Predictive learning analytics using deep learning model in MOOCs' courses videos

Ahmed Ali Mubarak[1,2] · Han Cao[1] · Salah A.M. Ahmed[3]

Abstract

Analysis of learning behavior of MOOC enthusiasts has become a posed challenge in the Learning Analytics field, which is especially related to video lecture data, since most learners watch the same online lecture videos. It helps to conduct a comprehensive analysis of such behaviors and explore various learning patterns for learners and predict their performance by MOOC courses video. This paper exploits a temporal sequential classification problem by analyzing video clickstream data and predict learner performance, which is a vital decision-making problem, by addressing their issues and improving the educational process. This paper employs a deep neural network (LSTM) on a set of implicit features extracted from video clickstreams data to predict learners' weekly performance and enable instructors to set measures for timely intervention. Results show that accuracy rate of the proposed model is 82%–93% throughout course weeks. The proposed LSTM model outperforms baseline ANNs, Super Vector Machine (SVM) and Logistic Regression by an accuracy of 93% in real used courses' datasets.

✉ Han Cao
caohan@snnu.edu.cn

Ahmed Ali Mubarak
ahmedmubarak@ibbuniv.edu.ye

Salah A.M. Ahmed
salah.a@snnu.edu.cn

[1] School of Computer Science, Shaanxi Normal University, Xi'an 710119, China

[2] Faculty of Education-Alnadra, Department of Computer, Ibb University, Ibb, Yemen

[3] School of Education, Shaanxi Normal University, Xi'an 710062, China

1 Introduction

Recently, technological advances have had a significant impact on the pedagogical process to develop e-learning methods. One of the outstanding innovations which aroused interest in online education is Massive Open Online Courses (MOOC)(Brinton et al. 2015). MOOC is a platform that provides various courses over the Internet, which allows learning to a very large number of learners from anywhere and anytime such as Coursera, Edx, and Udacity, which present courses essentially based on video lessons, forums, and quizzes (Yu et al. 2019). Video tutorials are a primary part of MOOC course design and an incredibly important aspect of it. They serve as the principal axis to draw students to the course. Learning platforms store web log data that includes basic learner information and their experiences with course material (e.g., clickstream videos, forum posts, and quizzes). Clickstream events (e.g. play, search, pause, etc.) are recorded for video along with the time they occurred at which are then processed. Many students spend much of their time watching video lectures. Therefore, many questions regarding the engagement of learners with video gradually emerged such as "Do learners view the whole video lecture? What portions of the video do learners repeat to view, and why?; What video section are more engaging or difficult?; Is there any relationship between learners' viewing and learning performance?; How can learners' performance be evaluated through their video-clickstreams?.(Giannakos et al. 2015).

Educational video analytics concentrate on gathering, analyzing, and describing data related to learners' interaction with videos. This can offer an opportunity for researchers and educators to understand how learners' interaction with video content in the learning field and their behavior analysis. A variety of studies on learners' participation and investigating behavior patterns in learner interaction with video have focused on the big data obtained from the relationship of learners with the different ways of the course. (Kizilcec et al. 2013; Wang and Baker 2015). Some other studies have investigated the clickstream interactions within videos. According to Halawa et al. (2014) and Sinha et al. (2014a, 2014b), leaving videos (i.e., the video exists without being watched) results in significant likely dropout from MOOC courses. Video interaction and engagement terms interchangeably have been utilized in related studies to refer to activities associated with clickstreams, views, and in-video quizzes, etc. Other related studies about video learning analytics have investigated learners' performance through video engagement behavior using explicit features such as views and annotations (De Barba et al. 2016; Hsin and Cigas 2013; Risko et al. 2012). For recent research, Guo et al. (2014) and (Zhang et al. 2006) pointed out that learners are more viable to engage when the contexts of video: 1) are short; 2) video building form (slides, code, khan academy style, etc.); and 3) setup interactive activities like in-video quizzes. Despite these academic efforts, yet, there are dearth research investigating several complex aspects of click-stream interactions emerging from students' actions within different MOOC videos (Guo et al. 2014; Sinha et al. 2014a, 2014b. Also, in-video, dropout is one of the outstanding challenging problems regardless of video design considerations. Furthermore, MOOC learners are expected to reduce their video engagement as they proceed during the course (Halawa et al. 2014). These challenges present important opportunities for learning analytics to study video clickstream data (e.g., when playing videos, pause, forward-back seek, re-playing the video, re-changing speed) and the influence on video contents under (i.e., verbal and visual features).

Regardless of the different data analytic methods, Deep Artificial Neural Networks (ANNs) are employed in the prediction of different measures (Coelho and Silveira 2017). Deep Learning architecture consists of multifarious layers of neural networks interlinked through neurons. The Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) are various kinds of ANNs and associated with the sequential data. Several studies are based upon considering MOOCs' data being time series on the prediction of dropouts. They exploit the power of deep learning models in the learning analytics field to predict at-risk students. Coelho and Silveira 2017 studied a literature review related to the studies learning analytics with deep learning techniques. They tracked some important areas including learners' performance (Baker and Inventado 2014; Corrigan and Smeaton 2017; Palmer 2013) and their learning evaluation. Hence, deep learning methods exceeded the conventional machine learning and other baseline methods.

Deep learning utilizes methods that comprise developing a model encompassing many layers to learn from raw data. Every layer converts the representation to a more abstract pattern for the following layer (Lecun et al. 2015). Corrigan and Smeaton (2017) predicted students' performance in online learning. They proposed Recurrent Neural Networks (RNN) variation Long Short-Term Memory (LSTM) to analyze students' interaction in the online learning environment and predict their performance. LSTM outperformed Random Forests by 13.3% of the variance of the model corresponding to 8.1%.

In addition, Limited studies discussed implicit events in video learning. For instance, researchers like Giannakos, Guo, Kim, Li, and their collaborators explored the correlation among video interaction patterns and the difficulty of perceived video and affirmed that patterns such as repeating/skips, repeating/re-watching and pause repetition mean the difficulty of the video(Giannakos et al. 2015; Guo et al. 2014; Kim et al. 2014; Li et al. 2015a, 2015b).

The present study analyzes clickstream behaviors of learners during video viewing as implicit features. Instead of using explicit features that express the learners' engagement behaviors (views and annotations), we employ extracted implicit features from behaviors of learners with click-video to create their video viewing profile. To be precise, our focus is on predicting learners' performance by video-clickstream. The following question states clearly what our focus is which is" How is learners' performance correlated to their behaviors in MOOC course videos?"

To answer that question, deep learning models (Hochreiter and Schmidhuber 1997; Sak et al. n.d.; Tang et al. 2016) are employed to process the dataset as a time series sequence by exploiting extracted implicit features as features of the model, which inherently serve in the compilation of interactional learners' clickstream for each video in a weekly format to predict learners' performance for timely intervention by the instructors. Moreover, developing such an approach in the form of extended learning and content analytics would have implications not only for theories about how information is processed by humans, but also systems striving to enhance learning outcomes for students and mitigate the learning burden.

The main contributions of this study are summarized in the following points:

- Extracting implicit features generated by video-clickstream data through learners' interaction from main events (play, pause, and seek).

- Ascertaining the effectiveness of a Long short-term Memory (LSTM) using the extracted features for course videos to predicting weekly learners' performance.
- Extending the effectiveness of the LSTM model in predicting learners' performance compared to baseline Artificial neural networks (ANNs), a support vector machine (SVM) and Logistic Regression models.

