Deep and recurrent neural networks: two real-world use cases

Current Trends in AI

Guillaume LEVASSEUR & Cédric GILON

IRIDIA,
Université libre de Bruxelles

March 4th 2022

Table of contents

Learning sequences

Index i	*1*	*2*	*3*	*4*	*5*
Variable Y_i	2	4	6	8	Y_5

Learning sequences

Index i	*1*	*2*	*3*	*4*	*5*
Variable Y_i	2	4	6	8	Y_5

$$Y_5 = 10$$

Learning sequences

Index i	*1*	*2*	*3*	*4*	*5*
Variable Y_i	2	4	6	8	Y_5

$$Y_5 = 10$$

Index i	*1*	*2*	*3*	*4*	*5*
Variable Y_i	SI	VIS	PACEM	PARA	Y_5

Learning sequences

Index i	*1*	*2*	*3*	*4*	*5*
Variable Y_i	2	4	6	8	Y_5

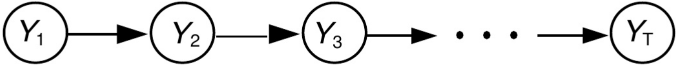
$$Y_5 = 10$$

Index i	*1*	*2*	*3*	*4*	*5*
Variable Y_i	SI	VIS	PACEM	PARA	Y_5

$$Y_5 = DOX$$
$$Y_5 = PLU(IE)$$
$$Y_5 = BELLVM$$

Markov chains

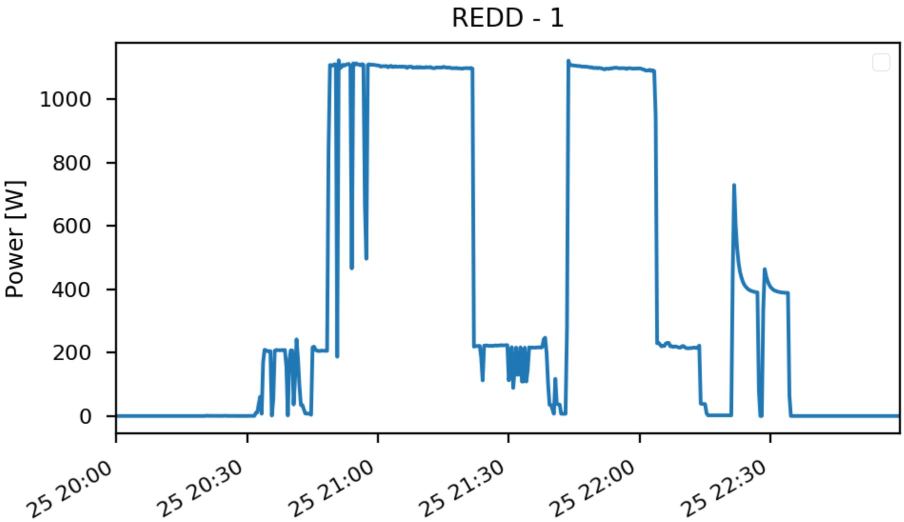
Index i	*1*	*2*	*3*	*4*	*5*
Variable Y_i	2	4	6	8	Y_5



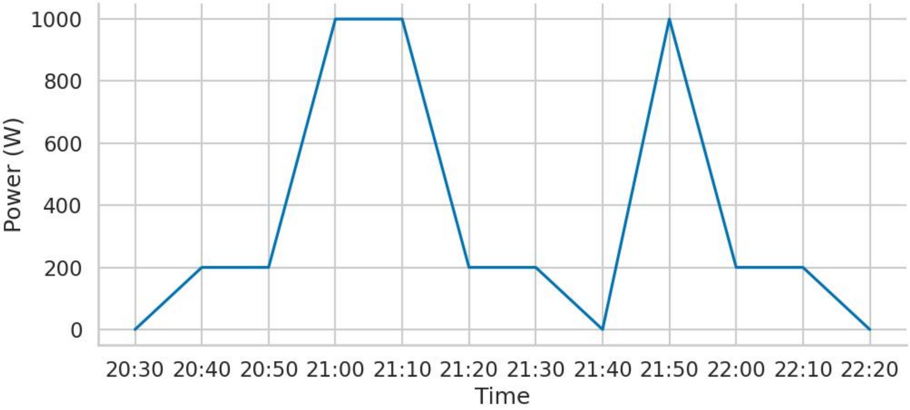
Source: [1]

