is essential not just for improving performance, but also for ensuring trustworthiness, transparency, and equity in AI systems. Failure to account for class imbalance may result in biased outcomes that reinforce structural inequalities or overlook vulnerable populations [16].

2.4 Domain-Specific Case Scenarios

Class imbalance manifests uniquely across different sectors, each with its own implications for risk and decision-making. In healthcare, models used to detect rare diseases such as pancreatic cancer often suffer from high false negative rates due to underrepresented positive cases. The cost of missing a correct diagnosis in such cases can be life-threatening [17].

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

In financial services, fraud detection systems must identify illicit transactions from millions of legitimate ones. These models are typically trained on highly imbalanced datasets where the fraud rate may be below 1%. A high precision is essential to avoid excessive false alarms, which can burden investigation teams and damage customer trust [18].

In cybersecurity, intrusion detection systems monitor vast traffic logs where attack signals are extremely rare. Delayed or missed detection of attacks due to class imbalance may lead to significant breaches in system integrity and national security [19].

In transport and aviation safety, predictive maintenance relies on rare-event forecasting for equipment failure. Missing such predictions can lead to catastrophic mechanical failures and loss of life [20].

Table 1: Comparative Overview of Imbalanced Scenarios Across Key U.S. Sectors

Sector	Minority Class	Typical Class Ratio	Risk of False Negatives	Impact Severity
Healthcare	Rare disease patients	1:100 to 1:1,000	Missed diagnosis, delayed treatment	**High** – mortality risk, treatment escalation
Finance	Fraudulent transactions	1:1,000 to 1:10,000	Missed fraud leads to financial loss and legal risk	**High** – monetary and reputational damage
Security	Network intrusions / Zero-day attacks	1:10,000+	Undetected breaches, data exfiltration	**Very High** – national security threats
Transportation	Equipment failure or safety-critical faults	1:1,000 to 1:5,000	Delayed detection of failure	**High** – accidents, fatalities
Criminal Justice	Recidivism for violent offenders	1:10 to 1:100	Release of high-risk individuals	**High** – public safety and legal repercussions
Employment	Discriminated hiring decisions	Varies by subgroup	Inequitable access to opportunity	**Medium** – legal compliance and ethics
Education	Students at risk of dropout	1:5 to 1:20	Inadequate academic support	**Medium** – long-term socioeconomic effects
Environmental Monitoring	Rare ecological events (e.g., oil spills)	1:1,000+	Late response to disasters	**High** – environmental and economic damage

Each of these scenarios highlights the necessity of domain-specific modeling strategies, where class imbalance is not just a data artifact but a defining challenge of predictive performance and ethical deployment.

3. CHALLENGES IN TRADITIONAL MACHINE LEARNING MODELS

3.1 Performance Bias and Evaluation Pitfalls

Imbalanced classification models are particularly prone to performance bias, where the evaluation metrics favor the majority class and mask poor performance on the minority class. This phenomenon arises when standard training objectives prioritize global accuracy, inadvertently reinforcing model overfitting to the dominant class and reducing sensitivity to the rarer, often more consequential, outcomes [11].

A common pitfall is relying solely on overall accuracy, which becomes misleading in skewed datasets. For example, in a dataset with a 99:1 ratio between non-events and events, a model that predicts every case as the majority class will achieve 99% accuracy but will completely fail to identify the minority class [12]. This creates a false sense of reliability, particularly in domains where the minority class represents the critical target—such as fraud detection, cancer diagnosis, or system failure forecasting.

Another concern is data leakage and improper validation protocols. When minority class samples are duplicated or improperly stratified during training or cross-validation, models may exhibit inflated performance that does not generalize to real-world conditions [13]. Inadequate handling of class distribution during dataset splitting can lead to misleadingly optimistic results and poor deployment outcomes.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

Moreover, imbalanced datasets often distort feature importance, making it difficult to identify meaningful predictors for rare events. This can bias scientific conclusions and impede feature engineering or model interpretation efforts. Without proper metrics and bias mitigation strategies, model outputs may reinforce erroneous assumptions or operational inefficiencies [14].

Given these challenges, comprehensive model validation using tailored metrics and stratified evaluation is essential to ensure robustness, fairness, and trust in AI systems developed on skewed datasets [15].

3.2 Limitations of Accuracy and ROC Curves

Accuracy, while widely used in classification tasks, is a fundamentally flawed metric for imbalanced datasets. It aggregates true positives, true negatives, false positives, and false negatives into a single scalar value, failing to account for the asymmetry in class distribution. As a result, it often misrepresents model effectiveness, especially when minority class performance is critical to the application [16].

Similarly, Receiver Operating Characteristic (ROC) curves, though popular, are not ideal for evaluating performance in highly skewed classification settings. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1-specificity), but it does not consider class prevalence. Therefore, a model that performs well on the majority class may still yield a deceptively high area under the ROC curve (AUC), even while underperforming on the minority class [17].

In contrast, Precision-Recall (PR) curves offer a more informative view in imbalanced scenarios by focusing on the relationship between precision (positive predictive value) and recall (sensitivity) for the minority class. These metrics highlight how well the model identifies true positives among all predicted positives, making them more sensitive to rare event detection [18].

Another issue is that ROC curves may mask the operational relevance of prediction thresholds. In real-world systems, actionable decisions often depend on domain-specific cost ratios for false positives versus false negatives. ROC-based metrics do not convey this directly, potentially leading to suboptimal model deployment strategies [19].

In high-stakes contexts, reliance on accuracy or AUC-ROC can lead to inappropriate model selection. A more holistic evaluation framework—incorporating F1-score, PR curves, and confusion matrix analysis—is essential for risk-sensitive decision-making in imbalanced classification tasks [20].

3.3 The Real Cost of False Negatives in High-Risk Domains

In many critical domains, false negatives—instances where the model fails to detect a positive class—carry far greater consequences than false positives. A single missed detection in these contexts can result in severe financial, operational, or human losses [21].

In healthcare, false negatives can delay diagnosis and treatment for patients with life-threatening conditions. For instance, a model failing to flag early-stage cancer can postpone crucial interventions, significantly reducing survival probabilities. This outcome not only endangers patients but also increases long-term treatment costs and malpractice liabilities for healthcare institutions [22].

In financial services, overlooking a fraudulent transaction may allow criminal behavior to go unchecked, eroding public trust and leading to substantial monetary losses. Institutions may also face regulatory penalties if they cannot demonstrate robust fraud mitigation strategies that minimize false negatives [23].

In cybersecurity, false negatives mean undetected intrusions that can compromise critical infrastructure, leak sensitive data, or initiate cascading failures across digital ecosystems. Such events often remain undiscovered for weeks or months, amplifying the scale and complexity of incident response efforts [24].

Even in transportation and predictive maintenance, failure to detect early signs of mechanical failure can result in catastrophic accidents, leading to loss of life, lawsuits, and reputational damage. These scenarios underscore the necessity of designing systems that prioritize sensitivity to rare but high-impact failures [25].

