# RoBell-RVFL: A Robust Generalized Bell Random Vector Functional Link Network

A. Rahaman
*Department of Mathematics*
*Indian Institute of Technology Indore*
Indore, India
phd2401141001@iiti.ac.in

A. Quadir
*Department of Mathematics*
*Indian Institute of Technology Indore*
Indore, India
mscphd2207141002@iiti.ac.in

M. Tanveer
*Department of Mathematics*
*Indian Institute of Technology Indore*
Indore, India
mtanveer@iiti.ac.in

***Abstract*—The dominance of majority classes in real-world datasets poses a fundamental challenge to randomized neural networks, often biasing decision boundaries and overlooking critical minority samples. Existing remedies, such as synthetic minority over-sampling (SMOTE) and class-weighted loss functions, primarily address class proportions while neglecting intra-class distribution, making them vulnerable to label noise and outliers. In this paper, we propose RoBell-RVFL, a robust and lightweight *quality-aware* generalized bell random vector functional link network that redefines how randomized models handle class imbalance and noisy data. RoBell-RVFL employs a dual-strategy, sample-level weighting mechanism that strictly preserves minority class information using unit weights, while adaptively regulating the influence of majority class samples through a probability-weighted generalized bell (gbell) membership function in a kernel-induced feature space. This design effectively suppresses noisy, boundary, and outlier samples within the majority class, enabling the network to learn from informative samples rather than merely abundant ones. By explicitly incorporating local class probability and class distribution information into the learning process, RoBell-RVFL achieves adaptive control over sample contributions without sacrificing the closed-form learning efficiency of RVFL networks. Extensive evaluations on UCI and KEEL benchmark datasets, along with robustness tests under up to 40% label noise, demonstrate that RoBell-RVFL consistently and significantly outperforms recent state-of-the-art RVFL variants. The results indicate that adaptive, quality-aware sample weighting is essential for robust RVFL learning, rendering conventional global weighting schemes ineffective in noisy and imbalanced environments. The supplementary material is available in an anonymous repository at https://github.com/AnonymousAuthor0011/RoBell-RVFL.**

***Index Terms*—Randomized neural networks (RdNNs), Random vector functional link network(RVFL), Robustness, Bell function.**

## I. Introduction and Motivation

THE random vector functional link (RVFL) network [1] is a prominent randomized neural network (RdNN) model that has gained considerable attention due to its architectural simplicity and high computational efficiency. The defining characteristic of RVFL lies in its learning paradigm, in which the weights connecting the input and hidden layers are randomly initialized and kept fixed throughout training, while only the output-layer weights are analytically learned via a closed-form solution using ridge regression [2]. By eliminating iterative weight updates and gradient-based optimization, RVFL enables extremely fast training with low computational overhead, while ensuring stable and reproducible learning behavior.

Moreover, RVFL adopts a shallow architecture with a single hidden layer and direct links between the input and output layers, enabling the preservation of raw input information and improved generalization. Owing to its closed-form learning of output weights and compact network structure with few trainable parameters, RVFL achieves rapid training and low computational complexity [3]. Guided by principles such as Occam's razor and probably approximately correct (PAC) learning theory [4], RVFL can be viewed as a lightweight yet effective alternative to conventional artificial neural networks.

Motivated by its architectural simplicity and analytically tractable learning framework, numerous RVFL-based variants have been proposed to further enhance predictive accuracy and robustness across diverse application scenarios [5]. Owing to its direct input-to-output connections and efficient information flow, RVFL exhibits strong generalization capability with high learning efficiency.

Recent studies have sought to enhance the representational power, robustness, and adaptability of RVFL networks through diverse architectural refinements and learning strategies. For instance, total-var-RVFL and class-var-RVFL [6] exploit statistical dispersion in both the input and randomized feature spaces to enable more effective estimation of output-layer weights. The RVFL+ model [7], developed under the learning using privileged information (LUPI) paradigm, leverages auxiliary training-only information to guide learning and improve convergence and generalization; its kernelized extension, KRVFL+, further enhances nonlinear modeling capability through the kernel trick. Another prominent research direction integrates fuzzy logic into the RVFL framework to better handle uncertainty and imprecision. The neuro-fuzzy RVFL (NF-RVFL) [8] embeds fuzzy inference within RVFL, improving robustness and interpretability.

