Learning Representations from Perturbation: A Novel Matrix-View Weighting Framework for Naive Bayes

Siyao Wu[1] **Huan Zhang***[1] **Kexin Meng**[1] **Zhipeng Ding**[1] **Pei Lv**[1]

[1]School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou, China

Abstract

Numerous attribute weighting methods have recently been proposed to alleviate the attribute conditional independence assumption in naive Bayes. Among them, multi-view attribute weighting framework has achieved state-of-the-art performance by constructing additional latent views beyond the raw attribute view, thereby capturing more comprehensive data characteristics. However, these latent views are usually derived from base classifiers trained on a fixed input distribution, which greatly limits the diversity of generated attributes. Additionally, in most cases, the latent views consist solely of hard labels, disregarding the predicted posterior probability of the base classifiers, resulting in incomplete utilization of discriminative evidence. To address these issues, we propose a novel framework called Perturbation-driven Matrix-view Weighted Naive Bayes (PMWNB). In PMWNB, diverse input perturbations are first applied to the raw attribute view to generate multiple base views. Subsequently, each base view independently trains multiple heterogeneous base classifiers, whose hard label and soft probability outputs are jointly leveraged to construct four latent views. Finally, class-specific attribute value weights in each view are respectively optimized by minimizing the negative conditional log-likelihood. Extensive experiments conducted on a collection of 59 benchmark datasets demonstrate the superiority of PMWNB. The source code and datasets are available at https://github.com/zhanghuan1994/PMWNB.

*Corresponding author

1 INTRODUCTION

In recent years, Bayesian network classifiers (BNCs) have received much attention due to their explicit interpretability and remarkable classification performance. Among various BNCs, naive Bayes (NB) stands out as a simple yet effective classifier, based on the assumption that all attributes are conditionally independent given the class label. Due to its efficiency and robustness, NB has already shown exceptional classification performance and continues to be one of the top 10 algorithms in data mining [Wu et al., 2008]. However, in real-world applications, its attribute conditional independence assumption rarely holds true. To mitigate this, numerous enhancements have been proposed, which are generally categorized into six groups: structure extension [Friedman et al., 1997, Webb et al., 2005, Zhang et al., 2020a, Zhao et al., 2025, Meng et al., 2026], fine-tuning [Diab and Hindi, 2018, Zhang and Jiang, 2022], instance selection [Kohavi, 1996, Frank et al., 2003, Wang et al., 2026], instance weighting [Xu et al., 2019, Zhang et al., 2026], attribute selection [Hall, 2000, Chen et al., 2017, 2020], and attribute weighting [Zaidi et al., 2013, Jiang et al., 2019, Zhou et al., 2025].

Among these categories, attribute weighting has long been a particularly important line of research owing to its considerable flexibility and strong performance. For instance, Hall [2007] proposed a Decision Tree-based Attribute Weighted Naive Bayes (DTAWNB), which utilizes decision tree depth to determine attribute weights. Jiang et al. [2019] developed a Correlation-based Attribute Weighted Naive Bayes (CAWNB), which considers both attribute-class relevance and attribute-attribute redundancy in weight learning. Zaidi et al. [2013] adopted gradient descent to optimize attribute weights by maximizing the conditional log-likelihood or minimizing the classification error rate. Zhang et al. [2020b] proposed a Class-specific Attribute Value Weighted Naive Bayes (CAVWNB), which enables more fine-grained modeling by assigning a unique weight to each attribute value for each class.

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:7416–7432.

Unlike the aforementioned methods that mainly focus on designing novel weighting strategies, recent studies have argued that relying solely on the raw attribute view is usually insufficient to fully capture the intrinsic data characteristics in complex real-world applications [Zhang et al., 2023, Zhou et al., 2025, Zhang et al., 2026]. To address this, numerous multi-view weighting frameworks have been introduced, among which the Multi-view Attribute Weighted Naive Bayes (MAWNB) [Zhang et al., 2023] is one of the most well-known. In MAWNB, multiple base classifiers are first trained on the raw attribute view. Subsequently, the hard labels predicted by each base classifier for each instance are used to construct two additional label views. By introducing label views, MAWNB enriches the view representation, thereby usually achieving superior performance over traditional single-view attribute weighting methods.

Despite their preliminary success, current multi-view weighting frameworks still suffer from two critical limitations. First, all base classifiers used for view construction are trained on a fixed input distribution, i.e., on the same raw attribute view, which greatly limits the diversity of generated attributes in the latent view space. Second, the constructed latent views usually consist solely of hard labels, which completely disregards the predicted posterior probability of the base classifiers and results in an incomplete utilization of discriminative evidence.

To address these issues, this study proposes a novel framework called Perturbation-driven Matrix-view Weighted Naive Bayes (PMWNB), which further extends the conventional multi-view representation into a matrix-view framework. Specifically, PMWNB first applies diverse input perturbations to the raw attribute view to generate multiple base views. Subsequently, multiple Super-Parent One-Dependence Estimators (SPODEs) and Random Trees (RTs) are trained on each base view independently. Next, to more comprehensively exploit the latent space information captured by the base classifiers, hard label and soft probability outputs are jointly leveraged to construct four latent views. Finally, PMWNB optimizes class-specific attribute value weights by minimizing the negative conditional log-likelihood in each view. By extending multi-view representation into a matrix-view framework, PMWNB integrates distinct partitions of the raw attribute view with richer discriminative evidence. Extensive experimental results on a collection of 59 benchmark datasets from the University of California at Irvine (UCI) Machine Learning repository demonstrate the superior performance of PMWNB.

The main contributions of this study are as follows:

- We extend the conventional multi-view representation into a matrix-view framework, and propose the Perturbation-driven Matrix-view Weighted Naive Bayes (PMWNB), in which diverse input perturbations are applied to the raw attribute view to generate multiple base views.
- We propose a novel view construction strategy that fully exploits the latent view space information through the joint utilization of hard label and soft probability outputs by base classifiers.
- We conduct extensive experiments to validate the superiority of PMWNB. We also perform an ablation study, an error-ambiguity decomposition analysis and an extension experiment to provide some in-depth discussions of PMWNB.

2 RELATED WORK

Naive Bayes. Given a test instance $\mathbf{x}$ represented by an attribute value vector $\langle a_1, a_2, \ldots, a_m \rangle$, NB uses Eq. (1) and Eq. (2) to predict its class label.

$$P(c \mid \mathbf{x}) = \frac{P(c) \prod_{j=1}^{m} P(a_j \mid c)}{\sum_{c \in \mathcal{C}} P(c) \prod_{j=1}^{m} P(a_j \mid c)} \tag{1}$$

$$c(\mathbf{x}) = \arg\max_{c \in \mathcal{C}} P(c \mid \mathbf{x}) \tag{2}$$

where m is the number of attributes, a_j is the jth attribute value of $\mathbf{x}$, $\mathcal{C}$ denotes the set of all possible class labels, $c(\mathbf{x})$ is the predicted class label of $\mathbf{x}$, while $P(c)$ and $P(a_j \mid c)$ correspond to the needed prior and conditional probability, respectively, both of which can be estimated using Laplace smoothing as follows:

$$P(c) = \frac{\sum_{i=1}^{n} \delta(c_i, c) + 1}{n + q} \tag{3}$$

$$P(a_j \mid c) = \frac{\sum_{i=1}^{n} \delta(a_{ij}, a_j)\delta(c_i, c) + 1}{\sum_{i=1}^{n} \delta(c_i, c) + n_j} \tag{4}$$

where n is the number of training instances, q is the number of classes, n_j is the number of values for the jth attribute A_j, c_i is the class label of the ith training instance, a_{ij} is the jth attribute value of the ith training instance, and the indicator function $\delta(x, y)$ is one if $x = y$ and zero otherwise.

Multi-view Weighted Naive Bayes. To alleviate the attribute conditional independence assumption in NB, numerous enhancements have been proposed from diverse perspectives. While most enhancements focus exclusively on the raw attribute view, recent studies argue that the raw attribute space is often insufficient to capture the intrinsic data characteristics in complex real-world applications. To uncover some hidden information, Zhang et al. [2022] proposed Attribute Augmented and Weighted Naive Bayes (A^2WNB), which treats predicted class labels from multiple random one-dependence estimators as latent attributes. These labels are concatenated with the original attributes to form an augmented attribute view. Subsequently, Zhang et al. [2023] introduced a multi-view learning framework for NB, termed Multi-view Attribute Weighted Naive Bayes

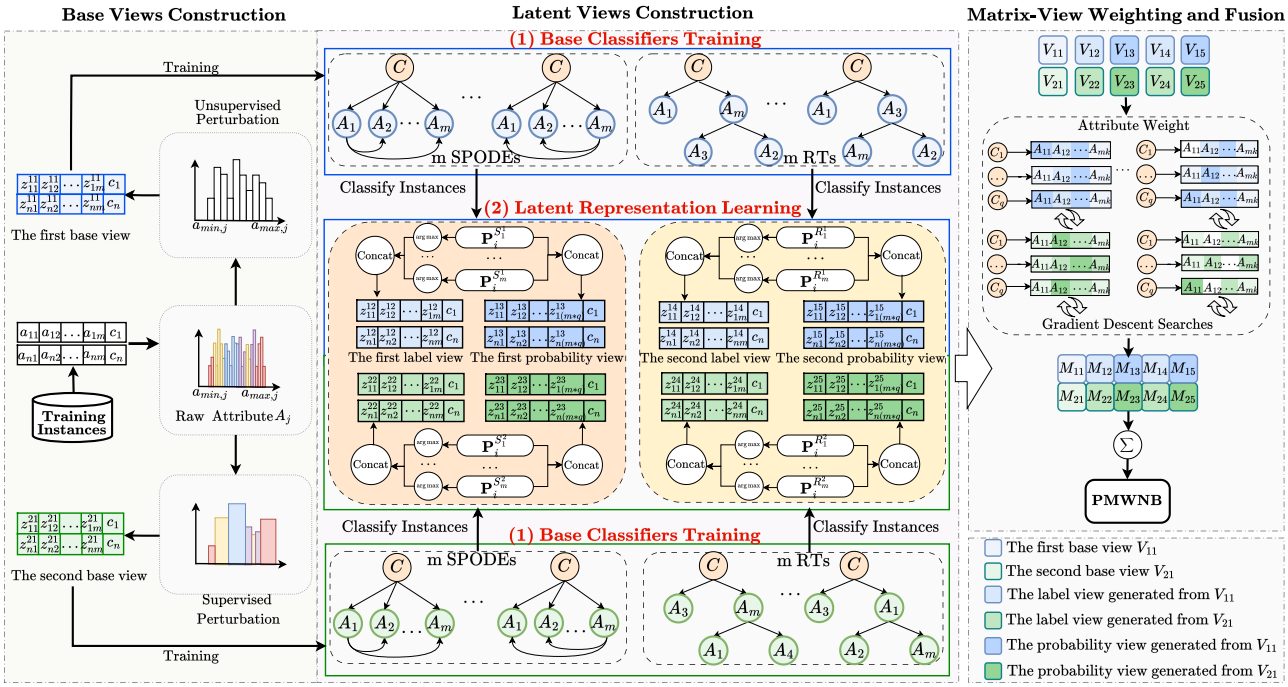


Figure 1: Overall framework of PMWNB. (1) Base Views Construction: Applying unsupervised and supervised perturbation to construct two base views. (2) Latent Views Construction: Training m SPODEs and m RTs on each base view, and then using predicted hard labels and soft probabilities to construct two latent views. (3) Matrix-View Weighting and Fusion: Optimizing class-specific attribute value weights in each view and fusing all class-membership probabilities for classification.

