Facilitating Experimental Reproducibility in Neural Network Research with a Unified Framework

Anestis Kaimakamidis
Aristotle University of Thessaloniki
Thessaloniki, Greece
akaimak@csd.auth.gr

Ioannis Pitas
Aristotle University of Thessaloniki
Thessaloniki, Greece
pitas@csd.auth.gr

ABSTRACT

In the realm of neural network research, achieving experiment reproducibility is paramount for building upon existing knowledge and advancing the field. This paper examines a multi-agent neural network framework on its ability to facilitate the reproduction of experiments. Also, we address the reproducibility problem when there are data or source code limitations. The framework offers crucial functionalities for facilitating experiment reproducibility achieved through data, layer outputs, architectures, and weights exchange among the framework's agents. Through the integration of these functionalities, this framework empowers researchers to reproduce and validate experimental results consistently, fostering a more robust and collaborative research environment in the field of neural networks. The experimental results demonstrate the framework's reproducibility abilities. Furthermore, we test the framework in terms of reproducibility in an emergency natural disaster management situation. Finally, we analyze how the privacy limitations of the original neural network affect the reproducibility results.

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**.

KEYWORDS

Reproducibility, Natural Disaster Management, Fire Classification, Teacher-Student Learning, Data Privacy

ACM Reference Format:

Anestis Kaimakamidis and Ioannis Pitas. 2023. Facilitating Experimental Reproducibility in Neural Network Research with a Unified Framework. In *IEEE/ACM 10th International Conference on Big Data Computing, Applications and Technologies (BDCAT '23), December 4–7, 2023, Taormina (Messina), Italy.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3632366.3632368

1 INTRODUCTION

Machine Learning (ML) and Deep Learning (DL) methods find application across a diverse spectrum of fields and contemporary challenges. ML has played a crucial role in domains including healthcare [4, 5], material sciences [8], economics [1], radiology [9], civil engineering [26], and sales and marketing [29]. ML algorithms have been used to solve recommendation systems [6, 28], image recognition [12], sentiment analysis [16, 25], and natural disaster management [20, 21].

Deep Neural Networks (DNNs) have achieved impressive results in various complex tasks, and much effort has gone into improving their predictive accuracy. However, an equally crucial aspect of any machine learning system is its ability to produce stable and reproducible predictions. The challenge of ensuring prediction reproducibility is a concern even when the network architecture and training data remain constant across different training runs. Unfortunately, two critical elements that contribute to the high accuracy of deep networks —over-parameterization and the randomization of training algorithms— present substantial challenges to achieving reproducibility. Over-parameterization means that neural networks often have multiple solutions that minimize the training objective, leading to different model solutions [23, 34]. Randomization, on the other hand, stems from various sources of uncertainty in standard neural network training, such as weight initialization, mini-batch ordering, non-deterministic aspects of training platforms, and, in some cases, data augmentation. When combined, these factors mean that training neural networks can result in vastly different solutions in each run, even when the training data remains the same, creating a significant challenge for reproducibility.

On the other hand, achieving reproducible experimental results can be particularly challenging because most papers do not provide the source code or data used for their experiments. Most reproducibility techniques require the exact original experimental setting (dataset, source code) and fail to proceed if any requirement is missing. This paper identifies the parts of the source code needed to reproduce the original deep learning experiment and uses a multi-agent neural network framework to facilitate the process. To this end, we define these availability limitations as *privacy* and we introduce the five different privacy levels the framework can handle.