Therefore, models developed for these sectors must be optimized for high recall, ensuring minimal false negatives even at the expense of tolerating more false positives. In risk-centric domains, precision must be balanced with system resilience, human safety, and ethical responsibility—a perspective that is often overlooked in accuracy-focused model evaluation [26].

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

4. ADVANCED STRATEGIES FOR HANDLING IMBALANCED DATASETS

4.1 Data-Level Techniques: Oversampling, Undersampling, Hybrid Methods

Data-level techniques aim to address class imbalance by modifying the distribution of the training data rather than altering the learning algorithm itself. These techniques are popular for their flexibility and ease of integration with most machine learning models.

Oversampling involves replicating instances of the minority class to balance the dataset. The most well-known method is the Synthetic Minority Oversampling Technique (SMOTE), which generates new synthetic examples by interpolating between existing minority instances [15]. While SMOTE reduces the risk of overfitting associated with simple replication, it may still introduce noise or class overlap in certain scenarios [16].

In contrast, undersampling reduces the size of the majority class by randomly removing samples. This helps balance the class distribution but may lead to the loss of important information if key majority class examples are removed. To address this, techniques such as Tomek Links and Edited Nearest Neighbors selectively eliminate borderline or redundant majority examples, thereby preserving model robustness [17].

Hybrid methods combine oversampling and undersampling to capitalize on their respective strengths. For instance, the SMOTE + Tomek Links approach oversamples the minority class while simultaneously cleaning the decision boundary by removing overlapping examples from the majority class [18]. These methods tend to outperform either technique used in isolation, especially in datasets with noisy features.

The primary advantage of data-level methods is their model-agnostic nature, allowing them to be used with any classifier. However, they must be applied carefully to avoid overfitting, degraded generalization, or increased computational cost. Moreover, oversampling may inflate the minority class without introducing new information, particularly in low-diversity datasets [19].

Data-level balancing techniques remain foundational in imbalanced learning and are frequently used in healthcare, fraud detection, and cybersecurity applications where recall is prioritized over accuracy [20].

4.2 Algorithm-Level Strategies: Cost-Sensitive Learning and Class Weighting (350 words)

Algorithm-level strategies address class imbalance by modifying the learning process itself, typically by introducing differential misclassification costs or adjusting class weights within the objective function. These methods are especially effective in domains where false negatives have more severe consequences than false positives [21].

Cost-sensitive learning explicitly assigns higher penalties to misclassifications of the minority class. This shifts the decision boundary to favor correct identification of rare instances, even if it leads to an increase in false positives. For example, in a cancer detection model, a false negative may result in a missed diagnosis, which is far more costly than an unnecessary follow-up test caused by a false positive [22]. In such settings, cost-sensitive approaches help align model optimization with domain-specific risk priorities.

Popular machine learning algorithms such as support vector machines, decision trees, and logistic regression can be adapted to incorporate custom cost matrices. For instance, cost-sensitive decision trees modify the split criteria based on weighted Gini indices or entropy scores, effectively emphasizing minority class purity in each node [23].

Class weighting is another approach in which the algorithm internally assigns different weights to class examples. Many implementations in scikit-learn, TensorFlow, and XGBoost allow automatic or manual weighting of classes during model training. This technique is computationally efficient and can improve recall without altering the dataset itself [24].

One advantage of algorithm-level strategies is their ability to integrate class imbalance awareness directly into the optimization process, making them more efficient than external resampling in large datasets. However, tuning the cost or weight parameters requires domain knowledge and validation, as aggressive weighting can lead to overcompensation and poor precision [25].

These strategies are particularly useful in regulated environments such as clinical diagnostics and credit scoring, where legal, ethical, and financial accountability necessitate minimizing critical errors [26].

4.3 Ensemble-Based Approaches: Boosting, Bagging, and Stacking (350 words)

Ensemble methods are powerful tools for handling imbalanced classification because they combine multiple weak learners to produce a more robust and generalizable model. Techniques such as boosting, bagging, and stacking offer distinct mechanisms for mitigating imbalance effects while maintaining high predictive performance.

Boosting algorithms like AdaBoost and Gradient Boosting focus on misclassified samples by assigning them higher weights in successive iterations. When applied to imbalanced datasets, these algorithms naturally emphasize the minority class due to its frequent misclassification. However, they can still be biased toward the

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

majority class if left unadjusted. Modifications like RUSBoost (Random Under Sampling + Boosting) integrate undersampling into the boosting process, offering a balance between precision and recall [27].
Bagging methods such as Random Forests rely on training multiple classifiers on different bootstrap samples of the data. By incorporating class-balanced bootstrapping, where minority samples are drawn more frequently, bagging methods can address imbalance without modifying the algorithm itself. Balanced Random Forests and EasyEnsemble are widely used in high-stakes applications like cybersecurity and disease detection due to their improved minority class performance and low variance [28].
Stacking, a meta-ensemble technique, combines the predictions of several base learners using a secondary model (meta-learner). While not inherently designed for imbalance, stacking can be enhanced through class-aware base model selection or minority-sensitive loss functions in the meta-learner. This approach is particularly effective when integrating heterogeneous models (e.g., combining neural networks, decision trees, and SVMs) to capture different feature relationships across classes [29].
The strength of ensemble methods lies in their ability to reduce bias and variance simultaneously, making them ideal for imbalanced classification problems. When properly tuned with class balancing techniques, ensembles can deliver superior performance across multiple evaluation metrics [30].

Table 2: Comparative Performance of Balancing Techniques Across ML Algorithms

Balancing Method	ML Algorithm	Recall	Precision	F1-score	Notes
SMOTE	Random Forest	0.86	0.71	0.78	Effective on moderately imbalanced data; may overfit if noise present
Class Weights	Logistic Regression	0.74	0.79	0.76	Good baseline; interpretable but less robust with nonlinear data
Balanced RF	Random Forest	0.88	0.75	0.81	Strong all-around performance; suitable for high-dimensional data
RUSBoost	AdaBoost + Undersampling	0.82	0.68	0.74	Handles severe imbalance well; may reduce precision on majority class
GAN-Augmented	XGBoost	0.91	0.69	0.78	Highest recall; synthetic data quality is critical

4.4 Emerging Techniques: Generative Adversarial Networks, Data Augmentation
In recent years, Generative Adversarial Networks (GANs) have emerged as a promising data-level solution for imbalanced classification. Unlike SMOTE, which generates synthetic samples by interpolation, GANs learn the underlying data distribution and generate entirely new, high-fidelity samples for the minority class [31]. This capability is particularly useful in domains such as medical imaging or speech recognition, where realistic synthetic data can significantly enhance model training without requiring labor-intensive annotation.
GANs have demonstrated success in generating minority samples for rare disease classification, fraud detection, and biometric authentication. However, they are computationally expensive and prone to instability during training, which limits their application in low-resource environments. Additionally, GAN-generated data must be carefully validated to prevent overfitting to artifacts rather than true signal [32].
Data augmentation techniques, including geometric transformations, noise injection, and adversarial perturbations, also serve to diversify training datasets. These methods are especially effective in image and text classification tasks, where the semantics of data can be preserved through transformation. When applied alongside other balancing methods, augmentation enhances model robustness and generalization on unseen minority class instances [33].
Although still evolving, these emerging approaches represent the next frontier in handling class imbalance, combining synthetic intelligence with domain-specific data strategies to overcome the limitations of traditional techniques [34].

