A Metalearning Approach to Personalized Automatic Assessment of Rectilinear Sketches

Laura Melissa Cruz Castro
School of Engineering Education
Purdue University
West Lafayette, IN, USA
lcruzcas@purdue.edu

Samantha Ray
Department of Computer Science & Engineering
Texas A&M University
College Station, TX, USA
sjr45@tamu.edu

Hillary E. Merzdorf
School of Engineering Education
Purdue University
West Lafayette, IN, USA
hmerzdor@purdue.edu

Kerrie A. Douglas
School of Engineering Education
Purdue University
West Lafayette, IN, USA
douglask@purdue.edu

Tracy Hammond
Department of Computer Science & Engineering
Texas A&M University
College Station, TX, USA
hammond@tamu.edu

***Abstract*—Sketchtivity is a stylus-based intelligent tutoring system that can help instructors automatically provide feedback to their students, saving them the time and effort of providing personalized feedback themselves. The system uses a generic evaluation of perspective, direction, and accuracy to give students feedback on the quality of their sketches. If instructors want to personalize the metrics, the system would require them to provide multiple sets of samples. Therefore, instructors may use instructional team members such as teaching and graduate teaching assistants to provide feedback on the required samples. Compared to that of assistants, the feedback they produce might vary due to expertise and create noise in the training data. To address this problem, we implement a deep neural network that leverages learning to reweight algorithms. The data collected by the instructor from undergraduate and graduate-level rectilinear perspectives sketching is considered the validated sample. In this study, we analyzed the training size requirement for a Multi-Layer Perceptron (MLP) to accurately predict whether or not a stroke was a perspective stroke. We observed that the training data required to predict stroke accuracy is small. In addition, the performance of the algorithm in terms of accuracy was good even under extreme conditions such as having highly unbalanced data and having a small valid set of data. The results from the study support the use of these types of algorithms for future system personalizing to support scalable feedback systems in education.**

I. Introduction

Sketching and learning in disciplines such as design, engineering, and creative professions are interconnected. Sketching is a vital part of the design process and an essential skill for team communication in design teams [1]. Before technological tools were widely available, the design was made mainly using paper and pencil, but with the endless growth of computer software and new technologies, more software was available for design. Therefore, in higher education, the use of design software has replaced in many spaces traditional sketching. Instructors have moved instruction and assessment from a paper-pencil space to a computer-aided design (CAD) [2].

Although the use of CAD brings multiple advantages to the design process, such as accuracy, quality of the prototypes, and efficiency [3], its use in education is controversial. Remarkably, some studies have argued that CAD systems in the early stages of the design process could potentially affect design skills acquisition for novice engineers. For example, in a study performed by [4], the authors found that the use of CAD in the design process for novice engineering designers affected design efficiency and effectiveness and premature design fixation, compared to engineers who designed using freehand sketching.

Multiple researchers addressed the need for intelligent systems that involve similar mechanics as paper as a pencil but with the advantages of using technology. From the need for a system in the middle ground between paper and pencil and CAD, sketch-based intelligent tutoring systems emerged. One of the most prominent systems in the literature is Sketchtivity. Sketchtivity is a web-based system design to provide feedback on students' practice of design sketching fundamentals [5]. Sketchtivity has proven to be a valuable tool in sketch education. The system has shown positive effects on students sketching ability and spatial visualization when studied in a pre-post scenario [6]. In addition to the potential of this tool on the increment of both accuracy and speed of the final sketches [7], Sketchtivity receives constant improvement in different aspects such as user interface, recognition systems, and training algorithms, among others.

Personalizing is one aspect that will be considered an improvement in the following milestones for the system. Personalization is vital for any intelligence system that wants to be incorporated into education. In 2020 a research group dedicated to evaluating the space that Artificial Intelligence (AI) system may have in education concluded that establishing collaborative relations between humans and machines is an essential step for AI educational applications [8]. In terms of automatic assessment, the confidence of instructors in automatic assessment systems may increase when they are involved in the process. Therefore, there is a need to connect instructors to the assessment of different abilities offered by Sketchtivity and other intelligent systems. The ability of an instructor to personalize the assessment system can help them trust the tool and be more connected with the assessment process. In addition, it gives instructors the ability to control the features assessed for different levels of expertise and learning.

