Supporting Document for

A Novel Fuzzy Large Margin Distribution Machine with Unified Pinball Loss

Libo Zhang, Denghao Dong, Lianyi Luo, and Dun Liu

This PDF file includes:

I. The Proof of Theorem 3

Proof. The matrix form of FUPLDM is

$$\begin{aligned}
\min_{\delta,\xi} \quad & \frac{1}{2}\delta^T Q\delta + E^T\delta + \sum_{i=1}^{m} C_i s_i \xi_i, \\
s.t. \quad & \xi_i \geq 1 - \Theta_i^{\mathbf{H_0}}, \\
& \xi_i \geq -\tau(1 - \Theta_i^{\mathbf{H_0}}), \qquad (\tau \in (0,1])
\end{aligned} \tag{1}$$

where $Q = 4\lambda_1(mK^TK - (KY)(KY)^T + 4\lambda_1 m^2 K)/m^2$ and $E = -\lambda_2 KY/m$.

The Lagrangian function of model (1) can be rewritten as

$$\begin{aligned}
\mathcal{L}(\delta, b, \xi; \alpha, \beta) = & \frac{1}{2}\delta^T Q\delta + E^T\delta + \sum_{i=1}^{m} C_i s_i \xi_i \\
& - \sum_{i=1}^{m} \alpha_i(\xi_i + y_i\delta^T K_i - 1) - \sum_{i=1}^{m} \beta_i(\xi_i + \tau(1 - y_i\delta^T K_i)).
\end{aligned} \tag{2}$$

Converting $\{\delta, b, \xi\}$ to $\{0\}$, the corresponding Karush-Kuhn-Tucker (KKT) condition of model (1) is as follows:

$$\begin{aligned}
& \tfrac{\partial\mathcal{L}}{\partial\delta} = Q\delta + E - \sum_{i=1}^{m} \alpha_i y_i K_i + \tau\sum_{i=1}^{m} \beta_i y_i K_i = 0, \\
& \tfrac{\partial\mathcal{L}}{\partial b} = \sum_{i=1}^{m} y_i(\alpha_i - \tau\beta_i) = 0, \\
& \tfrac{\partial\mathcal{L}}{\partial\xi_i} = C_i s_i - \alpha_i - \beta_i = 0, \ \forall \text{ i} = 1, 2, ..., \text{m}.
\end{aligned} \tag{3}$$

Substituting the Eq. (3) into Eq. (2), the objective function of model (1) can be transformed as follows:

$$\begin{aligned}
& \min_{\omega,b,\xi} \ \frac{1}{2}\|\omega\|^2 + \lambda_1\hat{\gamma}^{(\mathbf{H_0})} - \lambda_2\bar{\gamma}^{(\mathbf{H_0})} + \sum_{i=1}^{m} C_i s_i \xi_i \\
& = \frac{1}{2}(eYK\alpha - \tau eYK\beta - E)^T(Q^{-1})^T(eYK\alpha - \tau eYK\beta - E) \\
& \qquad + E^TQ^{-1}(eYK\alpha - \tau eYK\beta - E) + \sum_{i=1}^{m} C_i s_i \xi_i \\
& = \frac{1}{2}(eYK(\alpha - \tau\beta))^T(Q^{-1})^T eYK(\alpha - \tau\beta) \\
& \qquad - E^T(Q^{-1})^T eYK(\alpha - \tau\beta) + \sum_{i=1}^{m} C_i s_i \xi_i \\
& = \frac{1}{2}(\alpha - \tau\beta)^T A(\alpha - \tau\beta) + (\frac{\lambda_2 Ae - me}{m})^T(\alpha - \tau\beta),
\end{aligned} \tag{4}$$

where $A = [diag(Y)]KQ^{-1}K[diag(Y)]$, $e = [1, ..., 1]_{1\times m}$; α and β are both column vectors of

$m \times 1$. The constraint matrix form of FUPLDM can also be obtained by the above method, and we can get the initial form of the dual QPP of FUPLDM.

The result follows. $\square$

II. The Proof of Theorem 4

Proof. Incorporating the case of $\tau < 0$, the following complete FUPLDM dual problem can be obtained:

$$\begin{aligned}
\min_{\alpha,\mathbf{B}} \quad & \frac{1}{2}(\alpha - v\mathbf{B})^T A(\alpha - v\mathbf{B}) + (\frac{\lambda_2 Ae - me}{m})^T(\alpha - v\mathbf{B}),\\
s.t. \quad & \mathrm{C}_i s_i - \alpha_i - \frac{1}{|\tau|}\mathbf{B_i} = 0,\\
& \alpha_i \geq 0,\ \mathbf{B}_i \geq 0,\ i = 1, ..., m, \quad (\tau \in [-1, 0) \cup (0, 1])
\end{aligned} \tag{5}$$

where $\mathbf{B} = |\tau|\beta$. Assuming $\eta = \alpha - v\mathbf{B}$, the following equivalent formula can be obtained

$$\begin{aligned}
\min_{\eta} \quad & \frac{1}{2}\eta^T A\eta + (\frac{\lambda_2 Ae - me}{m})^T\eta,\\
s.t. \quad & \mathrm{C}_i s_i - \alpha_i - \frac{1}{|\tau|}\mathbf{B_i} = 0,\\
& \alpha_i \geq 0,\ \mathbf{B}_i \geq 0,\ i = 1, ..., m. \quad (\tau \in [-1, 0) \cup (0, 1])
\end{aligned} \tag{6}$$

The constraints of model (6) can be further simplified by inequality transformation, and the derivation process is demonstrated as follows:

$$\begin{aligned}
\eta_{\tau>0} = \alpha - \mathbf{B}: \quad & -|\tau|\,\alpha_i - \mathbf{B_i} = -|\tau|\,C_i s_i \leq \eta_{i(\tau>0)},\\
\eta_{\tau<0} = \alpha + \mathbf{B}: \quad & |\tau|\,\alpha_i + \mathbf{B_i} = |\tau|\,C_i s_i \leq \eta_{i(\tau<0)}.
\end{aligned} \tag{7}$$

Hence, the first equality constraint in model (6) can be transformed into: $\eta_i \geq -v\,|\tau|\,C_i s_i$. Moreover, the corresponding upper bounds can be derived: $\eta_i \leq C_i s_i$. In this way, we can get the final dual QPP of FUPLDM.

