Introducing Machine Learning to First-year Undergraduate Engineering Students Through an Authentic and Active Learning Labware

Lei Huang, Kuo-Sheng Ma
Department of Eletrical Engineering and Computer Science
Loyola Marymount University
Los Angeles, USA
lei.huang@lmu.edu, kuo-sheng.ma@lmu.edu

***Abstract*— This innovative practice work in progress paper presents our pilot experiment of teaching machine learning to first-year multidisciplinary undergraduate engineering students using an authentic and active learning tool, which consists of a public Google site repository and a course project. The preliminary evaluation we conducted by pre-post survey shows user experiences and effectiveness of this teaching/learning tool. Analysis of the evaluation results will be used to guide us for further development of the tool.**

Keywords—machine learning, active learning, authentic learning, labware

I. Introduction and Motivation

Recently, machine learning has attracted phenomenon attention and found many applications in every sector of our society, from data mining to automation, entertainment to the medical field, business to engineering [1-5]. Machine learning technology is enabling a paradigm shift in problem-solving from analytical to powerful data-driven approach, through computer programs that learn models from training data and predict results from new data. It is quickly becoming the essential modern engineering tool driven by the explosively increasing amount of data and high performance computing power available in various industries. Generally speaking, machine learning The rapid growth of machine learning technology and application resulted in a dearth of qualified engineers, who are able to develop and optimize machine learning tools to solve real-world problems. It has been well recognized that higher education in this field provides multitude opportunities to engage student interests, broaden student visions, increase student global competitiveness for all STEM students.

Universities have been putting efforts into developing and promoting relevant curriculum [6,7]. In the newest revision of ACM/IEEE-Computer Society's curriculum guidelines (CS2013), a much greater emphasis has been placed on machine learning than in the past [6]. However, most of these efforts are made either within the computer science/computer engineering programs or into an emerging new data science program, which is usually a joint force between computer science and mathematics. There is little effort being made in other traditional engineering disciplines such as mechanical engineering, civil engineering, and even electrical engineering, especially at the undergraduate level.

Although it is possible for undergraduate engineering students to take machine learning courses offered in computer science or data science program as electives, it is very challenging for them due to several reasons. First, machine learning is built upon a variety of theoretical background, including advanced math, linear algebra, statistics, and probabilities. Engineering students usually do not have all the prerequisites until toward the end of their academic program. In fact, machine learning courses are mostly taken by graduate engineering students. Second, it requires high computing and programming proficiency, which is usually lacking in the traditional engineering disciplines. Third, it is very hard to squeeze additional courses from other disciplines into their already fully packed curriculum. As a result, it is usually perceived by students as a very difficult subject and remains a foreign topic for those who do not take it as elective. For those who do, they usually will not have enough training and experience required by the industry workforce upon their graduation.

In this project, the authors systematically design, implement and evaluate innovative ways to integrate machine learning into multidisciplinary undergraduate engineering curriculum, using modern cyber-learning tools and resources, through effective engineering pedagogical approaches. As an initial step, we are introducing machine learning to first-year multidisciplinary engineering students taking a required algorithm and application course through an active and authentic learning labware. The main learning objective we want to achieve through this labware is to recognize the importance of machine learning as a data-driven approach to solving real-world problems, and consequently, attract students' interests in learning this important tool, change students' perception of its difficulty, motivate students to build up background and skills needed, and encourage them to dive deeper into this rapidly developing field throughout their undergraduate engineering program.

The following of the paper is organized as follows. Related education research that inspired our work will be briefly reviewed in Section 2. We will describe the details of our designed learning labware in Section 3. Section 4 presents our evaluation method and preliminary evaluation results. Section 5 concludes the current work and outlines our future plan.

II. RELATED WORK

In order to teach advanced machine learning techniques to freshman engineering students, it is crucial to employ effective pedagogy. In [8], the authors summarized different learning and teaching styles and pointed out that most engineering students are visual, sensing, inductive, and active learners. Therefore, active learning pedagogy is more effective than traditional lecture-based passive pedagogy. Other researchers have also demonstrated benefits of active learning in STEM educations [9,10]. Moreover, educational research has demonstrated that authentic learning, which involves real-world problems, open-ended thinking, collaborative learning, and self-directed learning, is stimulating and beneficial to all types of learners in engineering education [11,12].

