Iris Dataset Analysis

Wilson Borba da Rocha Neto
Prof. Fábio Rocha
Date: 10/08/2018

Introduction

This work presents the results obtained utilizing computational intelligence algorithms to classify the Iris Dataset. Three algorithms will be discussed, which are (i) Multi-Layer Perceptron, (ii) K-Nearest Neighbors and (iii) K-Means. To clarify what was done, a quickly summary of is presented in the section Discussion, which is followed by the Results section.

Discussion

The results are organized as following. First, figures containing plots of the maximum-, minimum-, mean- and standard deviation of the correct classification of each class (Setosa, Versicolor and Virginica) are presented. The horizontal axis of these figures indicates the variation of the test/train proportion, whilst its vertical axis represents the correct percentage (%) of the classification. These figures aim to illustrate how each algorithm performed regarding each class, which is important to perceive where can exist a high rate of misclassification between classes. Next, a confusion matrix for the mean classification success rate is presented, as a mean to provide an overall performance of the algorithm in question. Then, a table containing maximum-, minimum- and standard deviation of the classification is presented.

K-NN and K-Means

The K-NN and K-Means algorithm will be evaluated utilizing two metric's calculation method: Euclidean distance and Mahalanobis distance. Due to this, these algorithms will have one section for each method. Moreover, the number of Nearest Neighbors was variated, creating three scenarios where $K \in \{1,5,7\}$.

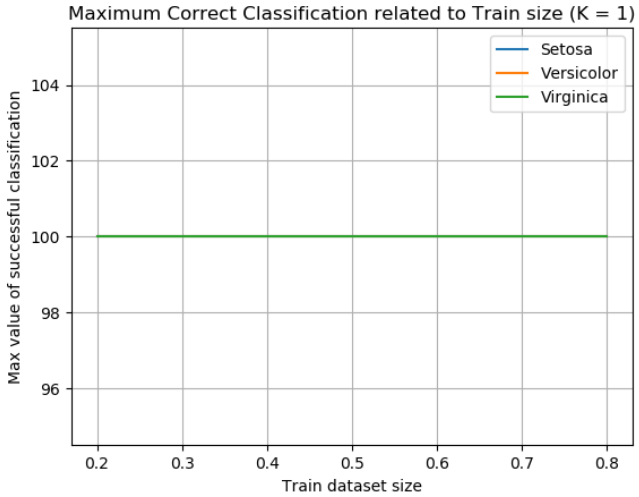
Multi-Layer Perceptron

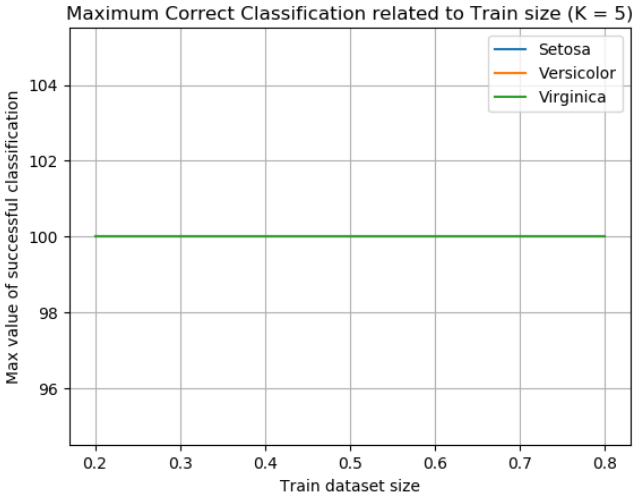
The number of neurons for the Input layer are equal to the number of parameters presented in our dataset, in this case four. The Output layer has three neurons, due to the use of the one-vs-all evaluation method. The Hidden layer size was calculated using classic methods, such as mean value and square root, resulting in four neurons. Moreover, the tests were conducted for 100, 1,000 and 10,000 epochs.

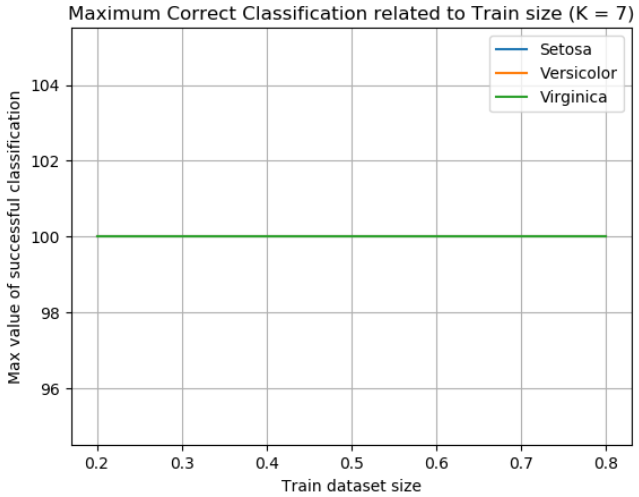
Results

K-NN (Euclidean Distance)

