Supplementary Materials for the manuscript: Multi View Twin Random Vector Functional Link with Pinball Loss

[1*]**Deepak Gupta** ***Sr. Member IEEE***, [1]**Roshan Kumar Gupta**, [1]**Jyoti Maurya**, [2]**Witold Pedrycz**
[1]*Department of Computer Science and Engineering,*
Motilal Nehru National Institute of Technology Allahabad, Prayagraj
[2]*Department of Electrical and Computer Engineering,*
University of Alberta, Canada
[1*]*deepakg@mnnit.ac.in,* [1]*roshan.2023ds19@mnnit.ac.in,* [1]*jyoti.2023rcs02@mnnit.ac.in,* [2]*wpedrycz@ualberta.ca*

A Mathematical Formulations of Related Models

A.1. Support Vector Machine (SVM)

SVM aims to find the maximum-margin hyperplane separating two classes [1]. Let us consider training data $x \in \mathbb{R}^{n\times d}$ with input weight $w \in \mathbb{R}^{n\times d}$ and bias $b \in \mathbb{R}^{n\times 1}$. Taking slack variable $\xi \in \mathbb{R}^{n\times 1}$ the primal optimization problem becomes:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}\|w\|^2 + C\sum_{j=1}^{n}\xi_i \\ \text{s.t.} \quad & y_i(w_i^\top x_i + b_i) \geq 1-\xi_i, \quad \xi_i \geq 0. \end{aligned} \tag{1}$$

The final decision function of separating hyperplane for a given data u will be:

$$Class(u) = sign(w^\top * u + b). \tag{2}$$

A.2. Twin Support Vector Machine (TWSVM)

TWSVM [2,3] constructs two non-parallel hyperplanes such that each is positioned closer to one class and farther from the other. Unlike the standard SVM, which involves solving a single large Quadratic Programming Problem or QPP, TWSVM formulates and solves two smaller-sized QPPs independently. Given a set of training samples from two classes, we partition the data x into positive and negative classes. Accordingly, we denote the positive class samples by $A \in \mathbb{R}^{m_A\times d}$ and the negative class samples by $B \in \mathbb{R}^{m_B\times d}$, where e represents a column vector of ones with appropriate dimension [2]. The primal problem of TWSVM for the positive class is given by:

$$\begin{aligned} \min_{w_A,b_A,\xi} \quad & \frac{1}{2}\|Aw_A + e_Ab_A\|^2 + C_1e_B^\top\xi \\ \text{s.t.} \quad & -(Bw_A + e_Bb_A) + \xi \geq e_B, \quad \xi \geq 0. \end{aligned} \tag{3}$$

Similarly for the negative class, it is given as:

$$\begin{aligned} \min_{w_B,b_B,\xi} \quad & \frac{1}{2}\|Bw_B + e_Bb_B\|^2 + C_2e_A^\top\xi \\ \text{s.t.} \quad & (Aw_B + e_Ab_B) + \xi \geq e_A, \quad \xi \geq 0. \end{aligned} \tag{4}$$

To classify a new sample u, we have to compute the function as:

$$Class(u) = \arg\min_{i=A,B} \frac{|w_i^\top u_i + b_i|}{\|w_i\|}. \tag{5}$$

A.3. Support Vector Machine with Two View Learning (SVM-2K)

In the SVM-2K [4] model, we work with data that has two different views or perspectives. For example, two sets of features describing the same object. Each view is represented as a matrix: $x^{(1)}$ for view A and $x^{(2)}$ for view B, where each row contains the feature values for one training example. We have a total of m training examples, and their class labels are stored in a label vector y. The model learns a set of weights $w^{(1)}$, $w^{(2)}$ and bias terms $b^{(1)}$, $b^{(2)}$ for each view.

To handle situations where data points aren't perfectly separable, the model uses slack variables $\xi^{(1)}$, $\xi^{(2)}$, and η. These allow for some error in classification and disagreement between the two views. The term ϵ sets a tolerance level for how much the predictions from the two views are allowed to differ. $e \in \mathbb{R}^{m\times 1}$ be the vector of ones. Lastly, the constants C_1, C_2, and D help balance the trade-offs between accuracy, model complexity, and how closely the two views should agree during training. Let the primal form for a given data is given as:

$$\begin{aligned} \min_{\substack{w^{(1)},w^{(2)},b^{(1)},b^{(2)},\\ \xi^{(1)},\xi^{(2)},\eta}} \quad & \frac{1}{2}\|w^{(1)}\|^2 + \frac{1}{2}\|w^{(2)}\|^2 + C_1e^\top\xi^{(1)} + \\ & C_2e^\top\xi^{(2)} + De^\top\eta \end{aligned} \tag{6}$$

$$
\begin{aligned}
\text{s.t.} \quad & |x^{(1)}w^{(1)} + b^{(1)}e - x^{(2)}w^{(2)} - b^{(2)}e| \leq \eta + \epsilon e, \\
& y * (x^{(1)}w^{(1)} + b^{(1)}e) \geq e - \xi^{(1)}, \\
& y * (x^{(2)}w^{(2)} + b^{(2)}e) \geq e - \xi^{(2)}, \\
& \xi^{(1)}, \xi^{(2)}, \eta \geq 0.
\end{aligned}
$$

The equation for final classifier for SVM-2K becomes for m samples:

$$
\begin{aligned}
f^{(1)}(x) &= \sum_{j=1}^{m} g_i^{(1)} \kappa^{(1)}(x_i, x) + b^{(1)} \\
f^{(2)}(x) &= \sum_{j=1}^{m} g_i^{(2)} \kappa^{(2)}(x_i, x) + b^{(2)} \\
f(x) &= \frac{1}{2}\left(f^{(1)}(x) + f^{(2)}(x)\right) \\
h(x) &= \text{sign}(f(x))
\end{aligned} \tag{7}
$$

A.4. Multi-view Twin Support Vector Machine (MvTWSVM)

Multi-view Twin SVM [5, 6] extends the traditional Twin SVM by incorporating multi-view learning, which improves the model's ability to analyze data and effectively manage heterogeneous information from multiple sources [7]. Given the source input data x, let $x^{(1)}, x^{(2)}$ be the dataset of two views i.e. view 1 and view 2 and $A^{(1)}, B^{(1)}$ be the positive/negative samples from view 1 , $A^{(2)}, B^{(2)}$ be positive/negative samples from view 2, $A^{(1)} = [A^{(1)} \ e_A]$, $B^{(1)} = [B^{(1)} \ e_B]$, $A^{(2)} = [A^{(2)} \ e_A]$, $B^{(2)} = [B^{(2)} \ e_B]$, $w_A^{(1)}, w_A^{(2)}$ are classifiers for positive class, $w_B^{(1)}, w_B^{(2)}$ are classifiers for negative class, Slack variables ξ_A, ξ_B, η for positive class; $\zeta^{(1)}, \zeta^{(2)}, \delta$ for negative class and Regularization constants are C_1, C_2, D for positive and D_1, D_2, H for negative class respectively.

The primal problem of MvTWSVM for the first hyperplane is given by:

$$
\begin{aligned}
\min_{w_A^{(1)}, w_A^{(2)}, \xi_A, \xi_B, \eta} \quad & \frac{1}{2}\|A^{(1)}w_A^{(1)}\|^2 + \frac{1}{2}\|A^{(2)}w_A^{(2)}\|^2 + \\
& C_1 e_B^\top \xi_A + C_2 e_B^\top \xi_B + D e_A^\top \eta \\
\text{s.t.} \quad & |A^{(1)}w_A^{(1)} - A^{(2)}w_A^{(2)}| \leq \eta, \\
& -B^{(1)}w_A^{(1)} + \xi_A \leq e_B, \\
& -B^{(2)}w_A^{(2)} + \xi_B \leq e_B, \\
& \xi_A, \xi_B, \eta \geq 0.
\end{aligned} \tag{8}
$$

Similarly for the second hyperplane it is given by:

$$
\begin{aligned}
\min_{w_B^{(1)}, w_B^{(2)}, \zeta^{(1)}, \zeta^{(2)}, \delta} \quad & \frac{1}{2}\|B^{(1)}w_B^{(1)}\|^2 + \frac{1}{2}\|B^{(2)}w_B^{(2)}\|^2 + \\
& D_1 e_A^\top \zeta^{(1)} + D_2 e_A^\top \zeta^{(2)} + H e_B^\top \delta \\
\text{s.t.} \quad & |B^{(1)}w_B^{(1)} - B^{(2)}w_B^{(2)}| \leq \delta, \\
& -A^{(1)}w_B^{(2)} + \zeta^{(1)} \leq e_A, \\
& -A^{(2)}w_B^{(2)} + \zeta^{(2)} \leq e_A, \\
& \zeta^{(1)}, \zeta^{(2)}, \delta \geq 0.
\end{aligned} \tag{9}
$$

Let test point be $u^{(1)}$ and $u^{(2)}$, consider $u^{(1)} = [u^{(1)}, e]$ and $u^{(2)} = [u^{(2)}, e]$, then $d_1 = \frac{1}{2}\left(|u^{(1)\top}w_A^{(1)}| + |u^{(2)\top}w_A^{(2)}|\right)$ and $d_2 = \frac{1}{2}\left(|u^{(1)\top}w_A^{(1)}| + |u^{(2)\top}w_B^{(2)}|\right)$. Then equation for separator hyperplane is stated as:

$$
\text{Class}(u) = \begin{cases} +1, & \text{if } d_1 < d_2 \\ -1, & \text{otherwise} \end{cases} \tag{10}
$$

A.5. Random Vector Functional Link (RVFL)

RVFL [8] is a single-layer feedforward neural network that enhances input features using randomized hidden nodes and directly connects the input to the output for better generalization.