(MAWNB), which constructs two additional label views using different classifiers, effectively capturing data characteristics from higher-level perspectives. To further improve the generalization and specificity of MAWNB, Liu et al. [2025a] proposed Enhanced Multi-view Attribute Weighted Naive Bayes (EMAWNB), which utilizes a cross-validation strategy to mitigate the overfitting risk in the multi-view construction module. Moreover, EMAWNB learns instance-specific view weights by evaluating the local classification performance of each view on the k-nearest neighbors of the test instance. Additionally, Zhou et al. [2025] proposed a Multi-view Collaborative Optimization framework for Weighted Naive Bayes (MCOWNB), which dynamically adjusts the weight of each view by iteratively minimizing the local mean squared error and optimizing the global generalization performance. Recently, multi-view learning has also been introduced into instance weighted naive Bayes. Zhang et al. [2026] proposed a Dual-view Instance Weighted Naive Bayes (DIWNB), which constructs a generated view based on k-nearest neighbors to capture local data characteristics within the raw attribute view. Unlike the above methods that rely on hard labels for latent view construction, DIWNB employs the summation of estimated class-membership probabilities generated by base classifiers to derive latent views, providing an alternative strategy under the multi-view weighting framework.

3 METHOD

3.1 MOTIVATION

As discussed above, the multi-view weighting framework substantially enhances NB by capturing more comprehensive data characteristics. However, most existing multi-view weighting frameworks, such as MAWNB, still exhibit two notable limitations that hinder their further improvement. First, as all base classifiers are trained on a single raw attribute view, the diversity of the generated attributes within the latent view space is substantially constrained, leaving potential discriminative patterns insufficiently explored. Second, MAWNB constructs the two label views solely based on the predicted hard labels of base classifiers, which completely disregards the fine-grained confidence information conveyed by the estimated class-membership probabilities, resulting in the insufficient exploitation of discriminative evidence.

To address these issues, we propose a novel framework called Perturbation-driven Matrix-view Weighted Naive Bayes (PMWNB). Unlike previous methods where all generated label views originate from a single raw attribute view, PMWNB extends the conventional multi-view learning framework to a matrix-view paradigm. It further leverages diverse input perturbation strategies, for example, different discretization methods, to construct multiple base

views and broaden the exploration of the latent attribute space. Additionally, PMWNB leverages both hard label and soft probability outputs by multiple heterogeneous base classifiers to construct the latent views. By integrating multiple forms of outputs, PMWNB captures the underlying attribute space uncovered by the base classifiers in a more comprehensive manner. Finally, class-specific attribute value weights within the matrix-view framework are optimized and all the class-membership probabilities are fused for classification.

3.2 OVERALL FRAMEWORK

PMWNB is built upon a matrix-view representation $\mathbf{V} = \{V_{gt} \mid g \in \{1, \ldots, G\}, t \in \{1, \ldots, T\}\}$, in which G denotes the total number of input perturbation strategies and T represents the number of view representation types. Specifically, the views within $\mathbf{V}$ are defined as follows:

- **Definition 1** (The base view). The view that is generated by applying distinct perturbation strategies to the raw attributes. It usually represents original attribute spaces, but can provide heterogeneous input distributions that drive the subsequent latent view construction.

- **Definition 2** (The latent view). The view that is generated by using the predictive outputs of base classifiers. It usually represents higher-level attribute spaces, which mainly consists of the hard label (label view) and the soft probability (probability view).

As illustrated in Figure 1, the framework consists of three modules. The Base Views Construction module transforms each raw instance $\mathbf{x}_i = \langle a_{i1}, \ldots, a_{im}, c_i \rangle$ into the base view $\langle z_{i1}^{g1}, \ldots, z_{im}^{g1}, c_i \rangle$ through G distinct input perturbations, where the upper branch employs an unsupervised perturbation to map continuous attributes to discrete values, while the lower branch utilizes a supervised perturbation strategy. Next, in the Latent Views Construction module, m Super-Parent One-Dependence Estimators (SPODEs) and Random Trees (RTs) are independently trained on each base view V_{g1}. Their predicted hard labels and soft probabilities construct the respective label and probability views, yielding the latent representation $\langle z_{i1}^{gt}, \ldots, z_{im_t}^{gt}, c_i \rangle$, where the dimension m_t is m for label views and $m \times q$ for probability views given q classes. Finally, in the Matrix-View Weighting and Fusion module, to mitigate attribute redundancy within each view V_{gt}, optimal class-specific weights for each attribute value are learned by minimizing the negative conditional log-likelihood via gradient descent search, forming the set of optimized models $\mathcal{M} = \{M_{gt}\}$. For each test instance, the class-membership probabilities estimated by all models in $\mathcal{M}$ are averaged. The following subsections describe the detailed processes of these modules.

3.3 BASE VIEWS CONSTRUCTION

To implement effective input perturbation, the first and second base views are generated by applying Equal Width Discretization (EWD) and the Minimum Description Length Principle (MDLP) [Fayyad and Irani, 1993], respectively. The selection of these specific strategies is mainly motivated by the following reasons: (1) As an unsupervised method, EWD partitions attributes based purely on distribution range, capturing global structural information without label bias. (2) In contrast, the supervised MDLP leverages information gain to determine cut-points that maximize class separability, focusing on the most informative decision boundaries. (3) The divergent principles of these strategies naturally yield distinct sets of cut-points for attribute partitioning, serving as effective perturbations to the raw attribute space.

For the construction of the first base view V_{11} via EWD, the raw attribute value a_{ij} of A_j is mapped to z_{ij}^{11} using Eq. (5).

$$z_{ij}^{11} = \begin{cases} K_{EWD}, & \text{if } a_{ij} = a_{\max,j} \\ \lfloor \frac{a_{ij} - a_{\min,j}}{\Delta_j} \rfloor + 1, & \text{otherwise} \end{cases} \tag{5}$$

where $a_{\max,j}$ and $a_{\min,j}$ denote the maximum and minimum values of the jth attribute in the training set, and the hyperparameter K_{EWD} represents the predefined number of intervals, which is set to 10 by default. The interval width Δ_j is calculated as Eq. (6).

$$\Delta_j = (a_{\max,j} - a_{\min,j}) / K_{EWD} \tag{6}$$

The construction of the second base view V_{21} utilizes MDLP. For each attribute A_j, candidate cut-points are defined as the midpoints between successive distinct values in the sorted attribute sequence. Let D be the set of training instances in the current interval, which initially consists of the entire training set. The optimal cut-point $\mathcal{T}$ is selected to maximize the information gain. If $\mathcal{T}$ satisfies Eq. (7), the current interval D is split into subsets D_1 and D_2 such that $A_j \leq \mathcal{T}$ and $A_j > \mathcal{T}$, and the process recurses on both subsets; otherwise, partitioning terminates.

$$Gain(A_j, \mathcal{T}; D) > \frac{\log(N-1)}{N} + \frac{\Delta(A_j, \mathcal{T}; D)}{N} \tag{7}$$

where N denotes the number of instances in the current interval D. The correction factor $\Delta(A_j, \mathcal{T}; D)$ is defined as:

$$\begin{aligned} \Delta(A_j, \mathcal{T}; D) = & \log \left[\sum_{q_1=1}^{q-1} \binom{q}{q_1} 2^{q_1} + 2^q - 1 \right] \\ & - [q \cdot Ent(D) - q_1 \cdot Ent(D_1) - q_2 \cdot Ent(D_2)] \end{aligned} \tag{8}$$

where $Ent(\cdot)$ denotes the Shannon entropy and q, q_1, q_2 represent the number of distinct classes present in D, D_1, D_2, respectively. The continuous value a_{ij} is mapped to z_{ij}^{21} determined by the interval it falls into:

$$z_{ij}^{21} = k, \quad \text{s.t. } \mathcal{T}_{k-1} < a_{ij} \leq \mathcal{T}_k \tag{9}$$

where $\mathcal{T}_0 = -\infty$ and $\mathcal{T}_{p+1} = +\infty$. The number of final cut points is denoted by p, resulting in $p+1$ disjoint intervals.

3.4 LATENT VIEWS CONSTRUCTION

To expand the latent attribute space through diverse input perturbations in the two base views, we construct the latent views in two phases: base classifiers training and latent representation learning.

3.4.1 Base Classifiers Training

In this phase, two distinct ensembles are trained on each base view V_{g1}: m SPODEs $\mathcal{S}^g = \{S_1^g, \ldots, S_m^g\}$ and m RTs $\mathcal{R}^g = \{R_1^g, \ldots, R_m^g\}$.

The jth SPODE S_j^g sets A_j as the super-parent for all other attributes. Let $\mathbf{y}_i = \mathbf{z}_i^{g1}$ denote the attribute vector of the ith instance within the base view V_{g1}. The posterior probability $P_i^{S_j^g}(c \mid \mathbf{y}_i)$ estimated by S_j^g is calculated by Eq. (10).

$$P_i^{S_j^g}(c \mid \mathbf{y}_i) = P(z_j, c) \prod_{r=1, r\neq j}^{m} P(z_r | z_j, c) \tag{10}$$

where $P(z_j, c)$ is the joint probability and $P(z_r|z_j, c)$ is the conditional probability, which can be estimated by the following Eq. (11) and Eq. (12), respectively.

$$P(z_j, c) = \frac{\sum_{i=1}^{n} \delta(z_{ij}, z_j)\delta(c_i, c) + 1}{n + n_j * q} \tag{11}$$

$$P(z_r|z_j, c) = \frac{\sum_{i=1}^{n} \delta(z_{ir}, z_r)\delta(z_{ij}, z_j)\delta(c_i, c) + 1}{\sum_{i=1}^{n} \delta(z_{ij}, z_j)\delta(c_i, c) + n_r} \tag{12}$$

where n_j, n_r is the number of values for the attribute A_j and A_r, respectively.

To construct the jth RT R_j^g, at each node, $\log m$ attributes are randomly selected from V_{g1} and the data is recursively partitioned by maximizing the information gain defined in Eq. (13).

$$Gain(D, A_j) = Ent(D) - \sum_{v=1}^{n_j} \frac{|D_v|}{|D|} Ent(D_v) \tag{13}$$

where A_j is the split attribute, n_j is the number of values for A_j, D_v denotes the subset of instances at the vth child node. The posterior probability $P_i^{R_j^g}(c \mid \mathbf{y}_i)$, estimated by R_j^g, is derived from the class frequency distribution in the leaf node that $\mathbf{y}_i$ falls into.

$$P_i^{R_j^g}(c \mid \mathbf{y}_i) = \frac{\sum_{r=1}^{l} \delta(c_r, c)}{l} \tag{14}$$

where l denotes the number of training instances in the leaf node that $\mathbf{y}_i$ falls into.

