# Supplementary Material of Learning Curves of Classification Metrics based on Confusion Matrices

## Appendix A. Technical Details.

We elaborate the technical details of the two synthetic, four real-world datasets, and the settings of the six non-neural, two neural classification algorithms used in our experiments.

The two synthetic datasets include a **simple** dataset and the **Exp2** dataset. For a **simple** dataset $D_n = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$, it has a binary class label $y_i \in \{+, -\}$ with $P(y_i = +) = P(y_i = -) = \frac{1}{2}$ and ten predictors jointly drawn from two conditional Gaussian distributions, i.e. $\mathbf{x}_i|y_i = + \sim N(\mathbf{0.5}_{10}, 2I_{10})$ and $\mathbf{x}_i|y_i = - \sim N(\mathbf{0}_{10}, I_{10})$, where $\mathbf{0}_{10}$ and $\mathbf{0.5}_{10}$ denote the ten-dimensional vector with the elements of all 0 and 0.5 and $I_{10}$ denotes the ten-order identity matrix. For the **Exp2** dataset, it is a binary-class variant of the EXP6 dataset. The EXP6 dataset has a six-class label $y_i \in \{Y_j\}_{j=1}^6$ and two predictors drawn with replacement from an uniform grid space of resolution 0.1 over the region $\mathbf{X} = [0, 15] \times [0, 15]$, where class label $y_i$ is decided as follows:

$$y_i = \begin{cases} Y_1 & x_2 - f_1(x_1) \geq 0 \ \wedge \ x_2 - f_2(x_1) \geq 0 \\ Y_2 & x_2 - f_1(x_1) < 0 \ \wedge \ x_2 - f_2(x_1) \geq 0 \\ & \wedge \, x_2 - f_3(x_1) \geq 0 \\ Y_3 & x_2 - f_1(x_1) \geq 0 \ \wedge \ x_2 - f_2(x_1) < 0 \\ Y_4 & x_2 - f_1(x_1) < 0 \ \wedge \ x_2 - f_2(x_1) < 0 \\ & \wedge \, x_2 - f_3(x_1) \geq 0 \\ Y_5 & x_2 - f_2(x_1) \geq 0 \ \wedge \ x_2 - f_3(x_1) < 0 \\ Y_6 & x_2 - f_2(x_1) < 0 \ \wedge \ x_2 - f_3(x_1) < 0 \end{cases}$$

where $f_1, f_2$ and $f_3$ are decision boundary functions:

$$\begin{aligned} f_1(x) &= x^2 - 4x + 6, \quad f_2(x) = 4\sin(x/2) + 8, \\ f_3(x) &= -\frac{1}{25}(x^2 - 108x) + 236. \end{aligned} \tag{30}$$

Here, we convert the EXP6 dataset to a binary-class dataset where the original six class labels of the EXP6 dataset are combined as two classes ($Y_1$ to $Y_3$) versus ($Y_4$ to $Y_6$).

For two synthetic datasets, six popular non-neural classification algorithms are considered as follows.

- **Logistic Regression** (LR): The *glm* package in R software with a binomial link function is used.
- **Linear Discriminant Analysis** (LDA): The *MASS* package in R software with default settings is used.
- **Naive Bayes** (NB): The *e*1071 package in R software with default settings is used.

Table 6: Logarithmic transformations of learning curves for the other nine baselines.

| Baseline | Formula | Transform |
|---|---|---|
| POW2 | $e_n = a \cdot n^{-b}$ | $e_n = a \cdot n^{-b}$ |
| LOG2 | $e_n = -alog(n) + c$ | $e_n = -alog(n) + c$ |
| EXP2 | $e_n = aexp(-bn)$ | $e_n = aexp(-bn)$ |
| EXP3 | $e_n = aexp(-bn) + c$ | $log(e_n - c) = log(a) - bn$ |
| LIN2 | $e_n = -an + b$ | $e_n = -an + b$ |
| VAP3 | $e_n = exp(a + \frac{b}{n} + clog(n))$ | $log(e_n) = a + \frac{b}{n} + clog(n)$ |
| WBL4 | $e_n = c - bexp(-an^d)$ | $log(c - e_n) = log(b) - an^d$ |
| ILOG2 | $e_n = c - \frac{a}{log(n)}$ | $e_n = c - \frac{a}{log(n)}$ |
| EXPD3 | $e_n = c - (c - a)exp(-bn)$ | $log(c - e_n) = log(c - a) - bn$ |

- **Classification Tree** (TREE): The *rpart* package in R software with default settings is used.
- **Random Forest** (RF): The *randomForest* package in R software with $ntree = 500$ and $nodesize = 8$ is used.
- **Multilayer Perceptron** (MLP): The *nnet* package in R software with a initial random weights on $[-0.5, 0.5]$ and 25 units in the hidden layer is used.

The four real-world datasets include Cat VS Dog dataset, Cifar10 dataset, COVID-19 Chest X-Ray dataset, and binary gender classification dataset. For Cat VS Dog dataset, COVID-19 Chest X-Ray dataset, and binary gender classification dataset, we use random-sized crop to $224 \times 224$ and random horizontal flipping. For Cifar10 dataset, we pad by 4 pixels and use a random $32 \times 32$ crop and random horizontal flipping. On these real-world datasets, two deep neural networks without pretrained, i.e., ResNet18 and DenseNet are used, where the default Pytorch implementations of networks, an optimizer of Stochastic gradient descent (SGD), a learning rate of 0.005, and a training epoch size of 200 are used.

## Appendix B. Results and Analysis

For the experimental results, a fraction of typical results of the comparisons between POW3 baseline and the proposed curve lines on the simple dataset, Cifar10 dataset and COVID-19 Chest X-Ray dataset have been listed in the Table 2,3 and 4 of the submitted paper. The remaining results of the comparisons between POW3 baseline and the proposed curve lines on the two synthetic datasets are showed in Table 7-19. In Table 7, the RMSE values of the POW3 baseline and our proposed method are compared for the metrics of test error, precision, recall, and $F_1$ score and the confidence bands of test error are evaluated on the simple dataset. From this Table, we can see that most values of RMSE of our proposed learning curves are lower than those in POW3 baseline, which illustrates that our proposed learning curves have a benefit in the fitting of learning curves. From the Table 11, 12 and 19, the similar results can be obtained from the different datasets and classifiers.

In Table 13, 14, 15 and 16, the comparison results of confidence bands of POW3 baseline and our proposed method are compared for the metrics of precision, recall, and $F_1$ score.

From these tables, we can see that our proposed confidence bands frequently possess DoC values higher than the desired degree $1-\alpha = 0.95$. It is illustrated that the confidence bands of the POW3 baseline are liberal and the proposed confidence bands are more reasonable than the baseline methods.