Reproducibility hinges on the ability to recreate a neural network model with minimum disagreement with the original one, ensuring that the model's behavior remains consistent. However, methods to mitigate this disagreement have only been presented on the over-parameterization and randomization problem. This paper examines the reproducibility solutions regarding the availability problem (privacy) using the framework described in [15]. This framework introduces a novel approach by creating a collaborative neural network environment composed of agents capable of exchanging information and learning tasks from each other. In this context, solving the reproducibility problem entails successful collaboration between a newly introduced agent and the agent to be reproduced. The framework offers three key mechanisms to enhance reproducibility: a) by enabling data exchange between the agents, b) by enabling weights exchange between the agents, and

BDCAT '23, December 4–7, 2023, Taormina (Messina), Italy

ACM ISBN 979-8-4007-0473-4/23/12.
https://doi.org/10.1145/3632366.3632368

c) by enabling learning via knowledge distillation techniques. This paper aims to assess and validate the reproducibility capabilities of this framework. By doing so, it seeks to aid the process of reproducing experimental results within the scientific community, thereby advancing modern research in the field of Deep Learning. For practical applications, these kinds of frameworks can be used to transfer knowledge safely, effectively, and quickly in emergencies, such as natural disasters.

2 RELATED WORK

2.1 Reproducibility

Reproducibility is a widely discussed [11] and concerning issue across various scientific disciplines [2, 30], highlighting a potential crisis in this regard. Many scientists have encountered difficulties when attempting to reproduce experimental findings [2], with failure rates exceeding 50% when attempting to reproduce their work in fields like medicine, physics, and engineering. The failure rate rises to more than 75% when attempting to reproduce results from the works of others in the same fields.

In the field of computational biology, reproducibility has been identified as a significant challenge [24]. Studies have shown that the percentage of reproducible studies in this field can be as low as 10% or less [24]. This lack of reproducibility has been attributed to factors such as unclear method descriptions and inherent variability in biological systems [24]. To address this issue, efforts have been made to improve reproducibility in computational biology research. Tools have been developed to facilitate reproducibility, such as comprehensive code libraries and online evaluation platforms [10]. These tools aim to provide researchers with the necessary resources to accurately reproduce and validate computational models and results [10].

Reproducibility is also a concern in the field of computer vision. In the context of face presentation attack detection competitions, reproducibility has been emphasized as a key aspect [33]. In a specific competition, the top-performing teams released their source code and summarized their approaches, promoting transparency and reproducibility [33]. This highlights the importance of sharing code and methodologies to enable others to reproduce and build upon research findings.

In the field of deep learning, related work until recently has been focused on the mitigation of the disagreement between the original and the reproduced neural network, due to over-parametrization and randomness [23, 34]. Another concern regarding the reproducibility of a research experiment is the dataset and code availability. Although works like [33] emphasize the code and dataset availability problem, there are limited to no ways of handling it.

To assess the different reproducibility techniques the need for a metric emerges. The metric that addresses this issue is defined as *churn* [22] and refers to the disagreement in predictions between two models. Essentially, churn represents the fraction of test examples where the predictions made by these two models do not align. It's important to note that when both models have perfect accuracy, which is often unattainable in practical scenarios, the churn is zero. While it is possible to reduce churn by removing all sources of uncertainty in the training process, such as controlling the random initialization seed and data order, it remains challenging due to inherent non-determinism in the current computational algorithms.

2.2 Teacher-Student Learning

The concept of teacher-student learning originated from the idea of condensing the knowledge stored in one or multiple complex neural networks into a single, simpler student network [7]. This approach balances efficiency and performance in learning tasks [3]. In a study conducted by [19], a large pre-trained network can provide labels for unlabeled data. Notably, [13] achieved significant results by distilling knowledge from a group of models into a single student model. Expanding on the work of [19] and [13], subsequent research focused on teacher-student interactions to refine the process of knowledge distillation, leading to highly proficient student models [32], [18], [35], [27], [3], [31].

In a study by [27], a novel framework is introduced to compress extensive and deep networks into narrower yet deeper ones. This method, known as FitNets, utilizes the teacher network's outputs and intermediate representations to train a more profound but narrower student network. Consequently, this approach enhances the training process and improves student performance. [32] defined distilled knowledge as the procedural flow learned through intermediate-layer feature representation. Considering this perspective, the student deep neural network surpasses the abilities of the teacher DNN. The capabilities of teacher-student interaction frameworks could be utilized to bolster deep neural network reproducibility.