More recently, complex-valued RVFL (CRVFL) [9] and its augmented extension (ACRVFL) have been introduced to address challenges related to data representation, incremental learning, and computational efficiency. Additionally, granular and graph-based RVFL variants, including GB-RVFL, GE-GB-RVFL [10], and the multiview graph-based RVFL (GRVFL-

MV) [11], have demonstrated improved discriminative capability by incorporating geometric and multiview information. Despite these advances, a fundamental challenge persists in real-world pattern classification: the presence of severely imbalanced data, often compounded by noise and outliers. In the real world, from face recognition systems to medical diagnostics [12], data is rarely perfect. We are frequently confronted with a majority class that overwhelms a minority class, often cluttered with noise and misleading outliers. This isn't just an inconvenience; it is a fundamental flaw in how standard models learn. Conventional RVFL networks are democratic to a fault, they treat every sample equally. When faced with a massive majority class, the network effectively ignores the minority, creating a biased decision boundary that renders the model useless for the cases that often matter most.

To date, existing approaches for addressing class imbalance can be broadly categorized into two groups: data-level and algorithm-level methods. Data-level approaches attempt to rebalance the training set through majority-class undersampling or minority-class oversampling techniques, such as SMOTE [13]. Although widely adopted, these methods suffer from inherent limitations: undersampling may discard informative samples, while oversampling can introduce redundancy and amplify noise [14].

Algorithm-level approaches, on the other hand, modify the learning mechanism itself. Within the RVFL family, several notable efforts have been reported. CSWRVFL [15] addresses power quality disturbance classification, and GE-IFRVFL-CIL [16] incorporates class-dependent weighting while preserving the topological structure of the data.

Despite these advances, a critical limitation remains. Existing methods primarily address the *quantity* of class imbalance, while largely overlooking the *quality* of training samples. In particular, they lack mechanisms to identify and suppress noisy or outlier samples within the majority class, while simultaneously preserving the discriminative information of minority samples [17]. Consequently, the majority-class noise continues to bias decision boundaries in imbalanced learning scenarios.

To address this gap, we propose the Robust Generalized Bell RVFL (RoBell-RVFL), which performs fine-grained, sample-level weighting rather than coarse class-level reweighting. The core innovation of RoBell-RVFL lies in a probability-weighted generalized bell (gbell) membership function [18], which integrates spatial distribution information with local class probability. Specifically, minority class samples are assigned unit membership values to ensure full information preservation, whereas majority class samples are adaptively weighted based on their gbell membership and local class probability. Samples located near cluster centers retain higher influence, while boundary, noisy, and outlier samples are effectively suppressed. By coupling this adaptive membership mechanism with a principled class-imbalance weighting scheme [19, 20], RoBell-RVFL enables the network to learn from informative samples rather than merely abundant ones.

The main contributions of this work are summarized as follows:

1) Adaptive Sample-Level Framework: We propose a dual-strategy learning framework that preserves minority class samples using unit weights, while adaptively reducing the influence of majority class samples according to their distribution characteristics.
2) Probability-Weighted Noise Suppression: A class probability–guided generalized bell (gbell) membership function is introduced to distinguish informative samples from noise and outliers, thereby preventing majority-class dominance in the learned decision boundary.
3) Extensive Empirical Validation: Comprehensive experiments on UCI and KEEL benchmark datasets, including evaluations under up to 40% label noise, demonstrate that RoBell-RVFL consistently outperforms state-of-the-art RVFL-based and imbalance-learning methods.

The remainder of this paper is organized as follows. Section III presents the mathematical formulation of RoBell-RVFL and its class-imbalance weighting mechanism. Related work, experimental setup, and experimental results on UCI and KEEL without label noise are reported in Section S.I., S.IV, and S.VII. of Supplementary Material. Finally, Section V concludes the paper.

## II. Notation

Consider the training dataset, denoted as $(a_i, b_i)$, where $a_i \in \mathbb{R}^{1\times B}$ and $b_i \in \mathbb{R}^{1\times D}$, for $i = 1, 2, \ldots, A$. Here, $B$ represents the number of features in each input sample, and $D$ is the total number of classes in the dataset, while $A$ indicates the total number of training samples. In the context of classification, $b_i$ refers to a one-hot encoded vector that signifies the class label corresponding to the input sample $a_i$. We define the collections of all input samples and target values as follows:

$$Z = \begin{bmatrix} a_1^t & a_2^t & \ldots & a_M^t \end{bmatrix}^t \in \mathbb{R}^{A\times B}, \quad Z_{out} = \begin{bmatrix} b_1^t & b_2^t & \ldots & b_M^t \end{bmatrix}^t \in \mathbb{R}^{A\times E},$$

where the operator $(\cdot)^t$ represents the transpose. These matrices, $Z$ and $Z_{out}$, encapsulate all input samples and their corresponding target values, respectively. $D$ is the number of hidden nodes.