The contribution is derived from the above research objectives in order to enable MOOCs' instructors and learning experts to have insight visibility on learners' behaviors in video-clickstream. Thus, they can make rational and timely interventions according to weekly predictions.

The organization of the paper is as follows: Existing related works to study MOOCs' clickstream and predict learners' performance in MOOC courses are described in Section 2. Section 3 presents our proposed model processes. In section 4, dataset is described. The experiment setup is shown in Section 5. Section 6 presents the experimental results. Section 7 provides the discussion of result implications. In the last section, conclusions are made.

2 Related work

Videos play a prominent role as the most instructive element in MOOC courses. In the beginnings of MOOC growth, related researches primarily focused on the quality of MOOC videos such as the length of the video and presentation (Guo et al. 2014; D. Zhang et al. 2006). Recently, researches on online learning analytics centered on creating a predictive model to predict dropout rates and students' performance by examining their participation in MOOC course video events (De Barba et al. 2016; Hsin and Cigas 2013). The recent research on video learning analytics has included three general ways: Firstly, visual content analysis studied visible pattern recognition, video indexing, and transcript analysis (Risko et al. 2012; Smoliar and Zhang 1994). Secondly, the analytics of explicit users' data or explicit video analytics provide inferences on video usage degree and popularity. This may affect videos learning outcomes. Examples of these explicit user data include clicking buttons, views, and votes (De Barba et al. 2016; Hsin and Cigas 2013). On the other hand, such superficial analytics lack in-depth video interaction analysis to discover learners behavior when they are interested, confused or misunderstood. Thirdly, in the same manner, the analytics of implicit users' data or implicit video analytics can be used in depth-video interaction to identify what kind of difficulties may encounter learners through video watching and to what degree they indicate the perceived video difficulty. Examples of these implicit users' data include interaction records in-depth videos (Chorianopoulos 2013; Mubarak et al. 2020; Shaw and Davis 2005; Yew et al. 2011). It can be used as a good indicator to predict the dropout of learners and their performance. Early studies on implicit video analytics focused certain reactions such play, pause, and seek events. Li et al. (2015a, 2015b) studied the relationships between video interaction patterns and some related aspects such as the perceived video difficulty, video re-watching behaviors, and learners' performance. They classified the in-video interactions into many classes such as infrequent/more skips, pause frequency, frequent pauses, and infrequent/large amount of re-watching as implicit

video learning analytics. They used statistical inferences in their study without considering the variations in video content or the nature of the course. Giannakos et al. (2015) improved an open-access video education analytics system to gather clickstream events of video interactions. They studied the relationship between video navigation (e.g., play, pause, skip, and stop) and the level of perception /thinking for a particular video segment. The authors concluded that the peaks of video interactions (e.g., repeated views) corresponded to answers of assessments. Kim et al. (2014) investigated large-scale analysis of in-video dropout and peaks of learners' activity in video watching, by using data from four courses received in edX platform. According to the measure of the authors, the peaks in sessions of play events and rewatching indicated states of interest or difficulty. They focused on the reasons why peaks happened, and they found that 61% of peaks occurred as a result of visual transitions in the video such as starting from the beginning of new material, returning to missed content, following a tutorial step, replaying a brief segment, and repeating a non-visual explanation. Sinha et al. (2014a, 2014b) constructed a quantitative Information Processing Index (IPI) h based on cognitively plausible behaviors by video clickstreams of learners. That could serve instructors' interventions for learners in real-time. They classified learners' clicks of video watching state sequence (e.g., play, pause, seek forward and seek backward) into higher-level behaviors such as long-time watching (e.g., playing and pausing) and re-watching (e.g., actions of seeking backward). Based on IPI, they also investigated whether the learner would continue till the end of the video and course and found that the learners' dropout in the MOOC course was 37% less likely if they had one standard deviation higher IPI than average. Brinton et al. (2015) presented two frameworks for studying the relationship between learners' video-watching behavior and performance in MOOC courses. The first framework was on groups of created events and the visited positions. According to research, they extracted the sequences to distinguish recurring motifs in learners' behavior and found that some of these features were significantly correlated with Correct on First Attempt (CFA) and non-CFA quiz submissions. Thus, a model was presented based on learner clickstreams of experience gained, and showed multi-aspects of prediction quality on individual videos and improved CFA outcome, which emphasizes the ability to associate behavior to performance. Li et al. (2015a, 2015b) screened out research on how video interactions reflected the perceived difficulty of the video for the learner. The authors extracted simple features of video interaction that determined higher-level of video difficulty. (Li et al. 2015a, 2015b)The Mixed Model analysis to show the variations in personal video difficulty for a learner from video to video was used. The authors found that some of the features such as frequent and long pauses, speed decreases, infrequent seeks were with high amount of skipping and re-watching which suggested higher-level of video difficulty. Atapattu and Falkner (2018) investigated the relationship between learner video inter-actions and discourse features patterns of video transcripts. They focused on language and discourse features of MOOC video content. According to them, discourse features (e.g., lexical diversity, causal connectives, and narrativity) promoted video discourse processing while long sentences, video length, and high speaking rate provided difficulties with video discourse processing. They found meaningful correlations between discourse features and video interactions. This correlation was shown by events (i.e., joined events except for stop/load and specific events like seek and pause).

In Yu et al. (2019), the authors used a 4-g approach to study the feature sequence of viewing learning behavior without caring about learners' cognitive participation on the clickstream data of MOOCs videos. They used machine learning methods to predict students' outcome by the click-stream data of MOOCs videos such as K-Nearest Neighbor Classification (KNN), which was commonly used to classify data. K represents a constant, and KNN takes the K points of the nearest distance to define to which class the object belongs. Support Vector Machines (SVM), as a supervised learning model which is often employed for classification, pattern recognition, and regression analysis (Xu and Yang 2016), and Artificial Neural Network (ANN) (Fauvel and Yu 2016), which included many neuron nodes, can be distributed into multilayers such as an input layer, hidden layers and an output layer. In addition to numerous data analysis methods, Deep Artificial Neural Networks (ANNs) are primarily used for the estimation of specific measures and attributes (Coelho and Silveira 2017). Deep learning, which consists of multiple types of neural networks and containing many computational layers, enables the paradigm to learn from current instances, overcoming traditional hand-functional computing practices (Q. Zhang et al. 2018). Substantial evidence of Deep Learning methods in the area of learning analytics can scarcely be seen in literature to forecast the performance of the students. Recently, LeCun & Silveira (2017) undertook a systematic literature review to analyze the research co-related with deep learning techniques. They mapped some important fields, including (Baker and Inventado 2014; Palmer 2013) student performance with learning assessment and handwriting recognition(Gross et al. 2015), where deep learning strategies overshadowed traditional machine learning and other benchmark approaches. The Long Short-Term Memory (LSTM) is one Deep learning method which is a hierarchal representational learning method, which consists of several non-linear layers that contribute to learning the representations from the input data. It is also a new technique in the testing context of learning analytics, by interpreting the data created as a time series problem by calculating learners' log data. Yet a limited number of studies applied these approaches for early prediction of dropout (Fei and Yeung 2015; Waheed et al. 2019).

Previous researches focused on learners' interaction with course videos and innovated the techniques and methods to predict learners' performance accurately. They emphasized that weak learners' performance was often attributed to factors related to their educational background and social information for learners plus extracted features selected from video-clickstream data. However, MOOC platforms still face challenges of high dropout rates and difficulty of early prediction for learners' performance. Moreover, there is no specific method to represent the best solution to predict students' behavior or their performance in MOOCs courses and identify students yet at risk. Therefore,, this study aimed to fill in the gap in this research area to enhance the efficiency of MOOC education.