3.4.2 Latent Representation Learning

In this phase, latent views are derived from the predictive outputs of the ensembles $\mathcal{S}^g$ and $\mathcal{R}^g$. Specifically, extracting both hard labels and soft probabilities from each ensemble produces two label views and two probability views. The reasons of selecting these two distinct forms are as follows: (1) Hard labels provide definitive classification results, which can establish clear decision boundaries and offer a robust representation for latent attribute space. (2) Soft probabilities preserve fine-grained confidence information and capture subtle distributional details that are inevitably lost in hard labels. (3) Instances with the same label or similar probability estimates are likely to possess common higher-level data characteristics.

For a training instance $\mathbf{y}_i$ within the base view V_{g1}, the label views are constructed by taking the class label with the highest estimated posterior probability, while the probability views are formed by concatenating posterior probability vectors of m base classifiers. Let $E \in \{S, R\}$ denote the base classifier type corresponding to the SPODEs and RTs, respectively. The SPODEs generate the first label and probability views at $t = 2$ and $t = 3$, whereas the RTs produce the second views at $t = 4$ and $t = 5$. The latent representation $\mathbf{z}_i^{gt}$ is defined as:

$$\mathbf{z}_i^{gt} = \begin{cases} \left[\hat{c}_i^{E_1^g}, \ldots, \hat{c}_i^{E_m^g}\right], & \text{if } t \in \{2, 4\} \\ \left[\mathbf{P}_i^{E_1^g}, \ldots, \mathbf{P}_i^{E_m^g}\right], & \text{if } t \in \{3, 5\} \end{cases} \tag{15}$$

where the predicted class labels $\hat{c}_i^{E_j^g}$ and posterior probability vectors $\mathbf{P}_i^{E_j^g}$ are defined as:

$$\hat{c}_i^{E_j^g} = \arg\max_{c \in \mathcal{C}} P_i^{E_j^g}(c \mid \mathbf{y}_i) \tag{16}$$

$$\mathbf{P}_i^{E_j^g} = \left[P_i^{E_j^g}(c_1 \mid \mathbf{y}_i), P_i^{E_j^g}(c_2 \mid \mathbf{y}_i), \ldots, P_i^{E_j^g}(c_q \mid \mathbf{y}_i)\right] \tag{17}$$

3.5 MATRIX-VIEW WEIGHTING AND FUSION

After constructing the view matrix $\mathbf{V}$, to mitigate attribute redundancy, a class-specific attribute value weighting module is employed for each view. As a prerequisite, the continuous distributions in the soft probability views are first discretized to align with their corresponding base views. Specifically, EWD with 10 equal-width bins in the $[0, 1]$ interval is applied to views generated by the first base view, while MDLP is utilized for those derived from the second base view. The weighted joint probability of class c is formulated by incorporating the weight $w_{c,jk}$ into the exponential part of the conditional probability, and then the class-membership probabilities of the test instance $\mathbf{x}$ belonging

to each class can be predicted by Eq. (18).

$$P(c \mid \mathbf{z}) = \frac{P(c)\prod_{j=1}^{m} P(z_{jk} \mid c)^{w_{c,jk}}}{\sum_{c' \in C} P(c')\prod_{j=1}^{m} P(z_{jk} \mid c')^{w_{c',jk}}} \quad (18)$$

where $w_{c,jk}$ is the weight specific to class c and $\mathbf{z} \in \{\mathbf{z}^{gt} \mid g \in \{1,\ldots,G\}, t \in \{1,\ldots,T\}\}$ denotes the representation of $\mathbf{x}$ mapped to the current view, in which each $\mathbf{z}^{gt}$ consists of the corresponding unified attribute values z_{jk}. Subsequently, the L-BFGS-M optimization procedure [Zhu et al., 1997] with L_2 regularization is employed to learn class-specific attribute value weights in each view, respectively. Initialized at 1.0, these weights are optimized via gradient descent by minimizing the negative conditional log-likelihood (-CLL):

$$\begin{aligned} & -CLL(\mathbf{w}_{gt}) \\ &= -\log P(C \mid \mathbf{Z}, \mathbf{w}_{gt}) + \lambda\|\mathbf{w}_{gt} - \mathbf{w_{one}}\|^2 \\ &= -\sum_{i=1}^{n} \log \frac{P(c_i, \mathbf{z}_i; \mathbf{w}_{gt})}{\sum_{c'} P(c', \mathbf{z}_i; \mathbf{w}_{gt})} + \lambda\|\mathbf{w}_{gt} - \mathbf{w_{one}}\|^2 \end{aligned} \quad (19)$$

where $\mathbf{w}_{gt}$ represents the attribute weight matrix, $\mathbf{w_{one}}$ is a matrix of the same size as $\mathbf{w}_{gt}$, with all elements equal to 1.0, λ is a hyperparameter whose default value is 1.0, and the weighted joint probability is:

$$P(c, \mathbf{z}_i; \mathbf{w}_{gt}) = P(c)\prod_{j=1}^{m} P(z_j \mid c)^{w_{c,jk}} \quad (20)$$

The gradient of $-CLL(\mathbf{w}_{gt})$ with respect to $w_{c,jk}$ can be calculated by Eq. (21).

$$\begin{aligned} & \frac{\partial}{\partial w_{c,jk}}[-CLL(\mathbf{w}_{gt})] \\ &= \frac{\partial}{\partial w_{c,jk}} \sum_{i=1}^{n} \left[\log\left(\sum_{c'} P(c', \mathbf{z}_i; \mathbf{w}_{gt})\right) - \log P(c, \mathbf{z}_i; \mathbf{w}_{gt})\right] \\ &\quad + 2\lambda(w_{c,jk} - 1) \\ &= \sum_{i=1}^{n} \left[\frac{P(c, \mathbf{z}_i; \mathbf{w}_{gt})\log(P(z_j \mid c))}{\sum_{c'} P(c', \mathbf{z}_i; \mathbf{w}_{gt})}\right. \\ &\quad \left. -\delta(c_i, c)\frac{P(c, \mathbf{z}_i; \mathbf{w}_{gt})\log(P(z_j \mid c))}{P(c, \mathbf{z}_i; \mathbf{w}_{gt})}\right] + 2\lambda(w_{c,jk} - 1) \\ &= \sum_{i=1}^{n} [P(c \mid \mathbf{z}_i, \mathbf{w}_{gt})\log(P(z_j \mid c))] \\ &\quad - \delta(c_i, c)\log(P(z_j \mid c)) + 2\lambda(w_{c,jk} - 1) \end{aligned} \quad (21)$$

Now, we have built all models $\mathcal{M}$ of the matrix-view $\mathbf{V}$. Given a test instance $\mathbf{x}$, it is first mapped to its corresponding representation $\mathbf{z}^{gt}$ within each view V_{gt}. Subsequently, we fuse the results by averaging their estimated class-membership probabilities. The predicted class label $c(\mathbf{x})$ for each test instance is defined as the class with the maximum class-membership probability, which can be represented as follows.

$$c(\mathbf{x}) = \arg\max_{c \in \mathcal{C}} \sum_{g=1}^{G}\sum_{t=1}^{T} P_{gt}(c \mid \mathbf{z}^{gt}) \quad (22)$$

4 EXPERIMENTS AND RESULTS

4.1 EXPERIMENTAL SETTING AND DATASETS

Experimental Setting and Competitors. We designed four groups of experiments on the Waikato Environment for Knowledge Analysis (WEKA) platform [Witten et al., 2011] to validate the effectiveness of PMWNB. (1) We compared PMWNB with its most related competitors EMAWNB [Liu et al., 2025a], MAWNB [Zhang et al., 2023], A^{2}WNB [Zhang et al., 2022], CAVWNB [Zhang et al., 2020b], AODE [Webb et al., 2005], RF [Breiman, 2001] and the standard NB [Langley et al., 1992]. (2) We conducted an ablation study to systematically assess the individual contributions of each component within the matrix-view construction module of PMWNB. (3) We performed an error-ambiguity decomposition analysis to offer a comprehensive evaluation of the diversity inherent in the matrix-view representation of PMWNB. (4) We combined various perturbation strategies to explore the extensibility of the proposed matrix-view framework.

Benchmark Datasets and Preprocessing. Our experiments were conducted on a collection of 59 benchmark datasets obtained from the University of California at Irvine (UCI) Machine Learning repository. Among them, 33 datasets were adopted from [Zhang et al., 2023], each characterized by numeric attributes comprising more than half of the attribute space. The remaining 26 datasets were fully numeric and sourced from [Liu et al., 2025b], with the constraint that each dataset contains more than five attributes. All missing attribute values in these datasets were replaced with the means for numeric attributes and the modes for nominal attributes from the available data. In our experiments, PMWNB is capable of directly handling numeric datasets, whereas all of its competitors are restricted to nominal inputs. To ensure a fair comparison, we standardized the preprocessing procedure by applying 10-bin equal-width discretization to all competitors, consistent with the settings adopted in their original studies.

4.2 EXPERIMENTAL RESULTS

Figure 2 visualizes the classification accuracy comparison between PMWNB and its competitors across 59 UCI datasets. All results were obtained via 10 separate runs of stratified 10-fold cross-validation. The detailed classification accuracy comparison for all datasets is provided in Appendix B.2, from which we conclude that PMWNB

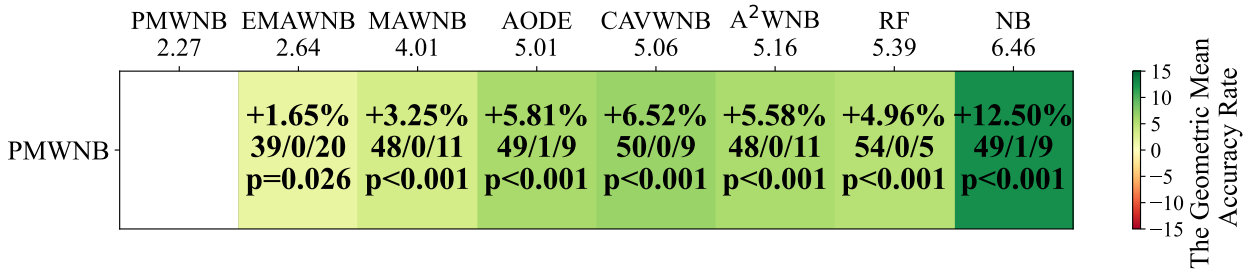


Figure 2: A summary of the classification accuracy comparison results for PMWNB versus its related competitors. Within each cell, the first row reports the geometric mean accuracy improvement of PMWNB compared with the respective method. The second row details the Wins/Draws/Losses (W/D/L) results, obtained by comparing classification accuracy across the 59 benchmark datasets. The third row presents the p-value calculated by a two-sided Wilcoxon signed-rank test. Statistically significant results ($p < 0.05$) are highlighted in bold, indicating that the performances are significantly different. The number below each algorithm's names represents its average rank across all datasets.

achieves the best overall performance among all competitors, highlighted by the following observations:

(1) **The overall performance of PMWNB:** As shown in Figure 2, PMWNB achieves the lowest average rank of 2.27, which is significantly better than its competitors: EMAWNB (2.64), MAWNB (4.01), AODE (5.01), CAVWNB (5.06), A^{2}WNB (5.16), RF (5.39), and NB (6.46). Moreover, the Wilcoxon signed-ranks test results confirm that PMWNB statistically outperforms all the compared state-of-the-art competitors at the $\alpha = 0.05$ significance level.