The result follows. □

III. THE RECURSIVE ALGORITHM OF FUPLDM

The final dual QPP of FUPLDM can be written as:

$$\begin{aligned} \min_{\eta} \quad & \frac{1}{2}\eta^T A\eta + (\frac{\lambda_2 Ae - me}{m})^T\eta, \qquad (\tau \in [-1,0) \cup (0,1]) \\ s.t. \quad & -v\,|\tau|\,C_i s_i \le \eta_i \le C_i s_i,\ \ i = 1, ..., m. \end{aligned} \tag{8}$$

Since (8) is a simple decoupled frame constraint and convex quadratic objective function, it can be solved efficiently by the two-coordinate descent method [1]. In the two-coordinate descent method [2], one variable is chosen to be minimized while the other variables are kept constant in each iteration, and a closed form solution is obtained in each iteration. Specifically, the following issues need to be addressed

$$\begin{aligned} \min_{t} \quad & f(\eta + te_i), \\ \text{s.t.} \quad & -v\,|\tau|\,C_i s_i \le \eta + te_i \le C_i s_i, \end{aligned} \tag{9}$$

where e_i represents a one-hot vector whose i-th component is 1 and the rest are 0. Let $A = [a_{ij}]_{i,j=1,\dots,m}$, we can get

$$f(\eta + te_i) = \frac{1}{2}a_{ii}t^2 + [\nabla f(\eta)]_i t + f(\eta), \tag{10}$$

where $[\nabla f(\eta)]_i$ is the i-th component of the gradient $\nabla f(\eta)$. Since $f(\eta)$ is a constant with respect to t, it can be removed. We can easily find that Eq. (9) has an optimal value at $t = 0$ if and only if $[\nabla^o f(\eta)]_i = 0$, which is as follows

$$[\nabla^o f(\eta)]_i = \begin{cases} [\nabla f(\eta)]_i, \ \ -v\,|\tau|\,C_i s_i \le \eta_i \le C_i s_i, \\ \min\left(-v\,|\tau|\,C_i s_i, [\nabla[f(\eta)]_i\right), \ \ \eta_i = -v\,|\tau|\,C_i s_i, \\ \max\left(-v\,|\tau|\,C_i s_i, [\nabla f(\eta)]_i\right), \ \ \eta_i = C_i s_i. \end{cases} \tag{11}$$

A closed-form solution to Eq. (8) is expressed as

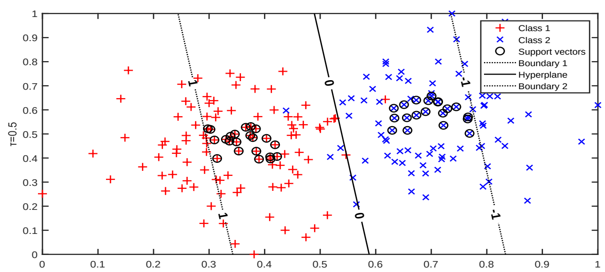
$$\eta_i^{new} = \min(\max(\eta_i - \frac{[\nabla f(\eta)]_i}{a_{ii}}, -v\,|\tau|\,C_i s_i), C_i s_i), \tag{12}$$

where $f(\eta)$ is the objective function of Eq. (8). Meanwhile, η also can be solved by using the programming function *quadprog* in the MATLAB toolbox.

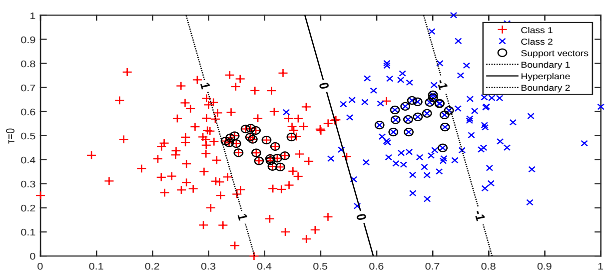
IV. THE PROPERTIES OF FUPLDM

To better illustrate the above discussion, the following examples are used for verification.

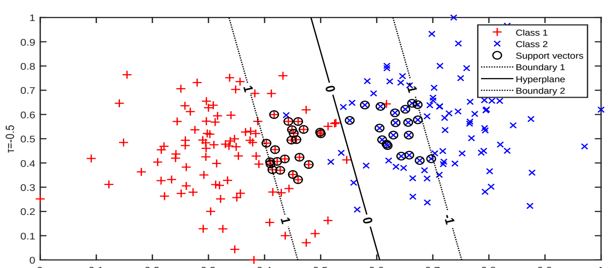
Fig. S1 shows the different classification effects with different τ values. When $\tau = 0$, FUPLDM degenerates into F-LDM with hinge loss, and the result is shown in Fig. S1(b). There is only a small number of points in $U^{+}_{\omega,b}$. When $\tau = 0.5$, the number of samples in $U^{+}_{\omega,b}$ increases, i.e., the difference between the number of samples in $U^{+}_{\omega,b}$ and $U^{-}_{\omega,b}$ becomes smaller, as shown in Fig. S1(a). When $\tau = -0.5 < 0$, the correctly classified points gain a definite gain, resulting in less points in $U^{-}_{\omega,b}$, as shown in Fig. S1(c).



(a) FUPLDM with $\tau = 0.5$



(b) FUPLDM with τ = 0



(c) FUPLDM with $\tau = -0.5$

Fig. S1: Illustrations of (a) FUPLDM with $\tau = -0.5$, (b) FUPLDM with $\tau = 0$ (F-LDM with hinge loss), and (c) FUPLDM with $\tau = 0.5$ on a 2-D artificial data set. FUPLDM gives weights of correctly classified and misclassified points via UP loss. For FUPLDM, as the parameter τ increases, the weight on correctly classified points becomes higher. Hence, the boundary between the positive hyperplane and the negative hyperplane becomes larger.