Although personalization is a vital improvement for intelligent systems, it does come with a cost. For instance, one-way personalization can be envisioned for the system is by giving instructors and stakeholders the complete ability to personalize the assessment by allowing them to provide a sample of assessed sketches. In that way, their criteria would be the ones used to train the system. Although this approach brings flexibility and allows a considerable level of personalization, the quality, quantity, and balance of the data could be affected by the training models. It would cause a more complex training process, starting with the algorithms used in production.

The training algorithms used for the production of assessment in Sketchtivity have varied in complexity [9]. In some cases, the selection of features collected by the user interface, the accuracy produced by the hardware, and the quality of data classification have reduced the complexity of the training models for some evaluated features. For instance, features such as perspective accuracy achieved good levels of precision and recall when using few features and a simple random forest algorithm [9].

The advantages of giving complete control on the training to stakeholders may outweigh the move from simple to more complex training. More complex modeling algorithms called meta-learning algorithms have the potential to generate knowledge about learning and therefore present a potential to predict with fewer valid observations. The learning to reweight algorithm proposed by Ren et al. [10] is one of the algorithms of this branch. Ren's method has proven effective in facial recognition [11], and it is considered an essential tool in the Meta-learning Branch of AI [12].

In this research paper, we use an algorithm of the meta-learning branch [13] of AI to explore the possibility of incorporating personalization in the assessment of sketches, particularly in recognition of rectilinear perspective accuracy recognition. This research paper, therefore, explores the following three questions:

1) Using data assessed by an expert, what would be an approximate number of sample sketches required for stakeholders to personalize Sketchtivity's rectilinear perspective accuracy classification?
2) How does noise in the classification classes affect the sample size requirement?
3) In case of unbalanced and noise data with few valid training samples, would the learning to reweight algorithm proposed by Ren et al. [10] have the potential to account for the noise and perform with good accuracy with a small valid sample size?

Our hypotheses are: (1) When the classification is performed using a valid set of samples, the sample size requirements provided by stakeholders can be very low, of the order of 50 samples. (2) Nevertheless, if some of those samples contain noise or the classes are unbalanced, the sample required might be more extensive. (3) The learning to reweight algorithm can increase the accuracy of the model under noise conditions, lowering the number of valid samples required.

II. Literature Review

This literature review will be divided into two sections. The first section will focus on the literature on learning and feedback, further focusing on sketching. The second section will focus on the literature on deep learning and meta-learning.

A. Learning, sketching and Feedback

Sketching is arguably one of the keystones of the engineering disciplines. It helps engineers consolidate ideas and communicate with others. Sketching is rooted in engineering identity, being noticed as the way engineers think and work [14]. Sketching has been integrated into engineering education, taking part in engineering training and designing courses. Its predominance in engineering education has highlighted its importance.

As with many other skills, feedback is essential in mastering the ability to sketch. Feedback is conceived as the notion of the dialogue that exists among different actors to support learning in formal and informal situations [15]. The relevance of feedback in learning draws from the complex relationship between teaching and learning; for instance, the co-constructive model of learning describes learning as a complex process in which social, emotional, and cognitive dimensions affect students learning [16]. Feedback takes a role in the model; as the communication between instructor and learner and is conceived as an additional opportunity for the students to learn, adding to the social dimension of learning [17].

Even more important than feedback, timely feedback is one of the most relevant resources in learning [17]. However, increasing classroom sizes makes giving individual feedback a problematic task, with time constrain affecting both quality and timely feedback. Nevertheless, automatic feedback, namely intelligence systems programmed to give automatic feedback to learners, have demonstrated to be a technological tool with the potential to help students master abilities [18] such as sketching [9]. The purpose of Sketchtivity is to give students opportunities to learn sketching skills while receiving continuous and timely feedback [5].

Students are trained to sketch using perspective and common exercises in beginner courses, including rectilinear perspective, providing the foundation for students to learn to draw simple sketches using point perspective. However, one of the most challenging tasks for novice learners is the sketching of strokes with good perspective technique [9]. Therefore, accuracy in rectilinear perspective is one of the features that Sketchtivity assesses. It helps students recognize whether a linear stroke has accurately drawn on perspective or not. For instance, this skill can be used when practicing sketches of streets and corners (see 2). The immediate feedback given by Sketchtivity has already proven been beneficial for novice learners [6], [9]; nevertheless, its adoption requires instructors confidence in the automatic feedback system.