## III. Proposed Method

Assume that the training dataset consists of $Z$ labeled samples given by $\{(a_1, b_1), (a_2, b_2), \ldots, (a_A, b_A)\}$, where $a_i \in \mathbb{R}^B$ denotes the $B$-dimensional feature vector of the $i^{th}$ sample and $b_i \in \{+1, -1\}$ represents its corresponding class label. Let $h_1 \in \mathbb{R}^{A_1\times B}$ and $h_2 \in \mathbb{R}^{A_2\times B}$ denote the matrices formed by the samples belonging to the positive $(+1)$ and negative $(-1)$ classes, respectively. Without loss of generality, it is assumed that the positive class constitutes the majority class $(A_1 \geq A_2)$, while the negative class represents the minority class. A sample $a_i$ belonging to the majority class is denoted by $a_i^m$, whereas a sample belonging to the minority class is denoted by $a_{i_m}$.

In the proposed framework, explicit use of the imbalance ratio is avoided at the sample-weight definition stage. Instead, the effect of class imbalance is addressed through adaptive sample weighting and class-dependent regularization in the learning phase. Furthermore, the data samples are implicitly mapped into a higher-dimensional feature space through a kernel-induced mapping, denoted by $K(\cdot)$, which enables effective modeling of nonlinear class distributions.

### *A. Determination of Membership Value*

The generalized bell (gbell) function is a flexible fuzzy membership function widely used in fuzzy systems due to its smooth controllability [18]. In the proposed RoBell-RVFL framework, the gbell function assigns adaptive membership values to training samples based on their distribution in a kernel-induced feature space. Specifically, it down-weights outliers and boundary samples from the majority class while preserving full membership for minority class samples to avoid information loss. Unlike conventional distance-based formulations, gbell memberships are computed using a kernel function, enabling implicit mapping into a higher-dimensional space. Let $K(\cdot,\cdot)$ denote a Gaussian kernel defined as $K(a_i,a_j)=\exp\left(-\frac{\|a_i-a_j\|^2}{\xi^2}\right)$, where $\xi$ is the kernel width parameter. The parameter $\xi$ controls the width of the Gaussian kernel and determines the scale of similarity in the kernel-induced feature space. Smaller values of $\xi$ emphasize local neighborhood structures, whereas larger values lead to smoother similarity distributions.

For the positive (majority) and negative (minority) class samples, the kernel-induced radii are computed as

$$r_i^{(+)}=\sqrt{1-2\frac{1}{t_1}\sum_{a_j\in T_1}K(a_i,a_j)+\frac{1}{t_1^2}\sum_{a_p,a_q\in T_1}K(a_p,a_q)}, \tag{1}$$

$$r_i^{(-)}=\sqrt{1-2\frac{1}{t_2}\sum_{a_j\in T_2}K(a_i,a_j)+\frac{1}{t_2^2}\sum_{a_p,a_q\in T_2}K(a_p,a_q)}. \tag{2}$$

Let

$$r_{\max}^{(+)}=\max_i r_i^{(+)}, \qquad r_{\max}^{(-)}=\max_i r_i^{(-)}. \tag{3}$$

The gbell membership functions for the positive and negative class samples are defined as

$$\xi_{\text{pw}}^{(+)}(a_i)=\frac{1}{1+\left(\frac{r_i^{(+)}}{c_1}\right)^{2d_1}}, \qquad c_1=d_1=\rho\, r_{\max}^{(+)}, \tag{4}$$

$$\xi_{\text{pw}}^{(-)}(a_i)=\frac{1}{1+\left(\frac{r_i^{(-)}}{c_2}\right)^{2d_2}}, \qquad c_2=d_2=\rho\, r_{\max}^{(-)}, \tag{5}$$

where $\rho>0$ is a scaling parameter controlling the effective width and steepness of the gbell function.

*1) Determination of Class Probability:* To further reduce the influence of noise and ambiguous samples during training, a class probability value is assigned to each training sample. This probability reflects the local class consistency of a sample within a kernel-induced neighborhood and is used to suppress noisy and mislabeled samples, particularly in the majority class. Unlike conventional approaches that employ a fixed-radius neighborhood in the input space, the proposed method computes the class probability using a kernel-based distance measure. The kernel-induced distance between two samples $a_i$ and $a_j$ is defined as $d(a_i,a_j)=\sqrt{2\big(1-K(a_i,a_j)\big)}$.

Let

$$\alpha_d=\max\Big(r_{\max}^{(+)},\, r_{\max}^{(-)}\Big) \tag{6}$$

denote a data-dependent neighborhood radius, where $r_{\max}^{(+)}$ and $r_{\max}^{(-)}$ are the maximum kernel-induced radii of the positive and negative class samples, respectively.

For a given sample $a_i$, the local neighborhood is defined as $\mathcal{N}(a_i)=\{a_j \mid d(a_i,a_j)\le\alpha_d\}$. The class probability of $a_i$ is then computed as the ratio of samples belonging to the same class within this neighborhood:

$$p(a_i)=\frac{|\{a_j\in\mathcal{N}(a_i)\ :\ b_j=b_k\}|}{|\mathcal{N}(a_i)|+\varepsilon}, \tag{7}$$

where $\varepsilon$ is a small positive constant introduced to avoid numerical instability.

