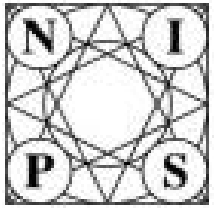
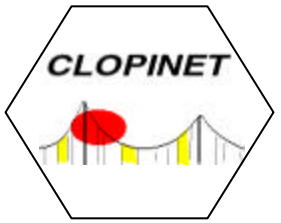



Neural Information Processing Systems Conference



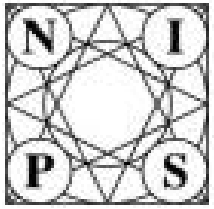

# RESULTS OF THE NIPS 2003 FEATURE SELECTION CHALLENGE

*Isabelle Guyon*

*Steve Gunn*

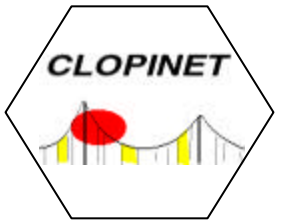
*Asa Ben Hur*

*Gideon Dror*

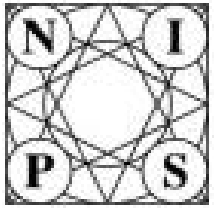



Neural Information Processing Systems Conference

# *Challenge*

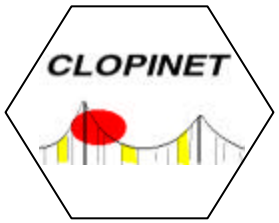



- **Date started**: Monday September 8, 2003.
- **Date ended**: Monday December 1, 2003 (+Dec. 8, entries using validation set labels).
- **Duration**: 12 (13) weeks.
- **Estimated number of entrants**: 78.
- **Number of development entries**: 1863.
- **Number of ranked participants**: 20 (16).
- **Number of ranked submissions**: 56 (36).




Neural Information Processing Systems Conference

# *Results*




Overall winners for ranked entries:

**Radford Neal and Jianguo Zhang**
**with BayesNN-DFT-combo (Dec 1 and 8)**

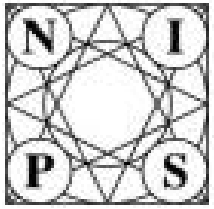
**Arcene**: (1) Neal&Zhang w. BayesNN-DFT-combo
(8) Radford Neal with BayesNN-small

**Dexter**: (1) Neal&Zhang w. BayesNN-DFT-combo
(8) Thomas Navin Lal with FS+SVM

**Dorothea**: (1&8) Neal&Zhang w. BayesNN-DFT-combo

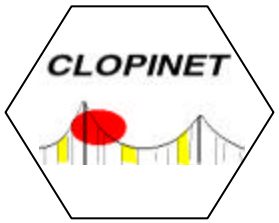
**Gisette**: (1&8) Yi-Wei Chen with final 2

**Madelon**: (1&8) Chu Wei with Bayesian + SVMs




Neural Information Processing Systems Conference

# *Part I*




# DATASET DESCRIPTION

# Domains

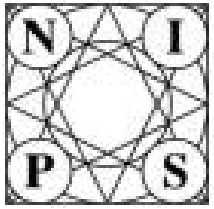


- **Arcene**: cancer vs. normal with mass-spectrometry analysis of blood serum.
- **Dexter**: filter texts about corporate acquisition from Reuters collection.
- **Dorothea**: predict which compounds bind to Thrombin from KDD cup 2001.
- **Gisette**: OCR digit “4” vs. digit “9” from NIST.
- **Madelon**: artificial data.

# *Data preparation*

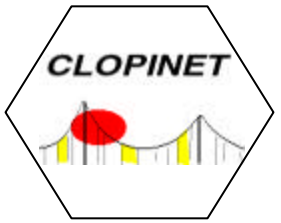


- **Preprocessing** and scaling to numerical range 0 to 999 for continuous data and 0/1 for binary data.
- **Probes**: Addition of "random" features distributed similarly to the real features.
- **Shuffling**: Randomization of the order of the patterns and the features.
- **Baseline error rates (errate)**: Training and testing on various data splits with simple methods.
- **Test set size**: Number of test examples needed using rule-of-thumb $\mathbf{n_{test} = 100/errate}$.




Neural Information Processing Systems Conference

# Data statistics




| Dataset | Size | Type | Features | Training Examples | Validation Examples | Test Examples |
|---|---|---|---|---|---|---|
| Arcene | 8.7 MB | Dense | 10000 | 100 | 100 | 700 |
| Gisette | 22.5 MB | Dense | 5000 | 6000 | 1000 | 6500 |
| Dexter | 0.9 MB | Sparse integer | 20000 | 300 | 300 | 2000 |
| Dorothea | 4.7 MB | Sparse binary | 100000 | 800 | 350 | 800 |
| Madelon | 2.9 MB | Dense | 500 | 2000 | 600 | 1800 |

# ARCENE

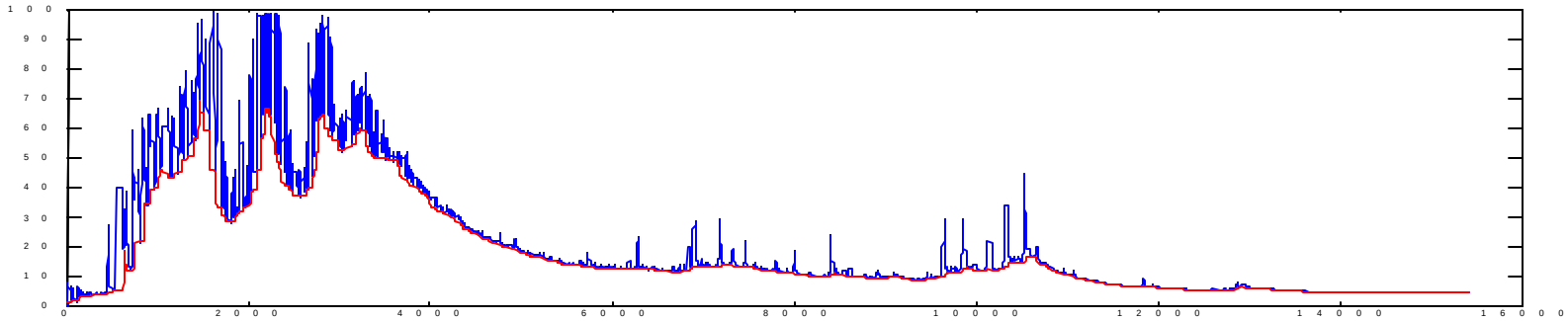
ARCENE is the **cancer** dataset

- **Sources**: National Cancer Institute (NCI) and Eastern Virginia Medical School (EVMS).
- **Three datasets**: 1 ovarian cancer, 2 prostate cancer, all preprocessed similarly.
- **Task**: Separate cancer *vs.* normal.

Neural Information Processing Systems Conference

# ARCENE

CLOPINET



- All SELDI mass-spectra.
- NCI ovarian cancer: 253 spectra (162 cancer, 91 control), 15154 feat.
- NCI prostate cancer: 322 spectra (69 cancer, 253 control), 15154 feat.

- EVMS prostate cancer: 652 spectra from 326 samples (167 cancer, 159 control), 48538 feat.

- Preprocessing including m/z 200-10000, baseline removal, alignment.

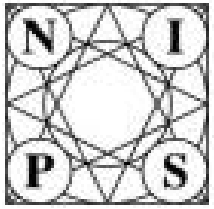
- Resulting dataset: 900 spectra (398 cancer, 502 control), 10000 features (7000 real features, 3000 random probes=permuted least-informative feat.).