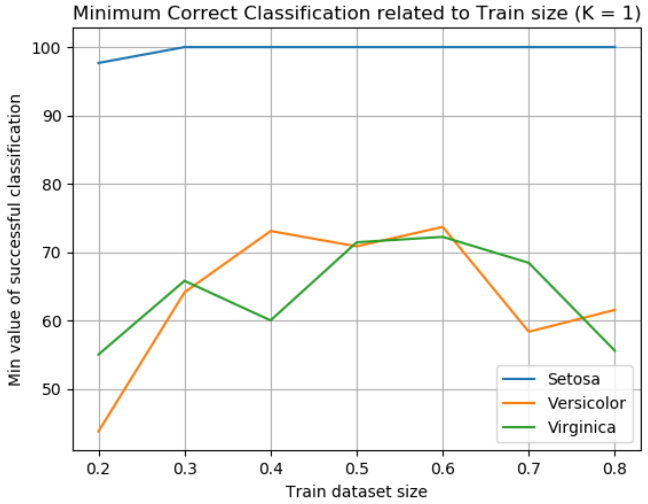
Maximum Success Rate

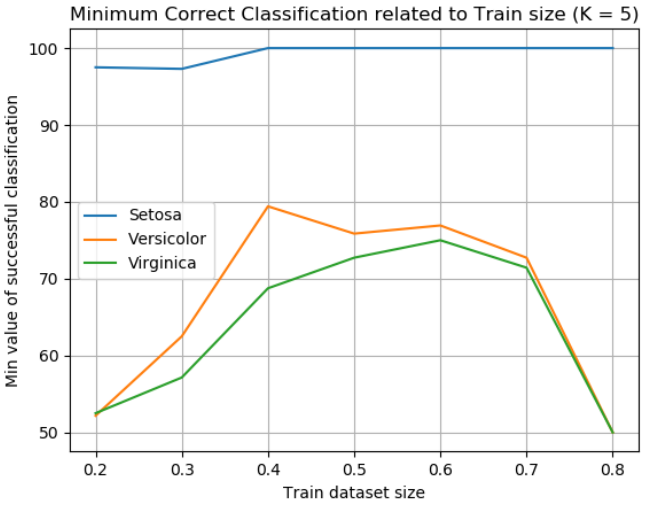


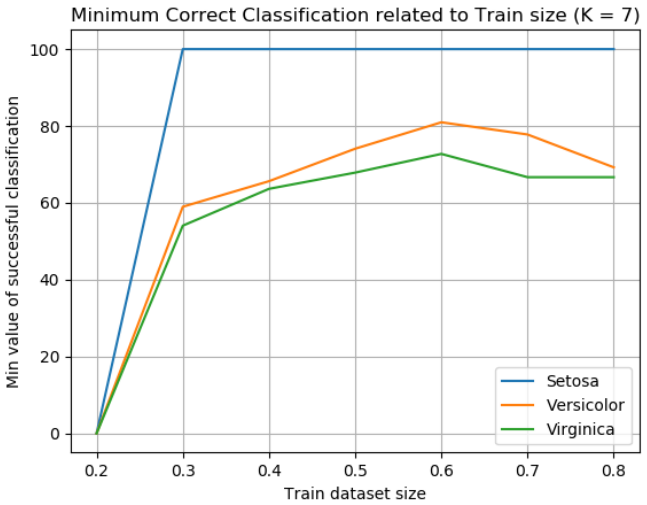




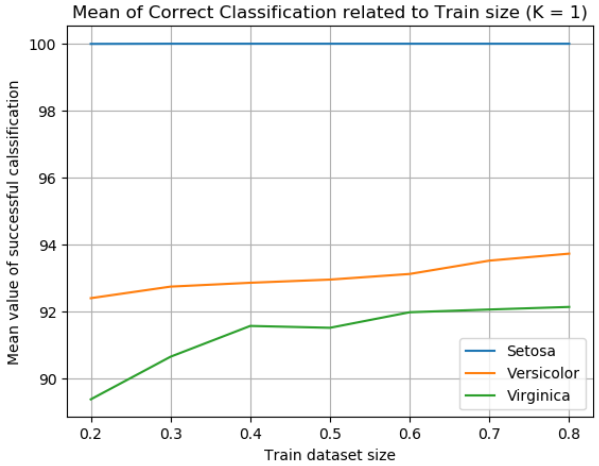
Minimum success rate

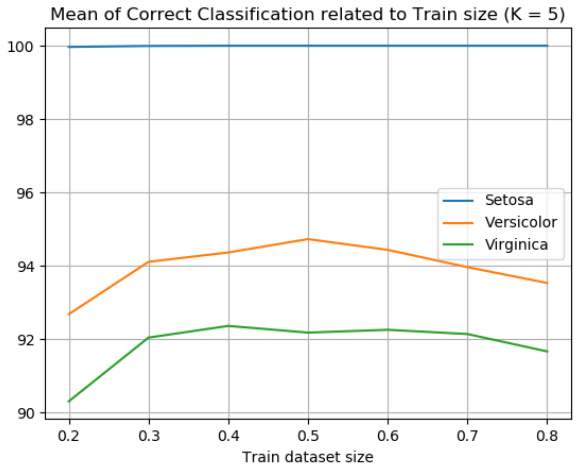


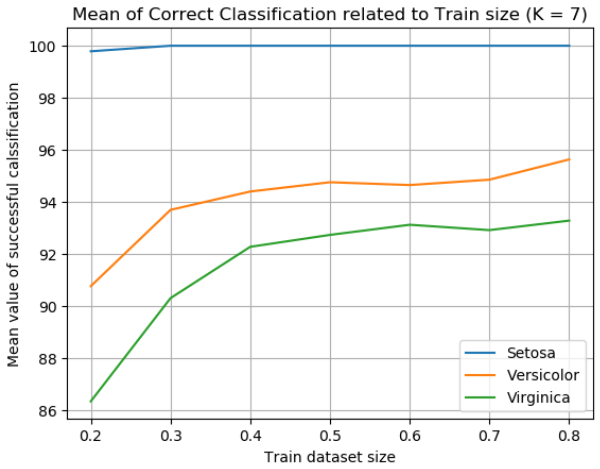




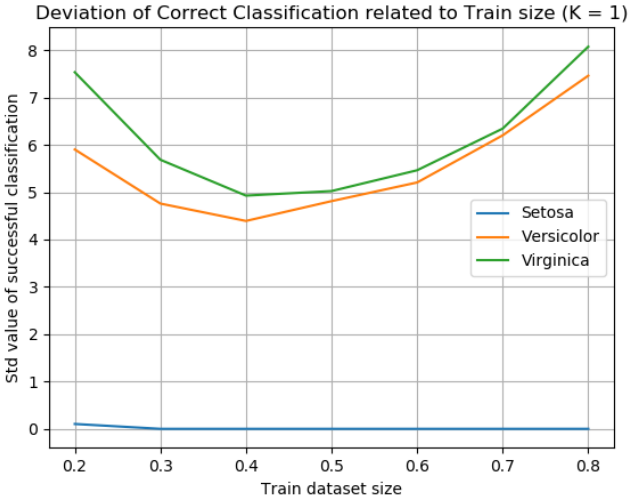
Mean of successful classification ratio

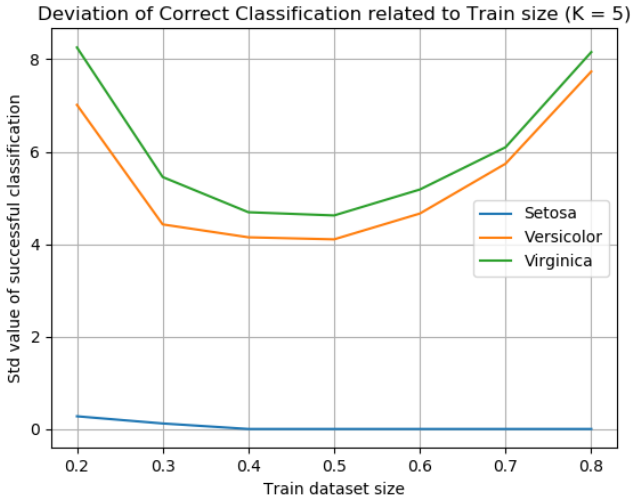






Standard deviation of successful classification ratio





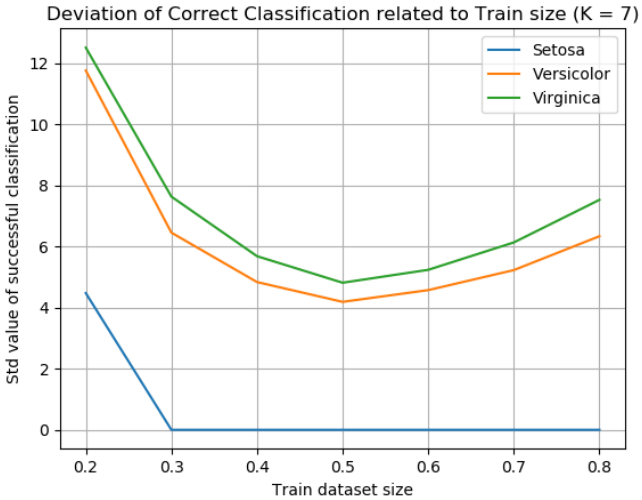


Table 1 – Confusion Matrix (K = 1)

Classes	*Setosa*	*Virginica*	*Versicolor*
Setosa	100	0	0
Virginica	0	92.5057	6.80011
Versicolor	0	7.49427	93.1999

Table 2 – Confusion Matrix (K = 5)

Classes	*Setosa*	*Virginica*	*Versicolor*
Setosa	100	0	0
Virginica	0	93.2477	7.38624
Versicolor	0	6.75234	92.6138

Table 3 – Confusion Matrix (K = 7)

Classes	*Setosa*	*Virginica*	*Versicolor*
Setosa	100	0	0
Virginica	0	95.0571	6.37779
Versicolor	0	4.94288	93.6222

Table 4 – Results (K = 1)

Algorithm	Maximum	Minimum	Standard Deviation
MLP	100	60	3.920

Table 5 – Results (K = 5)

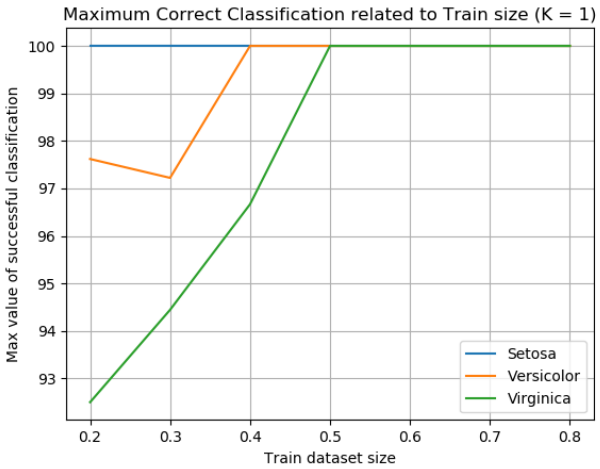
Algorithm	Maximum	Minimum	Standard Deviation
MLP	100	56.520	3.850

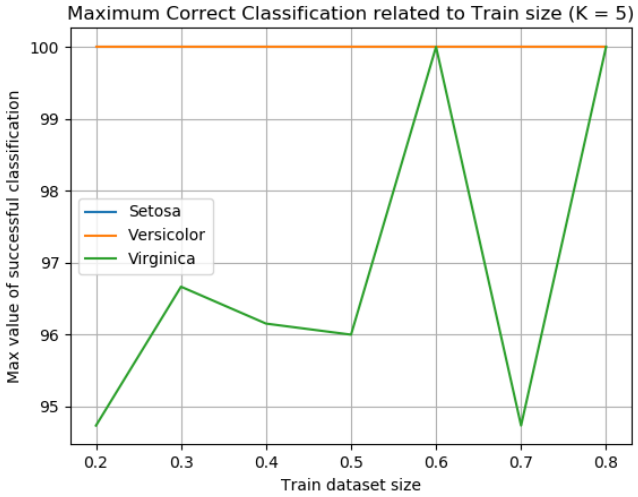
Table 6 – Results (K = 7)

