LUCY-learn: Visualising Image Classifiers for Education in Machine Learning

Isabel Millar
Computer Laboratory
University of Cambridge
im539@cam.ac.uk

Alan F. Blackwell
Computer Laboratory
University of Cambridge
afb21@cam.ac.uk

Abstract

We present LUCY-learn, an application designed for use in classrooms, that allows students to experiment with the application of different image classifiers to satellite images in the vicinity of the students' school or neighbourhood. The Land Use Classification sYstem (LUCY) illustrates fundamental probabilistic principles of machine learning by visualising the probability distributions of classifier features and output layers. Learning is scaffolded through visualisations ranging from a hue-based pixel classifier to comparison of local image features acquired by a simple convolutional neural network. The result has been evaluated in classroom trials, with successful learning outcomes.

1. Introduction

The constant advance of machine learning (ML) technology means we must begin to integrate machine learning education into K-12 computing education. Developing students' awareness of not only what ML applications can do but how they do it – crucially including their limitations – is critical to a comprehensive understanding of computer science. Despite government endorsement of the use of generative AI in education (Department of Education, 2025), there is no mention yet of ML in the English computer science curriculum (Department of Education, 2014). This is not due to a lack of interest or ability; Evangelista et al found that high school students have a desire to learn more about ML and proposed a workshop that students of this level are capable of understanding, presenting no reason *not* to teach it (Evangelista, Blesio, & Benatti, 2018).

This paper follows previous contributions at PPIG, including a presentation at PPIG 2024 of applications developed by Josephine Rey and Gemma Penson (Rey et al., 2024), one evaluating an interactive visualisation of Causal Bayesian Networks in a UK classroom context and the other extending the Scratch language with PPL functionality to investigate how classroom use of interactive statistical modeling tools in South Africa might support curriculum priorities in that country. In this paper we build on that work with LUCY-learn – the Land Use Classification sYstem – an interactive image classification for iPad, allowing students to visualise and compare different classes of image classifier, applied (as in Rey's work) to satellite images.

2. Previous approaches to ML education

To teach technical details of ML systems, many tools aim to increase transparency of algorithms and training processes, often allowing students to train models themselves. Jiang et al studied students' understanding of text classification models by creating StoryQ, a tool to allow students to investigate and create models, exposing the impact of individual feature weights in the classification of a sentence (Jiang et al., 2022). Focusing specifically on k-means clustering, 'SmileyCluster' introduces students to ML concepts by allowing them to group face glyphs together (Wan, Zhou, Ye, Mortensen, & Bai, 2020). 'Next world adventure' was created to help students conceptually understand n-gram models, allowing students to choose their own sentences (Vahedian Movahed & Martin, 2025). An image recognition tool for high school students was developed by Mariescu-Istodor and Jormanainen, aiming to explain how machine learning applications work, making classifications based on aspect ratio and fullness (Mariescu-Istodor & Jormanainen, 2019).

Tedre et al integrated ML concepts into the computational thinking agenda, highlighting the differences in testing, correctness and transparency of algorithms (Tedre, Denning, & Toivonen, 2021). The range of teaching methods in ML education, extends from workshops and discussion to more hands-on approaches that help engage students and make abstract concepts more tangible (Lee & Kwon, 2024). Broll et al aim to make complex ML principles available to students without previous mathematics knowledge, gradually exposing detail of the training and inference processes (Broll, Grover, & Babb, 2022). Similarly, Winn's *Model-Based Machine Learning* focuses on connecting the abstract mathematics behind ML systems to real-world scenarios where models are used (Winn, 2023). These strategies can be related to constructivist theories of learning that students learn most effectively when participating actively (Triantafyllou, 2022) so that new knowledge is built on previous understanding. Building an understanding of classification by starting with the most simple architecture possible can help to eliminate these misconceptions.

3. The LUCY-learn system

LUCY-learn provides four separate image classifiers, each including controls and output visualisations. Classifiers are displayed in order from simplest to most complex, providing a natural progression to help students build knowledge. Two classifiers are simply pixel based, the first classifying hues by ranges within a distribution, and the second a Bayesian multinomial classifier (Jurafsky & Martin, 2025). The third and fourth classifiers apply convolutional neural networks to demonstrate use of local image features.