Given a dataset with m training samples $[x \quad Y]$ or $\{(x_i, y_i)\}_{j=1}^m$, where $x_i \in \mathbb{R}^d$, $y_i \in \{-1, +1\}$, the RVFL network output is computed as $\theta = [\theta_1, \theta_2, \ldots, \theta_m]^T \in \mathbb{R}^m$ be the output weights, $Z = [x \quad Z_x] \in \mathbb{R}^{m \times 2d}$ be output from enhancement layer and original features. Here $Z_x = w^\top * x + b$ for input weights $w \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^{d \times 1}$ for the input feature x.

The regularized least squares objective is:

$$
\min_{\theta} \quad \|Z\theta - Y\|^2 + \lambda\|\theta\|^2. \tag{11}
$$

where $Y = [y^{(1)}, y^{(2)}, \ldots, y_m]^T$ and $\lambda > 0$ is a regularization parameter.

A.6. Twin Random Vector Functional Link (TRVFL)

TRVFL [9, 10] extends the RVFL model by training two independent RVFLs to construct two non-parallel hyperplanes, each closer to one class and farther from the other, similar to the Twin SVM formulation.

Let, A and B are positive and negative class samples respectively from the input data x. $Z_A \in \mathbb{R}^{m_A \times d}$ be the augmented feature matrix of hidden-layer output and input for positive class, $Z_B \in \mathbb{R}^{m_B \times d}$ be the augmented feature matrix for negative class, $\theta_A, \theta_B \in \mathbb{R}^d$ be output weight vectors for two hyperplanes, $\xi_A \in \mathbb{R}^{m_A}, \xi_B \in \mathbb{R}^{m_B}$ are slack variables, e_A, e_B be vectors of ones with appropriate dimension and $C, D > 0$ be the regularization parameters.

For a given sample primal optimization problems are

stated as:

$$\min_{\theta_A,\xi_B} \quad \frac{1}{2}\|Z_A\theta_A\|^2 + Ce_B^\top\xi_B \qquad (12)$$
$$\text{s.t.} \quad -Z_B\theta_A \geq e_B - \xi_B.$$

Similarly for the negative class as:

$$\min_{\theta_B,\xi_A} \quad \frac{1}{2}\|Z_B\theta_B\|^2 + De_A^\top\xi_A \qquad (13)$$
$$\text{s.t.} \quad Z_A\theta_B \geq e_A - \xi_A.$$

Now, the final weight vectors θ_A and θ_B are stated as $\theta_A = -(Z_A^\top Z_A + \epsilon I)^{-1}Z_B^\top\lambda_A$ and $\theta_B = (Z_B^\top Z_B + \epsilon I)^{-1}Z_A^\top\lambda_B$, where $\epsilon > 0$ is a small ridge term for stability. Given a test sample u with final augmented feature vector Z_u, assign the label:

$$Class(u) = \arg\min_{j=A,B} \frac{|\theta_j^\top Z_u|}{\|\theta_j\|}. \qquad (14)$$

A.7. Multi View Twin Random Vector Functional Link (MvTRVFL)

The MvTRVFL model improves classification by mapping multi-view inputs to a high-dimensional space using random features, enabling better class separation. It constructs separate hyperplanes for each class and integrates original and transformed features to capture view-specific and inter-view information, resulting in robust and generalizable performance across diverse datasets.

Let $x^{(j)}$ be the input data for view j for a given sample x and $A^{(j)}, B^{(j)}$ are positive and negative samples respectively from input data $x^{(j)}$. Suppose $Z_A^{(j)} \in \mathbb{R}^{m_A \times 2d}$ be augmented matrix $Z_A^{(j)} = [x^{(j)\prime} \quad Z_A^{(j)\prime}]$ where $Z_A^{(j)\prime} = [(w_A^{(j)\top} * A^{(j)} + b_A^{(j)}) \quad A^{(j)}]$ for input weights $w_A^{(j)}$ and input bias $b_A^{(j)}$ for positive class of view j. Also $x^{(j)\prime} = [(w^{(j)\top} * x^{(j)} + b^{(j)}) \quad x^{(j)}]$ for input weights $w^{(j)}$ and input bias $b^{(j)}$. The optimization problem for the positive class is formulated as:

$$\begin{aligned}\min_{\theta_A^{(j)},\xi^{(j)},\alpha^{(j)(k)}} \quad & \frac{1}{2}\sum_{j=1}^{v}\|Z_A^{(j)}\theta_A^{(j)}\|^2 + C_1\sum_{j=1}^{v}e_B^\top\xi^{(j)} + \\ & C_2\sum_{j=1}^{v}\sum_{j<k}^{v}e_A^\top\alpha^{(j)(k)} \qquad (15)\\ \text{s. t.} \quad & |Z_A^{(j)}\theta_A^{(j)} - Z_A^{(k)}\theta_A^{(k)}| \leq \alpha^{(j)(k)},\\ & -Z_B^{(j)}\theta_A^{(j)} \geq e_B - \xi^{(j)},\\ & \text{where, } \xi^{(j)} \geq 0, \quad \alpha^{(j)(k)} \geq 0\end{aligned}$$

Similarly for the negative class, the primal optimization problem is given as:

$$\begin{aligned}\min_{\theta_B,\chi^{(j)},\gamma^{(j)(k)}} \quad & \frac{1}{2}\sum_{j=1}^{v}\|Z_B^{(j)}\theta_B^{(j)}\|^2 + D_1\sum_{j=1}^{v}e_A^\top\chi^{(j)} + \\ & D_2\sum_{j=1}^{v}\sum_{j<k}^{v}e_B^\top\gamma^{(j)(k)} \qquad (16)\\ \text{s. t.} \quad & |Z_B^{(j)}\theta_B^{(j)} - Z_B^{(k)}\theta_B^{(j)}| \leq \gamma^{(j)(k)},\\ & -Z_A^{(j)}\theta_B^{(j)} \geq e_A - \chi^{(j)},\\ & \text{where, } \chi^{(j)} \geq 0, \quad \gamma^{(j)(k)} \geq 0.\end{aligned}$$

The objective function for separating hyperplane for an example data $u^{(j)}$, input weights $w^{(j)}$ and input bias $b^{(j)}$ of view j becomes:

$$Z_u^{(j)} = [u^{(j)} \quad Z_u^{(j)\prime}], \text{ where } Z_u^{(j)\prime} = w^{(j)\top}u^{(j)} + b^{(j)} \qquad (17)$$

$$\text{Class}(u) = \begin{cases} +1, & \text{if } \left|\sum_{j=1}^{v}\theta_A^{(j)}Z_u^{(j)}\right| \leq \left|\sum_{j=1}^{v}\theta_B^{(j)}Z_u^{(j)}\right| \\ -1, & \text{otherwise} \end{cases} \qquad (18)$$

B PMvTRVFL Derivations

B.1. Linear PMvTRVFL

The extended MvTRVFL model with pinball loss is formulated as:

$$\begin{aligned}\min_{\theta_A^{(j)},\xi_A^{(j)},\xi_B^{(j)},\alpha^{(j)(k)}} \quad & \frac{1}{2}\sum_{j=1}^{v}\|Z_A^{(j)}\theta_A^{(j)}\|^2 \\ & + C_1\sum_{j=1}^{v}e_B^\top(\xi_A^{(j)} + \xi_B^{(j)}) \\ & + C_2\sum_{j=1}^{v}\sum_{j<k}^{v}e_A^\top\alpha^{(j)(k)} \qquad (19)\\ \text{s.t.} \quad & |Z_A^{(j)}\theta_A^{(j)} - Z_A^{(k)}\theta_A^{(k)}| \leq \alpha^{(j)(k)},\\ & -Z_B^{(j)}\theta_A^{(j)} \geq e_B - \xi_A^{(j)},\\ & -Z_B^{(j)}\theta_A^{(j)} \leq e_B + \frac{1}{\tau}\xi_B^{(j)},\\ \textit{where,} \quad & \xi_A^{(j)} \geq 0, \quad \xi_B^{(j)} \geq 0, \quad \alpha^{(j)(k)} \geq 0, \quad \tau \geq 0.\end{aligned}$$

