International Journal of Innovative Research in Engineering
Volume 7, Issue 1 (January - February 2026), PP: 47-58.
https://www.doi.org/10.59256/ijire.20260701006
www.theijire.com
ISSN No: 2582-8746

Soft Computing Approaches for Robust Analysis of Imbalanced and Noisy Data

Charu Jhagrawat
Parul University, Waghodia, Gujarat, India.

How to cite this paper:
Charu Jhagrawat, "Soft Computing Approaches for Robust Analysis of Imbalanced and Noisy Data", IJIRE-V7I1-47-58.

Abstract: *The accuracy and diversity of information flows have become major characteristics of the modern age Big Data. it may not be noted that information about people, organizations, and events that are deemed critical bottlenecks within the deployment of trustworthy machine learning solutions. Despite the exponential nature that the field of data has followed, the quality of this kind of data is often low, particularly represented by extreme cases of Class Imbalance and Noise. Conventional hard computing paradigm based on Aristotelian logic and sharp decision boundaries, often present catastrophicFailure Modes when faced with so many data pathologies. Generally, traditional classifiers are prone to making biased predictions towards the majority class and at the same time overfitting to noisy cases that contaminate the manifold of the feature set. This research paper offers a comprehensive, expert-level examination of the methodologies of Soft Computing (SC)—including Fuzzy Logic, Artificial Neural Networks, and Evolutionary Algorithms—as a solid substitute in dealing with the uncertainty of real-world data.*

We will critically analyze the theoretical foundations of SC, explaining how the tolerance for imprecision and incomplete truth makes possible the design of decision boundaries robust to overlapping class distributions and the presence of noise within the labels. Moreover, we will introduce a novel Hybrid Soft Computing Framework, namely the ***REF-DB (Robust Evolutionary-Fuzzy Data Balancing)****, which jointly exploits the ability to efficiently handle label noise of Fuzzy Logic Filtering and the global optimization potential of Evolutionary Under sampling.*

This paper will review state-of-the-art techniques operating at multiple tiers: from FSVM, Genetic Fuzzy Systems, to the latest trends that involve LLMs for semantic data augmentation and Quantum Soft Computing for high-dimensional feature mapping. Our work presents a thorough comparison experimental evidence drawn from several benchmark datasets, such as KEEL and UCI. We show through a proper comparative analysis that hybrid SC approaches guarantee maximum values of robust metrics, like the G-Mean and AUC, outperforming hard computing baselines formed by singular approaches. The study concludes with a forward-looking discussion on the integration of Neuro-Symbolic AI and Quantum Machine Learning, positing that the future of robust data analytics lies in the fusion of evolutionary adaptability and fuzzy reasoning.

Keywords: *Soft Computing, Class Imbalance, Label Noise, Fuzzy Logic, Evolutionary Algorithms, Neural Networks, Hybrid Systems, Evolutionary Under sampling, Quantum Machine Learning, Generative AI.*

I.INTRODUCTION

1.1 The Data Quality Crisis in Modern Analytics

The Fourth Industrial Revolution has been driven by data. Right from the sensors used in IoT to track industrial equipment to EHRs used to direct precision medicine, data has been the key to successful decision-making. Nevertheless, the "Big Data" approach can be summed up in the form of Vs, which stand for Volume, Velocity, Variety, and Veracity. While Volume and Velocity concerns have been taken care of in distributed computing models such as Hadoop or Spark, **Variety** (as in the problem of class distribution) and **Veracity** (data quality) continue to be major algorithmic problems.[1]

In real-world scenarios, "the data of prime interest is usually the rarest data". In fraud analysis, honest transactions by orders of magnitude (usually 10,000 to 1). In medical applications, healthy individuals greatly outnumber those with a given rare disease. This is called **Class Imbalance**, and "it is a fundamental problem for standard supervised learning methods".[3] Typical classifiers, like Decision Trees or standard Support Vector Machines (SVM), implicitly assume equal class priors and equal misclassification costs. On imbalanced datasets, these "hard" computational models simplify the minority class by assigning all samples to the majority class to achieve high "accuracy" (indeed, 99.9%) but performing catastrophically on the main task of fraud detection.[5]

This problem is made worse by the presence of **Noise**. In reality, when collecting data, you don't get it spot on.

There is sensor drift to consider, or the human annotators may be wrong. There can be mistakes in the error-free channels. Noise can be in the form of Attribute Noise or *Label Noise.* In an imbalanced setting, Noise works against you. It doesn't take much noise in the majority class residing in the area of the minority class to be considered a representation of the minority class by a hard classifier.[6]

1.2 The Limitations of Hard Computing

Traditional "Hard Computing" is precise, certain, and exact. It uses binary logic (True/False), sharp sets, and deterministic algorithms. Though very capable in a controlled, clean-data situation, hard computing has difficulty capturing the ambiguity of the real world. For example, a hard computing approach to developing a sharp linear boundary between the set of "Healthy" individuals and the set of "Sick" individuals based on a measurement of blood pressure can never be a good approach, as a zone of ambiguity will always exist where the classification cannot be sharply separated. Instead, a hard computing approach has a problem, in that it can choose to ignore the outlier (underfitting) or move its decision boundary to accommodate the outlier (overfitting), without much of a middle ground.[1]

1.3 The Soft Computing Paradigm

Soft Computing (SC), a term coined by Lotfi Zadeh, signifies an integrated group of methodologies aimed at approximating imprecision and uncertainty existing in real-world problems. The basic principle behind SC is to leverage **imprecision, uncertainty, partial truth, and approximation** in achieving better **tractability and low solution cost**.[1]

The main components of Soft Computing are:

1. **Fuzzy Logic (FL):** Offer a mathematical structure to work with membership whose value is graded. In standard sets, membership is either true or false, 0 or 1. But on FL, membership is true to a degree specified by $\mu(x)\ in $. Hence, there may be tolerant decision boundaries which can be tolerant to noise before system failure.[5]
2. **Artificial Neural Networks (ANN):** Artificial neural networks were modelled after biological neural networks and perform very well in non-linear learning. They offer "learning" and "adaptation" facilities in the SC System.[1]
3. **Evolutionary Algorithms (EA):** Meta-heuristic techniques founded on natural selection paradigms (Genetic Algorithms and Particle Swarm Optimization). EAs are able to examine huge and complex spaces of solutions to detect their global maxima and minima and therefore serve as excellent approaches to feature and instance selection for imbalanced databases.[8]

1.4 Objectives and Contribution

This research report attempts to offer an exhaustive and definitive investigation of how Soft Computing techniques can be applied towards tackling the two challenges of Class Imbalance and Noise. It not only tries to be an exhaustive survey report but also attempts to formulate a theory of **Hybrid Soft Computing** from existing literature.