3 Methodology

Making a learner continue in an online learning environment is a significant challenge for research community with only numerous studies concentrating on different directions of learner retention and attrition as explained in the previous section. Recently,

researchers have laid focuses specifically on predicting at the end of the term or after a particular course through their legacy data, in contrast to predicting learners' performance through on-going courses. However, a research gap remains as final studies contribute to the early prediction of the performance of learners, which could ultimately help institutes to formulate early intervention strategies. This study aims to analyze the impact of learners' online participation on their performance and implementation techniques for early prediction of learners dropout.

The study presented in this paper is divided into three major phases as shown in Fig.1. The following subsections illustrate data processing procedure to ensure data integrity from missing values and nested data due to anomalies in some columns or rows in the first section. Features extraction of video-clickstream resulting from learners' interaction is explained in the second section along with including the transformation of the raw data into a sequential time-series format for early performance prediction. In the third section, we present an overview of the deep learning method (LSTM) discussion with the significance of the model for time-series data prediction.

3.1 Data cleaning

The dataset used is quite large and contains missing multiple values and interfering data within a tuple. To guarantee the accuracy of the model's results, it is necessary to consider these exceptions before analyzing and classification the data. Hence, data cleaning steps are required to ensure data integrity and validity when analyzing. The first way is to remove the records without a valid screen name or video code; anonymized name pre-defined are mapped for the learner as a unique id. The next step is to remove columns that add noise to the data (e.g., course name and resource name) and empty columns. The following step is extracting meaningful features from video's clickstreams.
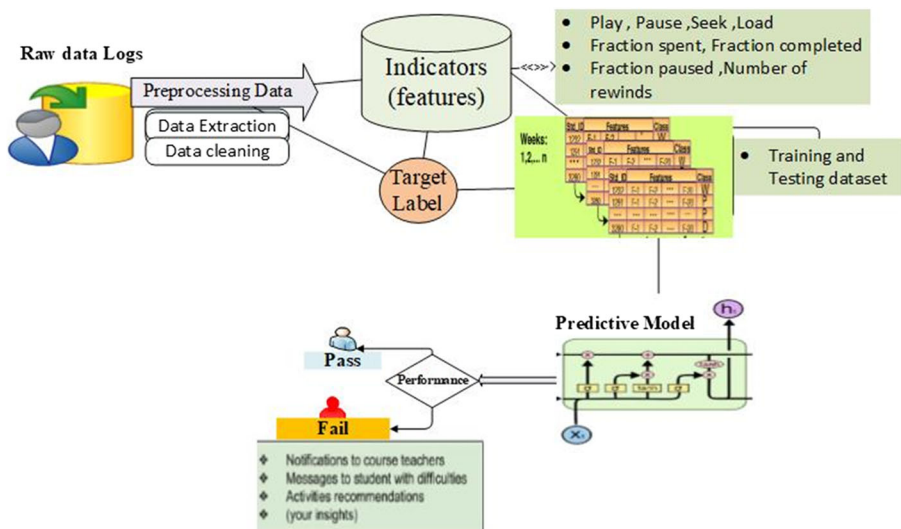

Fig. 1 Proposed Architecture of Methodology

3.2 Feature extraction

In this section, we discuss extracting implicit features from an error-free video-watching clickstream data resulting from the cleaning process, which were logged during learners' interaction with video player by following the main events: play, pause, rate-change, seek forward or backward and stop. Each time any of these events is fired, these events are logged with identification of the learners' IDs and video IDs, event type, event current time related to video time, and UNIX timestamp for the event. In addition to these events recorded as explicit features, we extracted implicit features through analyzing event current time which is relative to video time or timestamp for eventas follows:

	Feature	Description
1	Fraction completed (fracComp)	The percentage of the video which the learner watched, not counting repeated segment intervals. The completed fraction is a tentative measure of how closely learners are aligned with a video. So, values have to be like [0,1]. $fracComp = \frac{\sum current\ time\ watched}{real\ time\ of\ video}$
2	Fraction spent (fracSpent):	The amount of (real) time the learner spent watching the video (i.e. when playing or pausing), compared to the video's actual duration. FracSpent often reflects the amount of time that it took for the learner to grasp the given content. It can take more than 1 value. And is thus a reflection of the consistency and complexity of the delivery of the material. The high value fracSpent may mean that the l.
3	Fraction played (fracPL)	The cumulative amount of play time that the learner was watching, divided by a video's real duration. Playback periods of under 5 s have been deleted. The fracPL measures repetitive fragments that are watched as opposed to fracComp, which can thus take value greater than 1.
4	Number of played (NumPL):	The number of times that video was played by the learner. This function may indicate the learner's expended efforts to internalize the material further. An abnormally high-value NumPL can therefore suggest that the video quality is unclear, or that the quality is incredibly difficult.
5	Number of pauses (NumPa):	The amount of times the learner has tapped on the pause case in the video. Similar to NumPL, a high value of this attribute suggests the extra effort has been made by the learner to grasp the video material or to perform an off-task conduct.
6	Fraction paused (fracPa):	real-time fraction the learner spent pausing on the video, divided on its total playback time. It is possible for fracPa to get up value >1. This feature could take larger values, so that it indicates that the learner exhibited extra effort to understand the material. As can be seen in Fig. 2, the dashed lines represent the actual time spent by the learner after clicking on pause event on the video.
7	Fraction Forward (Fracfrwd)	The sum of the video the learner had to skip forward while the video was running or pausing; divided on its duration, with repetition. It needs values from 0 to 1
8	Fraction Backward (FracBkwd)	The sum of time the learner had to skip back when the time was playing or pausing, divided by its duration, with repetition. It takes values from 0 to 1 This feature can take values greater than 1 and may mean that the learner has expended extra time internalizing the video passages. It also indicates that the subject is more interested in or confided in these passages

	Feature	Description
9	Number of seek Backward (NumSBW)	The sum of points in the video the learner was skipping backward. A higher value of this feature indicates the video is interesting or difficult in content.
10	Number of fast forwards (NumFFW)	The number of times in the video the learner has hopped forward. In this event a high value indicates that the learner is aware that the content of this video clip is not important.
11	Average Change Rate (AvChR)	Time-average change of playback Rate selected in the play state by the learner. The Stanford MOOC player allows the default speed rates of between 0.50x and 2.0x. Analysis of average change rate may show that learners viewed the video content quickly or slowly.
12	Number of load (numLD)	The learner has loaded the video the number of times. A value in this feature indicates the learner can watch the video without having to associate it with the content of the course.
13	Number of complete(numCompt)	The sum of times a learner viewed a video in its entirety. This feature must have a value less than or equal to the number of video playbacks. It also indicates a learner is interested in viewing the whole video.

We primarily focused on implicit features about video-clickstream behavior that were extracted from the raw dataset. After extracting the video-clickstream information, learners' records were maintained in a week-wise fashion. Data of each week were appended with the video-clickstream behavior of the previous week. The transformed function to form a vector of extracted features for each video was used as time-series and fed to the machine learning algorithm.

3.3 Predictive Model (LSTM)

Deep learning architecture includes numerous nonlinear hidden layers between input and output layers interlinked through neurons. These layers are self-adaptive in inferring underlying functional associations from the input data (G. Zhang et al. 1998). They help the system to learn complex functions and features for better predictions (Graupe 2016), Contrary to the statistical and model-based methods.