A higher value of $p(a_i)$ indicates stronger local consistency of the sample with its class, whereas samples with low probability values are more likely to be noisy or located near class boundaries. Consequently, such samples are assigned lower influence during training, thereby improving the robustness of the proposed RoBell-RVFL model.

*2) PW-GB Membership Function:* By integrating the generalized bell (gbell) membership values with the local class probability, we propose a probability-weighted generalized bell (PW-GB) membership function. The PW-GB function is designed to suppress the influence of noisy and boundary samples in the majority class while preserving the full contribution of minority class samples.

The PW-GB membership function is defined as follows:

$$\xi(a_i)=\begin{cases}1, & \text{if } a_i=a_{i_m},\\ \xi_{\text{pw}}(a_i)\,p(a_i), & \text{if } a_i=a_i^m.\end{cases} \tag{8}$$

Here, $\xi_{\text{pw}}(a_i)$ represents the gbell membership value computed in the kernel-induced feature space, and $p(a_i)$ denotes the local class probability of the sample.

### *B. Class Imbalance Weighting Scheme*

[19] The class-dependent weight $l_+$ and $l_-/$ is defined as

$$l_-=1,\ \text{if } a_i \text{ is in the class of negatives}, \tag{9}$$

$$l_+=\frac{A_2}{A_1}, \quad \text{if } a_i \text{ is in the class of positives}. \tag{10}$$

In the adopted weighting strategy, all samples belonging to the minority class are given a weight of one, whereas the samples from the majority class are down-weighted proportionally according to the ratio between the number of minority and majority instances.

## IV. PROPOSED ROBELL-RVFL

This section presents a comprehensive formulation of the proposed RoBell-RVFL model. We begin by establishing a general mathematical framework designed to effectively manage class imbalance (CI). To mitigate the adverse effects of imbalanced data, the RoBell-RVFL model integrates a CI-aware weighting mechanism that adjusts sample contributions according to the imbalance ratio. Based on this formulation, the corresponding optimization problem of the proposed RoBell-RVFL model is expressed as follows:

$$\begin{aligned} min_V \ & \frac{1}{2}\|V\|_2^2 + \frac{\gamma}{2}(l_+)\left\|L_+^{\frac{1}{2}}\beta_+\right\|_2^2 \\ & + \frac{\gamma}{2}(l_-)\left\|L_-^{\frac{1}{2}}\beta_-\right\|_2^2 \\ \text{s.t.} \quad & X_+V = Y_+ - \beta_+, \\ & X_-V = Y_- - \beta_-. \end{aligned} \tag{11}$$

Here, $V$ denotes the weight matrix that links both the input and hidden layers to the output layer. $L-$ and $L+$ are diagonal matrices whose diagonal elements correspond to the PW-GB membership values of the negative and positive class samples, respectively.

$L_+ = \text{diag}[\xi(a_1^m), \xi(a_2^m), \cdots, \xi(a_{k+}^m)$, where $1 \leq i+ \leq A_1$ and $L_- = \text{diag}[\xi(a_{1_m}), \xi(a_{2_m}), \cdots, \xi(a_{i-_m})$, where $1 \leq i- \leq A_2$.

$\beta_-$ ($\beta_+$) denotes the error vector of the negative (positive) class samples. Here, $l_-$ ($l_+$) represents the weighting scheme of the negative (positive) class defined in Section III-B. $\gamma \in \mathbb{R}^+$ is the regularization parameter used to penalize the error variables. $X_-$ ($X_+$) denotes the nonlinear as well as linear projection of the negative (positive) class samples, defined as:

$$X_- = [Z_- \ Y_-], \quad X_+ = [Z_+ \ Y_+]. \tag{12}$$

In this formulation, $Z-$ ($Z+$) denotes the input data matrix corresponding to the negative (positive) class, while $Y-$ ($Y+$) represents the hidden-layer output matrix for the negative (positive) class. These hidden representations are obtained by projecting $Z-$ ($Z+$) through randomly initialized weights and biases, followed by the application of the nonlinear activation function $\phi$. The matrices $Y-$ and $Y+$ also serve as the target outputs for the negative and positive class samples, respectively.

Consequently, the incorporation of CI-aware weighting matrices ($l_+$ and $l_-$) enables the proposed model to regulate the influence of majority and minority class samples.

