Received: 9 June 2017 | Revised: 20 February 2018 | Accepted: 15 March 2018
DOI: 10.1111/jcal.12263

WILEY Journal of Computer Assisted Learning

ORIGINAL ARTICLE

Supervised machine learning in multimodal learning analytics for estimating success in project-based learning

Daniel Spikol[1] | Emanuele Ruffaldi[2] | Giacomo Dabisias[2] | Mutlu Cukurova[3]

[1]Malmö University, Sweden
[2]Scuola Superiore Sant'Anna, Italy
[3]UCL Knowledge Lab, University College London, UK

Correspondence
Daniel Spikol, MAU, Nordenskioldsgatan 1, 21119 Malmö, Sweden.
Email: daniel.spikol@mau.se

Funding information
Seventh framework programme of the European Community for research and technological development, Grant/Award Number: 619738

Abstract

Multimodal learning analytics provides researchers new tools and techniques to capture different types of data from complex learning activities in dynamic learning environments. This paper investigates the use of diverse sensors, including computer vision, user-generated content, and data from the learning objects (physical computing components), to record high-fidelity synchronised multimodal recordings of small groups of learners interacting. We processed and extracted different aspects of the students' interactions to answer the following question: Which features of student group work are good predictors of team success in open-ended tasks with physical computing? To answer this question, we have explored different supervised machine learning approaches (traditional and deep learning techniques) to analyse the data coming from multiple sources. The results illustrate that state-of-the-art computational techniques can be used to generate insights into the "black box" of learning in students' project-based activities. The features identified from the analysis show that distance between learners' hands and faces is a strong predictor of students' artefact quality, which can indicate the value of student collaboration. Our research shows that new and promising approaches such as neural networks, and more traditional regression approaches can both be used to classify multimodal learning analytics data, and both have advantages and disadvantages depending on the research questions and contexts being investigated. The work presented here is a significant contribution towards developing techniques to automatically identify the key aspects of students success in project-based learning environments, and to ultimately help teachers provide appropriate and timely support to students in these fundamental aspects.

KEYWORDS
machine learning, multimodal learning analytics, project-based learning

1 | INTRODUCTION

Over the last several years the field of learning analytics (LA) has grown rapidly in conjunction with massive open online courses and other technology systems. These systems include virtual learning environments, mobile applications, and student-response systems, which are rapidly becoming part of the everyday educational landscape. These systems collect and provide diverse types of data about learners' interactions that take place with the systems and among the learners, potentially allowing, new insights into education. Such systems often highlight the importance of big data in education that is of interest to diverse actors for using LA for educational management and policy making (Clow, 2013). However, from a learning sciences research perspective, the aim of LA is to understand and optimise the learning process, and most learning happens outside of these systems between people in face-to-face situations (Siemens & Baker, 2012; Greller & Drachsler, 2012). In this research paper, we investigate multimodal learning analytics (MMLA) to make sense of students' learning process in project-based learning activities with the purpose of optimising it for students and teachers.

Project-based learning activities have the potential to help educators to achieve high tier institutional and policy goals such as developing 21st century skills in science, technology, engineering, and mathematics subjects. More specifically for teaching technology subjects, such as computer science and information communication technology, project-based learning is a commonly employed approach, and its popularity is increasing, particularly after the introduction of the "makers movement." Most of these project-based approaches involve learning activities that combine hands-on computing technologies to explore various topics in both secondary and postsecondary learning institutions (Halverson & Sheridan, 2014). However, these hands-on activities introduce many challenges due to their dynamic and multifaceted nature, specifically regarding their design, implementation, and evaluation. Existing evidence shows, it becomes clear that students do not become effective learners when they are left on their own within such "student-led" learning environments (Kirschner & van Merriënboer, 2013). Therefore, appropriate monitoring and guidance of students' in these pedagogical approaches is an essential requirement for their success. Nevertheless, owing to practical challenges of project-based learning including the fact that teachers lack the required time and resources to attend and support each student group (or each student within groups) during their engagement with the projects, these type of project-based learning approaches often struggle to satisfy their common learning outcomes.

However, MMLA offers researchers new tools to capture different types of data from complex learning activities including project-based learning. The ability to collect multimodal data from bodily movements, face tracking, affective sensors, hardware and software log files, and user and research-generated data provides opportunities to obtain unique features, which can be interpreted to understand and appropriately support project-based learning. The multimodal data from these sensors provide new opportunities for investigating learning activities in the real-world between small groups of learners working on tasks with physical objects (Blikstein & Worsley, 2016). The automated collection and presentation of insights from MMLA to support project-based learning approaches is an exciting emerging field within the LA domain and it has the potential to provide the required support for students and teachers involved in project-based learning approaches to help them achieve their learning outcomes.

Starting from the initial assessment conducted by the authors (Spikol, Ruffaldi, Landolfi, & Cukurova, 2017), in this paper, we investigated how MMLA data can be used to support project-based learning from a specially designed worktable environment, where small groups of students use new physical computing components to solve open-ended tasks. To achieve this, we built a multimodal learning analytics system (LAS) that is part of the students' project worktable and collected diverse streams of data. We processed and extracted multimodal interactions to answer the following question: Which features of students' group work that can be automatically collected with our MMLA system are good predictors of students' project outcomes in open-ended learning activities with physical computing? To answer this question, we have explored different supervised machine learning approaches employing traditional and deep learning (DL) techniques to analyse the data coming from multiple sources. Our work is a significant contribution towards providing ways to automatically identify the key aspects of students success in project-based learning environments and to ultimately help teachers provide appropriate and timely support to students in these key aspects.

The paper is structured as an exploratory design work: The next section presents the background, followed by the system context and the material and methods, which includes the design of the intervention. We then present the results, followed by a discussion and our conclusions.

2 | BACKGROUND

The roots of project-based learning extends back almost a century to John Dewey's approach that argues for "laboratory schools" in which students are engaged with the process of inquiry in their learning activities (Dewey, 1938). The history of this approach is rich, and while a detailed literature review of the approach is outside the scope of this paper. Project-based learning is a form of situated learning, which allows students to engage in similar real-world activities to those in which the professionals engage (Krajcik & Blumenfeld, 2006). Project-based learning activities that support learners' participation in open-ended tasks are one of the most commonly used teaching approaches for improving 21st century skills (Bell, 2010); they emphasise the engagement of learners in projects that are personally meaningful, and they encompass driving questions, investigations, and collaboration (Krajcik, 2010). However, the hands-on and open-ended nature of project-based learning creates challenges for tracking the learning process. One of the key challenges faced in project-based work is the support of the group work and ensuring that students succeed in the planned learning outcomes (Blumenfeld et al. 1991; Krajcik & Blumenfeld, 2006).

Current research in MMLA focuses on developing a better understanding of the complexity of learning through the advances of high-frequency multimodal data capture, signal processing, and machine learning techniques (Ochoa & Worsley, 2016). MMLA offers an opportunity to capture different insights about learning in project-based learning tasks in which students have the opportunity to generate unique artefacts such as computer programs, robots, and small group collaboration to solve open-ended tasks (Blikstein & Worsley, 2016; Blikstein, 2013). MMLA builds upon multimodal human interaction, educational data mining, and many other fields that include learning sciences and cognitive sciences to capture the complexity of learning through data intensive approaches (Worsley, 2012; Siemens & Baker, 2012).

