JOSE
The Journal of Open Source Education

# Introduction to deep learning: Carpentries-style hands-on lesson material for introducing researchers to deep learning

**Sven A. van der Burg** [1], **Pranav Chandramouli** [1], **Anne Fouilloux** [2], **Cunliang Geng** [1], **Toby Hodges** [8], **Florian Huber** [1,3], **Dafne van Kuppevelt** [1], **Ashwin Vishnu Mohanan** [9], **Colin Sauze** [5], **Carsten Schnober** [1], **Djura Smits** [1], **Peter Steinbach** [4], **Berend Weel** [1], **Kjartan Thor Wikfeldt** [9], **and Samantha Wittke** [6,7]

**1** Netherlands eScience Center, Amsterdam, The Netherlands **2** Simula Research Laboratory, Oslo, Norway **3** Düsseldorf University of Applied Sciences, Düsseldorf, Germany **4** Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany **5** National Oceanography Centre, Liverpool, Great-Britain **6** CSC - IT center for Science, Espoo, Finland **7** Aalto University, Espoo, Finland **8** The Carpentries, USA **9** RISE Research Institutes of Sweden, Sweden

**DOI:** 10.xxxxxx/draft

**Software**

- Review
- Repository
- Archive

**Submitted:** 03 March 2025
**Published:** unpublished



## Summary

This article describes a hands-on introduction to the first steps in deep learning, intended for researchers who are familiar with (non-deep) machine learning.

The use of deep learning has seen a sharp increase in popularity and applicability over the last decade. While deep learning can be a useful tool for researchers from a wide range of domains, taking the first steps in the world of deep learning can be somewhat intimidating. This introduction aims to cover the fundamentals of deep learning in a practical and hands-on manner. By the end of the course, students will be able to train their first neural network and understand the subsequent steps needed to improve the model.

The lesson starts by explaining the basic concepts of neural networks, and then guides learners through the different steps of a deep learning workflow.
After following this lesson, learners will be able to prepare data for deep learning, implement a basic deep learning model in Python with Keras, and monitor and troubleshoot the training process. In addition, they will be able to implement and understand different layer types, such as convolutional layers and dropout layers, and apply transfer learning.

We use data with permissive licenses and designed for real world use cases:

- The Penguin dataset (Horst et al. (2020))
- The Weather prediction dataset (Huber et al. (2022))
- The Dollar Street Dataset (Gaviria Rojas et al. (2022)) is representative and contains accurate demographic information to ensure their robustness and fairness, especially for smaller subpopulations.

## Statement of Need

This lesson addresses the need for an introductory lesson on deep learning that is open-source, and can be used by instructors in a workshop as well as for self-study. While generally usable, its target audience are academic researchers.

JOSE
The Journal of Open Source Education

There are many free online course materials on deep learning, see for example: *Fast.ai - Practical Deep Learning for Coders* (n.d.); "Udemy - Basics of Deep Learning" (n.d.); "Udemy - Tensorflow 2.0 | Recurrent Neural Networks, LSTMs, GRUs" (n.d.); "Free Deep Learning Tutorial - Data Science" (n.d.); "Coursera - Deep Learning" (n.d.); "freeCodeCamp.org - Learn PyTorch for Deep Learning" (2022).

Nonetheless, these resources are often not available open-source and can thus not be easily adapted to the students' needs. Also, these resources are intended to use for self-study. Our material can be used for self-study, but it is primarily developed for instructors to use in a workshop. In addition, although a diverse range of online courses already exists, few are targeted towards academic researchers.

There is another Carpentries lesson on deep learning: Introduction to artificial neural networks in Python (Pollard et al. (2022)). That lesson takes a different angle to deep learning, focusing on computer vision with the application on medical images. Whereas this lesson is a general introduction to applied deep learning showing various applications and is more mature.

Many computing centers offer (local) deep learning courses, such as "CSC- Practical Deep Learning" (n.d.). But the lesson material, if it is available, is not easily adopted outside the course organisation.

The pedagogical approach of this lesson is both to make learners familiar with the key concepts, and let them practice with how to implement them – eventually resulting in an increase in confidence and the conviction that 'I can do this myself'. The key to getting there is live coding: before the course, learners have to setup a working environment on their own computer. During the course, learners type in the commands that are explained by the instructor on their own computer. This design is based on the Software Carpentry (Wilson, 2006) philosophy. Live coding ensures that learners master the programmatic implementation of deep learning at the end of the course. We believe that this makes our lesson a unique and crucial resource.