B. Deep Learning and Meta Learning

Machine learning is a field of artificial intelligence that aims to give machines the ability to make decisions and predictions without a specific programming routine designed for it [19]. Machine learning has been conceived as a way to replicate human thinking in an artificially intelligent system. In particular, deep learning is an area of machine learning compromised with the development of neural networks that contain hidden layers using linear and non-linear models and mainly used in the prediction of classification outcomes [20]. Finally, meta-learning is a branch of machine learning concerned with the ability of a machine to "learn how to learn.' The meta-learning branch of machine learning is in charge of creating algorithms that deal with tasks requiring the machine to learn multiple times and teach "itself" a task [21].

Some of the algorithms created in the meta-learning branch intend to solve problems that arise from work with real data. This is the case of the Learning to Reweight algorithm proposed by Ren et al. [10]. This algorithm intends to propose a solution to scenarios in which the training data contains noise, or the imbalance of the training data creates difficulties from deep learning algorithms.

The reasoning behind this algorithm is that if one has a small training sample that is clean, then a machine can learn from that small sample. It uses multiple runs to create weights, helping gather information from the rest of the training samples while remembering that there is a sub-sample of the data that is more accurate. This algorithm has been used in multiple fields such as image recognition [22], medicine [23] and biology [24], but to our knowledge, the current study is the first in which this algorithm is applied in an education context.

III. Methods

A. Data

The data for this study corresponds to the data collected from a convenience sampling of 40 students from undergraduate and graduate programs and from various majors, including engineering. The students were asked to sketch a rectilinear perspective sketch (see Fig.1). From the observation collected, 1210 strokes were manually coded by a design expert. Although multiple classifications were given to each stroke, this research paper focuses on classifying accurate rectilinear perspective strokes. The data was collected with Sketchtivity and the features used correspond to 25 features (see Table I) that have been highlighted as significant predictors for the rectilinear perspective [9].

From the 1210 observations obtained, 55% were classified as perspective strokes while 45% were classified as non-perspective strokes. Based on mutual information, the ten most important features for distinguishing perspective versus non-perspective are cosine of the initial angle (F1), length of bounding box diagonal (F3), angle of bounding box diagonal (F4), the distance between endpoints (F5), cosine of the angle between endpoints (F6), total stroke length (F8), total angle traversed (F9), the sum of absolute values of angles (F11), and line similarity ratio (F12).

TABLE I
FEATURES DEFINITION, ADAPTED FROM [9]

Feature	Definition
F1	Cosine of Initial Angle
F2	Sine of Initial Angle
F3	Length of Bounding Box Diagonal
F4	Angle of Bounding Box Diagonal
F5	Distance Between First and Last Point
F6	Cosine of Angle Between Endpoints
F7	Sine of Angle Between Endpoints
F8	Total Stroke Length
F9	Total Angle Traversed
F10	Sum of Absolute Values of Angles
F11	Sum of Squared Values of Angles
F12	Line Similarity Ratio
F13	# of Sub-strokes Siblings
F14	Above Horizon?
F15	Cosine of VP1
F16	Cosine of VP2
F17	Cosine to Vertical
F18	Cosine to Horizontal
F19	Sine to VP1
F20	Sine to VP2
F21	Sine to Vertical
F22	Sine to Horizontal
F23	Degrees to VP1
F24	Degrees to VP2
F25	Degrees to Vertical

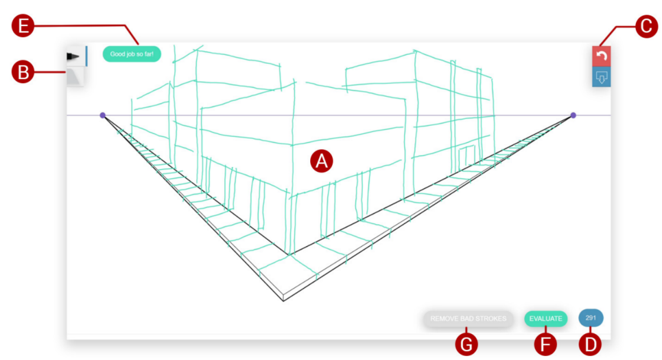

Fig. 1. Example of a sketching canvas with two vanishing points and a sidewalk [9]

B. Data sub-samples

In order to answer the first research question, the data is divided into two sets: train and test. All features are used, and the response variable is the binary classification of rectilinear perspective. For the second research question, an additional sub-sample is split from the training set called a meta sample. This sample will be a set of observations taken from the train sub-sample of research question 1, and that is left untouched by the addition of noise (see Fig.2).