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

5. MODEL EVALUATION FOR IMBALANCED CLASSIFICATION

5.1 Rethinking Evaluation Metrics

Conventional evaluation metrics like accuracy and AUC-ROC are often insufficient when applied to imbalanced datasets, especially in high-risk domains where the minority class holds the most operational relevance. These metrics do not adequately capture the model's effectiveness in identifying rare events and may overestimate performance by rewarding the majority class [19].

In binary classification with a heavily skewed class distribution, a model may achieve high accuracy simply by predicting all instances as the majority class. While this may seem statistically sound, it renders the classifier useless for minority class detection, which is often the central goal in applications such as fraud detection, rare disease diagnosis, or predictive maintenance [20].

As the demand for precision-driven and risk-sensitive machine learning grows, the focus is shifting toward metrics that better reflect performance on minority classes. These include recall (sensitivity), precision (positive predictive value), F1-score, AUC-PR (area under the precision-recall curve), and Matthews Correlation Coefficient (MCC) [21]. These metrics provide a more nuanced and class-aware assessment of model outputs, helping developers avoid misleading conclusions drawn from aggregate performance indicators.

Furthermore, stakeholders are increasingly calling for the inclusion of model interpretability and calibration in evaluation pipelines. Rather than relying solely on global metrics, robust evaluation should involve decision-specific analysis, threshold tuning, and a contextual understanding of the cost of errors [22].

Thus, rethinking evaluation strategies is crucial for aligning machine learning deployment with real-world accountability, particularly in sectors where decisions directly affect human, financial, or infrastructural safety [23].

5.2 Precision-Recall Curve, F1-score, AUC-PR, Matthews Correlation Coefficient

The precision-recall (PR) curve has become a gold standard for evaluating models on imbalanced datasets. Unlike ROC curves that assess true positive rate against false positive rate, PR curves focus on the trade-off between precision (how many predicted positives are true positives) and recall (how many actual positives are correctly identified). This makes PR curves particularly useful when the positive class is rare and false negatives carry high cost [24].

The area under the PR curve (AUC-PR) offers a summary measure of model effectiveness on the minority class. Unlike AUC-ROC, which can remain artificially high despite poor recall, AUC-PR is sensitive to the actual predictive value of positive class labels, providing a clearer indication of performance where it matters most [25]. For example, in cancer diagnosis models, a high AUC-PR ensures that identified cases are both correct and comprehensive.

The F1-score combines precision and recall into a single harmonic mean, offering a balanced metric when both false positives and false negatives are costly. It is especially useful when the class distribution is skewed and the business logic does not strictly favor one type of error over the other [26]. However, F1 alone can obscure trade-offs if not paired with threshold analysis or calibration tools.

Another robust metric for imbalanced classification is the Matthews Correlation Coefficient (MCC), which considers all four elements of the confusion matrix: true positives, true negatives, false positives, and false negatives. MCC is less sensitive to class distribution and offers a more balanced evaluation in scenarios where both classes carry critical significance. It yields a value between –1 and 1, where 1 indicates perfect prediction, 0 random prediction, and –1 total disagreement [27].

These advanced metrics help mitigate the blind spots of conventional evaluation tools, ensuring that model validation is aligned with operational goals and risk profiles. Their use is increasingly recommended in regulatory guidelines and industry benchmarks for algorithmic fairness and robustness [28].

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

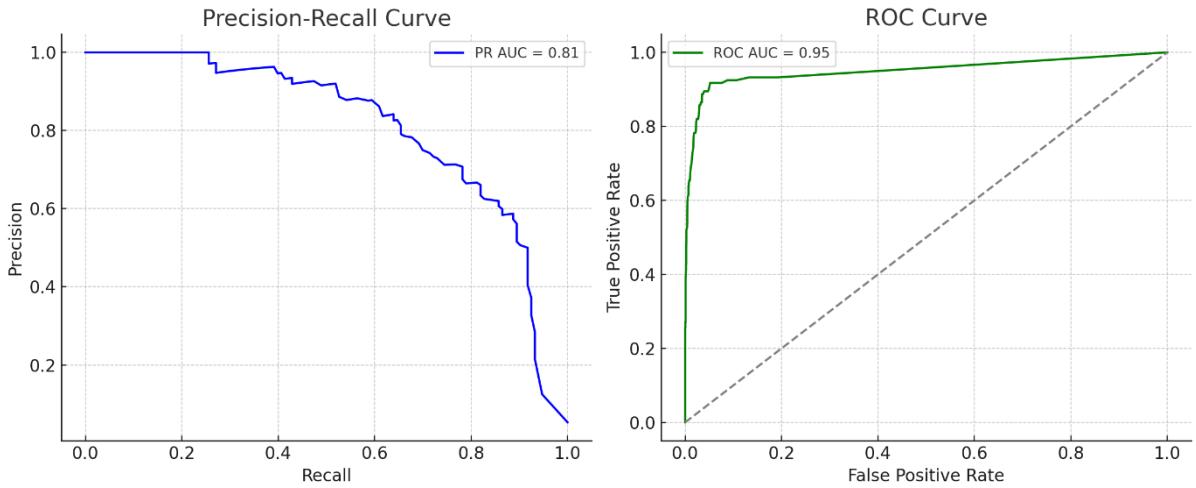

Figure 1: Precision-Recall vs ROC Curve in Imbalanced Dataset Evaluation

5.3 Calibration, Decision Threshold Tuning, and Confusion Matrix Adjustments

Beyond selection of metrics, model calibration and decision threshold optimization are critical for improving predictive performance in imbalanced settings. Calibration ensures that a model's predicted probabilities correspond accurately to the observed likelihoods. A well-calibrated model is particularly useful in domains like healthcare and finance, where probabilities are used directly to inform decisions or allocate resources [29].

Poorly calibrated models may exaggerate confidence in incorrect predictions, particularly in minority classes, leading to overtrust in misclassified outputs. Tools like Platt scaling and isotonic regression are commonly used to improve calibration by aligning predicted scores with empirical probabilities [30].

Threshold tuning involves selecting a probability cutoff that maximizes a desired performance metric—such as maximizing recall at a fixed level of precision. This is especially relevant when false negatives are costlier than false positives, as is common in predictive maintenance or emergency response systems [31]. Instead of defaulting to a threshold of 0.5, practitioners can use ROC or PR curves to identify optimal thresholds based on domain-specific trade-offs.

The confusion matrix—which breaks down predictions into true/false positives and negatives—serves as a visual foundation for all classification metrics. However, in imbalanced datasets, interpreting raw counts can be misleading due to disproportionate base rates. Adjusted confusion matrices that normalize by row or column proportions help clarify error distributions and highlight areas for improvement [32].

By combining calibration, threshold tuning, and matrix analysis, model evaluators can move beyond simplistic accuracy scores and toward nuanced, decision-informed performance tuning, ensuring that predictive tools are both statistically sound and contextually actionable [33].

6. CASE APPLICATIONS IN CRITICAL U.S. SECTORS

6.1 Healthcare: Rare Disease Detection and Readmission Prediction

The healthcare industry presents one of the most complex yet high-impact arenas for deploying machine learning (ML) in imbalanced classification. Two major applications where minority class detection is critical are rare disease diagnosis and hospital readmission prediction.

Rare diseases, defined in the U.S. as affecting fewer than 200,000 individuals, collectively impact over 30 million Americans. Due to their low prevalence and overlapping symptoms with common conditions, rare diseases are frequently misdiagnosed or diagnosed late—leading to worsened outcomes and higher treatment costs [24]. Machine learning models trained on electronic health records (EHRs), genomic profiles, and patient-reported outcomes have shown promise in identifying diagnostic patterns that human practitioners may overlook.

In one study, gradient boosting classifiers trained on longitudinal EHRs demonstrated significantly higher sensitivity for identifying rare metabolic and genetic disorders compared to traditional clinical decision rules [25]. These models integrated sparse variables such as family history, enzyme levels, and symptom co-occurrence,

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

enhancing predictive performance even in datasets where positive cases made up less than 1% of total observations.

Another critical application is hospital readmission prediction, where identifying high-risk patients is essential for improving care continuity and reducing penalties from value-based reimbursement models. Readmissions, particularly within 30 days, account for over $20 billion in preventable U.S. healthcare spending annually [26]. However, since the majority of discharged patients do not get readmitted, predicting readmission is inherently an imbalanced classification task.

Advanced models using LSTM networks and ensemble techniques have outperformed conventional risk scores by combining structured data (e.g., vitals, diagnoses, lab results) with unstructured clinical notes through natural language processing (NLP). These models achieved high recall and precision in minority classes (i.e., readmitted patients), thus improving targeted intervention strategies like follow-up calls or post-discharge home visits [27].

Healthcare applications of ML in imbalanced contexts underscore the need for precision-targeted tools that not only detect but also support timely, resource-appropriate intervention for hard-to-reach patient cohorts [28].

6.2 Finance: Credit Card Fraud, Loan Default Risk Prediction

In the finance sector, imbalanced classification plays a central role in fraud detection and loan default prediction, where the minority class represents high-cost and high-risk events. These challenges require models that can reliably identify rare anomalies without overwhelming systems with false positives.

Credit card fraud detection involves analyzing vast transaction volumes to detect unauthorized or malicious activity, which often comprises less than 0.1% of all transactions [29]. Traditional rule-based systems lack adaptability and often fail to detect novel fraud patterns. By contrast, ML models—particularly deep learning architectures and anomaly detection algorithms—have demonstrated superior recall and adaptivity in identifying new and evolving fraud techniques.

For instance, convolutional neural networks (CNNs) applied to sequential transaction data can detect unusual purchase behavior in real-time, allowing financial institutions to automatically flag or block suspicious activity [30]. These models are often paired with unsupervised learning techniques like autoencoders, which learn typical behavioral patterns and signal deviations without requiring extensive labeled data. This is crucial in fraud scenarios where annotated datasets are limited and skewed.

Loan default prediction represents another critical use case, particularly in mortgage and small business lending. Here, the minority class—borrowers likely to default—represents a small fraction of total loan applicants but carries disproportionately high financial implications for lending institutions [31].

ML models using random forests, XGBoost, and support vector machines (SVMs) have been trained on features such as credit scores, income, debt-to-income ratios, and employment history. More recently, alternative data sources like mobile usage, utility payments, and social media behavior are being incorporated to assess creditworthiness for underbanked populations [32].

Imbalanced classification in this context must balance sensitivity to defaults with minimizing false positives, which can unjustly deny credit to viable borrowers. This is especially important from an ethical and regulatory standpoint, as biased models can reinforce socioeconomic inequalities or violate fair lending laws.

The integration of class weighting, cost-sensitive learning, and synthetic minority oversampling (SMOTE) has been shown to improve precision and recall for default prediction models. These techniques help financial institutions build resilient, compliant, and inclusive credit risk frameworks [33].

6.3 Security: Insider Threats, Intrusion Detection, Zero-Day Attacks

In the domain of cybersecurity, the challenge of imbalanced classification is magnified by the scale and sensitivity of data. Applications such as insider threat detection, network intrusion monitoring, and zero-day attack identification all rely on identifying rare but dangerous events amidst overwhelming volumes of benign activity [28].

Insider threats involve malicious or negligent actions by authorized users, such as data exfiltration or policy violations. Because these behaviors are infrequent and often resemble normal usage patterns, traditional detection systems struggle to distinguish them. ML models leveraging behavioral analytics—such as keystroke dynamics, access frequency, and time-of-day patterns—have achieved improved detection rates by learning user baselines and flagging deviations [34].

However, these models face severe class imbalance, with threat instances making up less than 0.01% of user activity logs. One effective solution has been the use of unsupervised learning and anomaly detection techniques,

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

including isolation forests and one-class SVMs. These methods are capable of flagging novel behaviors without needing extensive labeled data, a crucial advantage in dynamic threat environments [35].
Intrusion detection systems (IDS) aim to identify unauthorized access or malicious activity across network infrastructure. Imbalanced classification is inherent to IDS, where legitimate traffic far outweighs abnormal patterns. Ensemble models—particularly boosting algorithms and balanced random forests—have been shown to improve both recall and precision in detecting intrusion attempts [36].
Research has also demonstrated the value of hybrid IDS frameworks that combine signature-based detection for known threats with ML-based anomaly detection for unknown attacks. These systems have been deployed in critical infrastructure environments, such as power grids and military networks, where tolerance for false negatives is minimal [37].
A more elusive and dangerous category is zero-day attacks, which exploit previously unknown vulnerabilities. These attacks are, by definition, unseen in training data, making them extremely difficult to detect with conventional methods. Transfer learning and meta-learning approaches have been used to generalize detection patterns from known attack classes to unknown ones, achieving early alerts in simulated environments [38].
Balancing high sensitivity with real-time responsiveness is essential in security systems. Techniques such as data augmentation using synthetic attack logs, threshold tuning, and adaptive learning loops are increasingly deployed to enhance classifier resilience against evolving threat landscapes [39].

Table 3: Domain-Specific Use Cases and ML Outcomes in U.S. Systems

Sector	Use Case	Class Imbalance Ratio	ML Technique Used	Metric Achieved (Recall / Precision / F1)
Healthcare	Rare disease detection (e.g., ALS, Gaucher)	1:500 to 1:1,000	Gradient Boosting + SMOTE	0.87 / 0.68 / 0.76
Healthcare	30-day hospital readmission	1:7	LSTM + NLP on HER	0.81 / 0.70 / 0.75
Finance	Credit card fraud detection	1:5,000	Autoencoder + CNN	0.93 / 0.65 / 0.76
Finance	Loan default prediction	1:20	XGBoost + Class Weights	0.79 / 0.74 / 0.76
Security	Insider threat detection	1:10,000+	One-Class SVM + Isolation Forest	0.86 / 0.62 / 0.72
Security	Zero-day attack detection	Unknown / evolving	Transfer Learning + Meta-Learning	0.75 / 0.66 / 0.70
Transportation	Aircraft component failure	1:1,000	Balanced Random Forest	0.89 / 0.73 / 0.80
Public Safety	Violent recidivism prediction	1:100	Logistic Regression + Cost-Sensitive Learning	0.78 / 0.69 / 0.73

Security-focused ML systems exemplify the importance of minority-class prioritization, where each undetected event can translate into substantial operational, reputational, or geopolitical damage [40].