The corresponding Lagrangian formulation of (11) is given by:

$$\begin{aligned} \mathcal{L} = & \frac{1}{2}\|V\|_2^2 + \frac{\gamma}{2}(l_+)\left\|L_+^{\frac{1}{2}}\beta_+\right\|_2^2 + \frac{\gamma}{2}(l_-)\left\|L_-^{\frac{1}{2}}\beta_-\right\|_2^2 \\ & - \alpha_+^\top(X_+V - Y_+ + \beta_+) - \alpha_-^\top(X_-V - Y_- + \beta_-), \end{aligned} \tag{13}$$

where $\alpha-$ and $\alpha+$ denote the Lagrange multipliers. Let

$$F_+ = \gamma \times l_+, \quad F_- = \gamma \times l_-. \tag{14}$$

Rewriting (13), we obtain:

$$\begin{aligned} \mathcal{L} = & \frac{1}{2}\|V\|_2^2 + \frac{F_+}{2}\left\|L_+^{\frac{1}{2}}\beta_+\right\|_2^2 + \frac{F_-}{2}\left\|L_-^{\frac{1}{2}}\beta_-\right\|_2^2 \\ & - \begin{bmatrix}\alpha_+^\top & \alpha_-^\top\end{bmatrix}\left(\begin{bmatrix}X_+ \\ X_-\end{bmatrix}V - \begin{bmatrix}Y_+ \\ Y_-\end{bmatrix} + \begin{bmatrix}\beta_+ \\ \beta_-\end{bmatrix}\right). \end{aligned} \tag{15}$$

By applying the Karush–Kuhn–Tucker (KKT) conditions to (15), we obtain:

$$V - X_+^\top\alpha_+ - X_-^\top\alpha_- = 0, \tag{16}$$

$$F_+L_+\beta_+ - \alpha_+ = 0, \tag{17}$$

$$F_-L_-\beta_- - \alpha_- = 0, \tag{18}$$

$$X_+V - Y_+ + \beta_+ = 0, \tag{19}$$

$$X_-V - Y_- + \beta_- = 0. \tag{20}$$

Rewriting (16)–(18), we have

$$IV = X_+^\top\alpha_+ + X_-^\top\alpha_-, \tag{21}$$

$$F_+L_+\beta_+ = \alpha_+, \tag{22}$$

$$F_-L_-\beta_- = \alpha_-, \tag{23}$$

where $I$ is an identity matrix of conformal dimension. Using (22) and (23) in (21), we obtain

$$IV = F_+X_+^\top L_+\beta_+ + F_-X_-^\top L_-\beta_-, \tag{24}$$

$$IV = F_+X_+^\top L_+(Y_+ - X_+V) + F_-X_-^\top L_-(Y_- - X_-V), \tag{25}$$

$$(I + F_+X_+^\top L_+X_+ + F_-X_-^\top L_-X_-)V = F_+X_+^\top L_+Y_+ + F_-X_-^\top L_-Y_-. \tag{26}$$

Multiplying $(F_+^{-1} + F_-^{-1})$ to (26), we have

$$\begin{aligned} & \left(\frac{1}{F_+} + \frac{1}{F_-}\right)(I + F_+X_+^\top L_+X_+ + F_-X_-^\top L_-X_-)V \\ & = \left(\frac{1}{F_+} + \frac{1}{F_-}\right)(F_+X_+^\top L_+Y_+ + F_-X_-^\top L_-Y_-). \end{aligned} \tag{27}$$

Thus,

$$\begin{aligned} V = & \left(\left(\frac{1}{F_+} + \frac{1}{F_-}\right)(I + X_+^\top L_+X_+ + X_-^\top L_-X_-\right)^{-1} \\ & \times \left(X_+^\top L_+Y_+ + X_-^\top L_-Y_-\right). \end{aligned} \tag{28}$$

Finally, we obtain the compact matrix form:

$$\begin{aligned} V = & \left(\left(\frac{1}{F_+} + \frac{1}{F_-}\right)I + [X_+^\top \ X_-^\top]\begin{bmatrix}\left(1 + \frac{F_+}{F_-}\right)L_+ & 0 \\ 0 & \left(1 + \frac{F_-}{F_+}\right)L_-\end{bmatrix}\begin{bmatrix}X_+ \\ X_-\end{bmatrix}\right)^{-1} \\ & \times [X_+^\top \ X_-^\top]\begin{bmatrix}\left(1 + \frac{F_+}{F_-}\right)L_+ & 0 \\ 0 & \left(1 + \frac{F_-}{F_+}\right)L_-\end{bmatrix}\begin{bmatrix}Y_+ \\ Y_-\end{bmatrix}. \end{aligned} \tag{29}$$

Using (29), we obtain the output layer weight matrix.

Section S.II, S.III, and S.V. of the Supplementary Material give the proposed methods' computational complexity, algorithm, and theoretical proof on the bound of the PW-GB membership function, respectively.