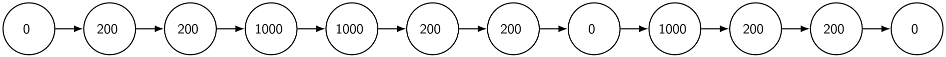
$$P(Y_5 = 10) = P(Y_5 = 10 | Y_4 = 8)$$

Markov chains

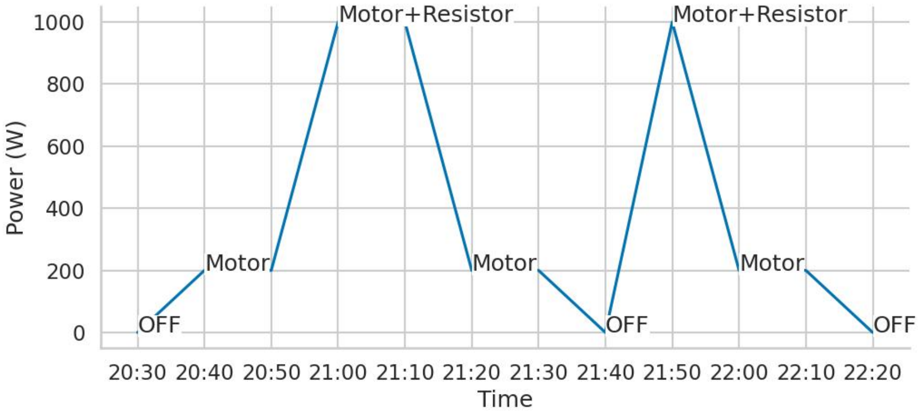


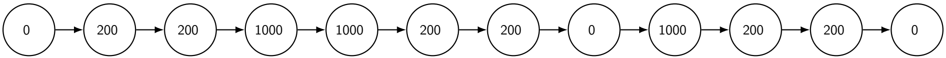
Markov chains



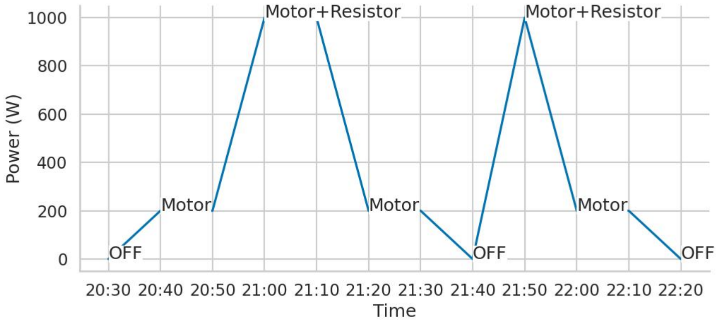


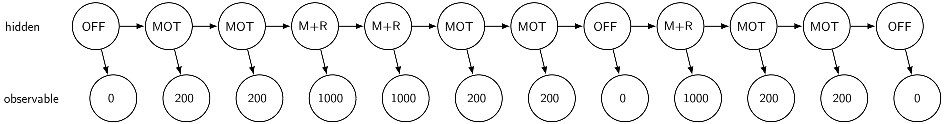
Markov chains



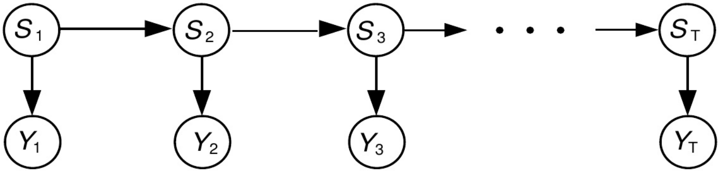


Hidden Markov models





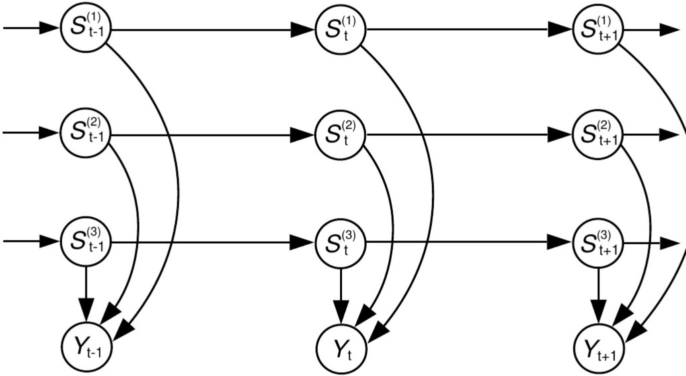
Hidden Markov models



Source: [1]

- State transition probability matrix & observation probability distribution.
- Left-right architecture.
- 1^{st} order Markov: $P(S_t) = P(S_t|S_{t-1})$
- Hypothesis: observations Y_t are independent.

Factorial HMM



Source: [1]

Markov models summary

Key features:

- Transition matrix can be constrained to the problem.
- Approximate learning.
- Generative models.
- Sucessfully applied to: signal denoising, molecular biology, electricity disaggregation, speech recognition, etc.

Limitations:

- Intractability for large numbers of states.
- Complexity: $O(TK^{2M})$
- Hypothesis: independence of observations.
- Hypothesis: 1^{st} order Markov

Learning sequences – Need for memory

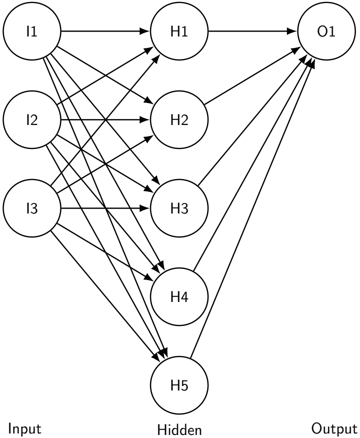
Index i	*1*	*2*	*3*	*4*	*5*
Variable Y_i	SI	VIS	PACEM	PARA	Y_5

$Y_5 = BELLVM$

Use Y_3 and Y_4 to deduce Y_5?

- 2^{nd} order Markov chain. $\Rightarrow$ Complexity!
- Store Y_3 in memory.

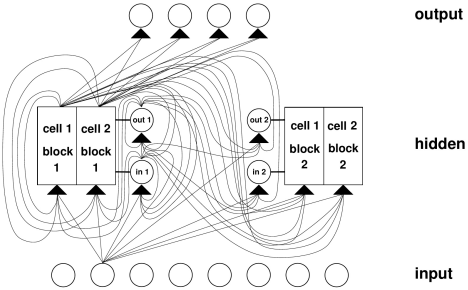
Neural networks: extension of HMMs



Multi-layer perceptron (MLP)

- Fully connected, feed-forward network.
- One hidden layer.
- Sigmoid activation.
- No memory.

Recurrent neural networks & decaying error

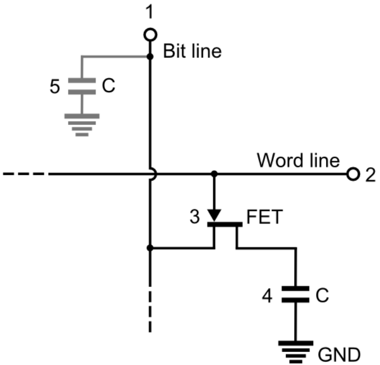


Source: [2]

- Fully connected network with feedback loops.
- One hidden layer.
- Sigmoid activation.
- Memory effect but exponential decay or blow-up of the error.