Long Short-Term Memory networks, which are usually called "LSTMs",are a particular kind of Deep Artificial Neural Networks which are capable of learning long-term dependencies in time series data(Chakraborty et al. 2017). LSTM consists of a recursive loop that helps the model to consider the prior inputs along with the current one. But the repeating module has a different structure. Instead of having a single neural network layer, there are three gates namely input, output and forget gate. A fourth unit which is the memory cell is also pointed to as Constant Error Carousel. The parameters in LSTM are learned by backpropagation through Time (BPTT) (Hochreiter and Schmidhuber 1997; Sak et al. 2014).
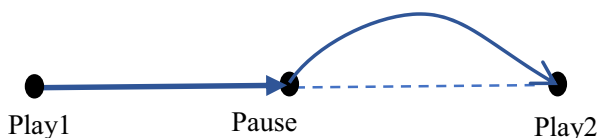

Fig. 2 illustration of fraction paused state.

The key to LSTMs is the memory cell, which controls the data to be written in the cell by running straight the entire chain, with some minor linear interactions from three gates. At some time interval t, the input gate i_t manages the data to be written in the cell, forget gate f_t determines the data to be ignored. A memory cell C_t controls the data to be written in the cell and output gate O_t controls the cell data to be drawn as output (Hochreiter and Schmidhuber 1997; Olah 2015). Mathematically, the gates are formulated as:

$$f_t = \sigma\left(W_f.[h_{t-1}, x_t] + b_f\right) \tag{1}$$

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{2}$$

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \tag{3}$$

At time interval t, the current cell state C_t is updated by multiplying the forget gate output with the previous cell state C_{t-1} and adding the previous hidden state defined by the input gate according to the equation formula 4.

$$C_t = f_t.C_{t-1} + i_t.\tanh(W_c.[h_{t-1}, x_t] + b_c) \tag{4}$$

The current cell state passes through a hyperbolic tangent nonlinearity limited by the output gate to produce the current hidden state as:

$$h_t = o_t.\tanh(C_t) \tag{5}$$

Where f_t; i_t and o_t represent the outputs of the forget, input and output gates respectively, σ represents the activation function, W_f, W_i, W_o, W_c, b_f, b_i, b_o, b_c represent weights and biases. One advantage of the powerful representation of LSTM is its ability to deal with information in the previous cells and a prediction based on each of the current and previous input (Okubo et al. 2017). Figure 3. illustrates visualization and in-depth understanding of the dependency of sequential data in the unrolling of LSTM cells. X represents the input data from the first layer, t timestamp and hidden values to hidden layers denoted by h, and C represents the memory cell. The LSTM model is deployed to analyze learners' performance based upon their interaction with video clickstream data. The clickstream data of the videos are stacked one on the other in each week to predict learners' performance in a particular week or course. Thus, the weekly data-stack of learners is fed to the model layers to process the data for each learner.

The input data of our model is a sequence of fixed length input vectors x_t. The features for X_t are features related to how the learners are engaged with video clickstream features (mentioned in extracted features section), and the final two features are related to the weekly quiz, which is taken as the probability of true values. The first 'attempt' indicates whether the quiz was attempted, and the other one is the weekly
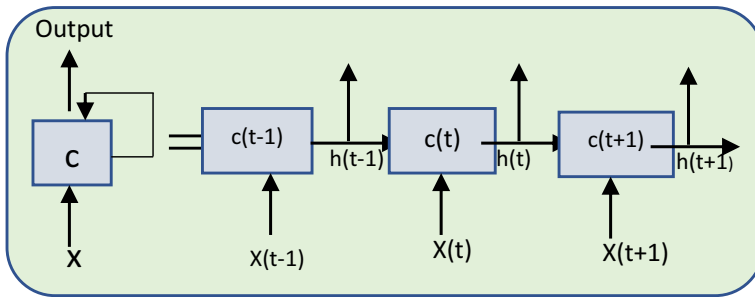
Fig. 3 Data in Unrolled LSTM Cells

quizzes where each quiz i covered the material in videos $(1, \ldots, m)$ to define a performance measure at "time" t. Fig. 4. demonstrates the overall structure of a MOOC with videos and a weekly quiz. The course is presented as a sequence of videos across sequential weeks with an in-week quiz. Thus, in our task of prediction, we take a week as a time step (t) and the output labels are tagged based on the input features of that week. According to this assumption, we obtain an input sequence $(x_1, x_2, \ldots, x_t)$ from learners' weekly learning interaction with videos in the course, and the corresponding output sequence labels $(y_1, y_2, \ldots, y_t)$, x_i refer to the input feature vector of week   , and y_i indicates the corresponding output label.

The model is designed in such a way that it adapts a variable window length automatically, which helps it to fit the RNN layer dimensions with input data. The architecture of the model is a recurrent layer with 128 hidden units, and *tanh* activation is used in each network that we train and one dense layer with a sigmoid activation as an output layer to predict learners' performance. Dropout layer in the last hidden layer h_t was used when computing the readout y_t to prevent overfitting during training. Since the model performs a process of binary classification (passing, failing), we use binary cross-entropy function with $L2$ regularization to render cost function. The $L2$ performs smoothed oscillations in the training loss. Besides, the Adam optimization algorithm (Kingma and Ba 2014) to adjust parameter updates during training was introduced in our model. To prevent the validation loss from failure during training, we used a variable learning rate that was automatically reduced during training. The value of dropout is 0.3.

$$\mathcal{L} = \sum_{j=1}^{n} \left(y\log\left(\hat{y}\right) + (1-y)\log\left(1-\hat{y}\right) \right) + \frac{\lambda}{2n} \sum_{i=1}^{i} \left\| w^i \right\|^2 \qquad (6)$$

4 Data description

The dataset used for this research was selected from two different MOOC courses designed and launched by the University of Stanford. Each of these two courses teaches
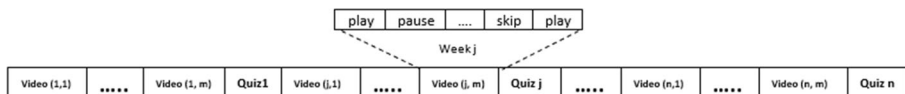

Fig. 4 general sequence of course videos and quizzes in a MOOC.

Computer Science topics, and each course has been taught multiple times. The first course is "Mining Massive datasets (SELF PACED and Fall 2016), and the second one is "Automata Theory" (SELF PACED and Fall 2016). The data was collected by the Center for Advanced Research Through Online Learning (CAROL)(Stanford 2017) at the University of Stanford. Data have been anonymized for the privacy protection of students' personal information. Dataset details shared by CAROL such as protocols to access data and table schemas are available online (University, S 2017). Table schema for each course includes three tables whichare eventsExtract, videoInteraction, and activityGrade. They contain learners' interaction logs and assessment grades. In this study, video interaction and activity grade tables are considered. In the video interaction table, each log entry contains temporal details about learners' interaction with video events such as clickstream events (load, play, pause, speed change, and so on.), learner/video identification information and the course. In activity grade table, the details for homework assessment grades are recorded along with its timestamp and answer selected. Table 1 shows a summary of the basic information on the data of the two courses after preprocessing.