In terms of the focus on purposes and context, an emerging body of work with in the field of MMLA captures small group work on project-based learning that has grown mainly out of the work of Worsley and Blikstein (2013) investigating engineering students' design activities (Blikstein, 2013; Chen et al., 2014; Ochoa et al., 2013). Within this research domain, Blikstein (2011) explored multimodal techniques for capturing code snapshots to investigate students learning computer programming and video and gesture tracking for engineering tasks; Worsley (2014) presented different approaches for data classification that included points about how these techniques have a significant impact on the relation of research and learning theories. Both of these initial approaches provided the means for other researchers to begin exploring MMLA with small groups of students across different subjects. Ochoa et al. (2013) presented the math data and oral presentation quality data corpora that has enabled the community to analyse and discuss the different requirements and results within this field. Moreover, Ochoa and colleagues' (2014) work used existing multimedia processing technologies to produce a set of features for accurate predictions of experts in groups of students solving math problems, which illustrated the benefits of MMLA to support students' learning in these contexts. Similarly, Chen and colleagues (2014)

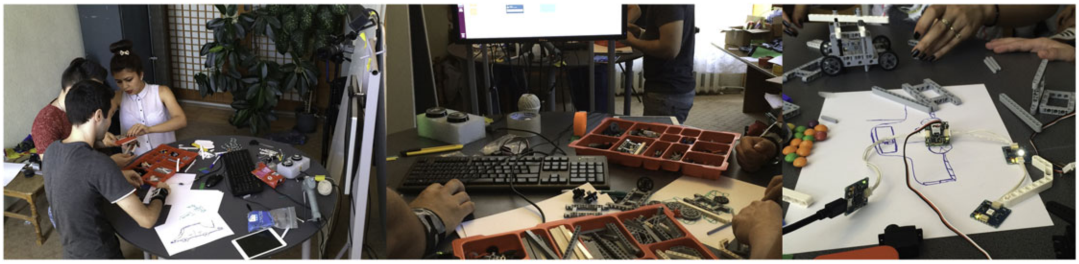

FIGURE 1 University engineering students working in the PELARS environment [Colour figure can be viewed at wileyonlinelibrary.com]

expanded from the oral presentation quality data corpus to further examine the feasibility of using multimodal technologies for the assessing of public speaking skills; Grover and colleagues (2016) explored how to develop computational models of social learning environments. In their work, Grover and colleagues (2016) managed to classify the quality of collaboration from body movement and gestures of pair programmers working together with acceptable accuracy rates. Although most of the existing MMLA research approaches focus on learners' data, Prieto and colleagues (2016) and Martinez-Maldonado and colleagues (2016) focused their research efforts on how MMLA can support teaching actions and orchestration in the classroom.

However, regarding the technical focus, and to make sense of complex data streams coming from multiple data sources, MMLA researchers employ various computational techniques. These approaches include logistic regressions (Ochoa et al., 2013), different feature reduction algorithms (Worsley, 2014; Schneider & Blikstein, 2014), and statistical models to investigate MMLA and to identify features and predict student performances (Schneider & Blikstein, 2014). These approaches all have advantages and disadvantages depending on the main research question and the purposes of data analysis and have the potential to provide insights into how to proceed with a multimodal data set. Regardless of which computational approach is taken, existing literature shows that MMLA has a role to play to support education in project-based learning approaches, and it has the potential to provide new means for gathering insights for complex open-ended learning activities (Blikstein & Worsley, 2016), which are otherwise extremely challenging to monitor and support with existing traditional standardised evaluation approaches.

3 | SYSTEM CONTEXT

The work discussed in this paper is based on the European project "practice-based experiential learning analytics research and support" (PELARS).[1] The central goal of the project was to develop LA tools for hands-on; open-ended science, technology, engineering, and mathematics; and science, technology, engineering, arts and math project-based learning activities using physical computing. The learning contexts we have investigated are high schools, engineering, and design departments at universities. The current system includes customised furniture with an integrated multimodal LAS such as tracking hands, faces and other objects, and the Arduino platform with a visual web-based integrated development environment (IDE) that captures interaction information of physical computing. The learners and observers use mobile devices to capture multimedia data (text, images, and video) to self-document the learning activities.

Overall, the PELARS project has developed an intelligent system for collecting activity data (LAS) for diverse LA (with data mining, reasoning, and visualisations) and active user-generated material and digital content (that include mobile tools and physical computing platform) for project-based learning activities (Cukurova, Avramides, Spikol, Luckin, & Mavrikis, 2016; Spikol, Ehrenberg, Cuartielles, & Zbick, 2015). See examples of the PELARS system in action with the university engineering students in Figure 1.

3.1 | PELARS LAS

The LAS collects multimodal data from different sensors and input from the learners and researchers. The learning environment is designed to foster collaboration and includes an integrated screen and standing round table to allow learners to share and work together. The LAS collects data from both ambient (sensors) and live sources (human interaction). The ambient collection of data includes a computer vision system that uses colour and depth cameras with audio for understanding how people interact around the workstation furniture. The LAS uses a web-based architecture in which a classroom located data collector performs data acquisition and vision processing, sending data to a remote server using WebSockets. The system has been designed to work in offline mode allowing to later synchronise the content on the remote server. The data on the server are further processed for extracting LA and statistics. For details about the architecture please refer to Ruffaldi, Dabisias, Landolfi, and Spikol (2016).

3.2 | Physical computing

A core part of the system are small Arduino-based boards that play a fundamental role in the project-based activity of the students. These boards use the TALKOO IDE, which has been designed to allow users to start building electronic devices without having to build circuits neither on breadboards nor prototyping boards and without having to write complex lines of code (Katterfeldt, Cuartielles, Spikol, & Ehrenberg, 2016). The visual programming interface is a web tool (HTML5 based) to the standard Arduino IDE. This platform has been developed for the project with plug-and-play sensors and actuators together with a flow-based visual programming IDE that allows learners to prototype

[1] http://www.pelars.eu

artefacts rapidly. A set of "sentiment/affective" buttons has also been developed with thundercloud and sunshine icons to allow the students to mark critical events in their activities.

3.3 | Mobile tools

The set of mobile tools has been developed to provide the means for the learners to self-document the learning process across the planning, building, and reflection phases of their projects with different content and multimedia data. Additionally, the mobile tools allow researchers and teachers to mark critical incidents, and to time stamp the different stages of the learners' project. The tool is developed on the basis of modern web technologies, which run across different platforms (Zbick, Vogel, Spikol, Jansen, & Milrad, 2016).

3.4 | Collected data

The automatically collected data include the capture of objects; the positions of people, hand movements, and faces; audio and video; interactions of plugged components from the Arduino-based physical computing platform; and interactions with the sentiment buttons. Additionally, the mobile-based tool can be used to collect self-documentation annotations from students and progress annotations from researchers or teachers examining the students. In particular, in the experimental settings employed in this work, the researchers have annotated the activity cycle marking the phases of planning, building, and reflecting. Figure 2 shows the PELARS system with the Arduino blocks, sentiment buttons, and students.

4 | MATERIALS AND METHODS

The automatic scoring presented in this work has been assessed by evaluating engineering students performing project work. This section discusses the subjects involved, the nature of the assessment, and the data acquired. Details about the data processing are discussed in the following section.