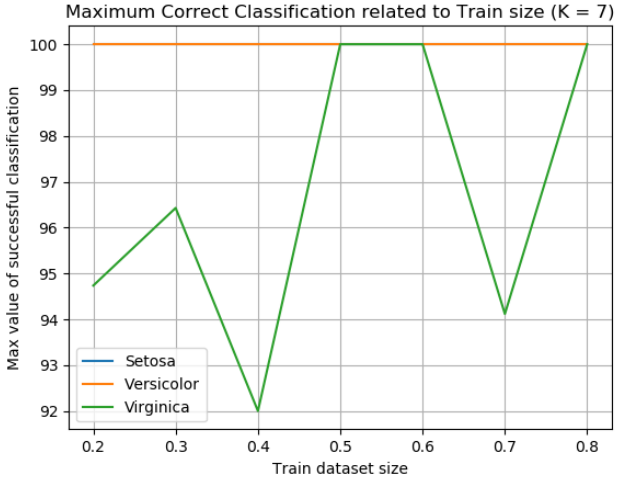
Algorithm	Maximum	Minimum	Standard Deviation
MLP	100	31.111	4.500

K-NN (Mahalanobis distance)

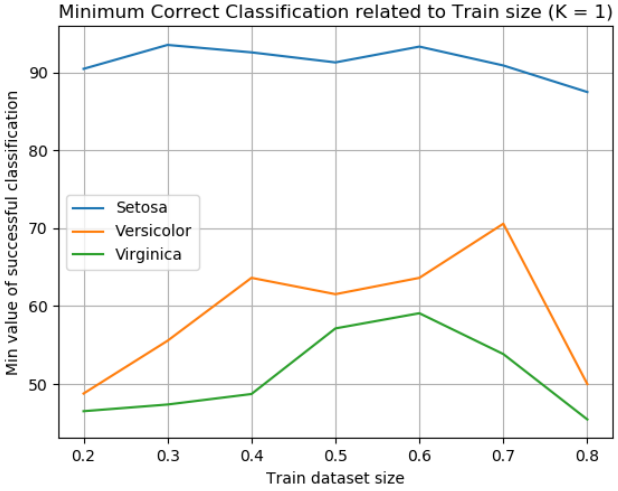
Maximum Success Rate

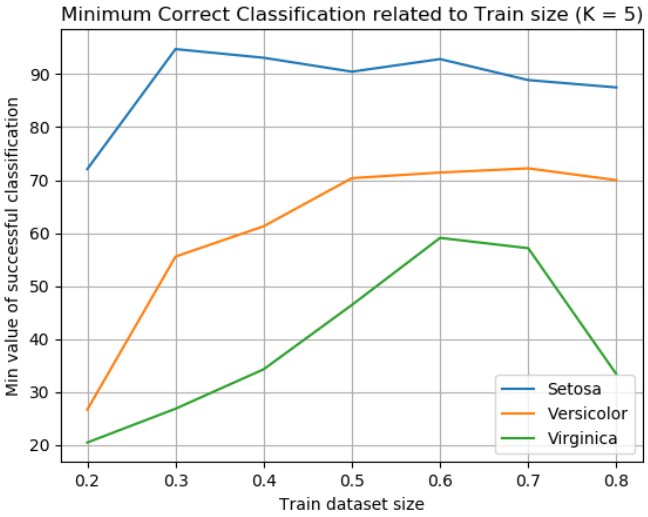


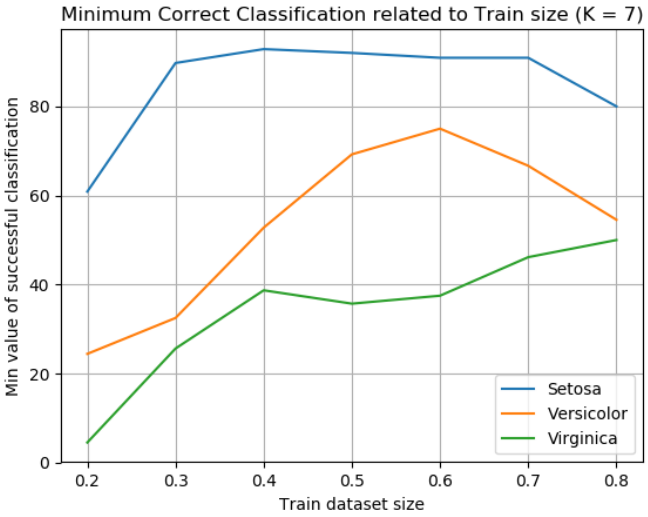




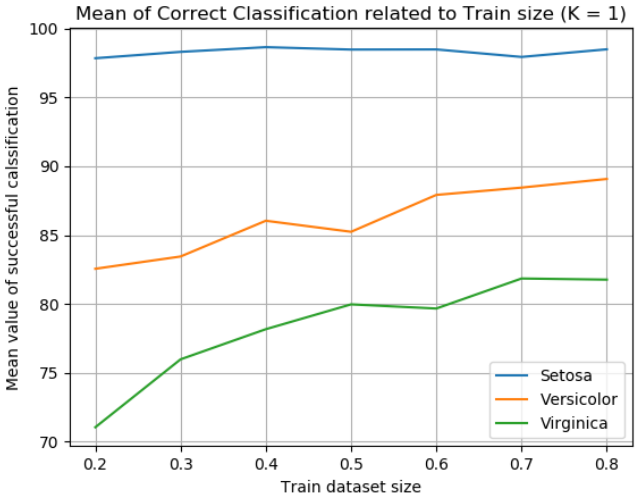
Minimum success rate

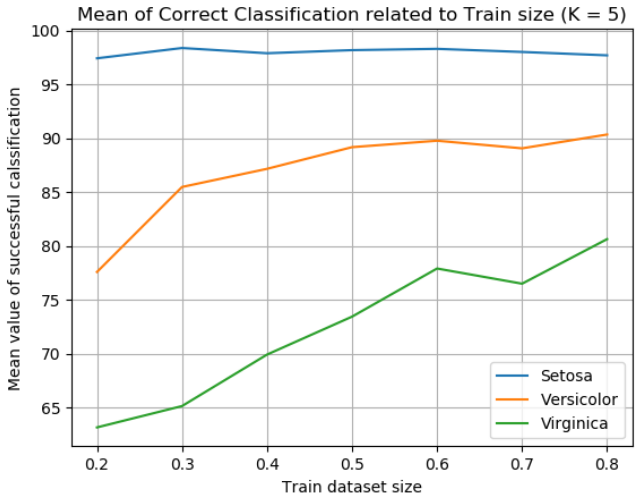


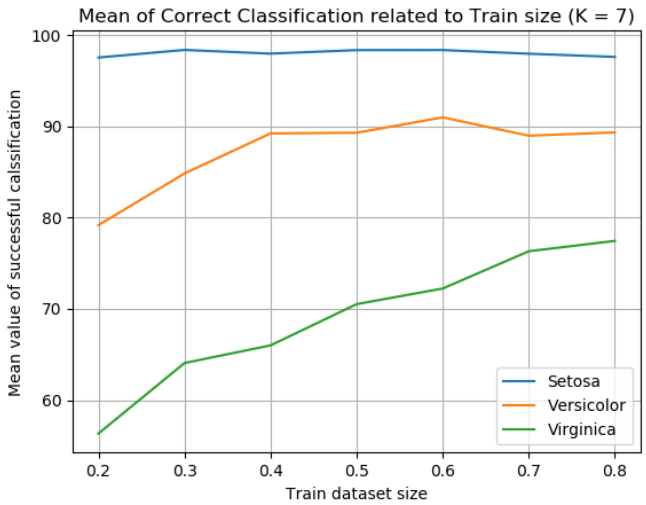




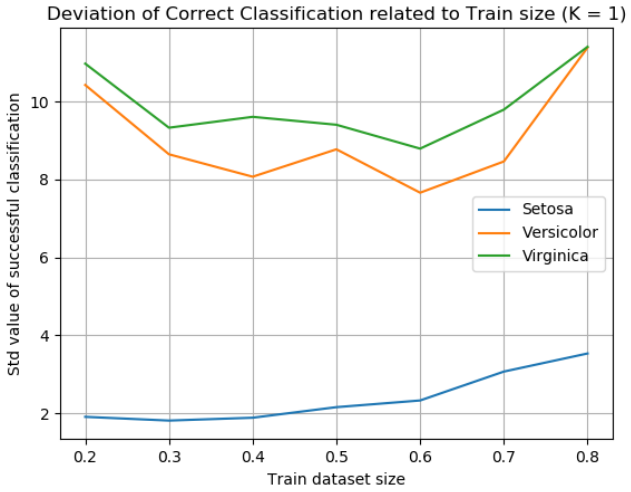
Mean of successful classification ratio

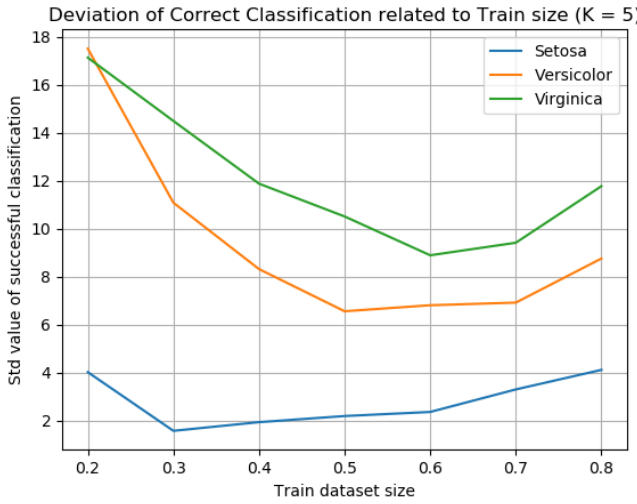






Standard deviation of successful classification ratio





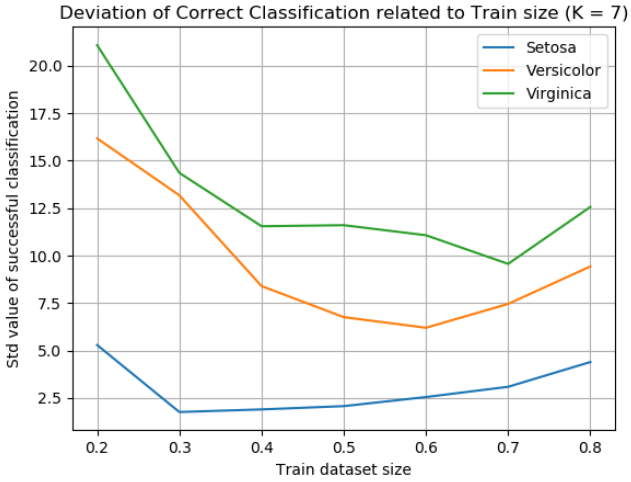


Table 1 – Confusion Matrix (K = 1)

Classes	*Setosa*	*Virginica*	*Versicolor*
Setosa	98.4809	0	0
Virginica	1.51909	89.0716	18.2333
Versicolor	0	10.9284	81.7667

