Federated Echo State Networks for Stress Prediction in the Automotive Use Case

Department of Computer Science

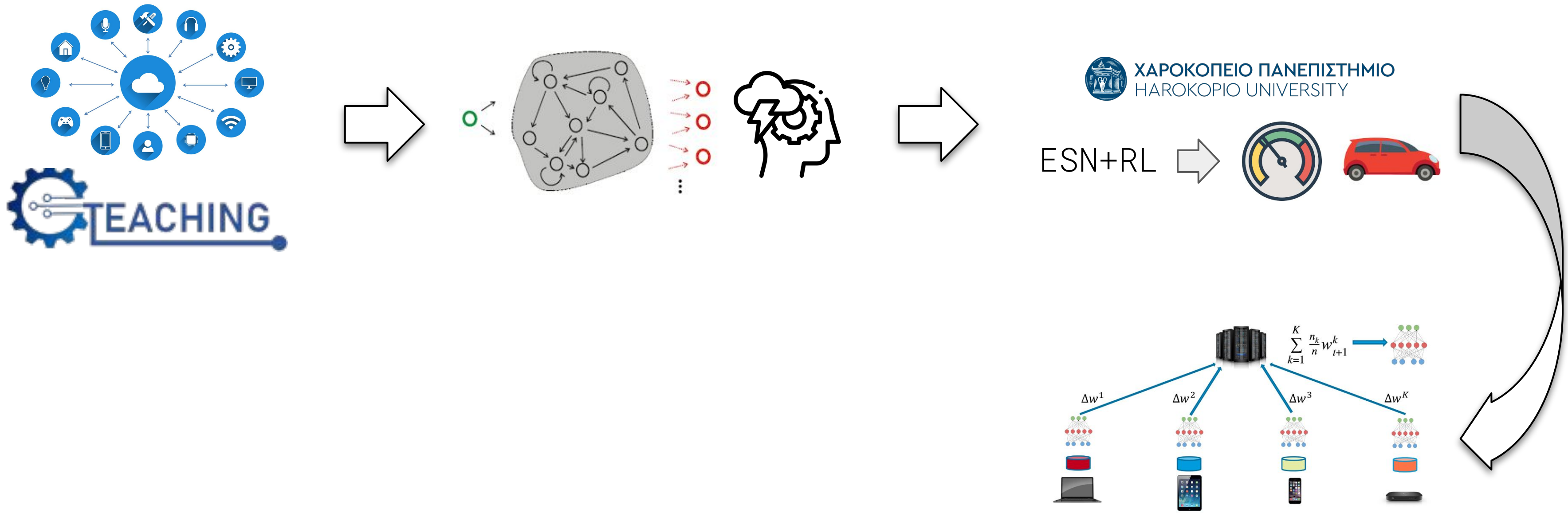
Master Degree in Computer Science: Artificial Intelligence curriculum

A.Y. 2021/22

Supervisors:
Prof. Davide Bacciu
Prof. Iraklis Varlamis (HUA)
Dr. Valerio De Caro