To derive the dual, we introduce Lagrange multipliers: $\beta_1^{(j)(k)}, \beta_2^{(j)(k)} \in \mathbb{R}^{m_A}$ for equality constraints, $\eta_1^{(j)}, \eta_2^{(j)} \in \mathbb{R}^{m_B}$ for inequality constraints related to $\xi_A^{(j)}$ and $\xi_B^{(j)}$, $\lambda_1^{(j)}, \lambda_2^{(j)}$ for non-negativity of slack variables, and $\sigma^{(j)(k)}$ for non-negativity of $\alpha^{(j)(k)}$. The Lagrangian is then formulated accordingly for the positive class as:

$$
\begin{aligned}
L = {} & \frac{1}{2}\sum_{j=1}^{v}\|Z_A^{(j)}\theta_A^{(j)}\|^2 + C_1\sum_{j=1}^{v} e_B^\top\left(\xi_A^{(j)}+\xi_B^{(j)}\right) \\
& + C_2\sum_{j=1}^{v}\sum_{j<k}^{v} e_A^\top\alpha^{(j)(k)} \\
& - \sum_{j=1}^{v}\sum_{j<k}^{v}\beta_1^{(j)(k)\top}\left(\alpha^{(j)(k)} - Z_A^{(j)}\theta_A^{(j)} + Z_A^{(k)}\theta_A^{(k)}\right) \\
& - \sum_{j=1}^{v}\sum_{j<k}^{v}\beta_2^{(j)(k)\top}\left(\alpha^{(j)(k)} + Z_A^{(j)}\theta_A^{(j)} - Z_A^{(k)}\theta_A^{(k)}\right) \\
& - \sum_{j=1}^{v}\eta_1^{(j)\top}\left(-Z_B^{(j)}\theta_A^{(j)} - e_B + \xi_A^{(j)}\right) \\
& - \sum_{j=1}^{v}\eta_2^{(j)\top}\left(Z_B^{(j)}\theta_A^{(j)} + e_B + \frac{1}{\tau}\xi_B^{(j)}\right) \\
& - \sum_{j=1}^{v}\lambda_1^{(j)\top}\xi_A^{(j)} - \sum_{j=1}^{v}\lambda_2^{(j)\top}\xi_B^{(j)} - \sum_{j=1}^{v}\sum_{j<k}^{v}\sigma^{(j)(k)}\alpha^{(j)(k)}.
\end{aligned} \tag{20}
$$

Now, to remove Lagrangian multipliers we have to apply KKT condition so we need to obtain partial derivatives of the above created Lagrangian form.

$$
\begin{aligned}
\frac{\partial L}{\partial\theta_A^{(j)}} = {} & Z_A^{(j)\top}Z_A^{(j)}\theta_A^{(j)} + \sum_{j<k}^{v} Z_A^{(j)\top}\beta_1^{(j)(k)} - \sum_{j<k}^{v} Z_A^{(j)\top}\beta_1^{(j)(k)} \\
& + Z_B^{(j)\top}\eta_1^{(j)} - Z_B^{(j)\top}\eta_2^{(j)} = 0
\end{aligned} \tag{21}
$$

$$
\frac{\partial L}{\partial\theta_A^{(k)}} = \sum_{j<k}^{v} -Z_A^{(k)\top}\beta_1^{(j)(k)} + \sum_{j<k}^{v} Z_A^{(k)\top}\beta_2^{(j)(k)} = 0 \tag{22}
$$

$$
\frac{\partial L}{\partial\xi_A^{(j)}} = C_1 e_B - \eta_1^{(j)} - \lambda_1^{(j)} = 0 \tag{23}
$$

$$
\frac{\partial L}{\partial\xi_B^{(j)}} = C_1 e_B - \frac{1}{\tau}\eta_2^{(j)} - \lambda_2^{(j)} = 0 \tag{24}
$$

$$
\frac{\partial L}{\partial\alpha^{(j)(k)}} = C_2 e_A - \beta_1^{(j)(k)} - \beta_2^{(j)(k)} - \sigma^{(j)(k)} = 0 \tag{25}
$$

Solving for $\theta_A^{(j)}$ for view j to obtain the dual formulation:

$$
\begin{aligned}
\theta_A^{(j)} = {} & (Z_A^{(j)\top}Z_A^{(j)})^{-1}\left(\sum_{j<k}^{v} Z_A^{(j)\top}\left(-\beta_1^{(j)(k)} + \beta_2^{(j)(k)}\right)\right) \\
& - (Z_A^{(j)\top}Z_A^{(j)})^{-1}\left(Z_B^{(j)\top}\left(\eta_1^{(j)} - \eta_2^{(j)}\right)\right)
\end{aligned} \tag{26}
$$

Although the formula seems complex. So, to simplify the things, we give following definitions for any view j:

$$
A^{(j)} = \begin{pmatrix} A_{1,1} & \cdots & A_{1,d} \\ \vdots & \ddots & \vdots \\ A_{m_A,1} & \cdots & A_{m_A,d} \end{pmatrix},
$$

$$
B^{(j)} = \begin{pmatrix} B_{1,1} & \cdots & B_{1,d} \\ \vdots & \ddots & \vdots \\ B_{m_B,1} & \cdots & B_{m_B,d} \end{pmatrix},
$$

$$
A^{(j)} = \begin{bmatrix} A^{(j)} & (w_A^\top * A^{(j)} + b_A^{(j)}) \end{bmatrix},
$$

$$
B^{(j)} = \begin{bmatrix} B^{(j)} & (w_B^\top * B^{(j)} + b_B^{(j)}) \end{bmatrix},
$$

$$
x^{(j)} = \begin{bmatrix} x^{(j)} & (w^\top * x^{(j)} + b^{(j)}) \end{bmatrix},
$$

$$
Z_A^{(j)} = A^{(j)} * x^{(j)\top} \tag{27}
$$

$$
Z_B^{(j)} = B^{(j)} * x^{(j)\top} \tag{28}
$$

$$
\theta_A^{(j)} = \begin{pmatrix} \theta_1^+ \\ \vdots \\ \theta_{m_A}^+ \end{pmatrix}, \qquad \theta_B^{(j)} = \begin{pmatrix} \theta_1^- \\ \vdots \\ \theta_{m_B}^- \end{pmatrix}
$$

Let us define $\zeta^{(j)} = \left(\sum_{j<k}^{v} Z_A^{(j)\top}\left(-\beta_1^{(j)(k)} + \beta_2^{(j)(k)}\right)\right) - \left(Z_B^{(j)\top}\left(\eta_1^{(j)} - \eta_2^{(j)}\right)\right)$. Then after applying KKT conditions we get the dual of the primal form of positive class as:

$$
\begin{aligned}
\min_{\zeta^{(j)},\eta_1^{(j)},\eta_2^{(j)}} \quad & \frac{1}{2}\sum_{j=1}^{v}\left(\zeta^{(j)\top}\left(Z_A^{(j)\top}Z_A^{(j)}\right)^{-1}\zeta^{(j)}\right) \\
& - \frac{1}{2}\sum_{j=1}^{v}\left(\eta_1^{(j)} - \eta_2^{(j)}\right)^\top e_B \\
\text{s.t.} \quad & 0 \le \eta_1^{(j)}, \eta_2^{(j)} \le C_1 e_B, \\
& -C_2\tau e_A \le \beta_1^{(j)(k)}, \beta_2^{(j)(k)} \le C_2 e_A.
\end{aligned} \tag{29}
$$

In the similar manner we can obtain the dual form for negative class also as:

$$
\begin{aligned}
\theta_B^{(j)} = {} & (Z_B^{(j)\top}Z_B^{(j)})^{-1}\left(\sum_{j<k}^{v} Z_B^{(j)\top}\left(-\kappa_1^{(j)(k)} + \kappa_2^{(j)(k)}\right)\right) \\
& - (Z_B^{(j)\top}Z_B^{(j)})^{-1}\left(Z_A^{(j)\top}\left(\phi_1^{(j)} - \phi_2^{(j)}\right)\right)
\end{aligned} \tag{30}
$$

Let us define $\delta^{(j)} = \left(\sum_{j<k}^{v} Z_B^{(j)\top}\left(-\kappa_1^{(j)(k)} + Z_B^{(j)\top}\kappa_2^{(j)(k)}\right)\right) - \left(Z_A^{(j)\top}\left(\phi_1^{(j)} - \phi_2^{(j)}\right)\right)$. Then in similar manner we can obtain dual of the primal form for negative class as:

$$
\min_{\delta^{(j)},\phi_1^{(j)},\phi_2^{(j)}} \quad \frac{1}{2}\sum_{j=1}^{v}\left(\delta^{(j)\top}\left(Z_B^{(j)\top}Z_B^{(j)}\right)^{-1}\delta^{(j)}\right) \tag{31}
$$

$$
\begin{aligned}
&- \sum_{j=1}^{v} \left(\left(\phi_1^{(j)} - \phi_2^{(j)} \right)^{\top} e_A \right) \\
\text{s.t.} \quad & 0 \leq \kappa_1^{(j)(k)}, \kappa_2^{(j)(k)} \leq D_2 e_B, \\
& -\tau D_1 e_A \leq \phi_1^{(j)}, \phi_2^{(j)} \leq D_1 e_A.
\end{aligned}
$$

where, $\kappa_1^{(j)(k)}, \kappa_2^{(j)(k)}$, $\phi_1^{(j)}, \phi_2^{(j)}$ be Lagrangian multipliers for negative class.

