S1 – Examine the effect of dataset size on algorithm performance

- Automated quality control inspection of geometric tip defects in medical needle manufacturing

This supplementary material presented a further investigation into the effect of dataset size on algorithm performance in terms of false-positive (FP) and false-negative (FN) results. False-positive and false-negative results, classifying a qualified sample into the defective category and vice versa, in the manufacturing process are the most unacceptable errors. This work stochastically selected 10%, 30%, 50%, 70% and 90% samples from the whole training dataset. And used these small-sized datasets in training the models.

Fig. 1- 5 shows the positions of all false-positive and false-negative results by the compared methods. According to Fig. 8 in the paper 'Automated quality control inspection of geometric tip defects in medical needle manufacturing', the Quadratic SVM (Q-SVM) model shown in Fig. 1, has FP and FN cases that cover the widest range of fail probability (including the fail probability within 70% and 90%), than the other models. It can be seen from Fig. 2 that the obtained number of FP and FN by the Medium Gaussian SVM (MG-SVM) model is decreasing with the increase of the dataset size. However, the number has not converged to zero. Thus, more training data is needed to train MG-SVM. Fig. 3 shows that the Optimised SVM (O-SVM) model obtains the most cases of FP and FN, 33 cases compared with the second largest number 24 resulted from the Decision Tree (DT) model, as shown in Fig.4. It is because the parameters of the O-SVM model are optimised to best fit each training group. However, the size of each training group was relatively too small for this model. This resulted in the strongest overfitting for the O-SVM model than the other models, so O-SVM generated the most cases of FP and FN than the other models.
It clearly shows that the proposed method (Fig. 5) has the smallest account of false-positive and false-negative results overall. This means the proposed method does not require as big a training dataset as the compared machine learning models, and it also does not has specific requirement on positive-and-negative sample ratios. Also, all the miss-classification of the proposed method came from classifying the qualified samples into the defective category. No defective samples were incorrectly passed by the proposed method. This satisfied the stringent quality requirement in manufacturing safety-critical medical devices.