4.1 | Dataset acquisition

The data analysed in this paper are from three sequential educational interventions with 18 engineering students at a European university (17 men and 1 woman, average age 20 years). The students were divided into six groups of three students. Each student group used the system over 3 days to complete one open-ended design task for each session. First, the students were introduced to the system in a workshop to familiarise them with it, and then their first task was to create a prototype of an interactive toy. The second task was the prototyping a colour sorter machine, and in the third task, the students were asked to build an autonomous automobile. Each of these design sessions ranged from 33 to 73 min. Each of the tasks introduced a more complex design concept to be solved with respect to the previous ones. Students were asked to perform an initial phase of planning, followed by an execution/building phase and finally a documentation/reflection phase. During the activity, the students had to document their planning, building, and reflecting phase using a mobile tablet. The tablet allowed the students to take photographs, record video, and report via a form and free text their plan, progress, and reflective thoughts. No specific instructions about the timing of these phases were given to the students. Additionally, the research observers used the mobile tool to divide the students' work flow into the planning, building, and reflecting phases.

4.2 | Scoring of the students' projects

To grade the students' design projects, a scoring scheme was developed that combined different approaches for collaborative problem solving (CPS) in small groups and guided by design thinking principles. We started with the seminal work done with engineering students (Atman et al., 2007) that was initially adopted by Worsley and Blikstein (2014) for MMLA. From these initial frameworks, we began to develop a framework for CPS (Cukurova et al., 2016) that we could apply to the PELARS context. We used a version of our CPS framework with the mobile system with an agreed set of codes for on-fly observations to initially grade the students' CPS competencies. However, this approach to coding student competencies on-the-fly was too challenging due to the dynamic nature of the student interactions in practice-based learning (PBL) and intercoder reliability issues.

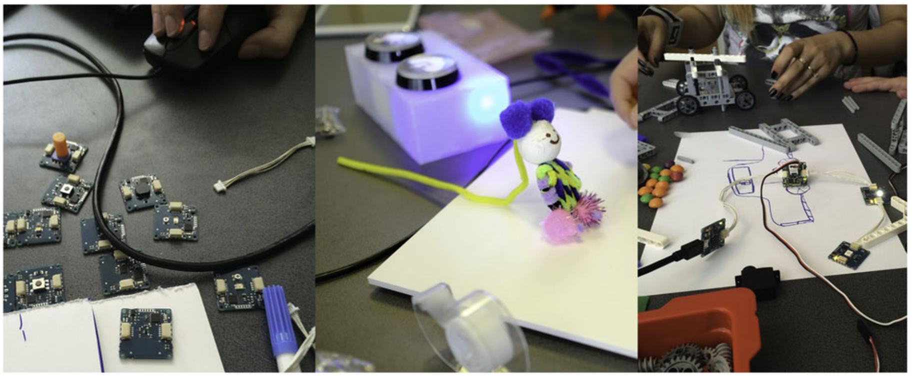

FIGURE 2 Details of the Arduino-based boards, sentiment buttons, and diverse student projects [Colour figure can be viewed at wileyonlinelibrary.com]

4.3 | Initial classification of the students' project outcomes

From the initial on-the-fly coding approach, we switched to postactivity coding. This grading of the student project outcomes were coded by two senior researchers and four research assistants at Malmö University and University College London using the below explained criteria. The team of researchers reviewed the students work collected in the LAS, which included snapshots of the students' plan, video of solutions, and learner's text input. The 18 sessions were graded, where 50% of the grade was the expert's opinion based on the documentation collected by students, 25% was how the students planned and delivered the artefact, and the remaining 25% was the student's self-assessment of the quality of their projects. Although, no statistical correlation was calculated among the raters, this new approached yielded a good amount of overlap among the raters and small discrepancies in the grades of the groups were resolved through discussions between the coders. The resulting scores were categorised into three classes: *poor*, *ok*, and *good*. This classification of the sessions was used as the reference point for the previous machine learning based classification work (Spikol et al., 2017) in which the nature of this evaluation allowed only to reliably classify the works in two classes: *good* and *poor*.

4.4 | Improved classification of the students' project outcomes

The binary approach used in the previous scoring each of sessions was then further reevaluated and rescored by the experts looking at videos, documentation (from the mobile tools) and the final project outcome (the artefact). The aim was to generate a richer scoring system that reflected the learning practises for engineering courses. The new scoring system was based on five different aspects expressed as a scale from 1 to 5:

- Level of clarity [Loc] (5=*very clear*, 3=*legible*, 1=*not understandable*)
- Independent thinking [InTh] (5=*independent*, 3=based off instruction, 1=*same as instruction*)
- Corresponds with plan [CorPl] (5=*Fully*, 3=*partially*, 1=*not at all*)
- Does it work? [DoWo] (5=*fully*, 3=*partially*, 1=*not working*)
- Quality of solution [QuaOS] (5=*great*, 3=*mediocre*, 1=*poor*)

Table 1 presents all the scores, and Figure 3 shows the quality score for the six teams.

4.5 | Acquired data: The investigated features of the LAS

For each sessions recorded, the LAS system collected data from the students comprising activity performed, user-generated content (text and

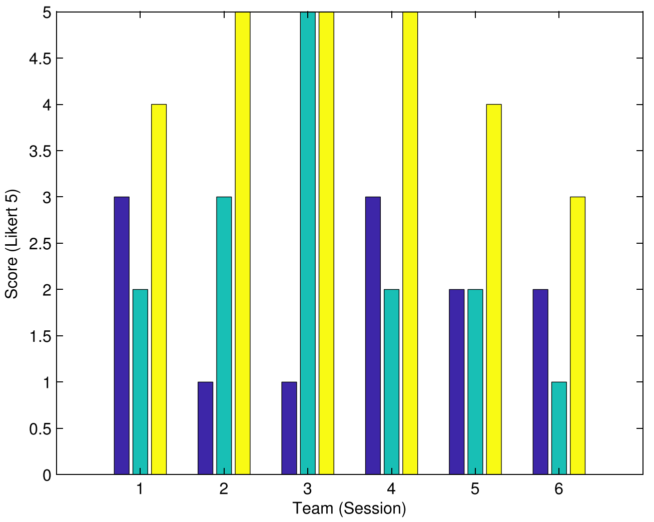

FIGURE 3 Quality of solution scores (QuaOS) of each team during the three sessions. Session 1 (blue), Session 2 (green), and Session 3 (yellow) [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Table of the 18 scores organised by team on the basis of a 5-point Likert scale

Team	Session	Clarity	Independent thinking	Plan	Solution working	Quality
A	1	5	2	5	4	3
A	2	1	1	5	5	5
A	3	5	3	5	4	5
B	1	2	3	3	3	2
B	2	1	3	3	1	1
B	3	2	4	1	3	2
C	1	1	4	3	5	4
C	2	2	1	5	5	5
C	3	5	3	2	2	2
D	1	4	5	1	1	1
D	2	5	3	4	4	4
D	3	5	4	3	3	3
E	1	4	4	4	3	3
E	2	2	1	3	3	3
E	3	2	2	3	4	2
F	1	2	5	3	5	5
F	2	3	5	2	1	2
F	3	1	3	2	1	1

multimedia), and actions on the Arduino visual IDE. In particular, the following data were acquired.