Problem: How to get a memory effect while avoiding decay or blow-up?

Computer memory cell (DRAM)

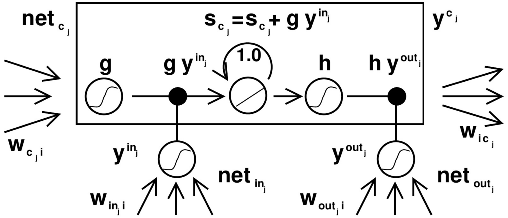


Source: wikimedia.org

1. Binary value to read/write
2. Trigger
3. Gate MOSFET
4. Memory capacitor
5. Line parasitic capacitance

- DRAM = dynamic random-access memory
- MOSFET = metal-oxide-silicon field-effect transistor

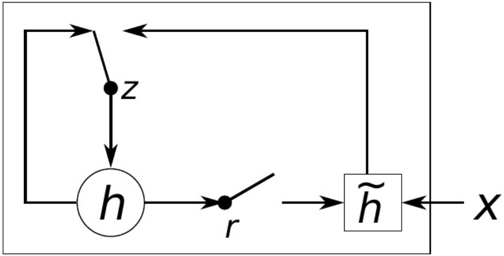
Long short-term memory (LSTM)



Source: [2]

- Gated unit for recurrent networks.
- Central self-recurring neuron with linear activation.
- net_c: input from the network.
- net_{in}: input gate.
- net_{out}: output gate.
- s_c: internal state.
- y_c: output to the network.

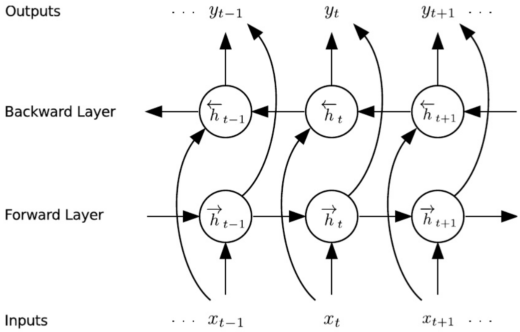
Gated recurrent unit (GRU)



Source: [3]

- Simplified LSTM for similar performance.
- X: input from the network.
- z: update gate.
- r: reset gate.
- h: hidden state.
- $\tilde{h}$: new state.
- Output: $zh^{t-1} + (1 - z)\tilde{h}^t$

Bi-directional recursion



Source: [4]

- Use when the whole sequence is known, i.e. no real-time processing.
- Can be used with LSTM or GRU.
- Possible aggregations:
  - sum
  - product
  - concatenation

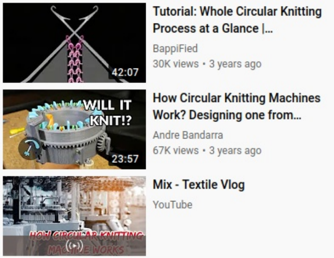
Deep recurrent neural networks summary

Key features:

- End-to-end learning of the problem.
- Generative models.
- Sucessfully applied to: signal denoising, electricity disaggregation, speech recognition, online translation, content recommendation, etc.

Limitations:

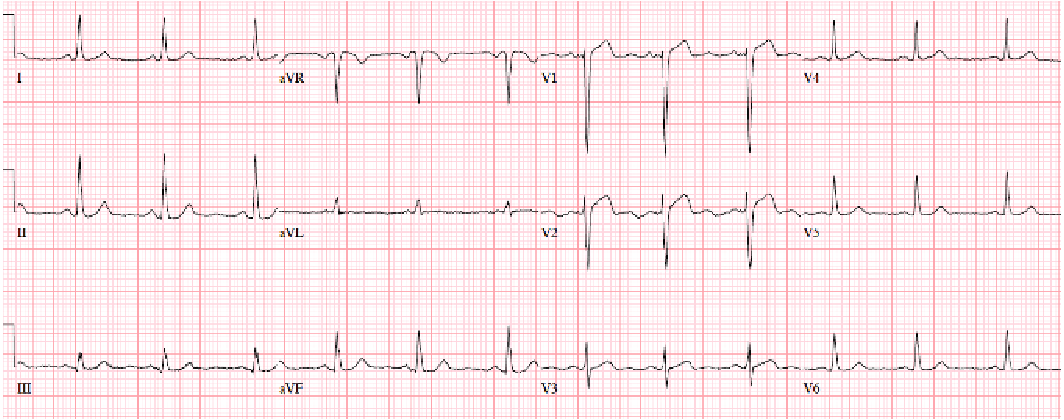
- Interpretability.
- Convergence is not guaranteed.
- More complex than HMMs.



