# AD-NEv: A Scalable Multi-Level Neuroevolution Framework for Multivariate Anomaly Detection

Marcin Pietroń, Dominik Żurek, Kamil Faber, and Roberto Corizzo, *Member, IEEE*

## I. DATASETS

TABLE I: Statistics of the datasets

| Dataset | Size of train datasets during evolution | Epochs evolution | Epochs final training |
|---|---|---|---|
| SWAT | 9 933 | 25 | 90 |
| WADI-2019 | 156 914 | 25 | 120 |
| SMAP | 138 004 | 25 | 70 |
| MSl | 58 317 | 25 | 70 |

## II. FINAL SUBSPACES AND MODELS

1) **SWAT** - [[5, 6, 7, 25, 26, 34, 36, 45], [0, 1, 2, 3, 8, 9, 12, 14, 16, 17, 20, 24, 43, 50], [4, 18, 21, 22, 23, 27, 28, 32, 33, 35, 37, 38, 39, 40, 41, 42, 44, 46, 49], [10, 11, 13, 15, 29], [19, 30, 31, 47, 48]]
2) **WADI** - [[8, 12, 20, 21, 29, 30, 41, 44, 52, 64, 67, 69, 75, 78, 87, 111, 112, 115, 119], [3, 4, 5, 6, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 24, 27, 33, 36, 38, 40, 42, 43, 45, 46, 47, 49, 60, 61, 70, 72, 73, 81, 82, 83, 85, 86, 88, 89, 90, 96, 98, 118], [48, 50, 53, 54, 58, 68, 71, 72, 84, 94, 102, 105, 107, 108, 113, 116, 117], [1, 22, 23, 25, 26, 28, 31, 32, 34, 35, 37, 39, 59, 62, 63, 65, 66, 76, 77, 79, 80, 92, 93, 95, 106, 109, 110], [0, 2, 7, 51, 57, 74, 91, 97, 99, 100, 101, 103, 104, 114]]
3) **SMAP** - [[0],[2, 5, 6, 7, 12, 13, 14, 15, 17, 18, 19, 21, 22, 23], [10], [1, 3, 4, 9, 16, 20, 24], [11]]
4) **MSL** - [[0], [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 39, 40, 41, 42, 44, 45, 47, 48, 49, 50, 51], [43, 52], [24, 53, 54], [20, 25, 38, 46]]

TABLE II: Hyperparameters of the most effective CNN 1D model for the SWAT dataset (learning rate = 0.01). **C1D** means Conv1D layer, while **BN1D** means Batch Norm 1D layer

| | Layer | Parameters | | | Layer | Parameters |
|---|---|---|---|---|---|---|
| 1 | C1D | $F[0]_{ic} = 5$ $F[0]_{oc} = 84$ $F[0]_K = 2$ $F[0]_P = 1$ | | 7 | C1D | $F[3]_{ic} = 205$ $F[3]_{oc} = 123$ $F[3]_K = 4$ $F[3]_P = 1$ |
| 2 | BN1D | $F[0]_c^B = 84$ | | 8 | BN1D | $F[3]_c^B = 123$ |
| 3 | C1D | $F[1]_{ic} = 84$ $F[1]_{oc} = 123$ $F[1]_K = 6$ $F[1]_P = 1$ | | 9 | C1D | $F[4]_{ic} = 123$ $F[4]_{oc} = 84$ $F[4]_K = 6$ $F[4]_P = 1$ |
| 4 | BN1D | $F[1]_c^B = 123$ | | 10 | BN1D | $F[4]_c^B = 84$ |
| 5 | C1D | $F[2]_{ic} = 123$ $F[2]_{oc} = 205$ $F[2]_K = 4$ $F[2]_P = 1$ | | 11 | C1D | $F[5]_{ic} = 84$ $F[5]_{ic} = 5$ $F[5]_K = 2$ $F[5]_P = 1$ |
| 6 | BN1D | $F[2]_c^B = 205$ | | 12 | BN1D | $F[5]_c^B = 5$ |

