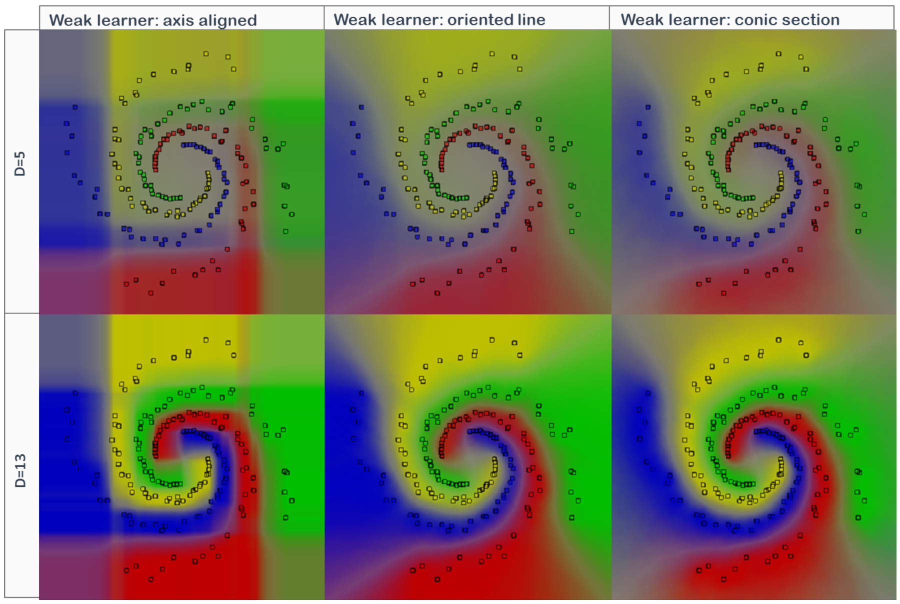

Fig. 3.7: **The effect of randomness.** The same set of 4-class training data is used to train 6 different forests, for 2 different values of D and 3 different weak learners. This experiment is identical to that in fig. 3.6 except that we have used much more training randomness. In fact $\rho = 5$ for all split nodes. The forest size is kept fixed at $T = 400$. More randomness reduces the artifacts of the axis-aligned weak learner a little, as well as reducing overall prediction confidence too. See text for details.

much lower overall confidence, especially noticeable in shallower trees (washed out colours in the top row).

A disadvantage of the more complex weak learners is that they are associated to a larger parameters space. Thus finding discriminative sets of parameter values may be time consuming. However, in this toy example the more complex conic section learner model works well for deeper trees ($D = 13$) even for small values of ρ (large randomness). The results reported here are only indicative. In fact, which specific weak learner to use depends on considerations of efficiency as well as accuracy and it is application dependent. Many more examples, ani-