4.5.1 | Face tracking

Using a frontal camera (Logitech C920, 960×540 resolution at 30 Hz) and the Viola-Jones algorithm (Viola & Jones, 2004), all the visible faces of students were tracked, and, through camera calibration and assumptions about face size, it was possible to estimate the 3D position of student's head with respect to the camera. Owing to per session calibration between the cameras and a fixed point on the table, it is possible to express the pose of student faces in a 3D reference frame of the table. From the face tracking data, two metrics were identified: the count of faces looking at the screen (FLS) and the distance between the faces, which provides an indicator of the distance between learners (DBL). We hypothesise that the measure DBL might be a proxy of collaboration, because students' physical proxy is a required but insufficient condition for students collaboration. We expect that when the DBL is small it is more likely that a collaboration will occur among students, and there is sufficient evidence to suggest that a collaboration has the potential to improve students' learning outcomes.

The adopted algorithm is quite robust to facial differences and illumination conditions, although it is primarily designed for frontal faces. Additional detectors are available for lateral faces. To compensate for sudden motions, we interpolated pose information when the face was not detected for a short period.

4.5.2 | Hand tracking

The top down colour with depth camera (Microsoft Kinect One, 1920×1080 resolution at 30 Hz) monitored the motion of the hands of the students that were wearing fiducial markers (Garrido-Jurado, Muñoz-Salinas, Madrid-Cuevas, & Marín-Jiménez, 2014) that disambiguate each primary hand. The pose is estimated by combining the image-based marker tracking with the depth information. Owing to the calibration of the camera and the size of the markers, the 3D position of the hands was obtained with respect to the table. Based on the 3D position of the hands, we were able to calculate two metrics: the distance between hands (DBH) and the hand motion speed (HMS).

In terms of tracking capabilities, wristbands with fiducial markers provide precise information when the marker is visible and with a non-lateral orientation. In comparison with markerless trackers, this solution is robust to object handling, although research is progressing well in this direction thanks to DL (Sridhar et al., 2016).

4.5.3 | Arduino IDE

The interface between the visual Arduino IDE and the data collection system provided information about the types of physical and software blocks used in the project and their connections. In particular, we counted the number of active blocks (IDEC), the variety of hardware (IDEVHW) and software blocks used (IDEVSW), and the number of interconnections between blocks as a measure of complexity (IDEX) in students' programming during their project-based activities.

4.5.4 | Audio level

Using the microphone included in one of the cameras and fast fourier transformation, we computed the sound level during the sessions. The resulting feature was a value sampled at 2 Hz called audio level (AUD).

5 | DATA PROCESSING FOR MMLA

5.1 | Experimental setup and research design

In the previous section, we presented two approaches for the classification of student projects together with the different MMLA features collected by the PELARS system. Figure 4 illustrates the different MMLA features (the independent variables) that are used as predictors of student project outcomes (dependent variables). The following section presents how these MMLA data were preprocessed and how different machine learning approaches were used to render them.

5.2 | Preprocessing

From all these MMLA data points, the data were collected at variable data rates (around 2 Hz), yet the data were not synchronised. For this reason, we needed a processing stage that aggregated indicators from the different variables in windows of same duration. The aggregation was performed on the basis of counting for most of the variables. We only employed averaging for the distance or proximity features. Considering that students' sessions were different lengths due to the open-ended nature of the project-based learning activities, we employed zero padding for sessions that were too short.

5.3 | Machine learning approach

A supervised machine learning approach was employed for associating the measured student actions with the resulting scores by the experts.

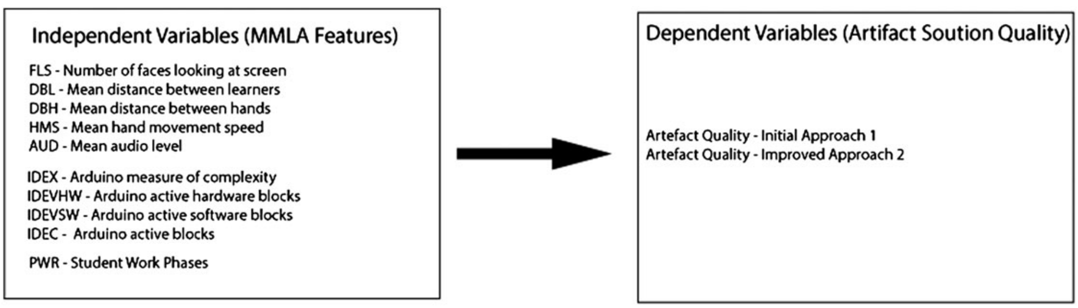

FIGURE 4 Independent and dependent variables

TABLE 2 Machine learning tasks performed over data

Method	Deep learning	Traditional
Task	Regression	Classification
Input	18 variables	9 variables per window
Output	6 scores over 5 levels	1 score with 3 levels
Metrics	Regression score	Classifier accuracy
Windowing	120,240 and 360 s	10,20,30,90 min
Phase exclusion	Reflection	Reflection
Method	Multiple layers	NB, LR, SVML, and SVMR

Note. NB = naive Bayesian; LR = logistic regression; SVML = support vector machines with linear kernel; SVMR = support vector machines for regression.

In particular, we performed a two stage approach using different techniques. One assessment was based on large data quantities and used DL for regressing the five scores by the experts (improved approach to artefact quality evaluation). The second, based on traditional machine learning, dealt with the simpler three-level assessment of the sessions (initial approach to artefact quality evaluation) and attempted to explain the causes of the outcome depending on measured features and phases. Table 2 shows a synthetic view of the two tasks together with the inputs, outputs, and details about the algorithms as discussed in the remainder of this section.

5.3.1 | Deep learning regression of outcome

DL has been tested to check the feasibility of nonlinear regression on the input data gathered from the sensors. A deep neural network (DNN) is composed using a graph of linear matrix multiplications, which are followed after each stage by a nonlinear function called *activation function*. The general behaviour can be synthesised as follows: given an input vector x, a series of matrices A_i composed of weights $w_{(k,j)}$, a bias vector b, an activation function F, and an output Y_i, it is possible to write stage i as

$$Y_i = F(A_i x + b).$$

The output Y_i will then be the input of the next stage of the pipeline, until reaching the end of it, where a classifier or regressor computes the final output. DNNs can be used for classification or regression: in the first case, the network is trained to obtain a label indicating the category to which the input belongs, whereas in the latter case, the network learns to fit an unknown function using the input and output data in order to estimate points that are not present in the input set. For the purpose of this study, regression was used because the output values can be a set of continuous values.

The input data are a set of time series that have different data rates and partial synchronisation. In this work, we decided to use a windowing approach with a dense network to compensate for such differences leaving the use of recurrent neural network techniques for future work. Given a session of duration T seconds, we split it into nonoverlapping windows of length L seconds (120,240 and 360) obtaining $\lceil T/L \rceil$ windows. For a given input, we compute an aggregated statistics for each window (averaging or summation). Each window is sent separately as input to the neural network. The following aggregated statistics (18 values in total) have been employed:

- Total number of faces looking toward the screen (FLLS)
- Total number of connected Arduino components (IDEVHW)
- Mean distance between learners (DBL)
- Mean distance between hands (DBH)
- Mean hand movement speed (HMS)
- Mean audio level (AUD)
- Mean hand positions (HP)
- Mean Arduino components activity (IDEC)

Given this, the network has been trained to fit a function, which has an 18-dimensional domain and a six-dimensional codomain. Several additional network parameters were tuned to obtain the best possible solution along with the window size for the input data creation. These parameters include: (1) dropout, (2) regularization method, (3) epochs, and (4) layers.