Inspired by these research results, we combined active learning and authentic learning into an integrative learning tool to introduce machine learning to freshman engineering students. We call it a labware since it is analogous to the engineering instructional laboratory, which has been the most critical active learning component in engineering education [13]. However, it does not have the physical location and time slot limitations as traditional lab activities. It is a web-supported and mobile-enabled cyber-learning tool, which allows students to learn anywhere, anytime, inside and outside the classroom. This labware developed in our project will be made open source and open access to the public, such that it can be shared among the engineering education community and easily adopted by other institutions. We would also like to invite contributions from other institutions to improve the tool to foster learning process and improve the learning experience.

III. AUTHENTIC AND ACTIVE LABWARE DESIGN

The prototype of our active and authentic learning labware has two components: A public Google Site repository, which is powered by Google Cloud technology and platform, and a final project associated with this website in a MATLAB based programming course. This course is required for all first-year undergraduate engineering students. A small portion of students from other disciplines, such as sciences and business, also take the class to learn practical programming. Thus, it provides a multidisciplinary learning environment that facilitates authentic learning through real-world applications in different disciplines.

Fig. 1 shows a screenshot of one page of the Google Site and its three main sections, including fundamental concepts of machine learning, modules of different machine learning algorithms, and real-world applications.The first section serves as preliminary material, where the big picture of machine learning is painted through a question-and-answer format. Its main purpose is to inspire students' interest and motivation to learn more about machine learning. The second section serves as instructional lab activities, with basic theories, including mathematical tools and algorithm development, followed by step-by-step instructions and hands-on activities with some simple examples. The third section provides real-world applications with real-world data. Students go through the entire process of an open-ended design laboratory experiment, during which a real-world problem is formulated, the initial solution is designed, tested, and evaluated, followed by improvement of design if necessary.

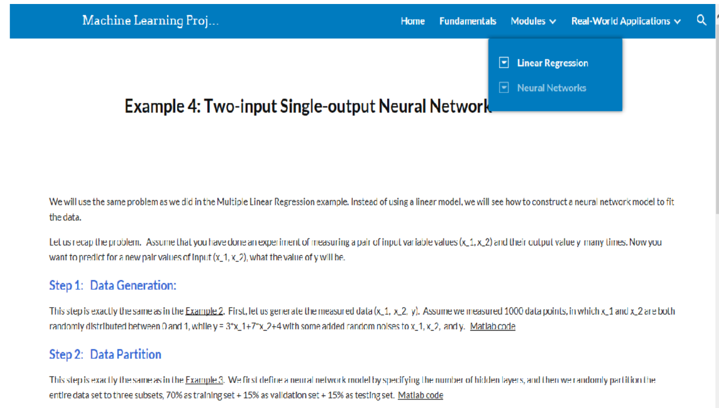

Fig. 1 Screenshot of the Google Site of the Labware Repository

Currently, we have implemented two modules in the second section: a linear regression module and a neural network module. In each module, we first introduced the basic theory, followed by two simple examples, which demonstrated the step-by-step development of a machine learning solution. Sample MATLAB code for each step is provided in a separate page.

In the real-world application section, there is currently only one problem provided. It is a housing price prediction problem. A real estate dataset consisting of around 10,000 houses in the current Southern California housing market is provided in a spreadsheet. Each house has feature data, such as square feet, lot size, year built, number of bedrooms, number of bathrooms, zip code, and its selling price.

In the final project, the students are required to read and learn from the Google Site, go through the modules and examples, and to build a housing price prediction model using machine learning approach. Students need to use their knowledge and techniques learned from the website to partition the data, choose a model, train the model using training data, and evaluate the trained model with validation/testing data. It is an open-ended problem, different models, different parameter settings, and different data processing methods can be explored when solving the problem.

When solving this problem, authentic learning is involved during the data preprocessing stage. Compared to the examples, where data are generated from a certain mathematical model with randomization, this real-world problem has real-world data that is noisy and biased with large variations. Therefore, it needs to be cleaned up and normalized before used as input to the model. Otherwise, the training may not converge.

Students work in groups in the project to promote collaborative learning. Each group has 2 or 3 students. Group discussions, group activities, progress, and each group member's responsibility are required to be documented in a group discussion page. The machine learning topics are not covered in class. Students are self-learning all the materials, modules, and examples at their own pace outside of the classroom, and are given four weeks to finish the project.

IV. PRELIMINARY EVALUATION

The prototype of the labware has been tested by three sections of students enrolled in the introduction to algorithms and MATLAB programming course. Most of them are first-year engineering students. Few of them had experiences in MATLAB programming before taking the course.