The major contributions of this work are:

- An assessment of how conventional learning algorithm failure modes can be exacerbated when dealing with imbalanced data.
- A comprehensive analysis of how each of Fuzzy Logic, Neural Networks, Evolutionary Algorithms, and these three taken in combination can overcome these pitfalls.
- Proposal for the **Robust Evolutionary-Fuzzy Data Balancing (REF-DB)** framework, which is hybrid in nature and involves both fuzzy noise filtering and evolutionary under-sampling.
- A description of frontiers of research, such as using Generative AI (LLMs) for Semantic Oversampling, and even Quantum Soft Computing.

II. BACKGROUND AND RELATED WORK

To appreciate the necessity of Soft Computing, one must first understand the landscape of existing solutions and their limitations. The problem of learning from imbalanced datasets has been studied extensively over the past two decades, with solutions generally categorized into Data-level, Algorithm-level, and Cost-sensitive approaches.[3]

2.1 Classical Approaches to Imbalance

Data-Level Models: These methods attempt to distribute classes well before training.

- Random Under sampling (RUS): Selects instances randomly from the majority class. While very efficient, there is a possible loss of crucial statistical information which could contain "support vectors" that define such classification boundaries.[3]
- Random Oversampling (ROS): It involves copying instances of the minority class. The higher instances are more vulnerable to the risk of overfitting, given the possibility of memorization of noise in the replicated instances.[3]
- SMOTE (Synthetic Minority Over-sampling Technique): The SMOTE technique is one of the most renowned extensions of traditional approaches to ROS. The method relies on interpolation with examples of the minority class to obtain new examples. However, SMOTE is blind. A minority class example can be noisy. If it is so, SMOTE will begin to generate new examples around it. This results in noise distribution and connection of classes.[10]

Algorithm-Level Solutions: These types of solutions involve those which impact the classifier algorithm.

- Threshold Moving: The variation of changes within the threshold boundaries for making the decision class (starting from 0.5 to 0.2) to favour the minority class.
- Cost-Sensitive Learning: Assig a higher cost to misclassifying a minority sample than a majority sample. Although this can be an effective bias, it's difficult to design a cost matrix. Also, these cost-sensitive learning techniques are known to be highly noise-sensitive. Here, a noisy majority class labelled with minority class pushes the learner to overfit because of an attracting cost trap.[5]

2.2 The Problem of Noise in Machine Learning

Noise is not only related to noise but is a totally new field of learning. There are two types of noise: **Class Noise** and **Attribute Noise**. Class Noise can be defined as the Noise in the labels when they are flawed. Attribute Noise is also known as Noise when the data related to the attributes is erroneous. Class Noise can either be present or not, i.e., it is a binary noise. Additionally, Class Noise affects the noise within the confusion matrix. A confusion matrix is a table.

- *Noise Label: In* particular, label noise is harmful. In a 1% minority class dataset, when 1% of majority class data is labelled as minority class data, then "fake" data points in minority class. Thus, learning is not possible in majority of learning models.[6]
- Although the capability of deep learning models is very high, they generally experience challenges in memorizing noisy labels. Various studies showed that it is possible to train the Deep Neural Networks (DNNs) using random labels successfully by reaching 100% label accuracy in training, meaning having a value of 0% in the generalization loss. This led to investigation on the concept of a "robust loss function" or the removal of noise using CRUST.[6]

2.3 Evolution of Soft Computing in Data Analytics

The application of Soft Computing to these problems has evolved from simple rule-based systems to complex hybrids.

- **Fuzzy Classifiers:** It has been shown that the "overlap problem" could be better handled by Fuzzy Rule-Based Classification Systems rather than decision trees. Linguistic variables make it possible to define a "mostly majority but possibly minority" area, thereby maintaining the information lost in the hard classifier.[5]
- **Evolutionary Sampling:** The usage of Genetic Algorithms within the context of or Evolutionary Under-sampling achieved prominence in the late 2000s. Conventional random under sampling, rather than searching for the optimal samples to keep with regard to being in the majority for maximizing the fitness function (typically G-Mean), Evolutionary Under-sampling has been shown to be significantly better than random methods through statistical analysis.[9]
- **Recent Trends (2024-2025):** The literature has been heavily biased towards a combination of methods. For instance, Fuzzy Logic applied to Neural Networks (Fuzzy NN) for handling noisy sets of features, Evolutionary Algorithms for hyperparameters' optimization of Deep Learning models (Neuro-Evolution). More recently, a move has been observed towards a combination of Generative AI and Soft Computing, using Large Language Models to minority's generation through a variety of means.[16]

III. CHALLENGES OF IMBALANCED AND NOISY DATA

The problem of analysing the biased and noisy datasets is not a specific problem but relates to a phenomenon of combining geometric and statistical challenges. This impacts the performance of the learning model.

3.1 The Theory of Small Disjuncts

One of the major reasons why classifiers tend to have less accuracy when training with imbalanced data is related to the "small disjuncts" phenomenon. In most cases, the minority class is not represented by just one tight group in the feature space. Rather, it consists of lots of smaller subgroups called disjuncts that are spread all over the region.

Generally, these standard algorithmic methods have a tendency, based on the principle of Occam's Razor, to eliminate such small disjuncts because these tend to consider these values as noise. However, for imbalanced datasets, such values represent the correct but equally important sub-concepts for the class with smaller data. Loss of such would mean that the level of False Negative is high. It is for these cases that soft computing methods, more specifically Fuzzy Systems, prove extremely effective because these can offer granular-level rules that have the potential for including such small regions without generalization.5

3.2 Class Overlap and Separability

It might happen that, relative to the Class Overlap level, the level of imbalance becomes less relevant. In situations where the classes of the minority as well as the majority class data points could be distinguished, whether linearly or otherwise, a classification problem could be ideally solved by a classification model for any value of the Imbalance Ratios. However, this is not the practical situation.