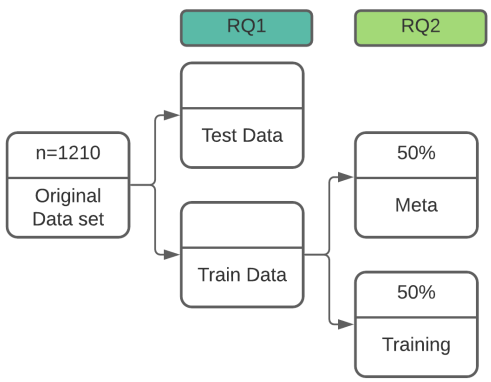

Fig. 2. Data sub-sampling to train and test the accuracy and recall

C. Noise

The training dataset corresponds to 50% of the total train data. This training dataset is modified to account for the noise that instructional teams might introduce in labeling the perspective strokes. For instance, one scenario that we can see as relevant is that of an assistant classifying more strokes as correctly rectilinear perspective when the strokes do not comply with all requirements. Therefore, the noise is created randomly, redirecting actual labels from 0 (non-perspective) to 1 (perspective) in 1%, 5%, 20%, 30%, 40%, and 50% of the cases assuming different levels of error. This noise would be testing the assumption of instructional teams being less accurate in classifying strokes that are not perspective.

D. RQ1 and RQ2: Approximate number of sketches required

To approximate the number of sketches, a multi-layer perceptron (MLP) will be used. An MLP is a series of activation functions stacked on top of each other to predict a classification, or linear outcome [25]. The choice of an MLP comes from three reasons (1) First, since the data used for this project is tabular with only 25 features, a standard feed-forward neural network fits the data (2) Second, a MLP is a universal approximator, meaning that it can model any suitable smooth function and produce any level of accuracy [25]. (3) Third, the data collected contains strokes from the same user produced in sequence. Nevertheless, the MLP processes each stroke independently without any time context, following the non-dependant nature of the data.

In this work, the model is implemented using PyTorch [26]. The MLP architecture has two hidden layers of sizes 20 and 10 and uses a rectified linear function (RELU) as the activation function. This architecture was determined empirically by testing between 1 and 3 layers and the hypertangent, sigmoid, and RELU activation functions. Following standard practice to create as generalizable a model as possible, each hidden layer decreases in size as data passes through the model. The MLP is trained using stochastic gradient descent with backpropagation and a momentum of 0.8. The loss function used is binary cross-entropy with logits. The training uses a batch size of 128, a learning rate of 0.001, and 100 epochs.

For both modeling scenarios, different training set sizes are tested (i.e., training on 80%, 50%, 20%, 10%, 5%, and 2% of the data). Each training configuration is run five times to get average performance. A summary table with the different results in terms of accuracy will be reported. The results will then be descriptively analyzed.

E. RQ2: Learning to reweight algorithm

Using the information gathered in the first research question, the experiment is repeated with the train/test split size held constant and the amount of noise in the training data varying from 0% to 50% as described earlier. Noise is added by dividing the training set into a training sub-sample and a meta sub-sample where the training sub-sample randomly has the noise percentage amount of the non-perspective labels flipped to perspective. This simulates the scenario in which the instructor provides some of the samples while less expert instructional team members provide others with possible errors. Finally, we report the accuracy results for the network without the additional meta-algorithm and then applied the algorithm reporting equivalent results.

IV. Results

In this study, we analyzed the training size requirement for an MLP to accurately predict whether or not a stroke was a perspective stroke. We replicated the technique proposed by Ren et al. [10] to reweigh the labels that might have noise in the data due to the diversity of graders. From the first part of the study, we concluded that the training data required to predict the accuracy of a rectilinear stroke is small. For instance, when looking at Table II it is possible to observe that the accuracy of the MLP stays relatively high even with a small training set. For example, from these experiments, an accuracy of nearly 80% is reached with a test proportion of 90%. In this case, the model is notably accurate while only having 10% of the observations to train.

From this observation, it was possible to conclude that a small valid training size might be sufficient to train the algorithm. Nevertheless, when accounting for how the data was obtained, it is possible to see that a unique design expert coded the data used in this study. In practice, this might not be the case. Therefore, a simulated scenario was proposed. Dividing the training sample between two sub-samples meta and training (see Fig 2) as specified in the methods section and carrying out different experiments, it was possible to see that when the test proportion was 0.9, even with a small proportion of noise the model lost accuracy at a great rate (see Table III).