Candidate:
Antonio Di Mauro

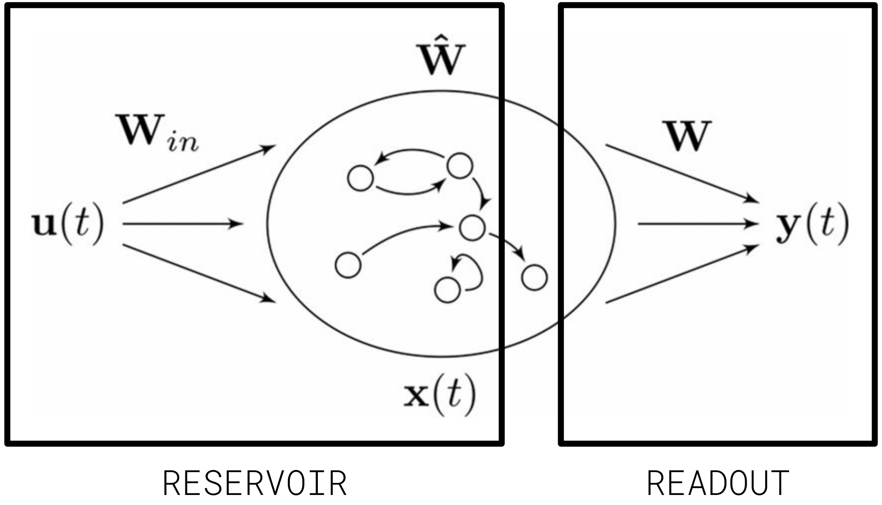
Introduction



Our Contributions

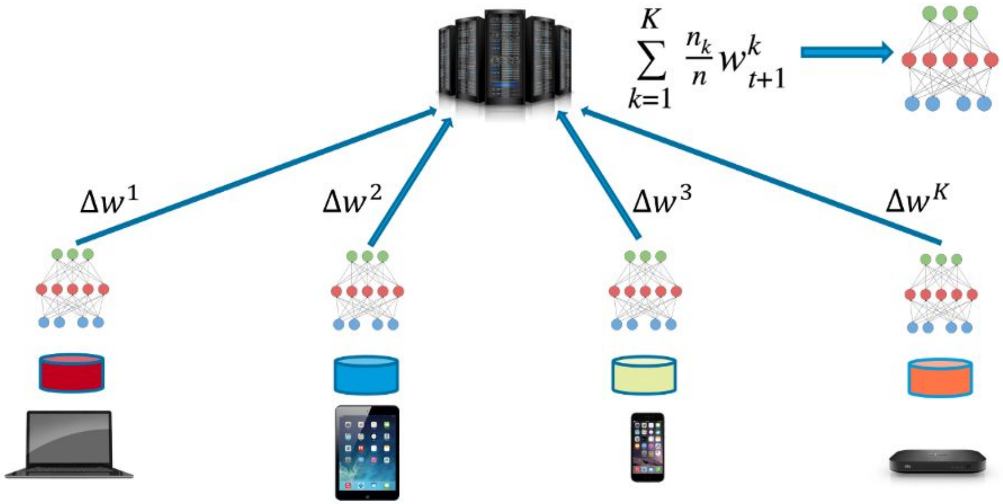
1. Gradient Descent + ESNs → Overcome SOTA
2. FedAvg + FedCurv + Tensorflow-Federated
3. Tensorflow-Federated + Anomaly Detector
4. Partial IncFed → Novel Approach

Echo State Networks

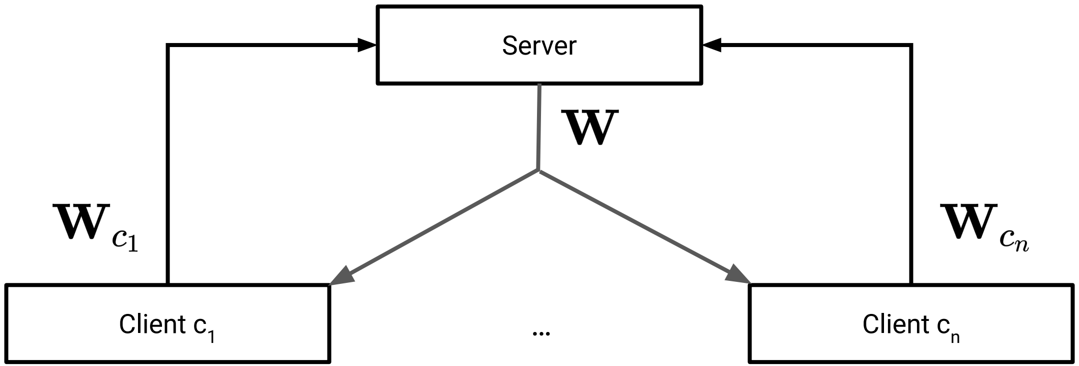


$$\mathbf{x}(t) = \tanh\left(\mathbf{W}_{in}\mathbf{u}(t) + \widehat{\mathbf{W}}\mathbf{x}(t-1)\right) \qquad \mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) \ \mathit{or} \ \mathbf{y}(t) = f(\mathbf{W}\mathbf{x}(t))$$

