Appendix A. Empirical Results - Effect of Varying Activation Functions

In this section, we study the effect of varying activation functions on the overall performance of the algorithm. In particular, the same architecture as studied in the Empirical Results section of the main paper is used, but this time a Rectified Linear Unit (RELU) activation is used. Tables 5 and 6 illustrate the performance of the distributed algorithm with (0.2%) and without overlap. It appears that with the RELU activation and no overlap of features, the distributed algorithm closely mimics the centralized one in almost all the datasets except HTSensor (which has a comparable performance with overlap). In two datasets, Dexter and Madelon, the performance of the distributed algorithm surpasses the centralized counterpart.

Dataset	Cent AUC	Dist. AUC	Cent. Itr. (I_C)	Dist. Itr. (I_D)
Arcene	0.94 ± 0.007	0.93 ± 0.003	292.33 ± 32.99	472.33 ± 4.71
Dexter	0.65 ± 0.004	0.84 ± 0.009	322.33 ± 26.67	549 ± 21.60
Dorothea Bal.	0.93 ± 0.002	0.93 ± 0.006	129 ± 8.16	135.67 ± 16.99
Gisette	0.97 ± 0.002	0.97 ± 0.002	415.67 ± 81.78	672.33 ± 83.39
Madelon	0.62 ± 0.002	0.64 ± 0.002	265.67 ± 61.28	412.33 ± 217.46
MNIST Bal.	0.96 ± 0.002	0.95 ± 0.002	145.67 ± 4.71	236.67 ± 33.99
HT Sensor	0.99 ± 0.002	0.99 ± 0.013	229 ± 45.46	249 ± 64.81

Table 5: Performance of the centralized(C) and distributed algorithms(D) with no overlap of features. The consensus neural network uses cross-entropy loss function, RELU activation for the hidden layer, and soft-max activation for the output layer. The results are averaged over three trials.

Dataset	Cent AUC	Dist. w/overlap AUC	Cent. Itr. (I_C)	Dist. Itr. (I_D)
Arcene	0.94 ± 0.007	0.91 ± 0.003	292.33 ± 32.99	472.33 ± 4.71
Dexter	0.65 ± 0.004	0.83 ± 0.009	322.33 ± 26.67	549 ± 21.60
Dorothea Bal.	0.93 ± 0.002	0.92 ± 0.006	129 ± 8.16	135.67 ± 16.99
Gisette	0.97 ± 0.002	0.97 ± 0.002	415.67 ± 81.78	672.33 ± 83.39
Madelon	0.62 ± 0.002	0.64 ± 0.002	265.67 ± 61.28	412.33 ± 217.46
MNIST Bal.	0.96 ± 0.002	0.95 ± 0.002	145.67 ± 4.71	236.67 ± 33.99
HT Sensor	0.99 ± 0.002	0.99 ± 0.013	229 ± 45.46	249 ± 64.81

Table 6: Performance of the centralized(C) and distributed algorithms(D) with 0.2% overlap of features. The consensus neural network uses cross-entropy loss function, RELU activation for the hidden layer, and soft-max activation for the output layer. The results are averaged over three trials.
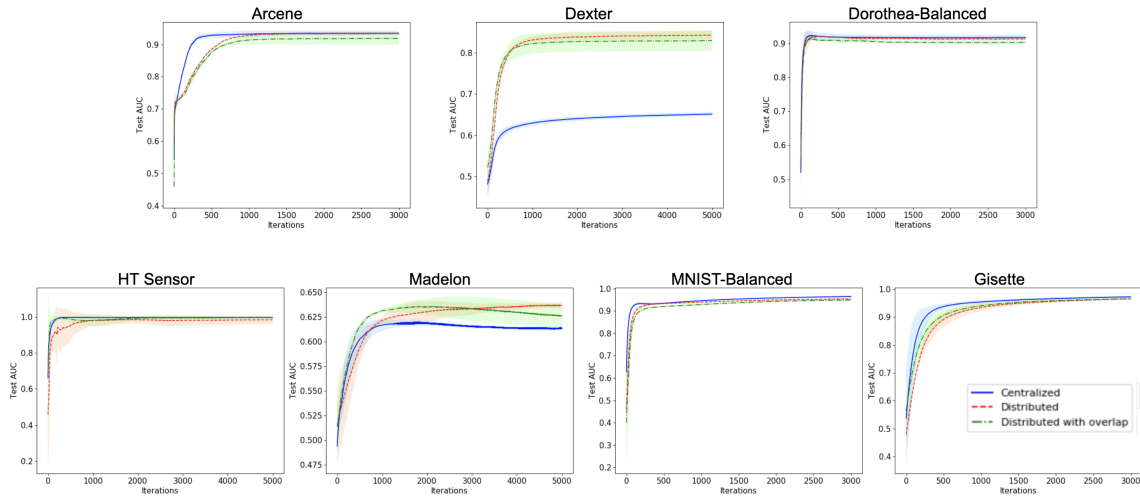
Figure 3: AUC on the test sets for both centralized and distributed settings, using RELU activations on the seven datasets presented in the main paper. For the distributed algorithm, test accuracy results averaged over three random vertical feature splits with and without overlap are presented.