Introduction to AI and its medical applications: Crash Course for an audience with diverse scientific backgrounds

Donatella Cea [1] **Helene Hoffmann** [2] **Marie Piraud** [1]

Abstract

Artificial intelligence (AI) and Machine Learning (ML) techniques have been developing more and more rapidly over the past few decades and teaching these methods can be very complicated even when students have good math and programming skills. Moreover, the background of the target group may be very diverse in terms of technical and coding skills, especially for a single introductory lecture. We describe here our experience in a three-hour crash course on introduction to AI and its medical applications, in which we alternated theoretical and practical sessions that could be engaging for students with different abilities and knowledge. The goals of the lecture were to demystify AI, introduce the challenges and current limitations of Deep Learning (DL), and present applications to the medical domain. The modularity of the course and choice of examples and tasks applied to the medical field interested the students who considered them authentic and relevant. The lesson has been positively evaluated by the students and their feedback identifies an NPS of 27 and an average of 8.2 over 10 when asked how likely they are to recommend the course to colleagues or friends.

1. Introduction

Online learning opportunities are increasing thanks to Massive Open Online Courses (MOOCs), video tutorials and plenty of other materials easily accessible on the web. Nonetheless, we believe that direct interaction with the experts and practitioners can still make a difference for a learner. But not everyone feels comfortable asking questions during or after the class in front of everyone. Moreover, sometimes doubts and questions occur at the moment of applying in practice what has been explained. For these reasons we decided to create an interactive lecture where three frontal lesson sessions were alternated with two tutorial sessions. Here we describe the material as well as our experience in teaching this once to a heterogeneous group of 45 students in a 3h crash course. During the tutorials the students worked in groups of three and two mentors were constantly walking around the class room to answer questions or give insights. This allows to create a relaxed atmosphere during which more students felt comfortable talking to the mentors in a smaller and more protected environment. Furthermore, group work has a positive effect on the students learning, enhancing motivation and social skills (Hassanien, 2006; 2007).

1.1. Target group

It is not always clear which audience to expect for a course. Often it can be very broad and with diverse backgrounds. Therefore, we designed the material to be suitable for students with various backgrounds, engaging students with a technical background as well as students with little or no prior knowledge in ML and coding. Not specific requirements or prior knowledge are needed. The course is an introduction to AI with focus on DL. In our first experience with the material we expected between 20 and 80 undergraduate and graduate students in medicine, medical computing, informatics, data science and different fields of engineering (45 students actually attended the lecture), hence a potentially extremely heterogeneous group. For this reason we believe that the material can be re-used for different audiences interested in an introductory lesson to AI with no or little previous background in ML and coding. In order to also involve learners with prior knowledge in this domain, we shared all the source files, such that they could have a deeper look at the models and experiment on their own. Therefore, we consider this material suitable for learners of different levels of skills and interests.

[1]Helmholtz AI, Helmholtz Zentrum München, Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany [2]Helmholtz AI, Helmholtz-Zentrum Dresden-Roßendorf, Bautzner Landstraße 400, 01328 Dresden, Germany. Correspondence to: Cea Donatella <donatella.cea@hemlholtz-muenchen.de>.

Proceedings of the 2^{nd} *Teaching in Machine Learning Workshop*, PMLR, 2022.

2. Lecture design

The goal of this crash course was not to enable learners to develop AI applications, but rather to prepare them as potential future users. The material is divided into five sections described below. We implemented it as a 3-hour lesson.

2.1. First frontal lesson: Demystifying AI

This is a first and generic introduction delivered in 30 minutes with the support of slides. During this introductory part, the speaker aims at demystifying AI, giving definitions, and explaining terminology, e.g. introduction to the neural network (NN) structure and accuracy metric. The speaker stresses that key historical elements in the advent of DL methods have been the introduction of new architectures (Convolutional NN and Recurrent NN), and the increasing availability of data and computing resources.