Tables 2 and 3 describes the content of datasets such as videos, students and clickstream events for each week of the course. Each week is in turn comprised of a set of videos and quizzes. 'MMDS' was a longer course, with 15 Modules on 7 weeks, whereas 'Automata Theory' had only 6. 'MMDS' had shorter videos in length with a total of 94 videos and the average (avg.) length was 10 min for most of the videos. Therefore, 'Automata Theory' had longer videos with a total of 26, and the average length for most of the videos was longer than 20 min. For the two courses, each week included one or two quizzes which were associated with videos of the week. Testing students' understanding was made by linking video concepts in the selected week and the material throughout the course. Each quiz consisted of multiple-choice questions, in radio-response format, with 4–5 possible answer choices and 5 tries for homework. Besides, there were summative assessments in the form of final exam to test learners' comprehension of the material throughout the course. The exam was taken in a three-hour period, and each question was submitted only once.

Learners interact with video player by clicking key events: play, pause, rate-change, seek forward or backward and stop. Each time one of these events is fired, it is stored in system logs. Table 4 shows details of the number and means for each click-event on videos for the four courses.

Table 1 Basic Information on four datasets

Name of course	Time frame	Modules Lectures	Videos	Students	Clickstream	Quizzes
MMDS SELF PACED "Mining of Massive Datasets"	7 weeks	15	94	5346	1,317,344	15 + final
MMDS+Fall2016 "Mining of Massive Datasets"	7 weeks	15	94	2135	503,697	15 + final
Automata SELF PACED "Automata Theory"	6 weeks	6	22	3022	297,748	6 + final
Automata+Fall2016	6 weeks	6	22	2497	377,958	6 + final

Table 2 Basic information on MMDS course weeks

Name of course weeks	MMDS SELF PACED			MMDS+Fall2016		
	videos	student	Processed Events	videos	student	Processed Events
1	11	4578	258,899	11	1993	148,583
2	12	1446	170,913	12	547	110,319
3	17	747	695,839	17	374	112,016
4	17	881	87,532	17	314	47,024
5	9	579	34,536	9	238	26,763
6	15	565	52,192	15	231	25,347
7	13	313	17,433	13	172	33,645

5 Experiments setup

To show our model's feasibility and efficiency, we performed a case study on two MOOC courses. As stated in Section 3, we first conducted preprocessing of data which included extracting features from video clickstream data. In the second phase, the data set resulting from the extraction process of features was transformed into suitable input data for the model, and padded vectors were constructed with a consistent shape, which was then masked before being fed to the model layers. Each course dataset was split into 60/30/10 training, testing and validation respectively. Several experiments were implemented with the number of epochs, the number of recurrent layers, the number of hidden units, and the learning rate parameter. During training, we tried to tune many hyperparameters on the model before finally settling on model architecture, with 128 hidden units and two layers which was a better balance of speed and accuracy than 64 or 256 hidden units. Besides, the evaluation metric was accurate during training. The parameters associated with Adam optimizer remained constant throughout the training period. We used the Scikit-Learn (Pedregosa et al. 2011), Keras(Clow 2013), and TensorFlow libraries developed with the Python programming language.

Table 3 Basic information on Automata Theory course weeks

Name of course weeks	Automata Theory SELF PACED			Automata Theory +Fall2016		
	videos	student	Processed Events	videos	student	Processed Events
1	5	2784	155,439	5	2260	162,138
2	5	716	54,608	5	943	94,021
3	3	498	29,707	3	325	29,152
4	5	368	21,589	5	407	39,279
5	4	362	26,780	4	300	29,646
6	4	199	9625	4	335	23,722

Table 4 Information on video click events

Event name	MMDS SELF PACED		MMDS+Fall2016		Automata SELFPACED		Automata +Fall2016	
	counts	means	counts	means	counts	means	counts	means
Play	234,346	0.178	259,535	0.515	108,975	0.366	173,012	0.458
Pause	194,159	0.147	95,950	0.19	92,835	0.312	94,740	0.251
Seek	139,807	0.106	65,386	0.13	61,302	0.206	59,324	0.157
Rate-Change	6628	0.005	3874	0.008	3997	0.013	3786	0.01
Load	69,626	0.053	37,570	0.075	23,900	0.08	22,073	0.058
Stop	672,778	0.511	41,382	0.082	6739	0.023	25,023	0.066

6 Experiments results

In this section, we explore the effectiveness and results of our proposal to study learners' video-watching behavior and illustrate likely relationships between students' behavior and their performance.

6.1 Experimental setup

To evaluate the effectiveness of the proposed model, the model was trained for each course on weekly basis. During the training process at each week *ith*, its previous weeks were appended in a homogenous single vector and fed to model layers. We applied padded on input data to obtain a vector of equal length, which was then masked before feeding it to the model layers. The padded values to an input vector were discarded by the masking layer, which only kept clickstream data as the sequence. The batch size was 50, and the larger batch of instances for each training step did not perform better. Due to the nature of the study problem which was to evaluate the proposed model for early prediction of the learner's performance, the problem was converted to a binary classification task 'pass' and 'fail' class. As it was also implemented on each complete course. We focused on learners who watched at least 20% of the videos of the week to ensure that we had sufficient data to train and test an effective model. The results in Figs. (5 and 6) presented the efficacy of the proposed model by the accuracy and loss values during training and validation of the model achieved at 50th epoch, concerning weeks for both datasets. These parameters were referred to during training to investigate the model performance. Loss value determined how well the LSTM model performance was after each iteration of optimization (One expects the decrease of loss after each, or several iterations).

We observed convergence of accuracy of the model during training which confirmed that a good balancing between validation and training model was achieved. It can be seen in (a, b) in Figs. (5 and 6) that the accuracy gradually progresses from the first weeks till the final weeks in both courses. This is due to the lack of sufficient and decisive data in the initial weeks. It can be concluded that dropouts' behavior can be defined more clearly during the final weeks when the dropout instances were less active
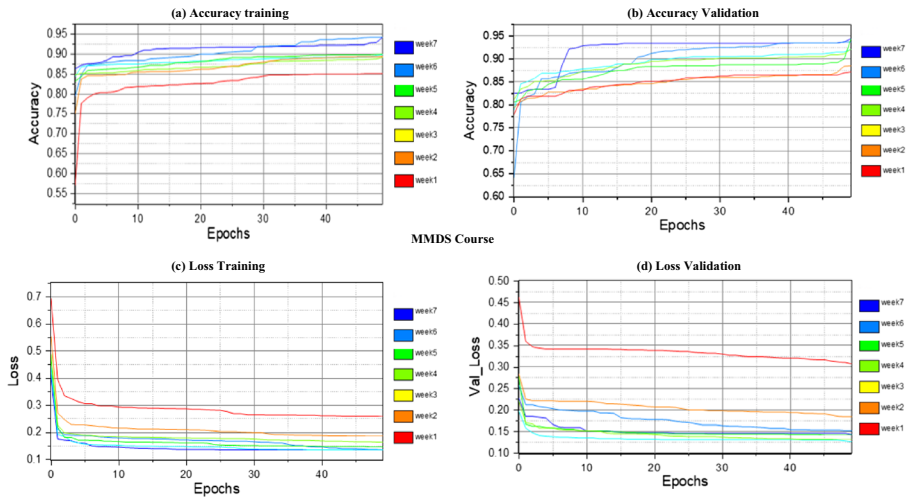
Fig. 5 Accuracy and loss Curves for training and validation for MMDS course

in the last weeks. On the other hand, the input pattern of the previous instance in week 1 isn't available.