Table 2 – Confusion Matrix (K = 5)

Classes	*Setosa*	*Virginica*	*Versicolor*
Setosa	97.7281	0	0
Virginica	2.27186	90.3757	19.3395
Versicolor	0	9.62433	80.6605

Table 3 – Confusion Matrix (K = 7)

Classes	*Setosa*	*Virginica*	*Versicolor*
Setosa	97.6152	0	0
Virginica	2.38475	89.3461	22.5445
Versicolor	0	10.6539	77.4555

Table 4 – Results (K = 1)

Algorithm	Maximum	Minimum	Standard Deviation
MLP	100	45.454	7.116

Table 5 – Results (K = 5)

Algorithm	Maximum	Minimum	Standard Deviation
MLP	100	20.454	8.07

Table 6 – Results (K = 7)

Algorithm	Maximum	Minimum	Standard Deviation
MLP	100	4.545	8.59

Multi-Layer Perceptron

Maximum success rate

Figure 1 - No. of Epochs = 100

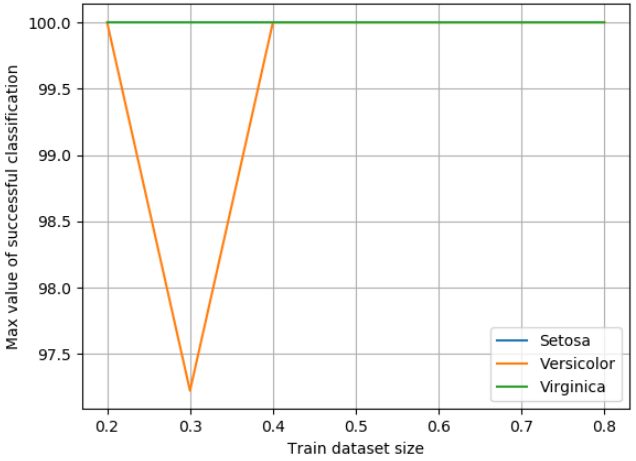


Figure 2 - No. of Epochs = 1000

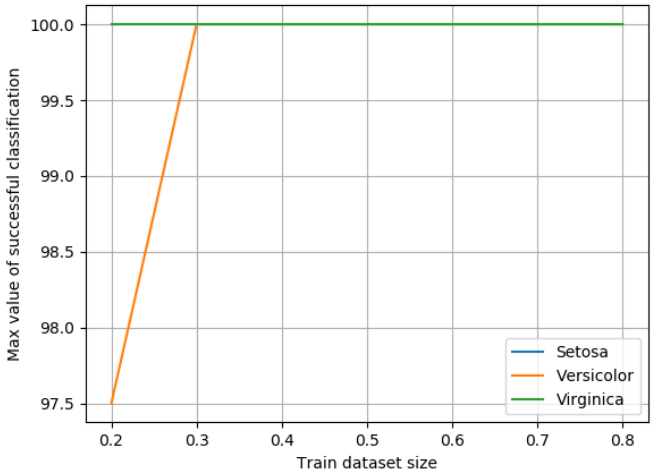
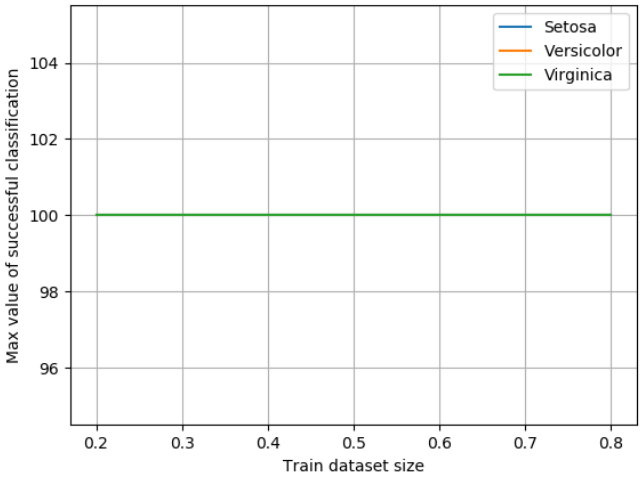