Table 7: Comparison of RMSE values of learning curves on the simple dataset.

| Classifier | POW3 | | | | | | Our proposed method | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 40 | 80 | 160 | 320 | 640 | 1280 | 40 | 80 | 160 | 320 | 640 | 1280 |
| | Test error | | | | | | | | | | | |
| LR | 3.26 | 2.48 | 2.59 | 2.72 | 2.94 | 3.83 | **2.88** | **2.37** | **2.48** | **2.67** | **2.93** | **3.34** |
| RF | 2.80 | 2.18 | 2.20 | 2.28 | 2.61 | 3.55 | **2.61** | **1.95** | **2.07** | **2.25** | **2.48** | **2.76** |
| | Precision | | | | | | | | | | | |
| LR | 4.25 | 3.01 | 2.97 | 3.15 | 3.52 | 4.38 | **3.24** | **2.78** | **2.88** | **3.12** | **3.42** | **3.79** |
| RF | 3.56 | 2.68 | 2.46 | 2.68 | 3.14 | 4.13 | **2.92** | **2.26** | **2.36** | **2.65** | **2.92** | **3.32** |
| | Recall | | | | | | | | | | | |
| LR | 5.20 | 3.78 | 3.89 | 4.13 | 4.56 | 5.82 | **4.61** | **3.59** | **3.78** | **4.06** | **4.41** | **4.87** |
| RF | 4.34 | 3.31 | 3.17 | 3.09 | 3.52 | 4.83 | **4.18** | **3.01** | **2.96** | **3.03** | **3.39** | **3.80** |
| | $F_1$ score | | | | | | | | | | | |
| LR | 3.69 | 2.81 | 2.87 | 2.98 | 3.20 | 4.13 | **3.36** | **2.63** | **2.73** | **2.91** | **3.17** | **3.60** |
| RF | 2.91 | 2.28 | 2.20 | 2.21 | 2.50 | 3.33 | **2.76** | **1.97** | **2.05** | **2.17** | **2.39** | **2.64** |

Table 8: Comparison of confidence bands of test error on the simple dataset.

| Measure | 40 | 80 | 160 | 320 | 640 | 1280 | 40 | 80 | 160 | 320 | 640 | 1280 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | LR | | | | | | | | | | | |
| DoC | 0.21 | 0.17 | 0.13 | 0.16 | 0.18 | 0.16 | **1.00** | **1.00** | **1.00** | **1.00** | **0.95** | 0.78 |
| IL($\times 10^{-1}$) | 0.15 | 0.10 | 0.08 | 0.10 | 0.13 | 0.15 | 4.53 | 3.25 | 2.32 | 1.60 | 1.14 | 0.81 |
| | RF | | | | | | | | | | | |
| DoC | 0.21 | 0.21 | 0.14 | 0.15 | 0.20 | 0.16 | **1.00** | **1.00** | **1.00** | **1.00** | **0.97** | 0.82 |
| IL($\times 10^{-1}$) | 0.14 | 0.10 | 0.08 | 0.09 | 0.13 | 0.15 | 4.24 | 3.11 | 2.15 | 1.48 | 1.04 | 0.75 |

Besides the POW3 baseline, we also compare our method with other nine types of parametric learning curves, namely POW2, LOG2, EXP2, EXP3, LIN2, VAP3, WBL4, ILOG2, and EXPD3 baselines. The specific formulae of these nine learning curves have been summarized in Table 6. For the fitting of learning curves, similar to the POW3 baseline, a logarithmic transformation of parametric formula as shown in the Table 6 and a weighted least squares method are used to fit these parametric learning curves. The comparison results of RMSE values of learning curves are showed in Table 20, 21, and 22. The bond values of RMSE represent the smallest RMSE values among the nine baseline learning curves

Table 9: Comparison of learning curves and confidence bands on Cifar10 dataset and DenseNet Classifier.

| Metric | POW3 | | | | | | Our proposed method | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_i(\times 10)$ | 125 | 250 | 500 | 1000 | 2000 | 4000 | 125 | 250 | 500 | 1000 | 2000 | 4000 |
| Test error | **0.99** | 1.89 | 1.15 | **1.93** | **2.57** | 4.02 | 1.18 | **1.63** | **1.10** | 2.13 | 2.66 | **2.29** |
| Precision | 2.02 | 2.00 | 0.92 | 2.35 | 3.11 | 4.32 | **1.12** | **1.71** | **0.86** | **2.24** | **2.90** | **2.37** |
| Recall | **1.80** | 2.19 | 2.15 | 2.13 | **3.25** | 3.30 | 1.82 | **1.61** | **1.74** | **2.09** | 3.42 | **3.13** |
| $F_1$ score | 1.32 | 1.88 | 1.35 | **2.04** | **2.60** | 3.42 | **1.32** | **1.60** | **1.22** | 2.09 | 2.68 | **2.31** |
| | Evaluation of the confidence bands of test error. | | | | | | | | | | | |
| DoC | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | **1.0** | **1.0** | **1.0** | **1.0** | 0.5 | 0.0 |
| IL($\times 10^{-1}$) | 0.03 | 0.02 | 0.02 | 0.03 | 0.04 | 0.04 | 2.70 | 1.79 | 1.15 | 0.70 | 0.42 | 0.25 |
| | Evaluation of the confidence bands of precision. | | | | | | | | | | | |
| DoC | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | **1.0** | **1.0** | **1.0** | **1.0** | 0.5 | 0.3 |
| IL($\times 10^{-1}$) | 0.03 | 0.02 | 0.02 | 0.03 | 0.04 | 0.04 | 2.75 | 1.82 | 1.18 | 0.72 | 0.44 | 0.27 |
| | Evaluation of the confidence bands of recall. | | | | | | | | | | | |
| DoC | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | **1.0** | **1.0** | **1.0** | **1.0** | 0.4 | 0.3 |
| IL($\times 10^{-1}$) | 0.03 | 0.02 | 0.02 | 0.03 | 0.04 | 0.04 | 2.75 | 1.82 | 1.17 | 0.69 | 0.40 | 0.23 |
| | Evaluation of the confidence bands of $F_1$ score. | | | | | | | | | | | |
| DoC | 0.1 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | **1.0** | **1.0** | **1.0** | **1.0** | 0.5 | 0.0 |
| IL($\times 10^{-1}$) | 0.03 | 0.02 | 0.02 | 0.03 | 0.04 | 0.04 | 2.75 | 1.82 | 1.17 | 0.71 | 0.42 | 0.25 |

and our proposed curve for a specific training set size. Compared our proposed method with each baseline method, most values of RMSE of our proposed learning curve are smaller than those in each baseline learning curve, which illustrates the proposed learning curves are better than the other existing learning curves in most experimental settings.