Researchers can often only free a limited amount of time (maximum 5 consecutive days), since they are so involved in their daily work. To accomplish this, we created a lesson that can be taught in 2 consecutive days or 4 half days.

Demand for our workshops and feedback gathered from students demonstrated the need for a low-threshold lesson that lets researchers take the first steps in the field of deep learning. This impression was validated by other instructors who taught the lesson independently to their own audiences and provided us with feedback on their experience.

## Lesson Development

In 2018, the Netherlands eScience Center initiated the development of this lesson to fill the gap identified above. Over the years, the lesson has attracted a broad community of individuals and organizations that have used the material for teaching workshops, and contributed to the improvement of the lesson significantly.

The diversity of the involved parties has facilitated the integration of various viewpoints on the lesson material. Apart from the feedback gathered from students while teaching the workshop (see below), the mix of contributors includes educators, data scientists, and, most prominently, (research) software engineers. Some of them have had years of experience in the deep learning domain, while others have used the lesson as a first step into the field.

Development sprints of typically two full working days have regularly facilitated focussed collaboration sessions that have brought together various contributors to tackle specific issues identified in the lesson material. These sessions have also provided a fruitful ground

JOSE
The Journal of Open Source Education

for discussing the various experiences with and insights about the material. They have facilitated the iterative improvement of the material, resulting in a mature and well-tested set of episodes.

## Instructional design

This lesson material was designed using the concepts from The Carpentries Curriculum Development Handbook (Becker & Michonneau, n.d.). Most importantly, we used 'backward design': we started with identifying learning objectives, the core skills and concepts that learners should acquire as a result of the lesson. Next, exercises were designed to assess whether these objectives are met. Eventually, the content is written to teach the skills and concepts learners need to successfully complete the exercises and, it follows, meet the learning objectives.

Live coding is central to this approach: the lesson is built up of small blocks. In each block first the instructor demonstrates how to do something, and students follow along on their own computer. Then, the students work independently on exercises individually or in groups to test their skills. This approach integrates opportunities for guided practice throughout the lesson, promoting learning by helping learners build up a functioning mental model of the domain and transfer new knowledge from working memory to long-term memory. This is in accordance with research-based successful teaching strategies (Lang, 2021).

The lesson material is built in the new lesson template: Carpentries Workbench (*The Carpentries Workbench*, n.d.). This makes the lesson material a complete self-study resource. But it also serves as lesson material for the instructor teaching the lesson through live-coding, in that case the lesson material is only shared with students after the workshop as a reference. The lesson material can be toggled to the 'instructor view'. This allows to provide instructor notes on how to approach teaching the lesson, and these can even be included at the level of the lesson content. In addition, the Carpentries Workbench prioritises accessibility of the content, for example by having clearly visible figure captions and promoting alt-texts for pictures.

The lesson is split into a general introduction, and 4 episodes that cover 3 distinct increasingly more complex deep learning problems. Each of the deep learning problems is approached using the same 10-step deep learning workflow (https://carpentries-lab.github.io/deep-learning-intro/1-introduction.html#deep-learning-workflow).

By going through the deep learning cycle three times with different problems, learners become increasingly confident in applying this deep learning workflow to their own projects. We end with an outlook episode. Firstly, the outlook eposide discusses a real-world application of deep learning in chemistry (Huber et al., 2021). In addition, it discusses bias in datasets, large language models, and good practices for organising deep learning projects. Finally, we end with ideas for next steps after finishing the lesson.

## Feedback

This course was taught 13 times over the course of 4 years, both online and in-person, by the Netherlands eScience Center (Netherlands, https://www.esciencecenter.nl/) and Helmholtz-Zentrum Dresden-Rossendorf (Germany, https://www.hzdr.de/). Apart from the core group of contributors, the workshop was also taught at at least 3 independent institutes, namely: University of Wisconson-Madison (US, https://www.wisc.edu/), University of Auckland (New Zealand, https://www.auckland.ac.nz/), and EMBL Heidelberg (Germany, https://www.embl.org/sites/heidelberg/).

JOSE
The Journal of Open Source Education

An up-to-date list of workshops that the authors are aware of having using this lesson can be found in a `workshops.md` file in the GitHub repository.

In general, adoption of the lesson material by the instructors not involved in the project went well. The feedback gathered from our own and others' teachings was used to polish the lesson further.