TABLE III: Hyperparameters of the most effective CNN 1D model for the WADI dataset (learning rate = 0.01). **C1D** means Conv1D layer, while **BN1D** means Batch Norm 1D layer

| | Layer | Parameters | | | Layer | Parameters |
|---|---|---|---|---|---|---|
| 1 | C1D | $F[0]_{ic} = 6$ $F[0]_{oc} = 91$ $F[0]_K = 7$ $F[0]_P = 1$ | | 7 | C1D | $F[3]_{ic} = 155$ $F[3]_{oc} = 153$ $F[3]_K = 4$ $F[3]_P = 1$ |
| 2 | BN1D | $F[0]_c^B = 91$ | | 8 | BN1D | $F[3]_c^B = 153$ |
| 3 | C1D | $F[1]_{ic} = 91$ $F[1]_{oc} = 153$ $F[1]_K = 3$ $F[1]_P = 1$ | | 9 | C1D | $F[4]_{ic} = 153$ $F[4]_{oc} = 91$ $F[4]_K = 3$ $F[4]_P = 1$ |
| 4 | BN1D | $F[1]_c^B = 153$ | | 10 | BN1D | $F[4]_c^B = 91$ |
| 5 | C1D | $F[2]_{ic} = 153$ $F[2]_{oc} = 155$ $F[2]_K = 4$ $F[2]_P = 1$ | | 11 | C1D | $F[5]_{ic} = 91$ $F[5]_{oc} = 6$ $F[5]_K = 7$ $F[5]_P = 1$ |
| 6 | BN1D | $F[2]_c^B = 155$ | | 12 | BN1D | $F[5]_c^B = 6$ |

Marcin Pietroń, Dominik Żurek, and Kamil Faber are with the Institute of Computer Science at AGH University of Science and Technology, Poland. Roberto Corizzo is with the Department of Computer Science at American University in Washington DC. Marcin Pietroń is a corresponding author: pietron@agh.edu.pl

TABLE IV: Hyperparameters of the most effective CNN 1D model for the MSL dataset (learning rate = 0.00001). **C1D** means Conv1D layer, while **BN1D** means Batch Norm 1D layer

| | Layer | Parameters | | Layer | Parameters |
|---|---|---|---|---|---|
| 1 | C1D | $F[0]_{ic} = 4$ <br> $F[0]_{oc} = 95$ <br> $F[0]_K = 7$ <br> $F[0]_P = 3$ | 13 | C1D | $F[6]_{ic} = 927$ <br> $F[6]_{oc} = 685$ <br> $F[6]_K = 2$ <br> $F[6]_P = 1$ |
| 2 | BN1D | $F[0]_c^B = 95$ | 14 | BN1D | $F[6]_c^B = 685$ |
| 3 | C1D | $F[1]_{ic} = 95$ <br> $F[1]_{oc} = 126$ <br> $F[1]_K = 2$ <br> $F[1]_P = 1$ | 15 | C1D | $F[7]_{ic} = 685$ <br> $F[7]_{oc} = 261$ <br> $F[7]_K = 3$ <br> $F[7]_P = 1$ |
| 4 | BN1D | $F[1]_c^B = 126$ | 16 | BN1D | $F[7]_c^B = 261$ |
| 5 | C1D | $F[2]_{ic} = 126$ <br> $F[2]_{oc} = 138$ <br> $F[2]_K = 4$ <br> $F[2]_P = 1$ | 17 | C1D | $F[8]_{ic} = 261$ <br> $F[8]_{oc} = 138$ <br> $F[8]_K = 3$ <br> $F[8]_P = 1$ |
| 6 | BN1D | $F[2]_c^B = 138$ | 18 | BN1D | $F[8]_c^B = 138$ |
| 7 | C1D | $F[3]_{ic} = 138$ <br> $F[3]_{oc} = 261$ <br> $F[3]_K = 3$ <br> $F[3]_P = 1$ | 19 | C1D | $F[9]_{ic} = 138$ <br> $F[9]_{oc} = 126$ <br> $F[9]_{ic} = 4$ <br> $F[9]_{ic} = 1$ |
| 8 | BN1D | $F[3]_c^B = 261$ | 20 | BN1D | $F[9]_c^B = 126$ |
| 9 | C1D | $F[4]_{ic} = 261$ <br> $F[4]_{oc} = 685$ <br> $F[4]_K = 3$ <br> $F[4]_P = 1$ | 21 | C1D | $F[10]_{ic} = 126$ <br> $F[10]_{oc} = 95$ <br> $F[10]_K = 2$ <br> $F[10]_P = 1$ |
| 10 | BN1D | $F[4]_c^B = 685$ | 22 | BN1D | $F[10]_c^B = 95$ |
| 11 | C1D | $F[5]_{ic} = 685$ <br> $F[5]_{oc} = 927$ <br> $F[5]_K = 2$ <br> $F[5]_P = 1$ | 23 | C1D | $F[11]_{ic} = 95$ <br> $F[11]_{oc} = 4$ <br> $F[11]_K = 7$ <br> $F[11]_P = 3$ |
| 12 | BN1D | $F[5]_c^B = 927$ | 24 | BN1D | $F[11]_c^B = 4$ |