V. The Scoring Strategy of FUPLDM on UCI Dataset

Assuming $N^{\bullet}_{\text{algi}}, N^{\Omega}_{\text{algi}}, N^{\circ}_{\text{algi}}$ denotes the number of wins, draws, and losses for algorithm i in the j-th dataset, respectively. Calculate the $\sum N^{\diamondsuit}_{\text{algi}}(\diamondsuit \in \{\bullet, \Omega, \circ\})$ of each algorithm, which represents the 'WIN/DRAW/LOSS' project score of the algorithm, respectively. The final score for each algorithm is indicated with $\sum(N^{\bullet}_{\text{algi}} + N^{\Omega}_{\text{algi}}/2 - N^{\circ}_{\text{algi}}) + 80$. ($0 \leq FinalScore \leq 160$). The scoring strategy on 6 algorithms are shown in Table S1.

TABLE S1: Scoring Strategy Sheet (#Style).

	SVM	Pin-SVM	UPSVM	LDM	F-LDM	FUPLDM	Total Score
WIN	$\sum N^{\bullet}_{\mathrm{SVM}}$	$\sum N^{\bullet}_{\mathrm{Pin-SVM}}$	$\sum N^{\bullet}_{\mathrm{UPSVM}}$	$\sum N^{\bullet}_{\mathrm{LDM}}$	$\sum N^{\bullet}_{\mathrm{F-LDM}}$	$\sum N^{\bullet}_{\mathrm{FUPLDM}}$	
DRAW	$\sum N^{\Omega}_{\mathrm{SVM}}$	$\sum N^{\Omega}_{\mathrm{Pin-SVM}}$	$\sum N^{\Omega}_{\mathrm{UPSVM}}$	$\sum N^{\Omega}_{\mathrm{LDM}}$	$\sum N^{\Omega}_{\mathrm{F-LDM}}$	$\sum N^{\Omega}_{\mathrm{FUPLDM}}$	
LOSS	$\sum N^{\circ}_{\mathrm{SVM}}$	$\sum N^{\circ}_{\mathrm{Pin-SVM}}$	$\sum N^{\circ}_{\mathrm{UPSVM}}$	$\sum N^{\circ}_{\mathrm{LDM}}$	$\sum N^{\circ}_{\mathrm{F-LDM}}$	$\sum N^{\circ}_{\mathrm{FUPLDM}}$	$\sum(N^{\bullet}_{\text{algi}} + N^{\Omega}_{\text{algi}}/2 - N^{\circ}_{\text{algi}}) + 80$
Final Score	-	-	-	-	-	-	

TABLE S2: The Algorithms Score Based on Scoring Strategies.

	SVM	Pin-SVM	UPSVM	LDM	FUPLDM	Total Score (S2)
WIN	4	20	32	36	60	
DRAW	22	19	21	17	19	
LOSS	54	41	27	27	1	451
Final Score (S1)	41	68.5	95.5	97.5	**148.5**	
Δ	9.09%	15.19%	21.18%	21.62%	**32.82%**	

* Δ = S1/S2.

The performance of each algorithm compared to the others can be further analyzed based on Table S1. Assuming $N^{\bullet}_{\text{algi}}, N^{\Omega}_{\text{algi}}, N^{\circ}_{\text{algi}}$ denotes the number of wins, draws, and losses for algorithm i in the j-th dataset, respectively. Calculate the $\sum N^{\diamondsuit}_{\text{algi}}(\diamondsuit \in \{\bullet, \Omega, \circ\})$ of each algorithm, which represents the 'WIN/DRAW/LOSS' project score of the algorithm, respectively. The final score for each algorithm is indicated with $\sum(N^{\bullet}_{\text{algi}} + N^{\Omega}_{\text{algi}}/2 - N^{\circ}_{\text{algi}}) + 80$. ($0 \leq FinalScore \leq 160$). The scoring results on 6 algorithms are shown in Table S2.

VI. The Parametric Asymptotic Stability Analysis

Since λ_1 and λ_2 control the margin variance and the margin mean, respectively, they have different effects on the model when they take on different values, and when any of them tends to zero, it means that the model no longer cares about the significance that this term brings to the model. For example, $\lambda_1 \to 0$ indicates that the model is only concerned with the margin mean and not the margin variance. Similarly, when the variable tends to infinity, it intensifies the model's concern for the term, e.g., when $\lambda_1 \to \infty$, the entire objective function is dominated by the margin variance and thus no longer concerns itself with the support vector and the margin mean. Both extreme cases are what we need to avoid, so in the experimental part of the model, we make λ take values within a certain range to ensure the asymptotic stability of the model.

In addition, the analysis is similar for the parameter C, which mainly characterizes how much the model penalizes classification errors as a way to tolerate some samples with linear classification errors, commonly known as soft intervals. When C obtains an extreme value $0 \; or \; \infty$, it will also make the model become less desirable and the above problems will occur.

For FUPLDM, C is a hyperparameter that controls the strength of the penalty strategy, so when $C = 2^3$, it means that the model penalizes the slack variables to a greater extent at this time and the effect of soft intervals is suppressed, but does not affect the fuzzy properties of our proposed fuzzy strategy, which is due to the fact that the fuzzy affiliation is calculated before training based on the characteristics of the data and is not affected during the training process.