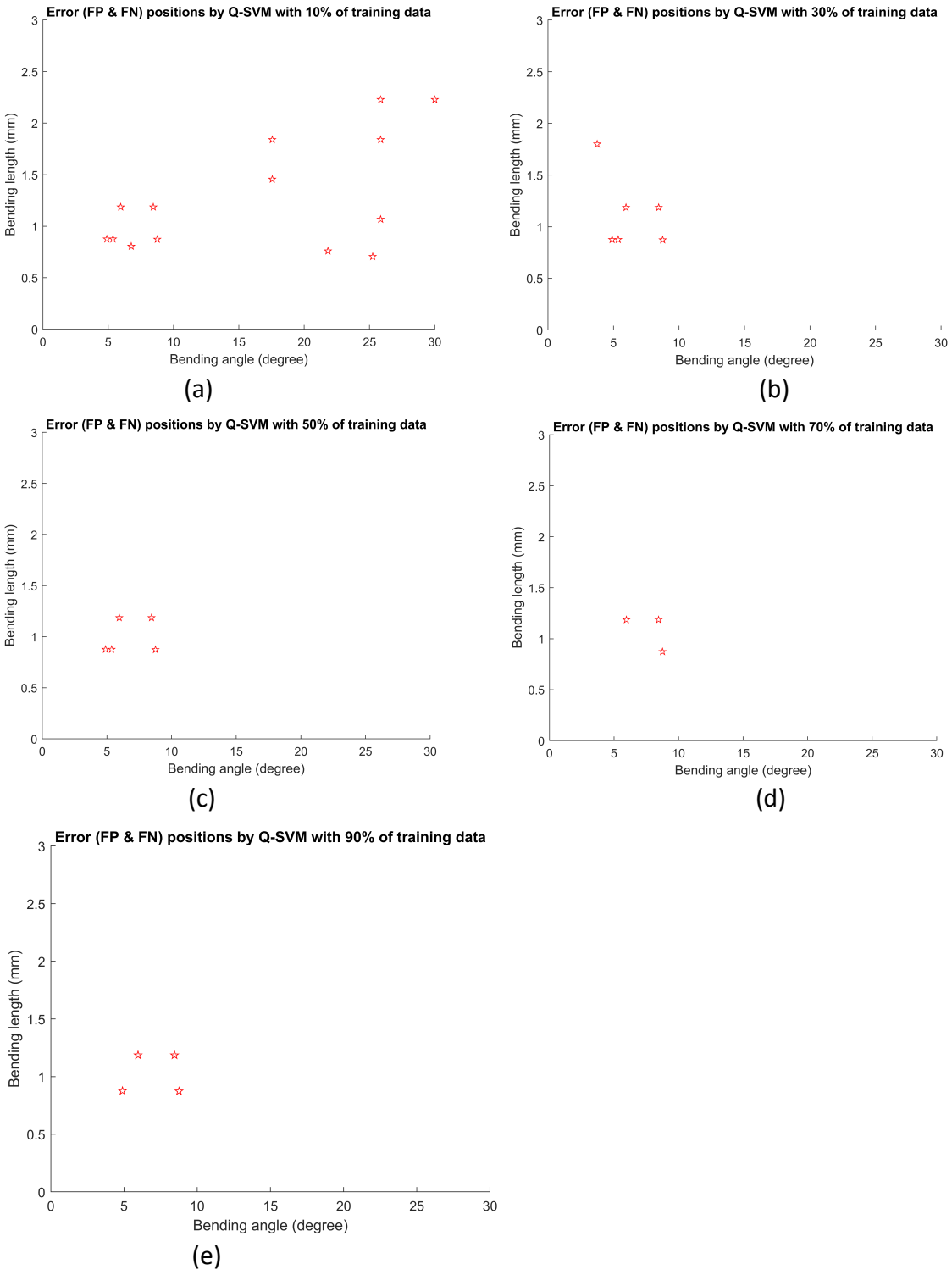

Figure 1. The obtained false-positive (FP) and false-negative (FN) results by Quadratic SVM (Q-SVM) model from the testing data group (132 data points) when the model was trained by: (a) 10%, (b) 30%, (c) 50%, (d) 70%, and (e) 90% of the training dataset. The total account of FP and FN obtained were 14, 6, 5, 3, 4 in (a), (b), (c), (d), (e), respectively.