TABLE V: Hyperparameters of the most effective CNN 1D model for the SMAP dataset learning rate = 0.000001). **C1D** means Conv1D layer, while **BN1D** means Batch Norm 1D layer

| | Layer | Parameters | | Layer | Parameters |
|---|---|---|---|---|---|
| 1 | C1D | $F[0]_{ic} = 5$ <br> $F[0]_{oc} = 75$ <br> $F[0]_K = 3$ <br> $F[0]_P = 1$ | 13 | C1D | $F[6]_{ic} = 1041$ <br> $F[6]_{oc} = 694$ <br> $F[6]_K = 2$ <br> $F[6]_P = 1$ |
| 2 | BN1D | $F[0]_c^B = 75$ | 14 | BN1D | $F[6]_c^B = 694$ |
| 3 | C1D | $F[1]_{ic} = 75$ <br> $F[1]_{oc} = 70$ <br> $F[1]_K = 2$ <br> $F[1]_P = 1$ | 15 | C1D | $F[7]_{ic} = 694$ <br> $F[7]_{oc} = 437$ <br> $F[7]_K = 2$ <br> $F[7]_P = 1$ |
| 4 | BN1D | $F[1]_c^B = 70$ | 16 | BN1D | $F[7]_c^B = 437$ |
| 5 | C1D | $F[2]_{ic} = 70$ <br> $F[2]_{oc} = 273$ <br> $F[2]_K = 5$ <br> $F[2]_P = 3$ | 17 | C1D | $F[8]_{ic} = 437$ <br> $F[8]_{oc} = 273$ <br> $F[8]_K = 3$ <br> $F[8]_P = 1$ |
| 6 | BN1D | $F[2]_c^B = 273$ | 18 | BN1D | $F[8]_c^B = 273$ |
| 7 | C1D | $F[3]_{ic} = 273$ <br> $F[3]_{oc} = 437$ <br> $F[3]_K = 3$ <br> $F[3]_P = 1$ | 19 | C1D | $F[9]_{ic} = 273$ <br> $F[9]_{oc} = 70$ <br> $F[9]_{ic} = 5$ <br> $F[9]_{ic} = 3$ |
| 8 | BN1D | $F[3]_c^B = 437$ | 20 | BN1D | $F[9]_c^B = 70$ |
| 9 | C1D | $F[4]_{ic} = 437$ <br> $F[4]_{oc} = 694$ <br> $F[4]_K = 2$ <br> $F[4]_P = 1$ | 21 | C1D | $F[10]_{ic} = 70$ <br> $F[10]_{oc} = 75$ <br> $F[10]_K = 2$ <br> $F[10]_P = 1$ |
| 10 | BN1D | $F[4]_c^B = 694$ | 22 | BN1D | $F[10]_c^B = 75$ |
| 11 | C1D | $F[5]_{ic} = 694$ <br> $F[5]_{oc} = 1041$ <br> $F[5]_K = 2$ <br> $F[5]_P = 1$ | 23 | C1D | $F[11]_{ic} = 75$ <br> $F[11]_{oc} = 5$ <br> $F[11]_K = 3$ <br> $F[11]_P = 1$ |
| 12 | BN1D | $F[5]_c^B = 1041$ | 24 | BN1D | $F[11]_c^B = 5$ |