Federated Learning

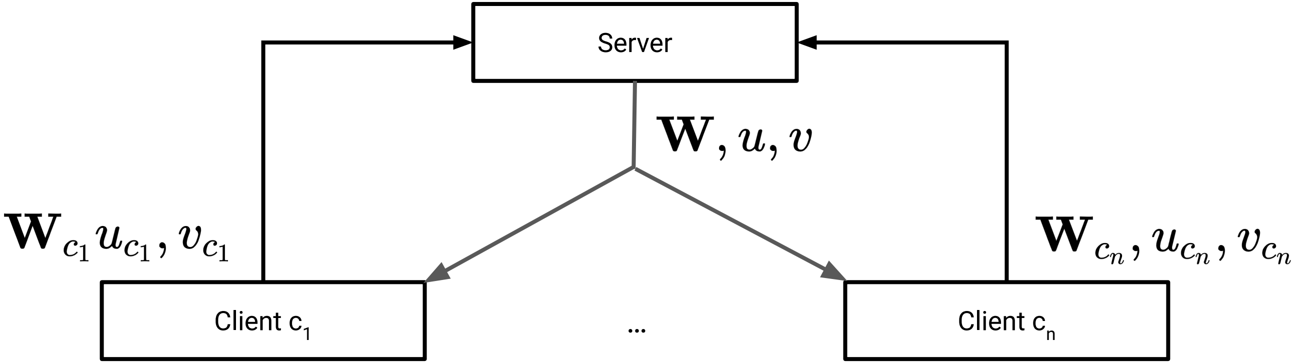


Federated Averaging (FedAvg)



$$\mathbf{W}_{t+1} \leftarrow \sum_{c_i} \frac{n_{c_i}}{n} \mathbf{W}_{t+1,c_i}$$

Federated Curvature (FedCurv)

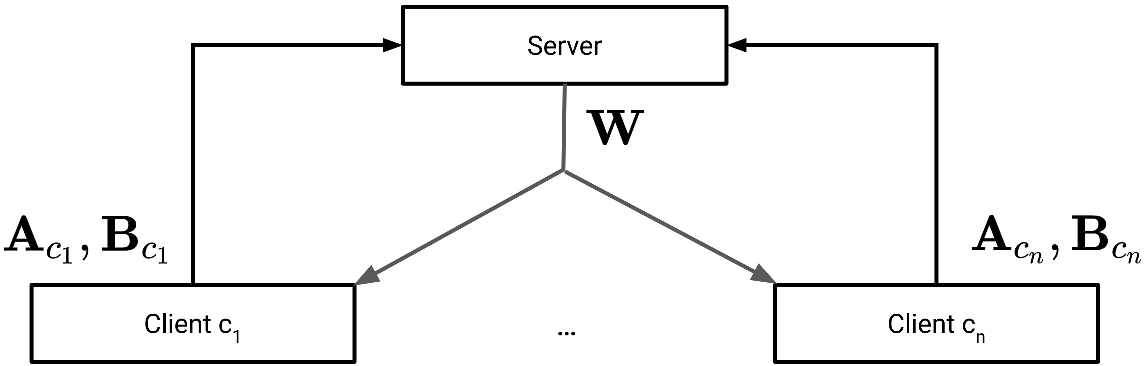


Each client computes → $\mathcal{L}_{t+1,c_i} = \mathcal{L}_{c_i}(\mathbf{W}_{c_i}) + \lambda \mathbf{W}_{t+1,c_i}^T \left[\sum_{c \in C \setminus c_i} u_{t+1,c}\right] \mathbf{W}_{t+1,c_i} - 2\lambda \left[\sum_{c \in C \setminus c_i} v_{t+1,c}\right] \mathbf{W}_{t,c_i}$

$$\mathbf{W}_{t+1} \leftarrow \sum_{c_i} \frac{n_{c_i}}{n} \mathbf{W}_{t+1,c_i}$$

$$u_{t+1} \leftarrow \sum_{c_i} u_{t+1,c_i} \qquad v_{t+1} \leftarrow \sum_{c_i} v_{t+1,c_i}$$