3 UTILISING FRAMEWORK FOR REPRODUCIBILITY

3.1 Brief framework description

The framework proposed in [15] facilitates efficient knowledge transfer and collaboration by establishing a network of *agents*. Each agent consists of an out-of-distribution detector, a classifier, and a set of rules for training, knowledge distillation, and out-of-distribution detection, as shown in Figure 1. The agents can use their set of rules to train their classifier, collaborate with other agents, and assess their knowledge of a given dataset. The framework also defines the rules for agent communication, that enable data, weights, feature maps, and soft-target transmission. By utilizing the framework communication rules, agents can exchange knowledge seamlessly in a collaborative environment. The agents either possess or need to acquire knowledge, as a result, they can act both as students and teachers. The framework also provides a library that consists of widely used architectures to avoid discrepancies between architecture variants.

3.2 Framework reproducibility functionalities

Experiment reproducibility is bolstered through a well-defined framework for neural network communications. The framework proposed in [15] offers a plethora of options for data and knowledge transmission among its agents. It defines the rules of communication and collaboration between different agents, facilitating the knowledge exchange among them. Let us now suppose that we

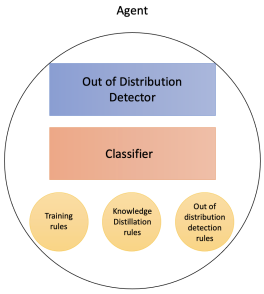

Figure 1: The representation of a framework's agent, containing an out-of-distribution detector, a classifier, and a set of rules. Knowledge distillation rules enable agents to change roles (teacher/student) depending on the experiment setting [15]

need to reproduce the experiments of a deep learning research paper. This paper can either propose a novel neural network architecture, a pre-processing method, a learning algorithm, or include an application on a specific dataset. Considering the framework's setting, we are going to refer to this paper as the *teacher agent*, which consists of a classifier model $\mathbf{a}^t = f(\mathbf{x}; \mathbf{w}^t)$, where f is the model decision function parametrized by $\mathbf{w}^t$, $\mathbf{x}$ is the model inputs and $\mathbf{a}^t$ is the soft-target model prediction. Thus, the model prediction is $y^t = \arg\max(\mathbf{a}^t)$. Let us now suppose another researcher (*student agent*) wants to reproduce the results of the teacher agent on its classifier model: $\mathbf{a}^s = f(\mathbf{x}; \mathbf{w}^s)$, where, the model now is parametrized by $\mathbf{w}^s$, $\mathbf{x}$ is the model inputs and $\mathbf{a}^s$ is the model prediction and $y^s = \arg\max(\mathbf{a}^s)$. With this notation, the disagreement (churn) between the teacher and the student agent for a set of n predictions $\mathbf{y}^t = \left[y_1^t, ..., y_n^t\right]$ and $\mathbf{y}^s = \left[y_1^s, ..., y_n^s\right]$ is defined as $Churn\left(\mathbf{y}^t, \mathbf{y}^s\right) = \frac{1}{n}\sum_{i=1}^{n} g(y_i^t, y_i^s)$ where,

$$g(y_i^t, y_i^s) = \begin{cases} 1, & y_i^t \neq y_i^s \\ 0, & y_i^t = y_i^s. \end{cases}$$

The teacher's classifier model is trained on sample training data $\mathcal{D}^t = \{(\mathbf{x_i}, y_i)\}_{i=0}^{M}$, where M is the total number of samples.

