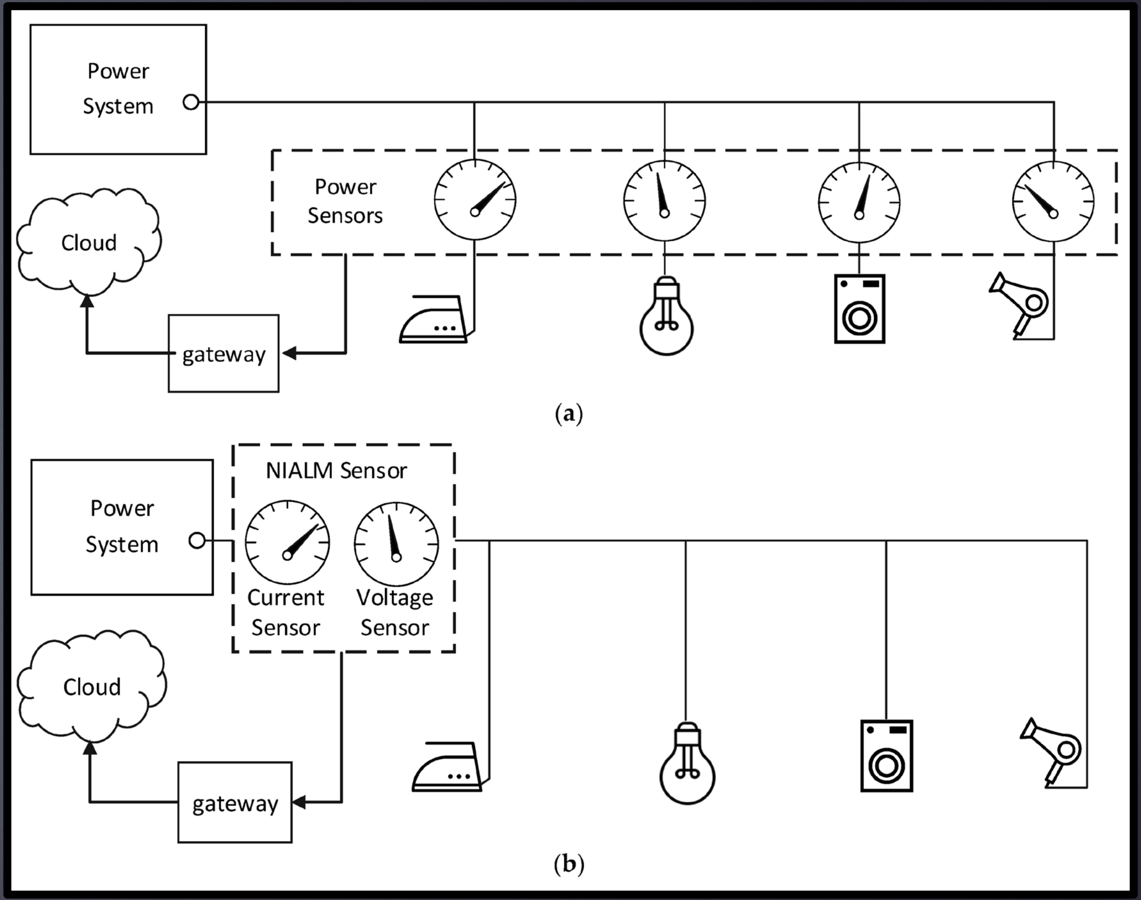
Learning to Automatically Identify Home Appliances

Dan Lorbek Ivančič, Blaž Bertalanič, Gregor Cerar, Carolina Fortuna

Introduction: What is ALM?

- Enables increasing the efficiency of domestic energy use
- Appliance consumption profiles
- Optimal energy utilization
- Useful information about appliance usage
  - Monitoring elderly
  - Theft detection
  - Running conditions
- **ILM** vs. **NILM**



Source: "Nonintrusive Appliance Load Monitoring: An Overview, Laboratory Test Results and Research Directions", 2019

Contributions

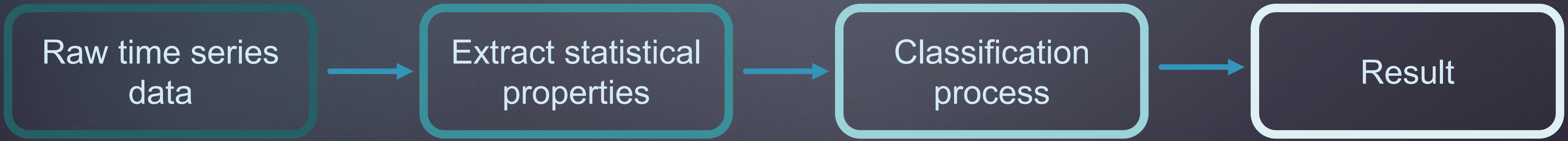
1) **Propose a new feature set for more accurate classification**
   - Features that better capture the shape of the time series yield better classification performance