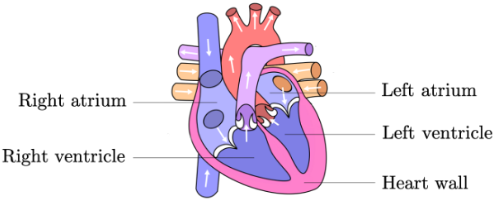
Use case 1: atrial fibrillation detection and forecast

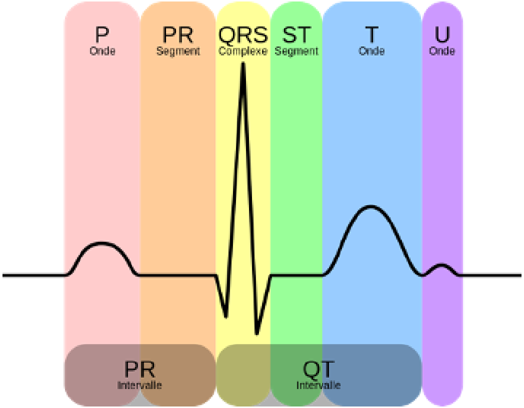
- Human heart and ECG
- Atrial fibrillation
- AF detection
- AF forecast

Electrocardiogram (ECG)



Human heart and ECG decomposition





Atrial fibrillation

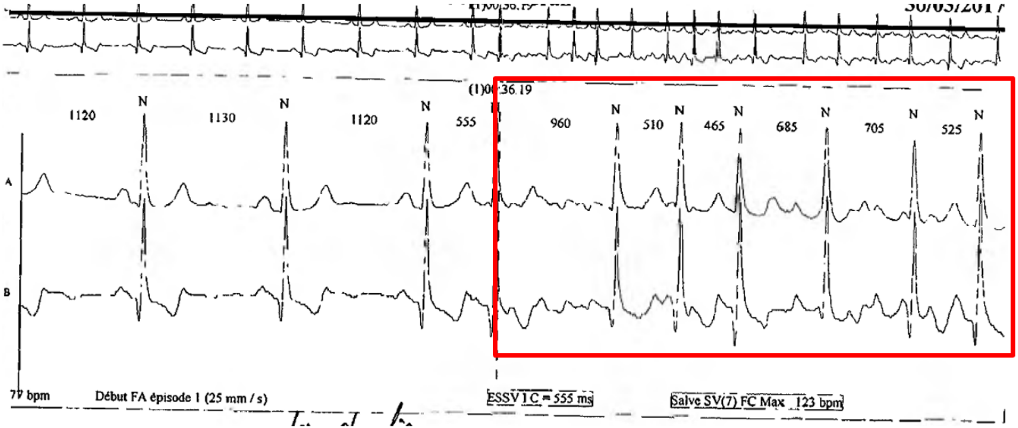
Medical definition [5]:

- No P wave
- ECG strip $>$ 30 seconds
- Irregular rhythm

Risks [5]:

- $\pm$ 50 millions patients worldwide
- Stroke risk x5
- Death risk x2

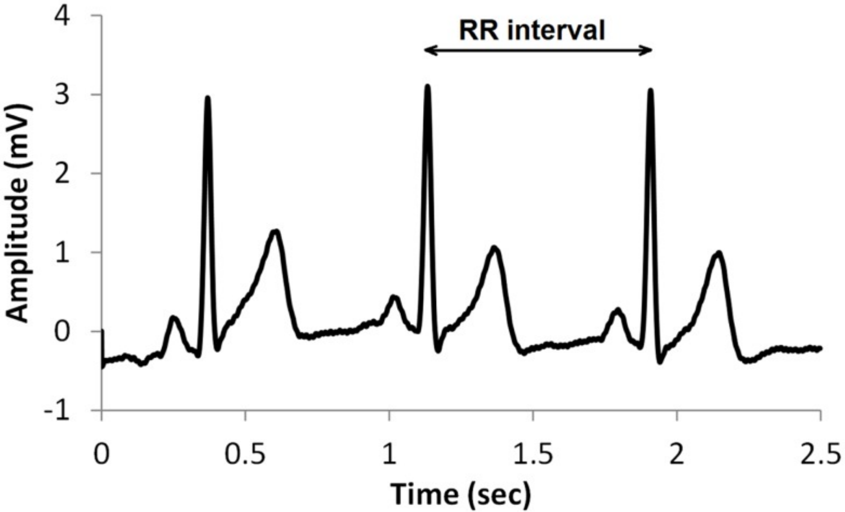
Transition sinus rhythm-AF on ECG



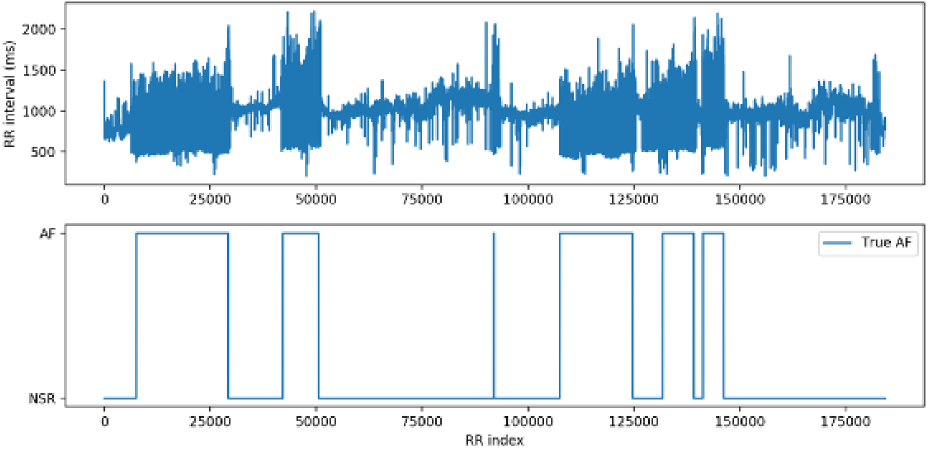
Holter recording with AF

Holter recording (3 days) with two episodes of AF

RR intervals

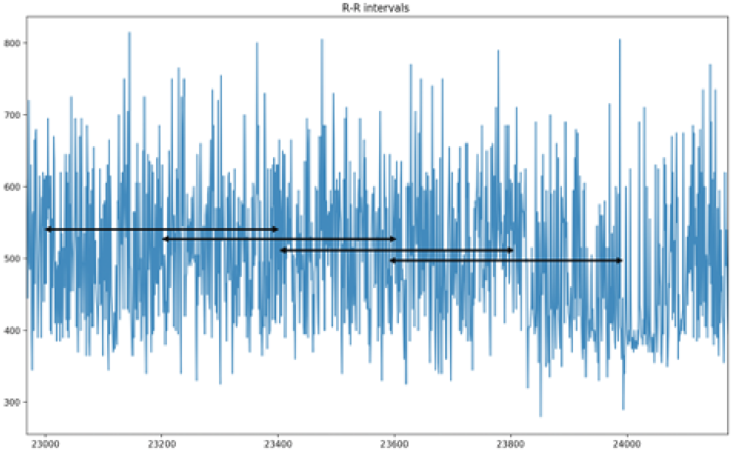


AF detection task



Detect at each instant t if AF of normal sinus rhythm (NSR)