TABLE II
PERFORMANCE METRICS BY TRAINING SIZE

TEST %	TRAIN SAMPLES	ACCURACY	RECALL	PRECISION
20	968	87.52	94.61	82.56
50	605	84.63	91.12	81.54
80	242	82.62	87.81	81.77
90	121	79.80	86.58	77.91
95	60	76.47	83.04	76.28
98	24	70.34	89.73	67.13

TABLE III
PERFORMANCE METRICS BY DIFFERENT LEVELS OF NOISE

TEST %	NOISE %	ORIG. ACC.	NEW ACC.	ORIG. RECALL	NEW RECALL	ORIG. PRECISION	NEW PRECISION
90	50	70.12	78.60	96.77	85.35	65.14	76.89
90	40	72.69	79.34	95.14	87.08	67.71	76.91
90	30	77.43	79.71	91.91	86.16	72.97	77.87
90	20	77.41	79.89	89.50	86.50	73.73	77.84
90	10	78.86	80.18	88.47	86.51	75.79	78.17
90	5	78.86	80.18	88.47	86.51	75.79	78.17
90	1	79.72	80.86	87.21	86.89	77.53	78.92
90	0	79.80	80.77	86.58	86.73	77.91	78.96

Thus, the importance of a meta-learning algorithm to account for the noise in the observations is evident.

Finally, the application of the meta-learning algorithm proposed by [10] in the noisy data seems promising. While keeping the proportion on training samples constant and varying the noisy order to simulate the error, the algorithm achieves a greater accuracy and precision in all scenarios. For instance, when looking at Table III it is possible to see that the algorithm maintains the performance of the network in the face of noise, keeping all performance metrics high and balanced while the accuracy and precision of the baseline model decrease.

V. DISCUSSION

The incorporation of AI in education is the next step for educational technologies. While there are multiple applications in which AI-like approaches may help instructors and other stakeholders encounter better alternatives to those used until now, it is necessary to encounter interaction between humans and technological tools for these implementations to be successful. AI-based technologies which provide high-quality feedback are a way of making engineering sketching instruction more scalable and personalized.

The assessment scenario is particularly challenging. Assessment is already a controversial topic, and the incorporation of automatic tools can create tensions between supporters and detractors. Nevertheless, the need for feedback and its importance [17], makes the use of intelligent feedback relevant for the educational community. Therefore, the more human interaction that exists between intelligent systems and humans, the better for the integration and acceptance of such systems into engineering education [8]

Nevertheless, the more interaction and flexibility of intelligent systems, the greater is the risk of finding extreme cases or difficulties in the training of the algorithms. In this research, we have demonstrated a specific case in which personalizing, defined as human participation in the system, can potentially carry noise. We have also experimentally shown how noise, even when in a small proportion, can lower the accuracy of an otherwise stable algorithm (see Table II).

Meta-learning, a branch of deep learning, has the potential to deal with some of the problems that might arise when using human interaction approaches. Furthermore, these types of algorithms open space to new approaches to the personalization of AI systems using fewer data points and learning from data provided by the stakeholder or users directly. This is the case of the learning to reweight algorithm proposed by [10]. We conclude that with the aid proportioned by this specific algorithm, a small proportion of valid classification samples, such as 61 samples, can be enough to train the network and achieve good accuracy results.

With these personalization features, instructors may use an AI system to support grading by first training on a small sample of assessed sketches. Sketching assessment can benefit from personalization when the system can accurately determine what level of learning students are at, providing more accurate information about student learning. Personalized assessment results can also inform higher-quality feedback provided by the system and/or instructions. The ability to implement personalized assessment in large classrooms is an advantage for making sketching assessment more scalable.

VI. LIMITATIONS

A comparison with other approaches for analyzing the data is necessary to assess other methodologies and make a holistic evaluation of the approach performance. The data and trained model apply only to perspective vs. non-perspective classification and have been proven beneficial when classifying rectilinear strokes. More studies need to be developed in order to argue its applicability to a broader spectrum of strokes.

ACKNOWLEDGMENT

This project was funded by the National Science Foundation "Collaborative Research: Fostering Engineering Creativity and Communication through Immediate, Personalized Feedback on 2D-Perspective Drawing": 2013612 (Texas A&M University), 2013504 (Georgia Tech), 2013575 (San Jose State University) and 2013554 (Purdue University). Special acknowledgment to Dr. Blake Williford for their contribution to the data collection process. In addition, a special acknowledgment to Yun-Wei Chu for their advice on the first implementation of the meta-learning algorithm.