2) **Propose new XGBoost-based and DNN-based models**
   - Present their development processes, make comparisons between the two and show that they both outperform the state of the art SVM-based model

3) **Consider multiple different evaluation metrics**
   - Accuracy, precision, recall, f1

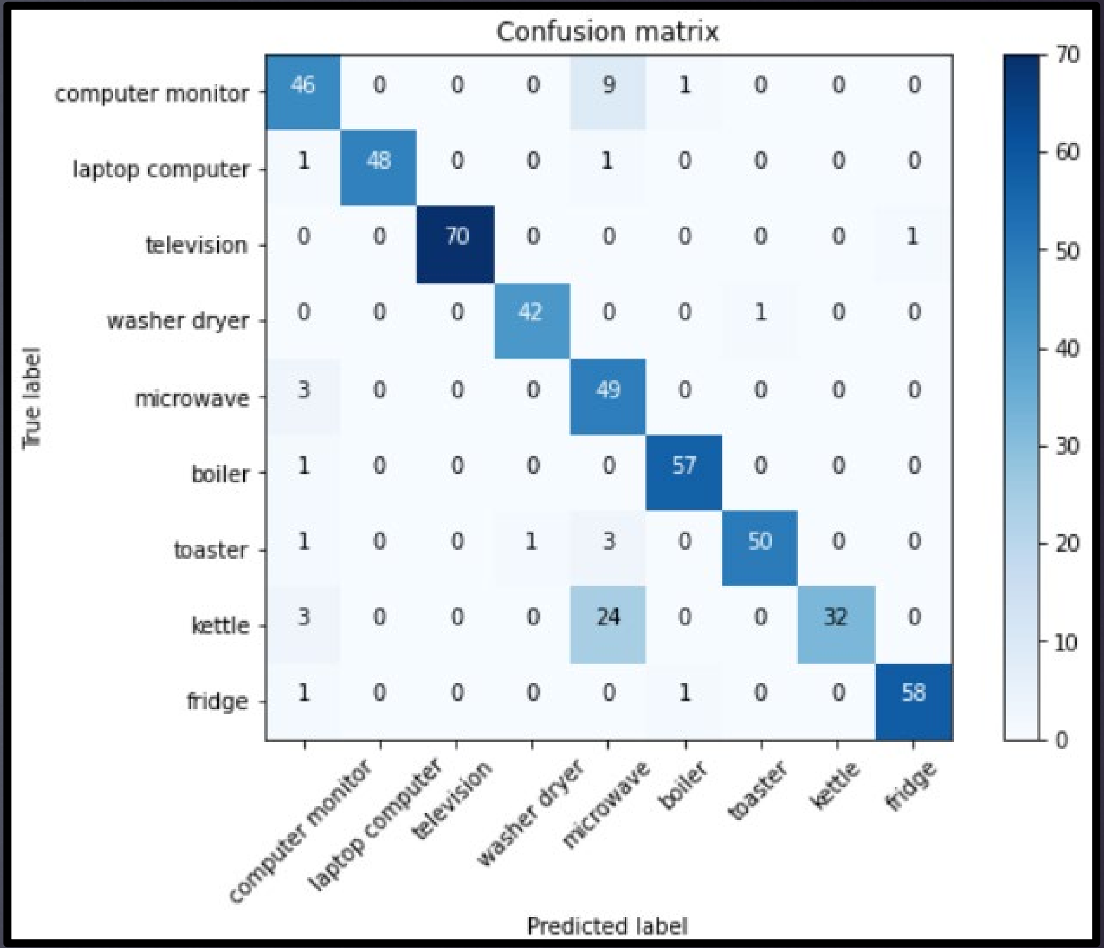
Problem statement

- Design a **9 class** classifier
  - Computer monitor, laptop computer, television, washer dryer, microwave, boiler, toaster, kettle, fridge
- Map time series to appliances
- Train on the **UK-Dale** dataset