In order to support classroom evaluations designed for schools in Oxford and Cambridge, a selection of satellite images from these areas were captured using the MapBox API, chosen to cover a range of different land uses. The project is written in Python and Swift, and the CNNs are implemented using timm. Fine tuning is completed using PyTorch. Models were converted to a format suitable for mobile deployment using CoreMLTools.

3.1. Simple pixel classifier

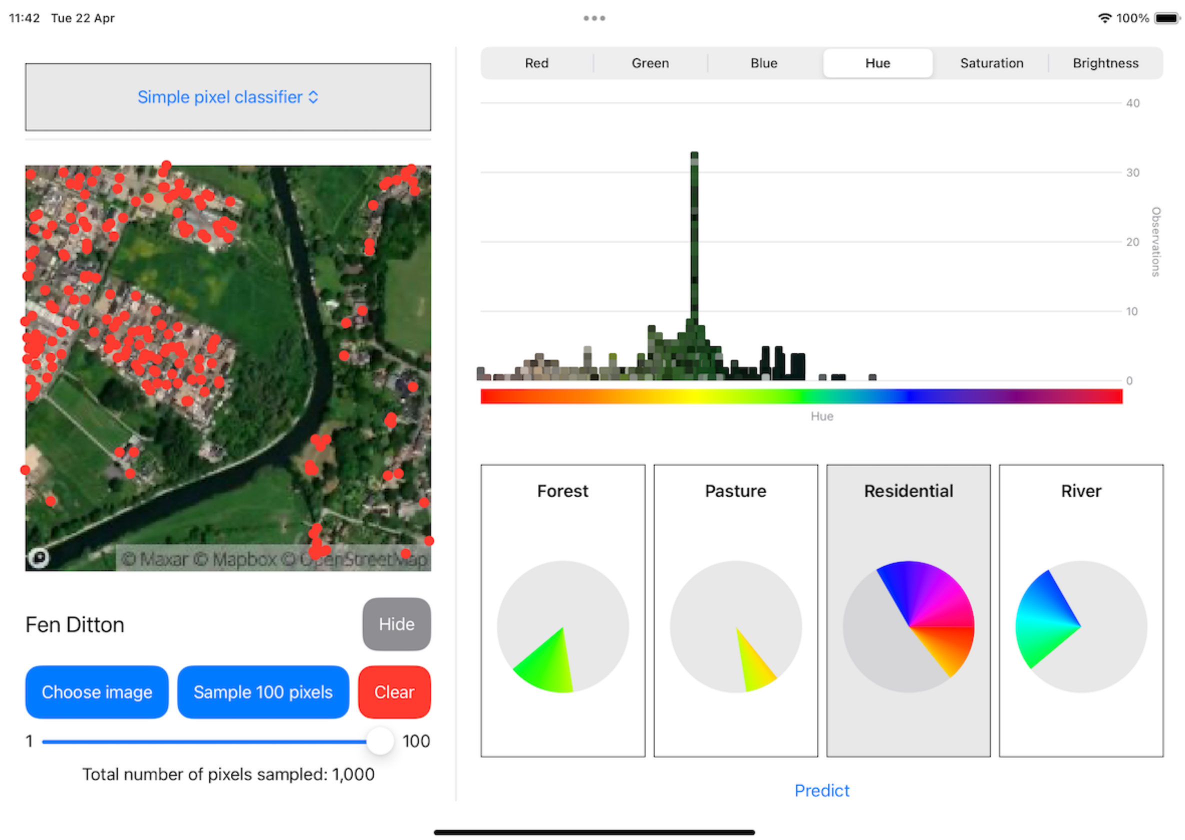


Figure 1 – Hue-based pixel classifier, showing pixels that have been predicted as 'Residential'

The simple pixel classifier tests the simple hypothesis that a hue range may be linked to land use, for example greenish hues may be forest. In many cases this simple assumption is incorrect, which stimulates a discussion about what can be done to improve accuracy. The layout of the dashboard encourages students to think critically about why results are produced.

For this introductory level, only four classes are used: Forest, Pasture, Residential, and River. They appear distinctive in satellite images, and represent a large proportion of the land use in Oxford and Cambridge. The classifier makes a judgement based on the hue values of a set of pixels sampled from an image. A method to train this classifier is not provided; instead, the hue ranges are hard-coded into the class, and a prediction function is implemented to calculate the probability of each class.