VII. THE ALGORITHM PERFORMANCE RANKING TABLE

TABLE S3: Accuracy Comparison of the nine algorithms on the UCI Datasets

Dataset	SVM	LDM	FLDM	PinSVM	UPSVM	FSVM	FTSVM	CDFTSVM	FUPLDM
Monk 1	64.35	68.06	67.59	66.67	67.13	65.05	63.89	63.66	**68.52**
Monk 2	67.13	67.36	67.13	67.13	67.13	67.13	46.53	67.13	**67.59**
Monk 3	81.02	84.03	81.94	84.49	**88.89**	81.02	77.08	81.48	**88.89**
Heberman	73.08	73.08	73.08	**73.72**	73.08	73.08	76.28	71.15	73.08
Statlog	**86.67**	85.00	85.83	**86.67**	**86.67**	85.83	80.00	87.50	**86.67**
Pima-Indian	67.09	79.91	79.70	69.87	75.21	67.09	74.79	74.79	**80.34**
Echo	70.59	88.24	90.20	84.31	90.20	70.59	88.24	88.24	**92.16**
Australian	84.48	86.55	85.86	84.48	**87.24**	84.83	81.72	81.45	**87.24**
Bupa	63.16	70.53	71.58	63.16	63.16	63.16	63.74	64.91	**72.63**
Daibetes	67.91	81.72	80.97	70.90	78.73	67.91	77.99	79.48	**83.21**
Fertility	**94.00**	**94.00**	**94.00**	**94.00**	**94.00**	**94.00**	94.00	94.00	**94.00**
Original	98.43	98.43	98.43	98.43	**98.69**	98.59	98.59	98.59	**98.69**
BUPA	60.00	**75.86**	75.17	60.00	60.69	57.31	63.74	64.91	**75.86**
Wine	77.27	73.86	73.86	**88.64**	86.36	80.68	40.91	42.05	**88.64**
Creditcard	93.13	96.56	96.20	94.36	93.74	93.74	98.05	**98.12**	96.20
Breast	98.12	**98.75**	98.43	98.12	98.12	96.55	88.86	93.15	98.43
Votes	60.43	**95.74**	**95.74**	94.04	95.32	73.62	94.47	94.47	**95.74**
WDBC	77.51	98.82	98.82	91.12	98.22	79.29	93.49	94.08	**100.00**
Patients	**100.00**	74.25	83.51	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Aps_failure	93.50	**98.35**	**98.35**	93.50	95.92	96.69	94.00	95.00	**98.35**

TABLE S4: Ranking of algorithms

Dataset	SVM	LDM	FLDM	PinSVM	UPSVM	FSVM	FTSVM	CDFTSVM	FUPLDM
Monk 1	7	2	3	5	4	6	8	9	1
Monk 2	5	2	5	5	5	5	9	8	1
Monk 3	7.5	4	5	3	1.5	7.5	9	6	1.5
Heberman	5.5	5.5	5.5	2	5.5	5.5	1	9	5.5
Statlog	3.5	8	6.5	3.5	3.5	6.5	9	1	3.5
Pima-Indian	8.5	2	3	7	4	8.5	5.5	5.5	1
Echo	8.5	4	2.5	7	2.5	8.5	5.5	5.5	1
Australian	6.5	3	4	6.5	1.5	5	8	9	1.5
Bupa	7.5	3	2	7.5	7.5	7.5	5	4	1
Daibetes	8.5	2	3	7	5	8.5	6	4	1
Fertility	4	4	4	4	4	4	8.5	8.5	4
Original	7.5	7.5	7.5	7.5	1.5	3	4.5	4.5	1.5
BUPA	7.5	1.5	3	7.5	6	9	5	4	1.5
Wine	5	6.5	6.5	1..5	3	4	9	8	1.5
Creditcard	9	3	4.5	6	7.5	7.5	2	1	4.5
Breast	5	1	2.5	5	5	7	9	8	2.5
Votes	9	2	2	7	4	8	5.5	5.5	2
WDBC	9	2.5	2.5	7	4	8	6	5	1
Patients	4.5	9	8	4.5	4.5	4.5	4.5	4.5	4.5
Aps_failure	8.5	2	2	8.5	5	4	7	6	2
Ave.rk	6.85	3.73	4.10	5.82	4.23	6.38	6.35	5.80	2.15

VIII. The Properties of Unbalanced Datasets

TABLE S5: A collection of dataset attributes

Category	Datasets	#Samples	#Features	#Imbalance ratio
Biomedical	Patients	1631	78	5.40
	Heberman	306	3	2.78
	Australian	690	14	1.25
	Bupa_Liver	345	6	1.38
	Breast_Cancer	683	9	1.86
	WDBC	569	29	1.68
	BUPA	341	6	1.40
	Statlog	270	13	1.25
Algorithms	Monk 1	556	6	1.00
	Monk 2	601	6	1.94
	Monk 3	554	6	1.08
Text	Votes	435	16	1.62
	Creditcart	1565	31	4.22
Others	Aps_failure	1707	169	3.55
	Wine	178	13	1.51
	Echo	327	7	2.05
	Pima_Indian	768	8	1.87
	Daibetes	768	8	1.87
	Fertility	150	9	7.33

IX. Experiments on Unbalanced Categorization Tasks

To test the performance of FUPLDM for the imbalance task, five datasets from different domains are selected for testing, and the experimental results are shown in Table S6.

TABLE S6: Comparison of SVM, SVM-SMOTE, SVM-KSMOTE and FUPLDM algorithms in imbalance classification

Dataset	Algorithm	Accurate	AUC	Sensitivity	Specificity	Fmeasure	Gmeans
Patients	SVM	99.52	98.96	**100.00**	99.57	98.47	99.40
	SVM-SMOTE	**99.99**	**99.99**	99.92	**100.00**	**99.96**	**99.96**
	SVM-KSMOTE	99.84	99.87	98.99	**100.00**	99.49	99.49
	FUPLDM	99.88	99.89	98.51	**100.00**	99.25	99.25
Creditcard	SVM	93.13	74.98	**100.00**	93.01	33.33	96.44
	SVM-SMOTE	97.05	**94.42**	93.79	**97.79**	92.17	95.77
	SVM-KSMOTE	**97.63**	94.40	97.35	97.70	**97.59**	**97.52**
	FUPLDM	97.18	82.06	**100.00**	95.13	61.54	97.36
Aps_failure	SVM	93.50	40.28	NaN	93.50	NaN	NaN
	SVM-SMOTE	95.70	**94.12**	89.45	97.49	90.28	93.39
	SVM-KSMOTE	96.17	92.99	**94.26**	96.66	**90.95**	**95.45**
	FUPLDM	**98.02**	92.41	86.79	**99.29**	84.13	92.45
Breast	SVM	98.12	**95.90**	**100.00**	**99.38**	**96.55**	97.89
	SVM-SMOTE	95.53	95.39	93.10	97.04	94.08	95.05
	SVM-KSMOTE	96.80	95.84	98.89	95.67	95.56	97.27
	FUPLDM	**98.43**	94.24	**100.00**	97.84	94.32	**97.90**
Diabetes	SVM	91.75	55.70	**100.00**	91.64	23.81	95.73
	SVM-SMOTE	91.75	**89.42**	51.88	98.87	65.58	71.62
	SVM-KSMOTE	95.41	76.67	83.52	96.21	**69.61**	89.64
	FUPLDM	**95.96**	87.29	**100.00**	**100.00**	48.12	**95.90**