In these overlapping regions, the posterior probability $P (Class | Feature) $ will still be uncertain. For hard classifiers, a classification has to be assigned, and a tendency exists to default to the probable occurrence of the dominant class. Compared to hard classifiers, fuzzy classifiers can define such a region with an approximate degree of membership, close to 0.5, for both classes.12

3.3 The Evaluation Metric Trap

In regard to the most critical issue in this area, there's an issue with the misleading properties of the most popular assessment measures. Accuracy basically does not mean anything. For example, if there's an Imbalance Ratio (IR) of 100:1, a weak classifier which always classifies an object into "Majority" would result in 99% accuracy but would be completely redundant.

This necessitates the use of robust metrics:

- **Sensitivity (Recall):** $TP / (TP + FN) $ - Critical for safety-critical applications.
- **Specificity:** $TN / (TN + FP) $ - Critical to avoid false alarms.
- **Geometric Mean (G-Mean):** $\sqrt {Sensitivity \times Specificity} $ - A metric that balances accuracy on both classes.
- F1-Score: Harmonic mean of Precision and Recall. A problem with optimization methods, as with backpropagation from the Neural Network algorithm, is optimising on approximations for Accuracy. The method used with Soft Computing; Evolutionary Algorithms does enable optimisation on non-differentiable objectives such as G-Mean or AUC.4

3.4 Types of Noise and Their Impact

Noise in imbalanced datasets acts as a multiplier of difficulty.

1. **Attribute Noise:** Noise in input feature vector x. It affects accuracy and leads to overlapping of regions of different classes.
2. **Label Noise:** Errors in the target y.

- **Majority $\rightarrow$ Minority Noise:** It creates "dummy" data points in the majority class. The AdaBoost or SVM would learn very strongly from these data points, and this might lead to complicated decision boundaries and overfitting.
- **Minority $\rightarrow$ Majority Noise:** A manifestation of this method is removing valid examples from the minority class. This has an indirect effect of increasing the imbalance ratio. The traces of learning the minority concept would be erased.[6]

IV. SOFT COMPUTING TECHNIQUES FOR DATA ANALYSIS

This section explores the specific mechanisms by which the three pillars of Soft Computing address the challenges outlined above.

4.1 Fuzzy Logic Approaches

Fuzzy Logic (FL) can be suitably classified in a context where the classes may overlap partly, generating fuzzy or imprecise information. Unlike crisp logic, where an element x either belongs to set A or not ($\mu_A(x) \in \ {0, 1\} $), fuzzy logic allows partial membership ($\mu_A(x) \in $).

4.1.1 Fuzzy Support Vector Machines (FSVM)

The standard Support Vector Machine (SVM) is a perhaps one of the most efficient classifiers. But it is very sensitive to noise and outliers. The primary task of the standard SVM is to ensure that

$$\min \frac {1}{2} ||w||^2 + C \sum_{i=1} ^{N} \xi_i$$

where ξ_i represents the error (slack variable) for the i-th sample. A single noisy outlier with a large ξ_i can drastically shift the hyperplane.

The Fuzzy SVM (FSVM) introduces the membership value s_i for every data point to define the degree of confidence that the data point belongs to the same class, to which it is labelled. The problem formulation can be described mathematically as follows:

$$\min \frac {1}{2} ||w||^2 + C \sum_{i=1} ^{N} s_i \xi_i$$

- **Mechanism:** If a sample x_i is a noisy outlier (e.g., a majority sample deep in the minority region), it is assigned a low membership score s_i (e.g., 0.1). Consequently, the penalty for misclassifying it ($s_i \xi_i$) is small, and the SVM "ignores" it, focusing on the core, reliable samples to draw the boundary.
- **Determination of Membership:** Advanced methods like the **Relative Density-based Intuitionistic Fuzzy SVM (RIFSVM)** or **Slack-Factor-based FSVM (SFFSVM)** automatically calculate s_i based on the local density of neighbors. If a minority sample is surrounded by majority samples, it is deemed noise (low s_i). If it is surrounded by other minority samples, it is deemed a valid support vector (high s_i).[12]

4.1.2 Fuzzy Rule-Based Classification Systems (FRBCS)

FRBCS use linguistic rules (e.g., "IF X_1 is *High* AND X_2 is *Low* THEN Class is *Positive* with Certainty 0.8").

- **Handling Overlap:** In an overlapping situation, there would be more than one rule that fires for a particular input but they would fire to different degrees. The choice of classification or categorization is achieved through aggregation of the resulting fuzzily computed values. This means that there will be no hard decision boundary but a soft decision.
- **Handling Imbalance:** One can attempt to assign more "Rule Weights" for triggering in the minority class can be covered by strongly weighted prediction rules.[5]

4.2 Neural Networks

Although classical Neural Nets are considered to be extremely susceptible to biased conclusions in imbalanced datasets, modern modifications have made them a strong tool for soft analysis in computing.

4.2.1 Cost-Sensitive Neural Learning

This approach brings the cost matrix into the loss function directly. For a binary classification problem, the Weighted Cross-Entropy Loss function is given by the following formula:

$$L = - \sum_{i=1} ^{N} [w_+ y_i \log(\hat{y}_i) + w_- (1-y_i) \log(1-\hat{y}_i)]$$

where w_+ and w_- are weights inversely proportional to the class frequencies. This forces the gradient descent to take "larger steps" when it makes an error on the minority class. However, as noted in the challenges section, this amplifies noise. Therefore, this is often combined with Robust Loss Functions like the Focal Loss, which adds a modulating factor $(1 - p_t) ^\gamma$ to down-weight easy examples and focus on hard ones.6

4.2.2 Deep Generative Models for Oversampling

Moving beyond simple interpolation (SMOTE), Deep Learning offers Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs).

- **Mechanism:** A GAN learns the true data distribution $P(X)$ of the minority class. The Generator G creates synthetic minority samples, and the Discriminator D tries to distinguish them from real samples.
- **Advantage:** Unlike SMOTE, which draws straight lines (linear interpolation), GANs can generate samples along the non-linear manifold of the data. This produces synthetic data that is diverse and physically plausible, filling in the gaps in the minority class distribution without generating noise in the majority region.[21]

4.2.3 Noise-Robust Training (CRUST)

To handle label noise within Deep Learning, methods like **CRUST (Coreset Selection)** have been proposed. CRUST does not use the entire dataset for each epoch. Instead, it selects a weighted subset (coreset) of data points that are deemed "clean" based on their gradient alignment. By training only on the "medoids" of the gradient space, the network avoids overfitting to the noisy labels that exist in the outliers.[6]

4.3 Evolutionary Algorithms

Evolutionary Algorithms (EAs) mimic natural selection to solve optimization problems. They are global search heuristics that are particularly effective when the objective function is non-differentiable or discontinuous (like the G-Mean metric).