2.2. First tutorial session: Teachable Machine and TensorFlow Playground

The theoretical introduction is followed by a 30-minutes hands-on tutorial, where the learners are free to choose one of two prepared tutorials. In both cases they do not have to code but will work on classification tasks on the Google platforms Teachable Machine (Carney et al., 2020) or TensorFlow Playground (Smilkov et al., 2017). It is possible and recommended to do the tutorials in groups of three learners.

Teachable Machine tutorial. This tutorial is recommended to learners with little or no experience in ML and DL. The task is to train a classifier to distinguish between images of open hands and fingers. The learners create the data sets by themselves using their cameras or uploading images from the web. The architecture of the NN and the training parameters are predefined. The training of the networks happens in the background with no expert knowledge needed. The aim of the tutorial is to give the students an idea of the power of a deep NN and guide them towards the understanding that only appropriate image acquisition will guarantee that the network will properly learn to recognize new images.

Tensorflow Playground tutorial. This tutorial is recommended for learners who already have a basic understanding of a deep NN and want to explore in more details the structure and the importance of the hidden layers. The tasks are sequential and the learners are provided with the setup and starting point, then they have to answer the questions and insert their answer in the respective cells of a Jupyter Notebook (see section 3.1). After running the cells, they get a feedback with a small comment whether the answer is correct or a hint if the answer is incorrect. The main goal is to understand the concept of proper initialization and feature creation inside the network.

2.3. Second frontal lesson: Health care applications

After this hands-on excursion, the speaker introduces a few recent research papers with relevant medical applications. This session focuses on the evaluation of the performance of a DL algorithm, with respect to the performance of medical practitioners in detecting diseases from images. The speaker presents the content in 20 minutes with the support of slides. The goal is to convey that classification accuracy is not always suited to assess performance (especially in the medical field), to introduce other metrics and to give an overview of applications that can be useful in the clinical routine.

2.4. Second tutorial session

In this section, the learners all work on the same tutorial - if possible still in groups of three persons-, in order to foster interactions and peer discussion. The tutorial is divided into two parts both focused on image classification. This tutorial session lasts about 1 hour, with 20 minutes for the first part and 40 for the second one. We recommend a 10-minutes break before starting with the tutorial.

Binary classification with Teachable machine. For the first part, the learners download the data set of lung CT scans to their drive and upload the images to the Teachable Machine to detect Covid-19 (see description in section 3.2). This gives learners who did the TensorFlow Playground tutorial also a chance to explore this other tool. Those less familiar with DL will continue to use the platform they have previously explored. In this section the focus is on learning curves, metrics (accuracy) and the confusion matrix that the platform creates directly. The goal is to use these indicators to understand if the network is learning or not and discuss the problem of overfitting.

Multi-class classification: An example of coding. In this part, learners can take a direct look at the code behind the NNs they have used in the precious tasks. Some parts of the code, including data handling, network definition and training, are exposed to the learners in a notebook. The parts considered to be more advanced are collected in a separate module that the most interested can explore or modify after the course.
For this tutorial we used the MedNIST data set, which is a collection of medical images for classification tasks (see

description in section 3.2). At first, we used an imbalanced data set without explicitly tell the learners, who are guided to discover that a very decent overall accuracy can hide a low accuracy for some of the individual classes. They are then asked to investigate the possible reasons, and encouraged to explore the data set through an interactive data frame (to filter the data set without explicitly writing lines of code) or by simply looking at the confusion matrix. In case they need input, they can either ask an instructor or run a cell that will show them a hint. The learners then observe how the accuracy per class improves when the whole, balanced data set is used.

Eventually, the learners are proposed to use the trained network on an image non-related to the MedNIST data set (e.g. an image from the MNIST data set). This illustrates that a network always makes a prediction and is not able to recognize when an image is completely different.
The goal of this session is to touch upon some risks of DL such as overfitting and imbalanced data sets. It should demonstrate that NN are powerful tools but that the user should be aware of its limits and potential biases.