(2) **The effectiveness of matrix-view construction:** Compared to the most related multi-view competitor EMAWNB and MAWNB, PMWNB yields a 1.65% and 3.25% increase in the geometric mean accuracy, achieving better classification accuracy on 39 and 48 out of 59 datasets, respectively. This substantial gain shows that driving the representation of hard labels and soft probabilities via diverse input perturbations enables the matrix-view framework to capture richer data characteristics than traditional multi-view methods.

(3) **The superiority over other state-of-the-art competitors:** PMWNB consistently dominates other compared baselines, yielding geometric mean accuracy improvements ranging from 4.96% to 12.50%. Moreover, out of the 59 datasets it achieves higher classification accuracy against AODE (49), CAVWNB (50), A^{2}WNB (48), RF (54) and NB (49). These results all demonstrate the superiority of PMWNB.

4.3 ABLATION STUDY

We conducted an ablation study to systematically assess the individual contributions of each component within the matrix-view construction module. We compared PMWNB with its three ablation variants: (1) PMWNB w/o GLV: Disables the generated latent views entirely, retaining only two base views. (2) PMWNB w/o MDLP: Disables the MDLP perturbation branch, retaining only the base view and the corresponding latent views derived from EWD. (3) PMWNB w/o EWD: Disables the EWD perturbation branch, retaining only the base view and the corresponding latent views derived from MDLP.

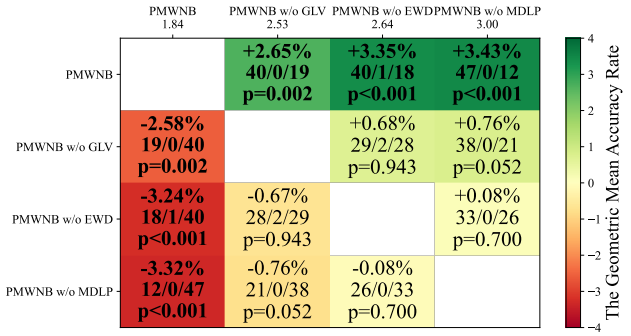
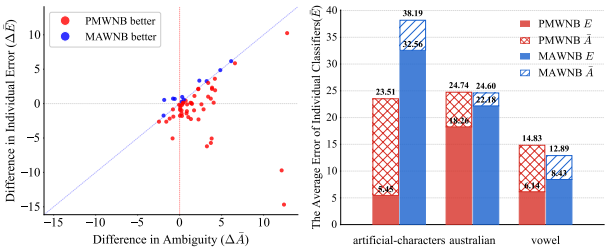


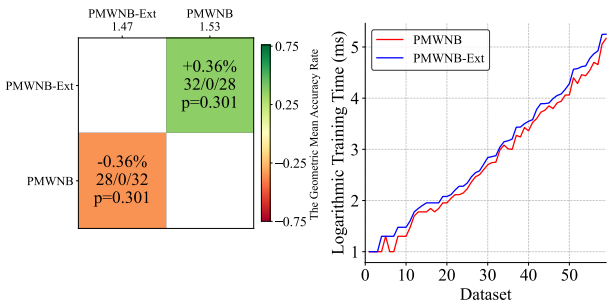
Figure 3: A summary of the pairwise classification accuracy comparison results among PMWNB and its three ablation variants. PMWNB w/o GLV disables the generated latent views; PMWNB w/o MDLP disables the MDLP perturbation branch, whereas PMWNB w/o EWD disables the EWD perturbation branch.

Figure 3 summarizes the classification performance of PMWNB and its three variants. PMWNB achieves the best average rank of 1.84, outperforming PMWNB w/o GLV (2.53), PMWNB w/o EWD (2.64), and PMWNB w/o MDLP (3.00). Compared to PMWNB w/o GLV, PMWNB yields a 2.65% improvement in geometric mean accuracy, achieving higher classification accuracy on 40 datasets. This confirms that jointly leveraging hard label and soft probability outputs of base classifiers can indeed capture richer data characteristics. Compared to PMWNB w/o EWD and PMWNB w/o MDLP, PMWNB achieves substantial accuracy gains of 3.35% and 3.43%, respectively. This demonstrates that relying on a fixed input distribution greatly limits the diversity of generated attributes, and fusing multiple perturbations



(a) Error-ambiguity difference (b) Error decomposition

Figure 4: Error-ambiguity analysis of PMWNB versus MAWNB. (a) Pairwise differences in ambiguity and individual error of PMWNB relative to MAWNB across 59 datasets. (b) Detailed decomposition of the average individual error ($\bar{E}$) into ensemble error (E) and ambiguity ($\bar{A}$) on three representative datasets.



(a) Classification accuracy (b) Logarithmic training time

Figure 5: A summary of the comparison results for PMWNB-Ext versus PMWNB. PMWNB-Ext refers to the extended version of PMWNB, which integrates three discretization methods including MDLP, PKID, and EWD. (a) Classification accuracy. (b) Logarithmic training time.

can effectively broaden the representation space. In conclusion, all these results validate the effectiveness and necessity of the matrix-view framework in PMWNB.

4.4 ERROR-AMBIGUITY DECOMPOSITION ANALYSIS

We further performed an error-ambiguity decomposition analysis to rigorously validate the effectiveness of the matrix-view in capturing diverse data representations. To this end, we employed the error-ambiguity decomposition formula [Krogh and Vedelsby, 1994] defined by Eq. (23).

$$E = \bar{E} - \bar{A} \tag{23}$$

where E, $\bar{E}$, and $\bar{A}$ denote the ensemble generalization error, the average individual error, and the ensemble ambiguity, respectively. We empirically approximated $\bar{E}$ and E using the classification errors to derive $\bar{A}$ for both PMWNB and MAWNB. Figure 4a compares the pairwise differences in ambiguity $\Delta\bar{A}$ and the individual error $\Delta\bar{E}$ achieved by PMWNB relative to MAWNB. In Figure 4a, red points below the blue dashed line indicate that PMWNB achieves lower ensemble error than MAWNB, whereas blue points above the line represent the contrary.

As shown in Figure 4a, PMWNB achieves higher ambiguity on 81.36% of the datasets (48/59), demonstrating its effectiveness in constructing distinct data representations. In particular, even among the 28 datasets in which PMWNB incurs higher individual errors, 64.29% of them still yield superior ensemble performance (18/28). This confirms that the increased ambiguity effectively compensates for the loss of individual accuracy, reducing the overall generalization error.

To provide an in-depth analysis of the sources of PMWNB's performance gains, Figure 4b presents three representative datasets that exhibit distinct error-ambiguity trade-off patterns. On the "artificial-characters" dataset, PMWNB achieves a simultaneous improvement by reducing the average individual error $\bar{E}$ while expanding the ambiguity $\bar{A}$, substantially lowering the final ensemble error E from 32.56 to 5.45. On the "australian" dataset, both models exhibit nearly identical individual performance; however, PMWNB yields a higher ambiguity $\bar{A}$, which in turn reduces the ensemble error E from 22.18 to 18.26. Finally, on the "vowel" dataset, PMWNB sacrifices individual precision with a higher $\bar{E}$ of 14.83 compared to 12.89 for MAWNB, yet it achieves a significantly larger $\bar{A}$ of 8.69, compare with 4.46 for MAWNB, ultimately securing a lower ensemble error of 6.14.

Overall, these results show that the matrix-view representation is highly effective in capturing diverse data characteristics, promoting a substantial increase in ensemble ambiguity to offset individual errors for superior generalization.

4.5 MULTI-PERTURBATION EXTENSIONS

To explore the extensibility and flexibility of the proposed matrix-view framework, we further extended the perturbation strategies in PMWNB by combining three widely adopted discretization methods: Minimum Description Length Principle (MDLP), Proportional K-Interval Discretization (PKID) [Yang and Webb, 2001], and Equal Width Discretization (EWD). We denote this extended variant as PMWNB-Ext. As shown in Figure 5a, PMWNB-Ext achieves a slightly lower average rank (1.47) than PMWNB (1.53), thereby indicating the framework's capacity to extend the latent space via additional perturbations.

However, PMWNB-Ext yields no statistically significant improvement over PMWNB ($p = 0.301$). As shown in Figure 5b, PMWNB-Ext incurs additional computational overhead compared to PMWNB. This additional overhead

arises because the multi-perturbation strategy requires constructing and learning extra views within the matrix-view framework. The complete classification accuracy and training time comparison for all 59 datasets is provided in Appendix B.2. Therefore, considering both predictive performance and computational efficiency, PMWNB adopts the combination of MDLP and EWD as its default configuration.

5 CONCLUSIONS AND FUTURE WORK

Most existing multi-view weighted NB methods rely on a fixed input distribution and exclusively utilize hard labels to construct latent views, which substantially constrains the diversity of generated attributes within the latent view space. To this end, we propose a novel framework called Perturbation-driven Matrix-view Weighted Naive Bayes (PMWNB). PMWNB employs diverse discretization strategies as input perturbations to construct a matrix view, and then jointly leverages both hard label and soft probability outputs to capture fine-grained discriminative evidence. Extensive experimental results on a collection of 59 UCI datasets show its superiority. The ablation study and in-depth analysis further validate the effectiveness and extensibility of the matrix-view weighting framework.

However, the perturbation strategies employed in this study are specifically designed for numeric attributes. Investigating more general perturbation techniques for mixed-type data is a primary direction for future work.

Author Contributions

Siyao Wu: Writing – original draft, Methodology, Conceptualization, Validation. Huan Zhang: Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. Kexin Meng: Writing – review & editing, Validation. Zhipeng Ding: Writing – review & editing, Validation. Pei Lv: Writing – review & editing, Supervision.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China (62406294), Henan Province Outstanding Youth Science Fund Program (262300421222), the Postdoctoral Fellowship Program of China Postdoctoral Science Foundation (2025M771579), the College Students' Innovation and Entrepreneurship Training Program of Zhengzhou University (202610459156) and the Scientific and Technological Development Scheme of Jilin Province (20250205059GH).

References

Leo Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, 2001.

Shenglei Chen, Ana M. Martínez, Geoffrey I. Webb, and Limin Wang. Sample-based attribute selective a*n*DE for large data. *IEEE Trans. Knowl. Data Eng.*, 29(1):172–185, 2017.

Shenglei Chen, Geoffrey I. Webb, Linyuan Liu, and Xin Ma. A novel selective naïve Bayes algorithm. *Knowl. Based Syst.*, 192:105361, 2020.

Janez Demsar. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.*, 7:1–30, 2006.

Diab M. Diab and Khalil M. El Hindi. Using differential evolution for improving distance measures of nominal values. *Appl. Soft Comput.*, 64:14–34, 2018.

Usama M. Fayyad and Keki B. Irani. Multi-interval discretization of continuous-valued attributes for classification learning. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, pages 1022–1029, 1993.