This adapts the approach of Rey et al (Rey et al., 2024), in which hue values sampled from a satellite image are visualised as a histogram and pie chart. Our version includes an option to hide the satellite image, encourage students to make their own judgement from the distribution of hues. Trying (and potentially failing) to guess what the image looks like with only limited information helps them understand why more features are needed for better classifications. The controls allow students to select a satellite image of their choice and sample from one to one hundred pixels.

Sampled pixels update the histogram and pie chart (figure 1) in real time. The histogram allows the distribution to be sorted by red, green, blue, hue, saturation or brightness. The UI displays each of the hues as a sector of the colour wheel, showing students which hues are classed as which land use. These can be selected to highlight which pixels in the image are in this range of hues (Figure 1).

The incorrect classifications made by this simple classifier exposes the problems of using local colour as the only feature for classification. For example, when provided with an image that students would label as ‘River’, the classifier predicts ‘Forest’ because the proportion of ‘blue’ pixels in the image is lower than the ‘green’ pixels of the banks.

3.2. Bayesian classifier

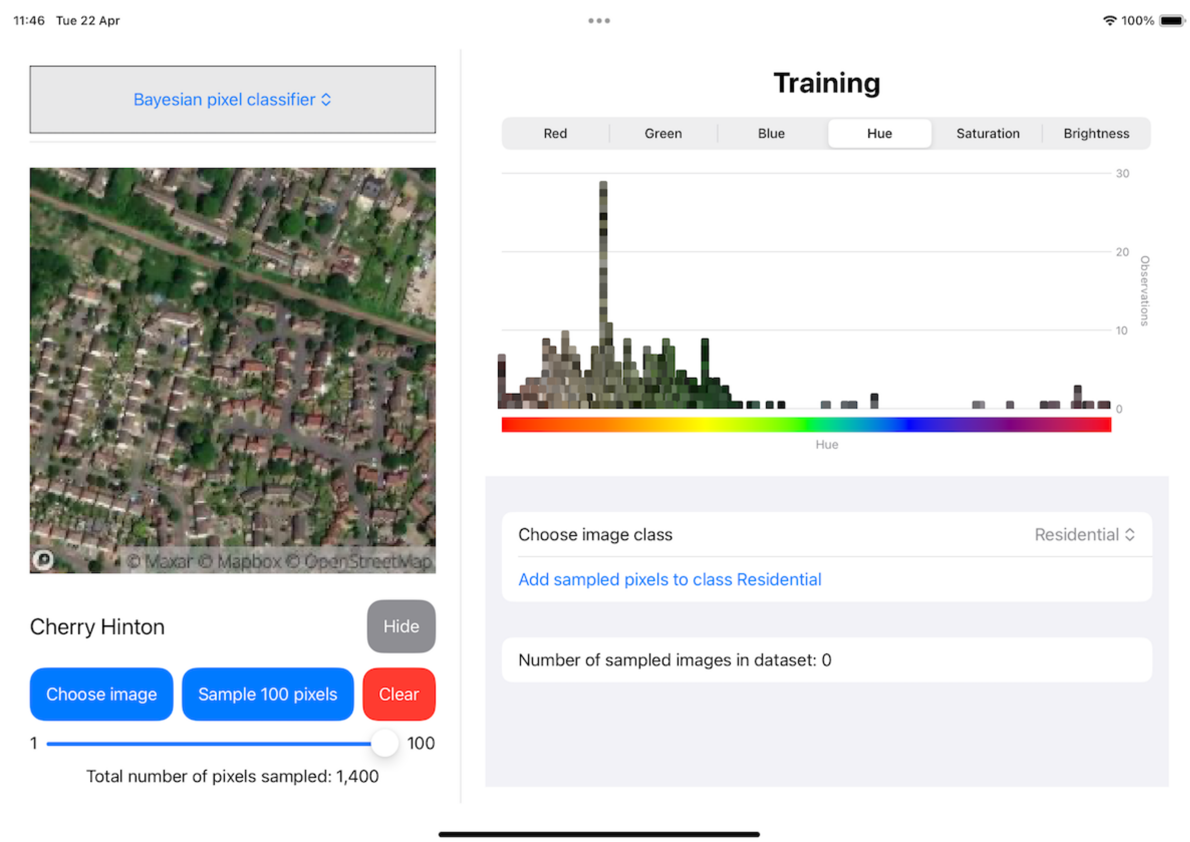


Figure 2 – The training interface for the Bayesian classifier.