A. *Evaluation method*

We evaluate the effectiveness of this labware in achieving the learning objectives as mentioned in Section 1 using a pre-post survey approach. The pre-survey was taken by the students right before the labware was published to the students. The post-survey is taken by the same group of students upon their completion of the project associated with the labware. We use the same sets of questionnaires in both surveys in order to measure the difference caused by the labware intervention. The evaluation is formative in the sense that the results are used to enhance the tool in the future.

A copy of the survey is shown in Fig. 2. There are 11 questions in total. Except for the first and last short-answer questions, all other questions require students to choose a rating between 5 and 1, where 5 means strongly agree, and 1 means strongly disagree. The last question is designed to evaluate the user experience of learning from the labware, and will be used to improve the labware in the future.

#	Questions	Strongly Agree Strongly Disagree 5-----4-----3-----2-----1
1	I plan to major in: (please write your intended major):	
2	I think Machine learning is very useful in my discipline.	
3	I know something about how to apply machine learning in my discipline.	
4	I am interested in learning more about machine learning topics.	
5	I think that machine learning is very difficult to learn.	
6	I learn better when difficult materials are extensively taught in lectures.	
7	I learn better from online materials at my own pace, with access to guidance and help.	
8	I learn better from working on real-world problems.	
9	I am good at math and data analysis.	
10	I like to take academic challenges.	
11	I have the following comments/suggestions	

Fig. 2 The Survey (both Pre-Project and Post-Project)

B. *Evaluation result*

Our pre-survey showed some interesting results. For example, around 70% of students showed strong interests in learning machine learning. However, only one-third of students had prior knowledge of machine learning applications in their discipline, and another one-third of students did not have any idea. Comparing students from different engineering disciplines, we found that there is a higher percentage of students in civil engineering than other engineering majors who have less interest and less knowledge in machine learning. Across all disciplines, most (about 90%) students prefer authentic learning through solving real-world problems. Moreover, most (about 90%) students like to take academic challenges.

Through further analysis of the data, we discovered that there is a strong correlation between the student's interests in learning machine learning and their knowledge of machine learning applications in their own discipline. Therefore, we have assigned a new task in the final project, requiring students to research on the recent development of machine learning applications in their own discipline. This helps provide us with ideas in developing more modules and applications to improve the labware.

The statistical results of the rating scores for each of the rating questions are compared between pre-survey and post-survey in Table 1.

Table 1 Statistical Results of Pre and Post-Project Surveys

Quest-ion #	Mean			Standard Deviation		
	Pre-project	Post-project	Change	Pre-project	Post-project	Change
2	3.83	4.33	0.50	1.09	0.91	-0.18
3	2.92	3.83	0.91	1.20	0.99	-0.21
4	3.82	4.06	0.24	1.18	1.00	-0.18
5	3.79	4.11	0.32	0.95	0.76	-0.19
6	3.72	4.11	0.39	1.04	1.18	0.14
7	3.07	3.44	0.37	1.19	1.20	0.01
8	4.13	4.28	0.15	0.92	0.89	-0.03
9	3.96	3.95	-0.01	0.88	1.05	0.17
10	4.17	4.23	0.06	0.80	0.73	-0.07

From Table 1, we can see that for most of the questions, the mean rating has increased after the project, except for Question 9, which asks students to rate their own proficiency in math and data analysis. The most significant increase of the mean rating is for Question 3, indicating that students think they have gained some knowledge on how to apply machine learning in their discipline. The second most significant increase of the mean rating is for Question 2, indicating that students have better recognized the importance and usefulness of machine learning in their discipline. From the results of Question 5, students think that machine learning is more difficult after the project than they thought before the project. Nonetheless, they are more interested in learning machine learning (Question 4) and more willing to take academic challenges (Question 10). In terms of learning methods (Questions 6, 7, and 8), students prefer learning from real-world problems consistently (pre-project and post-project). After the project, they appreciate lectures on difficult materials more than before the project, while on the other hand, they are embracing self-learning from online materials more than before the project. The slight decrease in students' self-assessment of their own proficiency in math and data analysis as shown in the results of Question 9 might be associated with the increased difficulty of machine learning students have perceived after the hands-on project. However, the increase in its standard deviation indicates possibly higher uncertainty of the mean result. Question 6 has a similar increase in standard deviation, thus higher uncertainty in the mean. For all other questions, we observe decrease or insignificant increase in the standard deviation, indicating higher confidence in the statistical results.