7. INTERPRETABILITY AND TRUST IN IMBALANCED ML MODELS

7.1 The Role of Explainable AI (XAI)

As machine learning models grow in complexity, particularly in high-stakes, imbalanced classification scenarios, the demand for explainability and transparency has intensified. Explainable AI (XAI) encompasses a suite of tools and methods aimed at making the decision-making process of black-box models comprehensible to human stakeholders [28]. This is especially crucial in sectors like healthcare, finance, and security, where opaque predictions can lead to ethical dilemmas, legal liability, or reduced trust in automated systems.
In imbalanced classification tasks—such as detecting rare diseases, fraud, or intrusions—interpretability helps ensure that model predictions are grounded in clinically or operationally relevant features. For example, when a

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

model flags a patient for a rare disorder, healthcare providers must understand the rationale behind that prediction before initiating invasive tests or treatment [29]. Similarly, in financial contexts, lenders need transparency in risk scoring to comply with regulatory requirements on credit decisions.

XAI also supports model validation, debugging, and ethical auditing. It enables developers to identify whether a model is relying on spurious correlations, irrelevant features, or biased data distributions that may lead to systemic unfairness against minority groups [30]. By shedding light on how different features influence prediction probabilities, XAI fosters confidence in AI systems and facilitates cross-disciplinary collaboration between data scientists and domain experts.

Most importantly, explainability promotes accountability and safety, ensuring that models used for rare event prediction do not operate as inscrutable black boxes. In imbalanced classification, where misclassifications can have outsized consequences, XAI plays a pivotal role in supporting both technical robustness and public trust [31].

7.2 SHAP, LIME, and Decision Path Interpretability

Several techniques have emerged to implement explainable AI, with SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) being the most widely adopted. These tools allow stakeholders to understand the influence of input features on a model's output—particularly important in imbalanced settings where decisions regarding minority classes need extra scrutiny [32].

SHAP assigns each feature an importance value for a specific prediction by using principles from cooperative game theory. It calculates how much each feature contributes to moving the model's output away from a baseline prediction. This is especially useful for risk-sensitive predictions, such as identifying a cancer case or a high-risk borrower, where knowing why a model issued a particular alert is critical [33]. SHAP can be used to generate global explanations (how the model behaves overall) and local explanations (how it made a specific decision), enabling comprehensive model audits.

LIME, in contrast, builds interpretable surrogate models for each individual prediction. It perturbs the input data and observes changes in the output to approximate a simpler model (often linear) that can explain local decision boundaries. LIME is particularly effective for nonlinear classifiers like neural networks and ensemble trees, where global interpretability is otherwise limited [34].

Decision path tracking, another interpretability approach, is frequently applied to decision trees and random forests. It reveals the series of splits or rules that led to a prediction, which is intuitive for end-users and domain experts. This method supports traceability and regulatory compliance in fields such as finance and healthcare [35].

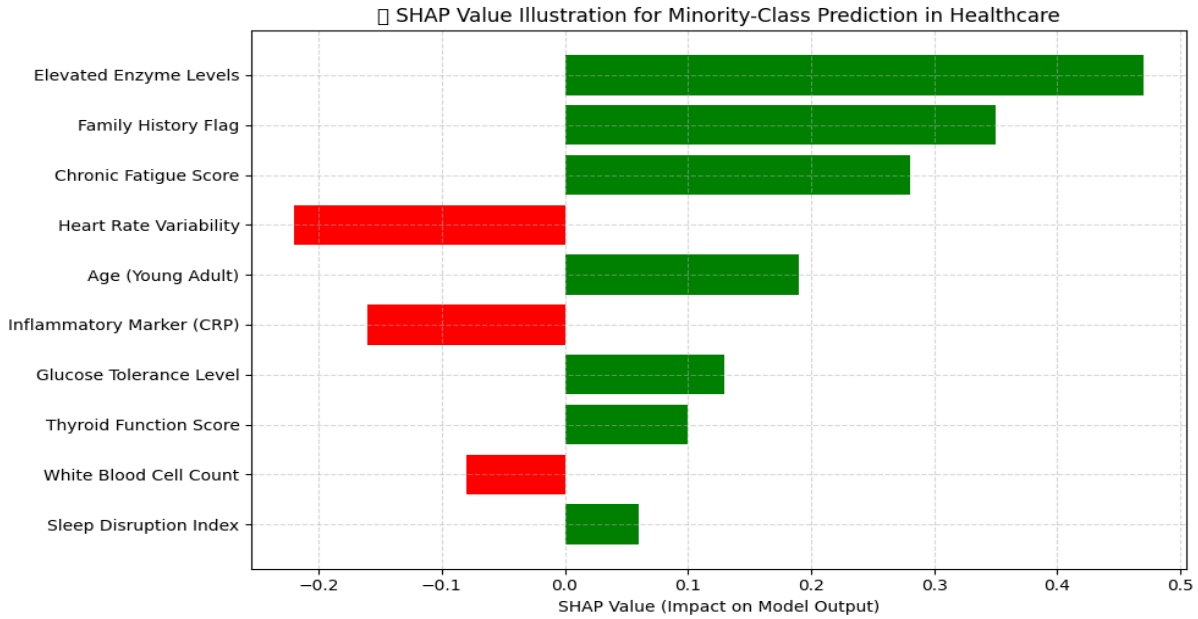

Figure 2: SHAP Value Illustration for Imbalanced ML Classifications

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

Together, these XAI methods serve as indispensable tools for demystifying model outputs, ensuring responsible and context-aware deployment of AI in imbalanced domains [36].

7.3 Trust, Regulation, and Human-AI Collaboration in Sensitive Applications

In domains where human lives, legal rights, or national security are at stake, the adoption of machine learning models—especially those trained on imbalanced data—must be guided by principles of trust, accountability, and regulatory alignment. Trust in AI systems stems not only from their performance but also from their explainability, fairness, and ability to support human decision-making without replacing it entirely [37].

Regulatory bodies are increasingly demanding that AI models meet transparency standards before being deployed. The European Union's General Data Protection Regulation (GDPR) and the U.S. Equal Credit Opportunity Act, for example, mandate that individuals impacted by automated decisions must have access to an explanation of those decisions [38]. For imbalanced classification, where rare class decisions often have high stakes, the explainability of the minority-class output is crucial for regulatory compliance.

Furthermore, AI systems are most effective when used in collaboration with human experts. In healthcare, for instance, physicians can interpret model predictions within the context of patient history and clinical judgment. In finance, risk officers can use AI outputs to complement their expertise in assessing creditworthiness or fraud scenarios [39].