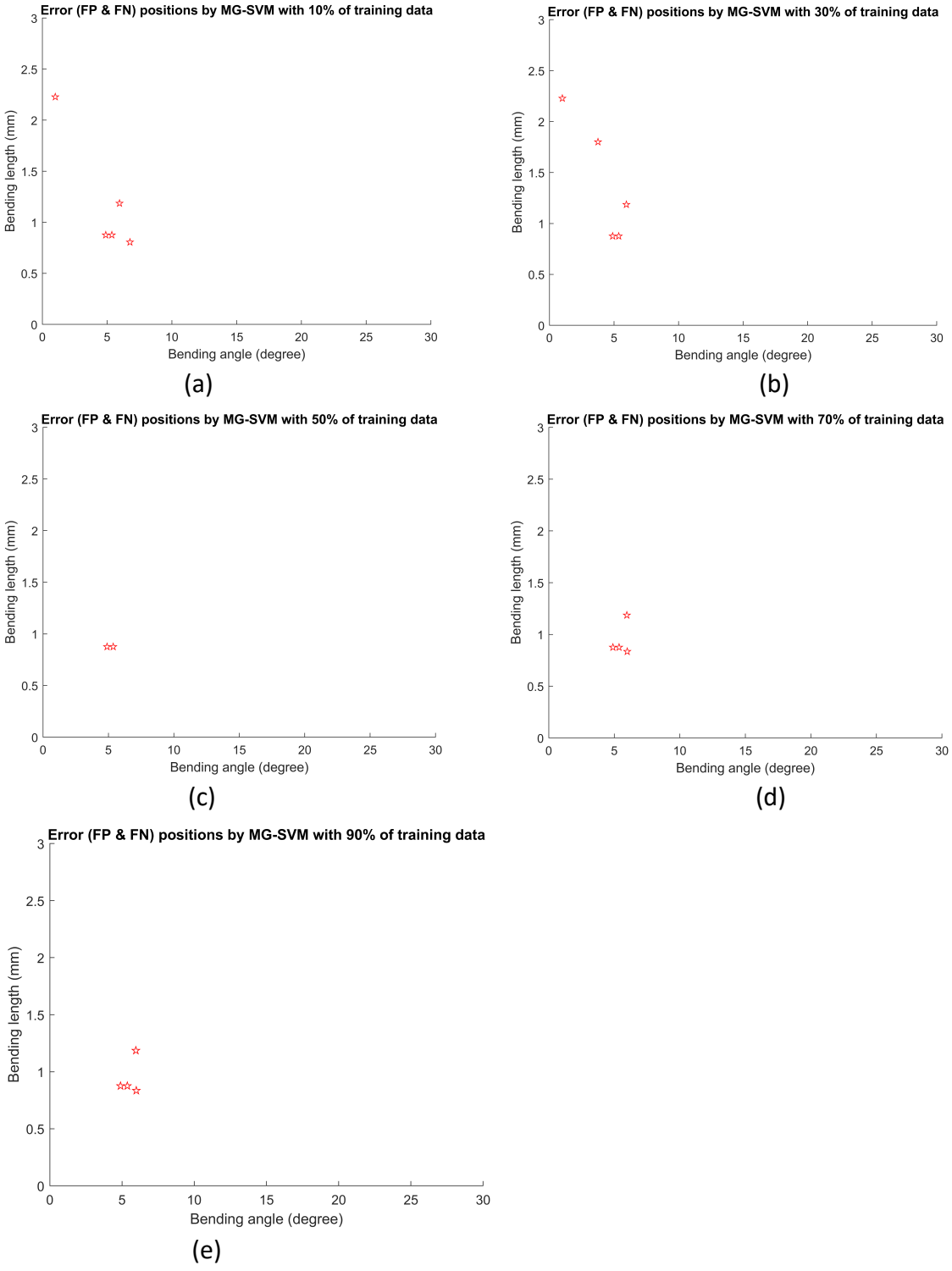

Figure 2. The obtained false-positive (FP) and false-negative (FN) results by Medium Gaussian SVM (MG-SVM) model from the testing data group (132 data points) when the model was trained by: (a) 10%, (b) 30%, (c) 50%, (d) 70%, and (e) 90% of the training dataset. The total account of FP and FN obtained were 5, 5, 2, 4, 4 in (a), (b), (c), (d), (e), respectively.

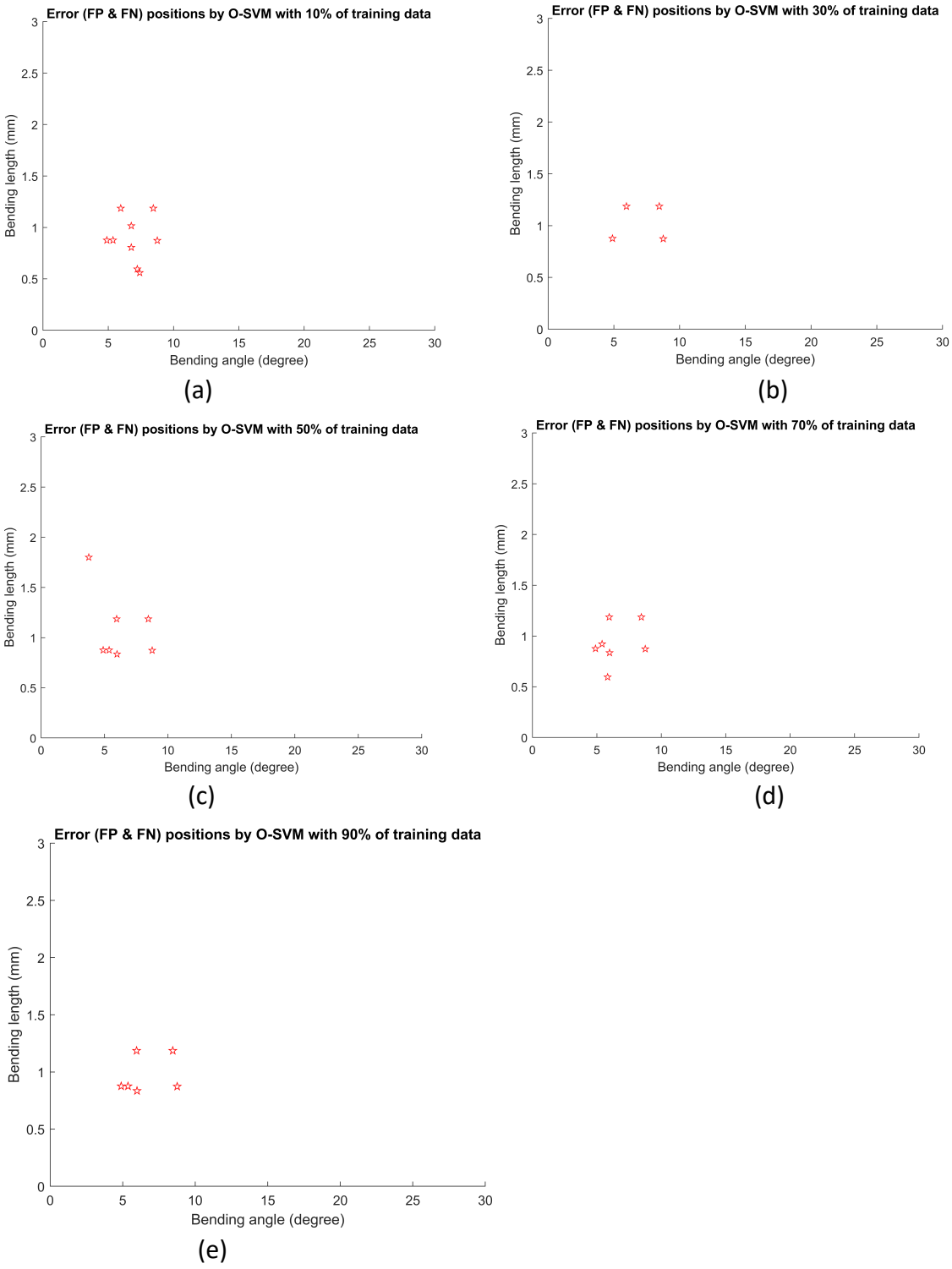

Figure 3. The obtained false-positive (FP) and false-negative (FN) results by Optimised SVM (O-SVM) model from the testing data group (132 data points) when the model was trained by: (a) 10%, (b) 30%, (c) 50%, (d) 70%, and (e) 90% of the training dataset. The total account of FP and FN obtained were 9, 4, 7, 7, 6 in (a), (b), (c), (d), (e), respectively.