The framework provides five options regarding reproducibility, also considering the privacy limitations of the teacher agent:

(1) **Training data and architecture**: The teacher agent sends its training data ($\mathcal{D}^t$) and network architecture and the student agent can reproduce the training process.
(2) **Knowledge distillation with teacher's training dataset**: The teacher agent sends its training data ($\mathcal{D}^t$), network architecture, and soft-targets ($\mathbf{a}^t$) for every sample of ($\mathcal{D}^t$) so that the student can reproduce the teacher's network using knowledge distillation.
(3) **Knowledge distillation without teacher's training dataset**: In case the teacher's training data are private, the student agent sends its training data to the teacher. The teacher agent sends its network architecture and soft-targets (of the student's training data $\mathcal{D}^s$) so that the student can reproduce the teacher's network using knowledge distillation.
(4) **Knowledge distillation without teacher's training dataset or architecture**: In case the teacher's training data and architecture are private, the student agent sends its training data ($\mathcal{D}^s$) to the teacher. The teacher agent sends the soft-targets ($\mathbf{a}^t$) produced from the student's training data. The student loads an architecture from the architecture library and uses knowledge distillation to reproduce the teacher's network.
(5) **Architecture and weights**: Having the minimum privacy possible, the teacher can send its architecture and weights ($\mathbf{w}^t$) for the exact teacher's network reproduction.

Table 1: Shared information of the teacher agent for reproducibility purposes according to the provided options

Options	Dataset	Architecture	Soft-Targets	Weights
1	✓	✓	✗	✗
2	✓	✓	✓	✗
3	✗	✓	✓	✗
4	✗	✗	✓	✗
5	✗	✓	✗	✓

In the context of the problem at hand, the term privacy refers to the degree of transparency the teacher agent has. Regarding the reproducibility topic, in the context of the framework examined, weights and architecture transmission is considered the most transparent (non-private) option. Option 4 constitutes the most private option. The three remaining options display a level of privacy with ranking: 2, 1, and 3 from lowest to highest privacy. In fact, option 2 is a special case of option 3 where the same dataset used for the teacher's training is used to distill the teacher's knowledge to the student. The difference between 1 and 2 is the contribution of the teacher agent in the training process of the student agent, which is really useful for reducing the disagreement between the two networks. The reason why option 1 is defined as a more private option than option 3 can be explained simply by understanding that actual datasets contain more information than the soft-target activations on the circumstances the teacher agent is trained, thus it facilitates the churn reduction. These remarks are illustrated in Table 1, which summarizes the information of the teacher agent shared for every option. In our experiments, we will compare the different reproducibility options in terms of privacy and churn.

3.3 Reproducibility for Natural Disaster Management

The framework's reproducibility options have direct application to Natural Disaster Management systems. The student agent in this case can either be a natural disaster control center, a drone, or another vehicle suitable for this situation. The student agent can collect data from the disaster at hand and accumulate the knowledge of the other agents seamlessly by distilling their knowledge. Another useful feature is the weights transmission, as it is quick and applicable directly, saving time which is crucial in situations like natural disasters. As extensively discussed before, the framework provides a plethora of options taking into consideration the privacy

limitations of the teacher agent. This feature ensures that the original neural network can be reproduced under many cases of privacy limitations, which is extremely useful for natural disaster management. Knowledge about wildfires, flooding, tropical cyclones, tsunamis, volcanic activity, and other kinds of natural disasters, can be transferred effectively and quickly using this framework. This enables the users of natural disaster management systems to acquire useful knowledge seamlessly and handle the situation more effectively. The experimental results will demonstrate the framework's capabilities in a wildfire situation.

4 EXPERIMENTAL RESULTS

The experimental setup, which verifies the effect of the framework's functionalities on the reproducibility problem, will be extensively discussed in this section. According to the previous rationale, the teacher agent is the agent to be reproduced on another agent (student). Let us now suppose that the teacher agent represents a neural network on a fire monitoring drone, performing fire classification [17]. The natural disaster control center claims that a second drone is required for monitoring. The problem at hand requires reproducing the neural network of the first drone (teacher agent) to the second drone (student agent) with the minimum possible disagreement.