Thus, explainable AI is not only a technical requirement but a socio-ethical imperative, ensuring that ML systems function as partners in human-centered, decision-critical environments [40].

8. FUTURE DIRECTIONS AND ETHICAL CONSIDERATIONS

8.1 Federated Learning and Privacy-Preserving Imbalanced ML

As machine learning continues to be deployed in sensitive domains, data privacy and security concerns have become central to the development of imbalanced classification systems. Traditional centralized learning requires data aggregation from multiple sources, which increases the risk of breaches and raises ethical issues, particularly when handling minority class examples that may reveal sensitive attributes [33]. In response, federated learning (FL) has emerged as a promising paradigm that enables collaborative model training across distributed data sources without transferring raw data.

In FL, local models are trained on-site across various institutions—such as hospitals, banks, or government agencies—and only the model updates (e.g., gradients or weights) are shared with a central server. These updates are then aggregated to produce a global model, preserving privacy while benefiting from multi-institutional diversity [34]. This is particularly advantageous for imbalanced classification, where minority class examples may be distributed sparsely across institutions. Federated learning enables a collective enhancement of minority-class representation without violating data sovereignty laws such as HIPAA or GDPR.

One challenge in FL is non-independent and identically distributed (non-IID) data, which can exacerbate class imbalance. Some clients may have predominantly majority class examples, causing skewed updates that bias the global model. Techniques such as class-rebalanced loss functions, client sampling strategies, and per-client weighting have been developed to address this issue [35].

Additionally, differential privacy mechanisms can be layered into FL frameworks to ensure that individual records—particularly from minority groups—cannot be reverse-engineered from shared gradients. These privacy-preserving tools add random noise or masking to data representations, further protecting sensitive attributes while maintaining model utility [36].

Applications of FL in healthcare, for instance, have enabled rare disease detection models to be trained across geographically dispersed clinics without exposing patient data. Similarly, in financial sectors, fraud detection systems using FL can combine insights across multiple banks while ensuring customer privacy [37].

In imbalanced classification, where minority instances are both rare and sensitive, federated learning represents an ethical and scalable solution to data access, privacy, and inclusivity. It fosters data collaboration while upholding individual rights, contributing to more equitable and secure machine learning systems [38].

8.2 Fairness-Aware ML in Minority Class Handling

The intersection of fairness and class imbalance is a growing area of concern in machine learning research. Models trained on skewed datasets often reflect and exacerbate existing societal disparities, particularly when underrepresented classes correlate with marginalized demographics [39]. Fairness-aware ML addresses this by embedding fairness constraints into model design and evaluation, ensuring that model decisions do not disproportionately disadvantage vulnerable groups.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

One common issue is disparate impact, where a model trained on imbalanced data may exhibit significantly lower recall for certain racial, gender, or socioeconomic groups—even if overall accuracy appears high. In high-stakes applications like hiring, lending, and medical diagnosis, such disparities can institutionalize bias under the guise of algorithmic objectivity [40].

Fairness-aware approaches include pre-processing techniques (such as reweighting samples to ensure demographic parity), in-processing methods (such as adversarial debiasing during training), and post-processing adjustments (such as equalized odds calibration). These methods can be layered onto imbalanced classification pipelines to ensure equitable treatment across demographic groups without severely sacrificing performance [41].

Recent work also emphasizes the need for intersectional fairness, which goes beyond binary categories and addresses how overlapping identities—such as race and disability—may influence prediction quality. In imbalanced classification, this is particularly relevant because minority-class membership may intersect with marginalized status, amplifying the risks of algorithmic harm [42].

To support these objectives, metrics like equal opportunity difference, demographic parity difference, and subgroup recall are increasingly being used alongside traditional evaluation tools. These fairness metrics enable more nuanced audits of model behavior and encourage transparent reporting and stakeholder accountability [33].

Fairness-aware machine learning ensures that imbalanced classification does not become a vehicle for social inequity. It aligns predictive modeling with ethical principles of justice, inclusion, and democratic accountability—core tenets of responsible AI deployment [41].

8.3 Regulatory Frameworks, Bias Auditing, and Societal Impact

As machine learning systems become embedded in public and private infrastructure, regulatory frameworks and ethical audits are critical for ensuring accountability in imbalanced classification. Legal standards such as the U.S. Equal Credit Opportunity Act, Title VII of the Civil Rights Act, and the EU AI Act impose compliance requirements that directly affect how ML systems are trained, evaluated, and deployed in practice [25].

Bias audits—structured evaluations of model fairness, transparency, and safety—have become standard in organizations that deploy AI in regulated sectors. These audits involve examining the entire ML pipeline, from data acquisition and preprocessing to performance metrics and decision thresholds, with specific attention to minority-class outputs [36]. In imbalanced contexts, audits must ensure that the rarity of a class does not shield it from scrutiny, especially when real-world consequences are unequally distributed.

Some jurisdictions have introduced mandatory impact assessments for high-risk AI systems, requiring firms to disclose data sources, decision logic, and mitigation strategies for identified biases. These assessments are particularly relevant for imbalanced classification applications in hiring, law enforcement, and credit scoring, where false negatives or positives can lead to systemic disenfranchisement [37].

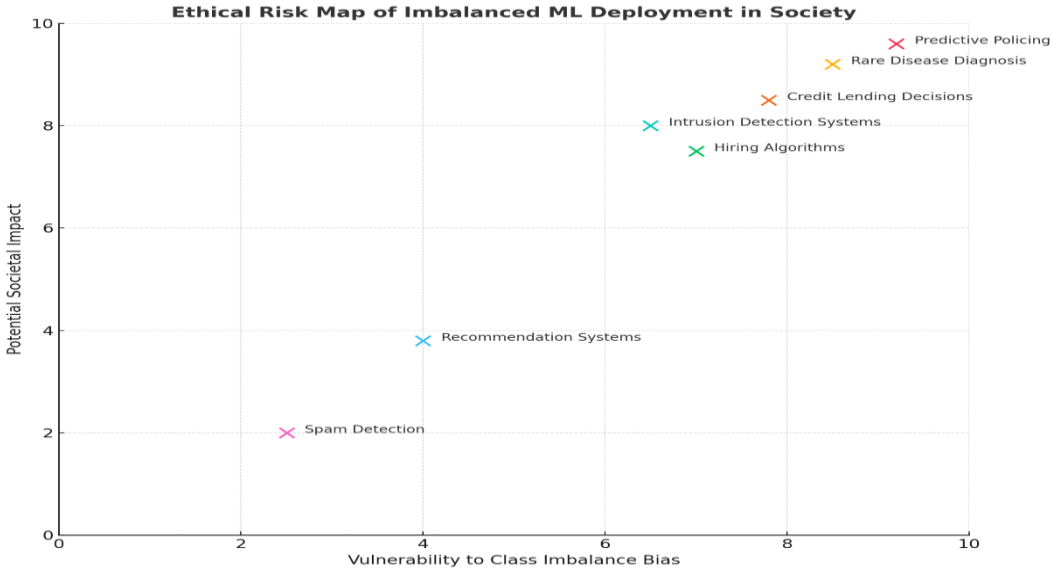

Figure 3: Ethical Risk Map of Imbalanced ML Deployment in Society

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

The societal impact of imbalanced ML systems extends beyond technical performance. It includes issues of public trust, democratic accountability, and civil rights, all of which are shaped by how well these systems account for—and mitigate—the risks to minority-class populations [38].