The second classifier is a Bayesian multinomial classifier (Jurafsky & Martin, 2025). This provides an introduction to both Bayesian probability and training and inference processes. By allowing students to build a dataset and train a classifier the concepts of prior and posterior probability are illustrated, as classifications for an image update based on previously seen data.

As this classifier also uses pixel values, the UI is very similar to the simple pixel classifier, with the same controls to allow the user to sample pixels and choose images, and the same hue distribution histogram (figure 2). Buttons are added to the main visualisation panel to label a set of pixels based on the image's class, add this to the training dataset, and train the classifier. Once the classifier has been trained, students can sample pixels and make predictions.

The small volume of data used to train this model allows students to perform data collection and labelling themselves. While this does not produce a very accurate classifier, it is sufficient to help students to understand training, inference and Bayesian probability. The relatively poor accuracy helped facilitate a discussion about the amount of data needed to train models, and the effects of a biased or incomplete dataset.

3.3. CNN land use classifier

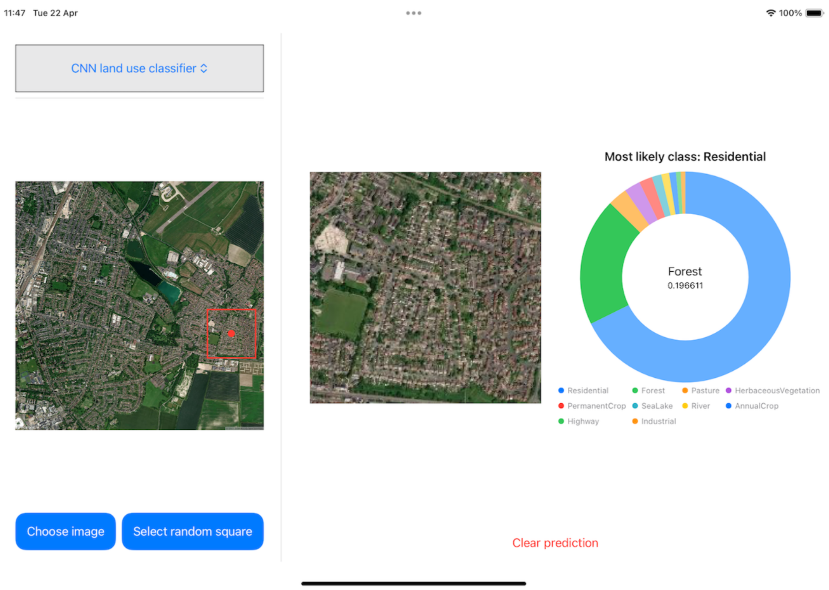


Figure 3 – CNN land-use classifier, showing part of a satellite image that has been selected from the larger area on the left with the 'Forest' section selected in the ring chart.

The CNN classifier employs a model pre-trained from EuroSat images as the basis for land use classification [1]. A visualisation of convolution features was also included in the app, and was used to support additional Year 12 curriculum content on the use of matrices in image processing (not discussed further here). The land use classifier dashboard shows magnified patches extracted from a larger satellite image. In the same style as the simple pixel classifier, a ring chart displays the probabilities of each land use class (figure 3). This reinforces that classifiers make classifications based on probabilities.

3.4. CNN school classifier

The final classifier is a CNN that we trained to classify image regions as either 'school' or 'not school', with a novel visualisation to help students assess why certain images may be seen as more or less 'school-like' (figure 5). We created a training dataset including images of all schools in Oxford and Cambridge, using coordinates from `https://get-information-schools.service.gov.uk/`, and images collected using the Mapbox API. A contrast set of not-school images from these regions was randomly generated. Once a model with suitable accuracy was obtained, it was converted to a CoreML

[1] `https://huggingface.co/cm93/resnet18-eurosat`

model and integrated into the app in the same way as the previous CNN.

We created a visualisation of ‘schoolness’ in the same style as the hue distribution histogram in the pixel classifiers, providing continuity throughout the app and helping students learn by building on their previous understanding of the pixel classifier. The schoolness scale on the x-axis corresponds to the output probability of the classifier. Values are sorted into ten bins, and the x-axis shows a thumbnail example of an image classified in each bin. This allows students to see trends in how the image changes to be more school-like, identifying features that might have more impact on schoolness. Students can then test other images with similar features to see if they are classified similarly.