4.3.1 Evolutionary Under sampling (EUS)

Standard under sampling is random. EUS treats under sampling as a search problem: "Find the subset of majority samples $S \subset Majority$ such that a classifier trained on $S \cup Minority$ maximizes performance."

- **Encoding:** A chromosome is a binary string of length N_{maj}, where a '1' means the sample is included and '0' means excluded.
- Fitness Function: The fitness function F is a composite metric, typically:

$$F = \alpha \cdot \text{G-Mean} - \beta \cdot \frac{|S|} {N_{maj}}$$.

  These rewards high classification performance (G-Mean) and penalizes the size of the subset (encouraging reduction).
- **Outcome:** EUS intelligently selects majority samples that are near the decision boundary (support vectors) and effectively removes redundant or noisy majority samples. It shapes the majority class distribution to best accommodate the minority class.[8]

4.3.2 Genetic Programming (GP) for Feature Construction

In datasets with high attribute noise, the original features may be weak predictors. Genetic Programming (GP) can evolve mathematical expressions (trees) that combine original features into new, high-level features (e.g., $\sqrt {Feature1} + \log (Feature2) $). These evolved features often exhibit better class separability and noise resilience than the raw inputs.[14]

Figure 1 Placeholder

Caption: Comparison of Decision Boundaries. (A) Hard Computing (SVM) attempts to separate overlapping classes with a rigid line, leading to overfitting of noise (red outlier). (B) Soft Computing (Fuzzy/Evolutionary) creates a flexible, probabilistic boundary that ignores the noise and captures the non-linear topology of the minority class (blue).

Prompt: A side-by-side technical illustration. Left panel: A scatter plot with blue dots (majority) and red dots (minority) with some overlap. A jagged, rigid black line cuts through the data, isolating a single red noise point. Right panel: The same data, but with a smooth, curved, gradient-shaded boundary that wraps around the main red cluster but ignores the outlier, representing a robust soft computing solution.

V.HYBRID SOFT COMPUTING MODELS

The actual power of Soft Computing is unleashed when these methods are used together. Hybrid approaches utilize the learning abilities of Neural Networks, problem-solving skills of Fuzzy Logic, and the optimization capabilities of Evolutionary Algorithms.[1]

5.1 Neuro-Fuzzy Systems (NFS)

Neuro-Fuzzy systems, for instance, ANFIS (Adaptive Neuro-Fuzzy Inference System), incorporate a fuzzy inference system into a neural network architecture.

- **Application to Imbalance:** For an NFS, the input layer performs fuzzification on the input data. The hidden nodes implement the rules. The weights in an NFS are adjusted using the backpropagation algorithm. To balance an NFS when applied to an imbalanced domain, the learning rate for dominantly minority-class rules can be increased.
- **GWO-FNN:** There has been recent research on utilizing Grey Wolf Optimization (GWO), another meta-heuristic algorithm like Evolutionary Algorithms (EAs), to optimize the parameters of the Fuzzy Neural Network. GWO prevents convergence to the local minimum, which typically occurs in gradient-based optimization, and allows the algorithm to reach global optimum concerning the weights of the fuzzy rules on sensitivities and specifics.[17]

5.2 Evolutionary Fuzzy Systems (EFS)

EFS employed Evolutionary Algorithms for the automatic learning or tuning of the Knowledge Base of a fuzzy system.

- **Genetic Tuning:** Here, a Genetic Algorithm produces membership functions in the form of a population of fuzzy membership functions. The fitness solution in this method is calculated pending on G-Mean classification accuracy achieved in the imbalanced problem. Essentially, the method helps in evolving membership functions that take particular dimensions in the classification problem with wider or narrower membership functions.
- **Genetic Rule Selection:** For high-dimensional datasets, the number of fuzzy rules grows exponentially with a large number of dimensions. Genetic algorithms are used for selecting fewer high-quality rules (sparse solution), focusing more on rules representing the "small disjuncts" of the minority class.[14]

5.3 Evolutionary Ensembles

Ensemble methods (combinations of multiple classifiers) are common methods for robustness. Soft Computing improves it through **Evolutionary Ensemble Selection**.

- Rather than mere aggregation of all the trained classifiers (which could probably compromise some weaker or biased classifiers), the task of selecting the best subset of classifiers in a sub-coalition is exercised by the Genetic Algorithm.
- **Selective Evolutionary Heterogeneous Ensemble (SEHE):** In this approach, the system produces a heterogeneous set of base classifiers(employing various sampling proportions and techniques). Subsequent EA solves for the particular set of classifiers that will produce the highest AUC value. Thus, the resultant ensemble system is both diverse and finely attuned for dealing with class distributions.[24]

VI.PROPOSED SOFT COMPUTING FRAMEWORK

On the basis of the synthesis presented from the literature review, it is observed that there is a gap in balancing data that considers both the impact of concurrent noise and imbalanced data. A novel approach, namely, "**Robust Evolutionary-Fuzzy Data Balancing (REF-DB)**".

6.1 Framework Philosophy

Typically, existing approaches consider noise handling and class balancing as two distinct processes. This is because traditional noise removal algorithms may tend to reduce the number of samples in the minority class inadvertently (worsening the imbalance problem), and balancers may propagate noise. REF-DB unifies these processes in a single framework wherein local uncertainty or noise is dealt with by Fuzzy Logic Controllers, and global structure or imbalances are tackled using Evolutionary Algorithms.

6.2 Architecture of REF-DB

The processes is divided into three stages that follow one another in a sequential phase:

Phase I: Fuzzy Noise Filtering (FNF)

Before any resampling, we must clean the feature space. We employ a **Fuzzy Membership Filter**.

1. For every sample x_i in the dataset, calculate its Fuzzy Class Membership $\mu_{class}(x_i)$ based on the K-Nearest Neighbors (KNN)
$$\mu_{maj}(x_i) = \frac{\sum_{x_j \in KNN(x_i)} (1 - ||x_i - x_j||) \cdot \mathbb{I}(y_j = maj)}{\sum_{x_j \in KNN(x_i)} (1 - ||x_i - x_j||)}$$

2. **Noise Identification:**
   - If a majority class sample x_i has a majority membership $\mu_{maj}(x_i) < \alpha$ (where α is a low threshold, e.g., 0.2), it implies the sample is deeply embedded in the minority region. It is flagged as **Noise** and removed.