As shown in Table S6, we analyzed the experimental results of FUPLDM using the AUC, Sensitivity, Specificity, Fmeasure and Gmeans. Here, sensitivity indicates the ability of the model to correctly classify positives. SVM-SMOTE and SVM-KSMOTE are algorithms dedicated to the classification of unbalanced data. Since the main advantage of FUPLDM does not lie in dealing with unbalanced classification tasks, its advantages are not obvious. However, FUPLDM still maintains a good performance in terms of classification accuracy and sensitivity. For the base model SVM, FUPLDM shows better unbalanced classification performance. This suggests that our proposed algorithm has a positive effect in handling unbalanced classification tasks.

X. Experiments on UCI Datasets with Noise and Parameter Sensitivity Analysis

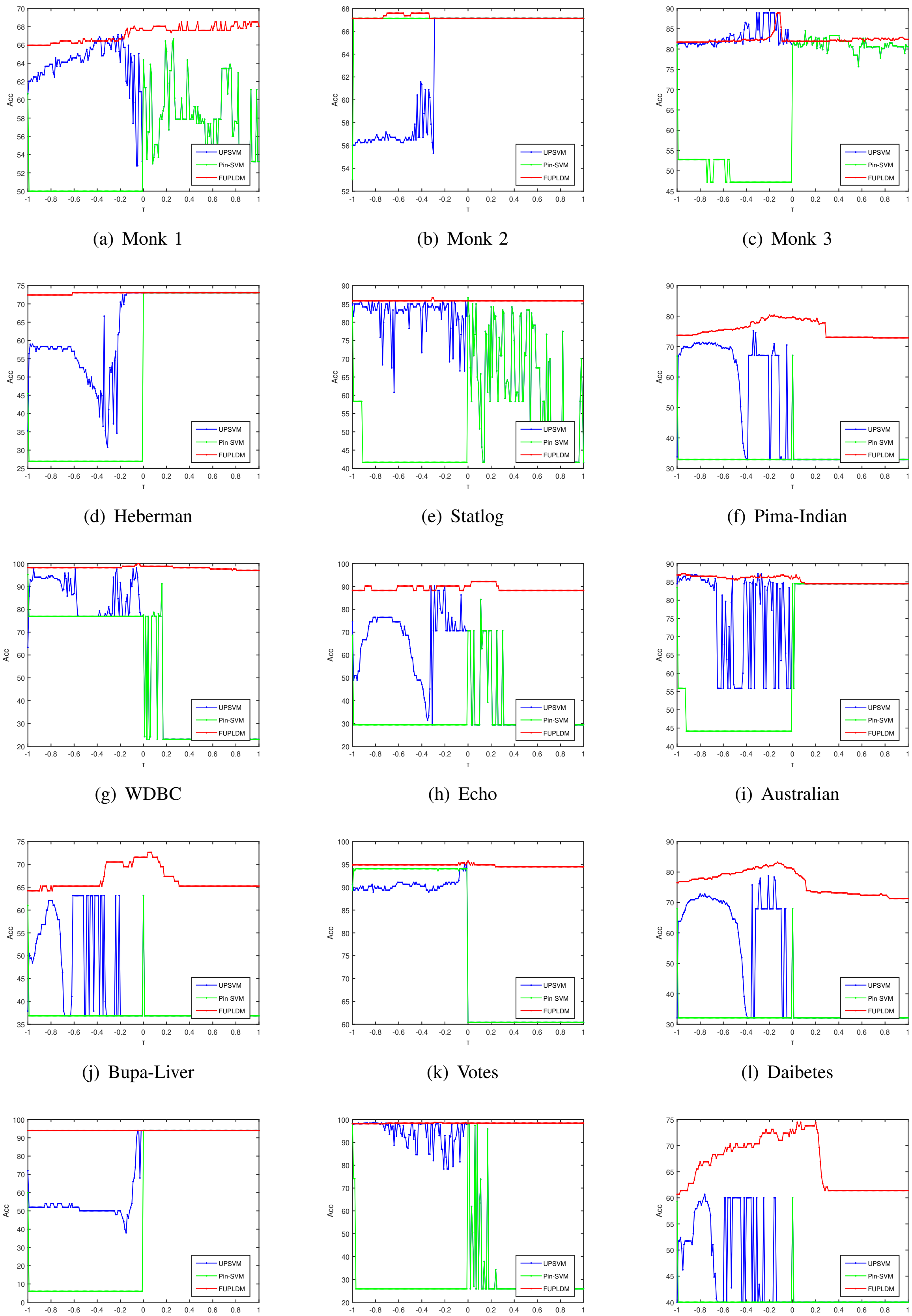
TABLE S7: Accuracy Comparison of SVM, Pin-SVM, UPSVM, LDM, F-LDM and FUPLDM on the Noisy UCI Dataset.