We also see in Figs. 5 and 6, parts (c, d) describe the depletion of the model in loss values. Ideally, the loss values should be degraded showing the disparity between the ground truth and the class label that the model predicted. We notice, in Figs. (5 and 6), the progress for accurate prediction of input instances cross weeks, for the model achieves predictions up to 80% in the first week, with accumulated video clickstream of each student from previous weeks, and the accuracy scales are up to 93% in the final week. It can be inferred that the state of learner behavior is more defined in the final weeks indicating the efficiency of the model in learning the distinct pattern till the last weeks.
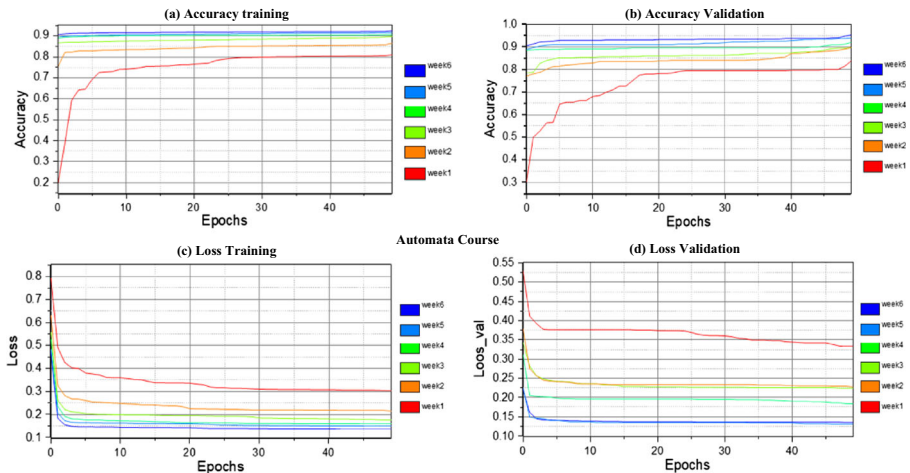

Fig. 6 Accuracy and loss Curves for training and validation for Automata course

AUC metric is considered to evaluate model performance on testing data throughout different weeks as shown in Fig. 7, which exhibits curves of the AUC metric. The (ROC) curve is a graphical plot employ to illustrate binary classifier diagnostic capabilities. A ROC curve plots the true positive rate (TPR) versus the false positive rate (FPR). In the current study which focuses on learners' performance (dropout prediction), TPR represents the proportion of learners' actual dropout that the classifier predicts as likely in dropout state, mathematically:(TP/ (TP + FN)). The FPR is the proportion of learners' actual non-drop and the classifier's prediction as dropouts (FP/ (TN + FP)) (Hanley and McNeil 1982). As revealed in Fig. 7, the ROC curve of the model is relatively more in a final week than in the first weeks and has a maximum area under the curve (AUC) of 0.85 at both courses. On the contrary, in the first-week, the model gained a minimum AUC of 0.73 in "MMDS" course and 0.75 at "Automata" course. One potential explanation of this is that in these weeks, learners interacted in video clickstreams more than final weeks as well referred to in Table 3 in the dataset section. The AUC gradually progresses through the consecutive weeks. This confirms why employing LSTM handling more extended memory has an important role in massive data.

In order to interpret the benefit of the LSTM model, we focused on learner performance prediction, since the target label was binary (1: pass, 0: fail, i.e., the number of 0 s exceeds the number of 1 s) called class imbalance. Therefore, other metrics to evaluate the goodness of the LSTM model predictions with the baseline models were employed to address imbalanced data problems as we will explain in the next section.

6.2 Evaluation of the model with the baseline

For the sake of evaluating the effectiveness of the LSTM model in identifying learners' performance and early dropout, the Deep ANN, SVM and LR models, as baseline models, was employed with the default hyperparameters. These models were selected because of their frequent selection by research community as baseline models (Marbouti et al. 2016). Table 5 provides an overview of metrics statistics values obtained from the baseline models and LSTM model for both courses where the results confirmed supremacy and effectiveness of the LSTM in the prediction of learners' performance. Overall, accuracy of the LSTM model is 93.3% in both courses in final
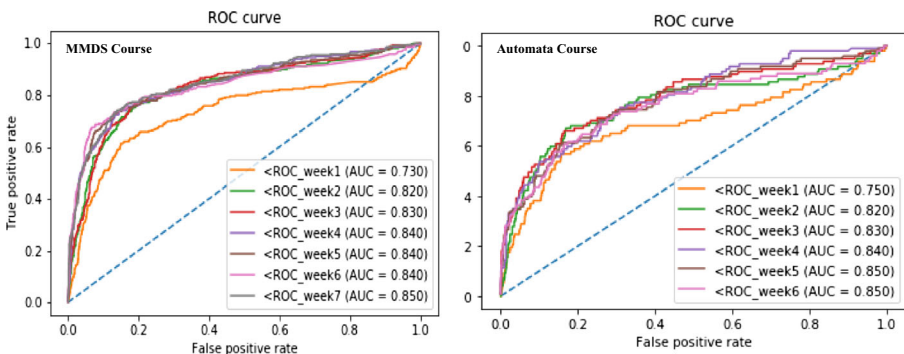

Fig. 7 AUC ROC curve for both courses

Table 5 Comparison of LSTM with Baseline Models

Model	MMDS				Automata			
	week	Accuracy	Precision	Recall	F 1 - Score	Accuracy	Precision	Recall	F 1 - Score
Long Short-Term Memory (LSTM)	1	0.86	0.90	0.86	0.88	0.82	0.96	0.80	0.87
	2	0.88	0.92	0.90	0.91	0.84	0.96	0.85	0.9
	3	0.89	0.91	0.92	0.91	0.89	0.96	0.9	0.93
	4	0.89	0.93	0.90	0.91	0.90	0.95	0.92	0.93
	5	0.88	0.94	0.89	0.91	0.91	0.95	0.91	0.93
	6	0.90	0.94	0.89	0.91	**0.93**	**0.95**	**0.95**	**0.95**
	7	**0.93**	**0.93**	**0.91**	**0.92**	–	–	–	–
Artificial Neural Networks (ANN)	1	0.76	0.76	0.85	0.8	0.8	0.8	0.8	0.8
	2	0.78	0.78	0.82	0.8	0.84	0.85	0.85	0.85
	3	0.78	0.78	0.77	0.77	0.84	0.84	0.85	0.84
	4	0.78	0.78	0.79	0.78	0.84	0.84	0.85	0.84
	5	0.79	0.79	0.79	0.79	0.85	0.85	0.86	0.85
	6	0.81	0.81	0.79	0.8	0.85	0.85	0.86	0.85
	7	0.85	0.85	0.77	0.81	–	–	–	–
Super vector machine (SVM)	1	0.75	0.87	0.77	0.82	0.77	0.81	0.83	0.82
	2	0.79	0.87	0.76	0.81	0.83	0.84	0.85	0.84
	3	0.8	0.89	0.81	0.85	0.83	0.84	0.8	0.82
	4	0.8	0.89	0.76	0.82	0.84	0.85	0.9	0.87
	5	0.8	0.89	0.81	0.85	0.84	0.84	0.83	0.83
	6	0.83	0.92	0.72	0.81	0.85	0.85	0.82	0.83
	7	0.85	0.95	0.76	0.84	–	–	–	–
Logistic Regression (LR)	1	0.75	0.75	0.77	0.76	0.8	0.8	0.81	0.8
	2	0.77	0.77	0.76	0.76	0.84	0.85	0.84	0.84
	3	0.77	0.77	0.81	0.79	0.84	0.85	0.84	0.84
	4	0.77	0.77	0.76	0.76	0.85	0.85	0.85	0.85
	5	0.78	0.78	0.81	0.79	0.85	0.86	0.84	0.85
	6	0.8	0.80	0.72	0.76	0.85	0.86	0.85	0.85
	7	0.84	0.84	0.76	0.80	–	–	–	–

week, while the accuracy of the baseline model does not exceed 86% in both courses. For each learner, the video-clickstream data values were aggregated until the current week as the baseline model input.