REFERENCES

[1] C. L. Dym, A. M. Agogino, O. Eris, D. D. Frey, and L. J. Leifer, "Engineering design thinking, teaching, and learning," *Journal of engineering education*, vol. 94, no. 1, pp. 103–120, 2005.

[2] L. Brooks, "Structuration theory and new technology: analysing organizationally situated computer-aided design (cad)," *Information Systems Journal*, vol. 7, no. 2, pp. 133–151, 1997.

[3] J. K. Euler, "Advantages of using 3-d cad plant models on the construction site," *Computer-Aided Civil and Infrastructure Engineering*, vol. 9, no. 6, pp. 435–444, 1994.

[4] D. Veisz, E. Z. Namouz, S. Joshi, and J. D. Summers, "Cad vs. sketching: An exploratory case study," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, vol. 26, no. Special Issue 03, 2012.

[5] T. Hammond, S. P. A. Kumar, M. Runyon, J. Cherian, B. Williford, S. Keshavabhotla, S. Valentine, W. Li, and J. Linsey, "It's not just about accuracy: Metrics that matter when modeling expert sketching ability," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 8, no. 3, pp. 1–47, 2018.

[6] S. Keshavabhotla, B. Williford, S. Kumar, E. Hilton, P. Taele, W. Li, J. Linsey, and T. Hammond, "Conquering the cube: learning to sketch primitives in perspective with an intelligent tutoring system," in *Proceedings of the Symposium on Sketch-Based Interfaces and Modeling*, 2017, pp. 1–11.

[7] E. C. Hilton, J. S. Linsey, M. A. Paige, B. Williford, W. Li, and T. A. Hammond, "Board# 52: engineering drawing for the next generation: students gaining additional skills in the same timeframe," in *2017 ASEE annual conference & exposition*, 2017.

[8] B. Cope, M. Kalantzis, and D. Searsmith, "Artificial intelligence for education: Knowledge and its assessment in ai-enabled learning ecologies," *Educational Philosophy and Theory*, pp. 1–17, 2020.

[9] B. Williford, M. Runyon, and T. Hammond, "Recognizing perspective accuracy: an intelligent user interface for assisting novices," in *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 2020, pp. 231–242.

[10] M. Ren, W. Zeng, B. Yang, and R. Urtasun, "Learning to reweight examples for robust deep learning," in *International Conference on Machine Learning*. PMLR, 2018, pp. 4334–4343.

[11] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Deep imbalanced learning for face recognition and attribute prediction," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 11, pp. 2781–2794, 2019.

[12] T. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, "Meta-learning in neural networks: A survey," *arXiv preprint arXiv:2004.05439*, 2020.

[13] C. Giraud-Carrier, R. Vilalta, and P. Brazdil, "Introduction to the special issue on meta-learning," *Machine learning*, vol. 54, no. 3, pp. 187–193, 2004.

[14] E. S. Ferguson, *Engineering and the Mind's Eye*. MIT press, 1992.

[15] S. Askew, *Feedback for learning*. Routledge, 2004.

[16] R. S. Prawat, "Teachers' beliefs about teaching and learning: A constructivist perspective," *American journal of education*, vol. 100, no. 3, pp. 354–395, 1992.

[17] S. Askew and C. Lodge, "Gifts, ping-pong and loops-linking feedback and learning," *Feedback for learning*, pp. 1–17, 2000.

[18] D. Fossati, B. Di Eugenio, S. Ohlsson, C. Brown, and L. Chen, "Data driven automatic feedback generation in the ilist intelligent tutoring system," *Technology, Instruction, Cognition and Learning*, vol. 10, no. 1, pp. 5–26, 2015.

[19] T. M. Mitchell *et al.*, "Machine learning," 1997.

[20] R. Vargas, A. Mosavi, and R. Ruiz, "Deep learning: a review," *Advances in intelligent systems and computing*, 2017.

[21] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.

[22] W. Zhang, Y. Wang, and Y. Qiao, "Metacleaner: Learning to hallucinate clean representations for noisy-labeled visual recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 7373–7382.

[23] T. Liu, W. Fan, and C. Wu, "A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset," *Artificial intelligence in medicine*, vol. 101, p. 101723, 2019.

[24] J. Becktor, E. Boukas, M. Blanke, and L. Nalpantidis, "Reweighting neural network examples for robust object detection at sea," *Electronics Letters*, 2021.

[25] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[26] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf