

	$N \times M$	Benchmark 1	Benchmark 2	Proposed 1	Proposed 2
Routine	100 x 5,000	0.58 ± 0.17	0.70 ± 0.24	0.74 ± 0.21	0.80 ± 0.13
Depresjon	55 x 10,000	0.67 ± 0.17	0.72 ± 0.17	0.73 ± 0.16	0.82 ± 0.16
ECG5Days	23 x 136	0.70 ± 0.31	0.87 ± 0.28	0.90 ± 0.21	0.90 ± 0.21
Earthquakes	139 x 512	0.65 ± 0.08	0.70 ± 0.08	0.71 ± 0.10	0.73 ± 0.15
Beef	60 x 470	0.47 ± 0.25	0.47 ± 0.25	0.48 ± 0.23	0.55 ± 0.20
ECG	200 x 96	0.83 ± 0.08	0.85 ± 0.09	0.87 ± 0.06	0.85 ± 0.11
Dodger	144 x 288	0.87 ± 0.06	0.89 ± 0.06	0.90 ± 0.06	0.83 ± 0.09
Melbourne	434 x 24	0.96 ± 0.03	0.95 ± 0.03	0.97 ± 0.03	0.94 ± 0.04
Coffee	56 x 286	0.96 ± 0.06	0.98 ± 0.06	0.98 ± 0.06	0.98 ± 0.03
Strawberry	983 x 235	0.91 ± 0.05	0.92 ± 0.04	0.93 ± 0.04	0.90 ± 0.04

Table 1: F-score from RFC trained on each dataset transformed into four feature sets.

Benchmark 1: all features as proposed by [Baldán and Benítez \(2020\)](#) $|K| = 42$. **Benchmark 2:** Best performing “Complexity feature” and all “Descriptive Features” as proposed by [Baldán and Benítez \(2020\)](#) $|K| = 34$. **Proposed 1:** Best performing proposed measure and all “Descriptive Features” $|K| = 34$. **Proposed 2:** Best performing proposed measure with top four features obtained from RFC feature importance $|K| = 5$

	LR	DT	RF	MLP	LSTM FCN
Depresjon Benchmark 2	0.73 ± 0.14	0.67 ± 0.25	0.72 ± 0.18	0.76 ± 0.14	0.64
Depresjon Proposed 2	0.76 ± 0.14	0.73 ± 0.17	0.85 ± 0.15	0.77 ± 0.18	0.64

Table 2: Accuracy from different classifiers on the Depresjon dataset transformed into feature sets using the method of [Baldán and Benítez \(2020\)](#) $|K| = 42$ and the proposed feature extraction where $|K| = 5$. Classifiers are listed in decreasing interpretability and include Logistic Regression (LR), Decision Tree (DT), Random Forest (RF) and Multi Layer Perceptron (MLP). For a comparison to the accuracy achieved through a state of the art deep learning approach, the hybrid LSTM, CNN architecture as proposed by [Karim et al. \(2018\)](#) was applied to the raw time series dataset with no feature extraction.

References

- Francisco J Baldán and José M Benítez. Complexity measures and features for times series classification. *arXiv preprint arXiv:2002.12036*, 2020.
- Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Shun Chen. Lstm fully convolutional networks for time series classification. *IEEE Access*, 6:1662–1669, 2018. doi: 10.1109/ACCESS.2017.2779939.