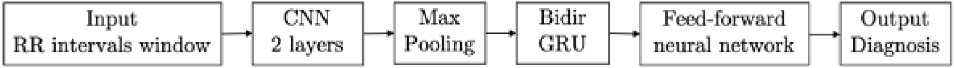
AF detection with sliding window



Detect for each window if AF signs or not
(binary classification - supervised learning)

DNN model

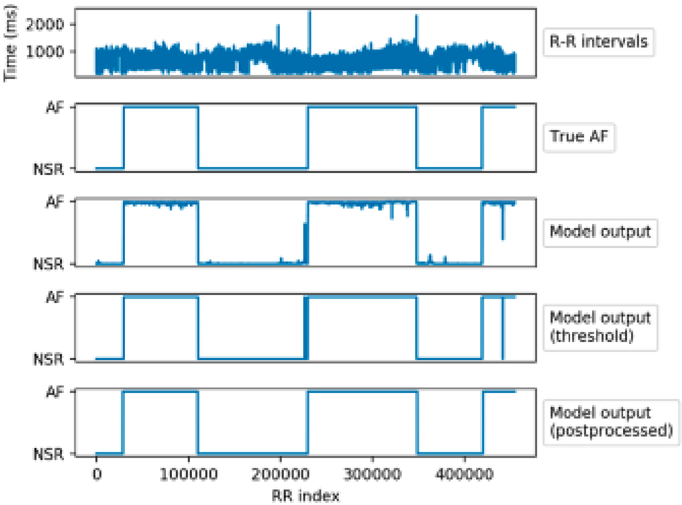
Layer	Type	Parameters	Output shape
1	Input layer		300 x 1
2	1D convolution	filers:100, kernel size: 3	298 x 100
3	1D convolution	filers:100, kernel size: 3	296 x 100
4	Global max pooling		100
5	Reshape		100 x 1
6	Bidirectional GRU	units: 100	200
7	Fully connected	units: 1, activation: sigmoid	1
	Total number of parameters	91 901	



Source: [6]

AF detection metric

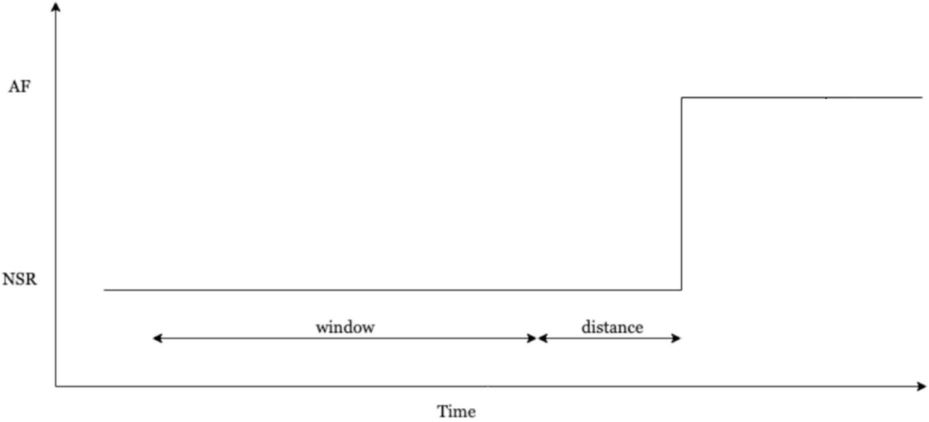
- AUC: 99.6
- Sensitivity: 94.9
- Specificity: 99.1



AF forecast

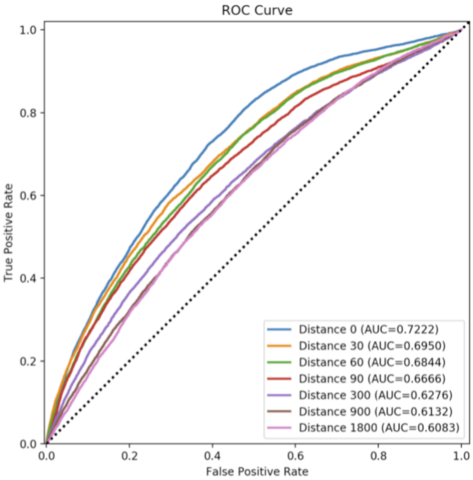
Is there information in the ECG previous to the AF onset?