Eibe Frank, Mark A. Hall, and Bernhard Pfahringer. Locally weighted naive Bayes. In *Proceedings of the 19th Conference in Uncertainty in Artificial Intelligence*, pages 249–256, 2003.

Nir Friedman, Dan Geiger, and Moisés Goldszmidt. Bayesian network classifiers. *Mach. Learn.*, 29(2-3): 131–163, 1997.

Mark A. Hall. Correlation-based feature selection for discrete and numeric class machine learning. In *Proceedings of the 17th International Conference on Machine Learning*, pages 359–366, 2000.

Mark A. Hall. A decision tree-based attribute weighting filter for naive Bayes. *Knowl. Based Syst.*, 20(2):120–126, 2007.

Liangxiao Jiang, Lungan Zhang, Chaoqun Li, and Jia Wu. A correlation-based feature weighting filter for naive Bayes. *IEEE Trans. Knowl. Data Eng.*, 31(2):201–213, 2019.

Ron Kohavi. Scaling up the accuracy of naive-Bayes classifiers: A decision-tree hybrid. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, pages 202–207, 1996.

Anders Krogh and Jesper Vedelsby. Neural network ensembles, cross validation, and active learning. In *Proceedings of the 7th International Conference on Neural Information Processing Systems*, pages 231–238, 1994.

Pat Langley, Wayne Iba, and Kevin Thompson. An analysis of Bayesian classifiers. In *Proceedings of the 10th National Conference on Artificial Intelligence*, pages 223–228, 1992.

Guanzhi Liu, Kexin Meng, Pei Lv, and Huan Zhang. EMAWNB: Enhanced multi-view attribute weighted naive Bayes. In *Proceedings of the 29th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 66–78, 2025a.

Han Liu, Yafei Zhang, Jingjing Xu, Pei Lv, and Huan Zhang. Adaptive non-disjoint discretization for tree augmented naive Bayes. In *Proceedings of the 21st International Conference on Advanced Data Mining and Applications*, pages 33–48, 2025b.

Kexin Meng, Huan Zhang, Liangxiao Jiang, Pei Lv, Shuo He, and Mingliang Xu. K-free dependence Bayesian classifiers. *IEEE Trans. Neural Networks Learn. Syst.*, 2026. doi: 10.1109/TNNLS.2026.3664196.

Baozheng Wang, Huan Zhang, Kexin Meng, and Yafei Zhang. Frequency-based lazy naive Bayes. *Appl. Soft Comput.*, 198:115249, 2026.

Geoffrey I. Webb, Janice R. Boughton, and Zhihai Wang. Not so naive Bayes: Aggregating one-dependence estimators. *Mach. Learn.*, 58(1):5–24, 2005.

Ian H. Witten, Eibe Frank, and Mark A. Hall. *Data mining: practical machine learning tools and techniques, 3rd Edition*. Morgan Kaufmann, Elsevier, 2011.

Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus F. M. Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael S. Steinbach, David J. Hand, and Dan Steinberg. Top 10 algorithms in data mining. *Knowl. Inf. Syst.*, 14(1): 1–37, 2008.

Wenqiang Xu, Liangxiao Jiang, and Liangjun Yu. An attribute value frequency-based instance weighting filter for naive Bayes. *J. Exp. Theor. Artif. Intell.*, 31:225–236, 2019.

Ying Yang and Geoffrey I. Webb. Proportional k-interval discretization for naive-Bayes classifiers. In *Proceedings of the 12th European Conference on Machine Learning*, pages 564–575, 2001.

Nayyar Abbas Zaidi, Jesús Cerquides, Mark James Carman, and Geoffrey I. Webb. Alleviating naive Bayes attribute independence assumption by attribute weighting. *J. Mach. Learn. Res.*, 14(1):1947–1988, 2013.

He Zhang, François Petitjean, and Wray L. Buntine. Bayesian network classifiers using ensembles and smoothing. *Knowl. Inf. Syst.*, 62(9):3457–3480, 2020a.

Huan Zhang and Liangxiao Jiang. Fine tuning attribute weighted naive Bayes. *Neurocomputing*, 488:402–411, 2022.

Huan Zhang, Liangxiao Jiang, and Liangjun Yu. Class-specific attribute value weighting for naive Bayes. *Inf. Sci.*, 508:260–274, 2020b.

Huan Zhang, Liangxiao Jiang, and Chaoqun Li. Attribute augmented and weighted naive Bayes. *Sci. China Inf. Sci.*, 65(12):222101, 2022.

Huan Zhang, Liangxiao Jiang, Wenjun Zhang, and Chaoqun Li. Multi-view attribute weighted naive Bayes. *IEEE Trans. Knowl. Data Eng.*, 35(7):7291–7302, 2023.

Huan Zhang, Kexin Meng, Pei Lv, Shuo He, and Mingliang Xu. A general dual-view framework for instance weighted naive Bayes. *Pattern Recognit.*, 171:112181, 2026.

Yuetan Zhao, Limin Wang, Xinyu Zhu, Taosheng Jin, Minghui Sun, and Xiongfei Li. Probability knowledge acquisition from unlabeled instance based on dual learning. *Knowl. Inf. Syst.*, 67(1):521–547, 2025.

Xiaoliang Zhou, Yongli Wang, Li Zhang, Anqi Huang, and Xiaoli Wang. An innovative multi-view collaborative optimization framework for weighted naive Bayes. *Knowl. Based Syst.*, 317:113378, 2025.

Ciyou Zhu, Richard H. Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-BFGS-B: fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.*, 23(4):550–560, 1997.

Learning Representations from Perturbation: A Novel Matrix-View Weighting Framework for Naive Bayes (Supplementary Material)

Siyao Wu[1] **Huan Zhang**[*1] **Kexin Meng**[1] **Zhipeng Ding**[1] **Pei Lv**[1]

[1]School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou, China

A ALGORITHM FRAMEWORK

The entire learning algorithm for PMWNB can be partitioned into training (PMWNB-Training) and classification (PMWNB-Classification) algorithms. They are in Algorithm 1 and Algorithm 2, respectively.

B MORE EXPERIMENTAL DETAILS

B.1 DATASET INFORMATION

Our experiments were conducted on a collection of 59 benchmark classification datasets from the University of California, Irvine (UCI) repository, which represents a wide range of domains, as listed in Table 1. All missing attribute values were replaced with modes for nominal attributes and the means for numeric attributes from the available data.

B.2 DETAILED EXPERIMENTAL RESULTS

In Table 2, we first present that the detailed classification accuracy comparisons for PMWNB versus their competitors: EMAWNB, MAWNB, A^{2}WNB, AODE, RF, CAVWNB, and NB. Then, the detailed classification accuracy comparisons for PMWNB versus its three ablation variants: PMWNB w/o GLV, PMWNB w/o MDLP and PMWNB w/o EWD are showed in Table 3. PMWNB w/o GLV disables the generated latent views entirely, retaining only two base views; PMWNB w/o MDLP disables the MDLP perturbation branch, retaining only the base view and the corresponding latent views derived from EWD, whereas PMWNB w/o EWD disables the EWD perturbation branch, retaining only the base view and the corresponding latent views derived from MDLP.

All results were obtained via 10 separate runs of stratified 10-fold cross-validation. The highest classification accuracy on each dataset is highlighted in bold. For each row, $\bullet$ and $\circ$ implies that the classification accuracy of PMWNB statistically upgrades or degrades compared to its competitors, based on a corrected paired two-tailed t-tests at the $p = 0.05$ significance level [Demsar, 2006]. Additionally, **Significant W/D/L** represents the number of datasets where PMWNB statistically wins, draws, or loses against a competitor based on the aforementioned t-test. Conversely, the **W/D/L** reports the strict wins, draws, and losses based on a direct comparison of the classification accuracy across 59 datasets. At last, the detailed classification accuracy and training time comparisons for PMWNB-Ext versus PMWNB are presented in Table 4.

[*]Corresponding author

Algorithm 1 PMWNB-Training(D, G, T)

Input: training dataset D, the number of input perturbation G, the number of representation types T.
Output: the sets of SPODE model $\mathcal{S}$, the sets of RT model $\mathcal{R}$, the sets of attribute weighted model in matrix $\mathcal{M}$.

//Base Views Construction
Map D into V_{11} by Eqs. (5)–(6);
Map D into V_{21} by Eqs. (7)–(9);
//Latent Views Construction
for $g = 1 \rightarrow G$ **do**
    //Base Classifiers Training
    for each attribute j $(j = 1, \ldots, m)$ in V_{g1} **do**
        Build a SPODE S_j^g with jth attribute as parent;
    end for
    for each attribute j $(j = 1, \ldots, m)$ in V_{g1} **do**
        Select $logm$ attributes from the raw attribute view randomly;
        Build a RT R_j^g using information gain selection measure using Eq. (13);
    end for
    //Latent Representation Learning
    for each instance $\mathbf{y}_i$ $(i = 1 \ldots n)$ in V_{g1} **do**
        for each SPODE model S_j^g $(j = 1, \ldots, m)$ **do**
            Predict posterior probabilities using Eq. (10);
        end for
        for each RT model R_j^g $(j = 1, \ldots, m)$ **do**
            Predict posterior probabilities using Eq. (14);
        end for
        Construct $\{\mathbf{z}^{gt} \mid t = 2, \ldots, 5\}$ using Eqs. (15)–(17);
    end for
end for
//Matrix-view Weighting
for $g = 1 \rightarrow G$ **do**
    for $t = 1 \rightarrow T$ **do**
        if V_{gt} is a probability view **then**
            Discretize V_{gt} based on strategy g;
        end if
        Build attribute weighted model M_{gt} using Eqs. (19)– (21);
    end for
end for
return $\mathcal{S}, \mathcal{R}, \mathcal{M}$

Algorithm 2 PMWNB-Classification($\mathbf{x}$, $\mathcal{S}$, $\mathcal{R}$, $\mathcal{M}$)

Input: a test instance $\mathbf{x}$, the sets of SPODE model $\mathcal{S}$, the sets of RT model $\mathcal{R}$, the sets of attribute weighted model in matrix $\mathcal{M}$.

Output: the predicted class label of the test instance $c(\mathbf{x})$.

Map $\mathbf{x}$ into $\mathbf{z}^{11}$ using Eqs. (5)–(6);
Map $\mathbf{x}$ into $\mathbf{z}^{21}$ using Eq. (9);
for $g = 1 \rightarrow G$ **do**
    for each SPODE model S_j^g $(j = 1, \ldots, m)$ **do**
        Predict the posterior probability of $\mathbf{z}^{g1}$ by Eq. (10);
    end for
    for each RT model R_j^g $(j = 1, \ldots, m)$ **do**
        Predict the posterior probability of $\mathbf{z}^{g1}$ using R_j^g by Eq. (14);
    end for
    Construct $\{\mathbf{z}^{gt}\}_{t=2}^{t=5}$ using Eqs. (15)–(17);
    if $g = 1$ **then**
        Discretize $\mathbf{z}^{g3}$ and $\mathbf{z}^{g5}$ into 10 equal-width bins over the $[0, 1]$ interval, using Eqs. (5)–(6);
    else
        if $g = 2$ **then**
            Discretize $\mathbf{z}^{g3}$ and $\mathbf{z}^{g5}$ using Eq. (9);
        end if
    end if
end for
for $g = 1 \rightarrow G$ **do**
    for $t = 1 \rightarrow T$ **do**
        Use the built M_{gt} to estimate the class membership probabilities of $\mathbf{z}^{gt}$ by Eq. (18);
    end for
end for
//Matrix-view Fusion
Fuse the estimated class-membership probabilities of all models by Eq. (22);
return $c(\mathbf{x})$

Table 1: Descriptions of 59 UCI benchmark datasets used in the experiments.