Input data are randomly split, as is usual in training and test data, with a minor split of the training data again into training and validation. In these experiments, 20% of the sessions were removed as test sessions, leaving 80%, for training. Of this 80% another 20% was used as a validation set during the training phase. Notably complete sessions were left out for testing and not just random inputs (windows) because they are usually correlated and could alter the final results if used. The results of the network were evaluated using a mean squared error distance between the predicted value vector, and the true value vector obtained in the test data set. A mean squared error was also computed for each of the six output values along with the variance to understand whether any of the output values had a different behaviour. Six different DNN architectures have been tested, growing from one to six fully connected layers starting with a size of 1,024 and decreasing at each layer by half. The best obtained network was created using the following parameters: Dropout 0.5, no regularisation, 100 epochs, 3 dense layers (1024,512,256), and 240 seconds window size. Figure 5 shows the the network structure.

The network was implemented using a Python library for DNNs called *Keras* (Chollet, 2015). This high-level library allows the use of the graphics processing unit optimised processing libraries *Tensorflow* Abadi (2016) and *Theano* (Bergstra et al., 2010).

5.3.2 | Outcome classification

At this stage, we performed a supervised classification task that matches the observers' scores. The purpose of this approach is to identify the data features that can support different score classifications that have been evaluated by human observers (artefact quality coders) as *poor*, *ok*, and *good*. Among the different families of classifiers available, we tested various parametric ones, namely, naive Bayesian (NB), logistic regression, and support vector machines with linear kernel (SVML) and Gaussian kernel (SVMR). We avoided the nonparametric ones (nearest neighbours) or decision trees with to reduce the overfitting effect. In particular, the NB is a simple classifier that employs a strong assumption about features, a condition that holds valid for most of the variables we employed in our investigation except for the ones related to the Arduino IDE. We decided not to use the ensemble of classifiers (Kotsiantis, Patriarcheas, & Xenos, 2010), as we would like to study the model behind these classifications as much as performing the

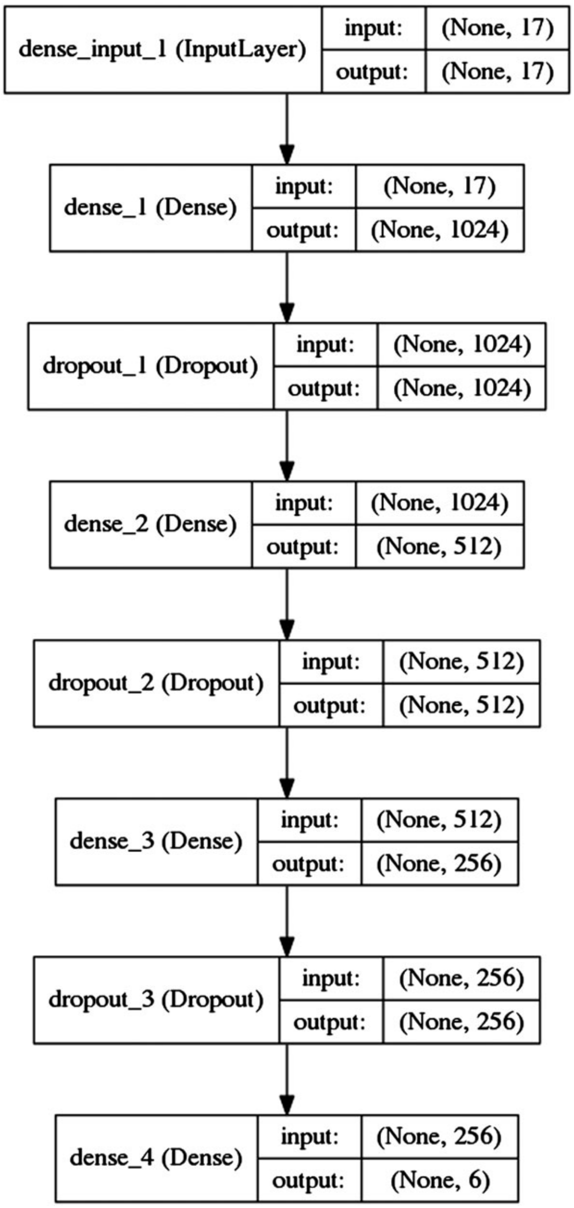

FIGURE 5 Neural network structure of the model that obtained the best results

classification itself. Understanding the model behind classifications is fundamentally important in social science contexts, particularly in education. For this classification task, time was considered by using larger windows of size (10, 20, 30 min, and whole session) aggregating the data similarly with the previous approach but considering the values from all the windows together and padding the sessions with a smaller number of windows.

We used cross-validation ($k = 4$) for understanding the effect of different parameters such as window size and the inclusion of different phases. Owing to the small sample size (18 sessions from 18 engineering students working in six groups of three students), we avoided the leave-one-out scheme. The data acquired from the PELARS LAS were exported and then processed in Python using the sklearn (Pedregosa et al., 2011) toolkit that provides state-of-the-art machine learning techniques integrated with a common interface. The test of the classifiers was performed by varying the window size, the score (binary or original three level), the inclusion of the different phases (planning, building, and reflecting) and, most importantly, the effect of features identified and described above (FLS, DBL, DBH, HMS, IDEC, IDEVHW, IDEVSW, IDEX, and AUD).

6 | RESULTS

In this section, we present the results of the evaluation comparing the DL-based regression with the classification.

6.1 | Deep learning regression of outcome results

Table 6 illustrates the overall results for the different network structures. Tables 3, 4, and 5 show the mean and variance for the error between expected output and predicted value. We compared 120, 240, and 360 s window sizes and the 240 s network and achieved a mean squared error of 0.13, as shown in Table 4, across the improved classification of the students' outcomes. We then investigated the different features by removing each feature individually (Table 6). In general, the results got worse as expected, except in the case of distance between faces DBF as reported in Table 7. This result illustrated that this feature of distance between faces is a substantial input for project-based work in the PELARS context. Additionally, the results show that the smallest window performs worse than the others (see Table 3). The network achieving the best results used a window size of 240 s as shown in Figure 5.

TABLE 3 Results for the 120s window, 0.242 overall accuracy

120s window	Loc	InTh	CorPi	DoWo	QuaOS	OG
Mean	0.182	0.238	0.166	0.197	0.155	0.228
Var	0.074	0.112	0.069	0.076	0.061	0.099

Note. CorPi = corresponds with plan; DoWo = does it work?; InTh = independent thinking; Loc = level of clarity; QuaOS = quality of solution.

TABLE 4 Results for the 240s window, 0.129 overall accuracy

240s window	Loc	InTh	CorPi	DoWo	QuaOS	OG
Mean	0.086	0.175	0.150	0.175	0.154	0.084
Var	0.074	0.056	0.084	0.092	0.062	0.048

Note. CorPi = corresponds with plan; DoWo = does it work?; InTh = independent thinking; Loc = level of clarity; QuaOS = quality of solution.

TABLE 5 Results for the 360s window, 0.193 overall accuracy

360s window	Loc	InTh	CorPi	DoWo	QuaOS	OG
Mean	0.213	0.077	0.237	0.147	0.196	0.181
Var	0.097	0.006	0.083	0.063	0.071	0.057

Note. CorPi = corresponds with plan; DoWo = does it work?; InTh = independent thinking; Loc = level of clarity; QuaOS = quality of solution.