To avoid imbalanced data problems, several metrics were selected based on data of the confusion matrix (Table 6) such as $precision = (TP/(TP + FP))$, $recall = (TP/(TP + FN))$ and $F1 - score = (2 * precision * recall)/(precision + recall)$(Perry et al. 1955). These metrics give an interesting overview of how well a model is doing for a two-class classification problem. By reviewing the values of the metrics in the table, there was relative convergence for the precision and recall rate of models in the last weeks. The F-score value reflects the overall effectiveness of the prediction models. Overall,

the predictive proposed model (LSTM) has the ability to predict learners' performance across the course weeks better. It can be one of the underlying explanations that LSTM has its memory cell that can follow the weekly behavior of each learner. On the contrary, the baseline models cannot follow learners' performance based on time. Baseline models' input vector comprises summation of feature values video-clickstream for each learner in a weekly manner. Therefore, they learn the behavior collectively except the first-week behavior. So, the traditional techniques may incorrectly predict that a learner is not active in the initial weeks or dropping out of a course during the course duration. Overall, the proposed model outperforms better than the baseline model.

7 Discussion

This study contributes in predicting learners' performance through a deep LSTM model by their interaction with course videos and the impact on their performance. Since learners spend much of their time watching video lessons, this research focuses on examining click-event that occur within MOOC videos through learners' interaction, including playing, pausing, skipping, and stopping. This was achieved by extracting important implicit features resulting from clicking events which contributed significantly to studying the success of learners. Navigating videos of the course before finishing the course is an indicator of commitment, but a question may be raised here which is "How much is the video used by learners? ". Furthermore, instructors can need to know the aspects that are the most engaging and most broadly watched.

We assume that the videos that were most often viewed were the shortest and most associated with quizzes. Specifically, some learners had click-events indicating they had different goals for either looking for or receiving any information. Therefore, in this research, we focused on (LSTM) model to model the behavior of learners in video-clickstreams and to increase the predictability of learners' performance. The model based on sequential data (till ith week) was applied in order for the extracted features to have positively influenced prediction of learners' performance. Thus, adopting the features of extraction method to perform the model is efficient. We obtained the best accuracy of the proposed model compared to baseline models, not only in terms of accurate prediction in learners' performance but also in its capacity to identify learners at the risk of dropping out in the early weeks.

Predicting the performance of learners is a must within the context of the MOOCs in order to avoid learners' dropout of courses. Teachers should understand the reasons for dropping the course and draw appropriate measures to maintain as many learners as possible. The prediction in this study was done by adopting the extracted feature from

Table 6 Confusion Matrix

	Predict Value	
Actual Value	True At-risk (TP)	False Non-At-risk (FN)
	False At-risk (FP)	True Non-At-risk (TN)

video-clickstream data of each learner in week *ith* and making a weekly prediction. At this point, instructors can early intervene and make the appropriate decision.

Overall, this makes our methodology more distinct from the literature, not just in terms of predictive accuracy, but also in extracting video-clickstream data implicit features. This methodology can be incorporated into the MOOC platforms that instructors can use as a tool to predict the performance of their learners and allow timely intervention by instructors. In addition, the findings of the analysis can be used for useful instructional guidance. Predicting the performance of learners would allow teachers and decision-makers to follow a realistic approach for timely interventions with learners and direct them efficiently by offering relevant guidance and advice. Such prediction model will help MOOC's educators to identify likely learners to early dropout providing them with supplementary support in an educational task, and then assisting educational organizations by encouraging them to continue their learning career.

8 Conclusion

This paper presents a contribution to early predict learners at-risk of low performance, determining learners likely to reduce their interaction with course videos by a predictive model construct depended on LSTM to track and manage learners' performance. The present study was applied on two courses taken from Stanford University MOOC (For more details, please see the Dataset section), which contained learners' interaction with videos' clickstream data and their quiz scores. We showed that video-clickstream events could be employed as learning features to predict learners' performance. In implementing a predictive model, we trained the model as a week sequence classification in order to capture the unique features of each learners' learning patterns. The deep learning model (LSTM) proved to be effective in the weekly prediction of learners' performance and outperformed better accuracy than a baseline. Performance evaluation model showed an accuracy of 86% in the first weeks to 93% in the final week at the "MMDS SELF PACED" course and 82% in the first week till 93% in the final week at the "Automata Theory" course to predict learners' performance.

Early prediction of the performance of learners will help teachers and educational experts analyze and understand the learning behavior of learners through the different social interactive activities of the learners. It will also support educational stakeholders to develop approaches to promote and deliver learner therapy, achieving optimal outcomes by engagement in a certain time to attract learners.

Overall, in MOOC platforms, the clickstream behavior of each learner is gathered continuously. Therefore, the task of learners' performance prediction is observed as sequence classification. The current approach employs only learners' interaction patterns with videos to confirm their performance. For future perspectives of this field, the activity patterns of learners at risk of failure can be established, and marginally passing learners can be identified, whereas variations in the actions of distinguished learners and those at risk of failure of their courses can be identified. In the near future, we intend to develop a platform for an instructional policy environment, fusing organized and unstructured heterogeneous behavior in addition to decision-making in predicting student's outcomes based on this Learning Analytics layer.

References

Atapattu, T., & Falkner, K. (2018). Impact of Lecturer's discourse for student video interactions: Video learning analytics case study of MOOCs. *Journal of Learning Analytics, 5*(3), 182–197. https://doi.org/10.18608/jla.2018.53.12.

Baker, R. S., & Inventado, P. S. (2014). *Educational data mining and learning analytics. In Learning Analytics: From Research to Practice (pp. 61–75)*. New York: Springer. https://doi.org/10.1007/978-1-4614-3305-7_4.

Brinton, C, G., Buccapatnam, S., Chiang, M., & Poor, H, V. (2015). Mining MOOC clickstreams: On the relationship between learner behavior and performance.

Chakraborty, S., Preece, A., Alzantot, M., Xing, T., Braines, D., & Srivastava, M. (2017). Deep learning for situational understanding. *20th International Conference on Information Fusion, Fusion 2017 - Proceedings*. https://doi.org/10.23919/ICIF.2017.8009785.

Chorianopoulos, K. (2013). Collective intelligence within web video. *Human-centric Computing and Information Sciences, 3*(1), 10.

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education, 18*(6), 683–695.

Coelho, O, B., & Silveira, I. (2017). Deep learning applied to learning analytics and educational data mining: A systematic literature review. *Anais Do XXVIII Simpósio Brasileiro de Informática Na Educação (SBIE 2017), 1*, 143. https://doi.org/10.5753/cbie.sbie.2017.143

Corrigan, O., & Smeaton, A, F. (2017). A course agnostic approach to predicting student success from vle log data using recurrent neural networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10474 LNCS*, 545–548. https://doi.org/10.1007/978-3-319-66610-5_59.