Development methodology

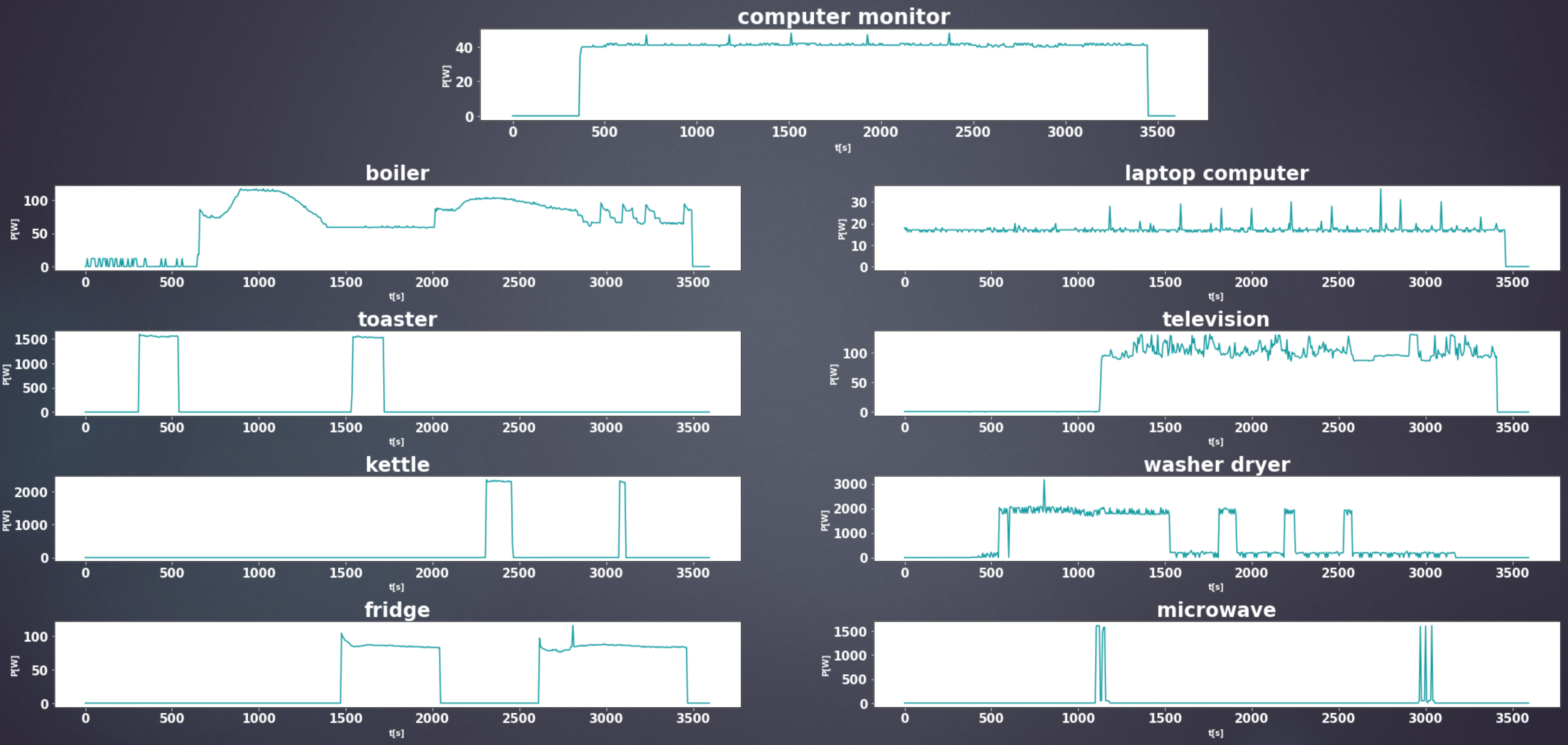
- Preprocess and clean the data
  - Remove “empty” samples
  - Make new feature sets by capturing different statistical properties
  - Robust scaling
- Build and evaluate the models
  - XGBoost, Deep Neural Network (DNN)
  - Calculate accuracy
  - Confusion matrices
- Compare the models
  - Compare with state of the art
  - Classification performance, complexity, computation time



UK-Dale dataset 1/2

- Domestic appliance-level electricity
- **5 houses** power demand
- **4.3 years** of data
- Sample rate **16Hz/0.1667Hz**
- **2700** samples
- **1 hour** segments

UK-Dale dataset 2/2



Feature engineering

- 3 feature sets
- Baseline raw time series
- Statistical properties of the data
- Extensive evaluation to identify best features
- Final feature set
  - Maximum, standard deviation, mean change, mean absolute change, longest strike above and below mean, kurtosis, absolute energy, number of peaks

Table 1: Feature comparison using the best models.

Model	**Feature set**	**Precision**	**Recall**	**f1**
DNN3	FeatureSet1	0.638	0.595	0.573
XGB3	FeatureSet1	0.799	0.769	0.779
DNN3	FeatureSet2	0.918	0.885	0.889
XGB3	FeatureSet2	0.869	0.864	0.867
DNN3	FeatureSet3	**0.931**	**0.898**	**0.902**
XGB3	FeatureSet3	**0.888**	**0.889**	**0.889**
DNN3	best[5]	0.893	0.887	0.888
XGB3	best[5]	0.861	0.860	0.861
SVM[5]	best[5]	0.851	0.835	0.834

Model structure

- XGBoost (XGB3)
  - Maximum depth 3
  - 500 estimators
  - Learning rate
  - Minimum child weight
- Deep Neural Network (DNN3)
  - 5 layers
  - 128 / 64 / flatten / 32 / 9
  - RMSprop