The fire dataset consists of 1900 images divided into 1520 for training and 380 for testing. The dataset contains annotated images with labels indicating the existence or the absence of fire in the image, as shown in Figure 2. The teacher's architecture is the ResNet18 architecture with a measured accuracy score of 96.05%. We are going to test the framework's ability to reproduce the teacher agent's results, according to the different privacy limitations. To this end, we measured the churn and accuracy metrics for all different reproducibility options. The experimental setting for options 1 and 2 is clear, option 1 is the repetition of the original training process and option 2 is the simplest knowledge distillation technique [13]. Regarding option 3, the student agent has managed to access only a subset of the teacher's training dataset, thus this subset is used for training. For option 4 the student uses the same subset as option 3 for training, however, the teacher's architecture is private. The student has the option to pick an architecture from the architecture library. For this experiment, the student picks MobileNetV3 [14]. Option 5 is also standard procedure and is applied by copying the teacher's architecture and weights.

Taking into consideration the definition of privacy as described above we can define different privacy levels for better results interpretation and visualization. To this end, we designate option 5 as having level 1 privacy, while options 1 to 3 are allocated level 3, level 2, and level 4 privacy, respectively. Option 4 is allocated level 5 privacy. Table 2 presents the accuracy and churn metrics for the options provided by the framework. The results show the deviation in churn and accuracy of the experiment reproducibility process, according to the different levels of privacy of the teacher agent. Table 2 along with Figure 3 indicate the effect of different levels of privacy on the churn metric. The more private the components of the teacher agent are, the more disagreement will emerge between the teacher and the student agent.

(a) No fire dataset example **(b) Fire dataset example**

Figure 2: Examples of the forest fire classification dataset [17]

Table 2: Accuracy and Churn measurements for the different options provided by the framework

Options	Privacy	Accuracy	Churn
1	level 3	94.92%	4.05%
2	level 2	96.05%	3.74%
3	level 4	93.87%	4.05%
4	level 5	92.82%	5.74%
5	level 1	96.05%	0

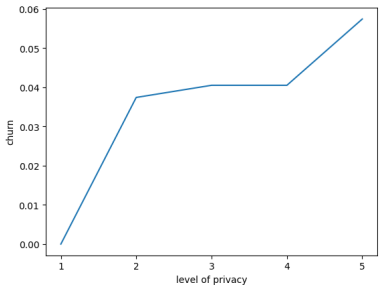

Figure 3: Relationship between different levels of privacy and the churn metric

The experimental results prove the framework's reproducibility abilities and its options constitute a useful solution for the privacy limitations of the teacher agent. The experiments exhibit good reproducibility results in terms of churn, i.e. the original and the reproduced model disagreement. The possibilities of the framework can be demonstrated by its ability to replicate deep learning models, taking into consideration real-world restrictions and limitations. Although the system is tested on a natural disaster management scenario, the method is directly applicable when reproducing the results of a research paper. The implementation of the framework on an accessible, worldwide scale promises to boost reproducibility and accelerate the research progress in every field using deep learning models, as the related work will be reproduced effortlessly, with the click of a button.

5 CONCLUSION

In conclusion, this paper has evaluated the reproducibility capabilities of a multi-agent neural network framework. We demonstrated the functionality of the examined framework in a natural disaster management setting, where the issue of safe and effective reproducibility is crucial. During the evaluation process, we considered privacy issues that are encountered frequently in natural disaster management and experiment reproducibility. The same privacy limitations can be applied to research papers, thus our work is directly applicable to research experiments. The framework empowers researchers to reproduce and validate experimental results consistently, fostering a more robust and collaborative research environment in the field of neural networks. The results are promising and future work should consider integrating churn reduction methods into the existing framework for achieving better reproducibility results.

6 ACKNOWLEDGEMENT

Funded by the European Union/European Commission (AI4EUROPE; Call: HORIZON-CL4-2021-HUMAN-01; Topic: HORIZON-CL4-2021-HUMAN-01-02; Type of action: HORIZON Coordination and Support Actions; Grant Agreement: 101070000). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union/European Commission. Neither the European Union nor the granting authority can be held responsible for them.