## Student responses

The feedback we gathered from students is in general very positive, with some responses from students to the question 'What was your favourite or most useful part of the workshop. Why?' further confirming our statement of need:

> *I enjoyed the live coding and playing with the models to see how it would effect the results. It felt hands on and made it easy for me to understand the concepts.*
>
> *Well-defined steps to be followed in training a model is very useful. Examples we worked on are quite nice.*
>
> *The doing part, that really helps to get the theory into practice.*

Below are two tables summarizing results from our post-workshop survey. We use the students' feedback to continuously improve the lesson.

| | STRONGLY DISAGREE | DISAGREE | UNDECIDED | AGREE | STRONGLY AGREE | TOTAL | WEIGHTED AVERAGE |
|---|---|---|---|---|---|---|---|
| I can immediately apply what I learned at this workshop. | 0 | 5 | 6 | 19 | 8 | 38 | 3,8 |
| The setup and installation instructions for the lesson were complete and easy to follow. | 0 | 0 | 4 | 13 | 21 | 38 | 4,4 |
| Examples and tasks in the lesson were relevant and authentic | 0 | 0 | 5 | 19 | 14 | 38 | 4,2 |

Table 1: Agreement on statements by students from 2 workshops taught at the Netherlands eScience Center. The results from these 2 workshops are a good representation of the general feedback we get when teaching this workshop.

| | POOR | FAIR | GOOD | VERY GOOD | EXCELLENT | N/A | TOTAL | WEIGHTED AVERAGE |
|---|---|---|---|---|---|---|---|---|
| Introduction into Deep Learning | 0 (0%) | 2 (5%) | 10 (27%) | 8 (22%) | 17 (46%) | 0 (0%) | 37 | 4,1 |
| Classification by a Neural Network using Keras (penguins dataset) | 0 (0%) | 1 (3%) | 5 (13%) | 16 (42%) | 16 (42%) | 0 (0%) | 38 | 4,2 |
| Monitoring and Troubleshooting the learning process (weather dataset) | 0 (0%) | 0 (0%) | 4 (11%) | 18 (47%) | 16 (42%) | 0 (0%) | 38 | 4,3 |

JOSE
The Journal of Open Source Education

| | POOR | FAIR | GOOD | VERY GOOD | EXCELLENT | N/A | TOTAL | WEIGHTED AVERAGE |
|---|---|---|---|---|---|---|---|---|
| Advanced layer types (CIFAR-10/Dollarstreet-10 datasets) | 0 (0%) | 2 (5%) | 5 (13%) | 7 (18%) | 16 (42%) | 8 (21%) | 38 | 4,2 |

Table 2: Quality of the different episodes of the workshop as rated by students from 2 workshops taught at the Netherlands eScience Center. The results from these 2 workshops are a good representation of the general feedback we get when teaching this workshop.

### Carpentries Lab review process

Prior to submitting this paper the lesson went through the substantial review in the process of becoming an official Carpentries Lab (https://carpentries-lab.org/) lesson. This led to a number of improvements to the lesson. In general the accessibility and user-friendliness improved, for example by updating alt-texts and using more beginner-friendly and clearer wording. Additionally, the instructor notes were improved and many missing explanations of important deep learning concepts were added to the lesson.

Most importantly, the reviewers pointed out that the CIFAR-10 (*CIFAR-10 and CIFAR-100 Datasets*, n.d.) dataset that we initially used does not have a license. We were surprised to find out that this dataset, that is one of the most widely used datasets in the field of machine learning and deep learning, is actually unethically scraped from the internet without permission from image owners. As an alternative we now use 'Dollar street 10' (burg, 2024), a dataset that was adapted for this lesson from The Dollar Street Dataset (Gaviria Rojas et al. (2022)). The Dollar Street Dataset is representative and contains accurate demographic information to ensure their robustness and fairness, especially for smaller subpopulations. In addition, it is a great entry point to teach learners about ethical AI and bias in datasets.

You can find all details of the review process on GitHub: https://github.com/carpentries-lab/reviews/issues/25.

## Conclusion

This lesson can be taught as a stand-alone workshop to students already familiar with machine learning and Python. It can also be taught in a broader curriculum after an introduction to Python programming (for example: Azalee Bostroem et al. (2016)) and an introduction to machine learning (for example: *Scikit-Learn Course* (2023)). Concluding, the described lesson material is a unique and essential resource aimed at researchers and designed specifically for a live-coding teaching style. Hopefully, it will help many researchers to set their first steps in a successful application of deep learning to their own domain.



