Combining Methods to Stabilize and Increase Performance of Neural Network-Based Classifiers

Fabricio A. Breve fabricio@dc.ufscar.br
Moacir P. Ponti Jr. moacir@dc.ufscar.br
Nelson D. A. Mascarenhas nelson@dc.ufscar.br

DC – Departamento de Computação
UFSCar - Universidade Federal de São Carlos, São Paulo, SP, Brasil

Goals

- Recognize materials in multispectral images (obtained with a tomograph scanner) using a Neural Network based classifier (Multilayer Perceptron)
- Investigate classifier combiner techniques in order to improve performance

Summary

- Image Acquisition
- Classification Methods
- Evaluation
- Experiments
- Results
- Conclusions
- Future Works

Image Acquisition

- First generation Computerized Tomograph developed by Embrapa in order to explore applications in soil science
  - X-Ray and γ-ray fixed sources
  - Object being studied is rotated and translated

Image Acquisition

- Phantom built with materials found in soil
- Plexiglass support
- 4 Cylinders containing: Aluminum, Water, Phosphorus and Calcium

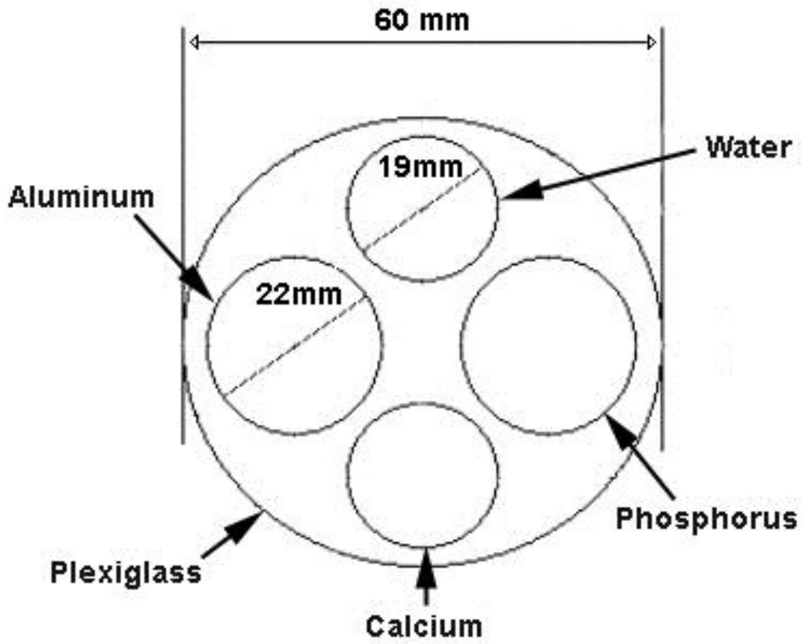


Image Acquisition

40keV

60keV

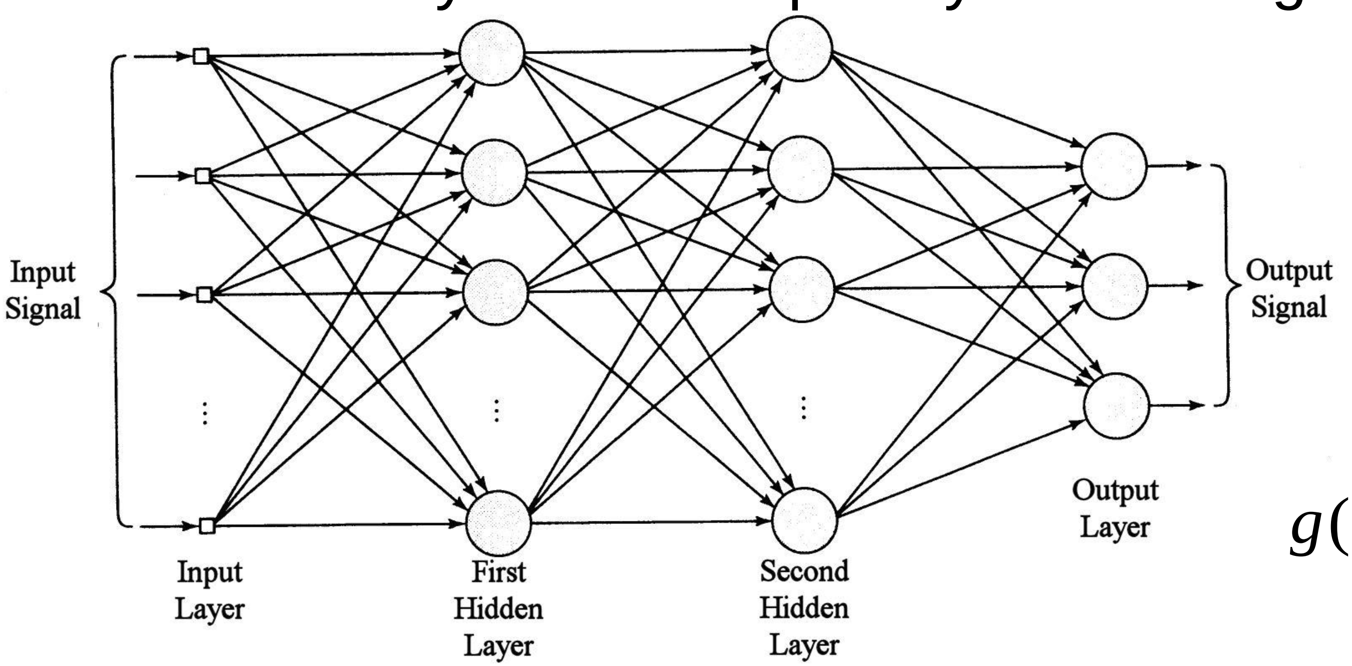
85keV

662keV

- X-ray sources:
  - 40 keV
  - 85 keV
- γ-ray sources
  - 60 keV (Americium)
  - 662 keV (Cesium)
- 65x65 pixels
- 3 seconds of exposure
- 256 levels of gray
- High level of noise

Classification Methods

- Multilayer Perceptron
  - Neural-Network Based Classifier
  - One Input Layer, one or more hidden layers, one Output Layer
  - Training: Error Backpropagation Algorithm
  - Hidden layers and output layers with Sigmoid functions