Figure 3 - No. of Epochs = 10000



Minimum success rate

Figure 4 - No. of Epochs = 100

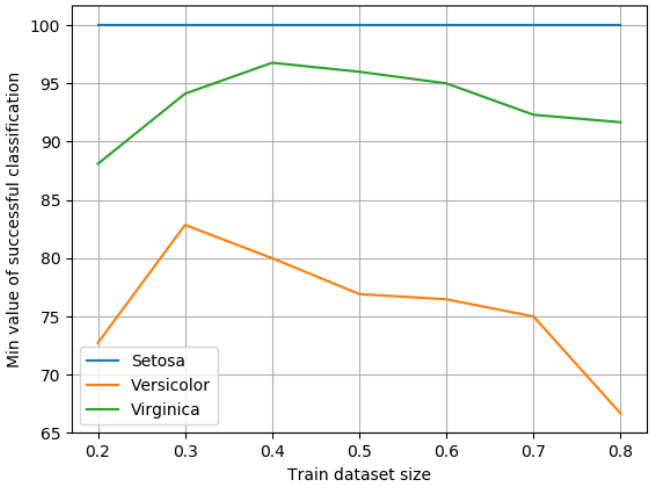


Figure 5 - No. of Epochs = 1000

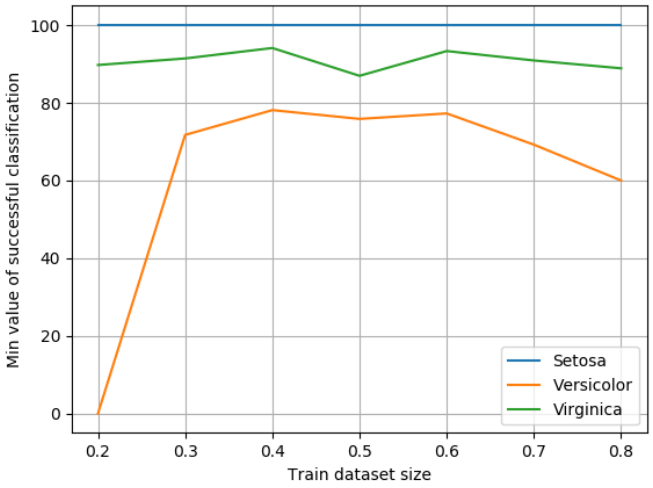
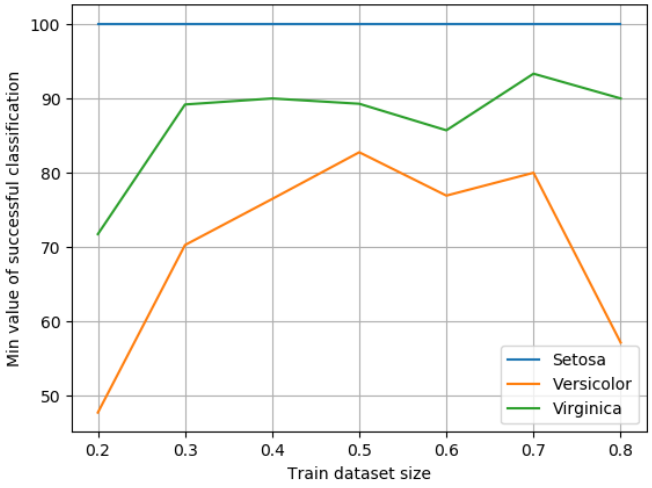


Figure 6 - No. of Epochs = 10000



Mean of successful classification ratio

Figure 7 - No. of Epochs = 100

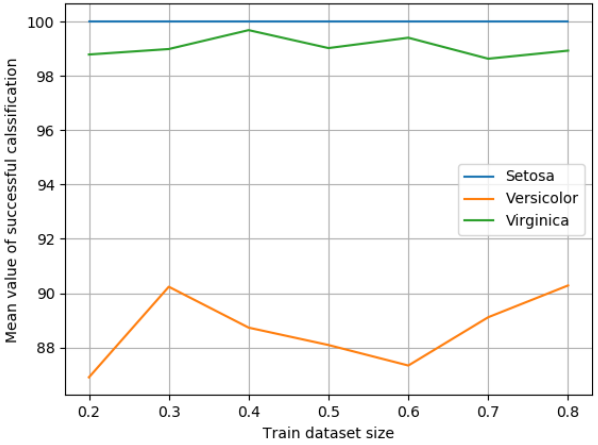


Figure 8 - No. of Epochs = 1000

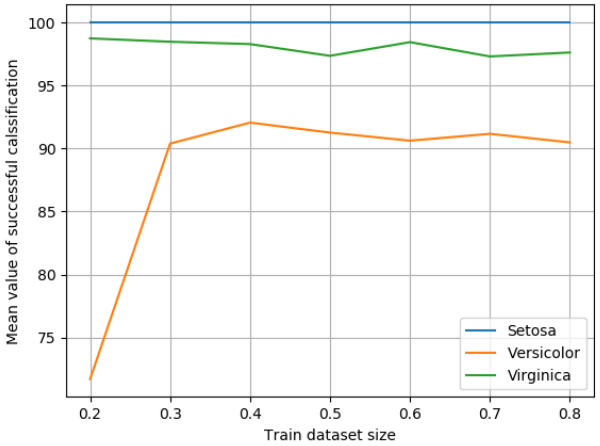
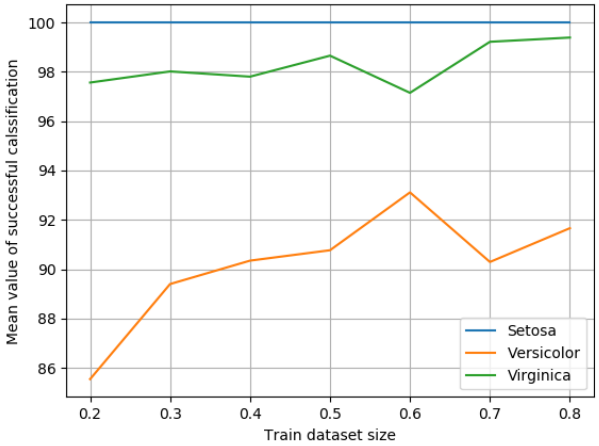


Figure 9 - No. of Epochs = 10000



Standard deviation of successful classification ratio

Figure 10 - No. of Epochs = 100

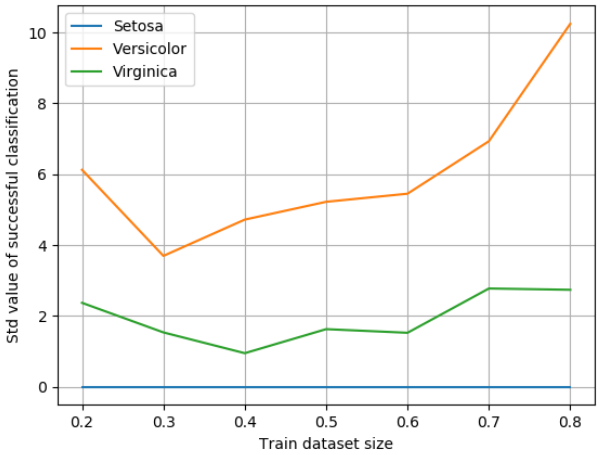


Figure 11 - No. of Epochs = 1000

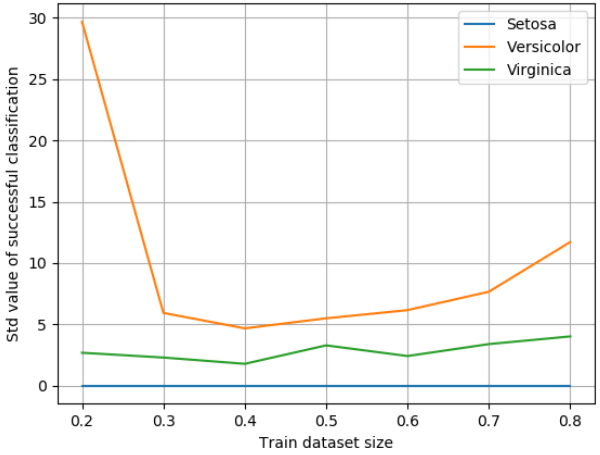


Figure 12 - No. of Epochs = 10000

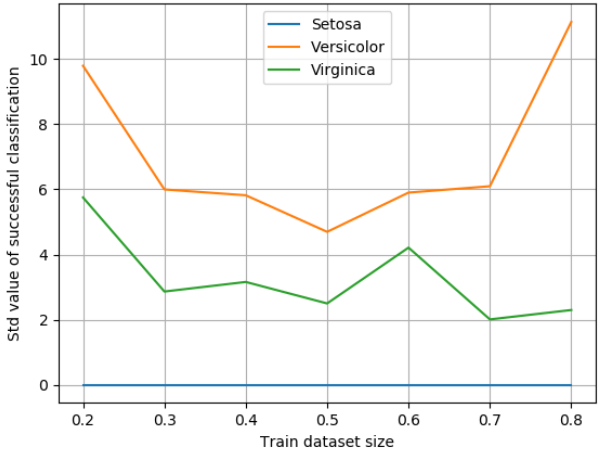


Table 1 – Confusion Matrix (100 epochs)

Classes	*Setosa*	*Virginica*	*Versicolor*
Setosa	100	0	0
Virginica	0	93.472	1.73945
Versicolor	0	6.528	98.2605

Table 2 – Confusion Matrix (1,000 epochs)

Classes	*Setosa*	*Virginica*	*Versicolor*
Setosa	100	0	0
Virginica	0	88.6385	1.18622
Versicolor	0	11.3615	98.8138

Table 3 – Confusion Matrix (10,000 epochs)

Classes	*Setosa*	*Virginica*	*Versicolor*
Setosa	100	2.80423	0
Virginica	0	89.6312	1.99375
Versicolor	0	7.56457	98.0063

Table 4 – Results (100 epochs)