- Rule-of-thumb: $n_{test}$=100/errate with errate=15% leads to 667 examples.

- Data split: Training 100, validation 100, test 700.

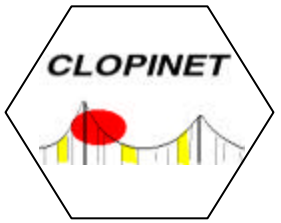
DEXTER filters **texts**

- **Sources**: Carnegie Group, Inc. and Reuters, Ltd.
- **Preprocessing**: Thorsten Joachims.
- **Task**: Filter “corporate acquisition” texts.




Neural Information Processing Systems Conference

# DEXTER




**NEW YORK, October 2, 2001 – Instinet Group Incorporated (Nasdaq: INET), the world's largest electronic agency securities broker, today announced that it has completed the acquisition of ProTrader Group, LP, a provider of advanced trading technologies and electronic brokerage services primarily for retail active traders and hedge funds. The acquisition excludes ProTrader's proprietary trading business. ProTrader's 2000 annual revenues exceeded $83 million.**

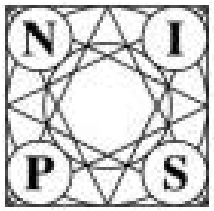
- 1300 texts about corporate acquisitions and 1300 texts about other topics.

- Bag-of-words representation prepared by Thorsten Joachims: 9947 features representing frequencies of occurrence of word stems in text.

- Probes: Added 10053 features drawn at random according to Zipf law.

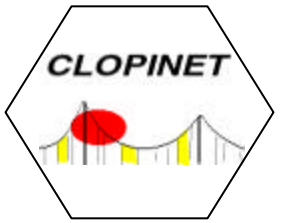
- Rule-of-thumb: $n_{test}$=100/errate with errate=5.8% leads to 1724 examples.

- Data split: Training 300, validation 300, test 2000.




Neural Information Processing Systems Conference

# *DOROTHEA*



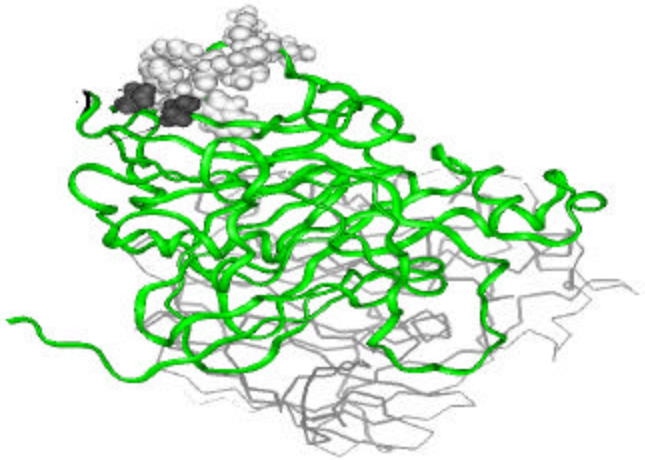

DOROTHEA is the **Thrombin** dataset

- **Sources**: DuPont Pharmaceuticals Research Laboratories and KDD Cup 2001.
- **Task**: Predict compounds that bind to Thrombin.

Neural Information Processing Systems Conference

# DOROTHEA

CLOPINET



- 2543 compounds tested for their ability to bind to a target site on thrombin, a key receptor in blood clotting; 192 “active” (bind well); the rest “inactive”.

- 139,351 binary features, which describe three-dimensional properties of the molecule.

- Preprocessing: Removed all-zero examples (except 1). Selected 100,000 features ranked with Weston et al. criterion, permuted randomly last 50,000 (probes).

- Rule-of-thumb: $n_{test}$=100/errate with errate=21% leads to 476 examples.

- Data split: Training 800, validation 350, test 800.

# GISETTE



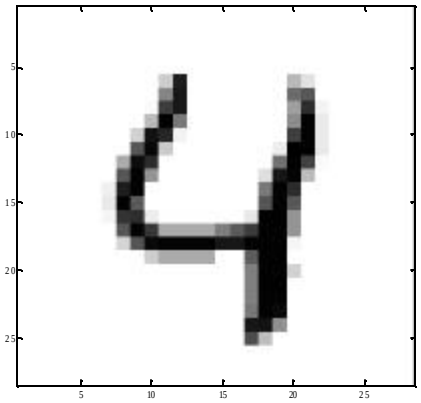
GISETTE contains handwritten **digits**

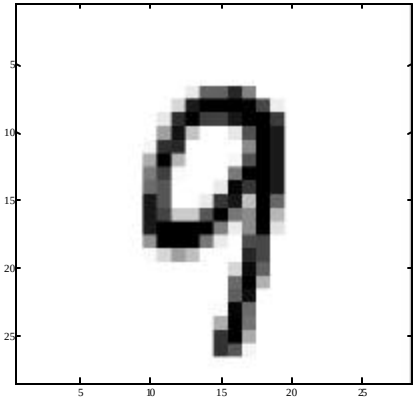
- **Source**: National Institute of Standards and Technologies (NIST).
- **Preprocessing**: Yann LeCun and collaborators.
- **Task**: Separate digits “4” and “9”.

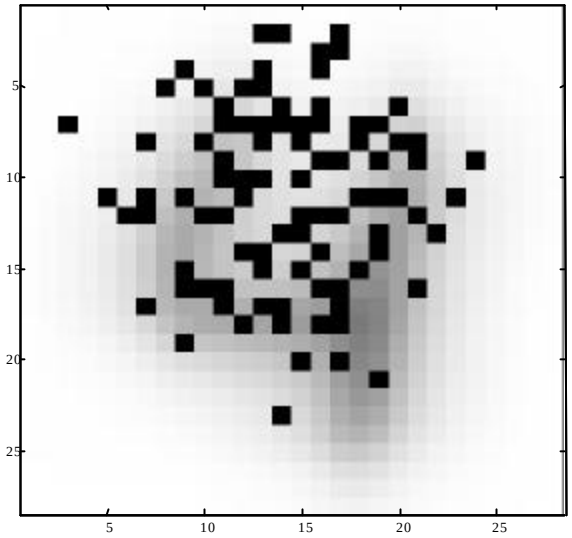
Neural Information Processing Systems Conference

# GISETTE

CLOPINET







- Original data: 13500 digits size-normalized and centered in a fixed-size image of dimension 28x28 .

- Constructed features: random selection of subset of products of pairs of variables.

- Feature set: 2500 features (pixels + pairs) + 2500 probes (permuted pairs).
- Rule-of-thumb: $n_{test}$=100/errate with errate=3.5% leads to 2857 examples.

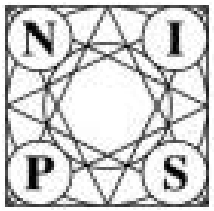
- Data split: Training 6000, validation 1000, test 6500.