TABLE 6 Best network results for the different network configurations

Layers	Error	Window (s)
1024	0.186	360
1024, 512	0.174	360
1024, 512, 256	0.129	240

TABLE 7 Best error scores after removing isolated features

Removed feature	Best result
No features removed	0.129
All faces data	0.21
All Arduino data	0.21
DBF	0.15
DBH	0.21
HMS	0.19
AUD	0.18
Hand pos	0.21
Arduino comp	0.19

6.2 | Outcome classification results

6.2.1 | Phases

Although, we had a small sample size of 18 sessions, the data generated from these sessions were rich and large due to the multimodal nature of our setup. The project-based learning activities lasted within the range of 33 to 75 min (median 63 min±13) with a total activity time of 17 hr and 10 min. Each project-based learning activity's project outcome was graded on the basis of the criteria described earlier, and different patterns along the three sessions were observed.

The design phases annotated by the observer (planning, building, and reflecting) varied broadly among the sessions and the groups. The mean scores for the time spent on these phases among the sessions are planning (11 min ± 10 min), building (41 min ± 16 min), and reflection (4 min ± 7 min). Figure 6 shows the duration of each session and the timing of the phases for different groups of students.

6.2.2 | Scoring

The three-level scoring we initially identified using human observation (*poor*, *ok*, and *good*) posed difficulties to the classification activity, and we needed to move to a binary version in which we aggregated ok graded groups with good graded groups. For example, NB and SVM classifiers score 0.8 and 0.75, respectively, with a window of 30 min and binary classification. However, this value decreases to 0.5 for both of the classifiers when we used a three-class classification. This situation is not ideal; and as explained earlier, in our further investigations, we have improved our human observer coding scheme to avoid such exclusions. However, to achieve adequate results, we took this binary approach, which still has great value to be able to identify project-based learning groups who perform poorly in comparison with others. Alternatively, this approach can be used to distinguish those group performances that are considered as good from the rest in a binary fashion. We see this as the first step towards obtaining more detailed classifications.

6.2.3 | Effect of phase

Across the different conditions, the selection of the phases was used to improved the capacity of the classifiers. For example, with a 30˜min window and binary classification, the exclusion of the reflection phase in student activities, provided stronger results across the different classifiers, while the exclusion of both the planning and the reflection phases reduced the classification power. Please note that the decision to omit the reflection phase from the data was taken due to statistical arguments. This decision does not reflect our lack of interest in the reflection stage from the research point of view. We think that reflection is an important phase of learning and would like to improve our algorithms in the future with additional data to generate meaningful results that include all significant phases (see Table 8 for the details).

TABLE 8 Effect of phases in the inclusion of the classifier with P = plan, W = work, and R = reflect

	PWR	PW	W	WR
NB	0.8	0.8	0.6	0.75
SVML	0.6	0.75	0.75	0.8
SVMR	0.75	0.75	0.75	0.75
LR	0.6	0.75	0.5	0.6

Note. NB = naive Bayesian; LR = logistic regression; SVML = support vector machines with linear kernel; SVMR = support vector machines for regression.

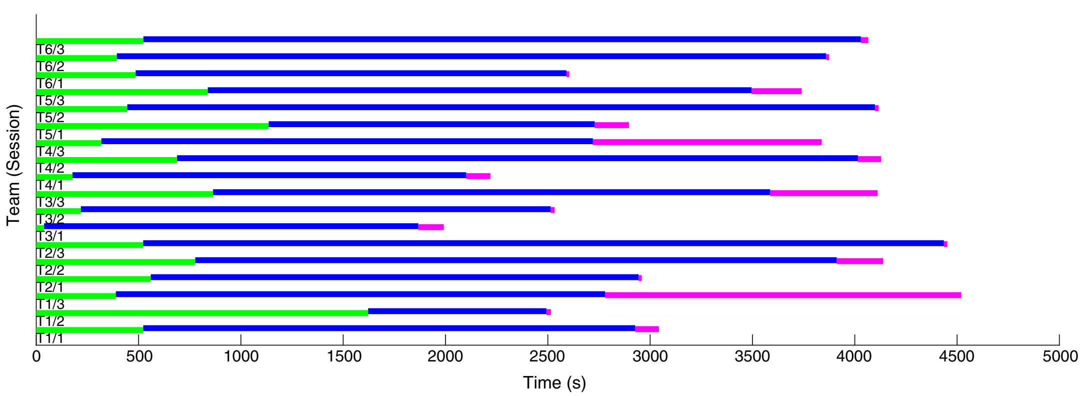

FIGURE 6 Distribution of phases among session of the six teams. Each session is split into the three phases, first plan (green), then build (blue), and finally reflect (purple) [Colour figure can be viewed at wileyonlinelibrary.com]

To provide the most reliable results and use the strongest classification power, we focus our results on data collected from the planning and working stages of the student activities and excluding the reflecting stages.

6.2.4 | Type of classifiers

As it can be seen in Table 1, across the different tests of the classifiers, those that behaved the most consistently were NB and the SVML.

6.2.5 | Effect of features

For better understanding, the role of each multimodal feature in the performance of the classifiers, we perform backward ablation of feature. This operation is performed over the configuration of the binary classifier on the basis of NB and SVML with window sizes of 30 min and the two phases of planning and building.

Regarding the effects of the MMLA features on predicting students group performances in open-ended project-based learning, the following results are found:

- IDEC (Arduino IDE) removal does not affect the results of the classifiers;
- Removal of all face and hand duration has very little effect on the classifiers;
- Distance measures DHB and DBL alone are capable of predicting the results with a high accuracy (0.75) across classifiers;
- The audio level feature AUD alone is currently a strong feature for classification (1.0 with naive Bayes) with time windows 5, 10, and 30 min and binary scoring.

The logistic regression is capable of an optimal result (1.0) when considering IDEX, IDEVHW, IDEVSW, and DBL, which are the main IDE features, except component counts and the DBL. As mentioned earlier, one of the main limitations of this approach is on the scoring of the sessions, which is limited to a binary classification.

7 | DISCUSSION

This article, started with the hypothesis that specific features in MMLA can provide useful information about the quality of groups' interactions and thus for the artefacts produced as part of the students' project-based learning. From the high-frequency multimodal data collected, we compared different machine learning approaches (that employed DL and traditional techniques) for their accuracy to predict human grading of the groups' artefact quality. In our first approach, using these classifiers, we identified the most effective features of MMLA to predict the students' group performances in project-based learning activities. More specifically, we used various machine learning classifiers to predict the poor student performances in terms of the groups' artefact quality on the basis of multimodal data. We were not satisfied with the binary grading system or the large time window, therefore we expanded the classification of the student's project outcomes to five categories. Then we used DNNs to evaluate student performances in project-based learning using multimodal data.