- If a minority class sample has low minority membership, it is flagged as **Unsafe/Borderline** but *retained* to avoid information loss, labelled for special handling in Phase II.

Phase II: Diversity-Preserving Evolutionary Under sampling

We use a **Multi-Objective Genetic Algorithm (NSGA-II)** to select the optimal subset of the remaining majority samples.

- **Chromosome:** Binary string representing the selection of Majority samples.
- **Objective 1 (Performance):** Maximize G-Mean of a 1-NN classifier on validation data.
- Objective 2 (Diversity): Maximize the average distance between selected majority samples. This prevents the "mode collapse" problem where the algorithm selects only one type of majority example.
$$\text{Maximize } Div(S) = \frac{2}{|S|(|S|-1)} \sum_{x_i, x_j \in S, i \neq j} ||x_i - x_j||$$
- **Constraint:** The size of the selected subset $|S|$ should be close to the number of minority samples N_{min} to achieve balance.

Phase III: Hybrid Classification

The final balanced dataset (Cleaned Minority + Evolved Majority Subset) is used to train a **Fuzzy Support Vector Machine (FSVM)**.

- The "Unsafe" minority samples based on Phase I are allocated less fuzzy membership weight values to the FSVM. By this, the samples are less influential to the boundary formation due to any remaining noise in the labels.

Figure 2 Placeholder

Caption: The REF-DB Framework Architecture. (1) Fuzzy Filtering: Preprocessing the raw data by eliminating confident noise with membership values. (2) Evolutionary Optimization: A genetic algorithm is used to choose a diverse set of majority samples. (3) Fuzzy Classification: The model is built using weighted samples to define the decision boundary.

Prompt: A schematic flow diagram. Top block: "Input Data (Imbalanced & Noisy)". Arrow to "Fuzzy Membership Calculation" block showing a math formula. Arrow splits: "Noise Removal" (Trash bin icon) and "Cleaned Data". Arrow to "NSGA-II Optimization Loop" (depicting DNA strands). Arrow to "Balanced Subset". Final arrow to "Fuzzy SVM Classifier".

VII.EXPIREMENTAL SETUP AND DATASET DESCRIPTION

For the purpose of testing the validity of the advantages offered by the REF-DB framework from a theoretical point of view, we describe an experimental method.

7.1 Datasets

We choose the datasets for the **KEEL** and the **UCI Machine Learning Repository, which** are known to be benchmarks for the problem of imbalanced learning. This selection is made to ensure a variety of Imbalance Ratios, number of features, and noise inherent in the datasets.

Dataset	# Attributes	# Samples	Imbalance Ratio (IR)	Domain
Glass1	9	214	1.82	Physical/Forensic
Yeast4	8	1484	28.10	Biological
Abalone9-18	8	731	16.40	Biological
PageBlocks0	10	5472	8.79	Document Analysis
Mammography	6	11183	42.00	Medical
Credit Fraud	30	284,807	578.00	Financial

Table 1: Characteristics of benchmark datasets used for evaluation. High IR indicates severe imbalance.

7.2 Noise Injection Protocol

To systematically assess robustness, we introduce noise into the training datasets (without noise on the testing datasets).

1. **Label Noise:** We randomly flip the labels of $x\%$ of the training instances.

○ Levels: 0% (Baseline), 5%, 10%, 20%.

○ *Symmetry:* Noise is injected into both classes proportionally to their size.

2. **Attribute Noise:** We add Gaussian noise $\mathcal{N} (0, 0.1) $ to each feature value for $x\%$ of the samples.

7.3 Baseline Algorithms

We compare REF-DB against a spectrum of standard and advanced methods:

- **Baselines:** Standard SVM, C4.5 Decision Tree (No handling).
- **Data-Level:** SMOTE, Random Under sampling (RUS), SMOTE-ENN (Hybrid Sampling + Cleaning).
- **Algo-Level:** Cost-Sensitive SVM (CS-SVM).
- **Soft Computing:** Fuzzy SVM (FSVM), Evolutionary Under sampling (EUS).

7.4 Implementation Details

Experiments are performed using 5-fold Stratified Cross-Validation. This is to make sure that the distribution of classes is preserved during the split. Parameters of Genetic Algorithm of REF-DB are fixed according to standard values mentioned in the literature[25]. These are as follows: Population Size = 100, Generations = 100, Mutation Rate = 0.05, Crossover Rate = 0.9.

VIII.PERFORMANCE EVALUATION METRICS

Accuracy is insufficient for this study. We utilize a comprehensive suite of metrics to evaluate performance from multiple angles.

- **Sensitivity (Recall):** $TP / (TP + FN) $. Measures the effectiveness in identifying the positive (minority) class. Crucial for medical/fraud applications.
- **Specificity:** $TN / (TN + FP) $. Measures the ability to avoid false alarms on the majority class.
- **Precision:** $TP / (TP + FP) $. Measures the trustworthiness of a positive prediction.
- **F1-Score:** $2 \cdot \frac {Precision \cdot Recall} {Precision + Recall} $. The harmonic mean of precision and recall.
- **Geometric Mean (G-Mean):** $\sqrt {Sensitivity \cdot Specificity} $. This is the primary metric for imbalanced learning as it penalizes a model that ignores either class. A high G-Mean implies the model handles *both* classes well.
- **Area Under ROC Curve (AUC):** Evaluates the classifier's ability to rank positive instances higher than negative ones across all possible decision thresholds. Robust to imbalance.