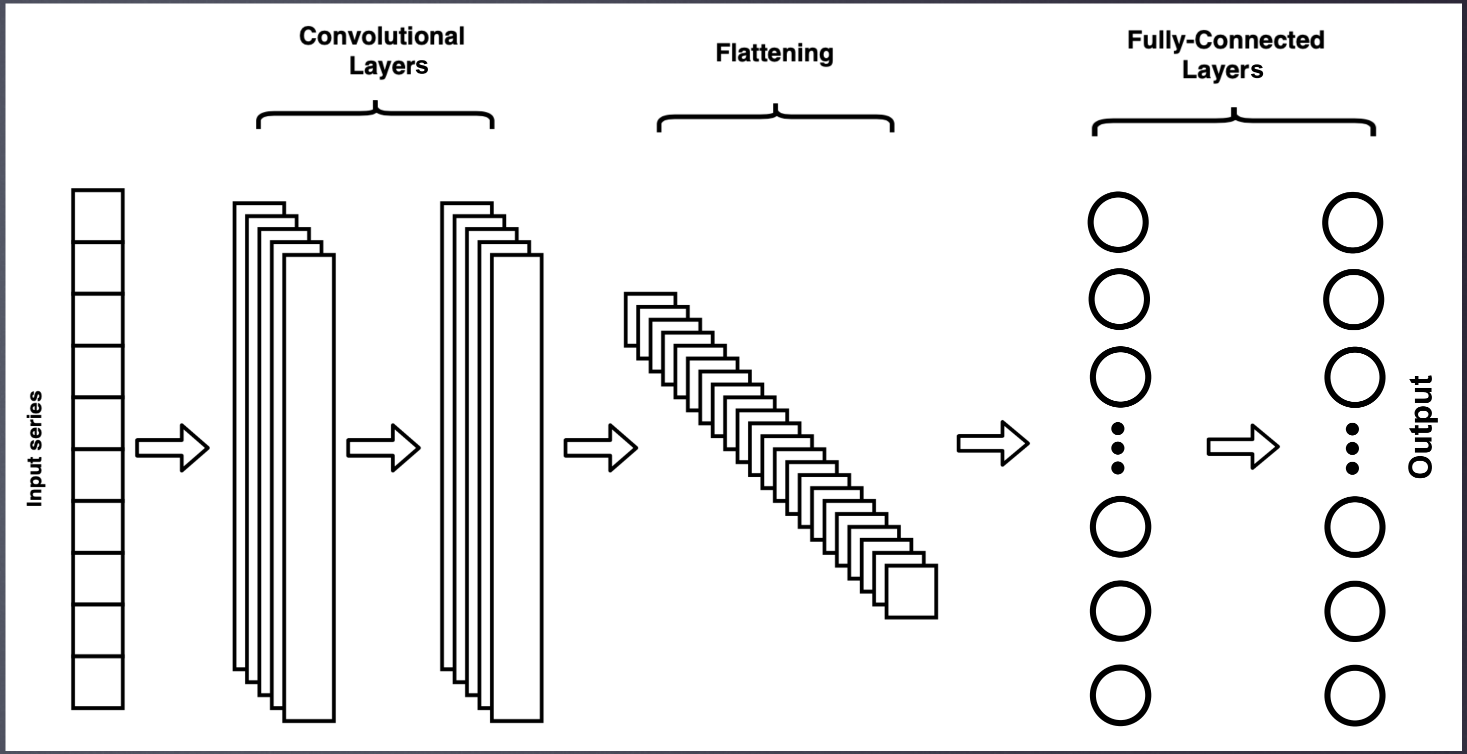


Figure: General CNN architecture
Figure: Our best XGB architecture

Source: “A hybrid ensemble method for pulsar candidate classification.” Astrophysics and Space Science, 2019. 364. 10.1007/s10509-019-3602-4.

Model selection

- 3 model configurations
- Tuning **hyperparameters** and evaluating performance
- Optuna
- Compare the 2 best models

Table 3: Model performance on FeatureSet3.

Model	Precision	Recall	f1	Comp. time
DNN1	0.866	0.851	0.846	10.972s
DNN2	0.900	0.887	0.889	21.026s
DNN3	**0.931**	**0.898**	**0.902**	21.124s
XGB1	0.876	0.863	0.864	1.126s
XGB2	0.884	0.881	0.882	2.518s
XGB3	**0.888**	**0.889**	**0.889**	3.225s
SVM [5]	0.878	0.852	0.852	0.301s

Conclusions

- Extracting **statistical features** from time series improves **f1 score** by up to **20%**
- **Model optimization** further increases performance by up to **4%**
- **Feature engineering** has much greater impact
- **XGBoost** vs. **Convolutional Neural Network**
  - CNN superior performance
  - XGB less computationally expensive

Thank you!