# *MADELON*



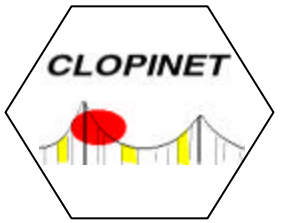
MADELON is **random** data

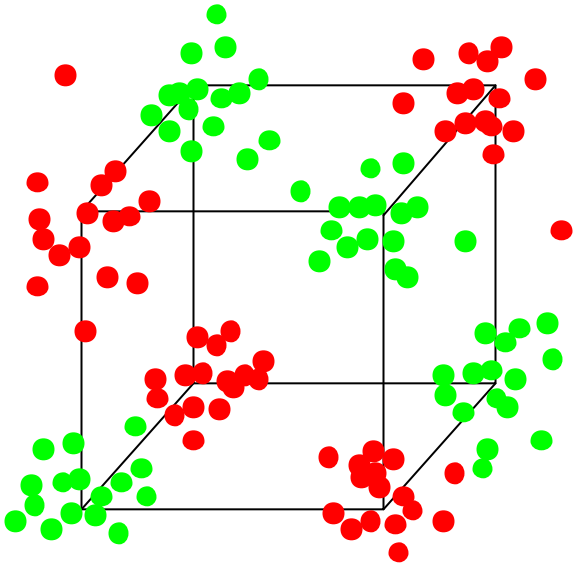
- **Source**: Isabelle Guyon, inspired by Simon Perkins et al.
- **Type of data**: Clusters on the summits of a hypercube.




Neural Information Processing Systems Conference

# MADELON



- Clusters placed on the summits of a five dimensional hypercube.

- 250 points per cluster; 16 clusters per class; 5 “useful” features; 5 “redundant” features; 10 “repeated” features; 480 “useless” features (probes).

- Rule-of-thumb: $n_{test}$=100/errate with errate=10% leads to 1000 examples.

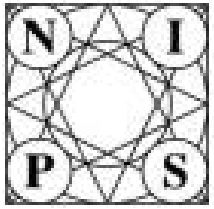
- Data split: Training 2000, validation 600, test 1800.

# Difficulties



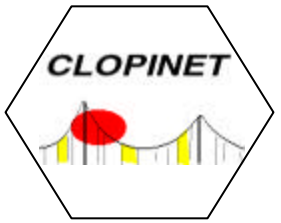
All 2-class classification problems.

| | Arcene | Dexter | Dorothea | Gisette | Madelon |
|---|---|---|---|---|---|
| **Sparsity** | **50%** | **99.5%** | **99%** | **87%** | **<1%** |
| **Binary** | **No** | **No** | **Yes** | **No (almost)** | **No** |
| **#feat / #patt** | **100** | **67** | **125** | **0.83** | **0.25** |
| **#probe / #feat** | **0.43** | **0.99** | **1** | **1** | **24** |
| **Cluster / class** | **>3** | **?** | **?** | **1-2?** | **16** |




Neural Information Processing Systems Conference

# *Part II*

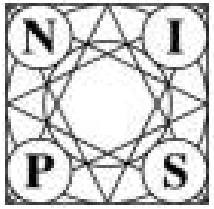



# SCORING METHOD

# Scoring steps

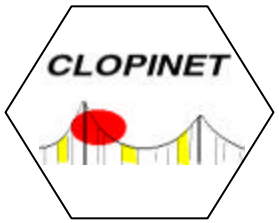


- Use test set results only (not training and validation set results).
- Make pairwise comparisons between classifiers for each dataset.
- Use McNemar test to determine whether A better than B according to BER with 5% risk. Score 1, 0 or –1.
- If score is 0, break tie with feature number if relative difference > 5%.
- If score still 0, break tie with fraction of probes.
- Overall score = sum of pairwise comparison scores.




Neural Information Processing Systems Conference

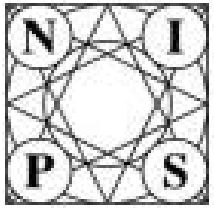
# *Observations*




- Positive and negative scores are obtained.
- Maximum score = num. submissions-1

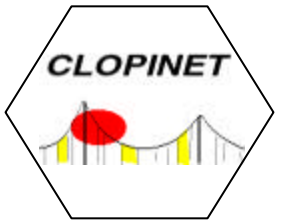
$\Rightarrow$ we normalize the score, then take the dataset average.

- Even a 0 score is good because we ranked only the 20 final participants / 75 total.
- Scoring/ranking is dependent on the set of submissions scored.
- The 5 top ranking people are consistently at the top and in the same order under changes of the set of submission.