Figure 4 – School classifier prediction with histogram showing bins for ’schoolness’

The visualisation includes a slider to change the schoolness threshold. This slider was used in lessons to help facilitate discussions about accuracy, and what might constitute a suitable value for different applications. This prompted conversation about precision and recall, for example we might want a classifier that never produces a false positive prediction, or in contrast we might want to be sure all schools have been selected and can tolerate selecting some extra images as well.

4. Learning Evaluation

Machine learning is not currently included in the English national curriculum, meaning there was no basis for comparison to current teaching. The goal was to develop students’ understanding of image classification and the probability concepts underlying classification methods. School trials were conducted by the first author in two classes – one year 10 computer science class and one year 12 further mathematics class. The variation in age and ability allowed us to focus on evaluating slightly different parts of LUCY-learn with each group, using lesson plans following a similar structure, but with more mathematical detail (of logarithms and matrices) for the more advanced class. The study was approved by the Ethics Committee of the Cambridge Department of Computer Science and Technology.

Students completed pre-lesson and post-lesson questionnaires, using a long-answer open question to

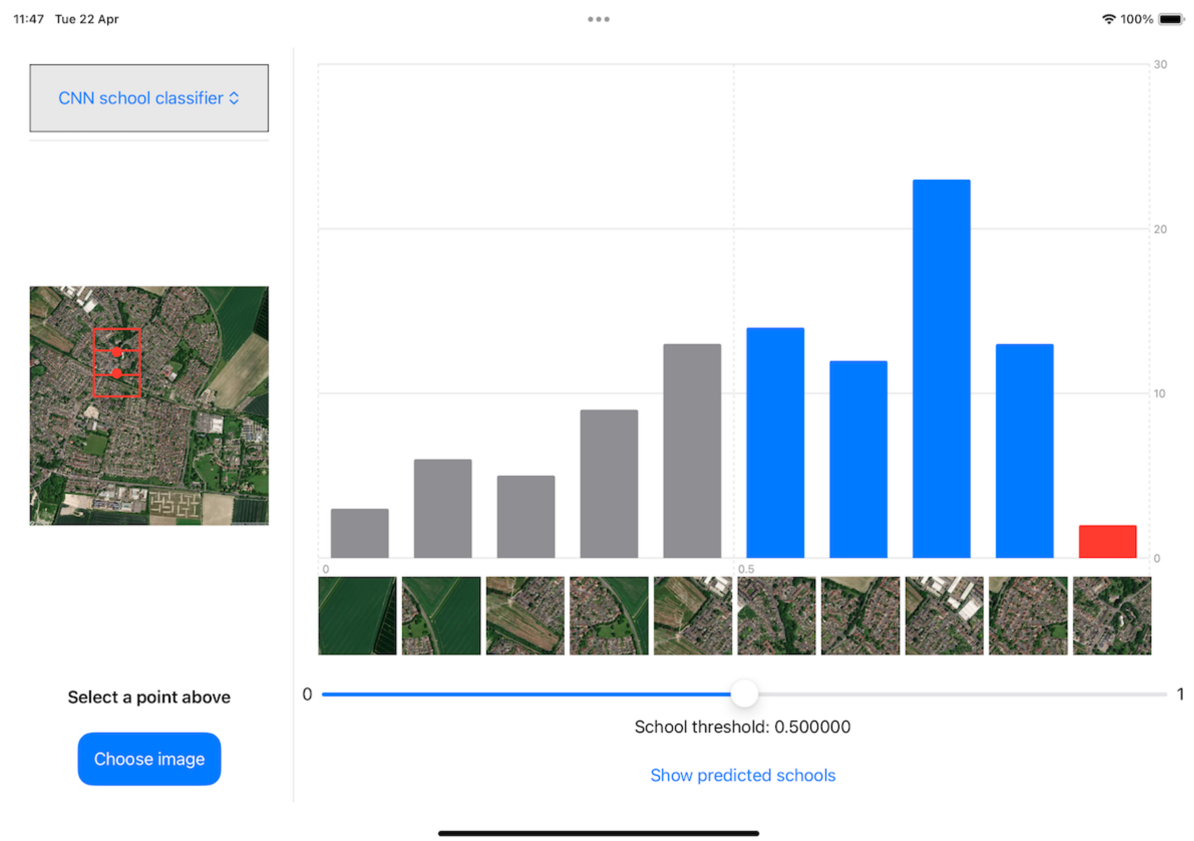


Figure 5 – The novel histogram visualisation of 'schoolness'.

enable qualitative evaluation of the development of understanding. Students also completed a Likert-scale evaluation, to report their assessment of LUCY-learn by comparison to any previous learning about image classifiers or ML.