Dataset	SVM	Pin-SVM		UPSVM		LDM		F-LDM		FUPLDM	
	Accuracy/tau	Accuracy/tau	ΔAcc	Acc/tau	ΔAcc	Accuracy/tau	ΔAcc	Accuracy/tau	ΔAcc	Accuracy/tau	ΔAcc
	Time/C	Time/C	(Vs. SVM)	Time/C	(Vs. SVM)	Time/C	(Vs. SVM)	Time/C	(Vs. SVM)	Time/C	(Vs. SVM)
Monk 1(r = 0.05)	57.87/0	62.04/0.07	4.17	60.65/-0.75	2.78	61.11/0	3.24	63.43/0	5.56	**64.12/-0.37**	**6.25**
	1.16/0.25	1.31/0.06		0.15/0.25		0.04/4.00		0.04/2.00		**0.04/4.00**	
r = 0.1	57.64/0	59.95/-0.29	2.31	**65.05/-0.34**	**7.41**	61.11/0	3.47	63.89/0	6.25	64.58/-0.29	6.94
	0.87/0.25	0.98/0.06		0.98/4.00		0.89/4.00		0.98/8.00		**0.87/0.50**	
r = 0.5	57.64/0	58.33/0.06	0.69	**65.28/-0.33**	**7.64**	61.34/0	3.70	64.35/0	6.71	64.58/-0.24	6.94
	0.99/0.06	1.09/0.06		1.09/4.00		0.04/2.00		**0.02/2.00**		0.03/4.00	
Monk 2(r = 0.05)	**67.13/0**	**67.13/-0.99**	0.00	**67.13/-0.71**	0.00	**67.13/0**	0.00	**67.13/0**	0.00	**67.13/-1**	0.00
	1.60/0.02	1.81/0.02		0.23/0.02		**0.05/0.02**		**0.05/0.02**		**0.05/0.02**	
r = 0.1	**67.13/0**	**67.13/-1**	0.00	**67.13/-0.49**	0.00	**67.13/0**	0.00	**67.13/0**	0.00	**67.13/-1**	0.00
	1.27/0.03	1.45/0.03		1.40/0.03		1.28/0.03		1.28/0.03		**1.26/0.03**	
r = 0.5	**67.13/0**	**67.13/-0.99**	0.00	**67.13/-0.49**	0.00	**67.13/0**	0.00	**67.13/0**	0.00	**67.13/-1**	0.00
	1.23/0.03	1.39/0.03		1.34/0.03		**0.04/0.03**		0.05/0.03		**0.04/0.03**	
Monk 3(r = 0.05)	72.45/0	**74.07/0.61**	**1.62**	**74.07/0.61**	**1.62**	72.45/0	0.00	72.45/0	0.00	72.92/0.96	0.47
	1.12/0.50	1.28/0.50		0.16/0.50		0.04/0.50		**0.03/0.50**		**0.03/4.00**	
r = 0.1	71.76/0	**73.61/0.91**	**1.85**	**73.61/0.78**	**1.85**	72.45/0	0.69	72.45/0	0.69	72.92/0.91	1.16
	0.87/0.25	0.97/0.50		0.93/0.50		0.69/0.25		**0.68/0.25**		**0.68/4.00**	
r = 0.5	72.69/0	73.61/0.37	0.92	73.61/0.37	0.92	72.69/0	0.00	72.92/0	0.23	**73.84/0.9**	**1.15**
	0.88/0.25	0.98/0.25		0.98/0.25		**0.02/0.13**		0.04/1.00		**0.02/8.00**	
Heberman(r = 0.05)	73.08/0	73.08/0	0.00	**76.28/-0.43**	**3.20**	73.08/0	0.00	73.08/0	0.00	73.08/-1	0.00
	0.55/0.02	0.74/0.02		0.12/0.02		0.05/0.02		**0.04/0.02**		0.05/0.02	
r = 0.1	**73.08/0**	**73.08/-1**	0.00	**73.08/-0.17**	0.00	**73.08/0**	0.00	**73.08/0**	0.00	**73.08/-1**	0.00
	0.43/0.03	0.56/0.03		0.51/0.03		0.45/0.03		0.45/0.03		**0.42/0.03**	
r = 0.5	73.08/0	**76.28/-1**	**3.20**	73.08/-0.16	0.00	73.08/0	0.00	73.08/0	0.00	73.08/-1	0.00
	0.46/0.03	0.55/0.06		0.54/0.03		0.04/0.03		0.04/0.03		**0.03/0.03**	
Statlog(r = 0.05)	84.17/0	84.17/0	0.00	84.17/0	0.00	**85/0**	**0.83**	84.17/0	0.00	**85/-0.25**	**0.83**
	0.46/0.06	0.73/0.06		0.12/0.06		0.05/0.03		**0.04/0.03**		0.05/0.03	
r = 0.1	84.17/0	**85.00/-0.46**	**0.83**	**85.00/-0.23**	**0.83**	**85.00/0**	**0.83**	84.17/0	0.00	**85.00/-0.46**	**0.83**
	0.35/0.06	0.53/0.25		0.52/0.25		**0.22/0.03**		**0.22/0.06**		**0.22/0.03**	
r = 0.5	83.33/0	83.33/0	0.00	**85.83/-0.78**	**2.50**	85/0	1.67	84.17/0	0.84	85/-0.62	1.67
	0.35/0.13	0.52/0.13		**0.52/0.50**		0.05/0.03		**0.03/0.03**		0.04/0.03	
Pima-Indian(r = 0.05)	67.09/0	69.87/-1	2.78	67.09/-0.37	0.00	78.85/0	11.76	78.85/0	11.76	**79.06/-0.14**	**11.97**
	3.13/0.02	3.97/0.13		0.58/0.02		0.15/4.00		**0.13/2.00**		**0.13/1.00**	
r = 0.1	67.09/0	67.09/-0.21	0.00	76.71/-0.28	9.62	78.85/0	11.76	**79.06/0**	**11.97**	**79.06/-0.21**	**11.97**
	2.54/0.03	3.27/0.03		3.30/2.00		**2.49/8.00**		2.50/4.00		2.50/1.00	
r = 0.5	67.09/0	69.87/-1	2.78	74.79/-0.2	7.70	79.06/0	11.97	79.06/0	11.97	**79.27/0.01**	**12.18**
	2.53/0.03	3.09/0.03		3.08/4.00		0.08/4.00		0.09/4.00		**0.07/4.00**	
WDBC(r = 0.05)	91.72/0	91.72/-1	0.00	95.86/-0.62	4.14	96.45/0	4.73	95.86/0	4.14	**97.04/0.32**	**5.32**
	2.17/0.02	4.64/0.03		0.85/0.02		**0.30/8.00**		**0.30/0.02**		0.33/8.00	
r = 0.1	90.53/0	95.86/0.32	5.33	96.45/-0.79	5.92	96.45/0	5.92	95.86/0	5.33	**97.04/0.32**	**6.51**
	1.67/0.03	4.19/1.00		4.09/2.00		3.69/8.00		3.69/0.03		3.69/8.00	
r = 0.5	89.94/0	94.08/-1	4.14	96.45/-0.78	6.51	96.45/0	6.51	95.86/0	5.92	**97.04/0.39**	**7.10**
	1.55/0.03	3.55/0.13		3.59/4.00		0.19/8.00		**0.16/0.03**		0.19/8.00	
Echo(r = 0.05)	76.47/0	90.2/0.05	13.73	90.2/-0.29	13.73	92.16/0	15.69	94.12/0	17.65	**96.08/-0.11**	**19.61**
	0.12/0.02	0.18/0.02		0.03/0.02		**0.02/0.02**		**0.02/0.50**		**0.02/2.00**	
r = 0.1	70.59/0	90.2/-0.11	19.61	92.16/-0.32	21.57	94.12/0	23.53	94.12/0	23.53	**96.08/-0.11**	**25.49**
	0.09/0.03	0.15/2.00		0.14/0.50		0.20/0.06		0.20/0.13		0.20/4.00	
r = 0.5	70.59/0	94.12/0.09	23.53	94.12/0.09	23.53	94.12/0	23.53	94.12/0	23.53	**96.08/-0.11**	**25.49**
	0.09/0.03	0.13/2.00		0.13/2.00		**0.01/0.06**		**0.01/0.13**		**0.01/4.00**	
Australian(r = 0.05)	55.86/0	75.17/0.18	19.31	67.93/-0.83	12.07	74.48/0	18.62	74.48/0	18.62	**75.86/-0.09**	**20.00**
	3.06/0.02	5.58/0.25		0.89/0.02		0.29/1.00		**0.28/2.00**		0.30/4.00	
r = 0.1	55.86/0	75.86/-0.1	20.00	**77.24/-0.97**	**21.38**	74.48/0	18.62	74.48/0	18.62	76.21/-0.1	20.35
	2.27/0.03	4.27/0.25		4.10/4.00		**2.26/1.00**		**2.26/2.00**		**2.26/4.00**	
r = 0.5	55.86/0	74.83/0.41	18.97	**77.24/-0.97**	**21.38**	74.83/0	18.97	74.48/0	18.62	76.21/-0.09	20.35
	2.36/0.03	4.28/0.06		4.35/0.50		**0.15/1.00**		0.18/2.00		0.18/4.00	