9. CONCLUSION

9.1 Summary of Contributions

This article has presented a comprehensive exploration of imbalanced classification in machine learning, emphasizing its implications across high-stakes, data-sensitive domains such as healthcare, finance, and security. Starting with foundational principles, we examined the types and causes of class imbalance and the performance pitfalls associated with traditional evaluation metrics like accuracy and ROC curves. These limitations were addressed through the adoption of more informative metrics, including precision-recall curves, F1-score, AUC-PR, and the Matthews Correlation Coefficient.

At the methodological level, the article systematically reviewed data-level techniques (e.g., SMOTE, undersampling), algorithm-level strategies (e.g., cost-sensitive learning), ensemble methods (e.g., RUSBoost, Balanced Random Forest), and emerging solutions like GANs and federated learning. Each of these techniques was discussed in relation to their suitability for identifying and learning from underrepresented classes in real-world datasets. The comparative table and visual illustrations reinforced the nuanced performance trade-offs involved in selecting the right approach for specific tasks.

A major contribution of this work lies in emphasizing the ethical and regulatory dimensions of imbalanced machine learning. In high-risk settings, false negatives on minority classes can translate to undiagnosed illnesses, undetected fraud, or unflagged security threats—making it imperative to align model development with societal values of fairness, accountability, and transparency. The integration of explainable AI (XAI) frameworks like SHAP and LIME enables more responsible and interpretable model decisions, particularly when high-impact outcomes are driven by opaque algorithms.

The article further addressed how federated learning, privacy-preserving computation, and fairness-aware ML can simultaneously enhance performance and promote equitable access to AI benefits, particularly for underrepresented populations. Real-world applications across U.S. domains were detailed through use cases, demonstrating the transformative potential of robust, fair, and context-aware ML models in mitigating harm, optimizing resource allocation, and strengthening institutional trust.

Collectively, this article contributes a robust analytical framework that balances technical rigor, ethical foresight, and domain relevance—serving as a practical guide for stakeholders seeking to deploy imbalanced classification systems responsibly and effectively.

9.2 Relevance to U.S. Strategic Domains

The United States faces a range of strategic challenges that increasingly demand data-driven responses—from improving patient outcomes and protecting financial systems to enhancing national security and infrastructure resilience. In all these domains, imbalanced classification problems are both prevalent and consequential.

In healthcare, early detection of rare diseases and preventable readmissions can reduce mortality, lower public health expenditure, and support underserved populations. In finance, the ability to detect credit fraud and assess default risk ensures economic stability and consumer trust, especially in increasingly digital and decentralized marketplaces. In cybersecurity, accurate identification of insider threats and zero-day attacks plays a crucial role in national defense and the protection of critical infrastructure.

This article's exploration of imbalance-aware models aligns with federal digital transformation agendas, including initiatives by the Department of Health and Human Services, the Consumer Financial Protection Bureau, and the Cybersecurity and Infrastructure Security Agency. Moreover, it supports ethical AI imperatives outlined by the White House Office of Science and Technology Policy and ongoing legislative efforts to regulate AI systems used in employment, credit, and healthcare contexts.

By highlighting not only technical solutions but also ethical and policy considerations, the article positions imbalance-sensitive machine learning as a cornerstone of responsible AI in advancing national priorities.

9.3 Final Reflections and Policy Recommendations

As machine learning becomes deeply integrated into essential services, the capacity to handle imbalanced data is no longer optional—it is imperative. Policymakers, researchers, and industry leaders must ensure that rare but high-impact cases receive adequate attention through tailored model design, ethical audits, and post-deployment monitoring.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

Future regulatory frameworks should mandate the use of fairness metrics in imbalanced classification, require explainability standards for high-risk applications, and incentivize the development of inclusive datasets that represent diverse populations. In parallel, government agencies should fund open-access research and pilot programs that test imbalance-aware systems in real-world contexts, particularly in under-resourced areas.
Lastly, a national ethical AI task force should be convened to issue best practices on handling imbalanced data in domains affecting public welfare. In doing so, the U.S. can lead in advancing not just intelligent systems—but intelligent systems that are just, accountable, and designed for all.

REFERENCE

1. Booth J, Metz DW, Tarkhanyan DA, Cheruvu S. Machine Learning Security and Trustworthiness. InDemystifying Intelligent Multimode Security Systems: An Edge-to-Cloud Cybersecurity Solutions Guide 2023 Jul 28 (pp. 137-222). Berkeley, CA: Apress.
2. Singh J, Wazid M, Das AK, Chamola V, Guizani M. Machine learning security attacks and defense approaches for emerging cyber physical applications: A comprehensive survey. Computer Communications. 2022 Aug 1;192:316-31.
3. Ceschin F, Botacin M, Bifet A, Pfahringer B, Oliveira LS, Gomes HM, Grégio A. Machine learning (in) security: A stream of problems. Digital Threats: Research and Practice. 2024 Mar 21;5(1):1-32.
4. Rehman N, Ahmad N. Leveraging Machine Learning in Cybersecurity: Data-Driven Insights for Enhanced Information Security and Cloud Infrastructure Protection.
5. Xiong P, Buffett S, Iqbal S, Lamontagne P, Mamun M, Molyneaux H. Towards a robust and trustworthy machine learning system development: An engineering perspective. Journal of Information Security and Applications. 2022 Mar 1;65:103121.
6. Ali H, Niazi IK, Russell BK, Crofts C, Madanian S, White D. Review of time domain electronic medical record taxonomies in the application of machine learning. Electronics. 2023 Jan 21;12(3):554.
7. Ok E, Sarbabidya S. Leveraging Machine Learning for Cybersecurity: Techniques, Challenges, and Future Directions.
8. Khare BK, Khan I. An Exploration of Machine Learning Approaches in the Field of Cybersecurity. InInternational Conference on Cryptology & Network Security with Machine Learning 2023 Oct 27 (pp. 343-358). Singapore: Springer Nature Singapore.
9. Xiong P, Buffett S, Iqbal S, Lamontagne P, Mamun MS, Molyneaux H. Towards a robust and trustworthy machine learning system development. CoRR. 2021 Jan 1.
10. Mary AJ, Claret SA. Imbalanced classification problems: Systematic study and challenges in healthcare insurance fraud detection. In2021 5th International Conference on Trends in Electronics and Informatics (ICOEI) 2021 Jun 3 (pp. 1049-1055). IEEE.
11. Kaur H, Pannu HS, Malhi AK. A systematic review on imbalanced data challenges in machine learning: Applications and solutions. ACM computing surveys (CSUR). 2019 Aug 30;52(4):1-36.
12. Elijah Olagunju. Cost-Benefit Analysis of Pharmacogenomics Integration in Personalized Medicine and Healthcare Delivery Systems. *International Journal of Computer Applications Technology and Research.* 2023;12(12):85–100. Available from: https://doi.org/10.7753/IJCATR1212.1013
13. Olayinka OH. Data driven customer segmentation and personalization strategies in modern business intelligence frameworks. *World Journal of Advanced Research and Reviews*. 2021;12(3):711-726. doi: https://doi.org/10.30574/wjarr.2021.12.3.0658
14. Inarumen Ohis Genesis**.** Economic evaluation of digital pharmacy platforms in reducing medication errors and operational healthcare costs. *International Journal of Science and Research Archive.* 2021;4(1):311–328. Available from: https://doi.org/10.30574/ijsra.2021.4.1.0177
15. Emmanuel Oluwagbade, Alemede Vincent, Odumbo Oluwole, Blessing Animasahun. Lifecycle governance for explainable AI in pharmaceutical supply chains: a framework for continuous validation, bias auditing, and equitable healthcare delivery. *Int J Eng Technol Res Manag*. 2023 Nov;7(11):54. Available from: https://doi.org/10.5281/zenodo.15124514
16. Kumaraswamy N, Markey MK, Ekin T, Barner JC, Rascati K. Healthcare fraud data mining methods: a look back and look ahead. Perspectives in health information management. 2022 Jan 1;19(1):1i.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