$$g(a) \equiv \frac{1}{1 + \exp(-a)}$$

Classification Methods

- Bagging
  - Bootstrap AGGregatING
  - Bootstrap sets built randomly from the original training set using substitution
  - Each bootstrap set trains a classifier
  - Outputs are combined using majority voting
  - Requires the base classifier to be unstable
    - minor differences in the training set can lead to major changes in the classifier

Classification Methods

- Decision Templates
  - Continuous-valued outputs from each classifier with a different initialization for a given sample are used to build a decision profile
  - The Decision Templates are the mean over all the decision profile from each training sample for each class
  - The label of a test sample is chosen by comparing its decision profile with each decision template and choosing the most similar one

$$DP(x) = \begin{bmatrix} d_{1,1}(x) & \cdots & d_{1,j}(x) & \cdots & d_{1,c}(x) \\ \vdots & & \vdots & & \vdots \\ d_{i,1}(x) & \ldots & d_{i,j}(x) & \cdots & d_{i,c}(x) \\ \vdots & & \vdots & & \vdots \\ d_{L,1}(x) & \ldots & d_{L,j}(x) & \cdots & d_{L,c}(x) \end{bmatrix}$$

Classification Methods

- Dempster-Shafer
  - Based on the Evidence Theory, a way to represent cognitive knowledge
  - It’s like the Decision Templates method, but for each test sample we calculate the proximity between the decision template and the output of each classifier.
  - These are used to calculate the belief degree for every class.
  - At last the final degrees of support of each test sample for each class are calculated from the belief degrees.