TABLE S8: ($Continued$) Accuracy Comparison of SVM, Pin-SVM, UPSVM, LDM, F-LDM and FUPLDM on the Noisy UCI Datasets.

Dataset	SVM	Pin-SVM		UPSVM		LDM		F-LDM		FUPLDM	
	Accuracy/τ	Accuracy/τ	$\Delta\tau$	Accuracy/τ	ΔAcc	Accuracy/τ	ΔAcc	Accuracy/τ	ΔAcc	Accuracy/τ	ΔAcc
	Time/C	Time/C	(Vs. SVM)	Time/C	(Vs. SVM)	Time/C	(Vs. SVM)	Time/C	(Vs. SVM)	Time/C	(Vs. SVM)
Bupa-Liver(r = 0.05)	63.16/0	63.16/-1	0.00	63.16/0	0.00	**73.68/0**	**10.52**	72.63/0	9.47	**73.68/-0.09**	**10.52**
	0.75/0.03	1.49/0.13		0.27/0.03		0.11/4.00		**0.09/8.00**		0.11/8.00	
r = 0.1	63.16/0	63.16/-0.1	0.00	63.16/-1	0.00	**73.68/0**	**10.52**	72.63/0	9.47	**73.68/-0.1**	**10.52**
	0.55/0.03	1.09/0.03		1.01/8.00		0.46/4.00		0.46/8.00		**0.40/8.00**	
r = 0.5	63.16/0	63.16/-1	0.00	63.16/-0.61	0.00	**72.63/0**	**9.47**	71.58/0	8.42	**72.63/-0.14**	**9.47**
	0.57/0.03	1.07/0.06		1.05/2.00		**0.06/8.00**		**0.06/8.00**		**0.06/8.00**	
Votes(r = 0.05)	80/0	82.13/-1	2.13	**83.83/-1**	**3.83**	80.43/0	0.43	80.43/0	0.43	81.28/0.11	1.28
	1.09/0.02	1.55/0.02		0.23/0.02		**0.06/0.50**		0.07/0.50		**0.06/2.00**	
r = 0.1	77.87/0	80.85/-0.07	2.98	**84.26/-1**	**6.39**	80.85/0	2.98	80.43/0	2.56	81.28/-0.07	3.41
	0.86/0.03	1.22/0.50		1.20/0.03		1.15/0.25		1.15/0.50		1.15/0.50	
r = 0.5	78.72/0	82.13/0.13	3.41	**85.11/-1**	**6.39**	81.28/0	2.56	81.7/0	2.98	82.13/0.08	3.41
	0.89/0.03	1.22/0.50		1.19/0.50		**0.04/0.25**		0.05/0.50		0.05/2.00	
Daibetes(r = 0.05)	67.91/0	70.15/-1	2.24	77.99/-0.33	10.08	80.22/0	12.31	78.73/0	10.82	**81.34/-0.35**	**13.43**
	3.66/0.02	6.28/2.00		1.17/0.02		0.39/2.00		0.41/1.00		**0.37/1.00**	
r = 0.1	67.91/0	68.28/-0.35	0.37	80.22/-0.23	12.31	80.22/0	12.31	78.73/0	10.82	**81.34/-0.35**	**13.43**
	2.81/0.03	5.02/0.50		4.88/0.25		3.46/2.00		3.46/8.00		**3.46/1.00**	
r = 0.5	67.91/0	70.15/-1	2.24	80.22/-0.22	12.31	80.22/0	12.31	78.73/0	10.82	**81.72/-0.35**	**13.81**
	2.91/0.03	4.99/0.13		4.99/8.00		0.31/2.00		0.27/8.00		**0.21/1.00**	
Fertility(r = 0.05)	94/0	94/0	0.00	94/-1	0.00	94/0	0.00	94/0	0.00	**96/-0.03**	**2.00**
	0.07/0.02	0.09/0.02		**0.01/0.02**		**0.01/0.02**		**0.01/0.02**		**0.01/2.00**	
r = 0.1	94.00/0	94/-0.04	0.00	94/-1	0.00	94/0	0.00	94/0	0.00	**96/-0.04**	**2.00**
	0.05/0.03	0.06/0.03		0.06/0.03		0.10/0.03		0.12/0.03		0.08/1.00	
r = 0.5	94/0	94/0	0.00	94/-1	0.00	94/0	0.00	94/0	0.00	**96/-0.04**	**2.00**
	0.05/0.03	0.06/0.03		0.06/0.03		**0.01/0.03**		**0.01/0.03**		**0.01/1.00**	
Breast-cancer(r = 0.05)	98.43/0	98.43/-1	0.00	98.43/-0.88	0.00	**98.69/0**	0.26	**98.69/0**	0.26	**98.69/-0.3**	0.26
	2.09/0.03	3.06/0.03		3.03/0.25		0.09/0.06		0.09/0.06		**0.08/0.06**	
r = 0.1	**98.43/0**	**98.43/-0.14**	0.00	**98.43/-1**	0.00	**98.43/0**	0.00	**98.43/0**	0.00	**98.43/-0.14**	0.00
	2.06/0.03	2.99/0.03		2.95/0.03		1.58/0.13		1.58/0.13		**1.57/0.13**	
r = 0.5	98.43/0	98.43/-1	0.00	98.43/-1	0.00	98.43/0	0.00	98.43/0	0.00	**98.69/-0.05**	**0.26**
	2.05/0.03	2.92/0.03		2.96/0.03		**0.08/0.13**		**0.08/0.13**		**0.08/0.25**	
BUPA(r = 0.05)	60/0	60.69/-1	0.69	**73.79/-0.41**	13.79	**73.79/0**	13.79	68.97/0	8.97	73.1/-0.19	13.10
	0.56/0.25	0.84/0.25		0.80/2.00		0.05/4.00		0.05/4.00		**0.04/2.00**	
r = 0.1	60.00/0	60/-0.19	0.00	60/-0.59	0.00	72.41/0	12.41	**73.79/0**	**13.79**	**73.79/-0.19**	**13.79**
	0.62/0.06	0.80/0.06		0.79/2.00		0.78/8.00		**0.78/8.00**		**0.62/4.00**	
r = 0.5	60/0	60/0	0.00	60/-0.59	0.00	73.1/0	13.10	73.79/0	13.79	**74.48/0.08**	**14.48**
	0.57/0.03	0.80/0.03		0.81/2.00		**0.05/8.00**		**0.05/8.00**		**0.05/8.00**	
Wine(r = 0.05)	77.27/0	77.27/0	0.00	**80.68/-0.01**	3.41	77.27/0	0.00	77.27/0	0.00	79.55/-0.09	2.28
	0.17/0.50	0.23/0.50		0.23/4.00		0.02/0.03		0.02/0.03		**0.01/4.00**	
r = 0.1	79.55/0	81.82/0.72	2.27	**84.09/-0.18**	**4.54**	81.82/0	2.27	81.82/0	2.27	**84.09/0.72**	**4.54**
	0.23/0.06	0.29/0.25		0.29/8.00		0.52/0.50		0.52/0.50		0.50/0.13	
r = 0.5	79.55/0	81.82/0.21	2.27	84.09/-0.24	4.54	81.82/0	2.27	82.95/0	3.40	**86.36/0.7**	**6.81**
	0.15/0.06	0.23/0.13		0.20/2.00		**0.01/0.06**		**0.01/0.50**		**0.01/0.25**	
Average of ΔAcc	-		3.42		5.29		6.32		6.25		**7.36**

* The traversal range of τ is [-1:0.01:1]. 'Time' denotes the CPU time consumption of the algorithm. ΔAcc is the accuracy improvement of the algorithms compared with the conventional SVM.



(a) Monk 1 (b) Monk 2 (c) Monk 3

(d) Heberman (e) Statlog (f) Pima-Indian

(g) WDBC (h) Echo (i) Australian

(j) Bupa-Liver (k) Votes (l) Daibetes

(m) Fertility (n) Breast-cancer (o) BUPA

Fig. S2: The influence of τ on the accuracy of Pin-SVM, UPSVM and FUPLDM algorithms respectively. (In different datasets)

XI. Experiments on Multiclassification

TABLE S9: Results of SVM, LDM and FUPLDM on a multiclassification task.

Dataset	Algorithm	Mean±Std	Min/Max
Beer	SVM	**99.17±1.71**	**93.33/100.0**
	LDM	99.00±0.97	**96.67/100.0**
	FUPLDM	98.25±1.71	**95.00/100.0**
Column_3C	SVM	68.39±5.14	58.87/75.81
	LDM	**70.89±6.45**	**59.68/82.26**
	FUPLDM	69.35±5.13	59.68/77.42
Archive	SVM	33.31±1.70	30.38/36.50
	LDM	86.25±0.00	86.25/86.25
	FUPLDM	**87.91±0.57**	**87.13/88.88**
Data_Cortex_Nuclear	SVM	26.48±2.66	21.30/27.78
	LDM	88.89±0.00	88.89/88.89
	FUPLDM	**95.05±1.74**	**90.15/96.53**

REFERENCES

[1] T. Zhang and Z. H. Zhou, "Large margin distribution machine," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 313–322.

[2] C. J. Hsieh, K. W. Chang, C.-J. Lin, S. S. Keerthi, and S. Sundararajan, "A dual coordinate descent method for large-scale linear svm," in *Proceedings of the 25th International Conference on Machine Learning*, 2008, pp. 408–415.