B.2. Kernel-PMvTRVFL

The kernel-PMvTRVFL can be stated as:

$$
\begin{aligned}
\min_{\theta_A^{(j)}, \xi_A^{(j)}, \xi_B^{(j)}, \alpha^{(j)(k)}} \quad & \frac{1}{2} \sum_{j=1}^{v} \left\| K_r \langle A^{(j)\top}, Z_A^{(j)\top} \rangle \theta_A^{(j)} \right\|^2 \\
& + C_1 \sum_{j=1}^{v} e_B^{\top} (\xi_A^{(j)} + \xi_B^{(j)}) \\
& + C_2 \sum_{j=1}^{v} \sum_{j<k}^{v} e_A^{\top} \alpha^{(j)(k)}
\end{aligned}
$$

$$
\begin{aligned}
\text{s.t.} \quad & |K_r \langle A^{(j)\top}, Z_A^{(j)\top} \rangle \theta_A^{(j)} - K_r \langle A^{(k)\top}, Z_A^{(k)\top} \rangle \theta_A^{(k)}| \\
& \leq \alpha^{(j)(k)}, \\
& -K_r \langle B^{(j)\top}, Z_B^{(j)\top} \rangle \theta_A^{(j)} \geq e_B - \xi_A^{(j)}, \\
& -K_r \langle B^{(j)\top}, Z_B^{(j)\top} \rangle \theta_A^{(j)} \leq e_B + \frac{1}{\tau} \xi_B^{(j)}, \\
& where, \xi_A^{(j)} \geq 0, \quad \xi_B^{(j)} \geq 0, \quad \alpha^{(j)(k)} \geq 0, \quad \tau \geq 0.
\end{aligned}
$$

The basic description of Kernel-PMvTRVFL is:

1. The regularization term $\frac{1}{2} \sum_{j=1}^{v} \left\| K_r \langle A^{(j)\top}, Z_A^{(j)\top} \rangle \theta_A^{(j)} \right\|^2$ controls the model complexity by penalizing large weights in the Reproducing Kernel Hilbert Space (RKHS). The slack penalty term $C_1 \sum_{j=1}^{v} e_B^{\top} (\xi_A^{(j)} + \xi_B^{(j)})$ handles asymmetric margin violations for the negative class using pinball loss. The inter-view consistency penalty $C_2 \sum_{j=1}^{v} \sum_{j<k}^{v} e_A^{\top} \alpha^{(j)(k)}$ encourages prediction agreement among different views by penalizing discrepancies.
2. The inter-view agreement constraint $|K_r \langle A^{(j)\top}, Z_A^{(j)\top} \rangle \theta_A^{(j)} - K_r \langle A^{(k)\top}, Z_A^{(k)\top} \rangle \theta_A^{(k)}| \leq \alpha^{(j)(k)}$ ensures that predictions from different views on the same positive samples remain consistent in the kernel-induced feature space.
3. The lower bound constraint $-K_r \langle B^{(j)\top}, Z_B^{(j)\top} \rangle \theta_A^{(j)} \geq e_B - \xi_A^{(j)}$ enforces the minimum required margin for negative class predictions, ensuring robustness against under-prediction.
4. The upper bound constraint $-K_r \langle B^{(j)\top}, Z_B^{(j)\top} \rangle \theta_A^{(j)} \leq e_B + \frac{1}{\tau} \xi_B^{(j)}$ imposes a quantile-based asymmetric margin, regulating over-prediction based on the quantile parameter τ.
5. Non-negativity conditions on all slack and auxiliary variables $\xi_A^{(j)} \geq 0, \xi_B^{(j)} \geq 0, \alpha^{(j)(k)} \geq 0$ and $\tau \geq 0$ ensure the feasibility of the optimization problem and the interpretability of the model components.

The formation of Lagrangian form becomes after applying Lagrangian multipliers for the positive class can be written as:

$$
\begin{aligned}
L = & \frac{1}{2} \sum_{j=1}^{v} \| K_r \langle A^{(j)\top}, Z_A^{(j)\top} \rangle \theta_A^{(j)} \|^2 + C_1 \sum_{j=1}^{v} e_B^{\top} (\xi_A^{(j)} + \xi_B^{(j)}) \\
& + C_2 \sum_{j=1}^{v} \sum_{j<k}^{v} e_A^{\top} \alpha^{(j)(k)} \\
& - \sum_{j=1}^{v} \sum_{j<k}^{v} \beta_1^{(j)(k)\top} (\alpha^{(j)(k)} - K_r \langle A^{(j)\top}, Z_A^{(j)\top} \rangle \theta_A^{(j)} \\
& \qquad + K_r \langle A^{(k)\top}, Z_A^{(k)\top} \rangle \theta_A^{(k)}) \\
& - \sum_{j=1}^{v} \sum_{j<k}^{v} \beta_2^{(j)(k)\top} (\alpha^{(j)(k)} + K_r \langle A^{(j)\top}, Z_A^{(j)\top} \rangle \theta_A^{(j)} \\
& \qquad - K_r (A^{(k)\top}, Z_A^{(k)\top}) \theta_A^{(k)}) \\
& - \sum_{j=1}^{v} \eta_1^{(j)\top} (-K_r \langle B^{(j)\top}, Z_B^{(j)\top} \rangle \theta_A^{(j)} - e_B + \xi_A^{(j)}) \\
& - \sum_{j=1}^{v} \eta_2^{(j)\top} (K_r \langle B^{(j)\top}, Z_B^{(j)\top} \rangle \theta_A^{(j)} + e_B + \frac{1}{\tau} \xi_B^{(j)}) \\
& - \sum_{j=1}^{v} \lambda_1^{(j)\top} \xi_A^{(j)} - \sum_{j=1}^{v} \lambda_2^{(j)\top} \xi_B^{(j)} - \sum_{j=1}^{v} \sum_{j<k}^{v} \sigma^{(j)(k)} \alpha^{(j)(k)}.
\end{aligned}
$$

Now, to obtain dual of the problem we have to obtain partial derivatives of the above created Lagrangian form.

$$
\begin{aligned}
\frac{\partial L}{\partial \theta_A^{(j)}} = & K_r \langle A^{(j)\top}, Z_A^{(j)\top} \rangle^{\top} K_r \langle A^{(j)\top}, Z_A^{(j)\top} \rangle \theta_A^{(j)} \\
& + \sum_{j<k}^{v} K_r \langle A^{(j)\top}, Z_A^{(j)\top} \rangle^{\top} \beta_1^{(j)(k)} \\
& - \sum_{j<k}^{v} K_r \langle A^{(j)\top}, Z_A^{(j)\top} \rangle^{\top} \beta_2^{(j)(k)} \\
& + K_r \langle B^{(j)\top}, Z_B^{(j)\top} \rangle^{\top} \eta_1^{(j)} \\
& - K_r \langle B^{(j)\top}, Z_B^{(j)\top} \rangle^{\top} \eta_2^{(j)} = 0 \\
\frac{\partial L}{\partial \theta_A^{(k)}} = & - \sum_{j<k}^{v} K_r \langle A^{(k)\top}, Z_k^{\top} \rangle^{\top} \beta_1^{(j)(k)} \\
& + \sum_{j<k}^{v} K_r \langle A^{(k)\top}, Z_k^{\top} \rangle^{\top} \beta_2^{(j)(k)} = 0
\end{aligned}
$$

$$\frac{\partial L}{\partial \xi_A^{(j)}} = C_1 e_B - \eta_1^{(j)} - \lambda_1^{(j)} = 0$$
$$\frac{\partial L}{\partial \xi_B^{(j)}} = C_1 e_B - \frac{1}{\tau}\eta_2^{(j)} - \lambda_2^{(j)} = 0$$
$$\frac{\partial L}{\partial \alpha^{(j)(k)}} = C_2 e_A - \beta_1^{(j)(k)} - \beta_2^{(j)(k)} - \sigma^{(j)(k)} = 0$$

$$\begin{aligned}\frac{\partial L}{\partial \theta_A^{(j)}} &= K_r\langle A^{(j)\top}, Z_A^{(j)\top}\rangle^\top K_r\langle A^{(j)\top}, Z_A^{(j)\top}\rangle\theta_A^{(j)} \\ &+ \sum_{j<k}^{v} K_r\langle A^{(j)\top}, Z_A^{(j)\top}\rangle^\top \beta_1^{(j)(k)} \\ &- \sum_{j<k}^{v} K_r\langle A^{(j)\top}, Z_A^{(j)\top}\rangle^\top \beta_2^{(j)(k)} \\ &+ K_r\langle B^{(j)\top}, Z_B^{(j)\top}\rangle^\top \eta_1^{(j)} \\ &- K_r\langle B^{(j)\top}, Z_B^{(j)\top}\rangle^\top \eta_2^{(j)} = 0\end{aligned} \tag{10}$$

Now, define the kernel matrices as: $\Theta_A^{(j)} = \left[K_r\langle A^{(j)\top}, Z_A^{(j)\top}\rangle, e_A\right]$ and $\Theta_B^{(j)} = \left[K_r\langle B^{(j)\top}, Z_B^{(j)\top}\rangle, e_B\right]$.