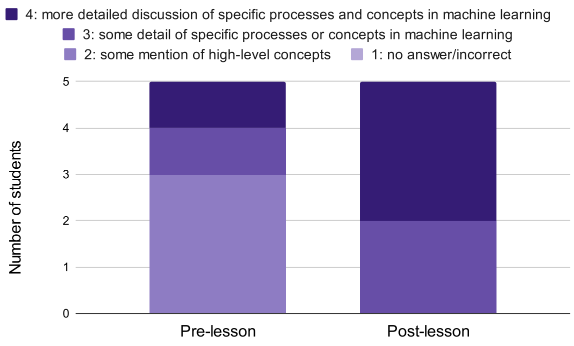
4.1. Year 12 lesson

The first trial was with a group of seven Year 12 students in a Further Mathematics class at Oxford High School (OHS). Students were allocated one school iPad between two, allowing them to collaborate on tasks. Students picked up details of the two pixel-wise classifiers quickly and were able to correctly use the app to train the Bayesian multinomial classifier, understanding why predictions were being made and the impact of the dataset in the prediction process. When discussing performance of the classifiers, students were able to identify issues with accuracy and suggest solutions for these problems. Overall, students were able to successfully use LUCY-learn to pick up concepts and put them into practice.

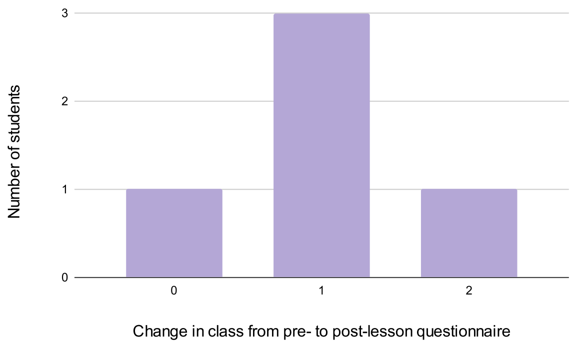
Students answered the following question before and after the lesson:

> Websites often use CAPTCHA tests – for example 'select all squares with traffic lights in' – to check if a user is human. Self-driving cars can use image classifiers to identify hazards on the road and therefore make decisions about actions to take. How can the answers collected from CAPTCHA tests help improve self-driving cars' ability to recognise objects on the road?

Students' answers were separated into four classes (figure 6): no answer/incorrect (1), some mention of high-level concepts (2), some detail of specific processes or concepts in machine learning (3), and more detailed discussion of specific processes and concepts in machine learning (4). The development of students' understanding through the lesson is evident in the figure. Answers to the pre-test question stated that CAPTCHAs collect labels from humans and can be used to train self-driving cars, but with



(a) Classes



(b) Change in class

Figure 6 – Evaluation of pre-lesson and post-lesson responses for year 12 class.

no detail of what this training entails. Some students mentioned that cars may make a prediction based on an image being "similar" to one it has already seen, but did not explain what constitutes "similar" images.

In the post-lesson questionnaire, all students mentioned probability or the inner workings of CNNs, as well as how we may use the output of a CNN to provide a confidence rating and make a decision of class based on this. Discussion of probabilistic qualities of classifiers, such as the fact that the and of confidence indicates understanding of the important distinction between data-driven systems and the rule-driven systems addressed in the current computer science curriculum (Caspersen, 2023).

4.2. Year 10 lesson

To evaluate the educational benefits provided by LUCY-learn over a range of ages and abilities, a second trial was completed with 17 mixed-ability Year 10 students in a computer science class at St Bede's School in Cambridge. The school does not have iPads, so the lesson was taught by projecting a demonstration, then handing the iPad around for students to try themselves. Students had previously completed a topic on data representation including images, so the lesson placed a greater focus on the simple pixel and Bayesian classifiers to tie into the year 10 (UK GCSE) curriculum. The lesson went well and students went from being able to identify barely any applications of machine learning to making intelligent analyses about why and how different classifiers make classifications.