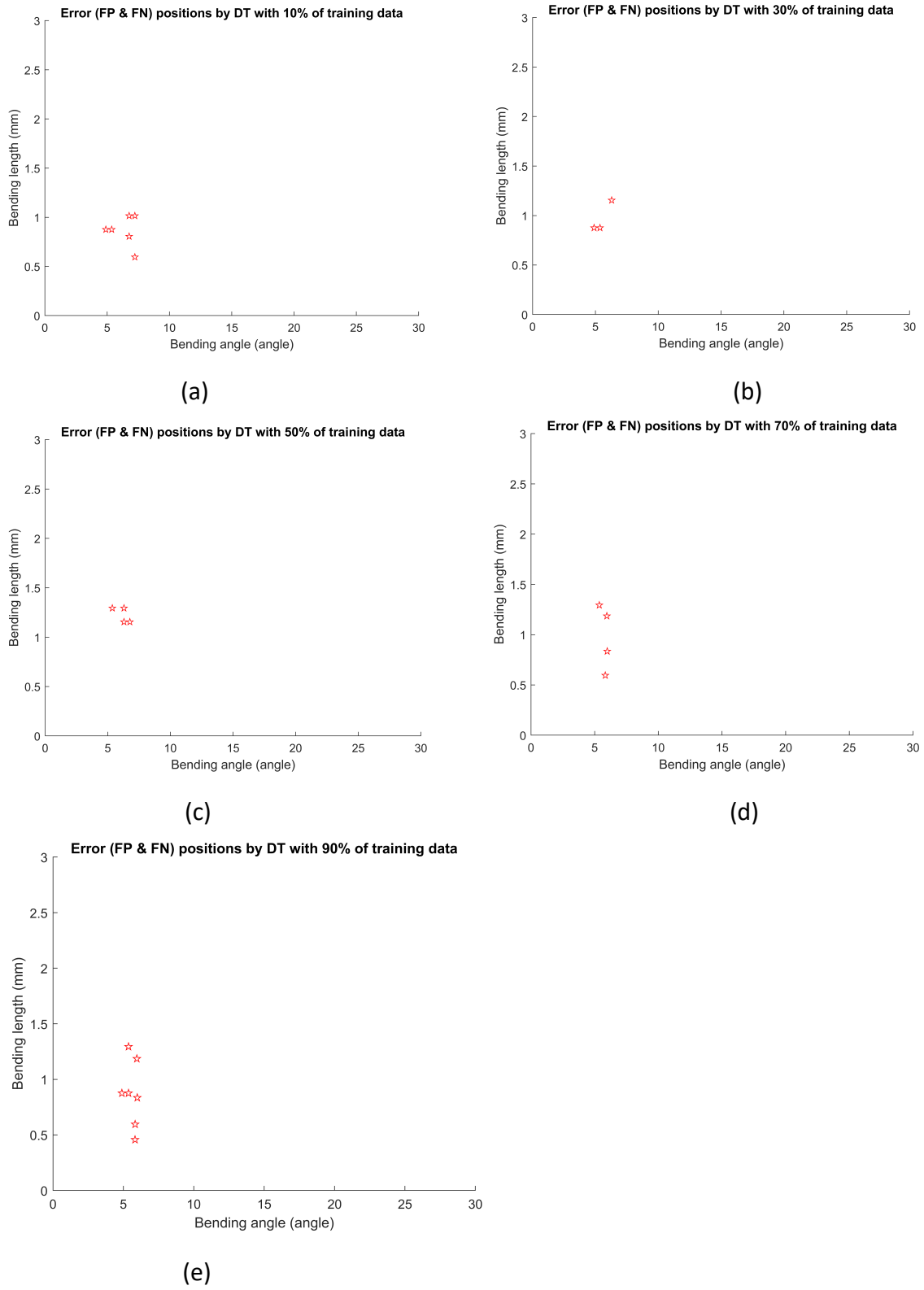

Figure 4. The obtained false-positive (FP) and false-negative (FN) results by Decision Tree (DT) model from the testing data group (132 data points) when the model was trained by: (a) 10%, (b) 30%, (c) 50%, (d) 70%, and (e) 90% of the training dataset. The total account of FP and FN obtained were 6, 3, 4, 4, 7 in (a), (b), (c), (d), (e), respectively.