TABLE I: Classification accuracies along with average accuracy and average rankfor RoBell-RVFL model evaluated against baseline models on 26 UCI and KEEL datasets.

| **Dataset** ↓ \| **Model** → | **RVFL** [1] | **RVFLwoDL** [21] | **NF-RVFL** [8] | **C-RVFL** [9] | **AC-RVFL** [9] | **GB-RVFL** [10] | **GE-GB-RVFL** [10] | **RoBell-RVFL**† |
|---|---|---|---|---|---|---|---|---|
| checkerboard_Data | 85.9423 | 85.9808 | 85.5769 | 83.1731 | 85.5769 | 87.0192 | 87.0192 | 87.5 |
| cleve | 81.1111 | 80 | 82.2222 | 73.3333 | 81.1111 | 75.5556 | 82.2222 | 83.3333 |
| cmc | 68.552 | 68.552 | 67.6471 | 60.181 | 57.6923 | 68.0996 | 68.552 | 66.9683 |
| congressional_voting | 61.0687 | 61.0687 | 51.9084 | 61.0687 | 64.1221 | 61.0687 | 61.0687 | 60.3053 |
| conn_bench_sonar_mines_rocks | 74.6032 | 71.4286 | 74.6032 | 65.0794 | 66.6667 | 74.6032 | 68.254 | 82.5397 |
| crossplane130 | 97.4359 | 97.4359 | 100 | 79.4872 | 89.7436 | 100 | 97.4359 | 97.4359 |
| crossplane150 | 81.1111 | 81.1111 | 88.8889 | 88.8889 | 73.3333 | 86.6667 | 73.3333 | 93.3333 |
| echocardiogram | 85 | 87.5 | 80 | 85 | 87.5 | 87.5 | 87 | 82.5 |
| ecoli0137vs26 | 85.7447 | 86.8085 | 94.6809 | 82.9787 | 87.234 | 87.234 | 88.2979 | 94.6809 |
| ecoli2 | 82.0792 | 81.0891 | 88.1188 | 89.1089 | 89.1089 | 87.1287 | 87.1287 | 88.1188 |
| fertility | 90 | 80 | 73.3333 | 90 | 90 | 86.6667 | 90 | 86.6667 |
| haberman | 76.087 | 76.087 | 77.1739 | 82.6087 | 82.6087 | 78.2609 | 77.1739 | 76.087 |
| haberman_survival | 76.087 | 78.2609 | 71.7391 | 82.6087 | 82.6087 | 78.2609 | 77.1739 | 76.087 |
| heart_hungarian | 72.7753 | 72.6517 | 79.7753 | 75.2809 | 76.4045 | 74.1573 | 74.1573 | 75.2809 |
| heart-stat | 81.8889 | 81.8889 | 87.6543 | 77.7778 | 82.716 | 82.7161 | 81.4815 | 87.6543 |
| iono | 88.6792 | 80.566 | 83.9623 | 86.7925 | 88.6792 | 83.9623 | 84.9057 | 90.566 |
| ionosphere | 82.7925 | 82.4528 | 87.7358 | 85.8491 | 83.9623 | 83.9623 | 83.0189 | 85.8491 |
| led7digit-0-2-4-5-6-7-8-9_vs_1 | 94.7368 | 94.7368 | 92.4812 | 93.2331 | 93.2331 | 93.985 | 94.7368 | 96.2406 |
| molec_biol_promoter | 71 | 71.875 | 62.5 | 62.5 | 71.875 | 71.875 | 75 | 65.625 |
| monk1 | 42.515 | 52.0958 | 44.3114 | 43.1138 | 44.3114 | 44.3114 | 47.9042 | 53.8922 |
| monk3 | 43.1138 | 43.1138 | 76.0884 | 43.7126 | 45.509 | 44.9102 | 45.509 | 45.509 |
| monks_3 | 90.4072 | 90.4072 | 90.6347 | 74.8503 | 85.0299 | 85.6287 | 91.018 | 97.006 |
| oocytes_merluccius_nucleus_4d | 80.7134 | 82.7362 | 75.1792 | 67.4267 | 77.1987 | 80.7818 | 78.5016 | 78.1759 |
| vehicle1 | 81.6772 | 81.8583 | 83.8583 | 77.5591 | 77.9528 | 81.1024 | 79.9213 | 83.8583 |
| vehicle2 | 90.8504 | 90.0315 | 97.6378 | 88.189 | 86.6142 | 91.7323 | 92.5197 | 97.6378 |
| yeast-0-2-5-7-9_vs_3-6-8 | 95.351 | 89.7086 | 95.6954 | 91.0596 | 91.3907 | 98.0133 | 97.351 | 96.3576 |
| **Average accuracy** | 79.2817 | 78.8248 | 80.5156 | 76.5716 | 78.5455 | 79.8155 | 79.6417 | **81.8927** |
| **Average rank** | 5.0385 | 5.0769 | 4.2692 | 5.7308 | 4.5962 | 4.0577 | 4 | **3.2308** |

† represents the proposed models. The boldface indicates the best in terms of accuracy.