7.1 | Traditional approach

In the linear regression approach, we focused on identifying the different phases of work in relation to the accuracy in predicting the groups' artefact quality. We found that the planning and building stages of students' learning activities are better predictors of their artefact quality than the reflection stage (in the intervention, the reflection phase signalled the end of making artefacts and coding, and the start of documenting the work with a mobile device). After looking at the different phases, we investigated the certain features of the MMLA, to determine which features can predict the students' artefact quality with higher accuracy. Our results show that the DBH and DBL are key features used to predict students' performances in project-based learning activities. In our case, these features highly correlate with the quality of the students' artefacts in project-based learning. These results align with existing research on PBL activities that show the value of nonverbal indexes of student interaction in estimating their success at learning processes (Cukurova, Luckin, Millan, & Mavrikis, 2018) as well as MMLA research findings that show the potential of hand motion and speed, and the location of the learners to predict student success at various learning outcomes (Blikstein, 2011; Ochoa et al., 2013; Grover et al., 2016). As mentioned in Section 2, there are three main aspects of PBL: students are asking driving questions, doing investigations to answer these questions, and collaborate together to solve these questions (Krajcik, 2010). It is important that MMLA research aims to support these three main aspects of PBL. The results presented here that show the value of the distance between students' hands and the distance between students to predict students' success at PBL are well aligned with the argument that closer students might potentially fruitfully collaborate, which is an important aspect of PBL. This does not necessarily mean that students who are at a close distance will always collaborate effectively. However, the proxy of a close distance, appears to be a necessary but insufficient feature for success in the context of PBL.

The other features of MMLA such as HMS and FLS, did not perform very well to predict students' artefact quality across the classifiers. Although the Arduino IDE the Number of active blocks (IDEC), the variety of hardware (IDEVHW) and software blocks used (IDEVSW), and the number of interconnections between blocks as a measure of complexity (IDEX) were able to predict students' outcomes, they were only marginal across the classifiers. Furthermore, the audio signal level (AUD) appears to be a promising feature to predict performance, however, further investigation is needed for using this feature in combination with others.

7.2 | Deep learning approach

The DNN results are more promising than the traditional techniques and show the feasibility of this method as an efficient approach for MMLA. In our investigation using this approach, we obtained net achieves a mean squared error of 0.13 with a window of size 240 s as shown in Table 4. One important result emerged from our study is that the smallest window performed worse than the others (see Table 3). This result is possibly caused by the low information amount in that time window. The 240 s interval performs the best, whereas the 360 s interval gives no performance gain as can be seen in Table 5. These findings indicate that the information gain from 240 to 360 s is negligible for our purposes. It is possible to see that (see Table 7) by removing a single feature, in general, the results get worse, except partially in the case of the DBL. This result shows that DBL is a very strong input feature. The network learned some higher level features, which do not consist of a single input, given that by removing any single input, we can not achieve the optimal results that we achieved using them all.

All results show a reasonably low variance evidencing the stability of the results, which is a positive sign in terms of the learned features. The fact that strong features have been trained is possibly due to the 0.5 dropout value, which "encourages" the network to find high level, strong features discarding the low level, weak features. Regularisation failed to boost the results significantly, but this is probably due to the relatively "small" amount of training data, partially avoiding the problem of overfitting. This parameter should become more relevant when more data will be added to the training set. A future step could involve removing pairs or triplets of features to understand the relationship and importance of the input features and make the factors on the learning process more visible. We aim to investigate these features in our immediate future work.

8 | CONCLUSION

The recent growing interest in project-based learning globally is, at least in part, due to an increased demand for the "21st century skills" and the potential of project-based learning to improve student skills. The evidence set out in recent influential reports (see, for instance, Luckin, Baines, Cukurova, and Holmes (2017)) confirm that these skills, look set to be increasingly relevant not just for many of the jobs that will survive new waves of automation but also for our ability to cope in everyday life. However, project-based learning requires giving students appropriate support while they are engaging with physical materials and with each other (Cukurova, Bennett, & Abrahams, 2018; Kirschner & van Merriënboer, 2013).

This study showed that MMLA and the state-of-the-art computational techniques can be used to generate insights into the "black box" of learning in students' project-based activities. These insights generated from multimodal data can be used to inform teachers about the key features of project-based learning and help them support students appropriately in similar pedagogical approaches. Towards achieving this ultimate aim, this paper makes three contributions to the field. First, we show that the distances between students' hands and faces while they are working on projects is a strong predictor of students' artefact quality, which indicates the value of student collaboration in these pedagogical approaches. Second, we show that both new and promising machine learning approaches, such as neural networks and traditional regression approaches, can be used to classify MMLA data, and they both have advantages and disadvantages depending on the research questions and contexts being investigated. Finally, although it is traditionally notoriously challenging to provide evidence about the robust and objective evaluations of project-based learning activities, the techniques and types of data we presented here can be the first step towards the evidence-informed and effective implementation and evaluation of project-based learning at a scale.

ACKNOWLEDGEMENT

The PELARS project received funding from the European Union's Seventh Framework Programme for research, technological development, and demonstrations under grant agreement 619738.

AUTHOR CONTRIBUTIONS

This work is the result of the collaborative effort between the institutions participating in the PELARS FP7 project. DS and MC designed the protocol with students, DS conducted the evaluation with students, and ER and GD designed the software and analysed data. All contributed to writing the manuscript.

ORCID

Daniel Spikol http://orcid.org/0000-0001-9454-0793
Mutlu Cukurova http://orcid.org/0000-0001-5843-4854

REFERENCES

Abadi, M. (2016). TensorFlow: learning functions at scale. *ACM SIGPLAN Notices*, *51*(9), 1–1. https://doi.org/10.1145/3022670.2976746.

Atman, C. J., Adams, R. S., Cardella, M. E., Turns, J., Mosborg, S., & Saleem, J. (2007). Engineering design processes: A comparison of students and expert practitioners. *Journal of Engineering Education*, *96*(4), 359–379. https://doi.org//10.1002/j.2168-9830.2007.tb00945.x.

Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F. J., & Marín-Jiménez, M. J. (2014). Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, *47*(6), 2280–2292. http://www.sciencedirect.com/science/article/pii/S0031320314000235.

Bell, S. (2010). Project-based learning for the 21st century: Skills for the future. *The Clearing House*, *83*(2), 39–43.

Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., ... Bengio, Y. (2010). Theano: A CPU and GPU math compiler in python. In *Proceedings of the 9th Python in Science Conference*, Austin, TX, pp. 1–7.

Blikstein, P. (2011). Using learning analytics to assess students' behavior in open-ended programming tasks. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge—LAK '11*, ACM Press, New York, New York, USA, pp. 110. https://doi.org/10.1145/3022670.2976746.

Blikstein, P. (2013). Multimodal learning analytics. In D. Suthers, & K. Verbert (Eds.), *Proceedings of the Third International Conference on Learning Analytics and Knowledge—LAK '13* (pp. 102). New York, New York, USA: ACM Press. https://doi.org/10.1145/2460296.2460316.

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, *3*(2), 220–238. Retrieved from October 12, 2016, http://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/4383/5596.

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychology*, *26*(3–4), 369–398. https://doi.org/10.1080/00461520.1991.9653139.

Chen, L., Feng, G., Joe, J., Leong, C. W., Kitchen, C., & Lee, C. M. (2014). Towards automated assessment of public speaking skills using multimodal cues, ACM Press, New York, New York, USA, pp. 200–203. https://doi.org/10.1145/2663204.2663265.

Chollet, F. (2015). Keras: Deep learning library for theano and tensorflow. https://keras.io/k, 7, 8.

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education*, *18*(6), 683–695. https:\ignorespaces//doi.org/10.1080/13562517.2013.827653.