## III. ADDITIONAL RESULTS AND CONVERGENCE PLOTS

TABLE VI: F-1 score for all entities from SMAP datasets, 1-$F_1$ score, 2-$F_1$ score **after two steps**, 3-$F_1$ score **fine-tuned**
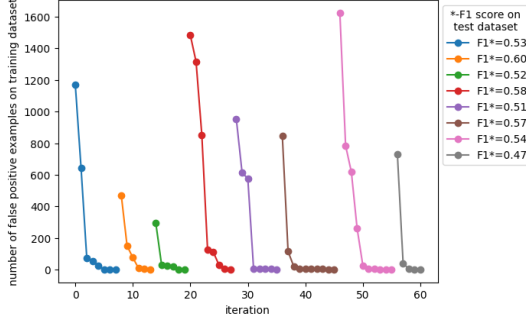
| Entity Name | 1 | 2 | 3 | Entity Name | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|
| P-1 | 0.15 | 0.15 | 0.17 | D-6 | 0.05 | 0.14 | 0.29 |
| S-1 | 0.16 | 0.71 | 0.74 | D-7 | 0.59 | 0.98 | 0.98 |
| E-1 | 0.20 | 0.21 | 0.21 | F-1 | 0.0 | 0.0 | 0.07 |
| E-2 | 0.46 | 0.51 | 0.72 | P-4 | 0.10 | 0.16 | 0.16 |
| E-3 | 0.13 | 0.14 | 0.36 | G-3 | 0.03 | 0.13 | 0.14 |
| E-4 | 0.50 | 0.54 | 0.64 | T-1 | 0.33 | 0.56 | 0.82 |
| E-5 | 0.07 | 0.07 | 0.08 | T-2 | 0.50 | 0.56 | 0.69 |
| E-6 | 0.10 | 0.11 | 0.38 | D-8 | 0.04 | 0.15 | 0.15 |
| E-7 | 0.07 | 0.32 | 0.34 | D-9 | 0.83 | 0.88 | 0.88 |
| E-8 | 0.27 | 0.44 | 0.53 | F-2 | 0.52 | 0.66 | 0.72 |
| E-9 | 0.11 | 0.28 | 0.30 | G-4 | 0.25 | 0.29 | 0.29 |
| E-10 | 0.21 | 0.21 | 0.28 | T-3 | 0.05 | 0.09 | 0.10 |
| E-11 | 0.21 | 0.27 | 0.27 | D-11 | 0.0 | 0.0 | 0.0 |
| E-12 | 0.12 | 0.13 | 0.22 | D-12 | 0.94 | 0.96 | 0.96 |
| E-13 | 0.06 | 0.06 | 0.07 | B-1 | 0.03 | 0.03 | 0.03 |
| A-1 | 0.0 | 0.10 | 0.10 | G-6 | 0.25 | 0.62 | 0.64 |
| D-1 | 0.77 | 0.94 | 0.96 | G-7 | 0.32 | 0.45 | 0.51 |
| P-3 | 0.31 | 0.62 | 0.63 | P-7 | 0.42 | 0.67 | 0.73 |
| D-2 | 0.84 | 0.95 | 0.95 | R-1 | 0.03 | 0.09 | 0.09 |
| D-3 | 0.68 | 0.95 | 0.97 | A-5 | 0.71 | 0.75 | 0.77 |
| D-4 | 0.69 | 0.98 | 0.99 | A-6 | 0.83 | 0.84 | 0.88 |
| A-2 | 0.0 | 0.0 | 0.10 | A-7 | 0.53 | 0.54 | 0.54 |
| A-3 | 0.0 | 0.0 | 0.15 | D-13 | 0.09 | 0.11 | 0.11 |
| A-4 | 0.08 | 0.19 | 0.22 | P-2 | 0.39 | 0.77 | 0.79 |
| G-1 | 0.0 | 0.0 | 0.17 | A-8 | 0.08 | 0.42 | 0.89 |
| G-2 | 0.0 | 0.19 | 0.19 | A-9 | 0.61 | 0.98 | 0.99 |
| D-5 | 0.20 | 0.21 | 0.68 | F-3 | 0.23 | 0.49 | 0.49 |