Dataset	Instance	Total Attribute	Numeric Attribute	Class
abalone	4177	8	8	3
appendicitis	106	7	7	2
artificial-characters	10218	7	7	10
australian	12546	8	8	3
autos	205	25	15	7
balance-scale	625	4	4	3
breast-c	699	9	9	2
breast-t	106	9	9	6
cardioto	2126	35	35	3
cardiotocography-2	2126	21	21	3
climate	540	20	20	2
connectionist-vowel	528	10	10	11
diabetes	768	8	8	2
ecoli	336	7	7	8
egg-eye	14980	14	14	2
energy-efficiency-y1	768	8	8	37
energy-efficiency-y2	768	8	8	38
fertility	100	9	9	2
glass	214	9	9	7
hayes-roth	160	4	4	3
heart-statlog	270	13	13	2
ionosphere	351	34	34	2
iris	150	4	4	3
labor	57	16	8	2
leaf	340	15	15	30
libras	360	90	90	15
liver	345	6	6	2
magic	19020	10	10	2
mammo	961	5	5	2
mfeat-fourier	2000	76	76	10
monks-1	556	6	6	2
new-thyroid	215	5	5	3
optdigits	5620	64	64	10
page-blocks	5473	10	10	5
parkinsons	195	22	22	2
planning	182	12	12	2
qsar-biodegradation	1055	41	41	2
robot24	5456	24	24	4
seeds	210	7	7	3
segment	2310	19	19	7
shuttle	58000	9	9	7
sonar	208	60	60	2
spectf-heart	349	44	44	2
statlog-g	1000	24	24	2
statlog-h	270	13	13	2
steel-plates-faults	1941	33	33	2
texture	5500	40	40	11
turkiye	5820	32	32	13
user	403	5	5	5
vehicle	846	18	18	4
vowel	990	13	10	11
waveform-5000	5000	40	40	3
wdbc	569	30	30	2
wholesale-c	440	7	7	2
wholesale-r	440	7	7	3
wilt	4839	5	5	2
wine-rec	178	13	13	3
wine-red	1599	11	11	6
wine-white	4898	11	11	7

Table 2: Classification accuracy comparisons for PMWNB versus EMAWNB, MAWNB, A^{2}WNB, AODE, RF, CAVWNB and NB.

Dataset	PMWNB	EMAWNB	MAWNB	A^{2}WNB	AODE	RF	CAVWNB	NB
abalone	54.50 ± 1.96	54.75 ± 2.01	54.52 ± 2.07	54.67 ± 2.40	**54.94** ± 2.09	53.58 ± 1.92	54.43 ± 2.10	52.07 ± 2.08 •
appendicitis	85.09 ± 8.72	85.95 ± 9.01	81.40 ± 8.56	84.17 ± 8.90	**86.20** ± 9.08	83.72 ± 7.62	80.43 ± 9.94	83.11 ± 9.66
artificial-characters	**94.55** ± 0.91	68.80 ± 1.37 •	67.44 ± 1.47 •	57.84 ± 1.77 •	56.62 ± 1.53 •	69.83 ± 1.47 •	46.44 ± 1.45 •	36.49 ± 1.35 •
australian	**81.74** ± 1.14	80.83 ± 1.20 •	77.82 ± 1.22 •	70.48 ± 1.36 •	70.81 ± 1.32 •	80.91 ± 1.09 •	64.69 ± 1.50 •	61.30 ± 1.51 •
autos	**83.75** ± 8.10	82.52 ± 8.70	81.36 ± 8.47	79.52 ± 9.03	73.91 ± 10.18 •	83.16 ± 8.27	77.91 ± 8.54 •	63.42 ± 11.59 •
balance-scale	85.50 ± 2.77	87.70 ± 2.44 ○	87.01 ± 2.41 ○	87.28 ± 3.64	89.78 ± 1.88 ○	72.96 ± 4.43 •	**91.94** ± 0.74 ○	91.44 ± 1.30 ○
breast-c	96.92 ± 1.86	96.90 ± 1.84	96.81 ± 2.01	96.02 ± 2.38	96.85 ± 1.90	96.12 ± 2.32	96.00 ± 2.06	**97.30** ± 1.75
breast-t	40.15 ± 11.56	40.87 ± 11.99	37.16 ± 11.53	34.02 ± 11.61	37.25 ± 12.80	34.15 ± 12.72	40.35 ± 12.67	**41.65** ± 12.77
cardioto	**98.47** ± 0.90	98.42 ± 0.95	98.18 ± 0.96	98.18 ± 0.95	97.08 ± 1.10 •	96.92 ± 1.39 •	98.33 ± 0.94	92.65 ± 1.80 •
cardiotocography-2	**93.47** ± 1.85	92.19 ± 1.68 •	92.14 ± 1.59 •	90.79 ± 1.86 •	89.40 ± 2.08 •	91.47 ± 1.72 •	91.90 ± 1.47 •	84.33 ± 2.32 •
climate	90.91 ± 1.65	91.19 ± 1.27	89.69 ± 1.90 •	83.50 ± 4.72 •	88.41 ± 2.57 •	**91.44** ± 1.08	87.11 ± 2.89 •	87.52 ± 3.22 •
connectionist-vowel	90.11 ± 4.00	**91.08** ± 3.97	90.13 ± 4.25	84.34 ± 4.91 •	89.51 ± 4.10	85.07 ± 4.86 •	83.84 ± 5.84 •	74.33 ± 6.26 •
diabetes	75.64 ± 4.64	75.01 ± 4.22	74.95 ± 4.88	76.36 ± 4.65	**76.38** ± 4.58	72.75 ± 5.22	75.17 ± 4.10	75.77 ± 4.86
ecoli	**84.07** ± 5.51	83.34 ± 5.68	82.00 ± 6.21	78.15 ± 6.39 •	81.52 ± 5.67	74.70 ± 6.27 •	83.40 ± 5.55	81.91 ± 5.76
egg-eye	**79.05** ± 1.09	57.19 ± 1.17 •	57.19 ± 1.17 •	57.18 ± 1.16 •	57.19 ± 1.15 •	57.20 ± 1.16 •	57.19 ± 1.16 •	57.15 ± 1.14 •
energy-eff-y1	67.95 ± 3.73	**72.60** ± 3.78 ○	67.78 ± 3.51	65.28 ± 4.05 •	58.58 ± 4.79 •	65.87 ± 4.16	53.25 ± 4.79 •	45.63 ± 5.14 •
energy-eff-y2	54.31 ± 4.47	**57.89** ± 4.75 ○	54.48 ± 5.03	52.00 ± 4.20	49.91 ± 4.78 •	52.84 ± 4.73	51.75 ± 4.73	47.42 ± 4.66 •
fertility	**88.00** ± 4.92	87.50 ± 5.75	85.00 ± 8.35	85.90 ± 6.21	**88.00** ± 4.02	85.80 ± 8.43	84.40 ± 8.57	87.90 ± 4.09
glass	**73.18** ± 9.57	63.88 ± 9.49 •	63.72 ± 8.90 •	61.99 ± 9.52 •	63.74 ± 9.30 •	62.01 ± 9.55 •	60.26 ± 9.90 •	59.93 ± 9.42 •
hayes-roth	82.37 ± 8.01	81.81 ± 8.10	79.81 ± 8.88	77.19 ± 8.45	71.00 ± 8.91 •	74.75 ± 10.36 •	**83.81** ± 7.85	82.63 ± 8.78
heart-statlog	**83.89** ± 6.06	83.52 ± 6.12	82.89 ± 6.31	81.26 ± 6.42	83.67 ± 5.16	78.44 ± 6.85 •	81.93 ± 5.70	**83.89** ± 5.38
ionosphere	92.20 ± 4.43	92.00 ± 4.26	91.89 ± 4.52	**92.28** ± 4.15	91.74 ± 4.28	90.92 ± 4.90	90.97 ± 4.49	90.86 ± 4.33
iris	94.27 ± 5.77	95.40 ± 5.08	**95.53** ± 4.84	94.47 ± 5.61	94.33 ± 5.87	94.53 ± 5.22	95.40 ± 5.25	94.60 ± 6.61
labor	92.60 ± 11.63	94.43 ± 9.16	94.97 ± 8.95	92.07 ± 11.20	94.57 ± 9.72	88.87 ± 12.88	96.00 ± 7.76	**96.70** ± 7.27
leaf	**66.18** ± 8.12	59.18 ± 7.79 •	56.74 ± 8.79 •	47.35 ± 8.61 •	55.97 ± 8.04 •	54.62 ± 9.42 •	57.15 ± 9.12 •	55.47 ± 7.95 •
libras	**79.47** ± 6.03	77.81 ± 6.34	75.25 ± 6.88 •	72.61 ± 6.74 •	75.81 ± 6.28	72.28 ± 7.68 •	68.50 ± 7.84 •	64.25 ± 8.15 •
liver	64.24 ± 7.33	66.16 ± 7.05	65.23 ± 7.70	65.29 ± 8.10	65.43 ± 7.28	61.71 ± 7.81	**67.03** ± 7.39	64.47 ± 7.45
magic	**85.26** ± 0.71	83.05 ± 0.75 •	83.29 ± 0.69 •	81.41 ± 0.75 •	81.15 ± 0.79 •	81.34 ± 0.79 •	82.72 ± 0.83 •	74.76 ± 0.68 •
mammo	**82.55** ± 3.57	80.53 ± 3.27 •	80.75 ± 3.36 •	79.81 ± 3.66 •	80.97 ± 3.64	78.39 ± 3.55 •	80.87 ± 3.73	80.59 ± 3.86
mfeat-fourier	**81.55** ± 2.52	81.09 ± 2.40	80.00 ± 2.55 •	78.81 ± 2.34 •	79.33 ± 2.32 •	80.01 ± 2.50 •	77.95 ± 2.52 •	76.60 ± 2.69 •
monks-1	99.14 ± 1.66	99.98 ± 0.18	99.17 ± 1.58	**100.00** ± 0.00	82.07 ± 4.61 •	96.96 ± 2.93 •	74.66 ± 4.54 •	74.64 ± 4.54 •
new-thyroid	**95.81** ± 4.06	94.65 ± 5.01	94.80 ± 4.61	93.30 ± 5.14	91.68 ± 4.90 •	93.20 ± 5.20	94.51 ± 4.39	91.95 ± 5.02 •
optdigits	96.28 ± 0.81	**96.40** ± 0.81	95.55 ± 0.93 •	96.14 ± 0.81	95.66 ± 0.85 •	94.07 ± 0.98 •	95.40 ± 0.86 •	92.27 ± 1.08 •
page-blocks	**96.43** ± 0.61	94.21 ± 0.75 •	94.18 ± 0.73 •	93.58 ± 0.76 •	93.27 ± 0.80 •	93.69 ± 0.77 •	93.93 ± 0.73 •	92.36 ± 0.91 •
parkinsons	90.39 ± 7.10	**92.03** ± 6.53	90.44 ± 6.83	89.87 ± 6.20	88.92 ± 6.37	91.88 ± 7.08	89.70 ± 6.22	76.53 ± 8.97 •
planning	65.62 ± 8.62	64.16 ± 7.70	56.73 ± 11.40 •	58.77 ± 9.62 •	63.26 ± 9.36	**68.07** ± 7.12	52.78 ± 10.84 •	56.45 ± 9.64 •
qsar-biodegradation	**86.38** ± 3.32	86.03 ± 3.37	84.87 ± 3.39 •	84.61 ± 3.58	81.99 ± 3.80 •	84.26 ± 3.51 •	83.68 ± 3.58 •	79.55 ± 4.16 •
robot24	**97.38** ± 0.73	95.12 ± 0.95 •	94.68 ± 1.00 •	89.91 ± 1.29 •	89.65 ± 1.35 •	95.01 ± 0.85 •	92.53 ± 1.10 •	80.30 ± 1.62 •
seeds	91.29 ± 5.58	91.90 ± 5.84	91.95 ± 5.93	**93.10** ± 5.26	90.33 ± 6.44	89.86 ± 6.79	92.38 ± 5.82	88.95 ± 7.06
segment	**96.61** ± 1.14	96.00 ± 1.41	95.43 ± 1.46 •	94.61 ± 1.48 •	92.91 ± 1.39 •	95.72 ± 1.44 •	93.97 ± 1.54 •	89.05 ± 1.62 •
shuttle	**99.86** ± 0.05	95.34 ± 1.63 •	95.31 ± 1.63 •	95.04 ± 1.76 •	94.32 ± 1.98 •	95.34 ± 1.63 •	95.25 ± 1.58 •	90.27 ± 2.74 •
sonar	79.78 ± 9.14	79.58 ± 8.92	79.36 ± 9.73	**82.67** ± 7.35	80.30 ± 9.54	76.95 ± 8.96	77.39 ± 10.17	76.11 ± 10.47
spectf-heart	91.38 ± 4.53	91.81 ± 4.05	91.30 ± 4.61	90.66 ± 4.66	87.46 ± 5.12 •	**92.35** ± 4.12	90.52 ± 4.61	77.26 ± 7.53 •
statlog-g	76.16 ± 3.24	75.86 ± 3.04	**76.23** ± 3.18	75.19 ± 3.18	75.63 ± 3.37	74.82 ± 2.96	75.05 ± 3.37	75.54 ± 3.62
statlog-h	83.74 ± 5.95	83.44 ± 5.92	82.59 ± 6.35	81.11 ± 5.92	83.67 ± 5.16	78.59 ± 6.61 •	81.93 ± 5.70	**83.89** ± 5.38
steel-plates-faults	99.22 ± 0.62	99.25 ± 0.66	98.74 ± 0.93	97.36 ± 1.18 •	89.96 ± 2.07 •	95.02 ± 1.89 •	**99.73** ± 0.35 ○	94.72 ± 1.68 •
texture	97.25 ± 0.76	**97.29** ± 0.64	97.13 ± 0.72	95.95 ± 0.92 •	94.49 ± 0.98 •	97.07 ± 0.75	96.31 ± 0.76 •	79.34 ± 1.69 •
turkiye	36.81 ± 1.68	36.17 ± 1.86	34.11 ± 1.79 •	37.12 ± 1.72	34.62 ± 1.61 •	35.06 ± 1.55 •	**37.95** ± 1.74 ○	22.12 ± 2.13 •
user	**89.50** ± 4.94	89.30 ± 4.87	87.78 ± 5.17	83.03 ± 5.79 •	80.42 ± 5.58 •	86.74 ± 5.73	83.93 ± 5.10 •	81.37 ± 4.61 •
vehicle	73.18 ± 4.05	**73.77** ± 3.58	72.53 ± 4.26	71.23 ± 4.16	71.48 ± 3.44	71.28 ± 4.05	71.06 ± 3.82	61.22 ± 3.58 •
vowel	**93.86** ± 2.71	93.49 ± 2.41	91.57 ± 2.92 •	88.39 ± 3.32 •	89.84 ± 3.17 •	89.73 ± 3.20 •	81.55 ± 4.19 •	65.94 ± 4.88 •
waveform-5000	**85.48** ± 1.56	84.99 ± 1.59	84.53 ± 1.62 •	84.00 ± 1.70 •	84.16 ± 1.65 •	82.20 ± 1.72 •	82.98 ± 1.63 •	80.02 ± 1.43 •
wdbc	95.96 ± 2.84	95.87 ± 2.72	95.64 ± 2.69	**96.36** ± 2.28	94.75 ± 2.74 •	94.98 ± 2.64	95.41 ± 2.76	94.29 ± 2.94 •
wholesale-c	**90.52** ± 4.16	87.25 ± 4.76 •	86.93 ± 4.72 •	86.41 ± 4.60 •	87.59 ± 4.48 •	86.14 ± 4.70 •	87.61 ± 4.30 •	87.02 ± 4.37 •
wholesale-r	70.95 ± 1.50	**71.23** ± 1.45	69.91 ± 2.53	69.64 ± 2.41	70.98 ± 1.93	66.45 ± 4.02 •	69.80 ± 2.45	69.11 ± 2.91 •
wilt	**94.67** ± 0.20	94.62 ± 0.19	94.62 ± 0.18	94.61 ± 0.17	94.60 ± 0.08	94.59 ± 0.20	94.60 ± 0.16	94.59 ± 0.09
wine-rec	**97.97** ± 3.04	96.45 ± 4.28	96.28 ± 3.93	95.32 ± 4.73	96.29 ± 4.18	94.03 ± 5.59 •	97.02 ± 3.86	96.63 ± 3.82
wine-red	**66.89** ± 3.55	66.82 ± 3.67	66.49 ± 3.79	62.31 ± 3.42 •	62.14 ± 3.63 •	66.45 ± 3.08	60.04 ± 3.63 •	58.18 ± 3.79 •
wine-white	**64.69** ± 2.24	64.57 ± 2.17	64.10 ± 1.98	53.73 ± 1.93 •	53.87 ± 1.96 •	64.13 ± 2.09	52.41 ± 1.88 •	47.86 ± 2.13 •
Average	**83.48**	82.22	81.15	79.56	79.26	79.84	78.90	75.31
Significant W/D/L	-	12/44/3	22/36/1	27/32/0	33/25/1	31/28/0	28/28/3	39/19/1
W/D/L	-	39/0/20	48/0/11	48/0/11	49/1/9	54/0/5	50/0/9	49/1/9