AF forecast parameters



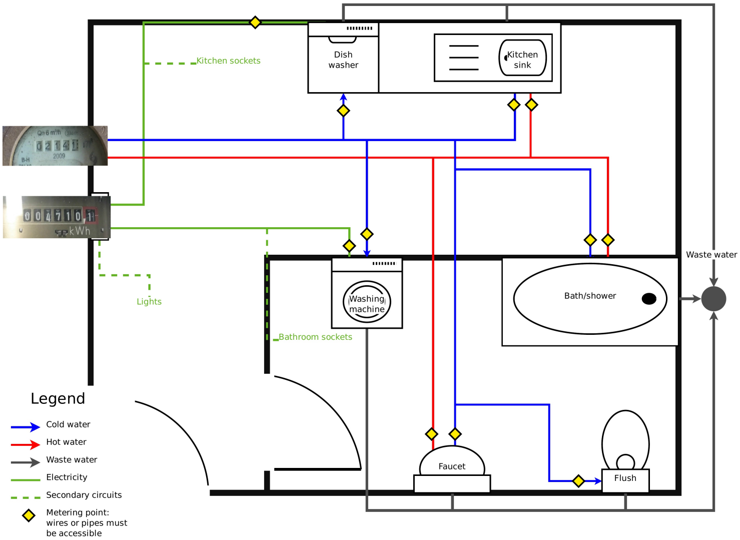
AF forecast

Distance	AUC	AUC CI (95%)
0	0.722	[0.721 - 0.724]
30	0.695	[0.693 - 0.697]
60	0.684	[0.683 - 0.686]
90	0.667	[0.665 - 0.668]
300	0.628	[0.626 - 0.629]
900	0.613	[0.611 - 0.615]
1800	0.608	[0.607 - 0.610]



Source: [6]

Use case 2: electricity disaggregation

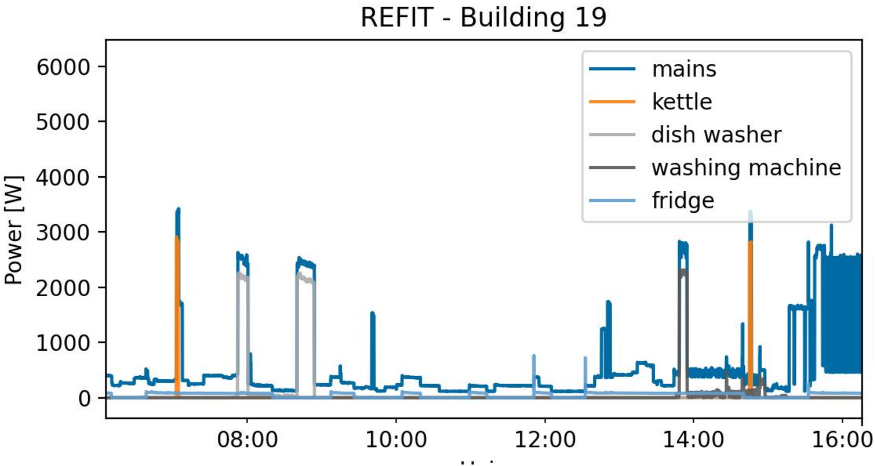


Feedback has an impact

Intervention type	Decrease range	Region	Sources
Comparison group	4–13 %	US, EU	[7, 8]
Total consumption	2–22 % [1]	EU	[8–11]
Consumption cost	0–23 % [1]	EU	[12, 13]
Appliance-level consumption	3–27 %	EU	[10, 11, 13]
Hawthorne effect	4–13 %	EU	[8, 11]

[1] Subject to damping as shown in [10].

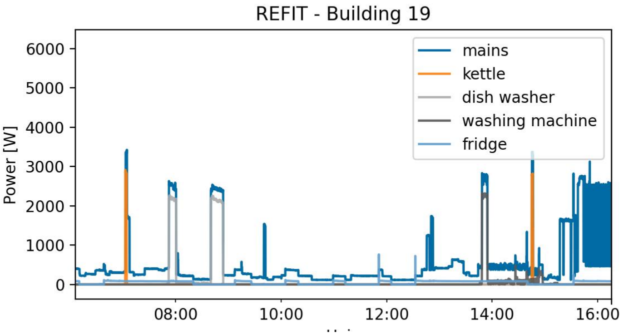
Data characteristics



Source: Data from [14]

C^0 continuity, aperiodic, class imbalance, 95% zeros.

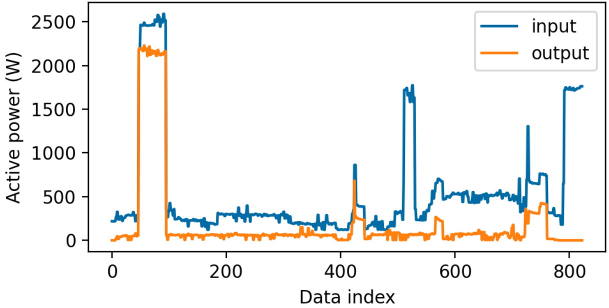
Multi-label classification approach



Source: Data from [14]

- Model type: factorial HMMs.
- Sensitive to class imbalance.
- Signal aggregate constrains can be added.