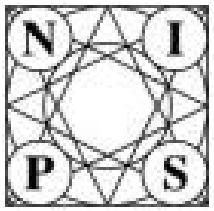



Neural Information Processing Systems Conference

# *Part III*

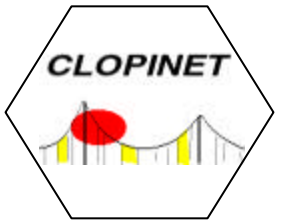



# ANALYSIS OF RESULTS

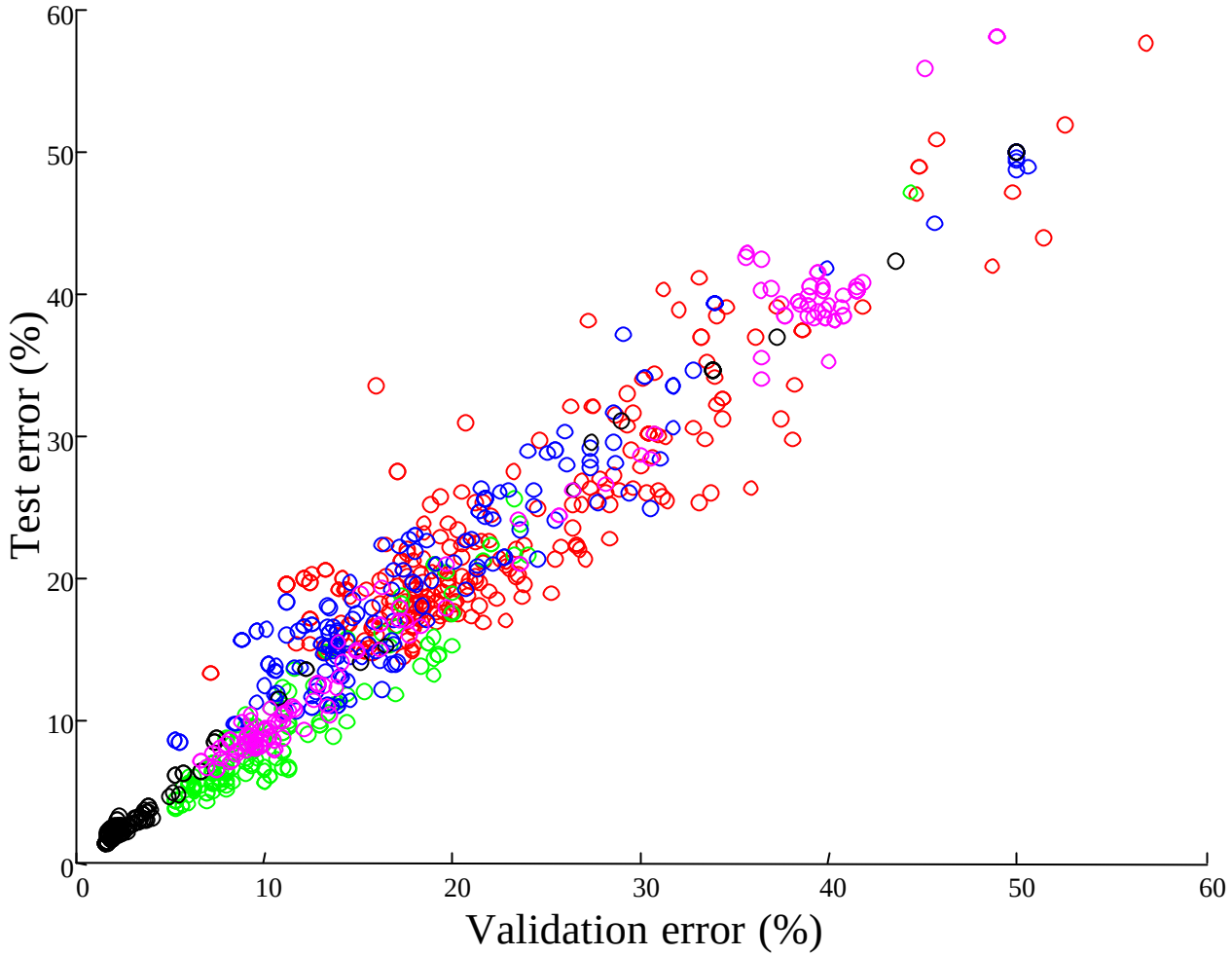



Neural Information Processing Systems Conference

# Test/Valid Correl



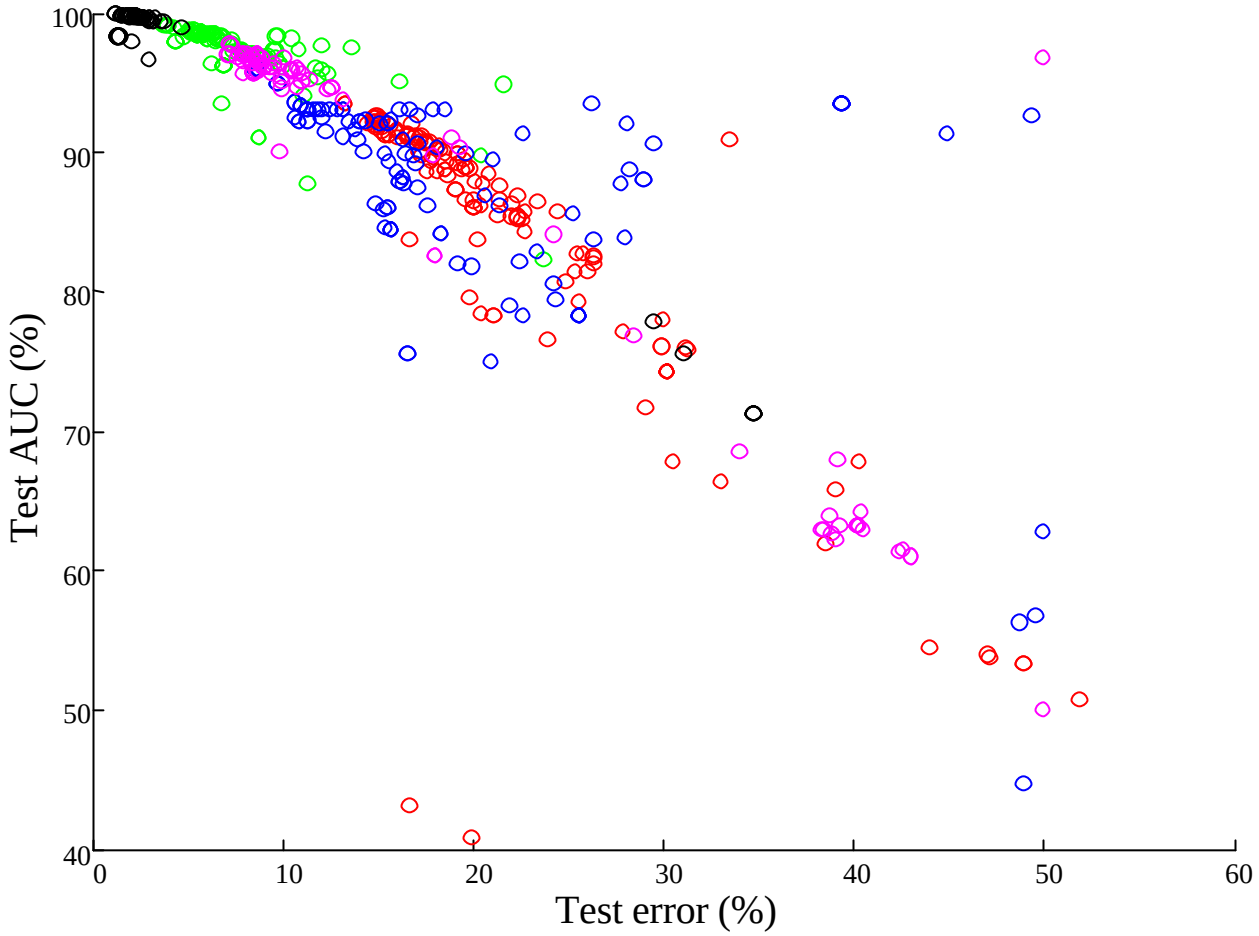

$R^2$%: Arcene 81.28, Dexter 94.37, Dorothea 93.11, Gisette 99.71, Madelon 98.62



# BER/AUC Correl

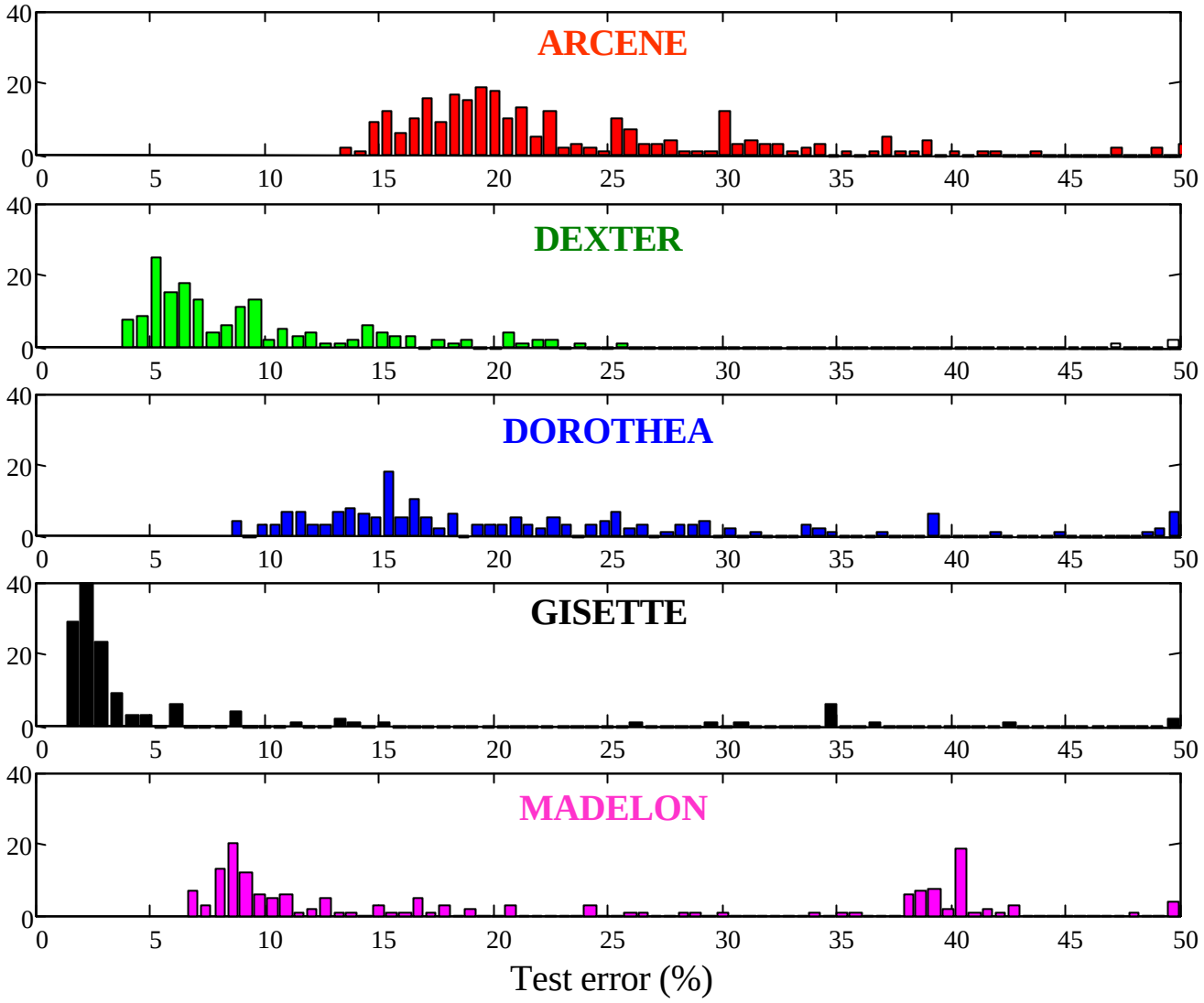


$R^2$%: Arcene 65.45, Dexter 53.5, Dorothea 29.57, Gisette 98.84, Madelon 89



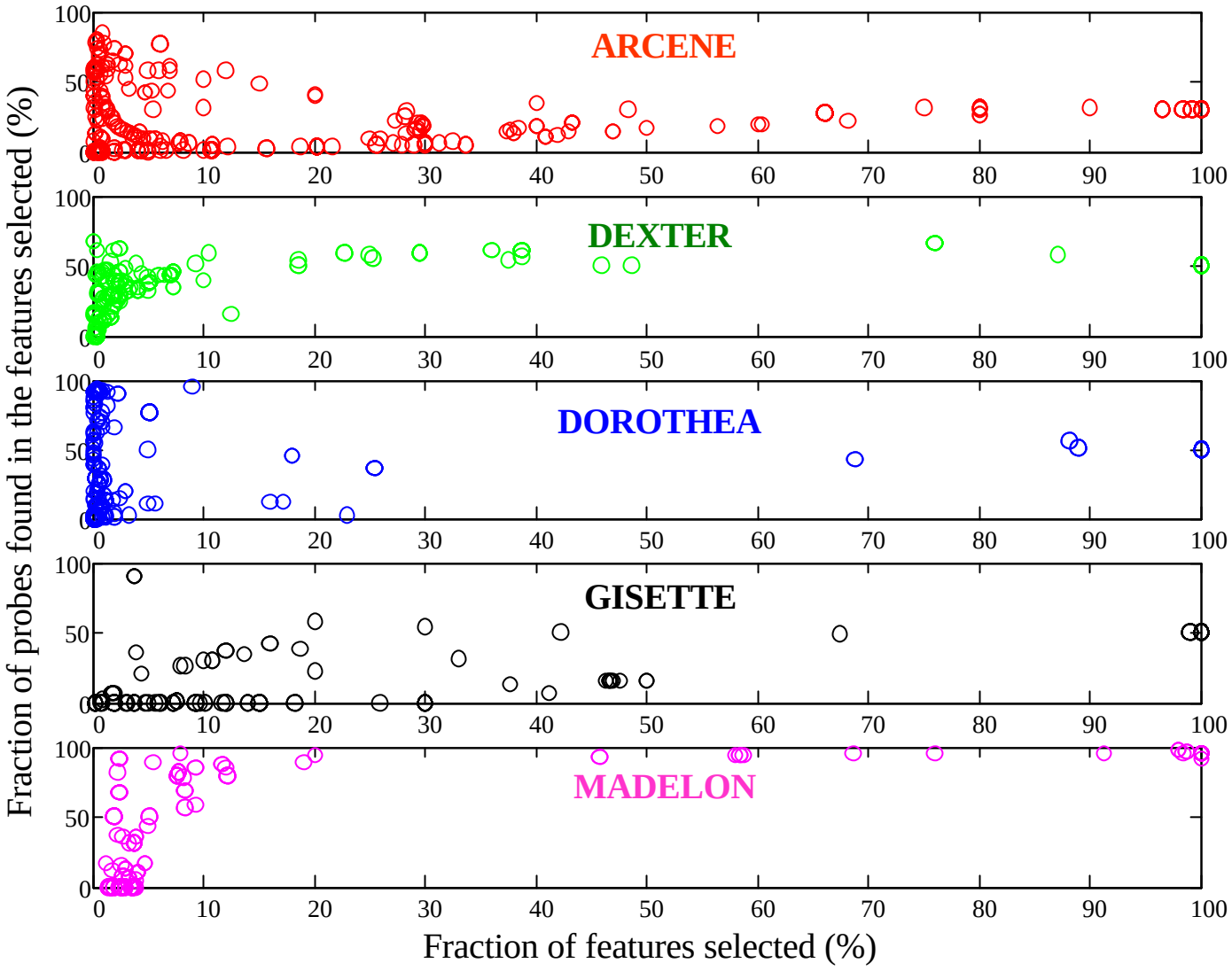
Neural Information Processing Systems Conference

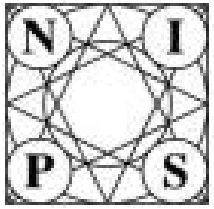
# BER distribution



Neural Information Processing Systems Conference

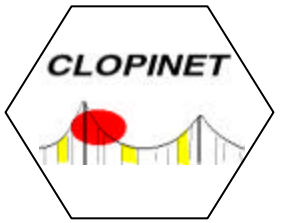
# Fraction of probes






Neural Information Processing Systems Conference

# *Global ranking*

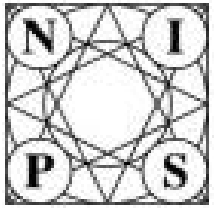



December 1st

| Method | People | Score | BER | AUC | Frac feat | Frac prob | McNemar |
|---|---|---|---|---|---|---|---|
| BayesNN-DFT-combo | Neal & Zhang | 88 | 6.84 (1) | 97.22 (1) | 80.3 | 47.77 | 0 |
| BayesNN-DFT-combo | Neal & Zhang | 86.18 | 6.87 (2) | 97.21 (2) | 80.3 | 47.77 | 0 |