Table 3: Classification accuracy comparisons for PMWNB versus its three variants.

Dataset	PMWNB	PMWNB w/o GLV		PMWNB w/o MDLP		PMWNB w/o EWD	
abalone	54.50 ± 1.96	**55.01** ± 2.10		54.36 ± 2.16		53.32 ± 2.04	
appendicitis	85.09 ± 8.72	84.54 ± 8.58		81.20 ± 8.67		**86.58** ± 8.69	
artificial-characters	**94.55** ± 0.91	83.37 ± 1.15	•	69.85 ± 1.40	•	94.13 ± 0.87	•
australian	**81.74** ± 1.14	67.58 ± 1.33	•	80.40 ± 1.18	•	79.81 ± 1.46	•
autos	**83.75** ± 8.10	79.78 ± 9.06		81.37 ± 8.50		82.77 ± 8.25	
balance-scale	85.50 ± 2.77	**90.16** ± 1.98	○	85.39 ± 2.87		71.13 ± 4.73	•
breast-c	96.92 ± 1.86	96.48 ± 1.92		**97.04** ± 1.90		96.48 ± 2.22	
breast-t	40.15 ± 11.56	**42.44** ± 13.86		37.27 ± 10.95		34.63 ± 10.00	
cardioto	98.47 ± 0.90	**98.76** ± 0.77		97.98 ± 1.01	•	98.51 ± 0.83	
cardiotocography-2	93.47 ± 1.85	**93.62** ± 1.47		92.09 ± 1.67	•	93.52 ± 1.79	
climate	90.91 ± 1.65	89.56 ± 2.59		88.96 ± 2.55	•	**90.98** ± 3.04	
connectionist-vowel	90.11 ± 4.00	84.50 ± 5.05	•	**91.16** ± 3.98		75.51 ± 5.67	•
diabetes	**75.64** ± 4.64	75.42 ± 4.23		74.65 ± 4.92		74.69 ± 4.94	
ecoli	84.07 ± 5.51	**85.05** ± 5.19		80.39 ± 5.86	•	80.50 ± 5.23	•
egg-eye	79.05 ± 1.09	72.88 ± 1.26	•	57.20 ± 1.16	•	**79.91** ± 1.02	○
energy-efficiency-y1	**67.95** ± 3.73	55.06 ± 5.08	•	67.55 ± 3.88		64.84 ± 4.31	•
energy-efficiency-y2	**54.31** ± 4.47	52.54 ± 4.37		53.98 ± 4.61		54.17 ± 4.29	
fertility	**88.00** ± 4.92	86.80 ± 5.66		84.00 ± 8.88		**88.00** ± 4.02	
glass	73.18 ± 9.57	70.93 ± 8.98		63.14 ± 8.44	•	**74.74** ± 9.86	
hayes-roth	82.37 ± 8.01	**84.81** ± 7.50		79.25 ± 9.40	•	60.00 ± 3.08	•
heart-statlog	83.89 ± 6.06	**84.67** ± 6.03		82.30 ± 6.37		84.56 ± 6.49	
ionosphere	92.20 ± 4.43	90.89 ± 4.82		**92.34** ± 4.45		91.48 ± 4.44	
iris	94.27 ± 5.77	94.93 ± 5.45		**95.20** ± 5.28		93.87 ± 5.82	
labor	92.60 ± 11.63	93.77 ± 10.98		**94.23** ± 10.37		89.30 ± 13.13	
leaf	**66.18** ± 8.12	65.65 ± 8.55		54.76 ± 7.63	•	64.88 ± 7.50	
libras	**79.47** ± 6.03	73.83 ± 6.86	•	75.56 ± 6.65	•	75.53 ± 6.46	•
liver	64.24 ± 7.33	**66.77** ± 7.05		64.56 ± 7.79		56.85 ± 4.20	•
magic	**85.26** ± 0.71	84.72 ± 0.77	•	83.21 ± 0.68	•	84.71 ± 0.71	•
mammo	**82.55** ± 3.57	82.40 ± 3.36		80.03 ± 3.52	•	82.48 ± 3.37	
mfeat-fourier	81.55 ± 2.52	80.58 ± 2.28		80.18 ± 2.58	•	**81.98** ± 2.56	
monks-1	99.14 ± 1.66	74.64 ± 4.54	•	**99.26** ± 1.49		74.64 ± 4.54	•
new-thyroid	95.81 ± 4.06	**95.99** ± 3.50		94.90 ± 4.51		94.74 ± 4.47	
optdigits	**96.28** ± 0.81	96.16 ± 0.71		95.34 ± 0.87	•	96.20 ± 0.83	
page-blocks	96.43 ± 0.61	96.33 ± 0.65		94.35 ± 0.71	•	**96.90** ± 0.69	○
parkinsons	90.39 ± 7.10	88.56 ± 7.40		**90.64** ± 6.86		87.42 ± 7.03	
planning	65.62 ± 8.62	60.08 ± 9.29		55.96 ± 11.10	•	**71.46** ± 1.53	
qsar-biodegradation	**86.38** ± 3.32	85.76 ± 3.30		84.61 ± 3.13	•	85.78 ± 3.27	
robot24	97.38 ± 0.73	97.68 ± 0.67		94.87 ± 0.98	•	**97.89** ± 0.62	○
seeds	91.29 ± 5.58	90.57 ± 5.90		**91.67** ± 6.07		89.10 ± 6.21	
segment	**96.61** ± 1.14	95.39 ± 1.31	•	95.95 ± 1.40		96.06 ± 1.23	
shuttle	99.86 ± 0.05	99.85 ± 0.04		95.33 ± 1.63	•	**99.94** ± 0.03	○
sonar	79.78 ± 9.14	79.12 ± 9.90		**80.08** ± 9.64		76.64 ± 9.74	
spectf-heart	**91.38** ± 4.53	89.40 ± 4.89		91.24 ± 4.65		90.38 ± 4.58	
statlog-g	**76.16** ± 3.24	75.66 ± 3.09		75.11 ± 3.33		73.71 ± 3.44	•
statlog-h	83.74 ± 5.95	**84.67** ± 6.03		81.93 ± 5.68		84.44 ± 6.82	
steel-plates-faults	99.22 ± 0.62	**99.78** ± 0.32	○	98.30 ± 1.09	•	99.27 ± 0.59	
texture	**97.25** ± 0.76	96.37 ± 0.85	•	97.07 ± 0.71		96.41 ± 0.79	•
turkiye	36.81 ± 1.68	**38.54** ± 1.49	○	34.29 ± 1.68	•	34.35 ± 1.88	•
user	**89.50** ± 4.94	85.78 ± 4.64	•	87.91 ± 5.00		85.41 ± 5.12	•
vehicle	**73.18** ± 4.05	72.84 ± 3.97		71.97 ± 4.16		71.39 ± 4.49	
vowel	**93.86** ± 2.71	83.79 ± 3.94	•	92.24 ± 2.76		86.16 ± 3.75	•
waveform-5000	**85.48** ± 1.56	84.95 ± 1.57		84.19 ± 1.70	•	84.69 ± 1.52	•
wdbc	95.96 ± 2.84	**96.29** ± 2.64		95.52 ± 2.81		95.59 ± 3.05	
wholesale-c	90.52 ± 4.16	90.50 ± 4.12		86.75 ± 4.39	•	**90.95** ± 4.32	
wholesale-r	70.95 ± 1.50	71.41 ± 1.34		67.80 ± 3.89	•	**71.82** ± 1.12	
wilt	94.67 ± 0.20	94.62 ± 0.10		94.61 ± 0.18		**96.49** ± 0.81	○
wine-rec	97.97 ± 3.04	**98.04** ± 3.12		96.28 ± 4.11		98.03 ± 3.15	
wine-red	66.89 ± 3.55	60.54 ± 3.44	•	**67.47** ± 3.40		60.29 ± 3.56	•
wine-white	64.69 ± 2.24	54.42 ± 1.69	•	**65.40** ± 1.97		58.56 ± 2.40	•
Average	**83.48**	81.44		81.03		81.09	
Significant W/D/L	-	14/42/3		25/34/0		19/35/5	
W/D/L	-	40/0/19		47/0/12		40/1/18	