Solving for $\theta_A^{(j)}$ for view j to obtain the dual formulation:

$$\begin{aligned}\theta_A^{(j)} =&(\Theta_A^{(j)\top}\Theta_A^{(j)})^{-1}\left(\sum_{j<k}^{v}\Theta_A^{(j)\top}\left(-\beta_1^{(j)(k)} + \beta_2^{(j)(k)}\right)\right) \\ &- (\Theta_A^{(j)\top}\Theta_A^{(j)})^{-1}\left(\Theta_B^{(j)\top}\left(\eta_1^{(j)} - \eta_2^{(j)}\right)\right).\end{aligned}$$

To simplify the things, we give following definitions for any view j: Let's say Kernel for a given data is defined as $Kr\langle t\rangle = \psi(t)$ for any data t.

$$Kr\langle A^{(j)}\rangle = \left[\psi(w_A^\top * A^{(j)} + b_A^{(j)}) \quad A^{(j)}\right],$$
$$Kr\langle B^{(j)}\rangle = \left[\psi(w_B^\top * B^{(j)} + b_B^{(j)}) \quad B^{(j)}\right],$$
$$Kr\langle x^{(j)}\rangle = \left[\psi(w^\top * x^{(j)} + b^{(j)}) \quad x^{(j)\top}\right],$$
$$Kr\langle A^{(j)\top}, Z_A^{(j)\top}\rangle = Kr\langle A^{(j)}\rangle * Kr\langle x^{(j)}\rangle^\top \tag{9}$$
$$Kr\langle B^{(j)\top}, Z_B^{(j)\top}\rangle = Kr\langle B^{(j)}\rangle * Kr\langle x^{(j)}\rangle^\top \tag{9}$$

$$\theta_A^{(j)} = \begin{pmatrix}\theta_1^+ \\ \vdots \\ \theta_{m_A}^+\end{pmatrix}, \qquad \theta_B^{(j)} = \begin{pmatrix}\theta_1^- \\ \vdots \\ \theta_{m_B}^-\end{pmatrix}$$

Let us define $\zeta^{(j)}$ as $\zeta^{(j)} = (\Theta_A^{(j)\top}\Theta_A^{(j)})^{-1}(-\Theta_A^{(j)\top}\beta_1^{(j)(k)} + \Theta_A^{(j)\top}\beta_2^{(j)(k)} - \Theta_B^{(j)\top}\eta_1^{(j)} + \Theta_B^{(j)\top}\eta_2^{(j)})$. Then after applying KKT conditions we get the dual form of the primal form of positive class as:

$$\begin{aligned}\min_{\zeta^{(j)},\eta_1^{(j)},\eta_2^{(j)}} \quad & \frac{1}{2}\sum_{j=1}^{v}\left(\zeta^{(j)\top}\left(\Theta_A^{(j)\top}\Theta_A^{(j)}\right)^{-1}\zeta^{(j)}\right) \\ & - \sum_{j=1}^{v}\left(\eta_1^{(j)} - \eta_2^{(j)}\right)^\top e_B \\ \text{s.t.} \quad & 0 \le \eta_1^{(j)}, \eta_2^{(j)} \le C_1 e_B, \\ & -C_2\tau e_A \le \beta_1^{(j)(k)}, \beta_2^{(j)(k)} \le C_2 e_A.\end{aligned}$$

In the similar manner for negative class we obtain $\theta_B^{(j)}$ for view j as:

$$\begin{aligned}\theta_B^{(j)} =&(\Theta_B^{(j)\top}\Theta_B^{(j)})^{-1}\left(\Theta_B^{(j)\top}\left(-\kappa_1^{(j)(k)} + \kappa_2^{(j)(k)}\right)\right) \\ &- (\Theta_B^{(j)\top}\Theta_B^{(j)})^{-1}\left(\Theta_A^{(j)\top}\left(\eta_1^{(j)} - \eta_2^{(j)}\right)\right).\end{aligned}$$

Let us define $\delta^{(j)} = \left(\Theta_B^{(j)\top}\left(-\kappa_1^{(j)(k)} + \kappa_2^{(j)(k)}\right)\right) - \left(\Theta_A^{(j)\top}\left(\eta_1^{(j)} - \eta_2^{(j)}\right)\right)$. Then in similar manner we can obtain dual of the primal form for negative class as:

$$\begin{aligned}\min_{\delta^{(j)},\phi_1^{(j)},\phi_2^{(j)}} \quad & \frac{1}{2}\sum_{j=1}^{v}\left(\delta^{(j)^\top}\left(\Theta_B^{(j)^\top}\Theta_B^{(j)}\right)^{-1}\delta^{(j)}\right) \\ & - \sum_{j=1}^{v}\left(\left(\phi_1^{(j)} - \phi_2^{(j)}\right)^\top e_A\right) \\ \text{s.t.} \quad & 0 \le \kappa_1^{(j)(k)}, \kappa_2^{(j)(k)} \le D_2 e_B, \\ & -\tau D_1 e_A \le \phi_1^{(j)}, \phi_2^{(j)} \le D_1 e_A.\end{aligned}$$

where, $\kappa_1^{(j)(k)}, \kappa_2^{(j)(k)}$ and $\phi_1^{(j)}, \phi_2^{(j)}$ are Lagrangian multipliers for negative class.

B.3. Dataset Description

Table I presents the datasets employed in this study, spanning domains such as handwritten digit recognition, bioinformatics, and others.Each dataset provides unique challenges in terms of class distribution, dimensionality, and modality, making them ideal for evaluating the robustness of multi-view classifiers. The inclusion of diverse datasets ensures a comprehensive assessment of the proposed model's generalization ability across various real-world applications.

1. Handwritten Digits

This dataset contains features extracted from handwritten digits (0-9), sourced from Dutch utility maps. Researchers digitized 200 samples for each digit class, resulting in a total of 2,000 binary images [11]. Each feature set captures different aspects of the digit shapes, such as pixel intensity, frequency domain components, and morphological characteristics. This diversity makes the dataset well-suited for evaluating multi-view learning approaches like MvTRVFL, which benefit from heterogeneous feature representations. These images are described using six dis-

tinct feature sets:

- **mfeat-fou:** Contains 76 Fourier coefficients representing character shapes.
- **mfeat-fac:** Includes 216 profile correlation values.
- **mfeat-kar:** Provides 64 Karhunen–Loève coefficients.
- **mfeat-pix:** Captures 240 average pixel values computed over 2×3 windows.
- **mfeat-zer:** Contains 47 Zernike moment features.
- **mfeat-mor:** Offers 6 features based on morphological analysis.

Each file stores all 2,000 patterns in ASCII format, where each line corresponds to one pattern. The data is organized sequentially by class, starting with 200 patterns for digit '0', followed by 200 each for digits '1' through '9'. We have taken 200 samples from a total of (1,7) pairs from 10 classes for training in first view and applied PCA for the second view data.

2. UCI Repository Datasets

The UCI Machine Learning Repository is one of the most widely recognized and reputable sources for benchmark datasets within the machine learning research community. Maintained by the University of California, Irvine, it provides a diverse collection of datasets suitable for tasks such as classification, regression, and clustering.

These datasets serve as standardized benchmarks for evaluating and comparing the performance of various machine learning algorithms. Owing to its accessibility and broad coverage of real-world applications, the repository plays a foundational role in empirical research and the development of robust machine learning methodologies. These datasets span multiple domains including biology, medicine, finance, image recognition, and text processing. Each dataset typically comes with structured features, target labels, and metadata including attribute descriptions and source references, making them ideal for model development, evaluation, and comparison. We have considered ecoli [12], sonar, wine [13], led7digit-0-2-4-5-6-7-8-9 vs_1, pima, ecoli-0-2-6-7vs3-5,ecoli-0-4-6vs_5, new-thyroid1, iono [14], wdbc, and votes datsets for our experiment.