| BayesNN-small | Neal | 68.73 | 8.20 (3) | 96.12 (3) | 4.74 | 2.91 | 0.8 |
| BayesNN-large | Neal | 59.64 | 8.21 (4) | 96.36 (4) | 60.3 | 28.51 | 0.4 |
| RF+RLSC | Torkkola & Tuv | 59.27 | 9.07 (7) | 90.93 (7) | 22.54 | 17.53 | 0.6 |
| final 2 | Chen | 52 | 9.31 (9) | 90.69 (9) | 24.91 | 11.98 | 0.4 |
| SVMbased3 | Zhili & Li | 41.82 | 9.21 (8) | 93.60 (8) | 29.51 | 21.72 | 0.8 |
| SVMBased4 | Zhili & Li | 41.09 | 9.40 (10) | 93.41 (10) | 29.51 | 21.72 | 0.8 |
| final 1 | Chen | 40.36 | 10.38 (23) | 89.62 (23) | 6.23 | 6.1 | 0.6 |
| transSVMbased2 | Zhili | 36 | 9.60 (13) | 93.21 (13) | 29.51 | 21.72 | 0.8 |
| myBestValidResult | Zhili | 36 | 9.60 (14) | 93.21 (14) | 29.51 | 21.72 | 0.8 |
| TransSVMbased | Zhili | 36 | 9.60 (15) | 93.21 (15) | 29.51 | 21.72 | 0.8 |
| BayesNN-E | Neal | 29.45 | 8.43 (5) | 96.30 (5) | 96.75 | 56.67 | 0.8 |
| Collection2 | Saffari | 28 | 10.03 (20) | 89.97 (20) | 7.71 | 10.6 | 1 |
| Collection1 | Saffari | 20.73 | 10.06 (21) | 89.94 (21) | 32.26 | 25.5 | 1 |

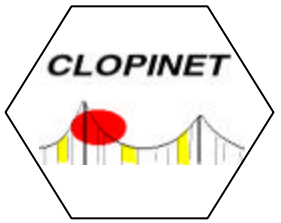
Neal and Zhang win in several respects: (1) best score, (2) best BER, (3) best AUC, (4) smallest feature set.