Incremental Federated Learning (IncFed)

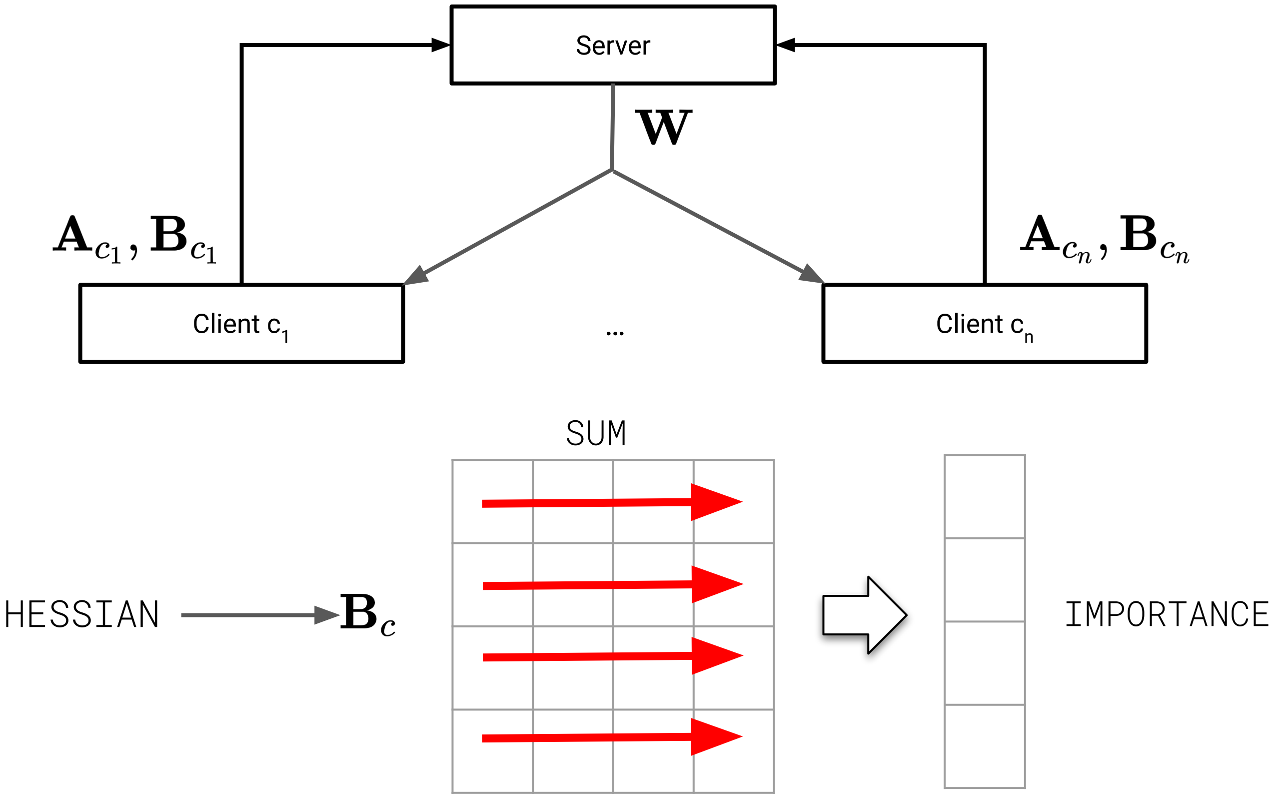


$$\mathbf{A}_c = \mathbf{Y}_c \mathbf{S}_c^T \ and \ \mathbf{B}_c = \mathbf{S}_c \mathbf{S}_c^T$$