Algorithm	Maximum	Minimum	Standard Deviation
MLP	100	0	5.657

Table 5 – Results (1,000 epochs)

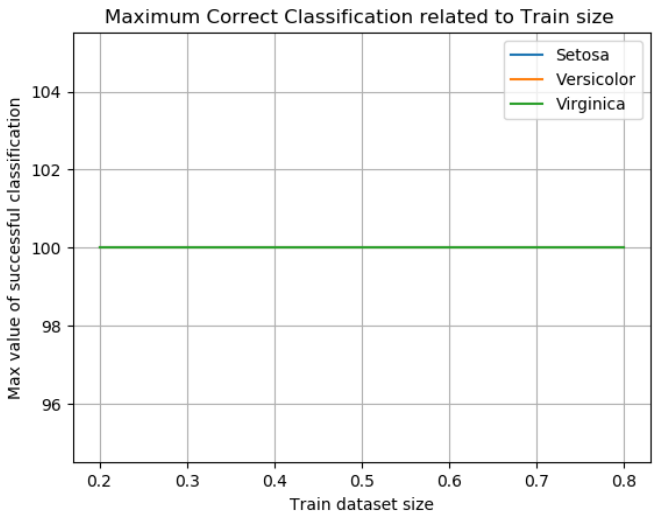
Algorithm	Maximum	Minimum	Standard Deviation
MLP	100	40	3.105

Table 6 – Results (10,000 epochs)

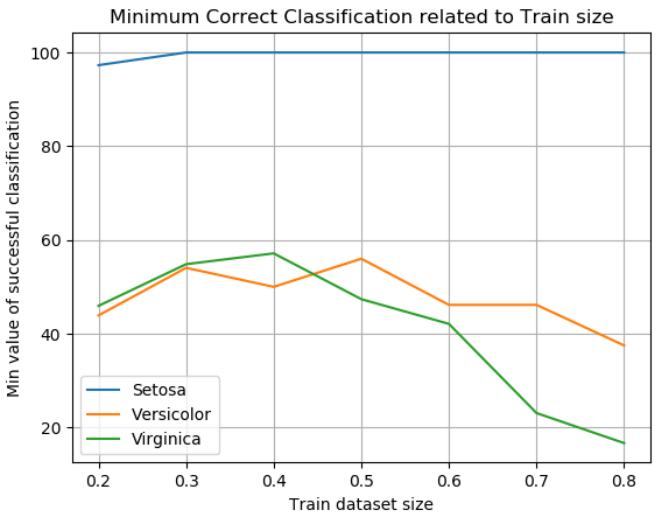
Algorithm	Maximum	Minimum	Standard Deviation
MLP	100	48.837	4.062

K-Means (Euclidean Distance)