REFERENCES

[1] Susan Athey. 2018. The Impact of Machine Learning on Economics. In *The Economics of Artificial Intelligence: An Agenda*. National Bureau of Economic Research, Inc, 507–547.

[2] Monya Baker. 2016. 1,500 scientists lift the lid on reproducibility. *Nature* 533, 7604 (2016).

[3] Andreas Bar, Fabian Huger, Peter Schlicht, and Tim Fingscheidt. 2019. On the robustness of redundant teacher-student frameworks for semantic segmentation. (2019), 0–0.

[4] David Ben-Israel, W. Jacobs, Steve Casha, Stefan Lang, W.H. Andrew Ryu, Madeleine de Lotbiniere-Bassett, and David Cadotte. 2019. The impact of machine learning on patient care: A systematic review. *Artificial Intelligence in Medicine* 103 (12 2019), 101785. https://doi.org/10.1016/j.artmed.2019.101785

[5] Stefano Bini. 2018. Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *The Journal of Arthroplasty* 33 (02 2018). https://doi.org/10.1016/j.arth.2018.02.067

[6] Jesus Bobadilla, Fernando Ortega, A. Hernando, and A. Gutiérrez. 2013. Recommender systems survey. *Knowledge-Based Systems* 46 (07 2013), 109–132. https://doi.org/10.1016/j.knosys.2013.03.012

[7] C Bucilua, R Caruana, and A Niculescu-Mizil. 2006. Model compression, in proceedings of the 12 th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. *New York, NY, USA* 3 (2006).

[8] Garry Choy, Omid Khalilzadeh, Mark Michalski, Synho Do, Anthony Samir, Oleg Pianykh, Jr Geis, Pari Pandharipande, James Brink, and Keith Dreyer. 2018. Current Applications and Future Impact of Machine Learning in Radiology. *Radiology* 288 (06 2018), 171820. https://doi.org/10.1148/radiol.2018171820

[9] Dennis Dimiduk, Elizabeth Holm, and Stephen Niezgoda. 2018. Perspectives on the Impact of Machine Learning, Deep Learning, and Artificial Intelligence on Materials, Processes, and Structures Engineering. *Integrating Materials and Manufacturing Innovation* 7 (08 2018), 1–16. https://doi.org/10.1007/s40192-018-0117-8

[10] J. Ding. 2021. Object detection in aerial images: a large-scale benchmark and challenges. (2021). https://doi.org/10.48550/arxiv.2102.12219

[11] D. o. Earth and C. B. o. Behavioral. 2019. Reproducibility and replicability in science. (2019). https://doi.org/10.17226/25303

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[13] Geoffrey Hinton, Oriol Vinyals, Jeff Dean, et al. 2015. Distilling the Knowledge in a Neural Network. *arXiv preprint arXiv:1503.02531* vol. 2, no. 7 (2015).

[14] Andrew Howard, Ruoming Pang, Hartwig Adam, Quoc Le, Mark Sandler, Bo Chen, Weijun Wang, Liang-Chieh Chen, Mingxing Tan, Grace Chu, Vijay Vasudevan, and Yukun Zhu. 2019. Searching for MobileNetV3. 1314–1324. https://doi.org/10.1109/ICCV.2019.00140

[15] Anestis Kaimakamidis, Ioanna Valsamara, and Ioannis Pitas. 2023. Learning by Education Multi-Agent Framework. (2023).

[16] Dionisis Karamouzas, Ioannis Mademlis, and Ioannis Pitas. 2022. Public opinion monitoring through collective semantic analysis of tweets. *Social Network Analysis and Mining* (07 2022). https://doi.org/10.1007/s13278-022-00922-8

[17] A. Khan, B. Hassan, S. Khan, R. Ahmed, and A. O. Abuassba. 2022. Deepfire: a novel dataset and deep transfer learning benchmark for forest fire detection. *Mobile Information Systems* 2022 (2022), 1–14. https://doi.org/10.1155/2022/5358359

[18] Nikos Komodakis and Sergey Zagoruyko. 2017. Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer. In *ICLR*.