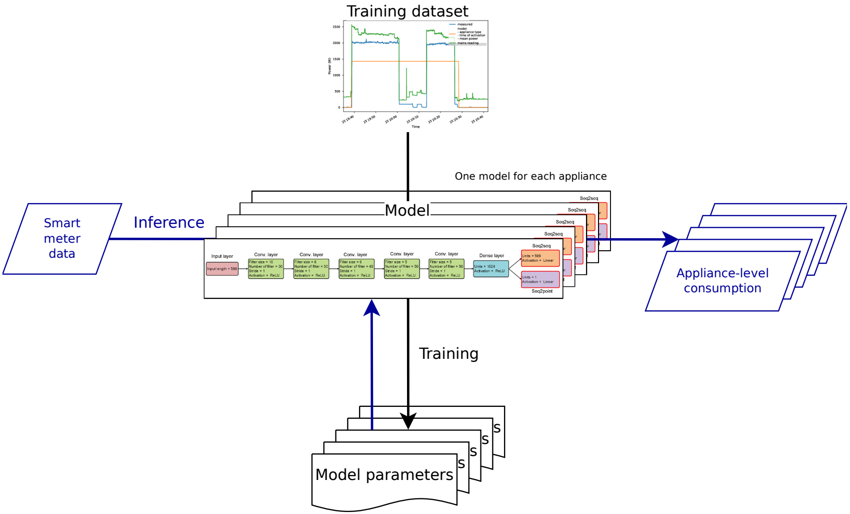
Denoising regression approach



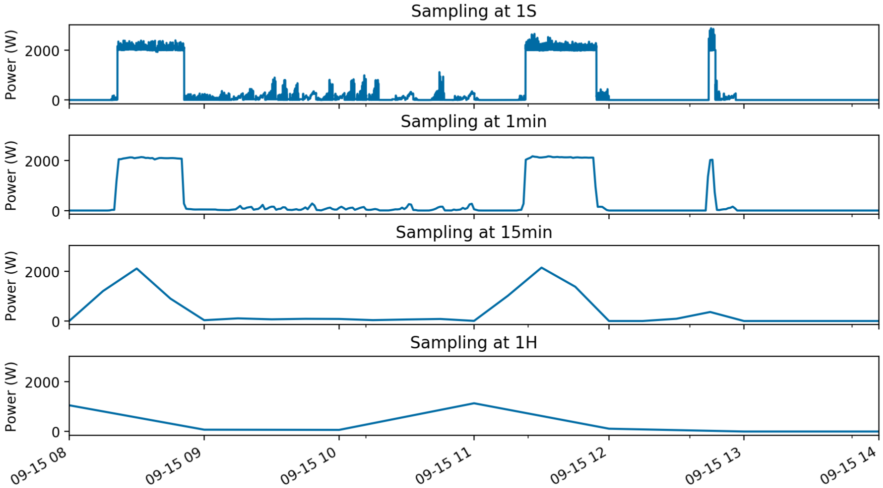
Source: Data from [14]

- Model type: deep neural networks.
- Requires data sub-metering for training.
- Better more homes than longer periods.

Denoising regression approach

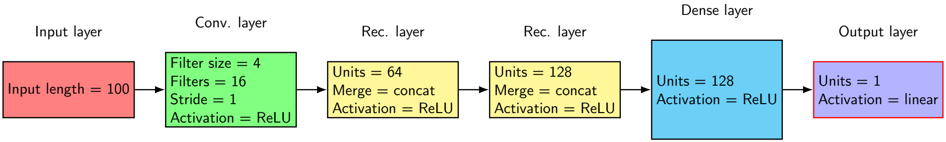


Necessary time granularity

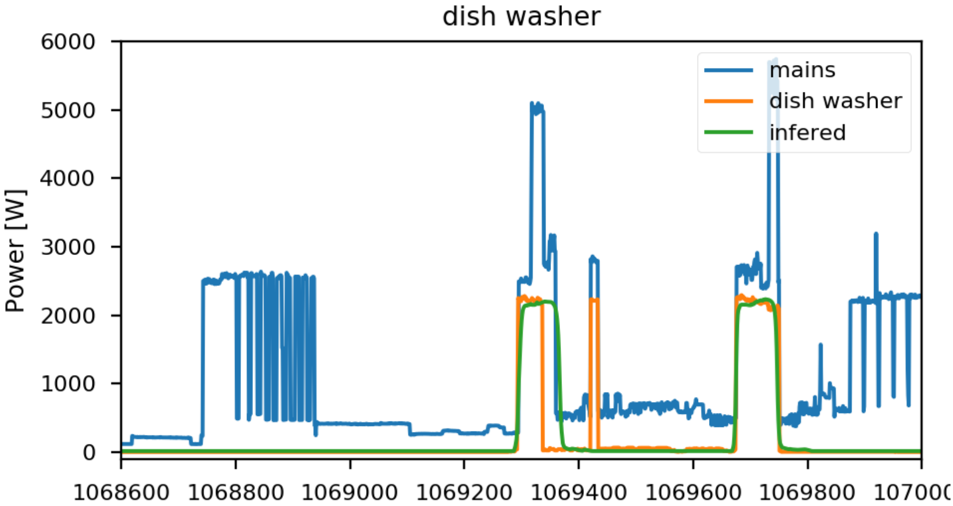


Deep recurrent net for disaggregation

WindowGRU [15]

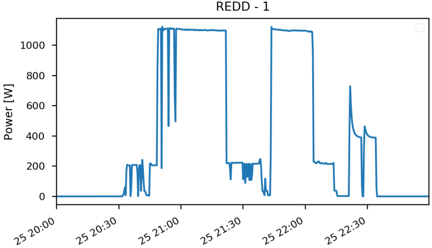


Model output



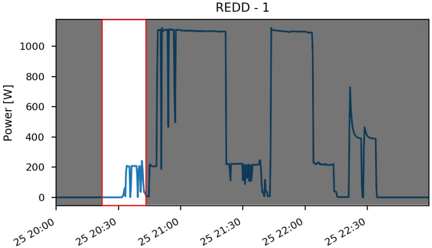
Sliding window processing

Human view



Source: Data from [16]

Algorithm view

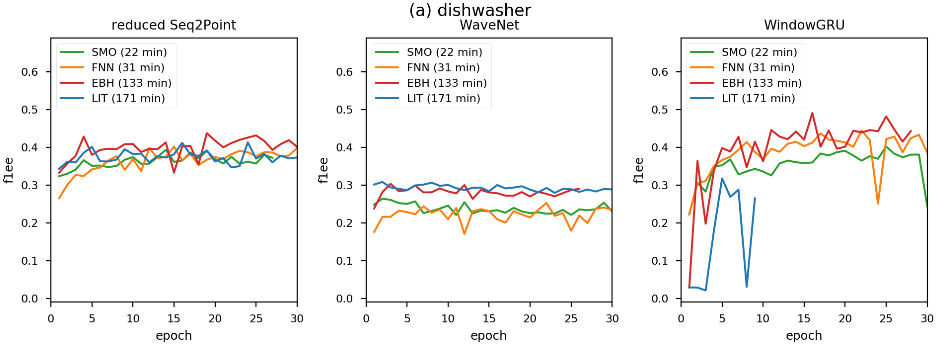


Source: Data from [16]

Mañé & Takens theorem $\Rightarrow$ A window size exists such that enough data is captured for the model to fit. Too long windows include noise and harm performance [17].