TABLE VII: F-1 score for all entities from MSL datasets, 1-$F_1$ score, 2-$F_1$ score **after two steps**, 3-$F_1$ score **fine-tuned**

| Entity Name | 1 | 2 | 3 | Entity Name | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|
| M-6 | 0.15 | 0.91 | 0.92 | T-13 | 0.35 | 0.35 | 0.35 |
| M-1 | 0.58 | 0.59 | 0.67 | F-4 | 0.11 | 0.27 | 0.27 |
| M-2 | 0.75 | 0.78 | 0.80 | F-5 | 0.47 | 0.70 | 0.70 |
| P-10 | 0.34 | 0.38 | 0.38 | P-14 | 0.31 | 0.48 | 0.49 |
| T-4 | 0.07 | 0.08 | 0.19 | T-8 | 0.12 | 0.17 | 0.17 |
| T-5 | 0.41 | 0.48 | 0.49 | P-11 | 0.11 | 0.17 | 0.17 |
| F-7 | 0.26 | 0.29 | 0.35 | D-15 | 0.66 | 0.71 | 0.82 |
| M-3 | 0.20 | 0.29 | 0.29 | D-16 | 0.31 | 0.32 | 0.33 |
| M-4 | 0.37 | 0.52 | 0.52 | M-7 | 0.09 | 0.73 | 0.73 |
| M-5 | 0.29 | 0.15 | 0.20 | F-8 | 0.32 | 0.32 | 0.42 |
| P-15 | 0.6 | 0.06 | 0.09 | D-14 | 0.86 | 0.93 | 0.93 |
| C-1 | 0.25 | 0.27 | 0.28 | T-9 | 0.23 | 0.40 | 0.47 |
| C-2 | 0.23 | 0.14 | 0.27 | T-12 | 0.11 | 0.13 | 0.13 |
| S-2 | 0.23 | 0.63 | 0.63 | | | | |