Table I: Dataset Descriptions and View-wise Features

Dataset	Domain	Instances	Features per View
mfeat-fou	Handwritten Digits	2000 (200 per class)	76 (Fourier)
mfeat-mor	Handwritten Digits	2000	6 (Morphological)
mfeat-fac	Handwritten Digits	2000	216 (Profile correlations)
mfeat-zer	Handwritten Digits	2000	47 (Zernike moments)
mfeat-kar	Handwritten Digits	2000	64 (Karhunen-Loève)
mfeat-pix	Handwritten Digits	2000	240 (Pixel averages)
ecoli	Bioinformatics	336	7 (Single view)
sonar	Signal Processing	208	60 (Single view)
wine	Chemistry / Agriculture	178	13 (Single view)
led7digit-0-2-4-5-6-7-8-9_vs_1	Image Analysis	443	7 (after view split)
pima	Medical (Diabetes)	768	8 (Single view)
ecoli-0-2-6-7_vs_3-5	Bioinformatics	224	7–8 (after view split)
ecoli-0-4-6_vs_5	Bioinformatics	205	7–8 (after view split)
new-thyroid1	Medical (Thyroid Disease)	215	5 (after view split)
iono	Signal Processing	351	34 (Single view)
wdbc	Medical (Breast Cancer)	569	30 (Single view)
votes	Political / Social	435	16 (Single view)

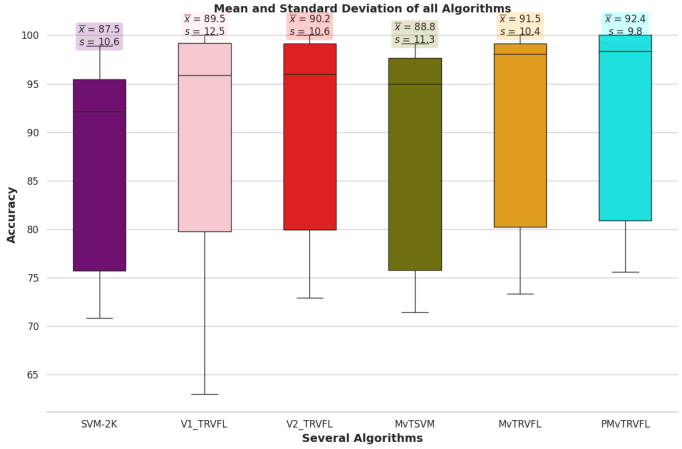

Figure I: Box plot representation of classifiers used

Figure I shows a box plot that compares how well each classifier performed in terms of accuracy and consistency. A box plot—also known as a box-and-whisker plot—is a simple yet powerful way to visualize how results are spread out. It helps us quickly see which models performed more steadily and which ones varied more. In each box, the line in the middle shows the median accuracy, giving a sense of the typical performance, while the top and bottom edges of the box mark the range where the middle 50% of the results fall, known as the interquartile range (IQR). This makes it easy to compare how tightly grouped or spread out the classifiers' performances were. The whiskers extend to show the range of the data, excluding outliers, and help in assessing the variability and consistency of each classifier's performance. A smaller box with shorter whiskers typically indicates higher stability and less variation in the model's predictions.

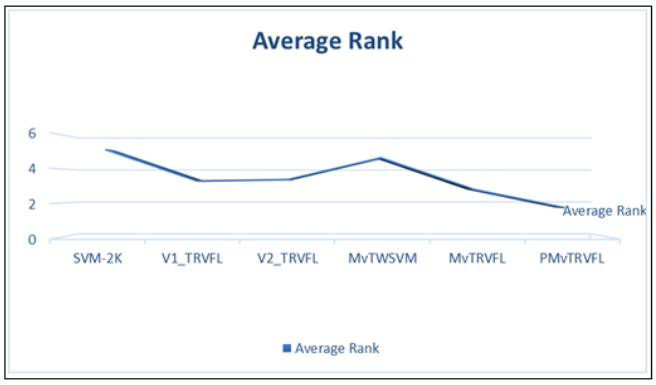

Figure II: Comparison of average ranks between the proposed model and existing models.

A rank-based comparison of six classifiers evaluated across 17 benchmark datasets is presented in Table **??**, where a lower rank indicates superior performance. The proposed PMvTRVFL model consistently secures top rankings, achieving the lowest overall average rank of 1.73 and delivering the best performance on the majority of datasets. This demonstrates the model's robustness, stability, and excellent generalization capability across diverse domains such as signal processing, image analysis, and medical diagnostics. In contrast, traditional methods like SVM-2K and MvTWSVM achieve significantly higher average ranks of 4.67 and 5.12, respectively, indicating comparatively weaker performance. These results highlight PMvTRVFL's strength in leveraging multi-view information more effectively than both single-view and earlier multi-view classifiers.

Figure II provides a line graph of the average ranks obtained by each model: SVM-2K 5.12, V1-TRVFL 3.29, V2-TRVFL 3.38, MvTWSVM 4.67, MvTRVFL 2.82, and PMvTRVFL 1.73. The performance gap between PMvTRVFL and other models—3.38, 1.56, 1.65, 2.94, and 1.09, respectively—clearly illustrates its superior and consistent performance. These rank differences underline the model's effectiveness in applications ranging from handwritten digit recognition to complex biomedical classification problems. The consistent top-tier ranking of PMvTRVFL reinforces its capability in exploiting multi-view data to achieve highly reliable, adaptable, and generalized classification outcomes.

Additionally, Table II provides a comprehensive overview of the classifiers and their associated hyperparameters, which directly influence learning behavior and predictive capacity. Key performance metrics such as AUC, Recall, F1-score, and G-mean are included, offering deep insight into the models' ability to handle imbalanced data and maintain accuracy across datasets. These indicators consistently reflect PMvTRVFL's superior performance, validating its reliability, efficiency, and practical utility in varied domains.

What makes PMvTRVFL stand out is its smart use of multiple data views through a twin-network structure, along with the pinball loss function that helps it handle uneven class distributions by adjusting the margin for each class. Its consistently high scores in AUC, Recall, F1-score, and G-mean prove that it not only works well in theory but also delivers stable and trustworthy results in real-world scenarios.

C Statistical Analysis

The Friedman test, originally introduced by Friedman (1937) [15] and further extended by Demšar (2006), is a non-parametric statistical method used to evaluate the performance of multiple classifiers across several datasets. In this study, we apply the Friedman test to assess and compare the ranks of six classifiers, SVM-2K, TRVFL, MvTWSVM, MvTRVFL and PMvTRVFL.

The Friedman test statistic χ^2_{FR} is computed using the following formula:

$$\chi^2_{\text{FR}} = \frac{12M}{n(n+1)} \left[\sum_{i=1}^{n} \bar{r}_i^2 - \frac{n(n+1)^2}{4} \right] \tag{1}$$

where M is the number of datasets, n is the number of classifiers, and $\bar{r}_i$ is the average rank of the i-th classifier.

To analyze the significance of the result, we use the F-distribution approximation given by:

$$F_{\text{FR}} = \frac{(M-1) \cdot \chi^2_{\text{FR}}}{M(n-1) - \chi^2_{\text{FR}}} \tag{2}$$

Substituting the experimental values, we compute:

$$\chi^2_{\text{FR}} = \frac{12 \times 17}{6 \times 7}(5.11^2 + 3.29^2 + 3.38^2 + 4.64^2 + 2.82^2 + 1.73^2). = 36.7127.$$

Table II: Performance of different models based on AUC, Recall, Precision, F1-score and Gmean

Dataset	**SVM-2K** [4]	**V1_TRVFL** [10]	**V2_TRVFL** [10]	**MvTWSVM** [5]	**MvTRVFL**	**PMvTRVFL**
	AUC **recall** **precision** **F1-score** **Gmean** **(C1, C2, D, epsilon, sigma)**	**AUC** **recall** **precision** **F1-score** **Gmean** **(C, L)**	**AUC** **recall** **precision** **F1-score** **Gmean** **(C, L)**	**AUC** **recall** **precision** **F1-score** **Gmean** **(C1, C2, D1, D2, D, epsilon, H)**	**AUC** **recall** **precision** **F1-score** **Gmean** **(C1, C2, D1, D2, alpha, epsilon, L)**	**AUC** **recall** **precision** **F1-score** **Gmean** **(C1, C2, D1, D2, alpha, epsilon, L, tau)**
mfeat-fou	92.8409 87.5 98.2456 92.562 92.6872 (8, 8, 0.0078125, 0.01, 2)	92.6573 90.7692 95.1613 92.9134 92.9393 (0.125, 20)	91.8881 89.2308 95.082 92.0635 92.1099 (32, 500)	93.6222 89.0625 98.2759 93.4426 93.5111 (0.5, 0.5, 0.5, 0.5, 0.5, 0.01, 32)	92.7131 89.0625 96.6102 92.6829 92.6412 (0.5, 0.5, 0.5, 0.5, 0.0078125, 100, 0.01)	88.1676 89.0625 89.0625 89.0625 88.1631 (0.0078125, 0.0078125, 0.0078125, 0.0078125, 0.0078125, 100, 0.01, 1.000000e-01)
mfeat-fac	99.2537 98.5075 100 99.2481 99.2509 (0.125, 0.125, 0.0078125, 0.01, 2)	100 100 100 100 100 (0.03125, 1000)	100 100 100 100 100 (0.125, 1000)	99.2537 98.5075 100 99.2481 99.2509 (0.0078125, 0.0078125, 0.0078125, 0.0078125, 0.0078125, 0.01, 64)	99.2537 98.5075 100 99.2481 99.2509 (2, 2, 2, 2, 0.03125, 100, 0.01)	99.2537 98.5075 100 99.2481 99.2509 (0.0078125, 0.0078125, 0.0078125, 0.0078125, 0.0078125, 100, 0.01, 0.1)
mfeat-kar	**100** 100 100 100 100 (2, 2, 0.0078125, 0.01, 2)	99.0741 98.1481 100 99.0654 99.0697 (0.0078125, 1000)	99.0741 98.1481 100 99.0654 99.0697 (0.0078125, 100)	96.5517 93.1034 100 96.4286 96.4901 (2, 2, 2, 2, 0.125, 0.01, 16)	99.0741 98.1481 100 99.0654 99.0697 (8, 8, 8, 8, 128, 100, 0.01)	**100** 100 100 100 100 (0.000488281, 0.000488281, 0.000488281, 0.000488281, 0.00195312, 500, 0.01, 0.0001)
mfeat-mor	96.5517 93.1034 100 96.4286 96.4901 (8, 8, 0.5, 0.01, 0.125)	98.3329 98.3051 98.3051 98.3051 98.3051 (32, 1000)	97.4576 94.9153 100 97.3913 97.4245 (2, 50)	95.7321 93.1034 98.1818 95.5752 95.696 (2, 2, 2, 2, 0.5, 0.01, 0.125)	96.5517 93.1034 100 96.4286 96.4901 (0.03125, 0.03125, 0.03125, 0.03125, 0.03125, 0.125, 100, 0.001)	97.4138 94.8276 100 97.3451 97.3795 (0.0078125, 0.0078125, 0.0078125, 0.0078125, 0.03125, 100, 0.001, 0.1)
mfeat-pix	**100** 100 100 100 100 0.125 0.125 0.0078125 0.01 8	**100** 100 100 100 100 128 500	**100** 100 100 100 100 128 1000	99.1071 100 98.4375 99.2126 99.1031 0.0078125 0.0078125 0.0078125 0.0078125 0.0078125 0.01 16	98.2143 100 96.9231 98.4375 98.1981 0.03125 0.03125 0.03125 0.03125 0.125 100 0.001	**100** 100 100 100 100 0.0000305176 0.0000305176 0.0000305176 0.0000305176 0.00012207 100 0.001 0.0001

Dataset	SVM-2K [4]	V1_TRVFL [10]	V2_TRVFL [10]	MvTWSVM [5]	MvTRVFL	PMvTRVFL
mfeat-zer	**100** 100 100 100 100 (0.5, 0.5, 0.0078125, 0.01, 0.5)	**100** 100 100 100 100 (0.03125, 1000)	**100** 100 100 100 100 (0.03125, 50)	**100** 100 100 100 100 (0.0078125, 0.0078125, 0.0078125, 0.0078125, 0.0078125, 0.01, 8)	**100** 100 100 100 100 (0.03125, 0.03125, 0.03125, 0.03125, 0.0078125, 0.01, 100)	**100** 100 100 100 100 (0.0078125, 0.03125, 0.0078125, 0.03125, 0.125, 100, 0.01, 0.1)
iono	87.8468 98.5507 89.4737 93.7931 87.1922 (8, 8, 0.0078125, 0.01, 2)	89.2857 92.8571 92.8571 92.8571 92.8571 (2, 500)	88.5714 94.2857 91.6667 92.9577 92.967 (2, 100)	89.9379 91.3043 94.0299 92.6471 89.9275 (0.125, 0.125, 0.125, 0.125, 0.5, 0.01 8)	92.1325 98.5507 93.1507 95.7746 91.9087 (0.5, 0.5, 0.5, 0.5, 0.0078125, 0.01, 100)	91.4079 97.1014 93.0556 95.0355 91.2304 (0.03125, 0.03125, 0.03125, 0.03125, 0.0078125, 100, 0.01 0.01)
ecoli	94.1102 95 90.4762 92.6829 94.106 (0.5, 0.5, 0.0078125, 0.01, 2)	96.25 97.5 92.8571 95.122 95.1503 (0.5, 50)	96.25 97.5 92.8571 95.122 95.1503 (8, 20)	92.8602 92.5 90.2439 91.358 92.8595 (0.03125, 0.03125, 0.03125, 0.03125, 0.0078125, 0.01, 16)	93.7076 92.5 92.5 92.5 93.6998 (0.03125, 0.03125, 0.03125, 0.03125, 0.125, 100, 0.001)	96.2076 97.5 92.8571 95.122 96.1989 (0.0078125, 0.0078125, 0.0078125, 0.0078125, 0.25, 10000, 0.0001, 0.1)
sonar	76.6234 71.4286 76.9231 74.0741 76.4471 (32, 32, 0.0078125, 0.01, 2)	64.7059 50 66.6667 57.1429 57.735 (32, 100)	69.7479 57.1429 72.7273 64 64.4658 (8, 200)	72.2403 53.5714 83.3333 65.2174 69.7863 (0.007813, 0.007813, 0.007813, 0.007813, 0.01, 8)	79.6537 71.4286 83.3333 76.9231 79.2279 (0.5, 0.5, 0.5, 0.5, 0.03125, 100, 0.01)	73.0519 64.2857 75 69.2308 72.5241 (0.0078125, 0.0078125, 0.0078125, 0.0078125, 0.0078125, 500, 0.001, 0.0001)
pima	73.8143 56.962 76.2712 65.2174 71.8649 (128, 128, 0.03125, 0.01, 2)	71.0412 49.3671 78 60.4651 62.0535 (2, 200)	72.2483 54.4304 74.1379 62.7737 63.5244 (0.5, 50)	73.7131 60.7595 70.5882 65.3061 72.566 (0.5, 0.5, 0.5, 0.5, 0.0078125, 0.01, 16)	77.0422 73.4177 66.6667 69.8795 76.9569 (0.03125, 0.03125, 0.03125, 0.03125, 0.03125, 100, 0.01)	**81.8397** 81.0127 71.1111 75.7396 81.8355 (0.000488281, 0.000488281, 0.000488281, 0.000488281, 0.00195312, 500, 0.01, 0.0001)
wine	97.8261 95.6522 100 97.7778 97.8019 (2, 2, 0.0078125, 0.01, 0.5)	97.9167 95.8333 100 97.8723 97.8945 (0.125, 20)	95.8333 91.6667 100 95.6522 95.7427 (128, 50)	96.1019 95.6522 95.6522 95.6522 96.1009 (0.0078125, 0.0078125, 0.0078125, 0.0078125, 0.0078125, 0.01, 8)	95.6522 100 97.7778 97.8019 (0.0078125, 0.0078125, 0.0078125, 0.0078125, 0.125, 0.01, 100)	97.8261 95.6522 100 97.7778 97.8019 (2048, 2048, 2048, 2048, 2048, 100, 0.01, 0.0001)