Table 4: Classification accuracy and training time comparisons for PMWNB-Ext versus PMWNB.

Dataset	Classification Accuracy PMWNB-Ext	PMWNB		Training Time PMWNB-Ext	PMWNB	
abalone	53.79 ± 2.04	54.50 ± 1.96		2.72± 0.05	1.74± 0.08	○
appendicitis	84.53 ± 8.73	85.09 ± 8.72		0.01± 0.00	0.01± 0.00	○
artificial-characters	95.86 ± 0.79	94.55 ± 0.91	●	42.68± 0.86	27.43± 0.93	○
australian	81.80 ± 1.13	81.74 ± 1.14		7.80± 0.12	5.31± 0.13	○
autos	83.64 ± 8.40	83.75 ± 8.10		0.75± 0.07	0.56± 0.04	○
balance-scale	86.75 ± 2.58	85.50 ± 2.77		0.09± 0.01	0.07± 0.01	○
breast-c	97.17 ± 1.81	96.92 ± 1.86		0.12± 0.01	0.09± 0.01	○
breast-t	40.55 ± 11.70	40.15 ± 11.56		0.06± 0.01	0.05± 0.01	○
cardioto	98.16 ± 1.02	98.47 ± 0.90		3.53± 0.27	2.30± 0.15	○
cardiotocography-2	93.51 ± 1.81	93.47 ± 1.85		2.71± 0.11	1.88± 0.06	○
climate	91.04 ± 2.05	90.91 ± 1.65		0.12± 0.02	0.09± 0.02	○
connectionist-vowel	91.46 ± 3.61	90.11 ± 4.00		1.41± 0.12	1.21± 0.09	○
diabetes	75.38 ± 4.87	75.64 ± 4.64		0.16± 0.01	0.13± 0.01	○
ecoli	84.59 ± 5.69	84.07 ± 5.51		0.51± 0.03	0.40± 0.02	○
egg-eye	76.71 ± 1.17	79.05 ± 1.09	○	6.07± 0.47	4.00± 0.48	○
energy-efficiency-y1	68.97 ± 3.86	67.95 ± 3.73		36.75± 0.98	25.04± 0.99	○
energy-efficiency-y2	55.24 ± 4.42	54.31 ± 4.47		41.28± 0.98	28.72± 1.03	○
fertility	87.20 ± 6.83	88.00 ± 4.92		0.01± 0.00	0.01± 0.00	○
glass	75.73 ± 8.50	73.18 ± 9.57		0.38± 0.02	0.32± 0.03	○
hayes-roth	82.25 ± 8.03	82.37 ± 8.01		0.02± 0.00	0.01± 0.00	○
heart-statlog	83.33 ± 5.87	83.89 ± 6.06		0.08± 0.01	0.06± 0.01	○
ionosphere	92.03 ± 4.13	92.20 ± 4.43		0.19± 0.02	0.14± 0.01	○
iris	94.73 ± 5.30	94.27 ± 5.77		0.02± 0.00	0.01± 0.00	○
labor	92.23 ± 12.15	92.60 ± 11.63		0.01± 0.00	0.01± 0.00	○
leaf	65.91 ± 7.68	66.18 ± 8.12		19.19± 3.14	11.58± 3.20	○
libras	79.28 ± 6.30	79.47 ± 6.03		37.78± 3.33	19.32± 3.18	○
liver	67.80 ± 7.64	64.24 ± 7.33		0.03± 0.00	0.02± 0.00	○
magic	85.30 ± 0.74	85.26 ± 0.71		7.83± 0.24	5.81± 0.46	○
mammo	82.85 ± 3.56	82.55 ± 3.57		0.19± 0.01	0.13± 0.01	○
mfeat-fourier	81.56 ± 2.47	81.55 ± 2.52		59.83± 1.98	34.96± 2.19	○
monks-1	99.91 ± 0.47	99.14 ± 1.66		0.03± 0.01	0.02± 0.00	○
new-thyroid	97.25 ± 3.31	95.81 ± 4.06		0.02± 0.01	0.02± 0.00	○
optdigits	95.98 ± 0.83	96.28 ± 0.81	○	83.89±14.27	45.24±12.37	○
page-blocks	96.73 ± 0.63	96.43 ± 0.61	●	9.82± 0.37	6.26± 0.17	○
parkinsons	91.26 ± 6.44	90.39 ± 7.10		0.07± 0.01	0.06± 0.01	○
planning	65.25 ± 7.85	65.62 ± 8.62		0.02± 0.00	0.01± 0.00	○
qsar-biodegradation	85.75 ± 3.61	86.38 ± 3.32		1.59± 0.08	1.00± 0.04	○
robot24	97.34 ± 0.71	97.38 ± 0.73		12.07± 0.44	8.73± 0.37	○
seeds	91.52 ± 6.08	91.29 ± 5.58		0.04± 0.01	0.03± 0.00	○
segment	96.42 ± 1.12	96.61 ± 1.14		11.27± 0.59	8.07± 0.51	○
shuttle	99.93 ± 0.03	99.86 ± 0.05	●	176.46± 9.71	113.43± 8.61	○
sonar	79.06 ± 9.10	79.78 ± 9.14		0.22± 0.02	0.17± 0.02	○
spectf-heart	91.29 ± 4.72	91.38 ± 4.53		0.35± 0.05	0.29± 0.03	○
statlog-g	75.70 ± 2.89	76.16 ± 3.24		0.69± 0.03	0.50± 0.03	○
statlog-h	83.19 ± 6.06	83.74 ± 5.95		0.09± 0.02	0.06± 0.01	○
steel-plates-faults	98.77 ± 0.89	99.22 ± 0.62	○	1.47± 0.11	1.02± 0.05	○
texture	96.56 ± 0.77	97.25 ± 0.76	○	73.77± 3.68	50.11± 3.67	○
turkiye	37.03 ± 1.78	36.81 ± 1.68		177.85± 8.90	147.55± 5.48	○
user	89.83 ± 4.74	89.50 ± 4.94		0.13± 0.01	0.11± 0.01	○
vehicle	72.52 ± 3.86	73.18 ± 4.05		1.11± 0.06	0.98± 0.08	○
vowel	93.88 ± 2.60	93.86 ± 2.71		3.81± 0.38	3.36± 0.28	○
waveform-5000	85.31 ± 1.59	85.48 ± 1.56		8.04± 0.85	7.07± 0.31	○
wdbc	95.87 ± 2.94	95.96 ± 2.84		0.29± 0.02	0.23± 0.02	○
wholesale-c	91.39 ± 4.26	90.52 ± 4.16		0.09± 0.01	0.07± 0.01	○
wholesale-r	71.59 ± 1.43	70.95 ± 1.50		0.09± 0.01	0.06± 0.01	○
wilt	95.36 ± 0.46	94.67 ± 0.20	●	0.72± 0.06	0.55± 0.06	○
wine-rec	97.69 ± 3.22	97.97 ± 3.04		0.03± 0.01	0.02± 0.01	○
wine-red	69.65 ± 3.63	66.89 ± 3.55	●	3.17± 0.07	2.67± 0.06	○
wine-white	68.65 ± 1.81	64.69 ± 2.24	●	14.95± 0.59	11.48± 0.24	○
Average	**83.73**	83.48		14.48	**9.84**	
Significant W/D/L	-	6/49/4		-	0/0/59	
W/D/L	-	31/0/28		-	0/0/59	