(a) Fine tuning convergence for the WADI dataset with a population size of 8, 64 iterations, $\tau = \frac{1}{256}$, and a mutation probability of 0.02.
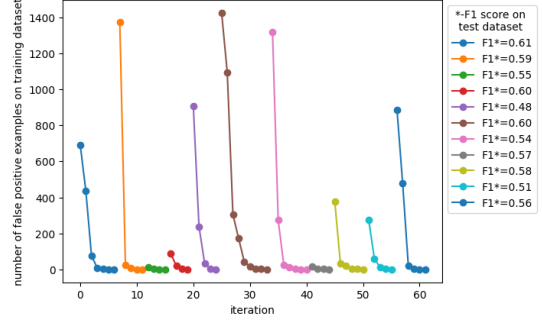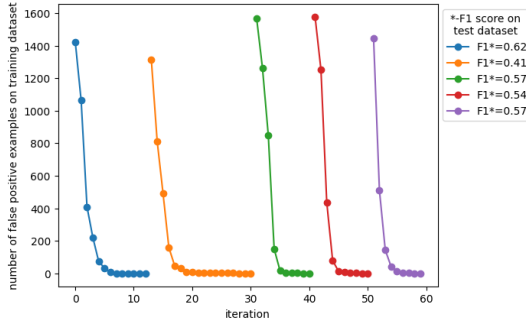
(b) Fine tuning convergence for the WADI dataset with a population size of 16, 64 iterations, $\tau = \frac{1}{256}$, and a mutation probability of 0.02.
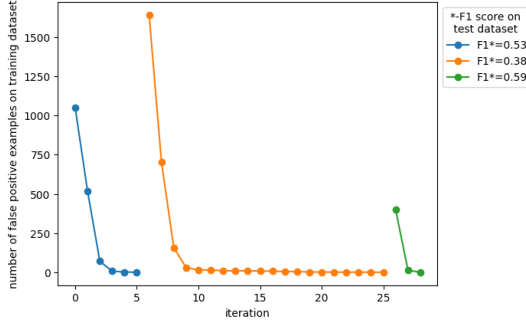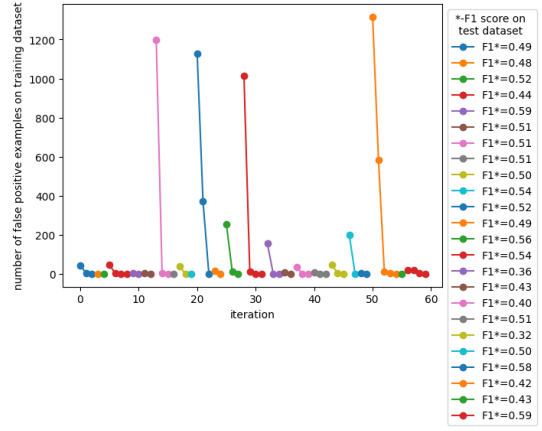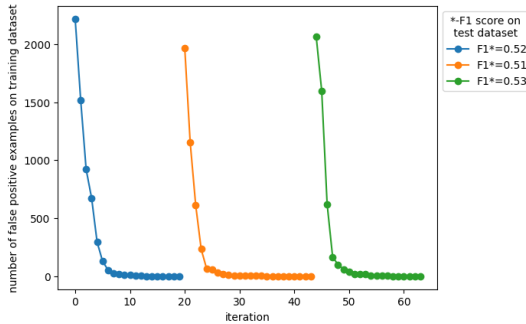
(c) Fine tuning convergence for the WADI dataset with a population size of 24, 32 iterations, $\tau = \frac{1}{256}$, and a mutation probability of 0.02.
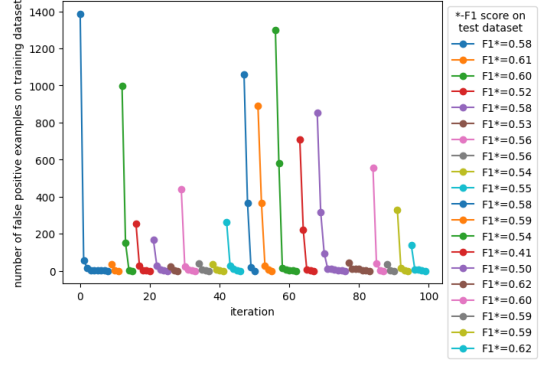
(d) Fine tuning convergence for the WADI dataset with a population size of 24, 100 iterations, $\tau = \frac{1}{256}$, and a mutation probability of 0.02
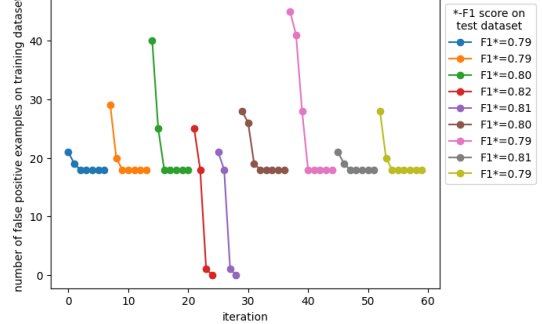
(e) Fine tuning convergence for the WADI dataset with a population size of 24, 64 iterations, $\tau = \frac{1}{512}$, and a mutation probability of 0.02.

(f) Fine tuning convergence for the WADI dataset with a population size of 24, 64 iterations, $\tau = \frac{1}{128}$, and a mutation probability of 0.02.

(g) Fine tuning convergence for the WADI dataset with a population size of 24, 64 iterations, $\tau = \frac{1}{256}$, and a mutation probability of 0.05.

(h) Fine tuning convergence for the SWAT dataset with a population size of 24, 64 iterations, $\tau = \frac{1}{256}$, and a mutation probability of 0.02.

Fig. 1: Additional convergence plots