From the above statistical analysis, we can assess the learning objectives as mentioned in Section I. Through the learning labware based project, students have better recognized the importance of machine learning as a data-driven approach in solving real-world problems. It has increased students' interests in learning this important tool, changed students' perception of difficulty, and motivated students to take more academic challenges in math and data analysis. The most surprising results we found is the increase in students perception of difficulty because we had expected a decrease instead. However, this increased difficulty perception has not prevented students from being more interested and motivated to learn.

Most feedbacks from students (Question 11) were positive and encouraging. Many students commented that they have realized the importance and usefulness of machine learning and more general programming and coding skills in their own discipline. "This realization has opened a world of possibility in real-world applications". Students view the introduction to machine learning as "opening the door to the intricacies of the modern technology". Some expressed their plans for advancing their study in this field in the future. Most students commented that the set up of this labware makes learning more effective than traditional lectures and homework assignments.

From students' feedback, we also found some limitations of this learning tool. Many students commented that they had difficulties in understanding difficult part of the material, especially the materials about neural networks. Since the Neural Netowrk toolbox in MATLAB is based on object-oriented programming (OOP) concept, which was not covered in the traditional offering of this programming course, students are having hard time to understand the net class and its attributes and methods, even though the link to the documentation resources from Mathworks are provided. We plan to cover OOP concept in lectures in the future offering of the course. Some students suggested adding more real-world data and problems from a scientific field to solve.

V. CONCLUSION AND FUTURE WORK

A prototype of authentic and active learning labware has been designed and pilot implementation has been conducted to expose machine learning to students in a first-year engineering algorithm and programming course. We are continuing to design more modules and real-world problems to improve the labware based on students survey results and feedbacks. Upon completion of the project, the labware will cover both basic and advanced machine learning topics, including some new modules in deep learning. Therefore, it can be used as an integrated and sequential lab material for multiple classes across multiple engineering curriculum. We highly expect to increase undergraduate engineering students' proficiency in modern engineering tools and improve their learning experiences.

REFERENCES

[1] I. H. Witten, et al. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.

[2] R. Carbonneau, K. Laframboise, and R. Vahidov. "Application of machine learning techniques for supply chain demand forecasting." *European Journal of Operational Research* 184.3, 2008, pp.1140-1154.

[3] T. Mueller, A. G. Kusne, and R. Ramprasad. "Machine learning in materials science: Recent progress and emerging applications." *Reviews in Computational Chemistry* 29, 2016, pp.186-273.

[4] A. Lavecchia, "Machine-learning approaches in drug discovery: methods and applications." *Drug discovery today*, 2015, pp.318-331.

[5] B. Narendra, et al. "Sentiment analysis on movie reviews: a comparative study of machine learning algorithms and open source technologies." *International Journal of Intelligent Systems and Applications* 8.8, 2016.

[6] ACM/IEEE-CS Joint Task Force on Computing Curricula, Computer Science Curricula 2013, ACM Press and IEEE Computer Society Press,

[7] P. Anderson, J. Bowring, R. McCauley, G. Pothering, and C. Starr. "An undergraduate degree in data science: curriculum and a decade of implementation experience", *Proceedings of the 45th ACM technical symposium on computer science education (SIGCSE '14).*, pp.145-150. ACM, New York, NY, USA, 2014.

[8] R. M. Felder, and L. K. Silverman. "Learning and teaching styles in engineering education." *Engineering education* 78.7 (1988): 674-681.

[9] S. Freeman, et.al., "Active learning boosts performance in STEM courses", *Proceedings of the National Academy of Sciences, 2014,* 111 (23) pp. 8410-8415.

[10] V. M. Cardullo, N. S. Wilson, and V. I. Zygouris-Coe. "Enhanced student engagement through active learning and emerging technologies." *Student Engagement and Participation: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*, 2017, pp. 399.

[11] J. Herrington, *Authentic learning environments in higher education*. IGI Global, 2006.

[12] A. Khanlari, "Robotics integration to create an authentic learning environment in engineering education." *Proc. IEEE Frontiers in Education Conference (FIE),* 2016.

[13] L. D. Feisel and A. J. Rosa. "The role of the laboratory in undergraduate engineering education." *Journal of Engineering Education* 94.1 ,2005, pp. 121-130.