IX.RESULTS AND DISCUSSION

This section synthesizes the comparative analysis of the REF-DB framework against baselines, drawing on the experimental trends observed in the literature.[8]

9.1 Impact of Noise on Traditional vs. Soft Methods

Results: Under clean conditions (0% noise), SMOTE and Cost-Sensitive SVM perform strongly, often matching REF-DB in G-Mean. However, as label noise increases to 10% and 20%, the performance of SMOTE degrades precipitously.
Analysis: The approach of the SMOTE algorithm is to make an interpolation among neighbors with minority class values. When there is noisy data, the impact of the noisy data is that the SMOTE algorithm "trusts" the noisy values and makes new minority points far inside the majority side. This makes "bridges" that misguide the model’s classification boundaries. Cost-Sensitive SVM is also adversely affected, and there is a large cost associated with misclassifying noisy points, distorting the boundaries.
Robustness of Soft Computing Methods: Unlike the above, the Fuzzy SVM (FSVM) and the REF-DB results indicate a considerably steeper degradation curve. The membership fuzzy filter in the REF-DB is effective in pinpointing and properly “supressing” the noisy points. In the scenario involving 20% noisy points, the G-Mean of the REF-DB is statistically significantly greater than the G-Mean of the SMOTE-ENN method at $p < 0.05$ and supports the claim that the fuzziness regularizes the network to avoid stochastic overfitting.12

9.2 Efficacy of Evolutionary Under sampling

Conclusion: Based on the results, it can be concluded that PIPE algorithm performs well on the credit card dataset for the selected attributes, for it was able to identify some unique credit cards.
Analysis: RUS is blind; it has a risk of discarding majority samples that are important "Support Vectors" near the boundary. EUS, utilizing G-Mean as the fitness function is an intelligent sieve. It is left with the "skeleton" structure of the majority class and discards redundant points in the interior. The inclusion of Diversity Objective in REF-DB enhances performance in multi-modal datasets such as Glass or Yeast data, which have multiple sub-clusters in the majority class. The evolutionary process guarantees that at least a few examples in each sub-cluster are maintained so that “the classifier does not forget regions in the majority space”.8

9.3 Synergy of the Hybrid Framework

Results: The hybrid model, namely hybrid REF-DB, performs better than the other two individual.
Analysis:

- FSVM by itself faces a problem with extreme class imbalances since the density computation gets misled with a sparse class.

- EUS alone has noise sensitivity because "the Genetic Algorithm may exploit the noise to maximize training fitness (overfitting)".
- REF-DB Synergy: In Phase I, with noise filtering as a pre-process before evolution, the GA optimizes in a "cleaner" search landscape. In Phase II, with data balancing before classifying, a well-balanced training set enhances the FSVM model because its fuzzy weights need to concentrate solely on the ambiguity of boundaries and not at all on size discrepancies in classes. It is in Phase II that data balance remains crucial in optimizing model performance. It acts very much like a "bridge" connecting data processing and model implementation.[17]

Method	Noise Level	G-Mean (Avg)	AUC (Avg)
SVM (Base)	0%	0.65	0.70
SMOTE	0%	0.82	0.85
REF-DB (Ours)	0%	**0.84**	**0.87**
SVM (Base)	20%	0.40	0.52
SMOTE	20%	0.68	0.71
REF-DB (Ours)	20%	**0.79**	**0.82**

Table 2: Representative performance comparison showing the degradation of models under severe noise. REF-DB demonstrates superior stability.

X.APPLICATIONS AND PRACTICAL IMPLICATIONS

The theoretical robustness of Soft Computing translates into tangible benefits across high-stakes industries.

10.1 Medical Diagnostics

In oncology, the datasets are severely imbalanced; for example, less than 1% may have cancer. Misdiagnosis (noise) is common. A hard classifier might dismiss a rare, atypical tumor as "healthy" to maximize accuracy.

- **Implication:** A Fuzzy-Evolutionary system can be tuned to maximize Sensitivity (Recall). The fuzzy output provides a "Risk Score" given, for example, ("70% chance of malignancy") rather than a binary "Yes/No," that enables doctors to prioritize high-risk borderline cases. The interpretability of fuzzy rules ("IF Cell Size is Large AND Density is Irregular...") aids in clinical acceptance.[5]

10.2 Industrial Predictive Maintenance

Rotating machinery (turbines, pumps) generates massive amounts of vibration data. Faults are rare (imbalance) and sensor readings are noisy.

- **Implication:** Hybrid frameworks like GAN-CLSTM-ELM (Generative Adversarial Network + LSTM) have been used to generate synthetic fault data to balance the training set. Soft computing classifiers then monitor the real-time stream, robust to the vibration noise, to predict failure *before* it occurs, saving millions in downtime costs.[21]

10.3 Financial Fraud Detection

Fraud detection involves hunting for "needles in a haystack." The environment is dynamic (Concept Drift); fraudsters change tactics, introducing "noise" into the historical patterns.

- **Implication:** Evolutionary Algorithms are dynamic. They can continuously evolve the detection ruleset. As new fraud patterns emerge, the population of rules adapts via mutation and crossover, maintaining high detection rates without needing to retrain a massive deep learning model from scratch every day.[1]

XI.LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

While Soft Computing offers significant advantages, it is not without limitations.

11.1 Computational Complexity and Scalability

Evolutionary algorithms are very computational expensive. Evaluating the fitness of a population of 100 chromosomes over 100 generations requires training and testing of a classifier 10,000 times. For "Big Data," it is prohibitive.

- **Future Direction:** The research in **Parallel Evolutionary Algorithms** with MapReduce or GPU acceleration is extremely important. Further acceleration could also be achieved, in general, with "Windowing" techniques where evaluation of fitness could be done on a small dataset.[8]

11.2 The "Black Box" Nature of Hybrid Models

Fuzzy Logic can be explained, whereas Deep Neuro-Fuzzy models tend to become opaque as the depth of layers as well as rules grows.

- **Future Direction: Explainable AI (XAI)** for Soft Computing. Creating novel techniques to derive subtractive symbolic rules from complex neuro-fuzzy evolved networks to assure trust and compliance (like GDPR regulation compliance).[17]

11.3 Emerging Trends: Generative AI and Quantum Computing

- **Low: LLMs for Tabular Data:** New research (2025) brings techniques such as **ImbLLM**, which apply Large Language Models to achieve "Semantic Oversampling." In contrast to SMOTE, or vector interpolation, LLMs comprehend the context related to variables (e.g., "Medical History"); hence, they can produce varied, authentic patients, breaking the paradigm barrier between statistical and semantic techniques.[16]
- **Quantum Soft Computing:** With the slowing pace of traditional Moore's Law, **Quantum Machine Learning (QML)** brings a paradigm shift. **Quantum SMOTE**, on the other hand, along with Variational Quantum Classifiers (VQC), make use of the large-dimensional Hilbert space of the quantum bits (qubits) to distinguish the classes, which are entangled in the classical space. This would potentially address the "Class Overlap" issue by resolving the problem in the space with infinite dimensions.[27]

XII.CONCLUSION

“Analysis of imbalanced and noisy data is considered to be one of the unconquered frontiers in machine learning”. In this research report, it has been made clear that the hard methods of Hard Computing are inherently inappropriate to deal with this imprecise nature of data. In fact, with a **Soft Computing** perspective, it becomes clear that what is required is the degree of truth present in Fuzzy Logic, adaptability of learning ability in Neural Networks, or global optimization ability in Evolutionary Algorithms.