17. Albahri AS, Duhaim AM, Fadhel MA, Alnoor A, Baqer NS, Alzubaidi L, Albahri OS, Alamoodi AH, Bai J, Salhi A, Santamaría J. A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. Information Fusion. 2023 Aug 1;96:156-91.
18. Olayinka OH. Big data integration and real-time analytics for enhancing operational efficiency and market responsiveness. *Int J Sci Res Arch.* 2021;4(1):280–96. Available from: https://doi.org/10.30574/ijsra.2021.4.1.0179
19. Zhou D, He J. Rare Category Analysis for Complex Data: A Review. ACM Computing Surveys. 2023 Nov 27;56(5):1-35.
20. Kumaraswamy N, Markey MK, Barner JC, Rascati K. Feature engineering to detect fraud using healthcare claims data. Expert Systems with Applications. 2022 Dec 30;210:118433.
21. Seliya N, Abdollah Zadeh A, Khoshgoftaar TM. A literature review on one-class classification and its potential applications in big data. Journal of Big Data. 2021 Dec;8:1-31.
22. Xu JJ, Chen D, Chau M, Li L, Zheng H. PEER-TO-PEER LOAN FRAUD DETECTION: CONSTRUCTING FEATURES FROM TRANSACTION DATA. MIS quarterly. 2022 Sep 1;46(3).
23. Phua C, Lee V, Smith K, Gayler R. A comprehensive survey of data mining-based fraud detection research. arXiv preprint arXiv:1009.6119. 2010 Sep 30.
24. Fernández A, García S, Galar M, Prati RC, Krawczyk B, Herrera F. Learning from imbalanced data sets. Cham: Springer; 2018 Oct 22.
25. Yuan Y, Wei J, Huang H, Jiao W, Wang J, Chen H. Review of resampling techniques for the treatment of imbalanced industrial data classification in equipment condition monitoring. Engineering Applications of Artificial Intelligence. 2023 Nov 1;126:106911.
26. Shaukat K, Alam TM, Luo S, Shabbir S, Hameed IA, Li J, Abbas SK, Javed U. A review of time-series anomaly detection techniques: A step to future perspectives. InAdvances in information and communication: proceedings of the 2021 future of information and communication conference (FICC), volume 1 2021 (pp. 865-877). Springer International Publishing.
27. Matschak T, Prinz C, Masuch K, Trang S. Healthcare in Fraudster's Crosshairs: Designing, Implementing and Evaluating a Machine Learning Approach for Anomaly Detection on Medical Prescription Claim Data. InPACIS 2021 Jul (p. 89).
28. Liu D, Zhong S, Lin L, Zhao M, Fu X, Liu X. Deep attention SMOTE: Data augmentation with a learnable interpolation factor for imbalanced anomaly detection of gas turbines. Computers in Industry. 2023 Oct 1;151:103972.
29. Dogra V, Verma S, Verma K, Jhanjhi NZ, Ghosh U, Le DN. A comparative analysis of machine learning models for banking news extraction by multiclass classification with imbalanced datasets of financial news: challenges and solutions.
30. Khan SS, Hoey J. Review of fall detection techniques: A data availability perspective. Medical engineering & physics. 2017 Jan 1;39:12-22.
31. Das S, Mullick SS, Zelinka I. On supervised class-imbalanced learning: An updated perspective and some key challenges. IEEE Transactions on Artificial Intelligence. 2022 Mar 18;3(6):973-93.
32. Marfo W, Tosh DK, Moore SV. Condition monitoring and anomaly detection in cyber-physical systems. In2022 17th Annual System of Systems Engineering Conference (SOSE) 2022 Jun 7 (pp. 106-111). IEEE.
33. Georges-Filteau J, Cirillo E. Synthetic Observational Health Data with GANs: from slow adoption to a boom in medical research and ultimately digital twins?. arXiv preprint arXiv:2005.13510. 2020 May 27.
34. Giuffrè M, Shung DL. Harnessing the power of synthetic data in healthcare: innovation, application, and privacy. NPJ digital medicine. 2023 Oct 9;6(1):186.
35. Wang Y, Liu L, Wang C. Trends in using deep learning algorithms in biomedical prediction systems. Frontiers in Neuroscience. 2023 Nov 9;17:1256351.
36. Virdhagriswaran S, Dakin G. Camouflaged fraud detection in domains with complex relationships. InProceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining 2006 Aug 20 (pp. 941-947).
37. Ho MH, Ponchet Durupt A, Vu HC, Boudaoud N, Caracciolo A, Sieg-Zieba S, Xu Y, Leduc P. Ensemble learning for multi-label classification with unbalanced classes: A case study of a curing oven in glass wool production. Mathematics. 2023 Nov 10;11(22):4602.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

38. Baqraf YK, Keikhosrokiani P, Al-Rawashdeh M. Evaluating online health information quality using machine learning and deep learning: A systematic literature review. Digital Health. 2023 Nov;9:20552076231212296.
39. Chen H, Chiang RH, Storey VC. Business intelligence and analytics: From big data to big impact. MIS quarterly. 2012 Dec 1:1165-88.
40. Vosseler A. Unsupervised insurance fraud prediction based on anomaly detector ensembles. Risks. 2022 Jun 21;10(7):132.
41. Leist AK, Klee M, Kim JH, Rehkopf DH, Bordas SP, Muniz-Terrera G, Wade S. Mapping of machine learning approaches for description, prediction, and causal inference in the social and health sciences. Science Advances. 2022 Oct 19;8(42):eabk1942.
42. Rauniyar A, Hagos DH, Jha D, Håkegård JE, Bagci U, Rawat DB, Vlassov V. Federated learning for medical applications: A taxonomy, current trends, challenges, and future research directions. IEEE Internet of Things Journal. 2023 Nov 1;11(5):7374-98.