### A. Evaluation Dataset

Table I summarizes the classification accuracies and average ranks of all compared methods on 26 benchmark datasets. Detailed results, including standard deviations and optimal hyperparameter settings for each dataset, are provided in the Supplementary Material (Table IV). Among all evaluated models, the proposed RoBell-RVFL achieves the highest average classification accuracy of 81.8927%, outperforming the second-best method, NF-RVFL, which attains an average accuracy of 80.5156%. These results demonstrate the consistent superiority of RoBell-RVFL over both baseline RVFL models and recent state-of-the-art variants.

In comparison, conventional RVFL-based methods, namely RVFL and RVFLwoDL, obtain lower average accuracies of 79.2817% and 78.8248%, respectively. Complex-valued extensions, including CRVFL and ACRVFL, further lag behind with average accuracies of 76.5716% and 78.5455%. Similarly, granular ball-based approaches, GE-GB-RVFL and GB-RVFL, achieve average accuracies of 79.6417% and 79.8155%, respectively. Overall, these results highlight the superior generalization capability and stable performance gains of RoBell-RVFL across diverse datasets, confirming its effectiveness in comparison with both classical and modern RVFL-based models.

**Statistical rank:** Since average accuracy may be biased by outlier datasets, a ranking-based evaluation is adopted following [22], where models are ranked independently on each dataset and lower ranks indicate better performance. Let $\delta$ denote the number of models evaluated over $\omega$ datasets, and $R_r^s$ represent the rank of the $r^{th}$ model on the $s^{th}$ dataset. The average rank is computed as $R_r = \frac{1}{\omega}\sum_{s=1}^{\omega} R_r^s$. Using this criterion, the proposed RoBell-RVFL achieves the lowest average rank (3.2308), outperforming all competing methods. In contrast, GE-GB-RVFL (4.0000) and GB-RVFL (4.0577) rank second and third, respectively, while NF-RVFL (4.2692), CRVFL (5.7308), and ACRVFL (4.5962) exhibit higher rank values, confirming the superior robustness and generalization capability of RoBell-RVFL across diverse datasets.

**Friedman test:** The Friedman test [23] is employed to statistically examine whether significant performance differences exist among the compared models. Under the null hypothesis, all models are assumed to have equal average ranks, indicating equivalent performance. The Friedman statistic follows a chi-squared distribution, denoted as $\chi_F^2$, with $(\delta - 1)$ degrees of freedom and is computed as$\chi_F^2 = \frac{12\omega}{\delta(\delta+1)}\left(\sum_{i=1}^{\delta} R_i^2 - \frac{\delta(\delta+1)^2}{4}\right)$, where $\delta$ denotes the number of models, $\omega$ represents the number of datasets, and $R_i$ is the average rank of the $i^{\text{th}}$ model. The corresponding F-statistic is given by $F_F = \frac{(\omega-1)\chi_F^2}{\omega(\delta-1)-\chi_F^2}$, which follows an $F$-distribution with $(\delta - 1)$ and $(\omega - 1)(\delta - 1)$ degrees of freedom. In our experimental setting, eight models are evaluated across 26 datasets, yielding $\chi_F^2 = 18.4495$ and $F_F = 2.8202$. At a 5% significance level, the critical value of the $F$-distribution with $(7, 175)$ degrees of freedom is 2.0622. Since $F_F > 2.0622$, the null hypothesis is rejected, confirming the existence of statistically significant performance differences among the compared models.

In Supplementary Material Section S.VI. and Section S.VIII., discusses Win–tie–loss (W–T–L) sign test and sensitivity analysis, respectively.

## V. Conclusion and Future Work

This paper introduced RoBell-RVFL, a probability-weighted generalized bell random vector functional link network designed to address the fundamental challenges of class imbalance, label noise, and outliers in real-world data. Unlike conventional RVFL-based models that assign uniform importance to all training samples, RoBell-RVFL employs a principled sample-level adaptive weighting mechanism by jointly integrating kernel-induced generalized bell (gbell) membership functions with local class probability estimation. This strategy enables effective suppression of noisy and boundary samples within the majority class while fully preserving minority class information, resulting in more reliable and unbiased decision boundaries. Notably, class imbalance is handled implicitly through class-dependent regularization, avoiding heuristic or ratio-based weighting schemes. Extensive experiments on 26 benchmark datasets from the UCI and KEEL repositories demonstrate that RoBell-RVFL consistently outperforms state-of-the-art RVFL variants, neuro-fuzzy models, complex-valued RVFLs, and recent granular and graph-based approaches in terms of classification accuracy, average rank, robustness to label noise, and statistical significance. These improvements are further supported by Friedman and win–tie–loss statistical tests, confirming that the observed performance gains are consistent and statistically meaningful. Overall, the results highlight that effective learning from imbalanced data requires simultaneous consideration of sample quality, class dominance, and noise suppression capabilities that are jointly realized by the performance of the proposed RoBell-RVFL framework.