Evaluation

- Hold-Out
  - Splits the set of available data in two halves:
    - Training set
    - Testing set
  - It’s a fast testing scheme, as required by Multilayer Perceptron
- Kappa Coefficient
  - Measures the agreement rating between the classification of the test samples and their true class

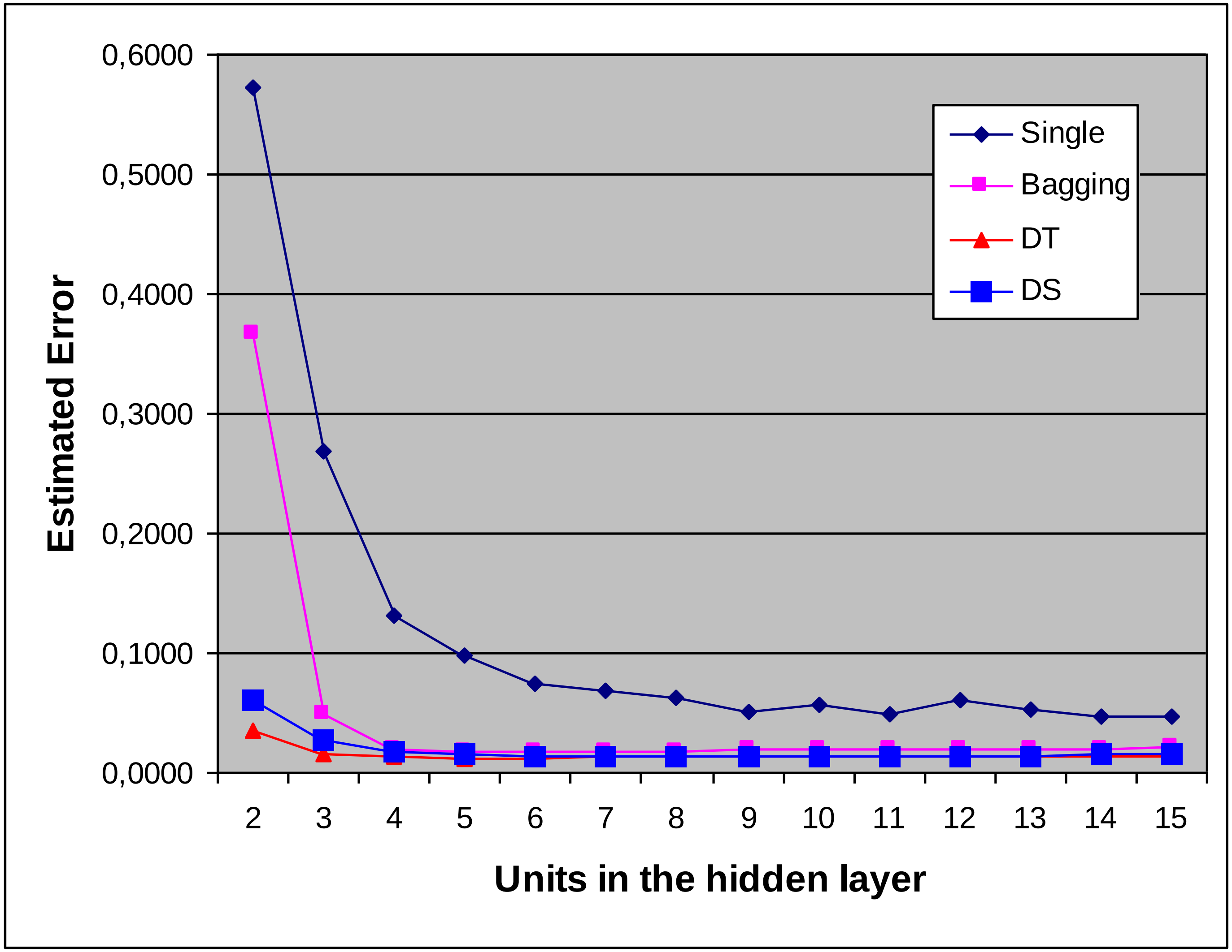
Experiments

- 480 samples (80 samples from each of the 6 class):
  - Aluminum
  - Water
  - Phosphorus
  - Calcium
  - Plexiglass
  - Background
- 240 samples (40 from each class) for training
- 240 samples for testing

Experiments

- Networks with 2 to 15 units in one single hidden layer
- Every experiment for all classification methods were executed 100 times
- Bagging using means instead of majority vote to take advantage of continuous-valued output
- All experiments with classifier combiners used 10 base classifiers with different initializations

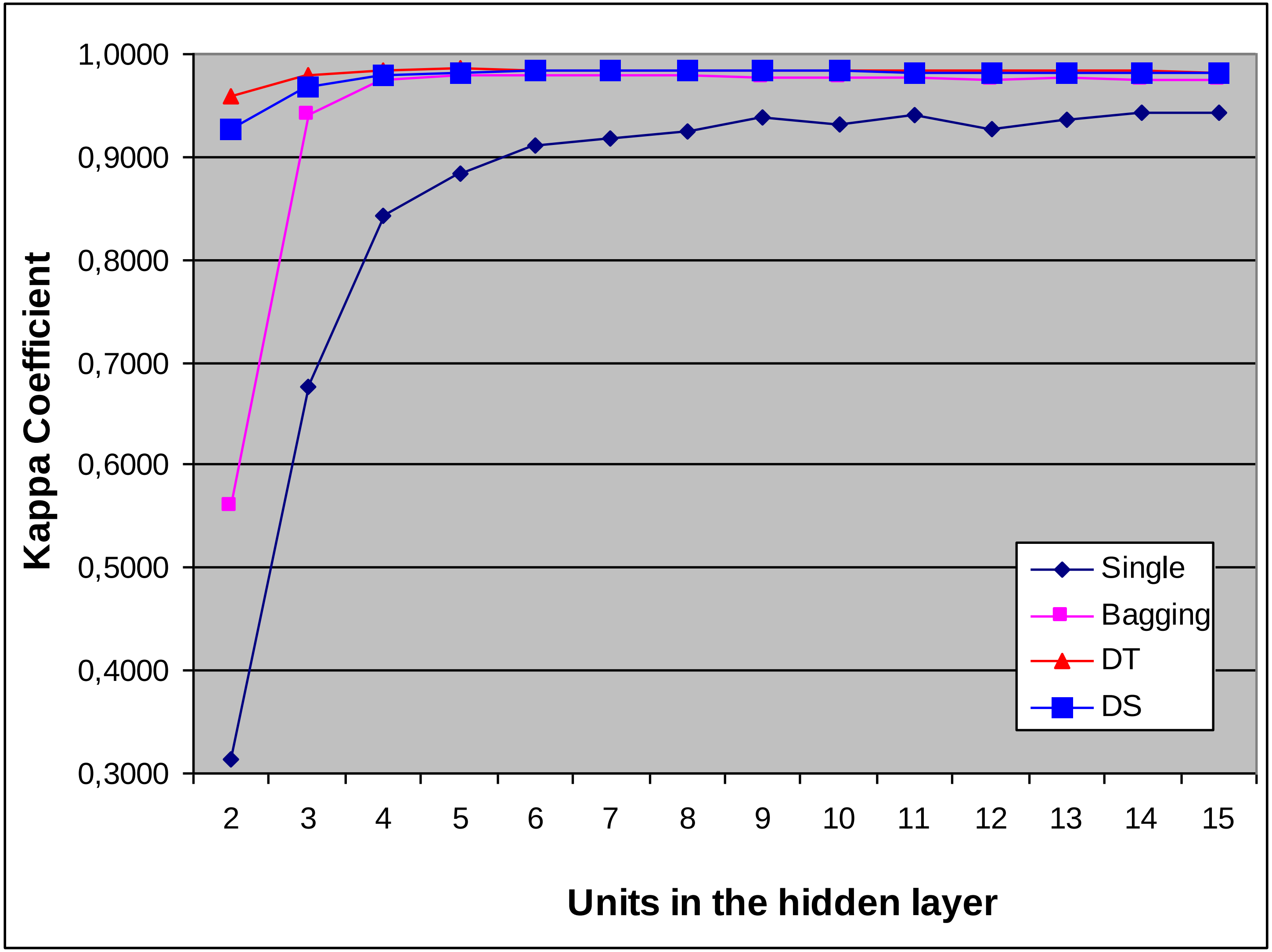
Results – Estimated Error



Results – Estimated Error

Units in the hidden layer	Single Classifier	Bagging	DT	DS
2	0.5720	0.3675	0.0349	0.0613
3	0.2689	0.0493	0.0163	0.0275
4	0.1318	0.0200	0.0141	0.0177
5	0.0976	0.0170	**0.0123**	0.0151
6	0.0741	**0.0168**	0.0127	0.0139
7	0.0681	0.0175	0.0129	0.0138
8	0.0636	0.0179	0.0130	**0.0137**
9	0.0511	0.0190	0.0134	0.0138
10	0.0570	0.0190	0.0135	0.0139
11	0.0497	0.0191	0.0136	0.0143
12	0.0603	0.0203	0.0136	0.0143
13	0.0525	0.0196	0.0137	0.0145
14	0.0477	0.0204	0.0140	0.0151
15	**0.0470**	0.0210	0.0143	0.0150

Results – Kappa Coefficient

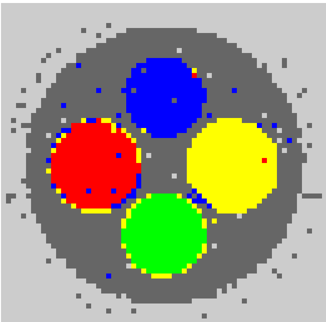


Results – Kappa Coefficient

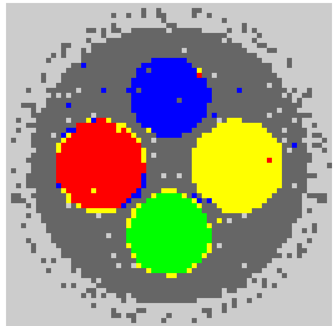
Units in the hidden layer	Single Classifier	Bagging	DT	DS
2	0.3137	0.5591	0.9581	0.9265
3	0.6773	0.9409	0.9805	0.9671
4	0.8419	0.9760	0.9831	0.9788
5	0.8829	0.9796	**0.9853**	0.9819
6	0.9111	**0.9799**	0.9848	0.9833
7	0.9183	0.9790	0.9845	0.9835
8	0.9237	0.9786	0.9844	**0.9836**
9	0.9387	0.9773	0.9840	0.9835
10	0.9316	0.9773	0.9838	0.9833
11	0.9404	0.9771	0.9837	0.9829
12	0.9277	0.9757	0.9837	0.9829
13	0.9371	0.9765	0.9836	0.9827
14	0.9428	0.9756	0.9832	0.9819
15	**0.9436**	0.9748	0.9828	0.9821

Results – Thematic Images

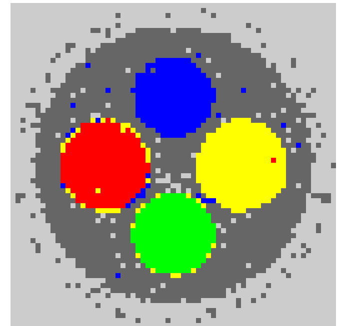
Thematic images for the best classifier of each group (best case)



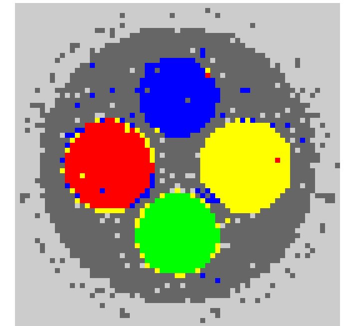
Single Classifier
Estimated Error: 0.0083
Kappa Coefficient: 0.9900



Bagging
Estimated Error: 0.0083
Kappa Coefficient: 0.9900

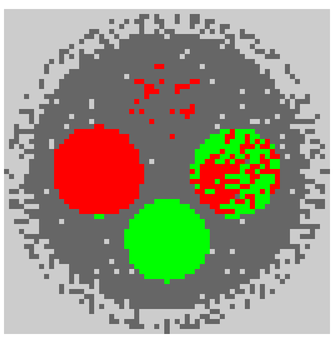


Decision Templates
Estimated Error: 0.0042
Kappa Coefficient: 0.9950

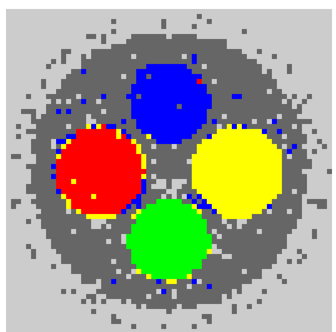


Dempster-Shafer
Estimated Error: 0.0125
Kappa Coefficient: 0.9850

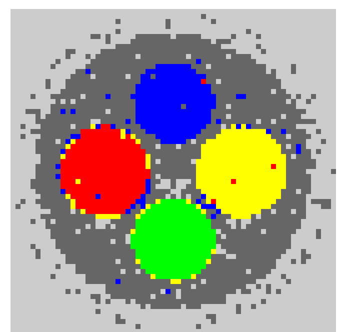
Thematic images for the best classifier of each group (worst case)



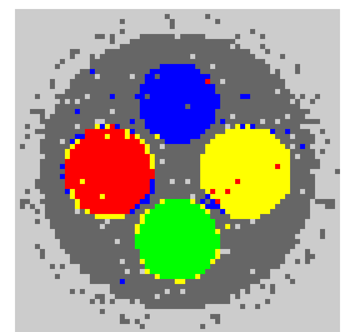
Single Classifier
Estimated Error: 0.3417
Kappa Coefficient: 0.5900