2.5. Third frontal session: Challenges and clinical solutions

In this last part of the course the speaker summarizes the main content of the tutorial and explains in more detail three of the main challenges in the application of DL methods namely overfitting, imbalance and biases in the data, and the black-box problem. Together with the challenges presentation, the speaker brings up existing solutions and describes famous cases in which these challenges lead to discriminatory behaviors. Finally, the speaker describes some success stories of DL applications that were certified for clinical application, to inform the students about the most innovative products and the state-of-the-art of AI in the medical domain. For this last section, the content is explained in 30 minutes with the support of slides.

3. Material

3.1. Coding environment

We chose to present all the tutorials in Jupyter Notebooks (Kluyver et al., 2016), as this is a popular framework suited for students with or without coding experience. For this one-time lecture, we wanted everyone to access the tutorials and computing resources easily. For this reason we suggested to open the notebooks in the Colab environment (`https://colab.research.google.com/`). With this solution, the students were able to run the code without the need of further setups on their own machines.

3.2. Data sets

The selected data sets meet criteria that we considered of fundamental importance for the course, such as free accessibility, small dimensions (to reduce download time and computing time) and relevance to the medical field.

- The CT lungs data set (Yang et al., 2020) is available on Kaggle and counts a total of 746 images divided as follows: 397 No Covid, 349 Covid. The images, i.e. 2D CT scans, are slices of the 3D images obtained through Computed Tomography, a medical imaging technique used in radiology to obtain non-invasive, detailed internal images of the body for diagnostic purposes. Only highly-trained experts are able to interpret the scans, such that it is difficult to detect the presence of Covid-19 on the scans, without the relevant medical background. The out-of-the-shelf NN algorithm used in the Teachable Machine, has an accuracy of 85 % in this task, which illustrates how DL can help technicians and doctors to diagnose this kind of disease.

- The MedNIST data set, was gathered from several sets from TCIA, the RSNA Bone Age Challenge, and the NIH Chest X-ray data set and is made available by Dr. Bradley J. Erickson M.D., Ph.D. (Department of Radiology, Mayo Clinic), under the Creative Commons CC BY-SA 4.0 license. The images belong to 6 different classes of radiological images (CT Head, CT Chest, CT Abdomen, CT Hand, Breast MRI, and Chest X-Ray). The original data set counts more than 50k images, which we reduced for the purposes of the tutorial. We kept 5% of the images in each class, for a total number of 2947 images. To show one of the effects of biased data, we created an imbalanced data set from the reduced data set: in three of the classes, we took only 20% of the images and 100% in the other three.

3.3. Availability

The presentation used to deliver the theoretical part is available on Zenodo at `https://doi.org/10.5281/zenodo.7053457`, and the Jupyter Notebooks can be found in the GitHub repository at `https://github.com/HelmholtzAI-Consultants-Munich/DL-lecture-tutorials`. All the materials are available under the Creative Commons Attribution License (CC-BY 4.0).

3.4. Student feedback

We sent the students a link to a form to collect their impression after the lecture. This included questions about the instructors, the atmosphere during the course, the structure and the learning experience. We got feedback from 11 students (24% of the audience). Although there were 1-2 votes against it most learners were satisfied with the instructors and the feedback was quite positive: the mix of theoretical and practical parts was appreciated and the examples and tasks were considered relevant and authentic. The question 'How likely are you to recommend this workshop to a friend or colleague' (on a 1-10 scale), got a mean rate of 8.2, with 64% of the respondents who gave more than 8, as shown in Figure 1. We also computed the Net Promoter Score (NPS) (Reichheld, 2003) by subtracting the percentage of detractors (students who answered the question with a 6 or lower) from the percentage of promoters (students who answered with a 9 or a 10). For this first experience we obtain a NPS of 27, which is considered as a good score.

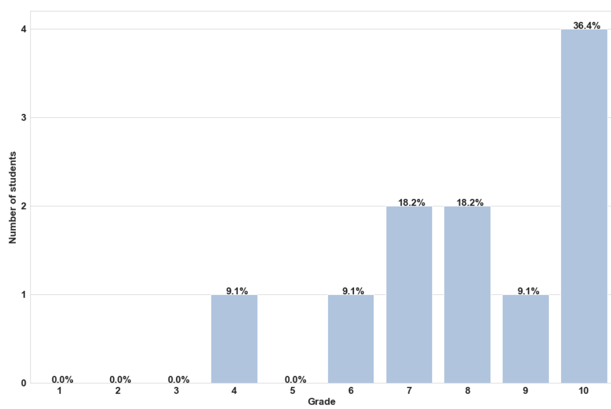

Figure 1. Answers to the question 'How likely are you to recommend this workshop to a friend or colleague' (mean value of 8.2 and NPS of 27).