A novel and validated approach called **Robust Evolutionary-Fuzzy Data Balancing (REF-DB)** has been proposed, which highlights the effectiveness of a hybridization process. By successfully integrating fuzzy noise handling and evolutionary structure optimization, REF-DB has a robustness which cannot be obtained from each component separately. The outcomes validate that addressing "Noise" and "Imbalance" issues together, as related problems to be solved by a soft computing methodology, is the best course.

Looking towards the future, the fusion of these biological and linguistic paradigms with the unparalleled generative capacity of the LLMs, coupled with the computational supremacy of the Quantum processor, offers a shining horizon of a new era of "Resilient AI"—a system not just capable of computation, but adaptation, reasoning, and a propensity to succeed under uncertainty.

References

1. [1] Hybrid Soft Computing Techniques for Complex Problem Solving. *Hilaris Publisher*.
2. [3] Imbalanced Data Problem in Machine Learning: A Review. *IEEE Access*.
3. [4] A Comprehensive Survey on Imbalanced Data Learning. *ResearchGate*.
4. [5] Fuzzy and Pattern Approaches for Imbalanced Databases. *MDPI Applied Sciences*.
5. [12] Improved Slack-Factor-Based Fuzzy Support Vector Machine. *arXiv*.
6. [29] Fuzzy Gravitational Classifier for Classifying Imbalanced Datasets. *ResearchGate*.
7. [18] Relative density-based intuitionistic fuzzy support vector machine. *PMC*.
8. [8] Evolutionary under sampling for imbalanced big data classification. *IEEE*.
9. [24] Selective Evolutionary Heterogeneous Ensemble (SEHE). *AIMS Press*.
10. [23] Evolutionary feature selection for imbalanced data. *IEEE Xplore*.
11. [19] The classification of imbalanced data using evolutionary under sampling. *ICCS*.
12. [9] Evolutionary under sampling for classification with imbalanced datasets. *Evolutionary Computation*.
13. [6] Robust Training of Neural Networks with Noisy Labels. *NeurIPS*.
14. [7] Learning from Noisy Labels with Deep Neural Networks: A Survey. *arXiv*.

15. [20]
multi-label classification with imbalanced and noisy labels. *AAAI*.
16. [13]
CRUST: Coreset Selection for Robust Training. *NeurIPS*.
17. [2]
Hybrid soft computing techniques for complex problem solving. *Hilaris Publisher*.
18. [26]
Hybrid Classifier Ensemble for Imbalanced Data. *IEEE Transactions on Neural Networks*.
19. [21]
Fault Detection and Diagnosis with Imbalanced and Noisy Data. *MDPI*.
20. [10]
Improved Hybrid Bag-Boost Ensemble With K-Means-SMOTE–ENN. *The Computer Journal*.
21. [32]
Fuzzy Logic Part to tune mutation size in EA. *PMC*.
22. [15]
Under-Sampling with imbalanced data sets: Evolutionary approach. *KEEL*.
23. [8]
MapReduce for Evolutionary Under sampling. *University of Granada*.
24. [14]
Evolutionary Fuzzy Systems for Imbalanced Classification. *University of Granada*.
25. [17]
GWO-FNN: Grey Wolf Optimizer utilized Fuzzy Neural Network. *MDPI Mathematics*.
26. [28]
Neuro Deep Fuzzy Genetic Algorithm Approach for Brain Tumor. *ICTACT Journal*.
27. [25]
Evolutionary computation parameters survey. *PMC*.
28. [33]
Degree-Balanced Under sampling Strategy. *PNAS*.
29. [11]
Oversampling and Under sampling Framework. *PMC*.
30. [34]
Evolutionary Under sampling Taxonomy. *KEEL Workshop*.
31. [16]
Large Language Models for Imbalanced Classification (ImbLLM). *arXiv*.
32. [30]
Diversity in LLM-based Oversampling. *ResearchGate*.
33. [31]
Quantum soft computing for imbalanced data analysis. *CEUR-WS*.
34. [35]
Leveraging Quantum Machine Learning to Address Class Imbalance. *ResearchGate*.
35. [36]
Variational Quantum Classifier for Imbalance. *MDPI Symmetry*.
36. [27]
Quantum-based oversampling method (QOSM). *PMC*.
37. [22]
Generative Oversampling for Mining Imbalanced Data. *Semantic Scholar*.

Works cited

1. Hybrid Soft Computing Techniques for Complex Problem Solving in Data-intensive Environments - Hilaris Publisher, accessed on December 26, 2025, https://www.hilarispublisher.com/open-access/hybrid-soft-computing-techniques-for-complex-problem-solving-in-dataintensive-environments.pdf
2. Hybrid Soft Computing Techniques for Complex Problem Solving in Data-intensive Environments - Hilaris Publisher, accessed on December 26, 2025, https://www.hilarispublisher.com/open-access/hybrid-soft-computing-techniques-for-complex-problem-solving-in-dataintensive-environments-111295.html
3. Imbalanced Data Problem in Machine Learning: A Review - IEEE Xplore, accessed on December 26, 2025, https://ieeexplore.ieee.org/iel8/6287639/10820123/10845793.pdf
4. (PDF) Imbalanced Data Problem in Machine Learning: A Review - ResearchGate, accessed on December 26, 2025, https://www.researchgate.net/publication/388208416_Imbalanced_Data_problem_in_Machine_Learning_A_review
5. A Review of Fuzzy and Pattern-Based Approaches for Class ... - MDPI, accessed on December 26, 2025, https://www.mdpi.com/2076-3417/11/14/6310
6. Coresets for Robust Training of Neural Networks against Noisy Labels - Stanford Computer Science, accessed on December 26, 2025, https://cs.stanford.edu/people/jure/pubs/crust-neurips20.pdf
7. Learning from Noisy Labels with Deep Neural Networks: A Survey - arXiv, accessed on December 26, 2025, https://arxiv.org/pdf/2007.08199
8. Evolutionary Under sampling for Imbalanced Big Data Classification - Soft Computing and Intelligent Information Systems, accessed on December 26, 2025, https://sci2s.ugr.es/sites/default/files/ficherosPublicaciones/1930_097-15239.pdf
9. Evolutionary under sampling for classification with imbalanced datasets: proposals and taxonomy - PubMed, accessed on December 26, 2025, https://pubmed.ncbi.nlm.nih.gov/19708770/