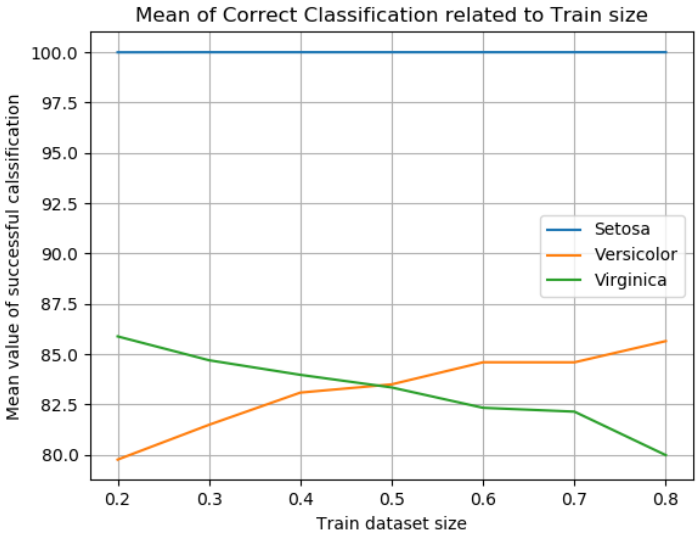
Maximum success rate



Minimum success rate



Mean of successful classification ratio



Standard deviation of successful classification ratio

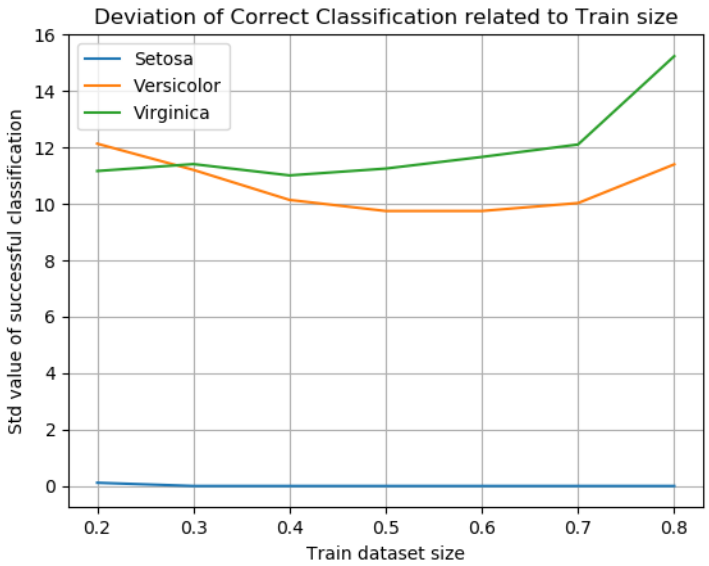


Figure 1 – Expected centroids and classification

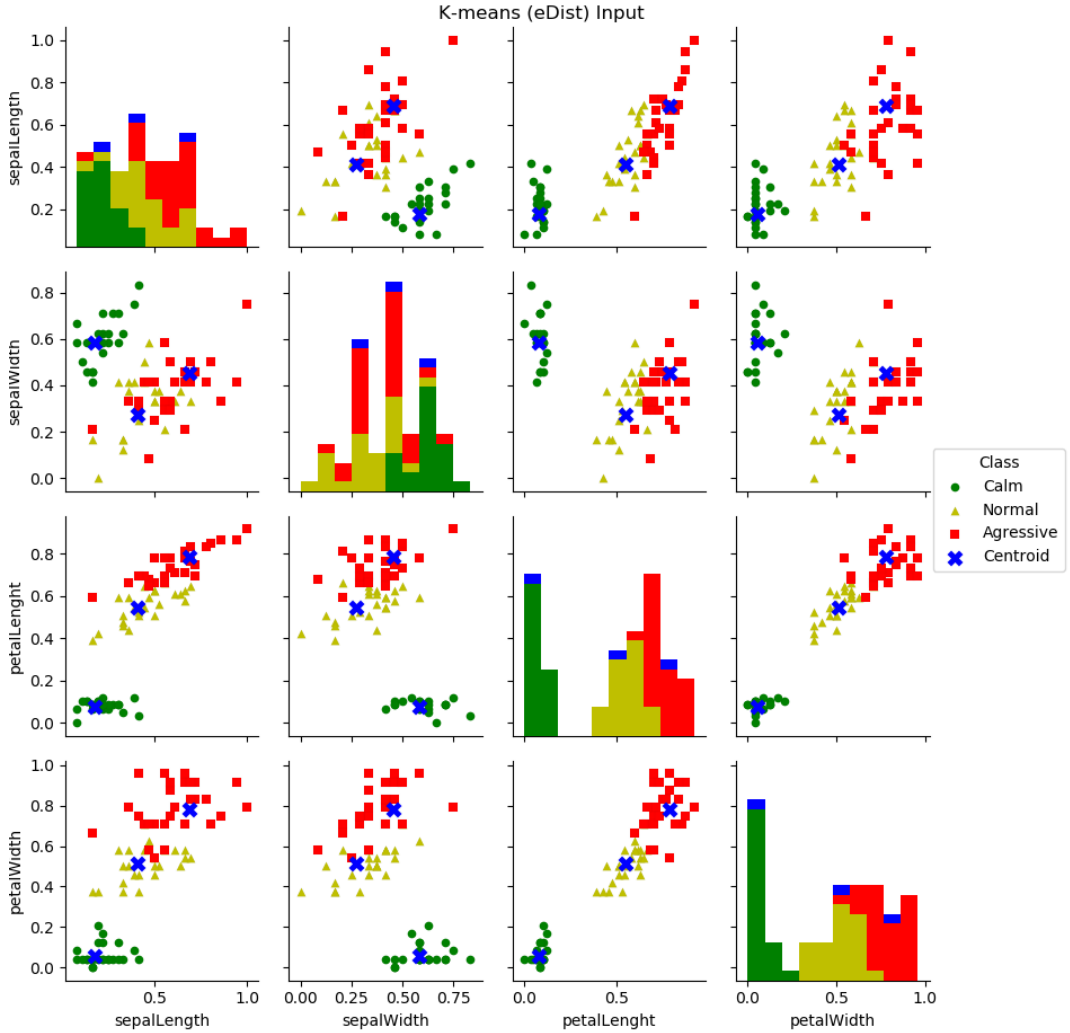


Figure 2 – Calculated centroids and classification

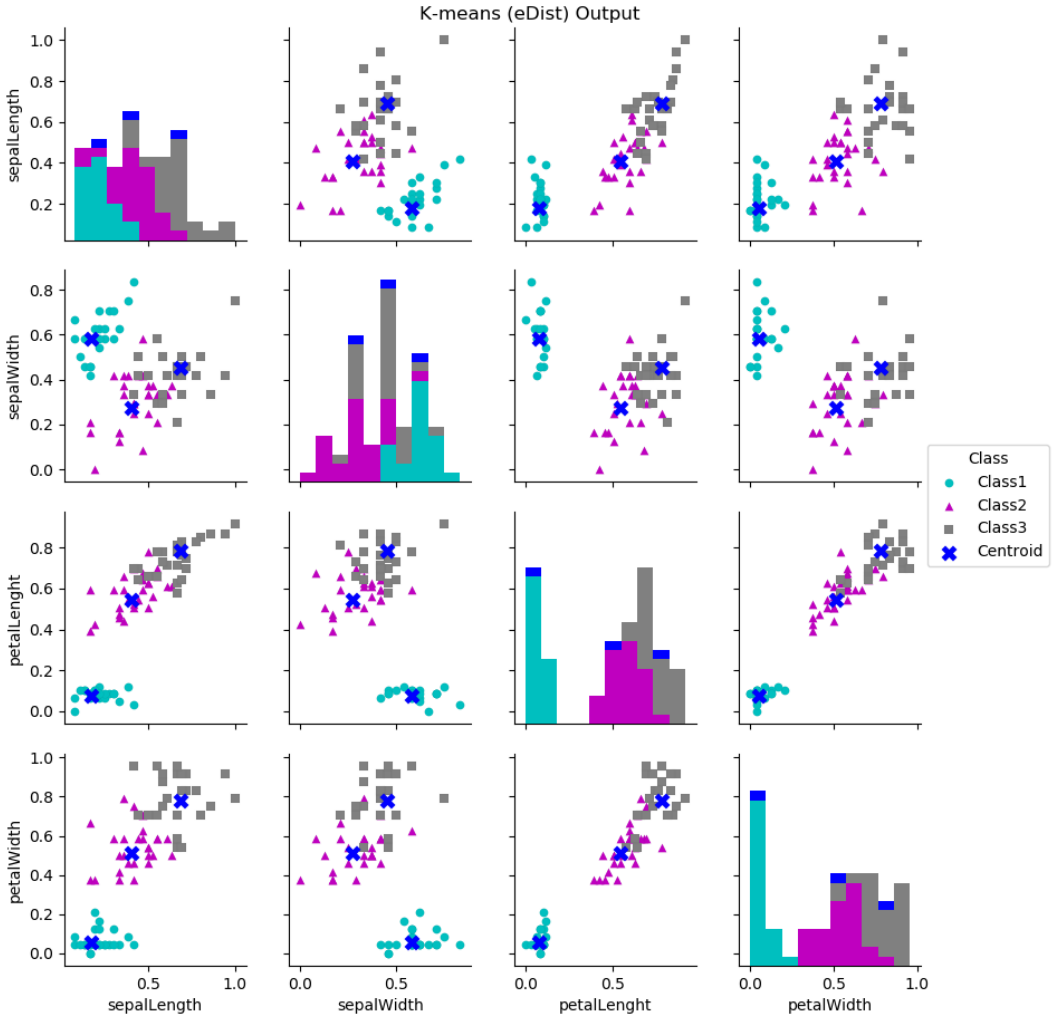


Table 1 – Confusion Matrix

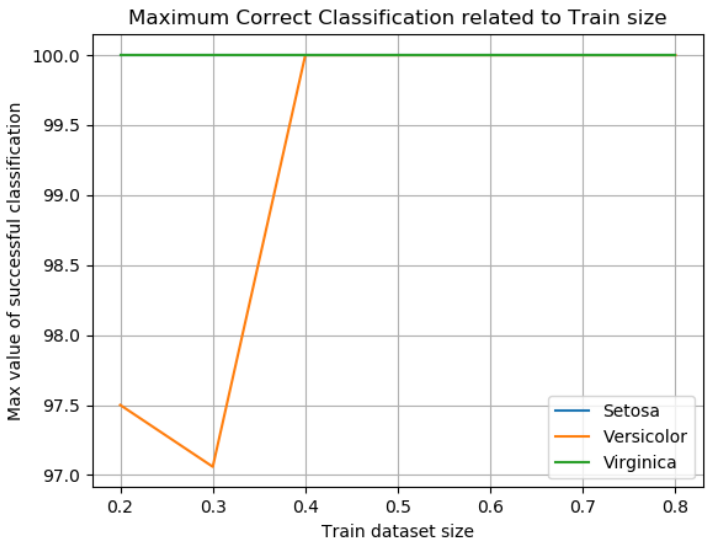
Classes	*Setosa*	*Virginica*	*Versicolor*
Setosa	100	0	0
Virginica	0	85.7351	18.0056
Versicolor	0	14.2649	81.9944

Table 2 – Results

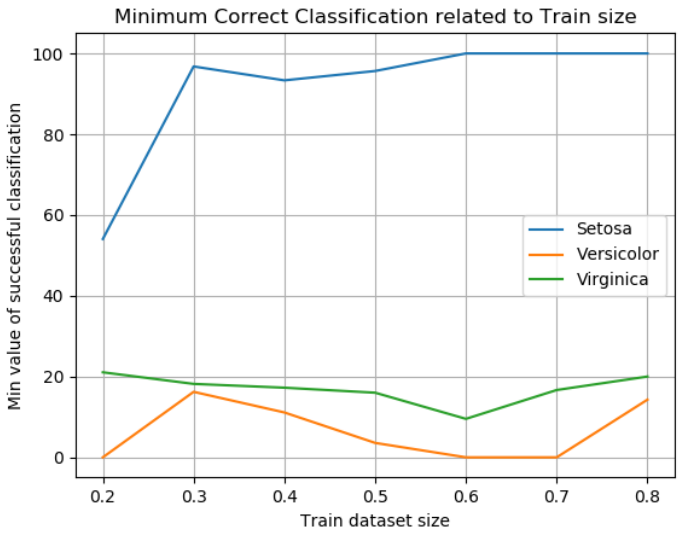
Algorithm	Maximum	Minimum	Standard Deviation
MLP	100	50	7.17

K-Means (Mahalanobis Distance)