De Barba, P. G., Kennedy, G. E., & Ainley, M. D. (2016). The role of students' motivation and participation in predicting performance in a MOOC. *Journal of Computer Assisted Learning, 32*(3), 218–231.

Fauvel, S., & Yu, H. (2016). *A survey on artificial intelligence and data mining for MOOCs.* http://arxiv.org/abs/1601.06862

Fei, M., & Yeung, D.-Y. (2015). Temporal models for predicting student dropout in massive open online courses. In *Data mining workshop (ICDMW), 2015 IEEE international conference on* (pp. 256–263).

Giannakos, M. N., Chorianopoulos, K., & Chrisochoides, N. (2015). Making sense of video analytics: Lessons learned from clickstream interactions, attitudes, and learning outcome in a video-assisted course. *International review of research in open and distributed learning, 16*(1), 260–283. https://doi.org/10.19173/irrodl.v16i1.1976.

Graupe, D. (2016). *Other neural networks for deep learning. In Deep Learning Neural Networks (pp. 101–109)*. WORLD SCIENTIFIC. https://doi.org/10.1142/9789813146464_0007.

Gross, E., Wshah, S., Skinner, G., & Simmons, I. (2015). A handwriting recognition system for the classroom. In *ACM international conference proceeding series, 16–20-Marc* (pp. 218–222). https://doi.org/10.1145/2723576.2723601.

Guo, P, J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. *Proceedings of the First ACM Conference on Learning@ Scale Conference*, 41–50.

Halawa, S., Greene, D., & Mitchell, J. (2014). Dropout prediction in MOOCs using learner activity features. *Proceedings of the Second European MOOC Stakeholder Summit, 37*(1), 58–65.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*(1), 29–36.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Hsin, W.-J., & Cigas, J. (2013). Short videos improve student learning in online education. *Journal of Computing Sciences in Colleges, 28*(5), 253–259.

Kim, J., Guo, P, J., Seaton, D, T., Mitros, P., Gajos, K, Z., & Miller, R, C. (2014). Understanding in-video dropouts and interaction peaks inonline lecture videos. *Proceedings of the First ACM Conference on Learning@ Scale Conference*, 31–40.

Kingma, D, P., & Ba, J. (2014). Adam: A method for stochastic optimization. ArXiv Preprint ArXiv: 1412.6980.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *ACM international conference proceeding series* (pp. 170–179). https://doi.org/10.1145/2460296.2460330.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539.

Li, N., Kidzinski, L., Jermann, P., & Dillenbourg, P. (2015a). How do in-video interactions reflect perceived video difficulty?

Li, N., Kidziński, Ł., Jermann, P., & Dillenbourg, P. (2015b). *MOOC video interaction patterns: What do they tell us? In Design for teaching and learning in a networked world (pp. 197–210).* Springer.

Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education, 103*, 1–15.

Mubarak, A. A., Cao, H., & Zhang, W. (2020). Prediction of students' early dropout based on their interaction logs in online learning environment. *Interactive Learning Environments.*, 1–20. https://doi.org/10.1080/10494820.2020.1727529.

Okubo, F., Yamashita, T., Shimada, A., & Konomi, S. (2017). Students' performance prediction using data of multiple courses by recurrent neural network. In *Proceedings of the 25th international conference on computers in education, ICCE 2017 - Main conference proceedings* (pp. 439–444) https://kyushu-u.pure.elsevier.com/en/publications/students-performance-prediction-using-data-of-multiple-courses-by.

Olah, C. (2015). Understanding LSTM Networks [Blog]. Web Page, 1–13. https://doi.org/10.1007/s13398-014-0173-7.2.

Palmer, S. (2013). Modelling engineering student academic performance using academic analytics. *International Journal of Engineering Education, 29*(1), 132–138 http://dro.deakin.edu.au/view/DU:30051021.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*(Oct), 2825–2830.

Perry, J. W., Kent, A., & Berry, M. M. (1955). Machine literature searching x. machine language; factors underlying its design and development. *American Documentation, 6*(4), 242–254.

Risko, E. F., Foulsham, T., Dawson, S., & Kingstone, A. (2012). The collaborative lecture annotation system (CLAS): A new TOOL for distributed learning. *IEEE Transactions on Learning Technologies, 6*(1), 4–13.

Sak, H, H., Senior, A., & Google, B. (2014). *Long short-term memory recurrent neural network architectures for large scale acoustic modeling.* https://research.google/pubs/pub43905.pdf

Sak, H, H., Senior, A., & Google, B. (n.d.). Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In *ieeexplore.ieee.org*. Retrieved April 28, 2020, from https://ieeexplore.ieee.org/abstract/document/7178816/

Shaw, R., & Davis, M. (2005). Toward emergent representations for video. In *Proceedings of the 13th annual ACM international conference on multimedia* (pp. 431–434).

Sinha, T., Jermann, P., Li, N., & Dillenbourg, P. (2014a). Your click decides your fate: Inferring information processing and attrition behavior from mooc video clickstream interactions. ArXiv Preprint ArXiv: 1407.7131.

Sinha, T., Li, N., Jermann, P., & Dillenbourg, P. (2014b). Capturing" attrition intensifying" structural traits from didactic interaction sequences of MOOC learners. ArXiv Preprint ArXiv:1409.5887.

Smoliar, S. W., & Zhang, H. (1994). Content based video indexing and retrieval. *IEEE Multimedia, 1*(2), 62–72.

Stanford, U. (2017). Center for advanced research through online learning | Institute for Research in the social sciences.]. In *Available: https://iriss.stanford.edu/carol* (center for).

Tang, S., Peterson, J, C., & Pardos, Z, A. (2016). Deep neural networks and how they apply to sequential education data. *L@S 2016 - Proceedings of the 3rd 2016 ACM Conference on Learning at Scale*, 321–324. https://doi.org/10.1145/2876034.2893444.

University, S. (2017). CAROL learner data documentation. In *Available: https://datastage.stanford.edu/* (center for).

Waheed, H., Ali, M., Hassan, S.-U., Ventura, S., & Herrera, F. (2019). Virtual learning environment to predict withdrawal by leveraging deep learning. *Article in International Journal of Intelligent Systems, 34*(8), 1935–1952. https://doi.org/10.1002/int.22129.

Wang, Y., & Baker, R. (2015). Content or platform: Why do students complete MOOCs. *MERLOT Journal of Online Learning and Teaching, 11*(1), 17–30.

Xu, B., & Yang, D. (2016). Motivation classification and grade prediction for MOOCs learners. *Computational Intelligence and Neuroscience, 2016*, 1–7. https://doi.org/10.1155/2016/2174613.

Yew, J., Shamma, D, A., & Churchill, E, F. (2011). Knowing funny: Genre perception and categorization in social video sharing. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 297–306.

Yu, C. H., Wu, J., & Liu, A. C. (2019). Predicting learning outcomes with MOOC clickstreams. *Education Sciences, 9*(2). https://doi.org/10.3390/educsci9020104.

Zhang, G., Eddy Patuwo, B., Hu, Y., & M. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14*(1), 35–62. https://doi.org/10.1016/S0169-2070(97)00044-7.

Zhang, D., Zhou, L., Briggs, R. O., & Nunamaker Jr., J. F. (2006). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information & Management, 43*(1), 15–27.

Zhang, Q., Yang, L, T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. In *Information Fusion* (Vol. 42, pp. 146–157). https://doi.org/10.1016/j.inffus.2017.10.006.

Publisher's note   Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.