[19] Jinyu Li, Rui Zhao, Jui-Ting Huang, and Yifan Gong. 2014. Learning small-size DNN with output-distribution-based criteria. In *Fifteenth annual conference of the international speech communication association*.

[20] Vasilis Linardos, Maria Drakaki, Panagiotis Tzionas, and Yannis Karnavas. 2022. Machine Learning in Disaster Management: Recent Developments in Methods and Applications. *Machine Learning and Knowledge Extraction* 4 (05 2022), 446–473. https://doi.org/10.3390/make4020020

[21] Vasilis Linardos, Maria Drakaki, Panagiotis Tzionas, and Yannis Karnavas. 2022. Machine Learning in Disaster Management: Recent Developments in Methods and Applications. *Machine Learning and Knowledge Extraction* 4 (05 2022), 446–473. https://doi.org/10.3390/make4020020

[22] Mahdi Milani Fard, Quentin Cormier, Kevin Canini, and Maya Gupta. 2016. Launch and Iterate: Reducing Prediction Churn. In *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Eds.), Vol. 29. Curran Associates, Inc.

[23] Behnam Neyshabur, Ryota Tomioka, and Nathan Srebro. 2014. In Search of the Real Inductive Bias: On the Role of Implicit Regularization in Deep Learning. (12 2014).

[24] J. Papin, F. Gabhann, H. Sauro, D. Nickerson, and A. Rampadarath. 2020. Improving reproducibility in computational biology research. *Plos Computational Biology* 16 (2020), e1007881. Issue 5. https://doi.org/10.1371/journal.pcbi.1007881

[25] Emmanouil Patsiouras, Ioanna Koroni, Ioannis Mademlis, and Ioannis Pitas. 2023. GreekPolitics: Sentiment Analysis on Greek Politically Charged Tweets. *31st European Signal Processing Conference (EUSIPCO)* (09 2023).

[26] Yoram Reich. 1996. Machine Learning Techniques for Civil Engineering Problems. *Microcomputers in Civil Engineering* 12 (12 1996). https://doi.org/10.1111/0885-9507.00065

[27] Adriana Romero, Nicolas Ballas, Samira Ebrahimi Kahou, Antoine Chassang, Carlo Gatta, and Yoshua Bengio. 2014. Fitnets: Hints for thin deep nets. *arXiv preprint arXiv:1412.6550* (2014).

[28] J Ben Schafer, Joseph Konstan, and John Riedl. 1999. Recommender systems in e-commerce. In *Proceedings of the 1st ACM conference on Electronic commerce*. 158–166.

[29] Keng Siau and Yin Yang. 2017. Impact of Artificial Intelligence, Robotics, and Machine Learning on Sales and Marketing.

[30] Aaron Stupple, David Singerman, and Leo Anthony Celi. 2019. The reproducibility crisis in the age of digital medicine. *NPJ digital medicine* 2, 1 (2019), 2.

[31] Fei Ye and Adrian G Bors. 2021. Lifelong teacher-student network learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 10 (2021), 6280–6296.

[32] Junho Yim, Donggyu Joo, Jihoon Bae, and Junmo Kim. 2017. A gift from knowledge distillation: Fast optimization, network minimization and transfer learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4133–4141.

[33] Z. Yu, J. Komulainen, X. Li, and G. Zhao. 2021. Review of face presentation attack detection competitions. (2021). https://doi.org/10.48550/arxiv.2112.11290

[34] Ying Zhang, Tao Xiang, Timothy M. Hospedales, and Huchuan Lu. 2018. Deep Mutual Learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[35] Guorui Zhou, Ying Fan, Runpeng Cui, Weijie Bian, Xiaoqiang Zhu, and Kun Gai. 2018. Rocket launching: A universal and efficient framework for training well-performing light net. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.