Dataset	SVM-2K [4]	V1_TRVFL [10]	V2_TRVFL [10]	MvTWSVM [5]	MvTRVFL	PMvTRVFL
ecoli-0-2-6-7_vs_3-5	**100** 100 100 100 100 (2, 2, 0.0078125, 0.01, 0.5)	**100** 100 100 100 100 (2, 100)	**100** 100 100 100 100 (0.5, 200)	99.2188 100 66.6667 80 99.2157 (0.5, 0.5, 0.5, 0.5, 0.0078125, 0.01, 8)	**100** 100 100 100 100 (0.5, 0.5, 0.5, 0.5, 0.0078125, 100, 0.01)	**100** 100 100 100 100 (0.0078125, 0.0078125, 0.0078125, 0.0078125, 0.0078125, 100, 0.01, 0.1)
ecoli-0-4-6_vs_5	**85.7143** 71.4286 100 83.3333 84.5154 (32, 32, 0.0078125, 0.01, 0.5)	78.5714 57.1429 100 72.7273 75.5929 (0.125, 100)	**85.7143** 71.4286 100 83.3333 84.5154 (0.125, 200)	71.4286 42.8571 100 60 65.4654 (0.0078125, 0.0078125, 0.0078125, 0.0078125, 0.0078125, 0.001, 0.125)	**85.7143** 71.4286 100 83.3333 84.5154 (0.5, 0.5, 0.5, 0.5, 0.0078125, 100, 0.01)	**85.7143** 71.4286 100 83.3333 84.5154 (0.0078125, 0.0078125, 0.0078125, 0.0078125, 0.0078125, 100, 0.01, 0.1)
new-thyroid1	99.0741 100 90 94.7368 99.0697 (8, 8, 0.0078125, 0.01, 0.5)	**100** 100 100 100 100 (2, 1000)	**100** 100 100 100 100 (0.5, 1000)	77.7778 55.5556 100 71.4286 74.5356 (0.00195312, 0.00195312, 0.00195312, 0.00195312, 0.125, 0.001, 0.125)	**100** 100 100 100 100 (0.5, 0.5, 0.5, 0.5, 0.0078125, 100, 0.01)	**100** 100 100 100 100 (0.0078125, 0.0078125, 0.0078125, 0.0078125, 0.0078125, 10, 0.01, 0.1)
led7digit-0-2-4-5-6-7-8-9_vs_1	93.3471 90 69.2308 78.2609 93.2871 (2, 2, 2, 0.01, 2)	93.7705 90 75 81.8182 82.1584 (0.5, 20)	93.7705 90 75 81.8182 82.1584 (2, 200)	93.7603 90 75 81.8182 93.6849 (0.03125, 0.03125, 0.03125, 0.03125, 0.03125, 0.01, 2)	92.9339 90 64.2857 75 92.8876 (2, 2, 2, 2, 0.03125, 100, 0.01)	93.7603 90 75 81.8182 93.6849 (0.5, 0.5, 0.5, 0.5, 8, 100, 0.0001, 0.01)
votes	**96.8148** 96.2963 96.2963 96.2963 96.8134 (2, 2, 0.0078125, 0.01, 2)	95.9064 94.4444 96.2264 95.3271 95.3313 (128, 50)	96.4425 98.1481 92.9825 95.4955 95.5304 (0.03125, 500)	95.8889 94.4444 96.2264 95.3271 95.878 (8, 8, 8, 8, 0.5, 0.01, 128)	95.8889 94.4444 96.2264 95.3271 95.878 (8, 8, 8, 8, 0.125, 0.01, 32)	**96.8148** 96.2963 96.2963 96.2963 96.8134 (0.125, 0.125, 0.125, 0.125, 0.03125, 100, 0.001, 0.05)
wdbc	98.8462 97.6923 100 98.8327 98.8394 (128, 128, 8, 0.01, 2)	96.5649 93.1298 100 96.4427 96.5038 (32, 1000)	98.4733 96.9466 100 98.4496 98.4614 (0.125, 100)	95.7692 99.2308 97.7273 98.4733 95.7067 (2, 2, 2, 2, 0.0078125, 0.01, 64)	98.4615 96.9231 100 98.4375 98.4495 (0.5, 0.5, 0.5, 0.5, 0.125, 100, 0.01)	97.5641 97.6923 99.2188 98.4496 97.564 (0.00195312, 0.00195312, 0.00195312, 0.00195312, 0.125, 500, 0.0001, 0.45)

Now, the F-statistic is calculated as:

$$F_{\text{FR}} = \frac{(17-1) \times 36.7127}{17 \times (6-1) - 36.7127} = 12.0521$$

Since we are evaluating 17 models over 6 datasets, the degrees of freedom F_{FR} are $(n-1) = 5$ and $(n-1)(M-1) = 80$. At a significance level of $\alpha = 0.05$, the critical F-value $F_{5,80}$ is 2.3287. Given that $F_{\text{FR}} > F_{\text{crit}}$, we reject the null hypothesis, which means there's strong evidence that the classifiers perform differently from one another.

To figure out which models truly perform differently from each other, we use the Nemenyi post-hoc test. This involves calculating something called the critical difference (CD), which tells us how big the gap in performance needs to be for it to count as:

$$\text{CD} = q_\alpha \cdot \sqrt{\frac{n(n+1)}{6M}} \tag{3}$$

Substituting $q_\alpha = 2.598$, $n = 6$, and $M = 15$, we obtain:

$$\text{CD} = 2.598 \cdot \sqrt{\frac{6 \cdot 7}{6 \cdot 15}} = 1.6671.$$

This critical difference value of 1.6671 allows us to visually compare the model rankings using the Nemenyi graph. Nemenyi test shows that PMvTRVFL significantly outperforms other existing models and their pairwise differences exceed the critical difference CD.

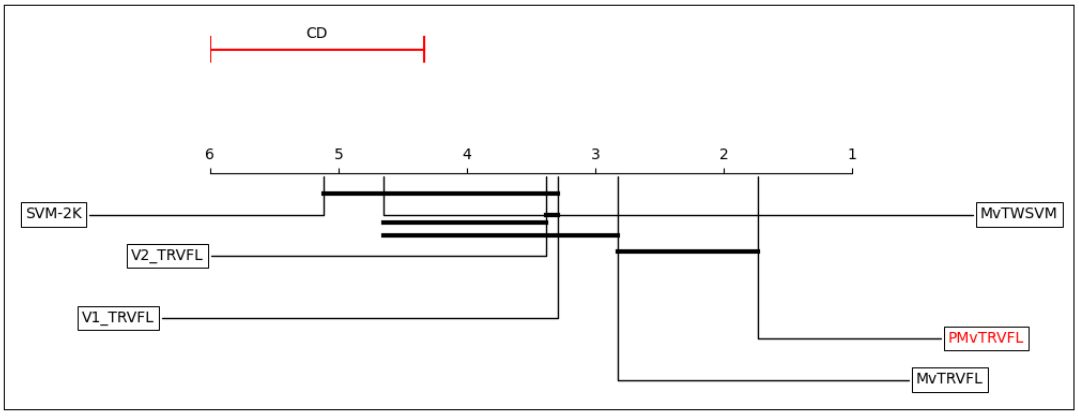

Figure III: Nemenyi test graph representation on various classifiers

The following conclusions can be drawn from the above tests:

- The differences between the average rank of all models, SVM-2K, V1_TRVFL, V2_TRVFL, MvTWSVM, and MvTRVFL compared to the proposed PMvTRVFL are 3.38, 1.56, 1.65, 2.94, and 1.09 respectively, all of which exceed the critical difference CD = 1.6671 in most cases. Hence, PMvTRVFL demonstrates superior generalization performance over other models based on average rank.
- Out of 17 datasets, the proposed PMvTRVFL achieves the best or second-best rank in 13 datasets, clearly indicating its consistent and robust classification capability across diverse domains. Figure III provides a statistical comparison of the six classifiers, highlighting the superior rank of PMvTRVFL.

References

[1] Corinna Cortes and Vladimir Vapnik. Support vector machine. *Machine learning*, 20(3):273–297, 1995.

[2] Reshma Khemchandani, Suresh Chandra, et al. Twin support vector machines for pattern classification. *IEEE Transactions on pattern analysis and machine intelligence*, 29(5):905–910, 2007.

[3] Divya Tomar and Sonali Agarwal. Twin support vector machine: a review from 2007 to 2014. *Egyptian Informatics Journal*, 16(1):55–69, 2015.

[4] Jason Farquhar, David Hardoon, Hongying Meng, John Shawe-Taylor, and Sandor Szedmak. Two view learning: Svm-2k, theory and practice. *Advances in neural information processing systems*, 18, 2005.

[5] Xijiong Xie and Shiliang Sun. Multi-view twin support vector machines. *Intelligent Data Analysis*, 19(4):701–712, 2015.

[6] Xijiong Xie and Shiliang Sun. Multi-view support vector machines with the consensus and complementarity information. *IEEE Transactions on Knowledge and Data Engineering*, 32(12):2401–2413, 2019.

[7] Chunling Lou and Xijiong Xie. Multi-view intuitionistic fuzzy support vector machines with insensitive

pinball loss for classification of noisy data. *Neurocomputing*, 549:126458, 2023.

[8] Ashwani Kumar Malik, Ruobin Gao, MA Ganaie, Muhammad Tanveer, and Ponnuthurai Nagaratnam Suganthan. Random vector functional link network: Recent developments, applications, and future directions. *Applied Soft Computing*, 143:110377, 2023.

[9] Upendra Mishra, Deepak Gupta, and Barenya Bikash Hazarika. An efficient angle-based twin random vector functional link classifier. *Applied Soft Computing*, 164:112005, 2024.

[10] Chittabarni Sarkar, Deepak Gupta, and Barenya Bikash Hazarika. 1-norm twin random vector functional link networks based on universum data for leaf disease detection. *Applied Soft Computing*, 148:110850, 2023.

[11] Robert Duin. Multiple Features. UCI Machine Learning Repository, 1998. DOI: https://doi.org/10.24432/C5HC70.

[12] Kenta Nakai. Ecoli. UCI Machine Learning Repository, 1996. DOI: https://doi.org/10.24432/C5388M.

[13] Stefan Aeberhard and M. Forina. Wine. UCI Machine Learning Repository, 1992. DOI: https://doi.org/10.24432/C5PC7J.

[14] Wing S. Hutton L. Sigillito, V. and K. Baker. Ionosphere. UCI Machine Learning Repository, 1989. DOI: https://doi.org/10.24432/C5W01B.

[15] Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research*, 7(Jan):1–30, 2006.