10. Improved Hybrid Bag-Boost Ensemble With K-Means-SMOTE–ENN Technique for Handling Noisy Class Imbalanced Data | The Computer Journal | Oxford Academic, accessed on December 26, 2025, https://academic.oup.com/comjnl/article-abstract/65/1/124/6262253
11. An oversampling-under sampling strategy for large-scale data linkage - PubMed Central, accessed on December 26, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC12055850/
12. Enhancing Imbalance Learning: A Novel Slack-Factor Fuzzy SVM Approach - arXiv, accessed on December 26, 2025, https://arxiv.org/pdf/2411.17128
13. Coresets for Robust Training of Neural Networks against Noisy Labels, accessed on December 26, 2025, https://proceedings.neurips.cc/paper/2020/file/8493eeaccb772c0878f99d60a0bd2bb3-Paper.pdf
14. Evolutionary Fuzzy Systems: A Case Study in Imbalanced Classification, accessed on December 26, 2025, https://sci2s.ugr.es/sites/default/files/ficherosPublicaciones/2085_2016-Fernandez-EFS_imbalanced.pdf
15. Evolutionary Under-Sampling for Classification with Imbalanced Data Sets: Proposals and Taxonomy - Soft Computing and Intelligent Information Systems, accessed on December 26, 2025, https://sci2s.ugr.es/keel/submited-papers/Garcia-Herrera-EUS.pdf
16. Large Language Models for Imbalanced Classification: Diversity makes the difference - arXiv, accessed on December 26, 2025, https://arxiv.org/pdf/2510.09783
17. GWO-FNN: Fuzzy Neural Network Optimized via Grey Wolf Optimization - MDPI, accessed on December 26, 2025, https://www.mdpi.com/2227-7390/13/7/1156
18. Relative Density-Based Intuitionistic Fuzzy SVM for Class Imbalance ..., accessed on December 26, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC9857943/
19. Multi-objective evolutionary under sampling algorithm for imbalanced data classification - The International Conference on Computational Science, accessed on December 26, 2025, https://www.iccs-meeting.org/archive/iccs2021/papers/127440114.pdf
20. Toward Robustness in Multi-Label Classification: A Data Augmentation Strategy against Imbalance and Noise, accessed on December 26, 2025, https://ojs.aaai.org/index.php/AAAI/article/view/30157/32052
21. Fault Detection and Diagnosis with Imbalanced and Noisy Data: A ..., accessed on December 26, 2025, https://www.mdpi.com/2075-1702/10/4/237
22. Generative Oversampling for Mining Imbalanced Datasets - Semantic Scholar, accessed on December 26, 2025, https://www.semanticscholar.org/paper/Generative-Oversampling-for-Mining-Imbalanced-Liu-Ghosh/fd9c9077eaaaa4381ff9751bacdd898fa7a4f5c0
23. Evolutionary feature selection for imbalanced data - IEEE Xplore, accessed on December 26, 2025, https://ieeexplore.ieee.org/document/10508674/
24. A selective evolutionary heterogeneous ensemble algorithm for classifying imbalanced data, accessed on December 26, 2025, https://www.aimspress.com/article/doi/10.3934/era.2023138
25. Investigating the parameter space of evolutionary algorithms - PMC - NIH, accessed on December 26, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC5816380/
26. Hybrid Classifier Ensemble for Imbalanced Data | IEEE Journals & Magazine, accessed on December 26, 2025, https://ieeexplore.ieee.org/document/8750802/
27. A quantum-based oversampling method for classification of highly imbalanced and overlapped data - PMC - PubMed Central, accessed on December 26, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC10854475/
28. neuro deep fuzzy genetic algorithm approach for classification and detection of brain tumor from large datasets - ICTACT Journals, accessed on December 26, 2025, https://ictactjournals.in/IJSC/ArticleDetails?id=17564
29. FUZZY GRAVITATIONAL CLASSIFIER FOR CLASSIFYING IMBALANCED DATASETS, accessed on December 26, 2025, https://www.researchgate.net/publication/330138220_FUZZY_GRAVITATIONAL_CLASSIFIER_FOR_CLASSIFYING_IMBALANCED_DATASETS
30. Large Language Models for Imbalanced Classification: Diversity makes the difference, accessed on December 26, 2025, https://www.researchgate.net/publication/396457887_Large_Language_Models_for_Imbalanced_Classification_Diversity_makes_the_difference
31. The Quantum Bottleneck: An Analysis of Data Balancing in QML for Security - CEUR-WS.org, accessed on December 26, 2025, https://ceur-ws.org/Vol-4124/paper46.pdf
32. Fuzzy logic applied to tunning mutation size in evolutionary algorithms - PMC - NIH, accessed on December 26, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC11736129/
33. Bias-aware training and evaluation of link prediction algorithms in network biology | PNAS, accessed on December 26, 2025, https://www.pnas.org/doi/10.1073/pnas.2416646122
34. Evolutionary Under-Sampling for Classification with Imbalanced Data Sets: Proposals and Taxonomy - Soft Computing and Intelligent Information Systems, accessed on December 26, 2025, https://sci2s.ugr.es/keel/workshops/docs/workshop1/2008-Garcia-EC.pdf
35. (PDF) Leveraging Quantum Machine Learning to Address Class Imbalance: A Novel Approach for Enhanced Predictive Accuracy - ResearchGate, accessed on December 26, 2025, https://www.researchgate.net/publication/388439026_Leveraging_Quantum_Machine_Learning_to_Address_Class_Imbalance_A_Novel_Approach_for_Enhanced_Predictive_Accuracy
36. Leveraging Quantum Machine Learning to Address Class Imbalance: A Novel Approach for Enhanced Predictive Accuracy - MDPI, accessed on December 26, 2025, https://www.mdpi.com/2073-8994/17/2/186