## References


[1] Y.-H. Pao, G.-H. Park, and D. J. Sobajic, "Learning and generalization characteristics of the random vector functional-link net," *Neurocomputing*, vol. 6, no. 2, pp. 163–180, 1994.

[2] P. N. Suganthan and R. Katuwal, "On the origins of randomization-based feedforward neural networks," *Applied Soft Computing*, vol. 105, p. 107239, 2021.

[3] A. Malik, R. Gao, M. Ganaie, M. Tanveer, and P. N. Suganthan, "Random vector functional link network: Recent developments, applications, and future directions," *Applied Soft Computing*, vol. 143, p. 110377, 2023.

[4] M. Kearns and U. Vazirani, *An Introduction to Computational Learning Theory*. MIT Press, 1994.

[5] D. Husmeier and D. Husmeier, "Random vector functional link (RVFL) networks," *Neural networks for conditional probability estimation: forecasting beyond point predictions*, pp. 87–97, 1999.

[6] M. A. Ganaie, M. Tanveer, and P. N. Suganthan, ""Minimum Variance Embedded Random Vector Functional Link Network"," in *Neural Information Processing*. Cham: Springer International Publishing, 2020, pp. 412–419.

[7] P.-B. Zhang and Z.-X. Yang, "A new learning paradigm for random vector functional-link network: RVFL+," *Neural Networks*, vol. 122, pp. 94–105, 2020.

[8] Sajid, M. and Malik, A. K. and Tanveer, M. and Suganthan, Ponnuthurai N., "Neuro-Fuzzy Random Vector Functional Link Neural Network for Classification and Regression Problems," *IEEE Transactions on Fuzzy Systems*, vol. 32, no. 5, pp. 2738–2749, 2024.

[9] C. Liu, H. Zhang, L. Chen, and F. Li, "Complex-valued random vector functional link neural network based on real augmented representation and its applications," *Applied Soft Computing*, vol. 170, p. 112682, 2025.

[10] M. Sajid, A. Quadir, and M. Tanveer, "GB-RVFL: Fusion of randomized neural network and granular ball computing," *Pattern Recognition*, vol. 159, p. 111142, 2025.

[11] M. Tanveer, R. Sharma, M. Sajid, and A. Quadir, "GRVFL-MV: Graph random vector functional link based on multi-view learning," *Information Sciences*, vol. 704, p. 121947, 2025.

[12] M. Hanmandlu *et al.*, "A new entropy function and a classifier for thermal face recognition," *Engineering Applications of Artificial Intelligence*, vol. 36, pp. 269–286, 2014.

[13] J. Laurikkala, "Improving identification of difficult small classes by balancing class distribution," in *Conference on artificial intelligence in medicine in Europe*. Springer, 2001, pp. 63–66.

[14] S. Rezvani and X. Wang, "A broad review on class imbalance learning techniques," *Applied Soft Computing*, vol. 143, p. 110415, 2023.

[15] M. Sahani and P. K. Dash, "Fpga-based online power quality disturbances monitoring using reduced-sample hht and class-specific weighted rvfln," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 8, pp. 4614–4623, 2019.

[16] M. Ganaie, M. Sajid, A. K. Malik, and M. Tanveer, "Graph embedded intuitionistic fuzzy random vector functional link neural network for class imbalance learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 9, pp. 11 671–11 680, 2024.

[17] M. Tanveer and A. Quadir, "Robust universum twin support vector machine for imbalanced data," in *2025 International Joint Conference on Neural Networks (IJCNN)*, 2025, pp. 1–9.

[18] A. Kumari, M. Tanveer, C. T. Lin, and the Alzheimer's Disease Neuroimaging Initiative, "Class Probability and Generalized Bell Fuzzy Twin SVM for Imbalanced Data," *IEEE Transactions on Fuzzy Systems*, vol. 32, no. 5, pp. 3037–3048, 2024.

[19] S. Rezvani and X. Wang, "Intuitionistic fuzzy twin support vector machines for imbalanced data," *Neurocomputing*, vol. 507, pp. 16–25, 2022.

[20] A. Quadir, M. Ganaie, and M. Tanveer, "Intuitionistic fuzzy generalized eigenvalue proximal support vector machine," *Neurocomputing*, vol. 608, p. 128258, 2024.

[21] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006, neural Networks.

[22] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *The Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

[23] M. Friedman, "A comparison of alternative tests of significance for the problem of m rankings," *The Annals of Mathematical Statistics*, vol. 11, no. 1, pp. 86–92, 1940.