Cukurova, M., Avramides, K., Spikol, D., Luckin, R., & Mavrikis, M. (2016). An analysis framework for collaborative problem solving in practice-based learning activities: A mixed-method approach. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, ACM, New York, pp. 84–88.

Cukurova, M., Bennett, J., & Abrahams, I. (2018). Students' knowledge acquisition and ability to apply knowledge into different science contexts in two different independent learning settings. *Research in Science & Technological Education*, *36*(1), 17–34.

Cukurova, M., Luckin, R., Millan, E., & Mavrikis, M. (2018). The NISPI framework: Analysing collaborative problem-solving from students' physical interactions. *Computers & Education*, *116*, 93–109. https://doi.org/10.1016/j.compedu.2017.08.007.

Dewey, J. (1938). *Experience and education* ([New ed.]). New York: Simon & Schuster.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics author contact details. *Educational Technology & Society*, *15*(3), 42–57. https://doi.org/hdl.handle.net/1820/4506.

Grover, S., Bienkowski, M., Tamrakar, A., Siddiquie, B., Salter, D., & Divakaran, A. (2016). Multimodal analytics to study collaborative problem solving in pair programming. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge—LAK '16*, ACM Press, New York, New York, USA, pp. 516–517. https://doi.org/10.1145/2883851.2883877.

Halverson, E. R., & Sheridan, K. (2014). The maker movement in education. *Harvard Educational Review*, *84*(4), 495–504. https://doi.org/10.17763/haer.84.4.34j1g68140382063.

Katterfeldt, E.-S., Cuartielles, D., Spikol, D., & Ehrenberg, N. (2016). Talkoo: A new paradigm for physical computing at school. In *Proceedings of the 15th International Conference on Interaction Design and Children*, ACM, New York, NY, pp. 512–517. Retrieved from http://dl.acm.org/citation.cfm?doid=2930674.2935990.

Kirschner, P. A., & van Merriënboer, J. J. G. (2013). Do learners really know best? Urban legends in education. *Educational psychologist*, *48*(3), 169–183.

Kotsiantis, S., Patriarcheas, K., & Xenos, M. (2010). A combinational incremental ensemble of classifiers as a technique for predicting students performance in distance education. *Knowledge-Based Systems*, *23*(6), 529–535.

Krajcik, J. (2010). Project-based science: Engaging students in three-dimensional learning. *The Science Teacher*, *82*(1), 1–25.

Krajcik, J., & Blumenfeld, P. (2006). Project-based learning. In R. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 317–334). New York, USA: Cambridge University Press.

Luckin, R., Baines, E., Cukurova, M., & Holmes, W. (2017). Solved! Making the case for collaborative problem-solving. London, UK: NESTA Retrieved from http://www.nesta.org.uk/sites/default/files/solved-making-case-collaborativeproblem-solving.pdf.

Martinez-Maldonado, R., Schneider, B., Charleer, S., Shum, S. B., Klerkx, J., & Duval, E. (2016). Interactive surfaces and learning analytics: Data, orchestration aspects, pedagogical uses and challenges. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge—LAK '16*, ACM Press, New York, New York, USA, pp. 124–133. https://doi.org/10.1145/2883851.2883873.

Ochoa, X., Chiluiza, K., Méndez, G., Luzardo, G., Guamán, B., & Castells, J. (2013). Expertise estimation based on simple multimodal features. In *Proceedings of the 15th ACM International Conference on Multimodal Iteraction (ICMI '13)*, ACM Press, New York, pp. 583–590. https://doi.org/10.1145/2522848.2533789.

Ochoa, X., Worsley, M., Chiluiza, K., & Luz, S. (2014). Mla'14: Third multimodal learning analytics workshop and grand challenges. In *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14)*, ACM, New York, NY, USA, pp. 531–532.

Ochoa, X., & Worsley, M. (2016). Editorial: Augmenting learning analytics with multimodal sensory data. *Journal of Learning Analytics*, *3*(2), 213–219.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., …, & Duchesnay, É. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

Prieto, L. P., Sharma, K., Dillenbourg, P., & Jesús, M. (2016). Teaching analytics: Towards automatic extraction of orchestration graphs using wearable sensors. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge—LAK '16*, ACM Press, New York, USA, pp. 148–157. https://doi.org/10.1145/2883851.2883927.

Ruffaldi, E., Dabisias, G., Landolfi, L., & Spikol, D. (2016). Data collection and processing for a multimodal learning analytic system, IEEE, pp. 858–863. https://doi.org/10.1109/SAI.2016.7556079.

Schneider, B., & Blikstein, P. (2014). Unraveling students' interaction around a tangible interface using gesture recognition. *Journal of Educational Data Mining*, *7*(2), 320–323. Retrieved from http://blog.bertrandschneider.com/wpcontent/uploads/2012/01/23%7B%5C_%7D102-844-3-CE.pdf.

Siemens, G., & Baker, R. S. J. D. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge—LAK '12*, ACM Press, New York, USA, pp. 252. https://doi.org/10.1145/2330601.2330661.

Spikol, D., Ehrenberg, N., Cuartielles, D., & Zbick, J. (2015). Design strategies for developing a visual platform for physical computing with mobile tools for project documentation and reflection. In *Proceedings of the Workshops at the 17th International Conference on Artificial Intelligence in Education AIED*, pp. 57–62. Retrieved from http://ceur-ws.org/Vol-1432/iseole_pap8.pdf.

Spikol, D., Ruffaldi, E., Landolfi, E., & Cukurova, M. (2017). Estimation of success in collaborative learning based on multimodal learning analytics features. In *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, pp. 269–273.

Sridhar, S., Mueller, F., Zollhöfer, M., Casas, D., Oulasvirta, A., & Theobalt, C. (2016). Real-time joint tracking of a hand manipulating an object from RGB-D input. In *European Conference on Computer Vision*, Springer, Amsterdam, The Netherlands, pp. 294–310.

Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, *57*(2), 137–154.

Worsley, M. (2012). Multimodal learning analytics. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction—ICMI '12*, ACM Press, New York, New York, USA, pp. 353. https://doi.org/10.1109/ICALT.2017.122.

Worsley, M. (2014). Multimodal learning analytics as a tool for bridging learning theory and complex learning behaviors. In *3rd Multimodal Learning Analytics Workshop and Grand Challenges, MLA 2014*, ACM, New York, NY, USA, pp. 1–4. https://doi.org/10.1145/2666633.2666634.

Worsley, M., Blikstein, P. (2013). Towards the development of multimodal action based assessment. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK '13* (p. 94). New York, New York, USA: ACM Press. https://doi.org/10.1145/2460296.2460315.

Worsley, M., & Blikstein, P. (2014). Analyzing engineering design through the lens of computation. *Journal of Learning Analytics*, *1*(2), 151–186.

Zbick, J., Vogel, B., Spikol, D., Jansen, M., & Milrad, M. (2016). Toward an adaptive and adaptable architecture to support ubiquitous learning activities. In A. Pena-Ayala (Ed.), *Mobile, ubiquitous, and pervasive learning* (pp. 193–222), Advances in Intelligent Systems and Computing, Vol. *406*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-26518-6_8.

How to cite this article: Spikol D, Ruffaldi E, Dabisias G, Cukurova M. Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. *Journal of Computer Assisted Learning*. 2018;1–12. https://doi.org/10.1111/jcal.12263