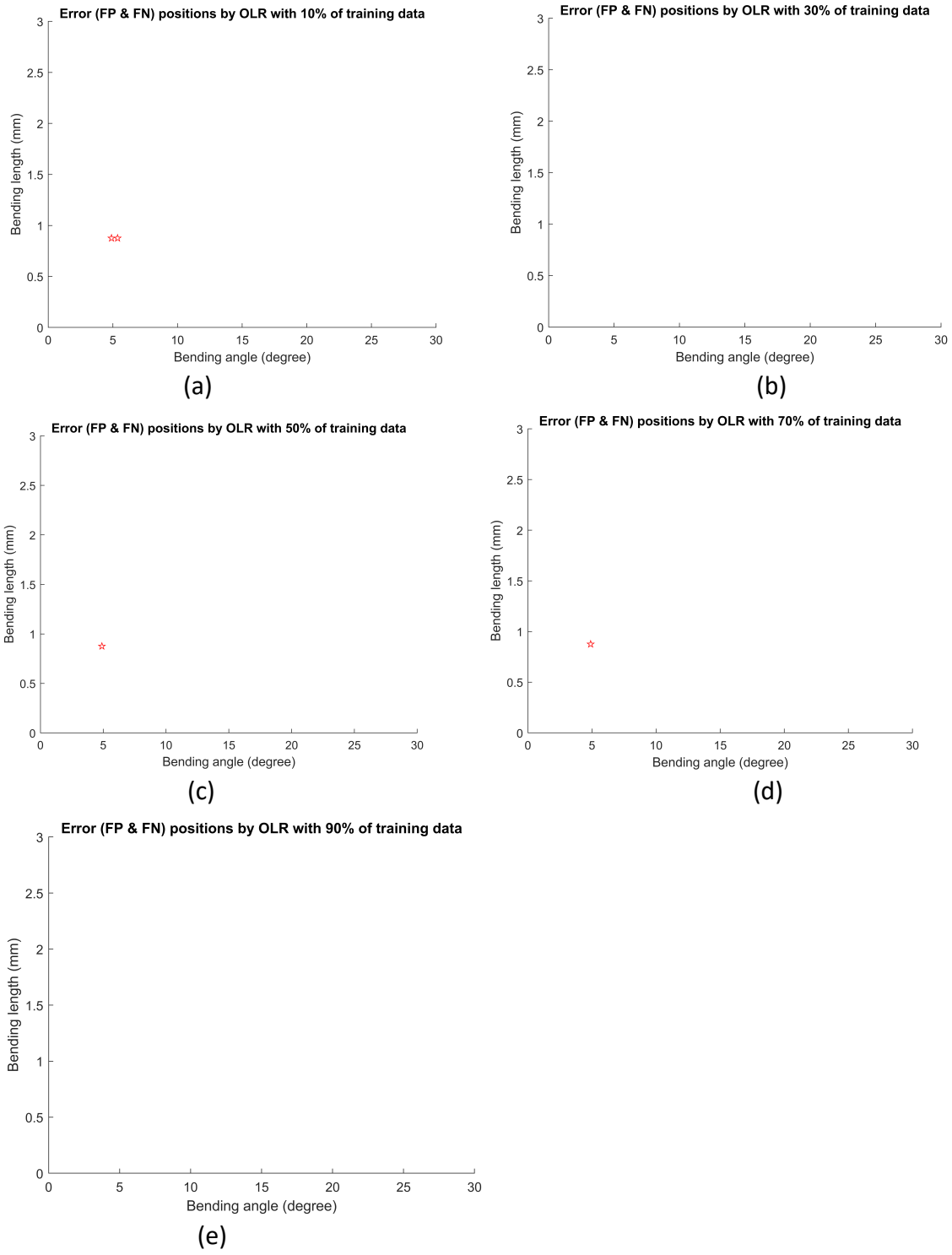

Figure 5. The obtained false-positive (FP) and false-negative (FN) results by the proposed Ordinal Logistic Regression (OLR) algorithm from the testing data group (132 data points) when the model was trained by: (a) 10%, (b) 30%, (c) 50%, (d) 70%, and (e) 90% of the training dataset. The total account of FP and FN obtained were 2, 0, 1, 1, 0 in (a), (b), (c), (d), (e), respectively.