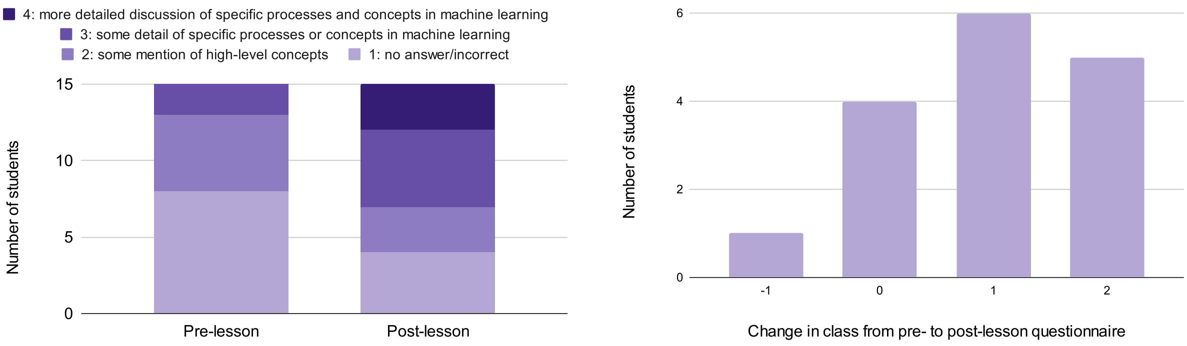
The evaluation question for this lesson was simplified to suit year 10 students' age and experience:

> Some smart cameras like ring doorbells can detect objects in the photos and videos they take, for example a package being delivered or a person at your door. This is called image classification. How does a computer figure out what object is in a picture?

In the pre-lesson questionnaire, only seven students attempted an answer, in contrast to the eleven attempts in the post-lesson questionnaire.

Most students were able to understand the role of colours in the classification of an image in the post-lesson questionnaire, with many making reference to proportions of each colour. Some commented only on different colours suggesting different classes – while others integrated more detail about probability and accuracy of classifiers.

Higher-achieving students attempted explanations in their pre-lesson questionnaires, spanning a range of "based off other similar images on the internet" (P13) to the general comment of "machine learning" (P6). Most of these students' answers were much more developed in the post-lesson questionnaire, referencing probabilities and in some cases more complex architectures. P14 started with an understanding of needing a database of images to check against, to an explanation of how a probability value for each class can be constructed in a pixel-level classifier and CNNs can extract other features in images.



(a) Classes *(b) Change in class*

Figure 7 – Evaluation of pre-lesson and post-lesson responses for year 10 class

Answers were grouped into the same classes as for the year 12 lesson (figure 7). Overall, eleven of the fifteen student questionnaires showed a development in understanding, with their post-lesson answer being classed higher than their pre-lesson one. These answers made comments about the use of colour in classification, as well as the need for labelled data, large datasets, and classifications' reliance on probability. The results show a clear development in understanding.

Category	LUCY-learn		Other ML	
	Mean	STD	Mean	STD
Understanding	4.6	0.52	4	0.96
Ease of use	4.47	0.52	5	0.52
Application	3.87	0.52	3.83	0
Interest	4.1	0.67	3.75	0.5

Table 1 – Summary of the results from Year 12 students' Likert-scale evaluations.

Likert scale responses showed that LUCY-learn was generally well received by year 12 students, with no negative values. Table 1 shows a summary of the mean and standard deviation for each category of questions. Students' self-evaluation of their progression was that LUCY-learn helped them to develop their own understanding, but they felt less confident to apply their knowledge to other situations or to teach other students. For year 10 students, only one student had learnt about machine learning before, so we do not report scale comparisons. Students in the year 10 class also reported that they might struggle to apply what they had learned.

5. Conclusion

We have described an app that builds on the PPIG 2024 paper by Rey et al (Rey et al., 2024). Rey's work presented an extension to Scratch, in which a Scratch background based on satellite images could be used to explore basic principles of Bayesian probability. The LUCY-learn app focuses specifically on satellite images as an example of machine learning classifiers. Where Rey's work was designed for deployment in schools in South Africa, informed by consultation with educators in that country, LUCY-learn focuses on extending the UK computer science and mathematics curriculum with fundamental principles of machine learning.

LUCY-learn provides a progressive model in which students are able to experiment with more sophisticated approaches to image classification, from simple hue categories to convolutional deep networks. The implementation of these models includes an opportunity for students to experiment with training their own classifier, evaluation of how pretrained classifiers generalise to new contexts, and critical assessment of the accuracy of specialised deep learning models via a novel visualisation of class likelihoods for sub-images.