For BER and AUC, ranks are shown in parentheses. Score, BER, AUC, frac feat & prob are in %. McNemar tests the significance of the diff. in BER with the smallest BER.




Neural Information Processing Systems Conference

# *Global ranking*




December 8th

| Method | People | Score | BER | AUC | Frac feat | Frac prob | McNemar |
|---|---|---|---|---|---|---|---|
| BayesNN-DFT-combo+v | Neal & Zhang | 71.43 | 6.48 (1) | 97.20 (1) | 80.3 | 47.77 | 0.2 |
| BayesNN-large+v | Neal | 66.29 | 7.27 (3) | 96.98 (3) | 60.3 | 28.51 | 0.4 |
| BayesNN-small+v | Neal | 61.14 | 7.13 (2) | 97.08 (2) | 4.74 | 2.91 | 0.6 |
| final_2-3 | Chen | 49.14 | 7.91 (8) | 91.45 (8) | 24.91 | 9.91 | 0.4 |
| BayesNN-large+v | Neal | 49.14 | 7.83 (5) | 96.78 (5) | 60.3 | 28.51 | 0.6 |
| final2-2 | Chen | 40 | 8.80 (17) | 89.84 (17) | 24.62 | 6.68 | 0.6 |
| GhostMiner Pack 1 | GhostMiner Team | 37.14 | 7.89 (7) | 92.11 (7) | 80.6 | 36.05 | 0.8 |
| RF+RLSC | Torkkola & Tuv | 35.43 | 8.04 (9) | 91.96 (9) | 22.38 | 17.52 | 0.8 |
| GhostMiner Pack 2 | GhostMiner Team | 35.43 | 7.86 (6) | 92.14 (6) | 80.6 | 36.05 | 0.8 |
| RF+RLSC | Torkkola & Tuv | 34.29 | 8.23 (12) | 91.77 (12) | 22.38 | 17.52 | 0.6 |
| FS+SVM | Lal | 31.43 | 8.99 (19) | 91.01 (19) | 20.91 | 17.28 | 0.6 |
| GhostMiner Pack 3 | GhostMiner Team | 26.29 | 8.24 (13) | 91.76 (13) | 80.6 | 36.05 | 0.6 |
| CBAMethod3E | CBAGroup | 21.14 | 8.14 (10) | 96.62 (10) | 12.78 | 0.06 | 0.6 |
| CBAMethod3E | CBAGroup | 21.14 | 8.14 (11) | 96.62 (11) | 12.78 | 0.06 | 0.6 |
| Nameless | Navot & Bachrach | 12 | 7.78 (4) | 96.43 (4) | 32.28 | 16.22 | 1 |

Neal and Zhang win again: (1) best score, (2) best BER, (3) best AUC, (4) smallest feature set.

Neural Information Processing Systems Conference

# ARCENE

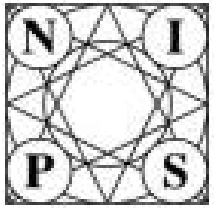
CLOPINET

Dec. 1st

| Method | People | Score | BER | AUC | Frac feat | Frac prob | McNemar |
|---|---|---|---|---|---|---|---|
| BayesNN-DFT-combo | Neal & Zhang | 98.18 | 13.30 (1) | 93.48 (1) | 100 | 30 | 0 |
| BayesNN-DFT-combo | Neal & Zhang | 98.18 | 13.30 (2) | 93.48 (2) | 100 | 30 | 0 |
| inf5 | Saffari | 85.45 | 17.30 (17) | 82.70 (17) | 5 | 0 | 1 |
| RF RLSC | Torkkola & Tuv | 81.82 | 15.14 (3) | 84.86 (3) | 100 | 30 | 0 |
| KPLS | Embrechts | 81.82 | 16.71 (12) | 83.67 (12) | 5.14 | 8.56 | 1 |
| BayesNN-small | Neal | 78.18 | 16.59 (10) | 91.15 (10) | 10.7 | 1.03 | 1 |
| Bayesian+SVM | Wei | 78.18 | 15.17 (4) | 91.52 (4) | 100 | 30 | 0 |
| final 2 | Chen | 74.55 | 15.27 (5) | 84.73 (5) | 100 | 30 | 0 |
| Bayesian+SVM | Wei | 70.91 | 15.55 (6) | 91.25 (6) | 100 | 30 | 0 |

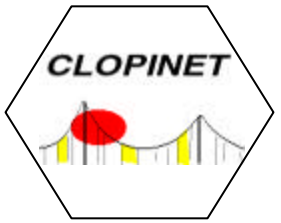
Dec. 8th

| Method | People | Score | BER | AUC | Frac feat | Frac prob | McNemar |
|---|---|---|---|---|---|---|---|
| BayesNN-small+v | Neal | 94.29 | 11.86 (7) | 95.47 (7) | 10.7 | 1.03 | 0 |
| RF w. feature select | Ng & Breiman | 88.57 | 12.63 (10) | 93.79 (10) | 3.8 | 0.79 | 1 |
| CBAMethod3E | CBAGroup | 85.71 | 11.12 (4) | 94.89 (4) | 28.25 | 0.28 | 0 |
| CBAMethod3E | CBAGroup | 85.71 | 11.12 (5) | 94.89 (5) | 28.25 | 0.28 | 0 |
| RF+RLSC | Torkkola & Tuv | 71.43 | 11.12 (3) | 88.88 (3) | 99.2 | 29.96 | 0 |
| final 2-2 | Chen | 68.57 | 10.73 (1) | 90.63 (1) | 100 | 30 | 0 |
| final 2-3 | Chen | 68.57 | 10.73 (2) | 90.63 (2) | 100 | 30 | 0 |
| FS+SVM | Lal | 65.71 | 12.76 (12) | 87.24 (12) | 47 | 5.89 | 1 |
| RF+RLSC | Torkkola & Tuv | 65.71 | 11.60 (6) | 88.40 (6) | 99.2 | 29.96 | 0 |
| BayesNN-DFT-combo+v | Neal & Zhang | 48.57 | 12.25 (8) | 93.01 (8) | 100 | 30 | 0 |




Neural Information Processing Systems Conference

# DEXTER

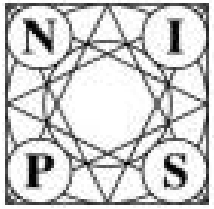



Dec. 1st

| Method | People | Score | BER | AUC | Frac feat | Frac prob | McNemar |
|---|---|---|---|---|---|---|---|
| BayesNN-DFT-combo | Neal & Zhang | 96.36 | 3.90 (1) | 99.01 (1) | 1.52 | 12.87 | 0 |
| BayesNN-large | Neal | 96.36 | 3.90 (2) | 99.01 (2) | 1.52 | 12.87 | 0 |
| BayesNN-DFT-combo | Neal & Zhang | 96.36 | 3.90 (3) | 99.01 (3) | 1.52 | 12.87 | 0 |
| BayesNN-small | Neal | 89.09 | 4.00 (4) | 99.03 (4) | 1.52 | 12.87 | 0 |
| FS+SVM | Lal | 85.45 | 4.20 (5) | 95.80 (5) | 18.57 | 49.78 | 0 |
| transSVMbased2 | Zhili | 70.91 | 4.40 (6) | 97.92 (6) | 29.47 | 59.71 | 0 |
| SVMbased3 | Zhili & Li | 70.91 | 4.40 (7) | 97.92 (7) | 29.47 | 59.71 | 0 |
| myBestValidResult | Zhili | 70.91 | 4.40 (8) | 97.92 (8) | 29.47 | 59.71 | 0 |
| TransSVMbased | Zhili | 70.91 | 4.40 (9) | 97.92 (9) | 29.47 | 59.71 | 0 |
| svmBased4 | Zhili & Li | 70.91 | 4.40 (10) | 97.92 (10) | 29.47 | 59.71 | 0 |

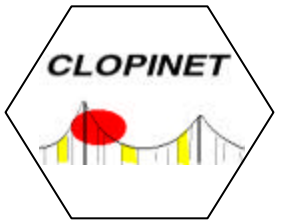
Dec. 8th

| Method | People | Score | BER | AUC | Frac feat | Frac prob | McNemar |
|---|---|---|---|---|---|---|---|
| FS+SVM | Lal | 100 | 3.30 (1) | 96.70 (1) | 18.57 | 42.14 | 0 |
| BayesNN-DFT-combo+v | Neal & Zhang | 85.71 | 4.05 (5) | 99.09 (5) | 1.52 | 12.87 | 1 |
| BayesNN-large+v | Neal | 85.71 | 4.05 (6) | 99.09 (6) | 1.52 | 12.87 | 1 |
| BayesNN-small+v | Neal | 85.71 | 4.05 (7) | 99.09 (7) | 1.52 | 12.87 | 1 |
| BayesNN-large+v | Neal | 85.71 | 4.05 (8) | 99.09 (8) | 1.52 | 12.87 | 1 |
| GhostMiner Pack 3 | GhostMiner | 71.43 | 3.50 (2) | 96.50 (2) | 100 | 50.27 | 0 |
| GhostMiner Pack 1 | GhostMiner | 65.71 | 3.60 (3) | 96.40 (3) | 100 | 50.27 | 1 |
| GhostMiner Pack 2 | GhostMiner | 54.29 | 3.80 (4) | 96.20 (4) | 100 | 50.27 | 1 |
| Sparse Bayes Logistic | DIMACS | 54.29 | 5.05 (14) | 94.37 (14) | 0.93 | 6.49 | 1 |
| RF+RLSC | Torkkola & Tuv | 48.57 | 4.65 (10) | 95.35 (10) | 2.5 | 28.4 | 1 |




Neural Information Processing Systems Conference

# DOROTHEA

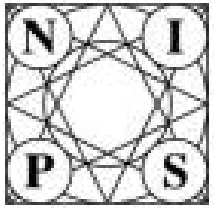



Dec. 1st

| Method | People | Score | BER | AUC | Frac feat | Frac prob | McNemar |
|---|---|---|---|---|---|---|---|
| BayesNN-DFT-combo | Neal & Zhang | 98.18 | 8.54 (1) | 95.92 (1) | 100 | 50 | 0 |
| BayesNN-large | Neal | 98.18 | 8.54 (2) | 95.92 (2) | 100 | 50 | 0 |
| BayesNN-E | Neal | 92.73 | 8.61 (3) | 95.98 (3) | 100 | 50 | 0 |
| BayesNN-DFT-combo | Neal & Zhang | 89.09 | 8.68 (4) | 95.86 (4) | 100 | 50 | 0 |
| greatest_hits_one | Navot & Bachrach | 85.45 | 10.86 (6) | 92.19 (6) | 0.3 | 0 | 1 |
| BayesNN-small | Neal | 81.82 | 10.63 (5) | 93.50 (5) | 0.5 | 0.4 | 1 |
| SVMbased3 | Zhili & Li | 78.18 | 11.52 (11) | 88.48 (11) | 0.5 | 18.88 | 1 |
| svmBased4 | Zhili & Li | 74.55 | 12.45 (12) | 87.55 (12) | 0.5 | 18.88 | 1 |

Dec. 8th

| Method | People | Score | BER | AUC | Frac feat | Frac prob | McNemar |
|---|---|---|---|---|---|---|---|
| BayesNN-DFT-combo+v | Neal & Zhang | 97.14 | 8.61 (1) | 95.92 (1) | 100 | 50 | 0 |
| BayesNN-large+v | Neal | 97.14 | 8.61 (2) | 95.92 (2) | 100 | 50 | 0 |
| IDEAL | BorisovEruhimovTuv | 85.71 | 8.92 (3) | 94.80 (3) | 100 | 50 | 0 |
| IDEAL | BorisovEruhimovTuv | 85.71 | 8.92 (4) | 94.80 (4) | 100 | 50 | 0 |
| BayesNN-large+v | Neal | 77.14 | 9.11 (5) | 95.98 (5) | 100 | 50 | 0 |
| A shot in the dark | Navot & Bachrach | 68.57 | 11.40 (7) | 93.10 (7) | 0.4 | 0 | 1 |
| Nameless | Navot & Bachrach | 68.57 | 11.40 (8) | 93.10 (8) | 0.4 | 0 | 1 |
| BayesNN-small+v | Neal | 60 | 11.07 (6) | 93.42 (6) | 0.5 | 0.4 | 1 |
| ESNB+NN | Boulle & Lemaire | 54.29 | 14.59 (17) | 91.50 (17) | 0.07 | 0 | 1 |




Neural Information Processing Systems Con

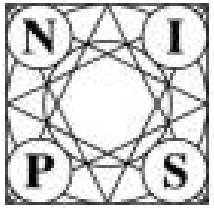
# GISETTE


Dec. 1st

| Method | People | Score | BER | AUC | Frac feat | Frac prob | McNemar |
|---|---|---|---|---|---|---|---|
| final 2 | Chen | 98.18 | 1.37 (8) | 98.63 (8) | 18.26 | 0 | 0 |
| final 1 | Chen | 98.18 | 1.37 (9) | 98.63 (9) | 18.26 | 0 | 0 |
| Depends II | Rosset & Zhu | 87.27 | 1.34 (4) | 98.26 (4) | 30 | 0 | 0 |
| Depends I | Rosset & Zhu | 87.27 | 1.34 (5) | 98.26 (5) | 30 | 0 | 0 |
| Depends III | Rosset & Zhu | 87.27 | 1.34 (6) | 98.26 (6) | 30 | 0 | 0 |
| Depends V | Rosset & Zhu | 87.27 | 1.34 (7) | 98.26 (7) | 30 | 0 | 0 |
| BayesNN-DFT-combo | Neal & Zhang | 70.91 | 1.29 (1) | 99.90 (1) | 100 | 50 | 0 |
| BayesNN-large | Neal | 70.91 | 1.29 (2) | 99.90 (2) | 100 | 50 | 0 |
| BayesNN-DFT-combo | Neal & Zhang | 70.91 | 1.29 (3) | 99.90 (3) | 100 | 50 | 0 |
| transSVMbased2 | Zhili | 56.36 | 1.58 (11) | 99.84 (11) | 15 | 0 | 1 |
| SVMbased3 | Zhili & Li | 56.36 | 1.58 (12) | 99.84 (12) | 15 | 0 | 1 |
| myBestValidResult | Zhili | 56.36 | 1.58 (13) | 99.84 (13) | 15 | 0 | 1 |
| Depends IV | Rosset & Zhu | 56.36 | 1.48 (10) | 98.26 (10) | 30 | 0 | 0 |
| TransSVMbased | Zhili | 56.36 | 1.58 (14) | 99.84 (14) | 15 | 0 | 1 |
| svmBased4 | Zhili & Li | 56.36 | 1.58 (15) | 99.84 (15) | 15 | 0 | 1 |

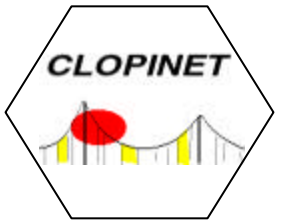
Dec. 8th

| Method | People | Score | BER | AUC | Frac feat | Frac prob | McNemar |
|---|---|---|---|---|---|---|---|
| final2 2 | Chen | 97.14 | 1.35 (7) | 98.71 (7) | 18.32 | 0 | 0 |
| final 2-3 | Chen | 97.14 | 1.35 (8) | 98.71 (8) | 18.32 | 0 | 0 |
| test | Chen | 88.57 | 1.37 (9) | 98.63 (9) | 18.26 | 0 | 0 |
| FS+SVM | Lal | 82.86 | 1.31 (6) | 98.69 (6) | 34 | 0.18 | 0 |
| BayesNN-DFT-combo+v | Neal & Zhang | 71.43 | 1.26 (1) | 99.92 (1) | 100 | 50 | 0 |
| BayesNN-large+v | Neal | 71.43 | 1.26 (2) | 99.92 (2) | 100 | 50 | 0 |
| BayesNN-large+v | Neal | 71.43 | 1.26 (3) | 99.92 (3) | 100 | 50 | 0 |
| GhostMiner Pack 1 | GhostMiner | 57.14 | 1.31 (4) | 98.69 (4) | 100 | 50 | 0 |
| GhostMiner Pack 3 | GhostMiner | 57.14 | 1.31 (5) | 98.69 (5) | 100 | 50 | 0 |
| P-SVM / nu-SVM 2 | Hochreiter | 37.14 | 1.82 (19) | 99.79 (19) | 4 | 0.5 | 1 |
| P-SVM / nu-SVM 1 | Hochreiter | 37.14 | 1.82 (20) | 99.79 (20) | 4 | 0.5 | 1 |
| P-SVM / nu-SVM too many | Hochreiter | 37.14 | 1.82 (21) | 99.79 (21) | 4 | 0.5 | 1 |
| GhostMiner Pack 2 | GhostMiner | 25.71 | 1.42 (10) | 98.58 (10) | 100 | 50 | 0 |




Neural Information Processing Systems Conference

# MADELON

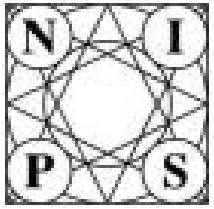



Dec. 1st

| Method | People | Score | BER | AUC | Frac feat | Frac prob | McNemar |
|---|---|---|---|---|---|---|---|
| Bayesian+SVM | Wei | 100 | 7.17 (5) | 96.95 (5) | 1.6 | 0 | 0 |
| RF+RLSC | Torkkola & Tuv | 96.36 | 6.67 (3) | 93.33 (3) | 3.8 | 0 | 0 |
| final 2 | Chen | 90.91 | 6.61 (1) | 93.39 (1) | 4.8 | 16.67 | 0 |
| final 1 | Chen | 90.91 | 6.61 (2) | 93.39 (2) | 4.8 | 16.67 | 0 |
| BayesNN-DFT-combo | Neal & Zhang | 76.36 | 7.17 (4) | 97.82 (4) | 100 | 96 | 0 |
| P-SVM/nu-SVM | Hochreiter | 76.36 | 8.67 (20) | 96.46 (20) | 1.4 | 0 | 1 |
| BayesNN-DFT-combo | Neal & Zhang | 76.36 | 7.17 (6) | 97.82 (6) | 100 | 96 | 0 |
| P-SVM/nu-SVM | Hochreiter | 76.36 | 8.67 (21) | 96.46 (21) | 1.4 | 0 | 1 |

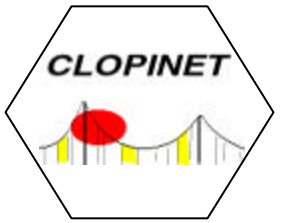
Dec. 8th

| Method | People | Score | BER | AUC | Frac feat | Frac prob | McNemar |
|---|---|---|---|---|---|---|---|
| Bayesian + SVMs | Wei | 94.29 | 7.11 (13) | 96.95 (13) | 1.6 | 0 | 1 |
| BayesNN-large+v | Neal | 85.71 | 6.56 (3) | 97.62 (3) | 3.4 | 0 | 0 |
| BayesNN-small+v | Neal | 85.71 | 6.56 (4) | 97.62 (4) | 3.4 | 0 | 0 |
| final 2-2 | Chen | 71.43 | 7.11 (12) | 92.89 (12) | 3.2 | 0 | 1 |
| RF+RLSC | Torkkola & Tuv | 71.43 | 6.67 (6) | 93.33 (6) | 3.8 | 0 | 0 |
| GhostMiner Pack 2 | GhostMiner | 65.71 | 7.44 (14) | 92.56 (14) | 3 | 0 | 1 |
| BayesNN-large+v | Neal | 60 | 6.78 (9) | 97.46 (9) | 3.4 | 0 | 1 |
| BayesNN-DFT-combo+ | Neal & Zhang | 54.29 | 6.22 (1) | 98.07 (1) | 100 | 96 | 0 |
| CBAMethod3E | CBAGroup | 51.43 | 6.72 (7) | 97.57 (7) | 4 | 0 | 0 |
| CBAMethod3E | CBAGroup | 51.43 | 6.72 (8) | 97.57 (8) | 4 | 0 | 0 |
| final_2 3 | Chen | 48.57 | 6.50 (2) | 93.50 (2) | 4.8 | 16.67 | 0 |
| RF+RLSC | Torkkola & Tuv | 48.57 | 7.00 (11) | 93.00 (11) | 3.8 | 0 | 1 |
| METHOD2 | CBAGroup | 42.86 | 6.83 (10) | 97.23 (10) | 4 | 0 | 0 |
| GhostMiner Pack 1 | GhostMiner | 37.14 | 7.67 (17) | 92.33 (17) | 3 | 0 | 1 |
| test | Chen | 31.43 | 6.61 (5) | 93.39 (5) | 4.8 | 16.67 | 0 |




Neural Information Processing Systems Conference

# *Part IV*




# CONCLUSIONS AND FURTHER WORK

# Conclusions



- Excellent results with no feature selection.
- In most cases, feature selection either helps or does not hurt.
- Wide variety of methods used.

# Future work



- Final check of feature set validity.
- Compute (more) statistics and write report.
- Publish the proceedings as a book.
- Use the challenge results as a benchmark: no release of test labels, leave the web site life.
- Organize another benchmark: Model selection? Feature selection for multi-class and regression? Feature construction?