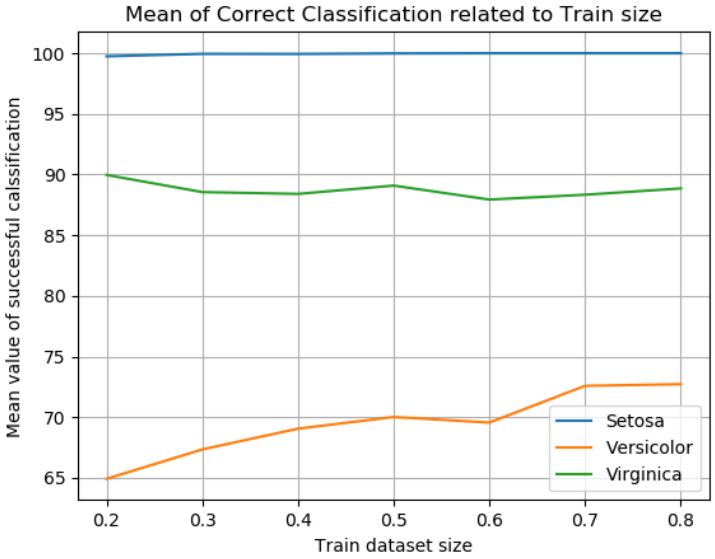
Maximum success rate



Minimum success rate



Mean of successful classification ratio



Standard deviation of successful classification ratio

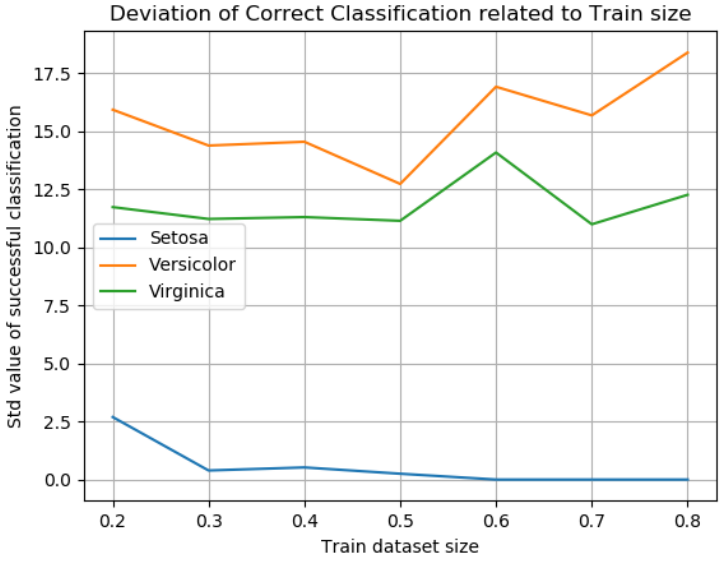


Figure 1 – Expected centroids and classification

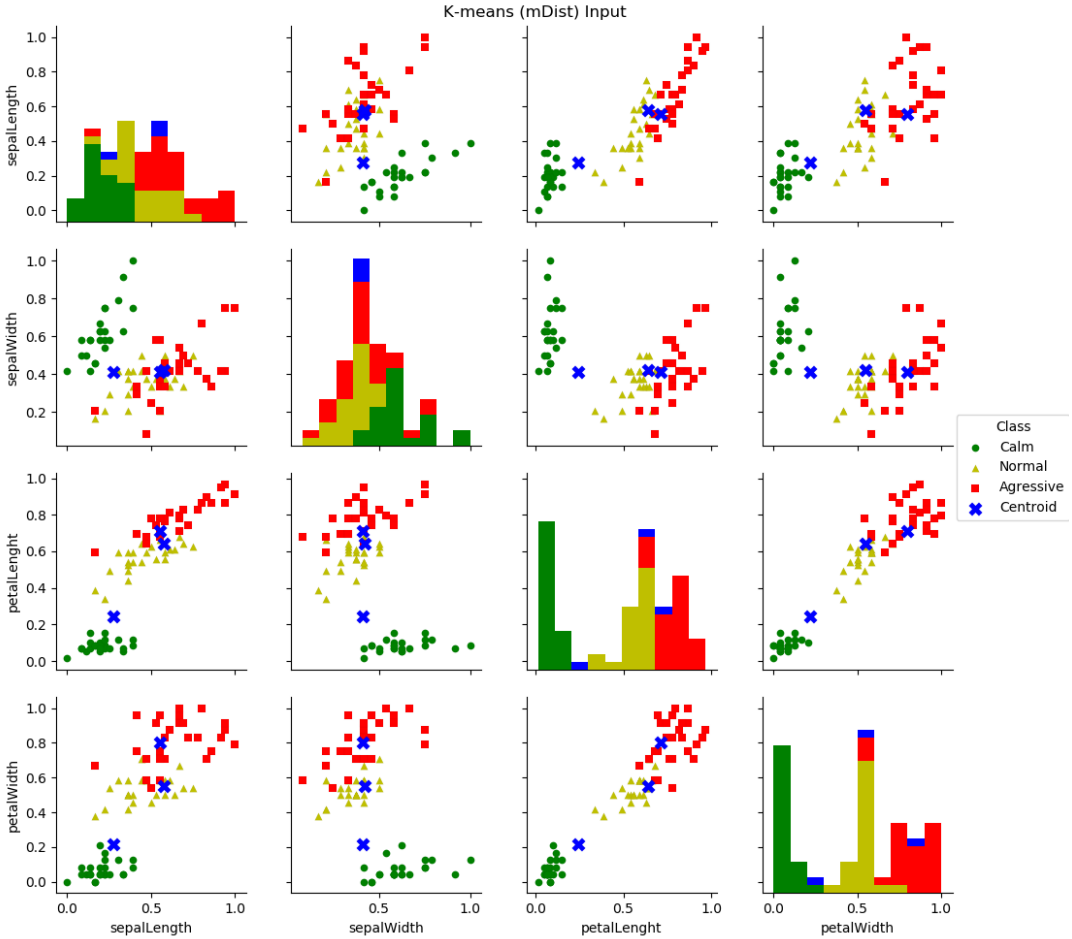


Figure 2 – Calculated centroids and classification

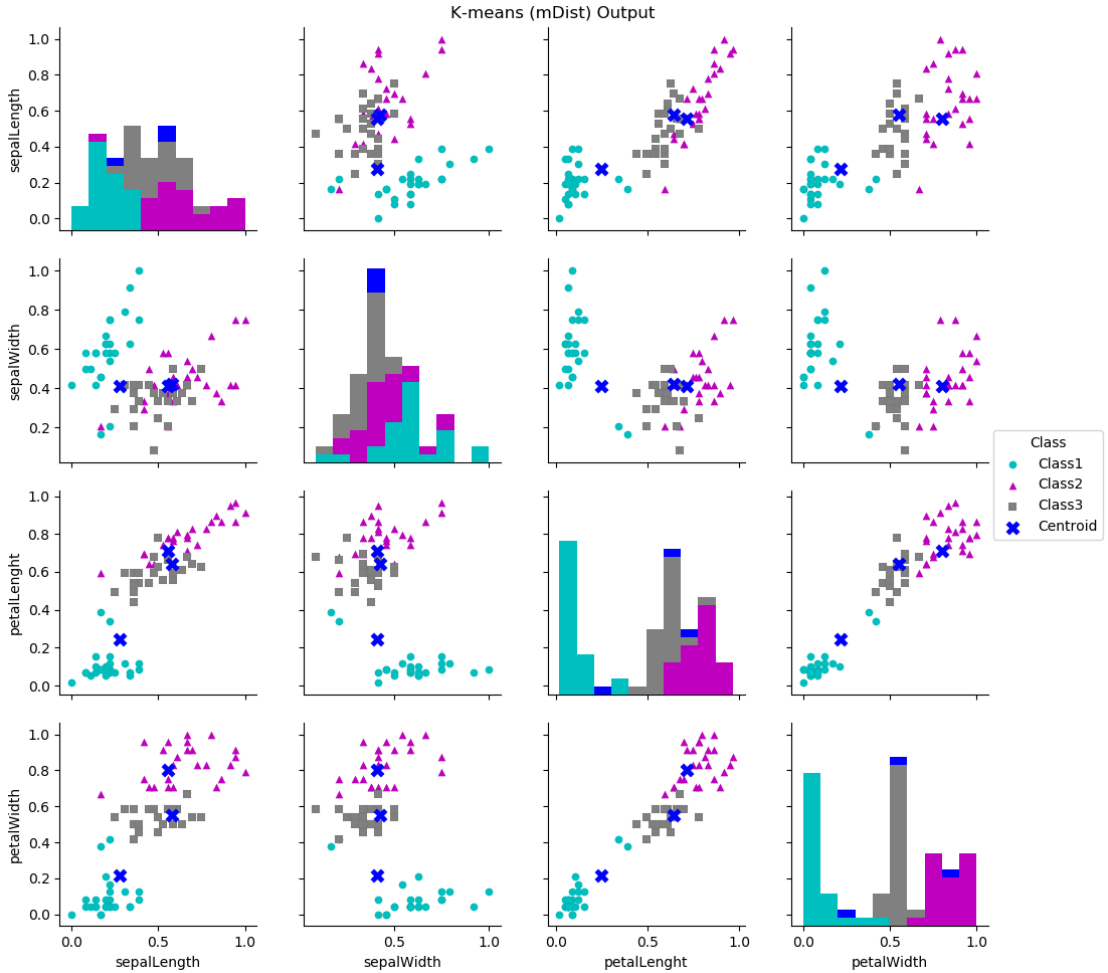


Table 1 – Confusion Matrix

Classes	*Setosa*	*Virginica*	*Versicolor*
Setosa	100	0	0
Virginica	0	72.8976	11.9904
Versicolor	0	27.1024	88.0096

Table 2 – Results

Algorithm	Maximum	Minimum	Standard Deviation
MLP	100	0	9.574