Classroom evaluation confirmed that students at different school levels and with a range of abilities were able to learn core principles of machine learning, beyond the current UK curriculum. This understanding was evaluated in relation to scenarios recognisable to students, although there is still room for development of LUCY-learn to allow students to test their understanding in different situations. One example of this would be to integrate another way to build their own classifier in the app, perhaps with more detail than the existing Bayesian classifier training, potentially exposing students to some code and allowing them to test their understanding. The single lesson that we taught did not yet give students confidence that they could apply this knowledge, but this could be addressed by integrating more practical activities into the app.

6. References

Broll, B., Grover, S., & Babb, D. (2022). Beyond black-boxing: Building intuitions of complex machine learning ideas through interactives and levels of abstraction. In *Proceedings of the 2022 acm conference on international computing education research - volume 2* (p. 21–23). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3501709.3544273` doi: 10.1145/3501709.3544273

Caspersen, M. E. (2023). Principles of programming education. In (2nd ed., chap. 18). Bloomsbury Publishing.

Department of Education. (2014). The national curriculum in england: Key stages 3 and 4 framework document [Computer software manual]. Retrieved from `https://www.gov.uk/government/publications/national-curriculum-in-england-secondary-curriculum`

Department of Education. (2025). *Generative artificial intelligence (ai) in education.* Retrieved from `https://www.gov.uk/government/publications/generative-artificial-intelligence-in-education`

Evangelista, I., Blesio, G., & Benatti, E. (2018). Why are we not teaching machine learning at high school? a proposal. In *2018 world engineering education forum - global engineering deans council (weef-gedc)* (p. 1-6). doi: 10.1109/WEEF-GEDC.2018.8629750

Jiang, S., Nocera, A., Tatar, C., Yoder, M. M., Chao, J., Wiedemann, K., ... Rosé, C. P. (2022). An empirical analysis of high school students' practices of modelling with unstructured data. *British Journal of Educational Technology*, *53*(5), 1114-1133. Retrieved from `https://bera-journals.onlinelibrary.wiley.com/doi/abs/10.1111/bjet.13253` doi: https://doi.org/10.1111/bjet.13253

Jurafsky, D., & Martin, J. H. (2025). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition with language models* (3rd ed.). Retrieved from `https://web.stanford.edu/~jurafsky/slp3/` (Online manuscript released January 12, 2025)

Lee, S. J., & Kwon, K. (2024). A systematic review of ai education in k-12 classrooms from 2018 to 2023: Topics, strategies, and learning outcomes. *Computers and Education: Artificial Intelligence*, *6*, 100211. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2666920X24000122` doi: https://doi.org/10.1016/j.caeai.2024.100211

Mariescu-Istodor, R., & Jormanainen, I. (2019). Machine learning for high school students. In *Proceedings of the 19th koli calling international conference on computing education research.* New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3364510.3364520` doi: 10.1145/3364510.3364520

Rey, J., Penson, G., Blackwell, A. F., Ge, H., Li, X., & Arnold, H. (2024). Educational tools for probabilistic machine learning curriculum in schools. In *Proceedings of the 35th annual workshop of the psychology of programming interest group (ppig).*

Tedre, M., Denning, P., & Toivonen, T. (2021). Ct 2.0. In *Proceedings of the 21st koli calling international conference on computing education research.* New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3488042.3488053` doi: 10.1145/3488042.3488053

Triantafyllou, S. A. (2022). Constructivist learning environments. In *Proceedings of the 5th international conference on advanced research in teaching and education.*

Vahedian Movahed, S., & Martin, F. (2025). From pre-conceptions to theories: How middle school student ideas about predictive text evolve after interaction with a new software tool. In *Proceedings of the extended abstracts of the chi conference on human factors in computing systems.* New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3706599.3719789` doi: 10.1145/3706599.3719789

Wan, X., Zhou, X., Ye, Z., Mortensen, C. K., & Bai, Z. (2020). Smileycluster: supporting accessible machine learning in k-12 scientific discovery. In *Proceedings of the interaction design and children conference* (p. 23–35). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3392063.3394440` doi: 10.1145/3392063.3394440

Winn, J. (2023). *Model-based machine learning*. CRC Press. Retrieved from `https://books.google.co.uk/books?id=y_jYEAAAQBAJ`