Table 10: The comparison results on Gender dataset and ResNet Classifier.

| Metric | POW3 | | | | | Our proposed method | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $N_i$ | 125 | 250 | 500 | 1000 | 2000 | 125 | 250 | 500 | 1000 | 2000 |
| Test error | 5.73 | 2.64 | 1.26 | 2.34 | 4.83 | **5**.59 | 4.05 | **0**.95 | 2.77 | 5.59 |
| Precision | 3.37 | 2.51 | 2.75 | 3.41 | 6.32 | **2**.98 | 2.75 | **2**.18 | 4.38 | 7.91 |
| Recall | 10.71 | 6.05 | 3.04 | 2.02 | 5.35 | 11.44 | 7.01 | 3.43 | **1**.75 | **3**.32 |
| $F_1$ score | 7.08 | 3.17 | 1.40 | 2.19 | 5.06 | 7.14 | 4.75 | **1**.27 | 2.58 | 5.34 |
| Measure | Evaluation of the confidence bands of test error. | | | | | | | | | |
| DoC | 0.0 | 0.0 | 0.2 | 0.1 | 0.1 | **1**.0 | **1**.0 | **1**.0 | 0.8 | 0.0 |
| IL($\times 10^{-1}$) | 0.67 | 0.60 | 0.48 | 0.80 | 0.74 | 24.10 | 15.80 | 10.74 | 6.97 | 5.35 |

Table 11: Comparison of RMSE values of learning curves on the Exp2 dataset.

| Classifier | POW3 | | | | | | Our proposed method | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 40 | 80 | 160 | 320 | 640 | 1280 | 40 | 80 | 160 | 320 | 640 | 1280 |
| | Test error | | | | | | | | | | | |
| LR | 2.39 | 2.04 | 1.97 | 2.01 | 2.16 | 2.47 | **2.24** | **1.88** | **1.95** | **2.01** | **2.09** | **2.18** |
| LDA | 2.44 | 2.05 | 2.02 | 2.04 | 2.23 | 2.52 | **2.31** | **1.90** | **1.99** | **2.04** | **2.12** | **2.20** |
| NB | 2.64 | 2.24 | 2.17 | 2.23 | 2.46 | 2.86 | **2.42** | **2.05** | **2.14** | **2.23** | **2.32** | **2.41** |
| TREE | **2.79** | **2.16** | 1.73 | 1.72 | **1.73** | 2.94 | 3.19 | 2.53 | **1.63** | **1.68** | 1.74 | **1.91** |
| RF | **2.11** | 1.61 | 1.52 | 1.34 | 1.37 | **1.93** | 2.59 | **1.58** | **1.34** | **1.34** | **1.37** | 2.08 |
| MLP | **2.16** | **1.49** | **1.30** | 1.15 | **1.13** | **1.93** | 3.43 | 1.85 | 1.51 | **1.14** | 1.31 | 2.19 |
| | Precision | | | | | | | | | | | |
| LR | 3.41 | 2.82 | **2.66** | **2.71** | 2.91 | 3.40 | **3.18** | **2.78** | 2.74 | 2.80 | **2.85** | **2.96** |
| LDA | 3.28 | 2.75 | **2.67** | **2.69** | 2.90 | 3.35 | **3.18** | **2.68** | 2.74 | 2.78 | **2.85** | **2.96** |
| NB | 3.40 | 2.88 | **2.78** | **2.84** | 3.09 | 3.67 | **3.25** | **2.85** | 2.90 | 2.95 | **3.02** | **3.15** |
| TREE | 4.45 | **3.07** | **2.31** | **2.22** | 2.57 | 3.07 | **4.19** | 3.61 | 2.37 | 2.25 | **2.36** | **2.58** |
| RF | **2.70** | 2.16 | 1.89 | 1.78 | **1.86** | **2.43** | 3.64 | **1.89** | **1.77** | **1.78** | 1.87 | 2.46 |
| MLP | **2.80** | 2.06 | 1.73 | 1.51 | **1.53** | **2.30** | 4.63 | **1.96** | **1.71** | **1.50** | 1.73 | 2.36 |
| | Recall | | | | | | | | | | | |
| LR | 3.53 | 2.95 | 2.78 | 2.83 | 3.07 | 3.66 | **3.24** | **2.64** | **2.73** | **2.83** | **2.93** | **3.13** |
| LDA | 3.50 | 2.95 | 2.79 | **2.89** | 3.17 | 3.67 | **3.26** | **2.68** | **2.76** | 2.90 | **3.01** | **3.12** |
| NB | 3.82 | 3.19 | 3.00 | 3.12 | 3.47 | 4.06 | **3.50** | **2.89** | **2.95** | **3.12** | **3.26** | **3.39** |
| TREE | **5.66** | **5.35** | 2.81 | **2.89** | **2.98** | 5.33 | 6.89 | 6.34 | **2.78** | 3.13 | 3.44 | **3.69** |
| RF | **3.35** | **2.43** | 2.16 | 1.90 | 1.87 | 2.71 | 4.02 | 2.50 | **1.93** | **1.90** | **1.86** | **2.57** |
| MLP | **3.27** | **2.29** | **1.93** | 1.58 | **1.66** | **2.65** | 4.79 | 2.81 | 2.20 | **1.57** | 2.02 | 2.80 |
| | $F_1$ score | | | | | | | | | | | |
| LR | 2.42 | 2.08 | 2.01 | **2.04** | 2.18 | 2.46 | **2.31** | **1.95** | **2.00** | 2.06 | **2.11** | **2.20** |
| LDA | 2.46 | 2.10 | 2.06 | **2.09** | 2.26 | 2.55 | **2.38** | **1.98** | **2.06** | 2.11 | **2.17** | **2.25** |
| NB | 2.71 | 2.31 | **2.23** | **2.30** | 2.50 | 2.87 | **2.55** | **2.17** | 2.24 | 2.33 | **2.40** | **2.47** |
| TREE | **3.14** | **2.41** | 1.81 | **1.79** | **1.76** | 3.14 | 3.84 | 2.75 | **1.70** | 1.83 | 1.88 | **2.02** |
| RF | **2.21** | 1.65 | 1.54 | 1.34 | **1.35** | **1.87** | 2.80 | **1.57** | **1.34** | **1.32** | 1.38 | 2.13 |
| MLP | **2.21** | **1.53** | **1.31** | 1.14 | **1.11** | **1.87** | 3.67 | 1.80 | 1.46 | **1.13** | 1.33 | 2.23 |

Table 12: Comparison of confidence bands of test error on the Exp2 dataset.

| Measure | 40 | 80 | 160 | 320 | 640 | 1280 | 40 | 80 | 160 | 320 | 640 | 1280 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | LR | | | | | | |
| DoC | 0.15 | 0.13 | 0.10 | 0.12 | 0.17 | 0.17 | **1.00** | **1.00** | **1.00** | **1.00** | **0.95** | 0.81 |
| IL($\times 10^{-1}$) | 0.10 | 0.07 | 0.05 | 0.06 | 0.09 | 0.11 | 3.21 | 2.19 | 1.48 | 1.03 | 0.79 | 0.52 |
| | | | | | | LDA | | | | | | |
| DoC | 0.15 | 0.15 | 0.10 | 0.13 | 0.14 | 0.19 | **1.00** | **1.00** | **1.00** | **1.00** | 0.94 | 0.81 |
| IL($\times 10^{-1}$) | 0.10 | 0.07 | 0.05 | 0.06 | 0.09 | 0.11 | 3.21 | 2.19 | 1.48 | 1.03 | 0.81 | 0.55 |
| | | | | | | NB | | | | | | |
| DoC | 0.17 | 0.14 | 0.10 | 0.13 | 0.16 | 0.18 | **1.00** | **1.00** | **1.00** | **1.00** | 0.94 | 0.79 |
| IL($\times 10^{-1}$) | 0.10 | 0.08 | 0.06 | 0.07 | 0.10 | 0.12 | 3.21 | 2.50 | 1.61 | 1.17 | 0.88 | 0.59 |
| | | | | | | TREE | | | | | | |
| DoC | 0.19 | 0.16 | 0.18 | 0.19 | 0.27 | 0.18 | **1.00** | **1.00** | **1.00** | **0.99** | 0.91 | 0.75 |
| IL($\times 10^{-1}$) | 0.13 | 0.08 | 0.07 | 0.08 | 0.11 | 0.13 | 3.82 | 2.19 | 1.48 | 0.97 | 0.63 | 0.39 |
| | | | | | | RF | | | | | | |
| DoC | 0.17 | 0.17 | 0.15 | 0.21 | 0.26 | 0.22 | **1.00** | **1.00** | **1.00** | **1.00** | 0.92 | 0.48 |
| IL($\times 10^{-1}$) | 0.10 | 0.07 | 0.06 | 0.07 | 0.10 | 0.11 | 3.82 | 2.19 | 1.33 | 0.78 | 0.47 | 0.32 |
| | | | | | | MLP | | | | | | |
| DoC | 0.18 | 0.21 | 0.18 | 0.26 | 0.35 | 0.27 | **1.00** | **1.00** | **1.00** | **1.00** | 0.87 | 0.32 |
| IL($\times 10^{-1}$) | 0.11 | 0.08 | 0.06 | 0.07 | 0.10 | 0.13 | 3.21 | 1.81 | 1.16 | 0.60 | 0.40 | 0.35 |

Table 13: Comparison of confidence bands of precision on the simple dataset.

| Measure | 40 | 80 | 160 | 320 | 640 | 1280 | 40 | 80 | 160 | 320 | 640 | 1280 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | LR | | | | | | |
| DoC | 0.17 | 0.17 | 0.15 | 0.15 | 0.18 | 0.21 | **1.00** | **1.00** | **1.00** | **0.98** | 0.92 | 0.71 |
| IL($\times 10^{-1}$) | 0.18 | 0.14 | 0.10 | 0.12 | 0.17 | 0.20 | 4.72 | 3.37 | 2.32 | 1.62 | 1.15 | 0.83 |
| | | | | | | LDA | | | | | | |
| DoC | 0.16 | 0.18 | 0.12 | 0.15 | 0.19 | 0.20 | **1.00** | **1.00** | **1.00** | **0.98** | 0.91 | 0.71 |
| IL($\times 10^{-1}$) | 0.18 | 0.14 | 0.10 | 0.12 | 0.17 | 0.20 | 4.72 | 3.37 | 2.32 | 1.58 | 1.10 | 0.81 |
| | | | | | | NB | | | | | | |
| DoC | 0.17 | 0.16 | 0.14 | 0.15 | 0.18 | 0.16 | **1.00** | **1.00** | **0.99** | **0.97** | 0.86 | 0.69 |
| IL($\times 10^{-1}$) | 0.16 | 0.10 | 0.09 | 0.10 | 0.14 | 0.16 | 4.24 | 3.11 | 2.08 | 1.36 | 0.95 | 0.66 |
| | | | | | | TREE | | | | | | |
| DoC | 0.22 | 0.22 | 0.19 | 0.22 | 0.21 | 0.16 | **1.00** | **1.00** | **1.00** | **1.00** | 0.86 | 0.70 |
| IL($\times 10^{-1}$) | 0.19 | 0.18 | 0.12 | 0.14 | 0.20 | 0.27 | 4.81 | 3.46 | 2.45 | 1.69 | 1.19 | 0.87 |
| | | | | | | RF | | | | | | |
| DoC | 0.21 | 0.20 | 0.15 | 0.16 | 0.20 | 0.16 | **1.00** | **1.00** | **1.00** | **0.99** | 0.93 | 0.75 |
| IL($\times 10^{-1}$) | 0.16 | 0.13 | 0.09 | 0.11 | 0.15 | 0.19 | 4.53 | 3.25 | 2.27 | 1.58 | 1.09 | 0.79 |
| | | | | | | MLP | | | | | | |
| DoC | 0.18 | 0.21 | 0.17 | 0.19 | 0.21 | 0.18 | **1.00** | **1.00** | **1.00** | **1.00** | **0.95** | 0.73 |
| IL($\times 10^{-1}$) | 0.19 | 0.15 | 0.10 | 0.13 | 0.17 | 0.21 | 4.72 | 3.37 | 2.41 | 1.68 | 1.18 | 0.82 |
| | | | | | | LR | | | | | | |
| DoC | 0.19 | 0.15 | 0.13 | 0.15 | 0.17 | 0.16 | **1.00** | **1.00** | **1.00** | **0.96** | 0.83 | 0.62 |
| IL($\times 10^{-1}$) | 0.22 | 0.15 | 0.12 | 0.15 | 0.20 | 0.23 | 4.53 | 3.37 | 2.37 | 1.62 | 1.13 | 0.77 |
| | | | | | | LDA | | | | | | |
| DoC | 0.18 | 0.14 | 0.12 | 0.16 | 0.15 | 0.15 | **1.00** | **1.00** | **1.00** | **0.96** | 0.84 | 0.63 |
| IL($\times 10^{-1}$) | 0.22 | 0.15 | 0.12 | 0.14 | 0.19 | 0.23 | 4.72 | 3.46 | 2.41 | 1.66 | 1.17 | 0.82 |

Table 14: Comparison of confidence bands of recall on the simple dataset.

| Measure | 40 | 80 | 160 | 320 | 640 | 1280 | 40 | 80 | 160 | 320 | 640 | 1280 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | LR | | | | | | |
| DoC | 0.19 | 0.15 | 0.13 | 0.15 | 0.17 | 0.16 | **1.00** | **1.00** | **1.00** | **0.96** | 0.83 | 0.62 |
| IL($\times 10^{-1}$) | 0.22 | 0.15 | 0.12 | 0.15 | 0.20 | 0.23 | 4.53 | 3.37 | 2.37 | 1.62 | 1.13 | 0.77 |
| | | | | | | LDA | | | | | | |
| DoC | 0.18 | 0.14 | 0.12 | 0.16 | 0.15 | 0.15 | **1.00** | **1.00** | **1.00** | **0.96** | 0.84 | 0.63 |
| IL($\times 10^{-1}$) | 0.22 | 0.15 | 0.12 | 0.14 | 0.19 | 0.23 | 4.72 | 3.46 | 2.41 | 1.66 | 1.17 | 0.82 |
| | | | | | | NB | | | | | | |
| DoC | 0.16 | 0.15 | 0.09 | 0.13 | 0.16 | 0.15 | **1.00** | **1.00** | **1.00** | **0.95** | 0.85 | 0.64 |
| IL($\times 10^{-1}$) | 0.18 | 0.14 | 0.10 | 0.12 | 0.16 | 0.20 | 4.24 | 3.11 | 2.21 | 1.51 | 1.05 | 0.74 |
| | | | | | | TREE | | | | | | |
| DoC | 0.18 | 0.10 | 0.14 | 0.16 | 0.16 | 0.14 | **1.00** | **1.00** | **1.00** | **0.95** | 0.76 | 0.49 |
| IL($\times 10^{-1}$) | 0.26 | 0.19 | 0.15 | 0.18 | 0.24 | 0.27 | 4.72 | 3.52 | 2.45 | 1.68 | 1.15 | 0.88 |
| | | | | | | RF | | | | | | |
| DoC | 0.18 | 0.17 | 0.14 | 0.19 | 0.22 | 0.18 | **1.00** | **1.00** | **1.00** | **0.97** | 0.86 | 0.65 |
| IL($\times 10^{-1}$) | 0.20 | 0.14 | 0.12 | 0.14 | 0.19 | 0.22 | 4.24 | 3.11 | 2.08 | 1.36 | 0.93 | 0.67 |
| | | | | | | MLP | | | | | | |
| DoC | 0.15 | 0.18 | 0.16 | 0.16 | 0.18 | 0.18 | **1.00** | **1.00** | **1.00** | **0.99** | 0.82 | 0.45 |
| IL($\times 10^{-1}$) | 0.25 | 0.18 | 0.13 | 0.14 | 0.20 | 0.29 | 4.81 | 3.52 | 2.48 | 1.71 | 1.24 | 0.86 |

Table 15: Comparison of confidence bands of $F_1$ score on the simple dataset.

| Measure | 40 | 80 | 160 | 320 | 640 | 1280 | 40 | 80 | 160 | 320 | 640 | 1280 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | LR | | | | | |
| DoC | 0.20 | 0.15 | 0.13 | 0.17 | 0.18 | 0.19 | **1.00** | **1.00** | **1.00** | **0.99** | **0.95** | 0.77 |
| IL($\times 10^{-1}$) | 0.17 | 0.11 | 0.10 | 0.12 | 0.16 | 0.18 | 4.72 | 3.37 | 2.37 | 1.62 | 1.14 | 0.81 |
| | | | | | | | LDA | | | | | |
| DoC | 0.20 | 0.16 | 0.12 | 0.15 | 0.18 | 0.18 | **1.00** | **1.00** | **1.00** | **0.99** | 0.94 | 0.77 |
| IL($\times 10^{-1}$) | 0.17 | 0.12 | 0.10 | 0.12 | 0.16 | 0.18 | 4.72 | 3.37 | 2.37 | 1.62 | 1.13 | 0.81 |
| | | | | | | | NB | | | | | |
| DoC | 0.20 | 0.16 | 0.13 | 0.13 | 0.18 | 0.16 | **1.00** | **1.00** | **1.00** | **0.99** | **0.95** | 0.78 |
| IL($\times 10^{-1}$) | 0.14 | 0.09 | 0.08 | 0.09 | 0.12 | 0.14 | 4.24 | 3.11 | 2.15 | 1.45 | 0.99 | 0.71 |
| | | | | | | | TREE | | | | | |
| DoC | 0.20 | 0.19 | 0.17 | 0.21 | 0.24 | 0.17 | **1.00** | **1.00** | **1.00** | **1.00** | 0.94 | 0.73 |
| IL($\times 10^{-1}$) | 0.19 | 0.15 | 0.12 | 0.14 | 0.19 | 0.22 | 4.81 | 3.52 | 2.45 | 1.69 | 1.18 | 0.87 |
| | | | | | | | RF | | | | | |
| DoC | 0.21 | 0.18 | 0.16 | 0.18 | 0.24 | 0.19 | **1.00** | **1.00** | **1.00** | **1.00** | **0.97** | 0.83 |
| IL($\times 10^{-1}$) | 0.15 | 0.11 | 0.09 | 0.11 | 0.15 | 0.17 | 4.53 | 3.25 | 2.21 | 1.48 | 1.03 | 0.74 |
| | | | | | | | MLP | | | | | |
| DoC | 0.20 | 0.19 | 0.19 | 0.17 | 0.23 | 0.19 | **1.00** | **1.00** | **1.00** | **1.00** | **0.96** | 0.67 |
| IL($\times 10^{-1}$) | 0.20 | 0.14 | 0.10 | 0.12 | 0.17 | 0.22 | 4.72 | 3.46 | 2.45 | 1.69 | 1.21 | 0.84 |

Table 16: Comparison of confidence bands of precision on the Exp2 dataset.

| Measure | 40 | 80 | 160 | 320 | 640 | 1280 | 40 | 80 | 160 | 320 | 640 | 1280 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | LR | | | | | | |
| DoC | 0.12 | 0.11 | 0.09 | 0.11 | 0.16 | 0.16 | **1.00** | **0.99** | **0.97** | 0.92 | 0.82 | 0.65 |
| IL($\times 10^{-1}$) | 0.12 | 0.09 | 0.06 | 0.08 | 0.11 | 0.13 | 3.82 | 2.50 | 1.61 | 1.08 | 0.82 | 0.54 |
| | | | | | | LDA | | | | | | |
| DoC | 0.11 | 0.11 | 0.08 | 0.10 | 0.13 | 0.15 | **1.00** | **0.99** | **0.97** | 0.92 | 0.80 | 0.65 |
| IL($\times 10^{-1}$) | 0.12 | 0.10 | 0.07 | 0.08 | 0.11 | 0.14 | 3.82 | 2.50 | 1.61 | 1.08 | 0.84 | 0.56 |
| | | | | | | NB | | | | | | |
| DoC | 0.12 | 0.13 | 0.09 | 0.10 | 0.14 | 0.14 | **1.00** | **0.99** | **0.98** | 0.94 | 0.81 | 0.63 |
| IL($\times 10^{-1}$) | 0.12 | 0.10 | 0.07 | 0.08 | 0.11 | 0.14 | 3.82 | 2.50 | 1.73 | 1.21 | 0.91 | 0.57 |
| | | | | | | TREE | | | | | | |
| DoC | 0.14 | 0.11 | 0.13 | 0.18 | 0.21 | 0.19 | **0.99** | **0.94** | **0.95** | 0.92 | 0.78 | 0.59 |
| IL($\times 10^{-1}$) | 0.15 | 0.09 | 0.09 | 0.11 | 0.14 | 0.14 | 4.24 | 2.50 | 1.48 | 0.91 | 0.56 | 0.42 |
| | | | | | | RF | | | | | | |
| DoC | 0.16 | 0.18 | 0.13 | 0.18 | 0.25 | 0.20 | **0.97** | **0.99** | **0.98** | 0.94 | 0.74 | 0.40 |
| IL($\times 10^{-1}$) | 0.12 | 0.09 | 0.07 | 0.08 | 0.11 | 0.13 | 3.82 | 2.19 | 1.33 | 0.85 | 0.58 | 0.43 |
| | | | | | | MLP | | | | | | |
| DoC | 0.19 | 0.18 | 0.16 | 0.22 | 0.34 | 0.23 | **0.97** | **1.00** | **1.00** | 0.92 | 0.63 | 0.25 |
| IL($\times 10^{-1}$) | 0.13 | 0.10 | 0.07 | 0.09 | 0.12 | 0.14 | 3.82 | 2.19 | 1.33 | 0.78 | 0.50 | 0.35 |

Table 17: Comparison of confidence bands of recall on the Exp2 dataset.

| Measure | 40 | 80 | 160 | 320 | 640 | 1280 | 40 | 80 | 160 | 320 | 640 | 1280 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | LR | | | | | |
| DoC | 0.15 | 0.15 | 0.12 | 0.12 | 0.17 | 0.15 | **0.99** | **1.00** | **0.98** | 0.93 | 0.82 | 0.64 |
| IL($\times 10^{-1}$) | 0.14 | 0.10 | 0.07 | 0.09 | 0.12 | 0.15 | 3.21 | 2.50 | 1.61 | 1.08 | 0.77 | 0.52 |
| | | | | | | | LDA | | | | | |
| DoC | 0.16 | 0.15 | 0.09 | 0.12 | 0.16 | 0.16 | **0.99** | **0.99** | **0.99** | 0.94 | 0.82 | 0.64 |
| IL($\times 10^{-1}$) | 0.14 | 0.11 | 0.07 | 0.09 | 0.12 | 0.15 | 3.82 | 2.50 | 1.61 | 1.13 | 0.81 | 0.55 |
| | | | | | | | NB | | | | | |
| DoC | 0.13 | 0.13 | 0.09 | 0.12 | 0.15 | 0.16 | **1.00** | **1.00** | **0.99** | **0.95** | 0.82 | 0.65 |
| IL($\times 10^{-1}$) | 0.15 | 0.12 | 0.08 | 0.09 | 0.13 | 0.17 | 3.82 | 2.74 | 1.83 | 1.25 | 0.88 | 0.62 |
| | | | | | | | TREE | | | | | |
| DoC | 0.13 | 0.11 | 0.13 | 0.14 | 0.24 | 0.17 | **0.96** | **1.00** | **0.98** | 0.77 | 0.60 | 0.44 |
| IL($\times 10^{-1}$) | 0.22 | 0.14 | 0.10 | 0.11 | 0.16 | 0.24 | 4.53 | 2.74 | 1.73 | 1.17 | 0.82 | 0.52 |
| | | | | | | | RF | | | | | |
| DoC | 0.19 | 0.21 | 0.15 | 0.21 | 0.30 | 0.24 | **0.99** | **1.00** | **0.98** | **0.95** | 0.77 | 0.39 |
| IL($\times 10^{-1}$) | 0.16 | 0.11 | 0.08 | 0.10 | 0.14 | 0.16 | 4.24 | 2.50 | 1.33 | 0.78 | 0.58 | 0.38 |
| | | | | | | | MLP | | | | | |
| DoC | 0.19 | 0.20 | 0.16 | 0.24 | 0.33 | 0.24 | **0.96** | **1.00** | **0.99** | 0.94 | 0.59 | 0.22 |
| IL($\times 10^{-1}$) | 0.14 | 0.11 | 0.08 | 0.09 | 0.13 | 0.17 | 3.82 | 2.19 | 1.16 | 0.78 | 0.50 | 0.37 |

Table 18: Comparison of confidence bands of $F_1$ score on the Exp2 dataset.

| Measure | 40 | 80 | 160 | 320 | 640 | 1280 | 40 | 80 | 160 | 320 | 640 | 1280 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | LR | | | | | | |
| DoC | 0.15 | 0.14 | 0.10 | 0.13 | 0.17 | 0.19 | **1.00** | **1.00** | **1.00** | **0.99** | 0.93 | 0.79 |
| IL($\times 10^{-1}$) | 0.10 | 0.07 | 0.05 | 0.06 | 0.09 | 0.11 | 3.82 | 2.50 | 1.61 | 1.08 | 0.79 | 0.52 |
| | | | | | | LDA | | | | | | |
| DoC | 0.17 | 0.16 | 0.11 | 0.13 | 0.16 | 0.17 | **1.00** | **1.00** | **1.00** | **0.99** | 0.93 | 0.80 |
| IL($\times 10^{-1}$) | 0.10 | 0.07 | 0.05 | 0.07 | 0.09 | 0.11 | 3.82 | 2.50 | 1.61 | 1.13 | 0.82 | 0.56 |
| | | | | | | NB | | | | | | |
| DoC | 0.16 | 0.14 | 0.12 | 0.12 | 0.16 | 0.18 | **1.00** | **1.00** | **1.00** | **0.98** | 0.92 | 0.76 |
| IL($\times 10^{-1}$) | 0.11 | 0.08 | 0.06 | 0.07 | 0.10 | 0.12 | 3.82 | 2.74 | 1.73 | 1.21 | 0.90 | 0.60 |
| | | | | | | TREE | | | | | | |
| DoC | 0.18 | 0.13 | 0.17 | 0.17 | 0.28 | 0.19 | **1.00** | **1.00** | **1.00** | **0.96** | 0.86 | 0.70 |
| IL($\times 10^{-1}$) | 0.14 | 0.09 | 0.07 | 0.08 | 0.12 | 0.14 | 4.24 | 2.50 | 1.61 | 1.03 | 0.65 | 0.40 |
| | | | | | | RF | | | | | | |
| DoC | 0.19 | 0.18 | 0.16 | 0.23 | 0.28 | 0.21 | **1.00** | **1.00** | **1.00** | **0.99** | 0.87 | 0.40 |
| IL($\times 10^{-1}$) | 0.10 | 0.08 | 0.06 | 0.07 | 0.10 | 0.11 | 4.24 | 2.50 | 1.48 | 0.85 | 0.50 | 0.33 |
| | | | | | | MLP | | | | | | |
| DoC | 0.18 | 0.21 | 0.16 | 0.26 | 0.36 | 0.28 | **0.99** | **1.00** | **1.00** | **0.97** | 0.79 | 0.26 |
| IL($\times 10^{-1}$) | 0.11 | 0.08 | 0.06 | 0.07 | 0.10 | 0.12 | 3.82 | 2.19 | 1.33 | 0.69 | 0.43 | 0.36 |

Table 19: Comparison of learning curves and confidence bands on the Cat VS Dog dataset.

| Metric | Classifier | POW3 | | | | | Our proposed method | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 625 | 1250 | 2500 | 5000 | 10000 | 625 | 1250 | 2500 | 5000 | 10000 |
| | | RMSE | | | | | | | | | |
| Test error | ResNet18 | **1.03** | 3.52 | 2.69 | 1.02 | **2.83** | 1.10 | **3.23** | **2.24** | **0.93** | 5.64 |
| | DenseNet | 2.13 | **2.86** | **1.74** | **0.89** | 2.46 | **1.40** | 3.07 | 2.24 | 1.02 | **2.30** |
| Precision | ResNet18 | **0.96** | 2.61 | 2.39 | 1.66 | **4.86** | 1.34 | **2.20** | **1.82** | **1.63** | 7.36 |
| | DenseNet | 1.69 | 3.16 | **1.43** | **1.17** | 2.41 | **1.29** | **3.16** | 1.64 | 1.23 | **2.23** |
| Recall | ResNet18 | 4.15 | 7.12 | 4.08 | 2.14 | **4.64** | **3.23** | **6.83** | **3.45** | **2.00** | 5.57 |
| | DenseNet | 2.79 | **3.12** | **3.26** | **1.49** | **2.70** | **1.96** | 3.33 | 3.71 | 1.60 | 2.86 |
| $F_1$ score | ResNet18 | 2.00 | 4.69 | 2.97 | 1.10 | **2.42** | **1.77** | **4.54** | **2.59** | **0.98** | 5.53 |
| | DenseNet | 2.40 | **2.88** | **1.92** | **0.89** | 2.56 | **1.54** | 3.13 | 2.45 | 1.05 | **2.33** |
| | | Evaluation of the confidence bands on the ResNet18. | | | | | | | | | |
| Test error | DoC | 0.1 | 0.0 | 0.0 | 0.1 | 0.0 | **1.0** | **1.0** | 0.9 | 0.1 | 0.3 |
| | IL($\times 10^{-1}$) | 0.04 | 0.02 | 0.03 | 0.05 | 0.06 | 1.29 | 0.87 | 0.59 | 0.38 | 0.28 |
| Precision | DoC | 0.1 | 0.1 | 0.0 | 0.2 | 0.1 | **1.0** | **1.0** | 0.9 | 0.8 | 0.1 |
| | IL($\times 10^{-1}$) | 0.04 | 0.02 | 0.03 | 0.05 | 0.07 | 1.30 | 0.86 | 0.57 | 0.37 | 0.27 |
| Recall | DoC | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | **1.0** | 0.2 | 0.6 | 0.7 | 0.2 |
| | IL($\times 10^{-1}$) | 0.06 | 0.03 | 0.03 | 0.06 | 0.08 | 1.30 | 0.89 | 0.61 | 0.39 | 0.28 |
| $F_1$ score | DoC | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | **1.0** | 0.6 | 0.8 | **1.0** | 0.3 |
| | IL($\times 10^{-1}$) | 0.05 | 0.02 | 0.03 | 0.05 | 0.07 | 1.30 | 0.88 | 0.59 | 0.38 | 0.28 |
| | | Evaluation of the confidence bands on the DenseNet. | | | | | | | | | |
| Test error | DoC | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | **1.0** | 0.9 | 0.6 | 0.9 | 0.1 |
| | IL($\times 10^{-1}$) | 0.04 | 0.02 | 0.03 | 0.05 | 0.06 | 1.14 | 0.72 | 0.44 | 0.28 | 0.18 |
| Precision | DoC | 0.0 | 0.0 | 0.0 | 0.2 | 0.3 | **1.0** | 0.7 | 0.9 | 0.8 | 0.5 |
| | IL($\times 10^{-1}$) | 0.04 | 0.02 | 0.03 | 0.05 | 0.06 | 1.14 | 0.70 | 0.44 | 0.27 | 0.18 |
| Recall | DoC | 0.1 | 0.1 | 0.1 | 0.0 | 0.1 | **1.0** | 0.6 | 0.4 | 0.7 | 0.0 |
| | IL($\times 10^{-1}$) | 0.04 | 0.03 | 0.03 | 0.05 | 0.07 | 1.17 | 0.75 | 0.45 | 0.28 | 0.17 |
| $F_1$ score | DoC | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | **1.0** | 0.9 | 0.5 | 0.9 | 0.1 |
| | IL($\times 10^{-1}$) | 0.04 | 0.02 | 0.03 | 0.05 | 0.06 | 1.16 | 0.73 | 0.45 | 0.28 | 0.18 |

Table 20: Comparison of RMSE values of learning curves between the other nine baselines and our proposed method on the simple dataset.

| Classifier | LR | | | | | | LDA | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method | 40 | 80 | 160 | 320 | 640 | 1280 | 40 | 80 | 160 | 320 | 640 | 1280 |
| POW2 | 5.70 | 3.52 | 2.55 | 3.78 | 7.12 | 11.54 | 5.97 | 3.25 | 2.53 | 3.76 | 6.94 | 11.19 |
| LOG2 | 3.20 | 3.01 | 3.03 | 2.74 | 3.09 | 6.03 | 2.99 | 2.74 | 2.87 | 2.71 | 3.03 | 5.41 |
| EXP2 | 5.15 | **2.32** | 2.95 | 3.04 | **2.82** | 9.76 | 4.72 | 2.32 | 2.75 | 2.86 | **2.79** | 8.88 |
| EXP3 | 4.60 | 2.51 | 3.72 | 3.57 | 2.91 | 15.30 | 4.19 | 2.40 | 3.41 | 3.36 | 2.90 | 13.79 |
| LIN2 | 4.32 | 2.78 | 4.16 | 3.86 | 2.92 | 19.55 | 3.92 | 2.60 | 3.80 | 3.62 | 2.91 | 17.49 |
| VAP3 | 4.26 | 2.80 | 2.70 | 2.69 | 2.93 | 4.54 | 4.14 | 2.78 | 2.69 | 2.67 | 2.94 | 4.48 |
| WBL4 | **2.81** | 2.59 | 2.55 | 2.67 | 3.02 | 3.70 | **2.73** | 2.44 | 2.51 | 2.65 | 2.95 | 3.45 |
| ILOG2 | 2.90 | 2.49 | 2.55 | 2.67 | 3.00 | 3.87 | 2.79 | 2.36 | 2.51 | 2.65 | 2.94 | 3.59 |
| EXPD3 | 4.23 | 2.93 | 4.35 | 3.99 | 2.92 | 21.49 | 3.84 | 2.71 | 3.97 | 3.73 | 2.91 | 19.14 |
| Ours | 2.88 | 2.37 | **2.48** | **2.67** | 2.93 | **3.34** | 2.80 | **2.29** | **2.47** | **2.65** | 2.88 | **3.19** |
| | NB | | | | | | TREE | | | | | |
| POW2 | 9.02 | 2.64 | 4.46 | 7.85 | 10.31 | 11.97 | 6.09 | 2.11 | 2.21 | 3.41 | 5.33 | 9.09 |
| LOG2 | 2.86 | 2.66 | 2.61 | 2.32 | 2.67 | 5.04 | **2.61** | 2.01 | 2.11 | 2.50 | 2.92 | 5.75 |
| EXP2 | 4.50 | **1.54** | **1.78** | **1.80** | **1.75** | 5.12 | 5.52 | 2.24 | 2.60 | 3.45 | 3.08 | 12.69 |
| EXP3 | 3.87 | 2.28 | 3.16 | 2.97 | 2.51 | 12.73 | 4.98 | 1.95 | 3.28 | 4.04 | 3.15 | 19.31 |
| LIN2 | 3.65 | 2.49 | 3.51 | 3.20 | 2.52 | 16.10 | 4.65 | 2.00 | 3.76 | 4.44 | 3.20 | 25.91 |
| VAP3 | 3.98 | 2.60 | 2.40 | 2.29 | 2.56 | 3.85 | 4.21 | 2.76 | 2.47 | 2.45 | 3.00 | 5.13 |
| WBL4 | **2.64** | 2.38 | 2.24 | 2.26 | 2.62 | 3.18 | 2.78 | **1.94** | 2.07 | 2.43 | 2.91 | **3.54** |
| ILOG2 | 2.70 | 2.26 | 2.23 | 2.25 | 2.59 | 3.27 | 2.66 | 1.99 | **2.06** | **2.42** | **2.89** | 3.58 |
| EXPD3 | 3.58 | 2.67 | 3.75 | 3.35 | 2.52 | 17.98 | 4.54 | 2.07 | 3.96 | 4.60 | 3.22 | 28.82 |
| Ours | 2.73 | 2.12 | 2.16 | 2.26 | 2.49 | **2.69** | 2.96 | 2.23 | 2.28 | 2.44 | 2.95 | 3.57 |
| | RF | | | | | | MLP | | | | | |
| POW2 | 8.71 | 2.55 | 2.72 | 5.36 | 8.90 | 11.59 | 5.51 | 2.43 | 1.97 | 2.96 | 5.45 | 8.38 |
| LOG2 | 2.79 | 2.45 | 2.47 | 2.31 | 2.69 | 5.16 | **2.60** | 2.05 | 2.10 | 2.23 | 2.69 | 4.16 |
| EXP2 | 4.99 | 2.01 | 2.30 | 2.47 | **2.32** | 7.94 | 4.25 | 2.26 | 2.13 | 2.40 | 2.58 | 8.20 |
| EXP3 | 4.36 | 2.11 | 3.29 | 3.31 | 2.59 | 15.01 | 3.75 | 2.05 | 2.58 | 2.84 | 2.65 | 12.16 |
| LIN2 | 4.09 | 2.31 | 3.71 | 3.62 | 2.61 | 19.54 | 3.47 | 2.06 | 2.94 | 3.12 | 2.67 | 15.60 |
| VAP3 | 4.04 | 2.65 | 2.40 | 2.26 | 2.60 | 4.11 | 4.26 | 2.77 | 2.30 | 2.20 | 2.59 | 3.51 |
| WBL4 | **2.50** | 2.06 | 2.07 | 2.25 | 2.57 | 3.01 | 2.61 | 1.90 | 1.92 | 2.19 | 2.59 | 3.51 |
| ILOG2 | 2.53 | 2.00 | 2.07 | **2.24** | 2.56 | 3.13 | 2.61 | **1.90** | **1.92** | 2.19 | 2.59 | **3.49** |
| EXPD3 | 4.01 | 2.45 | 3.92 | 3.76 | 2.62 | 21.59 | 3.38 | 2.11 | 3.10 | 3.23 | 2.67 | 17.16 |
| Our | 2.61 | **1.95** | **2.07** | 2.25 | 2.48 | **2.76** | 2.68 | 1.91 | 1.94 | **2.19** | **2.56** | 3.55 |

Table 21: Comparison of RMSE values of learning curves between the other nine baselines and our proposed method on the Cat VS Dog dataset.

| Classifier | ResNet18 | | | | | DenseNet | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Method | 625 | 1250 | 2500 | 5000 | 10000 | 625 | 1250 | 2500 | 5000 | 10000 |
| POW2 | 6.68 | 6.52 | 9.44 | 7.84 | 4.97 | **0.43** | 4.89 | 3.90 | 2.13 | 1.41 |
| LOG2 | 1.22 | 3.93 | 2.68 | 0.82 | 3.34 | 2.54 | 2.69 | 1.48 | 0.87 | 2.49 |
| EXP2 | 4.31 | 2.34 | **0.49** | 0.74 | 5.39 | 5.09 | 1.93 | **0.59** | 0.70 | 4.79 |
| EXP3 | 4.68 | 2.18 | 0.74 | 0.97 | 12.87 | 5.54 | 1.47 | 1.60 | 0.90 | 16.83 |
| LIN2 | 4.79 | 2.07 | 1.04 | 1.18 | 19.05 | 5.59 | 1.29 | 1.95 | 1.07 | 24.28 |
| VAP3 | 3.26 | **1.67** | 2.07 | 1.41 | **1.81** | 4.22 | 2.17 | 2.09 | 1.30 | **1.08** |
| WBL4 | 1.26 | 5.58 | 3.83 | 0.92 | 7.87 | 0.95 | 4.24 | 2.66 | **0.57** | 2.52 |
| ILOG2 | **0.85** | 4.93 | 3.45 | **0.72** | 6.12 | 1.58 | 3.62 | 2.21 | 0.72 | 1.13 |
| EXPD3 | 4.82 | 2.03 | 1.13 | 1.24 | 21.20 | 5.60 | **1.25** | 2.03 | 1.11 | 26.48 |
| Ours | 1.10 | 3.23 | 2.24 | 0.93 | 5.64 | 1.40 | 3.07 | 2.24 | 1.02 | 2.30 |

Table 22: Comparison of RMSE values of learning curves between the other nine baselines and our proposed method on the Cifar10 dataset.

| Classifier | ResNet18 | | | | | | DenseNet | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_i(\times 10)$ | 125 | 250 | 500 | 1000 | 2000 | 4000 | 125 | 250 | 500 | 1000 | 2000 | 4000 |
| POW2 | 10.07 | 4.28 | 8.03 | 4.73 | 2.32 | **1.16** | 1.21 | 3.72 | 3.54 | 3.97 | 3.65 | 2.14 |
| LOG2 | 2.71 | 1.54 | 2.60 | **0.47** | 1.93 | 1.36 | 3.43 | **0.71** | **0.71** | **1.18** | 3.05 | 7.24 |
| EXP2 | 4.09 | 1.01 | **1.00** | 3.57 | 1.85 | 7.07 | 9.24 | 1.91 | 2.86 | 2.26 | 1.81 | 6.64 |
| EXP3 | 4.87 | 1.00 | 2.11 | 5.70 | 2.71 | 25.04 | 9.51 | 1.25 | 4.46 | 4.36 | 1.30 | 26.82 |
| LIN2 | 4.98 | 0.84 | 2.69 | 6.64 | 3.16 | 44.09 | 9.45 | 0.86 | 5.09 | 5.11 | **1.25** | 45.21 |
| VAP3 | **1.75** | 3.04 | 1.67 | 0.98 | 1.73 | 1.19 | 2.30 | 1.24 | 1.22 | 2.41 | 2.75 | 1.61 |
| WBL4 | 6.29 | 4.47 | 5.43 | 0.78 | 4.27 | 8.37 | **0.30** | 3.08 | 2.19 | 2.31 | 1.67 | **0.90** |
| ILOG2 | 4.92 | 3.07 | 4.25 | 0.65 | 2.75 | 4.55 | 1.60 | 1.89 | 0.98 | 1.86 | 2.48 | 2.72 |
| EXPD3 | 5.00 | 0.78 | 2.85 | 6.89 | 3.29 | 51.67 | 9.43 | 0.75 | 5.26 | 5.30 | 1.25 | 51.77 |
| Our Method | 2.37 | **0.34** | 1.50 | 1.08 | **1.62** | 5.69 | 1.18 | 1.63 | 1.10 | 2.13 | 2.66 | 2.29 |