## Acknowledgements

We would like to thank all instructors and helpers that taught the course, and the community of people that left contributions to the project, no matter how big or small. Also, we thank Chris Endemann (University of Wisconson-Madison, US, https://www.wisc.edu/), Nidhi Gowdra (University of Auckland, New Zealand, https://www.auckland.ac.nz/), Renato Alves and Lisanna Paladin (EMBL Heidelberg, Germany, https://www.embl.org/sites/heidelberg/), that piloted this workshop at their

The Journal of Open Source Education

institutes. We thank the Carpentries for providing such a great framework for developing this lesson material. We thank Sarah Brown, Johanna Bayer, and Mike Laverick for giving us excellent feedback on the lesson during the Carpentries Lab review process. We thank all students enrolled in the workshops that were taught using this lesson material for providing us with feedback.

## References

Azalee Bostroem, Trevor Bekolay, & Valentina Staneva (eds). (2016). Software Carpentry: Programming with Python. In *GitHub*. https://github.com/swcarpentry/python-novice-inflammation, 10.5281/zenodo.57492

Becker, E., & Michonneau, F. (n.d.). *The Carpentries Curriculum Development Handbook*. Retrieved September 1, 2023, from https://cdh.carpentries.org/

burg, S. van der. (2024). *Dollar street 10 - 64x64x3*. Zenodo. https://doi.org/10.5281/zenodo.10970014

*CIFAR-10 and CIFAR-100 datasets*. (n.d.). Retrieved February 11, 2025, from https://www.cs.toronto.edu/~kriz/cifar.html

Coursera - Deep Learning. (n.d.). In *Coursera*. Retrieved September 1, 2023, from https://www.coursera.org/specializations/deep-learning

CSC- Practical Deep Learning. (n.d.). In *Eventilla*. Retrieved September 1, 2023, from https://ssl.eventilla.com/event/8aPek

*Fast.ai - Practical Deep Learning for Coders*. (n.d.). Retrieved September 1, 2023, from https://course.fast.ai/

freeCodeCamp.org - Learn PyTorch for Deep Learning. (2022). In *freeCodeCamp.org*. https://www.freecodecamp.org/news/learn-pytorch-for-deep-learning-in-day/

Gaviria Rojas, W., Diamos, S., Kini, K., Kanter, D., Janapa Reddi, V., & Coleman, C. (2022). The Dollar Street Dataset: Images Representing the Geographic and Socioeconomic Diversity of the World. *Advances in Neural Information Processing Systems*, *35*, 12979–12990. https://papers.nips.cc/paper_files/paper/2022/hash/5474d9d43c0519aa176276ff2c1ca528-Abstract-Datasets_and_Benchmarks.html

Horst, A. M., Hill, A. P., & Gorman, K. B. (2020). *Allisonhorst/palmerpenguins: v0.1.0*. Zenodo. https://doi.org/10.5281/zenodo.3960218

Huber, F., Burg, S. van der, Hooft, J. J. J. van der, & Ridder, L. (2021). MS2DeepScore: A novel deep learning similarity measure to compare tandem mass spectra. *Journal of Cheminformatics*, *13*(1), 84. https://doi.org/10.1186/s13321-021-00558-4

Huber, F., Kuppevelt, D. van, Steinbach, P., Sauze, C., Liu, Y., & Weel, B. (2022). *Weather prediction dataset*. Zenodo. https://doi.org/10.5281/ZENODO.4770936

Lang, J. M. (2021). *Small Teaching: Everyday Lessons from the Science of Learning*. John Wiley & Sons. ISBN: 978-1-119-75554-8

Pollard, T., Peru, G., & Pontes Reis, E. (2022). *Introduction to artificial neural networks in Python (Carpentries Incubator)* (Version 0.1.0). https://github.com/carpentries-incubator/machine-learning-neural-python

*Scikit-learn course*. (2023). Inria. https://github.com/INRIA/scikit-learn-mooc

*The Carpentries Workbench*. (n.d.). Retrieved September 1, 2023, from https://carpentries.github.io/workbench/

Udemy - Basics of Deep Learning. (n.d.). In *Udemy*. Retrieved September 1, 2023, from

https://www.udemy.com/course/basics-of-deep-learning/

Udemy - Data Science: Intro To Deep Learning With Python. (n.d.). In *Udemy*. Retrieved September 1, 2023, from https://www.udemy.com/course/complete-deep-learning-course-with-python/

Udemy - Tensorflow 2.0 | Recurrent Neural Networks, LSTMs, GRUs. (n.d.). In *Udemy*. Retrieved September 1, 2023, from https://www.udemy.com/course/tensorflow-20-recurrent-neural-networks-lstms-grus/

Wilson, G. (2006). Software Carpentry: Getting Scientists to Write Better Code by Making Them More Productive. *Computing in Science & Engineering*, *8*(6), 66–69. https://doi.org/10.1109/MCSE.2006.122