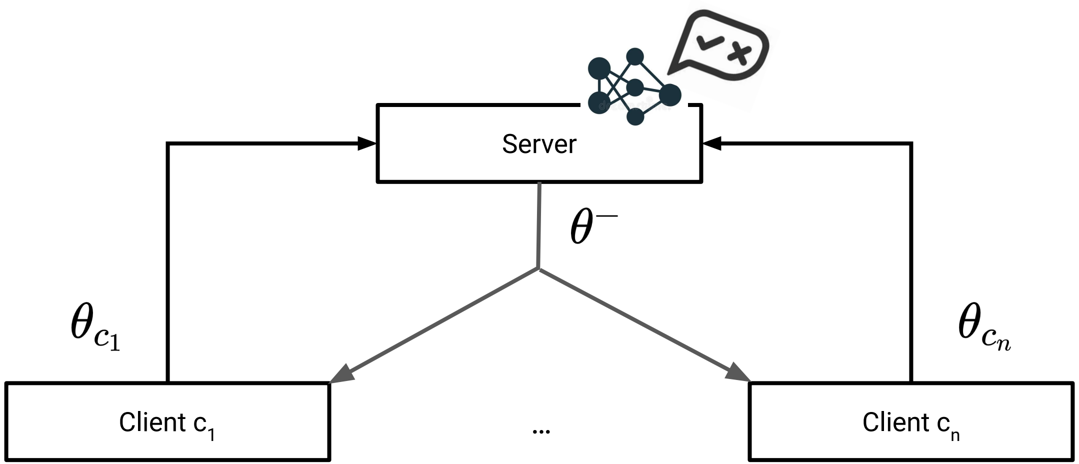
$$\mathbf{A} = \sum_{c \in \mathcal{C}} \mathbf{A}_c \ and \ \mathbf{B} = \left(\sum_{c \in \mathcal{C}} \mathbf{B}_c \right) + \beta \mathbf{I}$$

$$\mathbf{W} = \mathbf{A}\mathbf{B}^{-1}$$

Partial IncFed



Discriminator for Anomalous Clients



Dataset

- **WESAD** is a publicly available dataset for wearable **stress** and **affect detection**
- Physiological and motion data recorded from both a wrist/chest-worn device of **15 subjects**
- The signals are associated to specific cognitive state (label): **Baseline**, **Stress**, **Amusement**, **Meditation**
- **Reframing** the problem to **binary classification** (stressed/not stressed)

Experiments

- Choose the best ESN model on WESAD
- Model selection training/validation/test set splitting **of clients**: 10-4-1 subjects
- Best model chosen by F1 score on validation set

Units	α	η	ρ	Window	Batch	Features	Epochs
200	1.0	0.1	0.99	25×64	20	pEDA	1

Table 6.2: Best ESN model parameters associated with a validation F1 score of 0.9106.