Performance benchmark

Generalization test on an “unseen” home:



Parameters: 15,930,895 | 50,248 | 262,737

Design of ML experiments

- Inference of already seen patterns is easy.
- Evaluate generalization on unseen patients or homes.
- Group IDs for cross-validation.
- Select a period during which at least two events happen.

Group of 11

Home ID	1	2	3
0801	test	train	train
0809	test	train	train
0803	test	train	train
0807	test	train	train
0808	test	train	train
0804	test	train	train
0810	test	train	train
0802	test	train	train
3036	test	train	train
3021	test	train	train
3002	test	train	train
3023	train	test	train
3019	train	test	train
3014	train	test	train
3015	train	test	train
3005	train	test	train
3035	train	test	train
3039	train	test	train
3006	train	test	train
3001	train	test	train
3033	train	test	train
3037	train	test	train
3020	train	train	test
1203	train	train	test
1211	train	train	test
1206	train	train	test
1202	train	train	test
1208	train	train	test
1209	train	train	test
1207	train	train	test
1204	train	train	test
1205	train	train	test
1210	train	train	test

Training set

Testing set

Conclusion

- Neural networks extend Markov models.
- Memory cells are instrumental in the convergence of recurrent nets.
- GRU layer simplifies LSTM for a similar performance.
- Deep recurrent networks combine convolutional and recurrent layers.
- Deep RNN require less parameters than CNN.
- Two use-cases: fibrillation detection and disaggregation.

References I

[1] Zoubin Ghahramani. "An introduction to hidden Markov models and Bayesian networks". In: *Series in Machine Perception and Artificial Intelligence*. WORLD SCIENTIFIC, June 2001, pp. 9–41. DOI: 10.1142/9789812797605_0002.

[2] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Computation* 9.8 (Nov. 1997), pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

[3] Kyunghyun Cho et al. "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation". In: *Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*. ACL. June 2014.

[4] Alex Graves and Navdeep Jaitly. "Towards End-To-End Speech Recognition with Recurrent Neural Networks". In: *Proceedings of the 31st International Conference on Machine Learning*. Ed. by Eric P. Xing and Tony Jebara. Vol. 32. Proceedings of Machine Learning Research 2. Bejing, China: PMLR, 22–24 Jun 2014, pp. 1764–1772. URL: https://proceedings.mlr.press/v32/graves14.html.

[5] Gerhard Hindricks et al. "2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS)". In: *European Heart Journal* 42.5 (Aug. 2020), pp. 373–498. DOI: 10.1093/eurheartj/ehaa612.

[6] Cédric Gilon et al. "Forecast of paroxysmal atrial fibrillation using a deep neural network". In: *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, July 2020. DOI: 10.1109/ijcnn48605.2020.9207227.

[7] Maithili Iyer et al. "Comparison groups on bills: Automated, personalized energy information". In: *Energy and Buildings* 38.8 (Aug. 2006), pp. 988–996. DOI: 10.1016/j.enbuild.2005.11.009.

[8] Gary Raw and David Ross. *Energy Demand Research Project: Final Analysis*. Tech. rep. AECOM, 2011.

[9] Vojkan Tasic et al. "Self-powered water meter for direct feedback". In: *2012 3rd IEEE International Conference on the Internet of Things*. IEEE, Oct. 2012. DOI: 10.1109/iot.2012.6402300.

[10] Daire McCoy and Sean Lyons. "Unintended outcomes of electricity smart-metering: trading-off consumption and investment behaviour". In: *Energy Efficiency* 10.2 (June 2016), pp. 299–318. DOI: 10.1007/s12053-016-9452-9.

References II

[11] Adnane Kendel et al. “What do people ‘learn by looking’ at direct feedback on their energy consumption? Results of a field study in Southern France”. In: *Energy Policy* 108.Supplement C (2017), pp. 593–605. DOI: 10.1016/j.enpol.2017.06.020. URL: http://www.sciencedirect.com/science/article/pii/S0301421517303749.

[12] Jonathan William Stinson. “Smart energy monitoring technology to reduce domestic electricity and gas consumption through behaviour change”. PhD thesis. Edinburgh Napier University, 2015.

[13] Ales Podgornik et al. “Effects of customized consumption feedback on energy efficient behaviour in low-income households”. In: *Journal of Cleaner Production* 130 (Sept. 2016), pp. 25–34. DOI: 10.1016/j.jclepro.2016.02.009.

[14] David Murray et al. “An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study”. In: *Scientific Data* 4.1 (Jan. 2017). DOI: http://dx.doi.org/10.1038/sdata.2016.122. URL: https://pureportal.strath.ac.uk/en/datasets/refit-electrical-load-measurements-cleaned.

[15] Odysseas Krystalakos et al. “Sliding Window Approach for Online Energy Disaggregation Using Artificial Neural Networks”. In: *Proceedings of the 10th Hellenic Conference on Artificial Intelligence - SETN '18*. ACM Press, 2018. DOI: 10.1145/3200947.3201011.

[16] J. Zico Kolter and Matthew J. Johnson. “REDD: A Public Data Set for Energy Disaggregation Research”. In: *SustKDD 2011*. 2011.

[17] Matthew B Kennel et al. “Determining embedding dimension for phase-space reconstruction using a geometrical construction”. In: *Physical review A* 45.6 (1992), p. 3403.