4. Discussion

We are satisfied with the first teaching experience of this lecture structure. The students were interested and active during both the frontal lessons and the tutorials. The atmosphere was relaxed and open discussions came up spontaneously. Notwithstanding the positive general impression, we see room for improvement.

- Some students were reluctant to work in groups. Since peer interaction is crucial for the learning process, we recommend taking some time at the beginning of the lecture for an ice-breaker activity, if the students do not know each other. One option might be to dedicate 2-3 minutes to a personal presentation of the student to the closest person and let them talk about why they decided to participate in the lecture.
- We had 2 mentors for more than 40 students. As we had the feeling that we could not respond to all questions in enough detail, we would recommend to have 1 or 2 more mentors to give more feedback during the tutorials, especially in the case where learners work individually.
- We did not include information on the training process of a NN in the lecture, as a chosen position to avoid technical details but focus on awareness on the capabilities and limits of DL. It was therefore difficult for the utmost beginners to grasp concepts such as overfitting. This information could either be added in the first frontal lesson, if time allows or the tutorials could be modified (e.g. see point below).
- The introduction of several tools and data sets might be overwhelming for an three-hours lecture. The binary classification task on the Lung data set for Covid detection could be removed from the second tutorial and used in the first tutorial on the Teachable Machine. In this case, the focus would be on understanding the learning process, and the questions on overfitting can be kept as a bonus. In a longer format, the tutorials could be kept as such.
- As only 24% of the students answered the feedback-survey, one could dedicate a couple of minutes at the end of the lecture to answering it.

5. Conclusion

AI, ML and DL are becoming important tools for data analysis in many different fields. The presence of these methods and algorithms in our everyday life is growing and it is crucial that students from non-related fields do not feel overwhelmed by the amount of information or discouraged by the technical aspects. We propose an inclusive crash course for the introduction to DL and its medical applications. The goal is to show the power of these tools but also make them aware of the pitfalls and challenges that they could face in the application of these methods. The lecture is structured in a way that promotes the interaction between the students and with the instructors. The modularity of the lecture allows us to handle crowds of diverse, and even unknown, backgrounds, and to adapt it to different formats. Overall we are satisfied with the first experience with the material. We outlined different paths for improvement; our goal being to increase the NPS in the next iterations and make the students even more comfortable in this learning environment.

References

Carney, M., Webster, B., Alvarado, I., Phillips, K., Howell, N., Griffith, J., Jongejan, J., Pitaru, A., and Chen, A.

Teachable machine: Approachable web-based tool for exploring machine learning classification. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, apr 2020. doi: 10.1145/3334480.3382839.

Hassanien, A. Student experience of group work and group assessment in higher education. *Journal of Teaching in Travel & Tourism*, 6(1):17–39, jul 2006. doi: 10.1300/j172v06n01_02.

Hassanien, A. A qualitative student evaluation of group learning in higher education. *Higher Education in Europe*, 32(2-3):135–150, jul 2007. doi: 10.1080/03797720701840633.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., and Team, J. D. *"Jupyter Notebooks - a publishing format for reproducible computational workflows."*. ELUB, 2016.

Reichheld, F. F. The one number you need to grow. *Harvard business review*, 81:46–54, 124, December 2003. ISSN 0017-8012.

Smilkov, D., Carter, S., Sculley, D., Viégas, F. B., and Wattenberg, M. Direct-manipulation visualization of deep networks. August 2017.

Yang, X., He, X., Zhao, J., Zhang, Y., Zhang, S., and Xie, P. Covid-ct-dataset: A ct scan dataset about covid-19. March 2020.