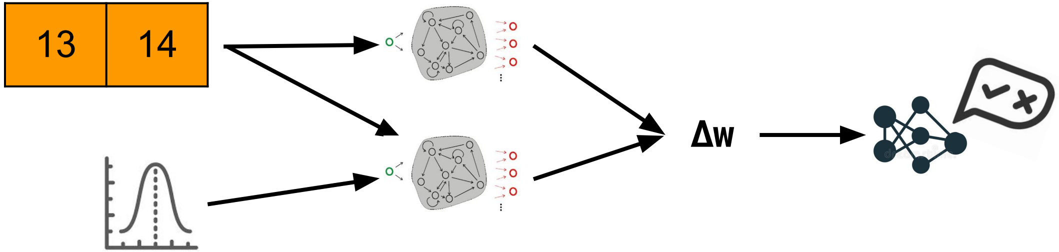
Federated Experiments

1. **Primal phase**: same dataset splitting end-to-end model

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

2. **Discriminative phase**: reduced validation set

1	2	3	4	5	6	7	8	9	10	11	12			15



Results

ESN Models	Train Acc.	Test Acc.	Train F1	Test F1
`SOTA` (multi-label class.)	0.8378	0.7792	-	-
`Central` (binary class.)	0.8986	0.9018	0.8986	**0.9018**
`FedAvg-1.0`	-	-	0.8878	**0.8693**
`FedAvg-0.1`	-	-	0.5913	0.4741
`FedAvg-0.01`	-	-	0.5784	0.4560
`FedAvg-0.001`	-	-	0.5711	0.4451
`FedCurv-1.0`	-	-	0.7299	0.5813
`FedCurv-0.1`	-	-	0.8878	**0.8693**
`FedCurv-0.01`	-	-	0.7646	0.6580
`FedCurv-0.001`	-	-	0.8140	0.7562
`FedAvg-1.0+Disc`	-	-	0.7601	0.6262
`FedAvg-0.1+Disc`	-	-	0.4415	0.5392
`FedAvg-0.01+Disc`	-	-	0.5652	0.5710
`FedAvg-0.001+Disc`	-	-	0.2972	0.4233
`FedCurv-1.0+Disc`	-	-	0.3427	0.4741
`FedCurv-0.1+Disc`	-	-	0.4456	0.5405
`FedCurv-0.01+Disc`	-	-	0.3251	0.4451
`FedCurv-0.001+Disc`	-	-	0.7136	0.5891

Table 6.4: Overall comparison of the best models with different FL algorithms. The reservoirs of the clients in all the FL algorithms are equal to the server.

ESN Models	Train Acc.	Test Acc.	Train F1	Test F1
`SOTA` (multi-label class.)	0.8378	0.7792	-	-
`Central` (binary class.)	0.8986	0.9018	0.8986	**0.9018**
`FedAvg-1.0`	-	-	0.8878	0.8693
`FedAvg-0.1`	-	-	0.5965	0.4772
`FedAvg-0.01`	-	-	0.5779	0.4538
`FedAvg-0.001`	-	-	0.5723	0.4457
`FedCurv-1.0`	-	-	0.8878	0.8693
`FedCurv-0.1`	-	-	0.8603	0.8693
`FedCurv-0.01`	-	-	0.8722	0.8693
`FedCurv-0.001`	-	-	0.5636	0.4379
`FedAvg-1.0+Disc`	-	-	0.9172	**0.9457**
`FedAvg-0.1+Disc`	-	-	0.8810	0.8693
`FedAvg-0.01+Disc`	-	-	0.3440	0.4744
`FedAvg-0.001+Disc`	-	-	0.4103	0.5187
`FedCurv-1.0+Disc`	-	-	0.8810	0.8693
`FedCurv-0.1+Disc`	-	-	0.9257	**0.8778**
`FedCurv-0.01+Disc`	-	-	0.8733	0.7955
`FedCurv-0.001+Disc`	-	-	0.3084	0.4332

Table 6.5: Overall comparison of the best models with different FL algorithms. The reservoirs of the clients in all the FL algorithms are different from the server.

Conclusions

- **Echo State Networks** suitable for FL because the aggregated model **is mathematically equivalent** to the **model trained** on the **whole dataset** and sends only the readout
- Our Echo State Network **outperforms** SOTA
- FedCurv more stable than FedAvg
- FedAvg+Disc **better** than the central model
- **Partial IncFed** has a better generalization and low communication cost w.r.t. **IncFed**

Training users	IncFed	Partial IncFed	Random
100%	0.7801±0.77	**0.8128±0.70**	0.8084±0.74
75%	0.7504±0.81	**0.8108±0.76**	0.7887±0.86
50%	0.7223±0.80	**0.7642±0.93**	0.7600±0.95
25%	0.7093±0.89	**0.7565±0.96**	0.7480±1.01

Future Works

- Use/create a federated learning framework for customized federated algorithms
- Explore different datasets
- Select the most important neurons using SVD, reducing the dimensionality of the model parameters

Thank you for your attention!