Bagging
Estimated Error: 0.0250
Kappa Coefficient: 0.9700



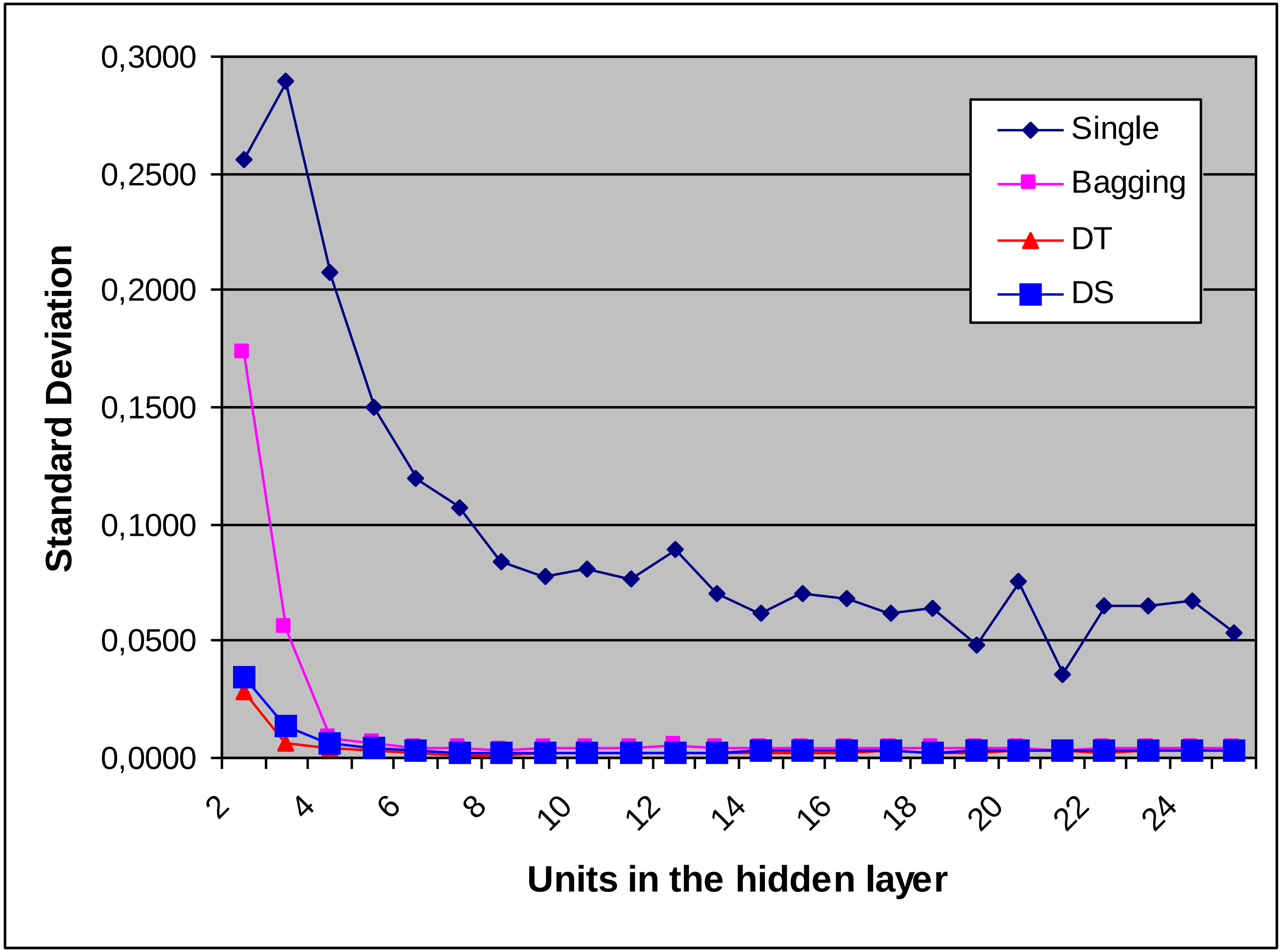
Decision Templates
Estimated Error: 0.0208
Kappa Coefficient: 0.9750



Dempster-Shafer
Estimated Error: 0.0208
Kappa Coefficient: 0.9750

water **aluminum** **phosphorus** **calcium** **plexiglass** **background**

Results – Standard Deviation



Conclusions

- Classifier combination led to more stable classifiers, where even the worst case still delivers a good classification
- The best results with all the combiners were achieved with few units in the hidden layer
- Dempster-Shafer and Decision Templates combiners showed relatively good results no matter how many units there were in the hidden layer

Conclusions

- Multilayer Perceptron based classifiers to identify materials on CT images is viable, even in images with high noise levels
- The use of classifiers combiners led to more stable systems and minimized the effects of the unstable nature of the individual MLP classifiers

Future Works

- experiments with more than 10 base classifiers to observe if the performance improvements could worth the extra time to train all the base classifiers
- a combination using Bagging techniques, but using Dempster-Shafer or Decision Templates as the combiner, instead of the mean combiner we used in our experiments

Acknowledgements

- Dr. Paulo E. Cruvinel for providing the multispectral images used in the experiments
- CAPES and FAPESP (grant n. 04/05316-7) for student scholarship
- This work was also partially supported by FAPESP Thematic Project 2002/07153-2

Combining Methods to Stabilize and Increase Performance of Neural Network-Based Classifiers

Fabricio A. Breve fabricio@dc.ufscar.br
Moacir P. Ponti Jr. moacir@dc.ufscar.br
Nelson D. A. Mascarenhas nelson@dc.ufscar.br

DC – Departamento de Computação
UFSCar - Universidade Federal de São Carlos, São Paulo, SP, Brasil