Bibliography I

1. Cisco annual internet report 2018–2023. https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html
2. Davide Bacciu, Daniele Di Sarli, Pouria Faraji, Claudio Gallicchio, and Alessio Micheli. Federated reservoir computing neural networks. In 2021 International Joint Conference on Neural Networks (IJCNN), pages 1–7. IEEE, 2021
3. Jie Lin, Wei Yu, Nan Zhang, Xinyu Yang, Hanlin Zhang, and Wei Zhao. A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications. IEEE internet of things journal, 4(5):1125–1142, 2017
4. Mehdi Mohammadi and Ala Al-Fuqaha. Enabling cognitive smart cities using big data and machine learning: Approaches and challenges. IEEE Communications Magazine, 56(2):94–101, 2018
5. Yaohua Sun, Mugen Peng, Yangcheng Zhou, Yuzhe Huang, and Shiwen Mao. Application of machine learning in wireless networks: Key techniques and open issues. IEEE Communications Surveys & Tutorials, 21(4):3072–3108, 2019
6. Cisco global cloud index 2016-2021. https://www.cisco.com/c/en/us/solutions/executive-perspectives/annual-internet-report/index.html
7. Jorge Granjal, Edmundo Monteiro, and Jorge Sá Silva. Security for the internet of things: a survey of existing protocols and open research issues. IEEE Communications Surveys & Tutorials, 17(3):1294–1312, 2015
8. Teaching unipi project. https://www.unipi.it/index.php/risultati-e-prodotti/item/17229-teaching
9. Mantas Lukoševičius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network training. Computer Science Review, 3(3):127–149, 2009
10. Kenji Doya et al. Bifurcations in the learning of recurrent neural networks 3. learning (RTRL), 3:17, 1992
11. Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks, 5(2):157–166, 1994
12. Felix A Gers, Jürgen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with lstm. Neural computation, 12(10):2451–2471, 2000
13. Wolfgang Maass, Thomas Natschläger, and Henry Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. Neural computation, 14(11):2531–2560, 2002

Bibliography II

1. Herbert Jaeger. The “echo state” approach to analysing and training recurrent neural networks-with an erratum note. Bonn, Germany: German National Research Center for Information Technology GMD Technical Report, 148(34):13, 2001
2. Herbert Jaeger. Echo state network. scholarpedia, 2(9):2330, 2007
3. Jacob Poushter et al. Smartphone ownership and internet usage continues to climb in emerging economies. Pew research center, 22(1):1–44, 2016
4. Monica Anderson. Technology device ownership: 2015. 2015
5. Jakub Konečnỳ, H Brendan McMahan, Felix X Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon. Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492, 2016
6. Dinh C Nguyen, Ming Ding, Pubudu N Pathirana, Aruna Senevi- ratne, Jun Li, and H Vincent Poor. Federated learning for internet of things: A comprehensive survey. IEEE Communications Surveys & Tutorials, 23(3):1622–1658, 2021
7. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In Artificial intelligence and statistics, pages 1273–1282. PMLR, 2017
8. James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. Proceedings of the na- tional academy of sciences, 114(13):3521–3526, 2017
9. Luc Pronzato and Andrej Pázman. Design of experiments in non- linear models. Lecture notes in statistics, 212:1, 2013
10. Varun Chandola, Arindam Banerjee, and Vipin Kumar. Outlier detection: A survey. ACM Computing Surveys, 14:15, 2007
11. Charu C Aggarwal. An introduction to outlier analysis. In Outlier analysis, pages 1–34. Springer, 2017
12. Raghavendra Chalapathy and Sanjay Chawla. Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407, 2019
13. Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In Proceedings of the 20th ACM international conference on multimodal interaction, pages 400–408, 2018
14. Sashank Reddi, Zachary Charles, Manzil Zaheer, Zachary Gar- rett, Keith Rush, Jakub Konečnỳ, Sanjiv Kumar, and H Bren- dan McMahan. Adaptive federated optimization. arXiv preprint arXiv:2003.00295, 2020